CVPR
JUNE 3-7, 2026

DENVER
COLORADO

IEEE/CVF Conference on
Computer Vision and Pattern Recognition

MAIN CONFERENCE PROGRAM

ULTIMATE SPONSORS

build

gm

Mohamed bin Zayed University of Artificial Intelligence | Institute of Foundation Models

Lambda

Meta

Tencent

TESLA

Uber

WAYMO

PLATINUM SPONSORS

Adobe

Alibaba

amazon

ANT GROUP

ByteDance

MINIMAX

ORACLE

Qualcomm

SONY

TORC

GOLD SPONSORS

Baidu 百度

baseten

contra LABS

Crusoe

DYNA

ENCORD

Luma

美团

NVIDIA

TURING

Matterport

UTPAI

Weights & Biases by CoreWeave

SILVER SPONSORS

Appen

Applied Intuition

Aurora

BagelLABS

GMI

HPC-AI.COM

Ideogram

kitware

LINKERBOT

Motional

nexdata

sudo

VOXEL51

waabi

ZOOX

Welcome to the 2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Denver, Colorado, the "Mile High City," located at an elevation of 5280 feet (1 mile)! CVPR is the premier and flagship annual meeting of IEEE/CVF and PAMI-TC, where researchers in our community present their latest advances in computer vision, pattern recognition, machine learning, robotics, and artificial intelligence, both in theory and practice. CVPR 2026 is primarily an in-person conference, but for those unable to attend in person, we are pleased to offer a virtual component with access to conference papers, posters, videos, and talks.

CVPR 2026 received 16,092 valid paper submissions that underwent the review process, a 24% increase from CVPR 2025. The review process was managed by the 6 Program Co-Chairs, 41 Senior Area Chairs, and 909 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 25,149 reviewers. As in prior years, after receiving these initial reviews, authors had the opportunity to submit a rebuttal to the reviews. The process concluded with discussions among reviewers and ACs, finalizing of reviews, and ACs working in triplets to make final accept/reject decisions for each paper. At the end of this process, 4,089 papers were accepted, for a 25.4% overall acceptance rate. In keeping with the CVPR tradition, the PCs did not pre-define any target acceptance rate or number of papers to be accepted; the resulting acceptance rate reflects the community consensus and is consistent with past CVPRs.

All the 4,072 accepted papers were invited to present posters at CVPR. In addition, 141 (3.4%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 578 (14.1%) papers were selected by ACs together with Senior ACs to be "highlights" due to their high quality and potential impact. This year, the oral presentations will be organized into four tracks. The highlights are also flagged with a special annotation in the program. 74 papers were nominated to be the best paper award candidates, from which a committee convened by the PCs selected the award winners to be announced during the conference.

Due to the dramatic increase in submissions and a commitment to improving the quality and accountability of the review process, the PCs adopted several measures introduced in recent CVPR review processes. These measures include imposing a limit on the number of submissions per author, requiring authors to participate in the review process when deemed qualified by the PCs, and rejecting papers submitted by highly irresponsible reviewers. In addition, we also adopted the outstanding reviewer recognition, highlighting 1,545 reviewers (6.1%) from our large pool of 25,149 reviewers. Papers authored by outstanding reviewers are flagged with a special annotation in the program. We further introduced the outstanding AC recognition, highlighting 303 ACs (33.3%) from our pool of 909 ACs for their timely and outstanding service.

New this year, CVPR 2026 introduces an experimental Compute Reporting Initiative, inviting authors to disclose the computational resources used in their work. This initiative reflects our community's growing interest in understanding the computational landscape of CV/AI research, from resource requirements to efficiency innovations. Compute reports were not visible to reviewers and did not influence acceptance decisions. We hope this data will enable community-wide benchmarking and inspire future work on computational efficiency and accessibility. Awards will be presented recognizing exceptional efficiency and transparency. Highlights from this initiative will be presented during the conference, and a full report will be disseminated to the community thereafter.

We would like to thank everyone involved in making CVPR 2026 a success. This includes the organizing committee, Senior Area Chairs, Area Chairs, reviewers, authors, demo session participants, contributing artists, and donors and exhibitors. It is exciting to be part of a team of people around the world united in a common goal. The scale and quality of this year's program is a testament to the dedication of this global community.

We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, Yoshitomo Matsubara for his amazing work as technical chair, Lee Campbell and the Event Hosts team for their work on the website and virtual platform, Mike Weil and Hall Erickson for handling sponsorships and the exhibition, and Luba Elliot as our AI Art Coordinator. Finally, we thank all of you for attending CVPR 2026 and making it one of the top venues for computer vision research in the world. We hope that you also have some time to explore Denver during the conference.

Enjoy CVPR 2026. We look forward to meeting you in person!

CVPR Program Chairs
Angela Dai *(TU Munich)*
Adriana Kovashka *(University of Pittsburgh)*
Chen Change Loy *(Nanyang Technological University),*
Vladimir Pavlovic *(Rutgers University),*
Alex Schwing *(University of Illinois Urbana-Champaign),*
Shaoting Zhang *(Shanghai Jiao Tong University)*

CVPR 2026 General Chairs
Chandra Kambhamettu *(University of South Florida)*
Dimitris N. Metaxas *(Rutgers University)*

CVPR 2026 QR CODES

CVPR 2026 Virtual Platform

Access schedules, papers, workshops, tutorials, etc

Exhibit Hall Floorplan

Access location of exhibitors on the main floor

CVPR 2026 ORGANIZING COMMITTEE

Honorary General Chairs
Takeo Kanade *(CMU)*
Shimon Ullman *(Weizmann Institute of Science)*

General Chairs
Chandra Kambhamettu *(University of South Florida)*
Dimitris N. Metaxas *(Rutgers University)*

Program Chairs
Angela Dai *(Technical University of Munich)*
Adriana Kovashka *(University of Pittsburgh)*
Chen Change Loy *(Nanyang Technological University)*
Vladimir Pavlovic *(Rutgers University)*
Alexander G. Schwing *(University of Illinois Urbana-Champaign)*
Shaoting Zhang *(Shanghai Jiao Tong University)*

Advisor to the Program Committee
David Forsyth *(University of Illinois Urbana-Champaign)*

Workshop Chairs
Erik Blasch *(AFRL)*
Mei Chen *(Dolby)*
Brian Clipp *(Motional)*
Sharon X. Huang *(Penn State University)*
Andreas Savakis *(Rochester Institute of Technology)*
Humphrey Shi *(Georgia Tech)*

Tutorial Chairs
Boqing Gong *(Boston University / Google)*
Hongsheng Li *(The Chinese University of Hong Kong)*
Paul Schrader *(Air Force Research Laboratory)*
Vitomir Struc *(University of Ljubljana)*

Demonstration Chairs
Sathyanarayanan N. Aakur *(Auburn University)*
Shu Kong *(University of Macau)*

Finance Chair
Walter Scheirer *(Notre Dame)*

Publicity Chairs
Deblina Bhattacharjee *(University of Bath)*
Kosta Derpanis *(York University)*
Antonino Furnari *(University of Catania)*
Zhengzhong Tu *(Texas A&M University, UT Austin)*
Yael Vinker *(MIT)*

Broadening Participation Chairs
Naresh Cuntoor *(BlueHalo Labs)*
Christopher Funk *(Kitware)*
Deepti Ghadiyaram *(Boston University)*
Michael King *(Florida Tech)*
Roni Sengupta *(UNC Chapel Hill)*

Senior PAMI-TC Ombuds
David Forsyth *(University of Illinois Urbana-Champaign)*
Linda Shapiro *(University of Washington)*

AI Art Curator
Luba Elliott *(Independent Curator)*

Technical Chair
Yoshitomo Matsubara *(Yahoo!)*

Doctoral Consortium Chairs
Paola Cascante-Bonilla *(Stonybrook University)*
Abby Stylianou *(St. Louis University)*

Local Chair
Brendan Klare *(ROC)*

Web Developer
Lee Campbell *(Eventhosts)*

Conference Producer
Nicole Finn *(c to c events)*

Compute Reporting Committee Chairs
Sourav Bhattacharya *(NatWest Group)*
Christopher Funk *(Kitware, Inc.)*
Farhad Pourkamali *(University of Colorado Denver)*
Roy Schwartz *(Hebrew University of Jerusalem)*
Emma Strubell *(Carnegie Mellon University)*
Sasha Luccioni (Hugging Face) - Advisor

Award Committee
Derek Hoiem *(Chair, University of Illinois Urbana-Champaign)*
Kristen Grauman *(UT Austin)*
Ko Nishino *(Kyoto University)*
Deqing Sun *(Google Deepmind)*
Tinne Tuytelaars *(KU Leuven)*
Angela Yao *(National University of Singapore)*
Jun-Yan Zhu *(Carnegie Mellon University)*

Oral Coaching Committee
Srinivasa Narasimhan *(Chair, Carnegie Mellon University)*
Dima Damen *(University of Bristol)*
Fernando De la Torre *(Carnegie Mellon University)*
Sven Dickinson *(University of Toronto)*
Alexei Efros *(UC Berkeley)*
David Forsyth *(University of Illinois Urbana-Champaign)*
Bill Freeman *(MIT)*
Ioannis Gkioulekas *(Carnegie Mellon University)*
Mohit Gupta *(University of Wisconsin, Madison)*
Kiriakos Kutulakos *(University of Toronto)*
Jean-Francois Lalonde *(Université Laval)*
Lana Lazebnik *(University of Illinois Urbana-Champaign)*
Ko Nishino *(Kyoto University)*
Devi Parikh *(Yutori)*
Deva Ramanan *(Carnegie Mellon University)*
Aswin Sankaranarayanan *(Carnegie Mellon University)*
Steve Seitz *(University of Washington, Google)*
Yaser Sheikh *(Sooth Labs)*
Ashok Veeraraghavan *(Rice University)*
Yair Weiss *(Hebrew University of Jerusalem)*
Richard Zhang *(Adobe Research)*
Todd Zickler *(Harvard University)*

CVPR 2026 SENIOR AREA CHAIRS

Andrea Tagliasacchi
Angela Yao
Anthony Hoogs
Bernt Schiele
Boqing Gong
David J. Crandall
Deqing Sun
Derek Hoiem
Dima Damen
Dimitris Samaras
Gang Hua
Georgia Gkioxari
Gim Hee Lee
Hongdong Li
Hongsheng Li
Ian Reid
Jianfei Cai
Jiaying Liu
Juergen Gall
Jun-Yan Zhu
Kate Saenko
Ko Nishino
Kristen Grauman
Lei Zhang
Lourdes Agapito
Maja Pantic
Manolis Savva
Marc Niethammer
Marcus Rohrbach
Matthias Nießner
Ming-Hsuan Yang
Nathan Jacobs
Pablo Arbelaez
Seon Joo Kim
Srinivasa Narasimhan
Stefan Roth
Svetlana Lazebnik
Tinne Tuytelaars
Vittorio Murino
Yoichi Sato
Yong Jae Lee

Aayush Bansal
Abby Stylianou
Abhinav Shrivastava
Adam Czajka
Adam Kortylewski
Adams Wai-Kin Kong
Adrian Barbu
Adrien Bartoli
Ahmet Iscen
Ailing Zeng
Aishwarya Agrawal
Ajay Kumar
Akihiro Sugimoto
Akshay Gadi Patil
Aleix M Martinez
Aleksander Holynski
Alex Colburn
Alex Wong
Ali Farhadi
Alimoor Reza
Alina Kuznetsova
Alireza Fathi
Aljosa Osep
Ameesh Makadia
Amir Zamir
Amit Roy-Chowdhury
Anand Bhattad
Anand Mishra
Anders Heyden
Andre Araujo
Andrea Fusiello
Andrei Bursuc
Andres Bruhn
Angel X Chang
Anh Tuan Tran
Anoop Cherian
Anpei Chen
Antoni B. Chan
Antonino Furnari
Anurag Arnab
Anurag Mittal
Anwesa Choudhuri
Anyi Rao
Arif Mahmood
Arun Mallya
Asako Kanezaki
Aswin C. Sankaranarayanan
Ayellet Tal
Babak Taati
Baoyuan Wu
Bastian Leibe
Basura Fernando
Benjamin Busam
Bin Fan
Bing Su
Binh-Son Hua
Björn Ommer
Bjorn Stenger
Bo Dai
Bohan Zhuang
Bolei Zhou
Boxin Shi
Boyi Li
Brais Martinez
Brian Clipp
Brian L. Price
Bruce Allen Maxwell
Bryan A. Plummer
Bryan Morse
Bryan Russell
Bumsub Ham
Byeongjoo Ahn
C.V. Jawahar
Carl Olsson
Cees G. M. Snoek
Chang Xu
Changqing Zou
Chao Chen
Chao Fan
Chaowei Xiao
Charles Herrmann
Chen Chen
Chen Feng
Chen Sun
Chen Wang
Cheng-Lin Liu
Chengjiang Long
Chenliang Xu
Chenyang Si
Chenyu You
Chetan Arora
Chi Zhang
Chi-Keung Tang
Chong Luo
Chongyi Li
Chris Thomas
Christian Richardt
Christian Rupprecht
Christian Wolf
Christoph Feichtenhofer
Christopher Funk
Chu-song Chen
Chuang Gan
Chuanmin Jia
Chuhang Zou
Chun-Guang Li
Chun-Le Guo
Chun-Yi Lee
Chunhua Shen
Cigdem Beyan
Cihang Xie
Clement Mallet
Concetto Spampinato
Constantin Pape
Cordelia Schmid
Cristian Canton Ferrer
Cuiling Lan
Dahua Lin
Dahun Kim
Dan Casas
Dan Guo
Dan Xu
Daniel Duckworth
Daniel Zoran
Danna Gurari
David B. Lindell
David Fouhey
David Jacobs
David Picard
Davide Modolo
Davide Moltisanti
Dayan Wu
Deepti Ghadiyaram
Deng-Ping Fan
Despoina Paschalidou
Devis Tuia
Di Huang
Diane Larlus
Dimitrios Tzionas
Dimosthenis Karatzas
Dingwen Zhang
Dong Chen
Dong Gong
Dor Verbin
Dorin Comaniciu
Doyup Lee
Du Tran
Edmond Boyer
Ehsan Adeli
Ehsan Elhamifar
Elahe Arani
Eli Shechtman
Elisa Ricci
Emily Morgan Hand
Eric Brachmann
Erik B. Sudderth
Eshed Ohn-Bar
Esube Bekele
Eunbyung Park
Evan Shelhamer
Evangelos Kalogerakis
Fabio Cuzzolin
Fabio Galasso
Fabio Poiesi
Fabio Tosi
Fangzhou Mu
Fartash Faghri
Fayao Liu
Federica Bogo
Federico Tombari
Feng Liu
Feng Liu
Feng Lu
Florian Bernard
Francis Engelmann
Frederic Jurie
Friedrich Fraundorfer
Fumio Okura
Gabriel Brostow
Gal Chechik
Gang YU
Gao Huang
Gedas Bertasius
George Kopanas
Georgios Pavlakos
Georgios Tzimiropoulos
Gerard Pons-Moll
Gianni Franchi
Giorgos Tolias
Giovanni Maria Farinella
Gordon Wetzstein
Greg Mori
Greg Shakhnarovich
Gregory Slabaugh
Guanbin Li
Guangcong Wang
Guansong Pang
Guanying Chen
Guillaume Couairon
Guillermo Gallego
Gunhee Kim
Guofeng Zhang
Guosheng Lin
Guotai Wang
Guoyu Lu
Gustavo Carneiro
Guy Gilboa
Gyeongsik Moon
Hadar Averbuch-Elor
Hadi Pouransari
Hae-Gon Jeon
Haibin Ling
Hailin Jin
Hajime Nagahara
Hakan Bilen
Hamed Pirsiavash
Hamid Rezatofighi
Han Hu
Han-Jia Ye
Hanbyul Joo
Hanwen Jiang
Hao Chen
Hao Dong
Hao Richard Zhang
Hao Wang
Haoliang Li
Haotong Qin
Haoxiang Li
Haoye Dong
Harsh Agrawal
Hazel Doughty
He Zhang
Hedvig Kjellstrom
Henghui Ding
Hengshuang Zhao
Hilde Kuehne
Hirokatsu Kataoka
Hiroshi Kawasaki
Hisham Cholakkal
Holger Caesar
Hong Chang
Hong Xuan
Hong Zhang
Hong-Han Shuai
Hongbin Zha
Hongyang Li
Hossein Azizpour
Hossein Rahmani
Hu Han
Huaizu Jiang
Huan Fu
Huan Liu
Huan Wang
Hugues Talbot
Hui Ji
Huijuan Xu
Hung-Yu Tseng
Hwann-Tzong Chen
Hyun Soo Park
Hyunjung Shim
Hyunwoo J. Kim
I. Zeki Yalniz
Iasonas Kokkinos
Idan Schwartz
Ilke Demir
In Kyu Park
Ioana Croitoru
Ioannis Patras
Ipek Oguz
Iro Armeni
Ismini Lourentzou
Ivan Laptev
Ivan Skorokhodov
Jacob C Walker
Jaesik Park
Jakob Verbeek
James Matthew Rehg
Jan Eric Lenssen
Jangwon Lee
Janne Heikkila
Jason J Corso
Jason Ren
Jason Y. Zhang
Jasper R. R. Uijlings
Javier Civera
Javier Vazquez-Corral
Jean Ponce
Jean-Charles Bazin
Jennifer J. Sun
Jeong Joon Park
Jia Deng
Jia-Bin Huang
Jiachen Li
Jiahuan Zhou
Jiajun Wu
Jiaming Zhang
Jian Liang
Jian Sun
Jiang Wang
Jiangmiao Pang
Jiangqun Ni
Jiangxin Dong
Jiankang Deng
Jianpeng Zhang
Jianwen Xie
Jianxin Wu
Jiaqi Wang
Jiasen Lu
Jiawen Chen
Jiaya Jia
Jiayi Ji
Jie Song
Jifeng Dai
Jimmy Ren
Jing Zhang
Jingdong Wang
Jingwan Lu
Jingya Wang
Jinshan Pan
Jinwei Ye
Jiri Matas
Jiwen Lu
Joachim Denzler
Joao P. Barreto
John Collomosse
Jon Sporring
Jonathan T. Barron
Jong Chul Ye
Jonghyun Choi
Jongwoo Lim
Joon-Young Lee
Joost van de Weijer
Joseph Tighe
Josephine Sullivan
Juan Carlos Niebles
Judy Hoffman
Jufeng Yang
Julian Straub
Jun Gao
Jun Liu
Jun-Cheng Chen
Junbin Xiao
Junchi Yan
Junfeng He
Jungseock Joo
Junhwa Hur
Junseok Kwon
Junsong Yuan
Junyi Zhu
Junzhou Huang
Justus Thies
Juyong Zhang
Jyh-Jing Hwang
Kai Chen
Kai Han
Kai Han
Kaicheng Yu
Kaifeng Chen
Kaihao Zhang
Kaiyang Zhou
Kaleem Siddiqi
Kalle Åström
Kang Liao
Karsten Roth
Katherine Bouman

Katja Schwarz
Kaustav Kundu
Ke Liang
Kede Ma
Keita Takahashi
Kelvin C.K. Chan
Kenneth Marino
Kent Fujiwara
Kevin J Liang
Kevin Smith
Kfir Aberman
Konstantinos G. Derpanis
Kostas Daniilidis
Kota Yamaguchi
Kris Kitani
Krishna Kumar Singh
Krystian Mikolajczyk
Kuk-Jin Yoon
Kushal Kafle
Kwan-Yee K. Wong
Kwang In Kim
Kwang Moo Yi
Kyle Genova
Kyungdon Joo
Lamberto Ballan
Laszlo A. Jeni
Laura Leal-Taixé
Le Lu
Le Wang
Lei Li
Lei Wang
Lei Zhu
Leonid Sigal
Lequan Yu
Li Erran Li
Li Shen
Li Yi
Li Zhang
Liang Zheng
Liang-Chieh Chen
Liangqiong Qu
Liangyan Gui
Lianli Gao
Lijun Yu
Limin Wang
Lin Gao
Lin Gu
Linchao Bao
Linchao Zhu
Lingjie Liu
Lingxi Xie
Linjie Yang
Linlin Shen
Liwei Wang
Loic Landrieu
Long Chen
Long Zhao
Longyin Wen
Lorenzo Torresani
Lu Jiang
Lu Qi
Lu Sheng
Lu Yuan
Luc Van Gool
Luigi Di Stefano
Luisa Verdoliva
Luping Zhou
Mahdi S. Hosseini
Makarand Tapaswi
Mandi Luo
Mang Ye
Manmohan Chandraker
Marc Habermann
Marc Pollefeys
Marcella Cornia
Marcello Pelillo
Marcus A Brubaker
Margret Keuper
Maria A Zuluaga
Maria Vakalopoulou
Marius Leordeanu
Mark Yatskar
Martin R. Oswald
Martin Weinmann
Massimiliano Mancini
Matej Kristan
Matheus Gadelha
Mathieu Aubry
Mathieu Salzmann
Matteo Poggi
Matthew B. Blaschko
Matthew Brown
Matthew Joseph Leotta
Matthew O'Toole
Matthew Turk
Matthieu Cord
Mengyu Wang
Miaomiao Liu
Michael Felsberg
Michaël Gharbi
Michael Maire
Michael Moeller
Michael Niemeyer
Michael S Ryoo
Michael S. Brown
Michael Wray
Michele Merler
Michele Nappi
Miguel Ángel Bautista
Mihalis Nicolaou
Mike Zheng Shou
Min H. Kim
Ming Yang
Ming-Ching Chang
Minghua Liu
Mingxing Tan
Mingyu Ding
Mingze Xu
Minh Ha Quang
Minh N. Do
Minjie Cai
Minsu Cho
Minye Wu
Mohamed Elhoseiny
Mohit Gupta
Mohsen Ali
Muhammad Haris Khan
Nanxuan Zhao
Nassir Navab
Naveed Akhtar
Neel Dey
Nico Lang
Nicolas Thome
Nicoletta Noceti
Nicu Sebe
Niki Martinel
Nikos Kolotouros
Ning Yu
Noah Snavely
Noha Radwan
Norimichi Ukita
Norman Müller
Oana-Maria Camburu
Octavia Camps
Ohad Fried
Oisin Mac Aodha
Olga Veksler
Olivia Wiles
Oncel Tuzel
Ondrej Chum
Or Litany
Or Patashnik
Orazio Gallo
Orchid Majumder
Oren Freifeld
Pan He
Paola Cascante-Bonilla
Paolo Rota
Pascal Fua
Pascal Mettes
Patrick Perez
Paul Albert
Paul Hongsuck Seo
Pedro O. Pinheiro
Peidong Liu
Peirong Liu
Peng Wang
Peter Hedman
Peter Vincent Gehler
Peter Wonka
Petia Radeva
Philippos Mordohai
Pinar Yanardag
Ping Hu
Ping Luo
Ping Tan
Piotr Koniusz
Pirazh Khorramshahi
Prateek Prasanna
Pratul P. Srinivasan
Qi Dai
Qi Dou
Qi Mao
Qi Shan
Qi Wu
Qi Ye
Qi Yu
Qian Yu
Qiang Ji
Qianqian Wang
Qianru Sun
Qibin Hou
Qifeng Chen
Qilong Wang
Qin Jin
Qing Guo
Qing Wang
Qingshan Liu
Qingyi Tao
Qiuhong Ke
Qixing Huang
Radu Timofte
Rahaf Aljundi
Raja Giryes
Ram Nevatia
Ran He
Ranjay Krishna
Raoul de Charette
Raymond A. Yeh
Rei Kawakami
Renaud Marlet
Rene Vidal
Renjie Liao
Richard Zhang
Risheng Liu
Robby T. Tan
Robert Pless
Rogerio Feris
Rohit Girdhar
Romain Brégier
Ronen Basri
Rongrong Ji
Roni Sengupta
Roozbeh Mottaghi
Rui Shao
Ruiping Wang
Ruiqi Gao
Ruohan Gao
Ryan Farrell
Ryoma Bise
Ryusuke Sagawa
S Kevin Zhou
Saeed Anwar
Sagie Benaim
Sai-Kit Yeung
Saining Xie
Saket Anand
Salman Asif
Samuele Salti
Sangdoo Yun
Sanjeev Jagannatha Koppal
Sareh Rowlands
Sathyanarayanan N. Aakur
Satoshi Tsutsui
Sayna Ebrahimi
Scott McCloskey
Scott Workman
Sebastiano Vascon
Sen Lin
Seong Jae Hwang
Ser-Nam Lim
Serena Yeung-Levy
Sergey Tulyakov
Sergio Escalera
Seunghoon Hong
Seungryong Kim
Seungyong Lee
Sewoong Oh
Shai Bagon
Shalini De Mello
Shang-Hong Lai
Shangchen Zhou
Shangzhe Wu
Shanshan Zhang
Shaodi You
Shayok Chakraborty
Shengfeng He
Shenghua Gao
Shenlong Wang
Shiguang Shan
Shizhe Chen
Shohei Nobuhara
Shuai Yang
Shubham Tulsiani
Shuting He
Sicheng Zhao
Sifei Liu
Silvia Cascianelli
Simon Niklaus
Simone Calderara
Simone Melzi
Simone Schaub-Meyer
Siyu Tang
Siyu Zhu
Slobodan Ilic
Song Bai
Songyou Peng
Spyros Gidaris
Srikrishna Karanam
Srinath Sridhar
Stamatios Georgoulis
Stavros Petridis
Stefan Lee
Stefano Mattoccia
Stella X. Yu
Stephan Alaniz
Stephan Richter
Stéphane Herbin
Stéphane Lathuilière
Stephen J. Guy
Stephen James
Stephen Lin
Stergios Christodoulidis
Stratis Gavves
Subhankar Roy
Subhashini Venugopalan
Subhransu Maji
Suha Kwak
Sunil Hadap
Supasorn Suwajanakorn
Tae Hyun Kim
Taesup Moon
Takayuki Okatani
Tal Hassner
Tammy Riklin Raviv
Tanmay Gupta
Tat-Jen Cham
Tat-Jun Chin
Tatiana Tommasi
Tatsuya Harada
Tengda Han
Thanh-Toan Do
Theo Gevers
Thibaut Durand
Thiemo Alldieck
Thomas Eboli
Thomas Mensink
Tianfan Xue
Tianfu Wu
Tianzhu Zhang
Ting Yao
Ting Zhang
Tolga Birdal
Tomas Hodan
Tomas Jakab
Tomas Pajdla
Tony C. W. MOK
Tony Tung
Torsten Sattler
Toshihiko Yamasaki
Trevor Darrell
Tushar Nagarajan
Ulugbek S. Kamilov
Umar Iqbal
Unnat Jain
Utkarsh Mall
Varun Jampani
Vasileios Belagiannis
Venkatesh Babu Radhakrishnan
Vicente Ordonez
Vicky Kalogeiton
Victor Adrian Prisacariu
Victor Fragoso
Victor Lempitsky
Vikram V. Ramaswamy
Viktor Larsson
Viktoriia Sharmanska
Vinay P. Namboodiri
Vincent Lepetit

CVPR 2026 AREA CHAIRS (continued)

Vineeth N. Balasubramanian
Viraj Shah
Vishal M. Patel
Vladislav Golyanik
Wang Yifan
Wangmeng Zuo
Wanli Ouyang
Waqas Sultani
Wayne Wu
Wei Ji
Wei Ji
Wei Li
Wei Shen
Wei Tang
Wei Yang
Wei-Chao Chen
Wei-Chih Hung
Wei-Chiu Ma
Wei-Shi Zheng
Weibin Wu
Weidong Cai
Weihong Deng
Wen Li
Wenhan Yang
Wenqi Ren
Wenwei Zhang
William A P Smith
William Robson Schwartz
Xavier Giró-i-Nieto
Xi Peng
Xi Yin
Xiang Bai
Xiangtai Li
Xiangyu Xu
Xiangyu Yue
Xiangyu Zhang
Xiao Luo
Xiaochun Cao
Xiaofan Zhang
Xiaoguang Han
Xiaojuan Qi
Xiaolin Huang
Xiaoling Hu
Xiaolong Wang
Xiaoming Liu
Xiaoming Zhao
Xiaoqian Wang
Xiaoshuai Zhang
Xiaoyi Dong
Xiaoyu Wang
Xiatian Zhu
Xihui Liu
Xin Yu
Xinchao Wang
Xingang Pan
Xinggang Wang
Xinlei Chen
Xinlong Wang
Xintao Wang
Xintong Han
Xinyu Zhang
Xiu-Shen Wei
Xu Yang
Xuefei Ning
Xuequan Lu
Xueting Li
Xuming He
Xun Xu
Yadan Luo
Yadong MU
Yajing Zheng
Yale Song
Yan Huang
Yan Huang
Yan Wang
Yan Wang
Yan Yan
Yan Zhang
Yanchao Yang
Yandong Wen
Yang Bai
Yang Wang
Yannick Strümpler
Yannis Avrithis
Yannis Kalantidis
Yao Qin
Yao Yao
Yaoyao Liu
Yapeng Tian
Yasamin Jafarian
Yaser Sheikh
Yasutaka Furukawa
Yasuyuki Matsushita
Yebin Liu
Yedid Hoshen
Yen-Yu Lin
Yezhou Yang
Yi Fang
Yi Wu
Yi Yang
Yi-Hsuan Tsai
Yi-Ting Chen
Yi-Zhe Song
Yibing Song
Yicheng Wu
Yifei Huang
Yihua Zhang
Yijun Li
Yiming Li
Yiming Wang
Yin Cui
Yin Li
Ying Fu
Ying Wu
Yinghao Xu
Yinghuan Shi
Yingli Tian
Yingying Li
Yinqiang Zheng
Yisen Wang
Yiwei Wang
Yixin Zhu
Yiyi Liao
Yogesh S Rawat
Yongtao Wang
Yosi Keller
Yu Cheng
Yu Kong
Yu Li
Yu Wu
Yu-Chiang Frank Wang
Yu-Chuan Su
Yu-Xiong Wang
Yuankai Huo
Yuankai Qi
Yuchao Dai
Yuexin Ma
Yuhang Zang
Yujiao Shi
Yujun Cai
Yuki M Asano
Yulan Guo
Yuliang Xiu
Yulun Zhang
Yumin Suh
Yung-Yu Chuang
Yunhui Guo
Yunzhong Hou
Yunzhu Li
Yuqian Zhou
Yuri Boykov
Yusuke Sugano
Yuyin Zhou
Zexiang Xu
Zhangyang Wang
Zhaopeng Cui
Zhaowei Cai
Zhaoxiang Zhang
Zhe Gan
Zhe Lin
Zhenan Sun
Zheng Xu
Zheng Zhang
Zhengqi Li
Zhengzhong Tu
Zhiding Yu
Zhijie Deng
Zhirong Wu
Zhixiang Wang
Zhizhen Zhao
Zhizhong Han
Zhonghua Wu
Zhongpai Gao
Zhuang Liu
Zhun Zhong
Zhuotao Tian
Ziad Al-Halah
Ziwei Liu
Zixiang Zhao
Zongsheng Yue
Zongwei Zhou
Zongxin Yang
Zsolt Kira
Zuxuan Wu
Zuzana Kukelova

CVPR 2026 OUTSTANDING AREA CHAIRS

Abhinav Shrivastava
Adam Czajka
Adams Wai-Kin Kong
Adrien Bartoli
Aishwarya Agrawal
Alex Colburn
Alex Wong
Andrea Fusiello
Andres Bruhn
Angel X Chang
Anh Tuan Tran
Antonino Furnari
Anwesa Choudhuri
Asako Kanezaki
Baoyuan Wu
Basura Fernando
Benjamin Busam
Bing Su
Binh-Son Hua
Bjorn Stenger
Bohan Zhuang
Brian L. Price
Bryan A. Plummer
Bumsub Ham
Carl Olsson
Cees G. M. Snoek
Chang Xu
Chaowei Xiao
Charles Herrmann
Chen Wang
Cheng-Lin Liu
Chenliang Xu
Chenyang Si
Chenyu You
Chi Zhang
Chi-Keung Tang
Chongyi Li
Chris Thomas
Christian Richardt
Chuanmin Jia
Chuhang Zou
Chun-Guang Li
Chun-Le Guo
Chun-Yi Lee
Cigdem Beyan
Constantin Pape
Cuiling Lan
Dan Casas
Dan Guo
Danna Gurari
David B. Lindell
David Picard
Davide Moltisanti
Deng-Ping Fan
Devis Tuia
Diane Larlus
Dimitrios Tzionas
Dong Gong
Dor Verbin
Ehsan Elhamifar
Emily Morgan Hand
Eric Brachmann
Fabio Cuzzolin
Fabio Tosi
Federica Bogo
Federico Tombari
Feng Lu
Florian Bernard
Gang YU
Gao Huang
George Kopanas
Georgios Pavlakos
Gordon Wetzstein
Guanying Chen
Guillaume Couairon
Guoyu Lu
Gustavo Carneiro
Gyeongsik Moon
Hailin Jin
Hamid Rezatofighi
Hanbyul Joo
Hao Chen
Hao Wang
Haoliang Li
Haotong Qin
Haoxiang Li
Hazel Doughty
Hong Xuan
Hong Zhang
Hongbin Zha
Hossein Azizpour
Huaizu Jiang
Huan Fu
Hui Ji
Ivan Skorokhodov
Jakob Verbeek
Janne Heikkila
Jason Ren
Jean Ponce
Jennifer J. Sun
Jiajun Wu
Jiaming Zhang
Jian Liang
Jiang Wang
Jiangqun Ni
Jiangxin Dong
Jiankang Deng
Jianxin Wu
Jiasen Lu
Jiaya Jia
Jie Song
Jingdong Wang
Jinshan Pan
John Collomosse
Jonathan T. Barron
Jong Chul Ye
Joon-Young Lee
Joost van de Weijer
Julian Straub
Jun Liu
Junseok Kwon
Juyong Zhang
Jyh-Jing Hwang
Kaifeng Chen
Kaihao Zhang
Kalle Åström
Katja Schwarz
Keita Takahashi
Kent Fujiwara
Kostas Daniilidis
Kota Yamaguchi
Krystian Mikolajczyk
Kwan-Yee K. Wong
Kwang In Kim
Kyle Genova
Le Wang
Lei Li
Lei Zhu
Li Shen
Liang-Chieh Chen
Lianli Gao
Lin Gao
Linchao Bao
Lingxi Xie
Linlin Shen

CVPR 2026 OUTSTANDING AREA CHAIRS (continued)

Longyin Wen
Luc Van Gool
Luisa Verdoliva
Mandi Luo
Marc Habermann
Maria A Zuluaga
Martin Weinmann
Matej Kristan
Mathieu Salzmann
Matteo Poggi
Matthew Joseph Leotta
Matthew Turk
Michael Niemeyer
Michael Wray
Mike Zheng Shou
Min H. Kim
Ming Yang
Minghua Liu
Mingxing Tan
Minjie Cai
Mohsen Ali
Naveed Akhtar
Nicolas Thome
Oana-Maria Camburu
Octavia Camps
Oisin Mac Aodha
Olga Veksler
Olivia Wiles
Or Litany
Oren Freifeld
Paolo Rota
Pascal Fua
Pascal Mettes
Peirong Liu
Philippos Mordohai
Pinar Yanardag
Qi Dai
Qi Mao
Qi Shan
Qi Wu
Qibin Hou
Qilong Wang
Qin Jin
Qing Guo
Qing Wang
Qixing Huang
Ran He
Rene Vidal
Renjie Liao
Richard Zhang
Rohit Girdhar
Rui Shao
Ryoma Bise
Sai-Kit Yeung
Saket Anand
Salman Asif
Sareh Rowlands
Sathyanarayanan N. Aakur
Scott Workman
Ser-Nam Lim
Shang-Hong Lai
Shangzhe Wu
Shizhe Chen
Shohei Nobuhara
Shuai Yang
Shuting He
Sicheng Zhao
Simon Niklaus
Simone Melzi
Songyou Peng
Stamatios Georgoulis
Stefan Lee
Stéphane Herbin
Stephen James
Stratis Gavves
Subhransu Maji
Takayuki Okatani
Tal Hassner
Tatsuya Harada
Tat-Jen Cham
Thomas Eboli
Tolga Birdal
Tomas Jakab
Tony C. W. MOK
Tony Tung
Tushar Nagarajan
Venkatesh Babu Radhakrishnan
Vicky Kalogeiton
Victor Lempitsky
Viktoriia Sharmanska
Vladislav Golyanik
Wanli Ouyang
Wayne Wu
Wei Ji
Wei Shen
Wei-Shi Zheng
Weidong Cai
Weihong Deng
Wenwei Zhang
Xi Peng
Xiang Bai
Xiaochun Cao
Xiaolin Huang
Xiaoming Zhao
Xinlei Chen
Xintong Han
Xu Yang
Xuefei Ning
Xuequan Lu
Xuming He
Yan Huang
Yan Huang
Yan Wang
Yan Yan
Yan Zhang
Yanchao Yang
Yannick Strümpler
Yannis Avrithis
Yaoyao Liu
Yasamin Jafarian
Yi Fang
Yi-Zhe Song
Yibing Song
Yijun Li
Yiming Li
Ying Wu
Yinghuan Shi
Yogesh S Rawat
Yu-Xiong Wang
Yuankai Huo
Yuhang Zang
Yujun Cai
Yuki M Asano
Yung-Yu Chuang
Yunhui Guo
Yunzhong Hou
Yusuke Sugano
Zhangyang Wang
Zhaowei Cai
Zhaoxiang Zhang
Zhengqi Li
Zhizhen Zhao
Zhizhong Han
Zhonghua Wu
Zhun Zhong
Zhuotao Tian
Zongsheng Yue
Zongwei Zhou
Zongxin Yang
Zuxuan Wu

CVPR 2026 OUTSTANDING REVIEWERS

We are grateful to all of the 17,491 reviewers who helped make CVPR 2026 possible. We are especially pleased to recognize the following Outstanding Reviewers, whose high-quality reviews (as judged by their Area Chairs) placed them among the top 5% of reviewers.

A. K. Qin
Aaron Carass
Aaron Sun
Aayush Prakash
Abdelrahman Eldesokey
Abdelrahman M Shaker
Abdul Bais
Abhijit Sarkar
Abhinav Mahajan
Abhishek Jha
Abir Das
Abner Guzman-Rivera
Abolfazl Hashemi
Achuta Kadambi
Adam Lilja
Adam Polyak
Adam Tonderski
Adam W Harley
Aditya Arora
Adnan Siraj Rakin
Adrian Penate-Sanchez
Adrian V Dalca
Ahmadreza Jeddi
Ahsan Baidar Bakht
Ajian Liu
Akhil Perincherry
Akshay Krishnan
Akshay R. Kulkarni
Alain Pagani
Alakh Desai
Alara Dirik
Albert Clapés
Albert Gu
Albert J. Zhai
Aleksandar Cvejić
Alessandro Fiorini
Alessio Tonioni
Alex Hanson
Alexander Carballo
Alexander H Berger
Alexander Mock
Alexander Sax
Alexandra Gomez-Villa
Alexandros Haliassos
Alexandru Paul Condurache
Alexei A Efros
Alexis Joly
Ali Garjani
Ali Mollaahmadi Dehaghi
Alvaro Budria
Alvaro Soto
Amandine Brunetto
Amaya Gallagher-Syed
Amelie Royer
Amin Parchami-Araghi
Amir Houmansadr
Amir Rasouli
Amirhossein Vahidi
Amirmojtaba Sabour
Amit Alfassy
Amita Kamath
Anbang Yao
Anders Johan Thunberg
André Mateus
Andrea Bottino
Andrea Caraffa
Andreas Floros
Andreas Kolb
Andreas Velten
Andrew D. Bagdanov
Anelia Angelova
Ang Cao
Angelica I Aviles-Rivero
Angjoo Kanazawa
Anh-Quan Cao
Aniello Panariello
Ankit Dhiman
Ankur Mali
Anna Frühstück
Anna-Maria Halacheva
Anqi Joyce Yang
Anthony Cioppa
Anthony Fuller
Antoine Guedon
Antoine Manzanera
Anton Obukhov
Antonin Vobecky
Antonio De Santis
Antonio Emanuele Cinà
Anup Kumar Gupta
Ao Li
Ao Luo
Arda Senocak
Ardhendu Behera
Arijit Ray
Arnas Uselis
Arnold Wiliem
Áron Monszpart
Artur Grigorev
Artur Jesslen
Artur Xarles
Aruni RoyChowdhury
Ashish Malik
Ashkan Ganj
Atri Rudra
Atsushi Hashimoto
Avinash Madasu
Awais Nizamani
Ayça Takmaz
Ayush Jain
Baihua Li
Bailin Deng
Balaji Krishnamurthy
BaoFeng Tan
Baptiste Brument
Bardia Safaei
Barry-John Theobald
Bastian Goldluecke
Behzad Bozorgtabar
Benjamin Attal
Benjamin Billot
Benjamin Planche
Benjamin Salmon
Bernhard Jaeger
Bharadwaj Ravichandran
Bin Chen
Bin Fan
Bin Ren
Bing Qin
Bing WANG
Biqing Qi
Björn Michele
Blake E. Dewey
Blaž Rolih
Bo Dai
Bo Lei
Bo Li
Bo Li
Bo Liu
Bo Peng
Bo Yan
Bo-Kai Ruan
Bohai Gu
Boris Ivanovic
Boris X. Vintimilla
Boseung Jeong
Boyang Sun
Brandon Zhao
Brian K. S. Isaac-Medina
Bruno Avignon
Buru Chang
Byeonghu Na
Byung Hyun Lee
Byung-Hoon Kim
Byung-Jun Yoon
Byungjun Kim
Caiyi Sun
Caleb Zheng
Carl Doersch
Chaitanya Patel

Chaithanya Kumar Mummadi
Changjiang Cai
Changwoon Choi
Changyeon Won
Chanho Ahn
Chanhwi Jeong
Charless Fowlkes
Charu Sharma
Che Liu
Chelsea Finn
Chen Guo
Chen Song
Chen Zhang
Chen-Yang Wang
Cheng Deng
Cheng Sun
Cheng Zhao
Cheng-Yao Hong
Cheng-Yen Yang
Cheng-Yi Lee
Chenghao Qian
Chenglong Ma
Chengxuan Qian
Chengyou Jia
Chengyu Wang
Chengze Miaomiao Li
Chengzu Li
Chenhan Jiang
Chenqi Kong
Chenyang Liu
Cheol-Ho Cho
Chi Liu
Chi Xie
Chiara Cappellino
Chih-Chung Hsu
Chih-Hao Lin
Chongjian GE
Chris Callison-Burch
Chris Xiaoxuan Lu
Christian Lange
Christoph Palm
Christoph Vogel
Christopher Pal
Christopher Tensmeyer
Christopher Wewer
Christos Matsoukas
Christos Sakaridis
Chuan Zhang
Chuanxia Zheng
Chuanyang Zheng
ChujieQin
Chun-Ta Lu
Chung-Hao Lee
Chunlong Xie
Chunyuan Li
Churan Wang
Chuyan Xiong
Claudia Cuttano
Clément Rambour
Cong Phuoc Huynh
Congyue Deng
Connelly Barnes
Connor Dunlop
Connor Hashemi
Conrad M Albrecht
Cristian Rodriguez-Opazo
Cristóvão Sousa
Daehyun Kim
Daeyoung Kim
Damian Borth
Damiano Marsili
Damien Muselet
Damien Robert
Dan Oneata
Dan Wang
Danda Pani Paudel
DanDan Zheng
Daniel Garibi
Daniel Kienzle
Daniel Lichy
Daniele Baieri
Danrui Li
Darshan Singh
David Ahmedt-Aristizabal
David Charatan
David Eigen
David Hurych
David M. Chan
David Mizrahi
David Paz
David Serrano-Lozano
David T. Hoffmann
David Tschirschwitz
Davide Bacciu
Davide Bucciarelli
Davide Caffagni
Davide Talon
Davis Rempe
Dayoung Gong
Dayun Ju
Dazhuang Liu
Debarshi Brahma
Decheng Liu
Deen Dayal Mohan
Deep Chakraborty
Delin Qu
Denis Rozumny
Deqian Kong
Derek Bradley
Derong Jin
Derong Xu
Dhruv Shah
Di Feng
Di Lin
Dian Chen
Dian Shao
Dibyadip Chatterjee
Diego Martín Arroyo
Diego Porres
Dimitri Gominski
Dinesh Jayaraman
Dingkang Liang
Diwei Sheng
Dmitrii Torbunov
Dogyun Park
Dominique Ginhac
Donggoo Jung
DongGuw Lee
Donghoon Lee
Donghyun Kim
Dongjin Kim
Dongyoon Han
Dosik Hwang
Dragomir Anguelov
Du Chen
Duan Mingxing
Duo Chen
Duolikun Danier
Duowen Chen
Duy Minh Ho Nguyen
Dylan Campbell
Edgar Simo-Serra
Eduarda Caldeira
Eduardo Pérez-Pellitero
Edward Choi
Ekta Prashnani
Eli Shlizerman
Emanuel Sanchez Aimar
Emeline Got
Eng Siong Chng
Enis Simsar
Enshen Zhou
Eric Peh
Eric Wang
Erwin Coumans
Esa Rahtu
Eslam Abdelrahman
Eva Almansa
Ewa Szczurek
Fabian Immel
Fabian Isensee
Fabien Despinoy
Fabio Cermelli
Fadi Boutros
Fangjinhua Wang
Fangqiang Ding
Farid Boussaid
Faris Janjos
Fatemeh Behrad
Fazel Arasteh
Federico Becattini
Federico Bolelli
Federico Paredes Valles
Federico Spurio
Fei Xue
Feitong Tan
Feiyang Kang
Felix Friedrich
Felix Hertlein
Felix X.-F. Ye
Feng Tian
Fengtao Zhou
Fengwei Zhou
Fengxiang Wang
Fengyi Shen
Ferda Ofli
Fernando Pérez-García
Filippo Aleotti
Filippo Maggioli
Filippo Ziliotto
Finn Wong
Fiona Ryan
Fiora Pirri
Firas Gabetni
Florent Bartoccioni
Florent Lafarge
Florian Fervers
Florian Hofherr
Florian Jug
Francesco Giuliari
Francisco M. Castro
Franck Mamalet
Francois Rameau
Fu-Yun Wang
Fucai Ke
Gabriel Herbert Sarch
Gabriele Trivigno
Gaku Nakano
Gang Wu
Garrett Tanzer
Gary K. L. Tam
Gaurav Patel
Gavriel Habib
Ge-Peng Ji
Gemma Canet Tarrés
Gengze Zhou
Georg Bökman
George Vosselman
Georgios Paschalidis
Geri Skenderi
German Barquero
Ghassan AlRegib
Gilad Lerman
Gilles Puy
Giuseppe Cartella
Görkay Aydemir
Goutam Bhat
Gregory Dudek
Grzegorz Cielniak
Guan Huankang
Guangkai Xu
Guangqian Guo
Guangrun Wang
Guangtao Lyu
Guanxiong Sun
Guanzhou Ke
Guido D'Amely
Guillaume Astruc
Guillaume Bono
Guillaume Jeanneret
Guillaume Sautiere
Guobao Xiao
Guolin Ke
Guorong Li
Haichuan Song
Haifeng Huang
Haifeng Zhao
Haiming Xu
Haisheng Su
Haitao Zheng
Haitao Zhou
Haiwei Wu
Haiyang Xu
Hallee E. Wong
Hamed Alemohammad
Hamid R. Rabiee
Hampus Linander
Han Xu
Hang Su
Hangyu Li
Hangyu Qin
Hannah Kerner
Hanno Gottschalk
Hanoona Abdul Rasheed
Hanqing Zhao
Hao Chen
Hao Cheng
Hao Ding
Hao Shi
Hao Sun
Hao Yang
Hao Yang
Haodong Jing
Haohong Lin
Haojie Zheng
Haoli Bai
Haoliang Sun
Haoming Song
Haonan Duan
Haonan Wang
Haopeng Zhang
Haoran Xie
Haotao Wang
Haotian Xia
Haotian Ye
Haoyang Li
Haoyu Chen
Haoyu Tang
Haoyue Liu
Hashmat Shadab Malik
He Wang
Hector Andrade-Loarca
Hee-Seon Kim
Heewon Kim
Hefeng Wu
Hemal Naik
Hemanth Saratchandran
Heng Guo
Hengfei Cui
Hengyuan Zhang
Hengyue Liang
Hermann Blum
Hervé Le Borgne
Hideo Saito
Hidir Yesiltepe
Hiroki Furuta
Hironobu Fujiyoshi
Hiroshi Kera
Hitesh Kandala
Ho Kei Cheng
Ho Yin Au
Hoang Anh Just
Hoin Jung
Homanga Bharadhwaj
Hong Yang
Hong Zhang
Hongchen Li
Hongfu Sun
Honghai LIU
Hongjie Wang
Hongjun Wang
Honglei Yan
Honglin He
Hongxin Wei
Hongyu Yang
Hongzhi Liu
Hossein Souri
Hsiang-Wei Huang
Huaming Chen
Huaxiaoyue Wang
Hui Liang
Hui Yin
Hui Yu
Huibin Li
Hunter Blanton
Hyeokjun Kweon
Hyeongyu Kim
Hyeonjun Sim
Hyun Kim
Hyungjun Joo
HyunJun Jung
Hyunjun Jung
Hyunsu Kim
Hyunsung Jang
Iacopo Masi
Iago Suárez
Ibne Farabi Shihab
Idil Esen Zulfikar
Ikuro Sato

In So Kweon
Inbar Huberman-Spiegelglas
Indro Spinelli
Ioannis Karamouzas
Ioannis Siglidis
Ishan Rajendrakumar Dave
Ismail Nejjar
Iuliia Kotseruba
Ivan Lopes
Jaehong Kim
Jaehoon Choi
Jaehyung Kim
Jaejun Yoo
Jaekoo Lee
Jaesung Huh
Jaewoo Park
Jaeyeul Kim
Jaeyong Song
Jaeyoung Chung
Jaeyoung Do
James Hays
James Tompkin
Jamie Wynn
Jan Kautz
Jan van Gemert
Jana Kosecka
Jason J. Yu
Jason Kuen
Jason Parham
Jason Rambach
Jay Zhangjie Wu
Jayakrishnan Unnikrishnan
Jaywon Koo
Jean-Marc Odobez
Jeannette Bohg
Jeany Son
Jefferson Hernandez
Jeffrey A Chan Santiago
Jeffrey Gu
Jenia Jitsev
Jens Behley
Jeong-gi Kwak
Jeonghun Kang
Jeonghwan Park
Jer Pelhan
Jessie Wang
Ji Du
Ji Woo Hong
Jia Jia
Jia Li
Jia Wang
Jia-Fong Yeh
Jia-Wei Liu
Jia-Xin Zhuang
Jiachen Tao
Jiacheng Wei
Jiaer Xia
Jiahang Zhang
Jianan Fan
Jianan Liu
Jianchuan Chen
Jianfei Yang
Jiangtao Li
Jiangtong Li
Jianhao Zheng
Jianing Sun
Jianqi Chen
Jianrong Zhang
Jianzhu Huai
Jiaru Zhang
Jiarui Hu
Jiarui Zhang
Jiawei Liu
Jiaxin Guo
Jiaxin Wang
Jie Chen
Jie Zhang
Jihang Wang
Jiho Choi
Jihwan Bang
Jihyong Oh
Jin Ye
Jinfeng Wang
Jing Qin
Jing Tang
Jingjing Fu
Jingjing Li
Jingjing Zhao
Jingkang Yang
Jinglin Xu
Jinglu Wang
Jinglun Li
Jingwei Tang
Jingxuan Wei
Jingyan Jiang
Jingyu Yang
Jinhyun Jang
Jinhyung Park
Jinjia Peng
Jinkun Cao
Jinlin Wu
Jinrui Yang
Jinsu Yoo
Jinxiu Liang
Jipeng Zhang
Jisang Han
Jiwoo Kang
Jiyeon Han
Jiyuan Wang
Joakim Bruslund Haurum
Joao F. Henriques
Johanna Wald
Johannes Betz
Johannes Künzel
Johannes Schönberger
John See
John Seon Keun Yi
Jon Tamir
Jonas Kulhanek
Jonathan Astermark
Jonathon Hare
Jong Hwan Ko
Jongseong Bae
Jongwon Choi
Joni Pajarinen
Jordi Gonzàlez
Jorge Calvo-Zaragoza
Joschka Boedecker
Joseph Campbell
Joseph Tung
Joshua C. Zhao
Joyce Jiyoung Whang
Juan-Manuel Perez-Rua
Julia Dietlmeier
Julia Grabinski
Julian Jorge Andrade Guerreiro
Julian Lorenz
Juliette Marrie
Jun Shu
Jun Wan
Jun Xiang
Jun Yang
Jun Yue
Jun-Hyuk Kim
Junbo Chen
Junheum Park
Junhyeog Yun
Junhyeok Kim
Junjie Wang
Junmo Kim
Junsheng Wang
Juntao Li
Junuk Cha
Junwu Weng
Junxuan Li
Junyi Li
Juzheng Miao
Kaede Shiohara
Kailai Zhou
Kaiming He
Kaixuan Zhang
Kaiyuan Chen
Kaiyue Sun
Kaleab A Kinfu
Kaleel Mahmood
Kang Eun Jeon
Kangil Kim
Kannappan Palaniappan
Kanzhi Cheng
Kaouther Messaoud
Karim Lekadir
Karsten Roscher
Katherine A. Skinner
Kavana Venkatesh
Keisuke Tateno
Keji He
Keling Yao
Kelvin Wong
Keming Wu
Kenji Enomoto
Khang Nguyen
Khanh-Duy Nguyen
Khiem Vuong
Kibum Kim
Kiet A. Nguyen
Kim Sung-Bin
Klara Janouskova
Konda Reddy Mopuri
Konstantinos Batsos
Konstantinos Georgiou
Krishna Murthy Jatavallabhula
Krishnamurthy Viswanathan
Kun Fang
Kun Wang
Kun Yan
Kun Zhang
Kunal Gupta
Kunming Luo
Kunyi Li
Kunyu Wang
Kwan-Yee Lin
Kwanghoon Sohn
Kwangsu Kim
Kyong Hwan Jin
Kyoung Mu Lee
Kyungtae Han
Lahiru Nuwan Wijayasingha
Lambert Mathias
Lan Emily Zhang
Lanqing Li
Lars Hammarstrand
Lars Jebe
Laurynas Karazija
Lavisha Aggarwal
Leander Girrbach
Lei Luo
Lei Zhang
Leif Van Holland
Lele Cao
Lennart Bastian
Leo Maxime Brunswic
Leonard Bruns
Leonid Keselman
Léopold Maillard
Levy Chaves
Li Tang
Li Wang
Liangqi Li
Libo Huang
Lijun Wang
Liliang Chen
Lilika Makabe
Lina Yu
Linfei Pan
Ling Li
Lingdong Kong
Lingfei Ma
Lingmin Ran
Linjie Li
Linshan Wu
Linwei Chen
Liqiang Nie
Liwei Liao
Liwei Wang
Liyuan Zhu
Lode Jorissen
Lorena Qendro
Lorenzo Bianchi
Lu Lin
Luca Bartolomei
Luca Di Giammarino
Luca Soldaini
Luca Zancato
Lucas Relic
Lucile Sassatelli
Lucy Chai
Luigi Piccinelli
Luigi Seminara
Luis Baumela
Luisa Vargas
Lukas Höllein
Lukas Picek
Lukas Schmid
Lumin Xu
Luoxin Ye
Luxi Zhao
M. Saquib Sarfraz
M. Usman Rafique
Maciej Zieba
Mackenzie Weygandt Mathis
Madalina Fiterau
Mahmoud Afifi
Mahmut Yurt
Malte Prinzler
Mani Ramanagopal
Manling Li
Manogna Sreenivas
Manthan Patel
Manuel Kaufmann
Marcus Magnor
Maria Alejandra Bravo
Maria Marrium
Maria Sofia Bucarelli
Mark Endo
Markus Enzweiler
Martin Benning
Martin Foltin
Martin JJ. Bucher
Martin Simon
Marwan Torki
Masashi Hatano
Matan Kichler
Matic Fučka
Matt De Vries
Matteo Bortolon
Matthieu Le
Maxime Pietrantoni
Mehdi Hosseinzadeh
Mehrdad Saberi
Mei Chen
Melinos Averkiou
Meng Du
Meng Wei
Meng Zheng
Mengmeng Ma
Mengyin Liu
Mengyu Yang
Meysam Madadi
Miao Zhang
Michael Baumgartner
Michael Bi Mi
Michael Firman
Michael Hubbertz
Michael Wand
Michael Weinmann
Michael Yu Wang
Michal Jan Tyszkiewicz
Mike Roberts
Min Liu
Minfeng Xu
Ming He
Ming Kong
Mingfei Chen
Minggui Teng
Minghan Li
Minghao Chen
Minghao Liu
Mingjia Li
Mingyang Chen
Minh-Quan Viet Bui
Minhyeok Lee
Minje Kim
Mischa Dombrowski
Mohamed A. Abdelsalam
Mohamed Sayed
Moncef Gabbouj
Moonkyung Ryu
Motoharu Sonogashira
Mridul Khurana
Mu Hu
Mubarak Shah
Muhammad Abdullah Jamal
Muhammad Atif Butt
Muhammad Kashif Ali

Nafie El Amrani
Nakul Sharma
Namyong Park
Nan Duan
Nan Ma
Nan Song
Naoki Murata
Naoshi Kaneko
Naoya Sogi
Nathaniel Burgdorfer
Nathaniel Eliot Chodosh
Naufal Suryanto
Naveen Venkat
Navin Ranjan
Navlika Singh
Nedyalko Prisadnikov
Neelu Madan
Neha Bhargava
Nermin Samet
Nhat Ho
Nianyi Li
Nicholas John Eliopoulos
Nicklas Hansen
Niclas Zeller
Nicolai Häni
Nicolas Audebert
Nicolas Dufour
Nicolas Michel
Nie Lin
Nikhil Gosala
Nikhil Varma Keetha
Niki Amini-Naieni
Nikita Araslanov
Nikola Popovic
Nikolaos Papamarkou
Nikolaos Passalis
Nikos Athanasiou
Nikos Deligiannis
Niloufar Alipour Talemi
Nina Miolane
Ning Xu
Ning Xu
Ning-Chi Huang
Niv Cohen
Noa Garcia
Nobuhiko Wakai
Nur Suriza Syazwany
Nyle Siddiqui
Olaf Dünkel
Oliver Zendel
Olivier Moliner
Omprakash Chakraborty
Ondrej Miksik
Orest Kupyn
Oscar Michel
Pablo Garrido
Pablo Marcos-Manchón
Pan Zhou
Panagiotis Papadakis
Partha Das
Patrick Bas
Patrick Beukema
Patrick J. Flynn
Pau de Jorge
Paul Bodesheim
Paul Bogdan
Paul Couairon
Paul Roetzer
Paul Vicol
Paul-Edouard Sarlin
Pei An
Pei Xu
Pei Zhou
Peilin Zhao
Peiqin Sun
Peiqing Yang
Peixuan Zhang
Peizheng Li
Peng Gao
Pengcheng Xu
Pengxiang Yan
Per-Erik Forssen
Petar Palasek
Peter Kontschieder
Petr Hruby
Philipp Lindenberger
Pier Luigi Dovesi
Pierluigi Zama Ramirez
Pierre Moulon
Pietro Morerio
Pingting Hao
Pinxin Liu
Pinxuan Dai
Piyush Nitin Bagad
Poorvi Hebbar
Praful Mathur
Prakash Chandra Chhipa
Priyank Pathak
Prune Truong
Przemysław Spurek
Puneet Gupta
Purvish Jajal
Qi Chen
Qi Tian
Qi WANG
Qi Yang
Qi Yang
qiang liu
Qiang Zhou
Qianglong Chen
Qianli Feng
Qianyu Zhou
Qifan Fu
Qiming Hu
Qing Yu
Qingpei Guo
Qingqiu Li
Qingtian Zhu
Qingyang Tan
Qingyong Hu
Qinsong Li
Qiong Zhang
Qirong Ho
Qirui Wu
Qisen Wang
Qixiu Li
Qiyuan Dai
Qiyuan He
Qiyuan Zhuang
Qizhen Lan
Quan Kong
Quentin Bammey
Quoc-Huy Trinh
Radek Danecek
Radu Marculescu
Rahul Tallamraju
Rainer Stiefelhagen
Rajeev Yasarla
Rajiv Soundararajan
Raman Dutt
Ranga Rodrigo
Rares Andrei Ambrus
Ravi Garg
Raviteja Vemulapalli
Rémi Pautrat
Rémy Sun
Ricardo Henao
Richard Jiang
Richard Newcombe
Richard Szeliski
Richard Tucker
Rico Sennrich
Rinat Abdrashitov
Rining Wu
Rishabh Jain
Rishubh Parihar
Riza Alp Guler
Robert Benavente
Robert K. Katzschmann
Robert Sablatnig
Robert T. Collins
Roberto Di Via
Robin Swanson
Rodrigo Mira
Rodrigo Santa Cruz
Rohit Mohan
Rolandos Alexandros Potamias
Romain Vo
Roman Jacome
Roman Klokov
Rongchang Li
Rosalie Martin
Rui Liao
Rui Liu
Rui Sun
Rui Yu
Rui Zhou
Ruihan Lu
Ruihao Gong
Ruixu Geng
Ruiyu Wang
Ruizhe Liu
Runkai Zhao
Runmin Cong
Runyu Ding
Ruqi Huang
Ryan Rabinowitz
Saarthak Kapse
Sabrina Schmedding
Sachin Kelkar
Saehoon Kim
Sai Kumar Dwivedi
SaiKiran Tedla
Sainan Liu
Sairam Vcr
Saksham Suri
Sakuya Ota
Salman Khan
Sammy Christen
Samuel Schulter
Samuli Laine
Samy Tafasca
Sang-Chul Lee
Sangheum Hwang
Sanghoon Lee
Sanghyuk Chun
Sanghyun Park
Sangyu Han
Sanjoy Chowdhury
Sanket Biswas
Sanqing Qu
Santhosh Kumar Ramakrishnan
Sara Beery
Sara Sabour
Sarah Ibrahimi
Sarah Rastegar
Saswat Subhajyoti Mallick
Savva Ignatyev
Sayan Deb Sarkar
Sayanta Adhikari
Scott C. Lowe
Scott Cohen
Scott S. Howard
Sebastian Koch
Sebastian Scherer
Sebastian Weiss
Sehoon Ha
Sehwan Kim
Seil Kang
Selim Engin
Senthil Yogamani
Seokju Cho
Seonwook Park
Sergey Prokudin
Seung-Hun Nam
SeungJun Moon
Seungwook Kim
Shahzad Ahmad
Shaofeng Yin
Shaogang Gong
Shaokun Wang
Shaoxiang Wang
Shaoyu Liu
Shariq Farooq Bhat
Shashank Agnihotri
Shashank Gupta
Shashank Tripathi
Shashanka Venkataramanan
Shelly Golan
Shen Chen
Shen Zhao
Sheng Yang
Shengbang Tong
Shengbo Eben Li
Shengxuming Zhang
Shengyu Huang
Sherry X. Chen
Shi Chen
Shijie Wang
Shimon Malnick
Shin seong Kim
Shin'ya Yamaguchi
Shir Amir
Shishir Shah
Shiwei Tan
Shiwei Zhang
Shohei Taniguchi
Shoma Iwai
Shraman Pramanick
Shreelekha Revankar
Shrisha Bharadwaj
Shu Zhao
Shuai Wang
Shuai Yuan
Shufan Li
Shuhong Zheng
Shulin Tian
Shuo Zhang
Shuyuan Tu
Shuzhe Wang
Shweta Bhardwaj
Shweta Mahajan
Si Li
Sicheng Zuo
Siddhartha Chaudhuri
Sidi Wu
Sidong Liu
Sifeng SHANG
Silvano Galliani
Silvia Sellán
Simon Boeder
Simon Giebenhain
Simon Osindero
Simon S. Woo
Simone Milani
Sirui Xu
Sixian Zhang
Siyou Lin
Siyu Ren
Siyuan Bian
Sizhuo Ma
Song Wang
Soo Kyung Kim
Souhaib Ben Taieb
Soumava Paul
Soumen Basu
Sourajit Saha
Spencer Whitehead
Srijan Das
Stefan Maria Ailuro
Stefan Oehmcke
Stefan Stojanov
Stefanos Zafeiriou
Stephan Mandt
Stephen Becker
Steve Cruz
Steve Drew
Steve Seitz
Steven McDonagh
Stylianos Moschoglou
Sujan Sai Gannamaneni
Sukrut Rao
Sumit Mamtani
Sunan He
Sunghyun Cho
Sungmin Cha
Sungnyun Kim
Sungrae Hong
Sungroh Yoon
Sunyoung Jung
Suorong Yang
Suparna Bhattacharya
Suprosanna Shit
Suqin Yuan
Suryansh Kumar
Sylvain Lobry
Szymon Rusinkiewicz
Tabish Syed
Taci Ata Kucukpinar
Taegyeong Lee
Tamas Matuszka
Taras Khakhulin
Teck-Yian Lim
Tejaswi Kasarla
Teodora Popordanoska
Tero Karras
Tetiana Martyniuk

Themos Stafylakis
Thomas Brox
Thomas De Min
Thomas Gossard
Thomas Hummel
Thomas Mitchel
Thomas Monninger
Tian Yun
Tian-Zhu Xiang
Tianhao Walter Wu
Tianyu Li
Tianyu Wang
Tien-Tsin Wong
Tieyuan Chen
Till Beemelmanns
Tim K. Marks
Timo Bolkart
Ting Liu
Ting Xiao
Ting-Feng Zhao
Tingting Liao
Tingwu Wang
Tobias Fischer
Tobias Kirschstein
Todd Zickler
Tom Tirer
Tomaso Poggio
Tommaso Apicella
Tomoki Ichikawa
Tong Chen
Toshiaki Fujii
Tracy Hammond
Trong-Tung Nguyen
Trung Tuan Dao
Trung-Nghia Le
Tsai-Shien Chen
Tuan Duc Ngo
Tuka Alhanai
Tuna Han Salih Meral
Tz-Ying Wu
Urs Waldmann
Utkarsh Aashu Mishra
Uttaran Bhattacharya
Valentinos Pariza
Vanessa Sklyarova
Varun Manjunatha
Vasileios Mygdalis
Victor Ion Butoi
Vincent Casser
Vincent Etter
Vivien Sainte Fare Garnot
Wankou Yang
Wei Cheng
Wei Ding
Wei Liu
Wei Liu
Wei Pan
Wei Qian
Wei Shang
Wei Ye
Weihao Li
Weihao Xia
Weijie Zhou
Weikai Huang
Weilin Wan
Weining Ren
Weining Wang
Weiqin Zhao
Wenbin Ouyang
Wengang Zhou
Wenhao Wu
Wenjie Wang
Wenjing Li
Wenliang Zhong
Wenming Cao
Wenshan Wang
Wentao Qu
Wenxuan Li
Wenya Wang
Wenyao Zhang
Wenyi Mo
Wenzhao Lian
Wolfgang Förstner
Won-Dong Jang
Wongi Jeong
Wonjun Hwang
Wonkwang Lee
Woochul Kang
Wujiang Xu
Wynne Hsu
Xavier Michel Tricoche
Xi Liu
Xi Xiao
Xi Yang
Xi-Le Zhao
Xiang Wan
Xiang Zheng
Xianghua Ye
Xiangru Huang
Xiangyu He
Xiangyu Zeng
Xianyu Chen
Xiao-Shan Gao
Xiaojing Ma
Xiaomeng Li
Xiaoming Zhai
Xiaopeng Ji
Xiaoqian Liang
Xiaotong Luo
Xiaoyang Lyu
Xiaoyu Dong
Xiaoyu Hu
Xiaoyu Liu
Xiaoyu Shen
Xin Chen
Xin Guo
Xin Huang
Xin Li
Xin Liao
Xin Luo
Xin Xing
Xincheng Yao
Xinde Li
Xingguang Zhang
Xingjian Bai
Xingjian Shi
Xingrui Wang
Xinliang Zhang
Xinwen Cheng
Xinxin Zuo
Xinyu Yan
Xinyuan Liu
Xirong Li
Xixi Liu
Xuchong QIU
Xudong Mao
Xuehui Wang
Xuelu Feng
Xuerui Qiu
Xuhang Chen
Xuhui Kang
Xun Guo
Xun Wang
Xunzhe Zhou
Yan Kang
Yan-Tsung Peng
Yanbin Hao
Yancheng Zhang
Yandong Tang
Yang Cao
Yang Cong
Yang Gao
Yang Yang
Yang Zhou
Yang-yang Li
Yangyi Huang
Yanshuo Chen
Yanyun Qu
Yao Zhu
Yao-Chih Lee
Yaojie Liu
Yaonan Wang
Yaqing Ding
Yawar Siddiqui
Yawen Lu
Yawen Zhang
Yaxiong Chen
Yaxiong Wang
Yaxu Xie
Ye Fang
Yechan Kim
Yefei He
Yi Li
Yi Wang
Yi Wu
Yi-Hao Peng
Yi-Ping Phoebe Chen
Yichang Shih
Yichen Sheng
Yiduo Hao
Yifan Peng
Yifan Wang
Yifeng Shi
Yifu Ding
Yihang Qiu
Yihong Xu
yilun chen
Yiming Huang
Yiming Yang
Yinghui Sun
Yingjie Victor Chen
Yinzhen Xu
Yiqun Zhao
Yitong Li
Yixiang Chen
Yixin Wang
Yixin Yang
Yizhi Song
Yizhuo Lu
Yohan Jo
Yohann Perron
Yonghan Lee
Yongjun Zhang
Yongle Zhao
Yongsheng Yu
Yoonki Cho
Youmin Xu
Young-Han Son
Youngho Yoon
Youngjin Oh
Younkwan Lee
Youpeng Wen
Yu Chen
Yu Fang
Yu He
Yu Zheng
Yu-Bin Yang
Yuan Cheng
Yuan Xie
Yuanbo Xiangli
Yuanfeng Zhou
Yuanhao Wang
Yuanlong Wang
Yuanpei Liu
Yubin Cho
Yubin Gu
Yuchen Yan
Yucheol Jung
Yudong Guo
Yueh-Hua Wu
Yueqi Duan
Yufan He
Yuge Huang
Yuhan Ping
Yuhang Liu
Yuhao Chen
Yuheng Qiu
Yuhui Yuan
Yujia Li
Yujie Wang
Yujun Liu
Yuki Kondo
Yulong Lu
Yumeng Liu
Yuming Du
Yuming Qiao
Yuming Wang
Yun He
Yun Liu
Yunbin Tu
Yunrong Guo
Yuntao Wang
Yunwen Zhou
Yusuf Dalva
Yusuke Mukuta
Yutaka Matsuo
Yutong Xie
Yuval Haitman
Yuwei Wu
Yuxiang Fu
Yuxiang Yan
Yuxing Tang
Yuxuan Zhang
Zaber Ibn Abdul Hakim
Zahra Gharaee
Zanming Huang
Ze Li
Zefan Cai
Zehua Chen
Zeming Li
Zengsheng Kuang
Zerui Chen
Zerui Gong
Zewei Zhang
Zeyu Wang
Zeyuan Yin
Zhakshylyk Nurlanov
Zhan Peng
Zhang Chen
Zhao-Min Chen
Zhaoyang Jia
Zhaoyang Lv
Zhaoyang Zhang
Zhaoyuan Fang
Zhe Li
Zhe Liu
Zhe Zhang
Zhehao Huang
Zhejun Zhang
Zhen Han
Zhenbo Xu
Zhengbao He
Zhenghao Chen
Zhengwei Yang
Zhengyuan Li
Zhenqi He
Zhenyu Li
Zhenyu Zhou
Zhi Gao
Zhida Sun
Zhifan Zhu
Zhifang Zhang
Zhigang Deng
Zhinan Qiao
Zhipeng Bao
Zhipeng Du
Zhipeng Zhang
Zhiqiang Yan
Zhirui Dai
Zhiwei Yang
Zhiwen Chen
Zhixiang Wei
Zhixin Wang
Zhixun Su
Zhiyuan Ren
Zhizhong Li
Zhong Li
Zhong-Qiu Zhao
Zhongbao Yang
Zhonghan Zhao
Zhongjie Mi
Zhulin An
Zhuo Su
Zicheng Wang
Zicong Fan
Zijian He
Zilei Wang
Zinuo Li
Zinuo You
Zipeng Wang
Ziqi Zhou
Ziqian Zeng
Ziren Gong
Zixin Wang
Zixuan Chen
Zixuan Jiang
Ziyi Chen
Ziying Song
Ziyuan Liu
Ziyun Zeng
Zonghao Guo
Zonglin Li
Zorah Lähner
Zuchao Li
Zuria Bauer

Friday, June 5

7:00 - 17:00	**Registration / Badge Pickup** (Lobby A)
7:00 - 17:00	**Press Room** (ExHall F)
7:00 - 17:00	**Mother's Room** (Adjacent to Room 102)
7:00 - 17:00	**Quiet Room** (Adjacent to Room 102)
7:00 - 17:00	**Prayer Room** (Room 206)
7:00 - 8:30	**Findings Posters** (ExHall A)
7:00 - 9:00	**Breakfast** (ExHall C)
7:30 - 17:00	**Poster Pickup / T-shirt Pickup** (ExHall A & F)
8:45 - 9:15	**Poster Setup** (ExHall A)
8:30 - 9:00	**Welcome & Awards** (Bluebird Ballroom)
9:00 - 9:15	**Courtesy Break**
9:15 - 10:30	**Oral Session 1A: Multimodal Vision** (Bluebird Ballroom) **Oral Session 1B: Visual Security** (Four Seasons Ballroom) **Oral Session 1C: Efficient Reasoning** (Mile High Ballroom 1A - 2A) **Oral Session 1D: Computational Imaging** (Mile High Ballroom 3A - 4A)
10:15 - 10:45	**Poster Setup** (ExHall A)
10:45 - 11:30	**Coffee** (ExHall F)
10:45 - 12:45	**Poster Session 1 and Exhibit Hall** (ExHall A-F)
10:45 - 18:00	**Art Program** (ExHall E)
10:45 - 12:45	**DEMOS** (ExHall F)
11:00 / 17:00	**Art Gallery Tour with Curator, Luba Elliott** (ExHall E) 30 mins each
11:45 - 13:30	**Lunch** (ExHall)
13:00 - 14:15	**Oral Session 2A: 3D Reconstruction** (Bluebird Ballroom) **Oral Session 2B: Materials & Lighting** (Four Seasons Ballroom) **Oral Session 2C: Gaussian Splatting & Reconstruction,** (Mile High Ballroom 1A - 2A) **Oral Session 2D: Spatio-Temporal Reconstruction,** (Mile High Ballroom 3A - 4A)
13:30 - 14:30	**Art Panel** (Room 201)
14:15 - 14:30	**Courtesy Break**
14:45 - 15:45	**KEYNOTE 1 - Simon Kohl, Founder & CEO, Latent Labs; Programmable Biology: Generative AI for Molecular Design** (Bluebird Ballroom)
15:30 - 16:00	**Poster Setup** (ExHall A)
16:00 - 18:00	**Poster Session 2 & Exhibit Hall w/ Coffee Break** (Exhall A & F)
16:00 - 18:00	**DEMOS** (ExHall F)

7:00 - 8:30 Findings Posters (ExHall A)

1 BLMT-Stereo: Breaking the Local Minima Trap of Iterative Stereo Matching, *Zhien Dai, Zhaohui Tang, Hu Zhang, Mingjun Pan, Jin Luo, Yongfang Xie*

2 SwiftNDC: Fast Neural Depth Correction for High-Fidelity 3D Reconstruction, *Kang Han, Wei Xiang, Lu Yu, Mathew Wyatt, Gaowen Liu, Ramana Rao Kompella*

3 ProDiG: Progressive Diffusion-Guided Gaussian Splatting for Aerial to Ground Reconstruction, *Sirshapan Mitra, Yogesh S Rawat*

4 GauSDF: Signed Distance Embedded Gaussian Surfels for 3D Reconstruction, *Minsol Kim, Usman Ali*

5 4D E-SloMo: 4D Reconstruction for High Speed Scene using a Hybrid RGB-Event Multi-View System, *Bo Xu, Jun Dai, Yutian Chen, Linning Xu, Mulin Yu, Yujin Wang, Shi Guo, Xinyi Le, Tianfan Xue*

6 MADrive: Memory-Augmented Driving Scene Modeling, *Polina Karpikova, Daniil Selikhanovych, Kirill Struminsky, Ruslan Musaev, Maria Golitsyna, Dmitry Baranchuk*

7 OnlineX: Unified Online 3D Reconstruction and Understanding with Active-to-Stable State Evolution, *Chong Xia, Fangfu Liu, Yule Wang, Yize Pang, Yueqi Duan*

8 SyncTrack4D: Cross-Video Motion Alignment and Video Synchronization with Multi-Video 4D Gaussian Splatting, *Yonghan Lee, Tsung-Wei Huang, Shiv Gehlot, Jaehoon Choi, Guan-Ming Su, Dinesh Manocha*

9 AR4D: Autoregressive 4D Generation from Monocular Videos, *Hanxin Zhu, Tianyu He, Zhibo Chen*

10 CLLAP: Contrastive Learning-based LiDAR-Augmented Pretraining for Enhanced Radar-Camera Fusion, *Bingyi Liu, Chuanhui Zhu, Hongfei Xue, Jian Teng, Jipeng Liu, Enshu Wang, Penglin Dai, Pu Wang*

11 Point2Gaussian: Point-Cloud-to-Gaussian Conversion for Efficient 3D Scene Rendering, *Powei Liao, Jiro Abe, Kazumine Ogura*

12 Speed3R: Sparse Feed-forward 3D Reconstruction Models, *Weining Ren, Xiao Tan, Kai Han*

13 FF3R: Feedforward Feature 3D Reconstruction from Unconstrained views, *Chaoyi Zhou, Run Wang, Feng Luo, Mert D. Pesé, Zhiwen Fan, Yiqi Zhong, Siyu Huang*

14 Generalizable Human Gaussian Splatting via Multi-view Semantic Consistency, *Jingi Kim, Wonjun Kim*

15 HEDA: Hyperbolic-Euclidean Dual Adaptation for Robust Real-World Point Cloud Completion, *Aihua Mao, Jun Yang, Yong-Jin Liu, Ying He*

16 WildAni4D: Towards 4D Animal Mesh Reconstruction, *Gyeongsu Cho, Hezhen Hu, Donghyeon Soon, Changwoo Kang, Kyungdon Joo*

17 Instant Colorization of Gaussian Splats, *Daniel Lieber, Alexander Mock, Nils Wandel*

18 Dynamic Scene Decomposition Beyond Moving Objects for High-Fidelity 3D Reconstruction in Autonomous Driving, *Mingbo Dai, Han Yan, Bolun Zhang, Wu Ran, Chao Ma*

19 LangFlash: Feed-forward 3D Language Gaussian Splatting from Sparse Unposed Images, *Yilong Liu, Wanhua Li, Chen Zhu-Tian, Hanspeter Pfister*

20 FACT-GS: Frequency-Aligned Complexity-Aware Texture Reparameterization for 2D Gaussian Splatting, *Tianhao Xie, Linlian Jiang, Xinxin Zuo, Yang Wang, Tiberiu Popa*

21 Affine Bases for Affine Spaces, *Gabriel Dogadov, Marc Alexa*

22 Improving Densification in 3D Gaussian Splatting for High-Fidelity Rendering, *Xiaobin Deng, Changyu Diao, Min Li, Ruohan Yu, Duanqing Xu*

23 OpenTrack3D: Towards Accurate and Generalizable Open-Vocabulary 3D Instance Segmentation, *Zhishan Zhou, Siyuan Wei, Zengran Wang, Chunjie Wang, Xiaosheng Yan, Xiao Liu*

24 GeoFusion-CAD: Structure-Aware Diffusion with Geometric State Space for Parametric 3D Design, *Xiaolei Zhou, Chuangjie Fang, Jie Wu, Jingyi Yang, Boyi Lin, Jianwei Zheng*

25 SPIDER: Spatial Image CorresponDence Estimator for Robust Calibration, *Zhimin Shao, Abhay Yadav, Rama Chellappa, Cheng Peng*

26 GEAR: GEometry-Motion Alternating Refinement for Articulated Object Modeling with Gaussian Splatting, *Jialin Li, Bin Fu, Ruiping Wang, Xilin Chen*

27 GlowGS: Generative Semantic Feature Learning for 3D Gaussian Splatting in Nighttime Glow Scenes, *Beibei Lin, Xiao Cao, Jingyuan Guo, Robby T. Tan*

28 2D Triangle Splatting for Direct Differentiable Mesh Training, *Kaifeng Sheng, Zheng Zhou, Yingliang Peng, Qianwei Wang*

29 ForgeDreamer: Industrial Text-to-3D Generation with Multi-Expert LoRA and Cross-View Hypergraph, *Junhao Cai, Deyu Zeng, Junhao Pang, Lini Li, Xiaopin Zhong, Zongze Wu*

30 3D Gaussian Splatting for Annular Dark Field Scanning Transmission Electron Microscopy Tomography Reconstruction, *Beiyuan Zhang, Hesong Li, Ruiwen Shao, Ying Fu*

31 IDSplat: Instance-Decomposed 3D Gaussian Splatting for Driving Scenes, *Carl Lindström, Mahan Rafidashti, Maryam Fatemi, Lars Hammarstrand, Martin R. Oswald, Lennart Svensson*

32 Stream3D: Streaming Zero-Shot 3D Instance Segmentation with Multi-View Noise Mask Filtering and Manifold Refining, *Jie Xu, Na Zhao*

33 Active Exploration for Sparse Visual Localization, *Johanna Lidholm, Ludvig Dillén, Zuzana Kukelova, Torsten Sattler, Viktor Larsson*

34 GRVS: a Generalizable and Recurrent Approach to Monocular Dynamic View Synthesis, *Thomas Tanay, Mohammed Brahimi, Michal Nazarczuk, Qingwen Zhang, Sibi Catley-Chandar, Arthur Moreau, Zhensong Zhang, Eduardo Pérez-Pellitero*

35 3DFA: Aligning the Features Between Point Cloud and Query Image for Scene-Specific Visual Localization, *Sizhe Song, Yankuan Chi, Shuhan Zhong, S.-H. Gary Chan*

36 Long-LRM++: Preserving Fine Details in Feed-Forward Wide-Coverage Reconstruction, *Chen Ziwen, Hao Tan, Peng Wang, Zexiang Xu, Li Fuxin*
37 Native3D: End-to-End 3D Scene Generation via Unified Mesh-Texture Modeling and Semantic Alignment, *Yibo Liu, Ziwei Zhang, Haozhou Pang, Menghao Li, Lanshan He, Gan Qi*
38 From Orbit to Ground: Generative City Photogrammetry from Extreme Off-Nadir Satellite Images, *Fei Yu, Yu Liu, Luyang Tang, Mingchao Sun, Zengye Ge, Rui Bu, Yuchao Jin, Haisen Zhao, He Sun, Yangyan Li, Mu Xu, Wenzheng Chen, Baoquan Chen*
39 NeVStereo: A NeRF-Driven NVS-Stereo Architecture for High-Fidelity 3D Tasks, *Pengcheng Chen, Yue Hu, Wenhao Li, Nicole M Gunderson, Andrew Feng, Zhenglong Sun, Peter Beerel, Eric J Seibel*
40 VGGT4D: Mining Motion Cues in Visual Geometry Transformers for 4D Scene Reconstruction, *Yu Hu, Chong Cheng, Sicheng Yu, Xiaoyang Guo, Hao Wang*
41 AvatarMix: Identity-Preserving Cross-Avatar Composition for Outfit Personalization, *Zhaorong Wang, Yoshihiro Kanamori, Yuki Endo*
42 Object Pose Transformer: Unifying Unseen Object Pose Estimation, *Weihang Li, Lorenzo Garattoni, Fabien Despinoy, Nassir Navab, Benjamin Busam*
43 SwiftVGGT: A Scalable Visual Geometry Grounded Transformer for Large-Scale Scenes, *Jungho Lee, Minhyeok Lee, Sunghun Yang, Minseok Kang, Sangyoun Lee*
44 CATRF: Codec-Adaptive TriPlane Radiance Fields for Volumetric Content Delivery, *Tung-I Chen, Lingdong Wang, Subhransu Maji, Ramesh K. Sitaraman*
45 PDF-GS: Progressive Distractor Filtering for Robust 3D Gaussian Splatting, *Kangmin Seo, MinKyu Lee, Tae-Young Kim, ByeongCheol Lee, JoonSeoung An, Jae-Pil Heo*
46 Three-Step Conditional Diffusion 3D Reconstruction for Light-Field Microscopy, *Qihong Zhao, Shaokang Yan, Zhimin Qiao, Jinjia Wang, Bo Xiong*
47 LTGS: Long-Term Gaussian Scene Chronology From Sparse View Updates, *Minkwan Kim, Seungmin Lee, Junho Kim, Young Min Kim*
48 Learning a Particle Dynamics Model with Real-World Videos, *Chanho Kim, Suhas V. Sumukh, Li Fuxin*
49 Softmax-GS: Generalized Gaussians Learning When to Blend or Bound, *Chen Ziwen, Peng Wang, Hao Tan, Zexiang Xu, Li Fuxin*
50 G2I: Transitioning a Generalized Monocular Depth Estimation Model to In-Domain Metric Depth Prediction, *Chao Ning, Naoto Yokoya*
51 3D-RE-GEN: 3D Reconstruction of Indoor Scenes with a Generative Framework, *Tobias Sautter, Jan-Niklas Dihlmann, Hendrik P A Lensch*
52 HiDiGen: Hierarchical Diffusion for B-Rep Generation with Explicit Topological Constraints, *Shurui Liu, Weide Chen, Ancong Wu*
53 RiGS: Rigid-aware 4D Gaussian Splatting from a Single Monocular Video, *Chenyu Wu, Wanhua Li, Chen Zhu-Tian, Hanspeter Pfister*
54 RAD: Retrieval-Augmented Monocular Metric Depth Estimation for Underrepresented Classes, *Michael Baltaxe, Dan Levi, Sagie Benaim*
55 WGS: Watertight Geometry Standardization for Scalable 3D Generation, *Dehao Hao, Tanghui Jia, Kaiyi Zhang, Weikai Chen, Zeyu Hu, Yingda Yin, Runze Zhang, Lingting Zhu, Li Yuan, Xin Wang, Long Quan*
56 Self-Evolving 3D Scene Generation from a Single Image, *Kaizhi Zheng, Yue Fan, Jing Gu, Zishuo Xu, Xuehai He, Xin Eric Wang*
57 Distill Any Depth: Distillation Creates a Stronger Monocular Depth Estimator, *Xiankang He, Dongyan Guo, Hongji Li, Ying Cui, Libo Weng, Ruibo Li, Chi Zhang*
58 Re-Depth Anything: Test-Time Depth Refinement via Self-Supervised Re-lighting, *Ananta R. Bhattarai, Helge Rhodin*
59 UniVerse3D: Emerging Properties of Unified Multimodal Models in 3D Understanding and Generation, *Junliang Ye, Zehuan Huang, Yansong Qu, Chunshi Wang, Yunhan Yang, Yang Li, Yawei Luo, Zhuo Chen, Sheng Lu, Jun Zhu, Chunchao Guo*
60 HumanOrbit: 3D Human Reconstruction as 360° Orbit Generation, *Keito Suzuki, Kunyao Chen, Lei Wang, Bang Du, Runfa Blark Li, Peng Liu, Ning Bi, Truong Nguyen*
61 Beyond Voxel 3D Editing : Learning from 3D Masks and Self-Constructed Data, *Yizhao Xu, Hongyuan Zhu, Caiyun Liu, Tianfu Wang, Keyu Chen, Sicheng Xu, Jiaolong Yang, Nicholas jing Yuan, Qi Zhang*
62 Adversarial Agents: Black-Box Evasion Attacks with Reinforcement Learning, *Kyle Domico, Jean-Charles Noirot Ferrand, Ryan Sheatsley, Eric Pauley, Josiah Hanna, Patrick McDaniel*
63 Defending CLIP via Noise-Induced Feature Dynamics for Training-Free, Zero-shot Adversarial Robustness, *Debarshi Brahma, Soma Biswas*
64 Jailbreaking Frontier Foundation Models Through Intention Deception, *Xinhe Wang, Katia Sycara, Yaqi Xie*
65 NSGuard: Null-Space Guided Robust Watermarking for Data Copyright Protection in Customized Generation, *Lizhi Xiong, Jianguo Feng, Ziqiang Li, Jun Li, Weiwei Jiang, Zhangjie Fu*
66 A Robust Out-of-Distribution Detection Framework via Synergistic Smoothing, *Maria Stoica, Abdelrahman Hekal, Alessio Lomuscio*
67 Phantasia: Context-Adaptive Backdoors in Vision Language Models, *Nam Duong Tran, Phi Le Nguyen*
68 BadRSSD: Backdoor Attacks on Regularized Self-Supervised Diffusion Models, *Jiayao Wang, Yiping Zhang, Mohammad Maruf Hasan, Xiaoying Lei, Jiale Zhang, Junwu Zhu, Qilin Wu, Dongfang Zhao*
69 CLIP-Inspector: Model-Level Backdoor Detection for Prompt-Tuned CLIP via OOD Trigger Inversion, *Akshit Jindal, Saket Anand, Chetan Arora, Vikram Goyal*
70 BadVLM: Towards Efficient and Resilient Backdoor Attacks on Large Vision-Language Models, *Ba Luan Dang, Vu Tuan Truong, Long Bao Le*
71 Finetune Like You Pretrain: Boosting Zero-shot Adversarial Robustness in Vision-language Models, *Songlong Xing, Weijie Wang, Zhengyu Zhao, Jindong Gu, Philip Torr, Nicu Sebe*
72 Breaking the Illusion: Consensus-Based Generative Mitigation of Adversarial Illusions in Multi-Modal Embeddings, *Fatemeh Akbarian, Anahita Baninajjar, Yingyi Zhang, Ananth Balashankar, Amir Aminifar*
73 When Data is Scarce, Learn to Adapt: Robust Federated Learning via Adversarial Meta-Optimization, *Md Zarif Hossain, Awal Ahmed Fime, Ahmed Imteaj*
74 Robust Alignment: Harmonizing Clean Accuracy and Adversarial Robustness in Adversarial Training, *Yanyun Wang, Qingqing Ye, Li Liu, Zi Liang, Haibo Hu*
75 DRA: Structure-Preserving Backdoor Erasure via Diagnosing, Recalibrating, and Adapting, *Minwei Wen, Yang Wei, Junhao Xiao, Xiuli Bi, Bin Xiao*
76 APC: Transferable and Efficient Adversarial Point Counterattack for Robust 3D Point Cloud Recognition, *Geunyoung Jung, Soohong Kim, Inseok Kong, Jiyoung Jung*
77 Cognitive Attack Detection in Augmented Reality (CADAR): A Neuro-Symbolic Approach with Particle Filtering on Perception Graphs, *Rongqian chen, Allison Andreyev, Yanming Xiu, Joshua Chilukuri, Shunav Sen, Mahdi Imani, Bin Li, Maria Gorlatova, Gang Tan, Tian Lan*
78 On Evaluating Stateful Defence Models against Query-Based Black-Box Attacks, *Ziad Tariq Muhammad Ali, Raja Muhammad Atif Azad, Muhammad Ajmal Azad, Iain Rice, Umar Daraz, Ali Shariq Imran, James Holyhead*
79 Optimizing Certified Radius of Zero-shot Composed Image Retrieval via Text Guidance, *Junyang Chen, Haomin Ni, Hanjiang Lai*
80 When Interpretability Becomes a Liability: Adversarial Attacks on CBM Concept Layers, *Aditya Sridhar*
81 Red-teaming the Multimodal Reasoning: Jailbreaking Vision-Language Models via Cross-modal Entanglement Attacks, *Yu Yan, Sheng Sun, Shengjia Cheng, Teli Liu, Mingfeng Li, Min Liu*
82 Phantom: A Unified Face-Swap Deepfake Protection Framework with Latent and Spatial Constraints, *Jungkon Kim, Cheolseung Jung, Jong-Min Choi, Juseong Lee*
83 Tap, Scan, Exploit: The Hidden Vulnerabilities of Everyday QR Codes, *Ashish Kumar, Aarthi S, Akshay Agarwal*
84 DeepFakeShield: A Proactive Defense Against Malicious Face Swapping, *Saeed Karimi-Bidhendi, Joseph DeGol, Eric Wengrowski, Dominic Roberts, Kristin Dana*
85 MDG: Masked Denoising Generation for Multi-Agent Behavior Modeling in Traffic Environments, *Zhiyu Huang, Zewei Zhou, Tianhui Cai, Yun Zhang, Jiaqi Ma*
86 LiDAR-to-4D Radar Synthesis for Building Large-Scale Tensor Datasets, *Woo-Jin Jung, Dong-Hee Paek, Seung-Hyun Kong*
87 SurfaceGS: Dynamic Surface Gaussian Splatting for Urban Driving Scenes, *Fudong Ge, Dingning Liu, Hanshi Wang, Yiwei Zhang, Jin Gao, Weiming Hu, Zhipeng Zhang*
88 JACoP: Joint Alignment for Compliant Multi-Agent Prediction,

Qingze Tony Liu, Alen Mrdovic, Danrui Li, Mathew Schwartz, Sejong Yoon, Mubbasir Kapadia

89 Fine-tuning is Not Enough: A Parallel Framework for Collaborative Imitation and Reinforcement Learning in End-to-end Autonomous Driving, *Zhexi Lian, Haoran Wang, Xuerun Yan, Weimeng Lin, Xianhong Zhang, Yongyu Chen, Jia Hu*

90 What Matters for Scalable and Robust Learning in End-to-End Driving Planners?, *David Holtz, Niklas Hanselmann, Simon Doll, Marius Cordts, Bernt Schiele*

91 Physics-Informed Reward Framework for Vision-Language Driven Safe Autonomous Driving, *Xuepei Yang, Mingtao Feng, Weisheng Dong, Lin Chen, Jie Feng, Fangfang Wu, Yufan Zhu, Ajmal Saeed Mian*

92 HorizonWeaver: Generalizable Multi-Level Semantic Editing for Driving Scenes, *Mauricio Soroco, Francesco Pittaluga, Zaid Tasneem, Abhishek Aich, Bingbing Zhuang, Wuyang Chen, Manmohan Chandraker, Ziyu Jiang*

93 VESPA: Open-World Auto-Labeling for 3D Object Detection in Autonomous Driving, *Levente Tempfli, Esteban Rivera, Markus Lienkamp*

94 IRL-VLA: Vision-Language-Action Training via Reward World Model, *Anqing Jiang, Gao Yu, Heng Yuwen, Yiru Wang, Wang Shuo, Jiang Hao, Sun Hao*

95 KnowMTP: A Knowledge-Guided Framework for Multi-Agent Trajectory Prediction in Autonomous Driving, *Rufan Bai, Tianyi Xue, Tiantian Zhou, Weiwei Wu, Changle Li, Yuhuan Lu*

96 MapGPT: A Vision-Language Model for Large-Scale High-Definition Map Generation, *Mengxi Wu, Long Zhou, Zhixia Li, Adrian Kwan, Denis Laprise, Hengyi Huang, Xiaqing Wu, Shuang Wu*

97 RoaD: Rollouts as Demonstrations for Closed-Loop Supervised Fine-Tuning of Autonomous Driving Policies, *Guillermo Garcia-Cobo, Maximilian Igl, Peter Karkus, Zhejun Zhang, Michael Watson, Yuxiao Chen, Boris Ivanovic, Marco Pavone*

98 PAVE: An End-to-End Dataset for Production Autonomous Vehicle Evaluation, *Xiangyu Li, Chen Wang, Yumao Liu, Dengbo He, Jiahao Zhang, Ke Ma*

99 RoadTones: Tone Controllable Text Generation from Road Event Videos, *Chirag Parikh, Siddhi Pravin Lipare, Ravi Kiran Sarvadevabhatla*

100 GRADE: Guiding Realistic Autonomous Driving with Adaptive Trajectory Evolution, *Zehong Ke, Zhiyuan Liu, Yuning Wang, Jinhao Li, Junkai Jiang, Yanbo Jiang, Zhenhua Xu, Jianqiang Wang*

101 SurfelOcc: Self-supervised Occupancy Prediction via 2D Surfel Splatting, *Jikai Wang, Xingtai Gui, Jiahao Gong, Feiyang Tan, Wencheng Han, Cheng-Zhong Xu, Jianbing Shen*

102 dVLM-AD: Enhance Diffusion Vision-Language-Model for Driving via Controllable Reasoning, *Yingzi Ma, Yulong Cao, Wenhao Ding, Shuibai Zhang, Yan Wang, Boris Ivanovic, Ming Jiang, Marco Pavone, Chaowei Xiao*

103 Devil is in Narrow Policy: Unleashing Exploration in Driving VLA Models, *Canyu Chen, Yuguang Yang, Zhewen Tan, Yizhi Wang, Ruiyi Zhan, Haiyan Liu, Xuanyao Mao, Jason Bao, Xinyue Tang, Linlin Yang, Bingchuan Sun, Yan Wang, Baochang Zhang*

104 Learning Vision-Language-Action World Models for Autonomous Driving, *Guoqing Wang, Pin Tang, Xiangxuan Ren, Guodongfang Zhao, Bailan Feng, Chao Ma*

105 AD-R1: Closed-Loop Reinforcement Learning for End-to-End Autonomous Driving with Impartial World Models, *Tianyi Yan, Tao Tang, Xingtai Gui, Yongkang Li, Jiasen Zheng, Weiyao Huang, Lingdong Kong, Wencheng Han, Xia Zhou, Xueyang Zhang, Yifei Zhan, Kun Zhan, Cheng-zhong Xu, Jianbing Shen*

106 Pseudo-Expert Regularized Offline RL for End-to-End Autonomous Driving in Photorealistic Closed-Loop Environments, *Chihiro Noguchi, Takaki Yamamoto*

107 OmniDrive-R1: Reinforcement-driven Interleaved Multi-modal Chain-of-Thought for Trustworthy Vision-Language Autonomous Driving, *Zhenguo Zhang, Haohan Zheng, Yishen Wang, Le Xu, Tianchen Deng, Xuefeng Chen, Qu Chen, Bo Zhang, Wuxiong Huang*

108 CoRT-Predictor: Chain of Risk Thought Autoregressive Trajectory Predictor for Autonomous Driving, *Yanlin Jiang, Yuchen Liu, Mingren Liu*

109 C^2T: Captioning-Structure and LLM-Aligned Common-Sense Reward Learning for Traffic–Vehicle Coordination, *Yuyang Chen, Kaiyan Zhao, Yiming Wang, Ming Yang, Bin Rao, Zhenning Li*

110 PEARL: A Lightweight Prompt-based Feature Interpreter Framework for Real-Time, Anonymous, and Heterogeneous Collaborative Perception, *Armin Maleki, Hayder Radha*

111 Variable-View Diffusion with Geometric Uncertainty Unlocks LiDAR Upsampling, *Pengfei Yang, Sifu Luo, Feng Wu, Fan Zhou, Ting Zhong*

112 RQR3D: Reparametrizing the regression targets for BEV-based 3D object detection, *Ozsel Kilinc, Cem Tarhan*

113 On the Feasibility and Opportunity of Autoregressive 3D Object Detection, *Zanming Huang, Jinsu Yoo, Sooyoung Jeon, Zhenzhen Liu, Mark Campbell, Kilian Q Weinberger, Bharath Hariharan, Wei-Lun Chao, Katie Z Luo*

114 See Tomorrow, Act Today: Foresight-Driven Autonomous Driving, *Bozhou Zhang, Nan Song, Yuang Wang, Jiankang Deng, Xiatian Zhu, Li Zhang*

115 Spatial Transcriptomics as Images for Large-Scale Pretraining, *Yishun Zhu, Jiaxin Qi, Jian Wang, Yuhua Zheng, Jianqiang Huang*

116 DiffGradCAM: A Class Activation Map Using the Full Model Decision to Solve Unaddressed Adversarial Attacks, *Jacob Piland, Christopher Sweet, Adam Czajka*

117 Modality-Aware and Anatomical Vector-Quantized Autoencoding for Multimodal Brain MRI, *Mingjie Li, Edward Kim, Yue Zhao, Ehsan Adeli, Kilian M. Pohl*

118 Fingerprint Fragment Expansion using Image Outpainting Approach based on Spectral Normalization PatchGAN, *C. Zaghetto, A. Purim, W. Oliveira, J. R. Ribeiro, H. Nolla, F. Santos, M. Chang, R. H. Vareto*

119 Improving Autoregressive Image Generation Through Coarse-to-Fine Token Prediction, *Ziyao Guo, Kaipeng Zhang, Michael Qizhe Shieh*

120 Intelligent Photo Retouching with Language Model-Based Artist Agents, *Haoyu Chen, Keda Tao, YiZao Wang, Xinlei Wang, Lei Zhu, Jinjin Gu*

121 Guided Lensless Polarization Imaging, *Noa Kraicer, Erez Yosef, Raja Giryes*

122 Blockwise Divide-and-Aggregate for Image Restoration using Diffusion Priors, *Vishal Purohit, Wei Chen, Qiang Qiu*

123 Towards Imperceptible Watermarking Via Environment Illumination for Consumer Cameras, *Hodaka Kawachi, Tomoya Nakamura, Hiroaki Santo, SaiKiran Kumar Tedla, Trevor D Canham, Yasushi Yagi, Michael S. Brown*

124 Adaptive Continuous Kernel Networks for Image Reconstruction from Non-Uniform Sampling, *Camille Biscarrat, Michaël Gharbi, Rahul Goel, Jonathan Ragan-Kelley, Frédo Durand, Tzu-Mao Li*

125 FreqAdapt: Frequency-Adaptive Processing for RAW Object Detection, *Hanxi Li, Huiling Li*

126 Stability and Non-Local Modeling in Hybrid Convolution–Transformer Networks for Snapshot Hyperspectral Reconstruction, *Xian-Hua Han*

127 Breaking Degradation Coupling: A Structural Entropy–Guided Decoupled Framework and Benchmark for Infrared Enhancement, *Pu Li, Huafeng Li, Yafei Zhang, Yu Liu, Wen Wang*

128 Fast Generative DeOcclusion for Visual Geometry and Robotics, *Jieneng Chen, Tiezheng Zhang, Xiwei Xuan, Ju He, Yifan Yin, Haojun Shi, Suyu Ye, Xinyi Li, Ruisheng Yuan, Tianmin Shu, Alan Yuille*

129 FineCog-Nav: Integrating Fine-grained Cognitive Modules for Zero-shot Multimodal UAV Navigation, *Dian Shao, Zhengzheng Xu, Peiyang Wang, Like Liu, Yule Wang, Jieqi Shi, Jing Huo*

130 Unlocking Single-View Constraints for Efficient Camera Relocalization with Keypoint-Level Multi-View Geometric Consistency in Training, *Hu Lin, Chengjiang Long, Jiqing Zhang, Chuanlu Jiang, Huilin Ge, Erwei Yin, Baocai Yin, Xin Yang*

131 Evolve Vision-Language-Action Model into an Agent with On-the-fly Tool-use, *Ding Yi, Yanzhao Yu, Xili Dai, Xianbiao Qi, Peiwen Sun, Xueqian Wang, Xiangyu Yue, Jianan Wang*

132 Retrieval-VLA: Training-Free In-Context Adaptation for Vision-Language-Action Models, *Yue Zhang, Rui Wang, Jiehong Lin, Zhongrui Wang, Xiaojuan Qi*

133 Revisiting Articulated Parts Perception in Robot Manipulation, *Xiaoqian Wu, Yejie Guo, Xiaoyang Chen, Lixin Yang, Cewu Lu, Yong-Lu Li*

134 Re^2MoGen: Open-Vocabulary Motion Generation via LLM Reasoning and Physics-Aware Refinement, *Jiakun Zheng, Ting Xiao, Shiqin Cao, Xinran Li, Zhe Wang, Chenjia Bai*

135 Teleoperation, Simulation, or Human Video? Data Utilization Law for Robot Manipulation, *Chenhao Shi, Yichen Zhu, Junjie Wen, Yefei Chen, Ziang Liu, Faming Fang*
136 ShelfGaussian: Shelf-Supervised Open-Vocabulary Gaussian-Based 3D Scene Understanding, *Lingjun Zhao, Yandong Luo, James Hays, Lu Gan*
137 RoboTransfer: Controllable Geometry-Consistent Video Diffusion for Manipulation Policy Transfer, *Liu Liu, Xiaofeng Wang, Guosheng Zhao, Keyu Li, Wenkang Qin, Jiagang Zhu, Jiaxiong Qiu, Guan Huang, Zhizhong Su*
138 Switch-JustDance: Benchmarking Whole-Body Motion Tracking Controllers Using a Commercial Console Game, *Jeonghwan Kim, Wontaek Kim, Yidan Lu, Jin Cheng, Fatemeh Zargarbashi, Zicheng Zeng, Zekun Qi, Zhiyang Dou, Nitish Sontakke, Donghoon Baek, Li Yi, Sehoon Ha, Tianyu Li*
139 OminiMAG-SLAM : Unified Online Dual Graph Optimization for Multi-Agent Gaussian SLAM, *Leqian Ding, Caibo Li, Yu Guo, Fei Wang*
140 ReaAct: Bridging Robotic Reasoning and Action Generation Toward Real-World Spatial Generalization, *Yanzhao Yu, Yi Ding, Peijun Tang, Haotian Yang, Xianbiao Qi, Jianan Wang, Xueqian Wang*
141 Learning Multi-Task Robot Trajectory Segmentation from Visual and Kinematic Streams, *Kaiyuan Chen, Shuangyu Xie, Andrew Goldberg, Ken Goldberg*
142 LP3: LLM-based Potential Prediction Policy for Object Navigation using a Scene-Object Semantic Map, *Wei Luo, Xiaohan Wang, Yuehu Liu*
143 RoboScape-R: Unified Reward-Observation World Models for Generalizable Robotics Training via RL, *Yinzhou Tang, Yu Shang, Yinuo Chen, Bingwen Wei, Xin Zhang, Shu'ang Yu, Liangzhi Shi, Chao Yu, Chen Gao, Wei Wu, Yong Li*
144 CoTFly: Making UAVs Think Where to Fly Next Through Visual Chain-of-Thought Reasoning, *Meiqi Wang, Longnyu Xu, Jun Liu, Hewu Li, Han Qiu*
145 RU4D-SLAM: Reweighting Uncertainty in Gaussian Splatting SLAM for 4D Scene Reconstruction, *Yangfan Zhao, Hanwei Zhang, Ke Huang, Qiufeng Wang, Zhenzhou Shao, Dengyu Wu*
146 A1: Adaptive Truncated Vision-Language-Action Model from Affordance to Action, *Kaidong Zhang, Jian Zhang, Rongtao Xu, Yu Sun, Youpeng Wen, Shuoshuo Xue, Xiaoyu Guo, Minghao Guo, Weijia Liufu, Liu Zihou, Kangyi Ji, Zihang Li, Ruiyi Chen, Meng Cao, Jingming Zhang, Shen Zhao, Xiaojun Chang, Feng Zheng, Ivan Laptev, Xiaodan Liang*
147 Ego-Pi: VLA Fine-Tuning for Ego-Centric Human and Robot Data, *Ji Woong Kim, Ke Wang, Zipeng Fu, Sirui Chen, Cong zhao, Jeff Lai, Chelsea Finn*
148 RACE-6D: Real-time Accurate Coarse-to-finE Object 6D Pose Transformer, *Yoonwoo Ha, Hyungpil Moon*
149 MiVLA: Towards Generalizable Vision-Language-Action Model with Human-Robot Mutual Imitation Pre-training, *Zhenhan Yin, Xuanhan Wang, Jiahao Jiang, Kaiyuan Deng, Pengqi Chen, Shuangle Li, Chong Liu, Xing Xu, Jingkuan Song, Lianli Gao, Heng Tao Shen*
150 Riemannian Score-Based Diffusion for Language-Conditioned Grasp and Affordance Detection, *Yan Li, Zhouchao Fu, Wenbin Lu, Junjie Zheng, Junnan Xu, Junjie Liao, Jianwei Zheng*
151 DINO-VO: Learning Where to Focus for Enhanced State Estimation, *Qi Chen, Guanghao Li, Sijia Hu, Xin Gao, Junpeng Ma, Xiangyang Xue, Jian Pu*
152 Temporally-Smooth Global Bundle Adjustment for Real-Time Dense Visual SLAM, *Cabrel Wouladje, Golden Tendekai Mumanikidzwa, Md Apon Islam, Huiying Xu, Hongbo Li, Wenzhe Tan, Zhendong Chen, Xinzhong Zhu*
153 Masked Next-Scale Prediction For Self-Supervised Scene Text Recognition, *Zhuohao Chen, Zeng Li, Yifei Zhang, Chang Liu, Yu Zhou*
154 iTCTSL: Interpretable Tropical Cyclone Track and Intensity Forecasting via Task Sensitive Learning, *Pan Mu, Yuchao Zhu, Shiqi Zhang, Hanting Yan, Jinglin Zhang, Cong Bai*
155 Machine Vision-Oriented Appearance Design: Generate Natural And Robust Textures For 3D Meshes, *Weihang Ran, Qingtian Zhu, Mingdeng Cao, Wei Yuan, Isao Echizen, Yinqiang Zheng*
156 Bridge Your Fields: MeteoNet for Efficient Non-Uniform Meteorological Field Reconstruction, *Xuanming Jiang, Baoyi An, Dingyu Nie, Haoyu Ren, Zhengwei Zou, Yizhe Yang, Jialie Shen, Zhiwen Jin, Xueming Qian, Zhongyu Yang, Guoshuai Zhao*
157 Catalyst: Out-of-Distribution Detection via Elastic Scaling, *Abid Hassan, Tuan Ngo, Saad Shafiq, Nenad Medvidovic*
158 LUMINA: Learning and Understanding of Multimodal Information for Narrative and Affect-based Virality Prediction, *Jiazhou Lin, Zhongyi Liu, Ying Shi, Zhichun Zhao, Zhuoyu Wang, Yuhang Zhou, Huanling Hu, Guangnan Ye, Mengtian Li, Lei Guo*
159 LOOPE: Learnable Optimal Patch Order for Positional Encoders in Vision Transformers, *Md Abtahi Majeed Chowdhury, Md Rifat Ur Rahman, Akil Ahmad Taki*
160 Watermarking Matters for Deepfake Detection: A Proactive Method for Detecting Forgeries under Conventional Attacks, *Zhiqiu Xia, Furong Mu, Qi Li, Shanshan Zhang, Jie Gui, Chunpeng Wang, Yunan Liu*
161 CTFS : Collaborative Teacher Framework for Forward-Looking Sonar Image Semantic Segmentation with Extremely Limited Labels, *Ping Guo, Chengzhou Li, Guanchen Meng, Qi Jia, Jinyuan Liu, Zhu Liu, Yu Liu, Zhongxuan Luo, Xin Fan*
162 TPTransformer: Tensor–Tensor Product Transformer for Hyperspectral Image Super-Resolution, *Honghui Xu, Chuangjie Fang, Yiqun Meng, Jiawei Jiang, Sixian Chan, Shiqing Zhang, Jianwei Zheng*
163 Co-Adaptive Graph Learning Through Coupled Spectral Refinement for 3D Anomaly Detection, *Hanvitha Saraswathi Mukkamala, Arun K Pujari*
164 The Mechanics of CNN Filtering with Rectification, *Liam Frija-Altarac, Matthew Toews*
165 AndroidLong: LLM-based Android Agents Struggle with Long Looping Tasks, *Xinghan Liu, Xiao Liu, Yifan Xu, Jiaqi Fu, Jiayu Huang, Yixuan Liu, Yuxiao Dong, Jie Tang*
166 Multimodal Large Language Models as Image Classifiers, *Nikita Kisel, Illia Volkov, Klara Janouskova, Jiri Matas*
167 Beyond Recognition: Evaluating Visual Perspective Taking in Vision Language Models, *Gracjan Góral, Alicja Ziarko, Piotr Miłoś, Michał Nauman, Maciej Wołczyk, Michał Kosiński*
168 Exploring the best way for UAV visual localization under Low-altitude Multi-view Observation Condition: a Benchmark, *Yibin Ye, Xichao Teng, Shuo Chen, Leqi Liu, Kun Wang, Xiaokai Song, Zhang Li*
169 Benchmarking Layout-Guided Diffusion Models through Unified Semantic-Spatial Evaluation in Closed and Open Settings, *Luca Parolari, Nicla Faccioli, Lamberto Ballan*
170 RefDrone: A Challenging Benchmark for Referring Expression Comprehension in Drone Scenes, *Zhichao Sun, Yepeng Liu, Zhiling Su, Huachao Zhu, Yuliang Gu, Yuda Zou, Zelong Liu, Gui-Song Xia, Bo Du, Yongchao Xu*
171 Vision Language Models are Confused Tourists, *Patrick Amadeus Irawan, Ikhlasul Akmal Hanif, Muhammad Dehan Al Kautsar, Genta Indra Winata, Fajri Koto, Alham Fikri Aji*
172 LenghuSky-8: An 8-Year All-Sky Cloud Dataset with Star-Aware Masks and Alt-Az Calibration for Segmentation and Nowcasting, *Yicheng Rui, Xiao-Wei Duan, Licai Deng, Fan Yang, Zhengming Dang, Zhengjun Du, Junhao Peng, Wenhao Chu, Umut Mahmut, Kexin Li, Yiyun Wu, Fabo Feng*
173 Through the PRISM: Principle-Aware, Interpretable, and Multi-Scale Evaluation of Visual Designs, *Mona Gandhi, K.J. Joseph, Srinivasan Parthasarathy, Sayan Nag*
174 AEGIS: Exploring the Limit of World Knowledge Capabilities for Unified Multimodal Models, *Jintao Lin, Bowen Dong, Weikang Shi, Chenyang Lei, Suiyun Zhang, Rui Liu, Xihui Liu*
175 Name That Part: 3D Part Segmentation and Naming, *Soumava Paul, Prakhar Kaushik, Ankit Vaidya, Anand Bhattad, Alan Yuille*
176 Do MLLMs Exhibit Human-like Perceptual Behaviors? HVSBench: A Benchmark for MLLM Alignment with Human Perceptual Behavior, *Jiaying Lin, Shuquan Ye, Dan Xu, Wanli Ouyang, Rynson W. H. Lau*
177 Memorization in 3D Shape Generation: An Empirical Study, *Shu Pu, Boya Zeng, Kaichen Zhou, Mengyu Wang, Zhuang Liu*
178 Shape and Texture Recognition in Large Vision-Language Models, *Sagi Eppel, Mor Bismut, Alona Strugatski*
179 U-SEG: Uncertainty in SEGmentation - A systematic multi-variable exploration, *Michael Smith, Frank P. Ferrie*
180 UDVideoQA: A Traffic Video Question Answering Dataset for Multi-Object Spatio-Temporal Reasoning in Urban Dynamics, *Joseph Raj Vishal, Nagasiri Poluri, Katha Naik, Rutuja Patil, Kashyap Hegde Kota, Krishna Vinod, Prithvi Jai Ramesh, Mohammad Farhadi, Yezhou Yang, Bharatesh Chakravarthi*
181 GOVTrack: Towards Generative Open-Vocabulary Multi-Object

Tracking, *Zekun Qian, Ruize Han, Zhixiang Wang, Liang Wan, Wei Feng*
182 Towards Text-Guided Attribute-Disentangled Multimodal Representation Learning, *Yibing Wei, Sudeep Katakol, Manuel Brack, Jinhong Lin, Haoyue Bai, Yu-Teng Li, Richard Zhang, Eli Shechtman, Hareesh Ravi, Ajinkya Kale*
183 The DeepSpeak Dataset, *Sarah Barrington, Maty Bohacek, Hany Farid*
184 AndroidLens: Long-latency Evaluation with Nested Sub-targets for Android GUI Agents, *Yue Cao, Yingyao Wang, Pi Bu, Jingxuan Xing, Wei Jiang, Zekun Zhu, Junpeng Ma, Sashuai Zhou, Tong Lu, Jun Song, Yu Cheng, Yuning Jiang, Bo Zheng*
185 A2Z-10M+: Geometric Deep Learning with A-to-Z BRep Annotations for AI-Assisted CAD Modeling and Reverse Engineering, *Pritham K Jena, Bhavika Baburaj, Tushar Anand, Vedant Dutta, Vineeth Ulavala, Sk Aziz Ali*
186 MolRecBench-Wild: A Real-World Benchmark for Optical Chemical Structure Recognition, *Haote Yang, Hui Wang, Chen Zhu, Jingchao Wang, Linye Li, Hongbin Lai, Huijie Ao, Yongxuan Lv, Jiang Wu, Jiaxing Sun, Lua Chen, Yuanyuan Cao, Ruijie Zhang, Shengxin Lu, Lijun Wu, Bin Wang, Conghui He*
187 FinChart-Multimodal: A Dataset for Context-Injected Financial Chart Understanding with Aligned OHLCV Time Series, *Devansh Garg*
188 THEval. Evaluation Framework for Talking Head Video Generation, *Nabyl Quignon, Baptiste Chopin, Yaohui Wang, Antitza Dantcheva*
189 PolyReal: A Benchmark for Real-World Polymer Science Workflows, *Wanhao Liu, Weida Wang, Jiaqing Xie, Suorong Yang, Jue Wang, Benteng Chen, Guangtao Mei, Zonglin Yang, Shufei Zhang, Yuchun Mo, Lang Cheng, Jin Zeng, Houqiang Li, Wanli Ouyang, Yuqiang Li*
190 OutSafe-Bench: A Benchmark for Multimodal Offensive Content Detection in Large Language Models, *Yuping Yan, Yuhan Xie, Yuanshuai Li, Yingchao Yu, Lingjuan Lyu, Yaochu Jin*
191 PureSpace: A Benchmark for Abstract Spatial Reasoning in Vision-Language Models, *Jinkai Li, Zhenliang Zhang, Lifeng Fan, Wei Wang*
192 Native Visual Understanding: Resolving Resolution Dilemmas in Vision-Language Models, *Junbo Niu, Yuanhong Zheng, Ziyang Miao, Hejun Dong, Chunjiang Ge, Hao Liang, Ma Lu, Bohan Zeng, Qiahao Zheng, Conghui He, Wentao Zhang*
193 Beyond 3D Geometry: M3FD, a Large-Scale Dataset and Benchmark for Multimodal 3D Perceptual Understanding, *Huan Hu, Ping Chen, Zezhou Chen, Zhaoxiang Liu, Zipeng Wang, Xiang Liu, Xin Wang, Kai Wang, Shiguo Lian*
194 Paper2SysArch: Structure-Constrained System Architecture Generation from Scientific Papers, *Ziyi Guo, Zhou Liu, Wentao Zhang*
195 WildRelight: A Real-World Dataset and Benchmark for Single-Image Relighting, *Lezhong Wang, Mehmet Onurcan Kaya, Siavash Arjomand Bigdeli, Jeppe Revall Frisvad*
196 EgoTL: Egocentric Think-Aloud Chains for Long-Horizon Tasks, *Lulin Liu, Dayou Li, Yiqing Liang, Sicong Jiang, Hitesh Vijay, Hezhen Hu, Xuhai Xu, Zirui Liu, Srinivas Shakkottai, Manling Li, Zhiwen Fan*
197 VibraVerse: A Large-Scale Geometry–Acoustics Alignment Dataset for Physically-Consistent Multimodal Learning, *Bo Pang, Chenxi Xu, Jierui Ren, Guoping Wang, Sheng Li*
198 When Harmful Content Goes Invisible: Unveiling Perception Failure of LVLMs with CAMOUHARMTI, *Yanhui Li, Qi Zhou, Zhihong Xu, Huizhong Guo, Wenhai Wang, Dongxia Wang*
199 Step-CoT: Stepwise Visual Chain-of-Thought for Medical Visual Question Answering, *Lin Fan, Yafei Ou, Zhipeng Deng, Pengyu Dai, Chongxian Hou, Jiale Yan, Yaqian Li, Kaiwen Long, Xun Gong, Masayuki Ikebe, Yefeng Zheng*
200 Real-IAD MVN: A Multi-View Normal Vector Dataset and Benchmark for High-Fidelity Industrial Anomaly Detection, *Wenbing Zhu, Jianing Liang, Linjie Cheng, Yurui Pan, Zhuhao Chen, Qingwang Yan, Yudong Cheng, Jianghui Zhang, Mingmin Chi, Bo Peng*
201 The Unwritten Benchmark: A New Challenge for Multimodal Machine Learning in Abstract Perceptual Reasoning, *Garima Arya Yadav, Nilay Yilmaz, Yezhou Yang*
202 Unifying Scientific Communication: Fine-Grained Correspondence Across Scientific Media, *Megha Mariam K.M, Vineeth N. Balasubramanian, C.V. Jawahar*
203 MathAll: A Real-World Benchmark for Mathematical Reasoning and Cross-Modal Understanding Evaluation in Omni-MLLMs, *Zhilin Lin, Zhihui Zhang, Shiliang Sun, Jing Zhao, Hao Yang*
204 Safe-LLaVA: A Privacy-Preserving Vision Language Dataset and Benchmark for Biometric Safety, *Younggun Kim, Sirnam Swetha, Fazil Kagdi, Mubarak Shah*
205 DR-DPO: Dual-Regularized DPO for Efficient Dataset Condensation, *Haiduo Huang, Jiangcheng Song, Yadong Zhang, Guansu Wang, Pengju Ren*
206 DrawingVQA: A Real-World Benchmark for Multi-Depth Visual–Textual Reasoning on Construction Drawings, *Yoonhwa Jung, Junryu Fu, Mani Golparvar-Fard*
207 SciPostGen: Bridging the Gap between Scientific Papers and Poster Layouts, *Shun Inadumi, Shohei Tanaka, Tosho Hirasawa, Atsushi Hashimoto, Koichiro Yoshino, Yoshitaka Ushiku*
208 Some Modalities are More Equal Than Others: Decoding and Architecting Multimodal Integration in MLLMs, *Tianle Chen, Chaitanya Chakka, Arjun Reddy Akula, Xavier Thomas, Deepti Ghadiyaram*
209 Towards Reliable Human Evaluations in Gesture Generation: Insights from a Community-Driven State-of-the-Art Benchmark, *Rajmund Nagy, Hendric Voss, Thanh Hoang-Minh, Mihail Tsakov, Teodor Nikolov, Zeyi Zhang, Tenglong Ao, Sicheng Yang, Shaoli Huang, Yongkang Cheng, M. Hamza Mughal, Rishabh Dabral, Kiran Chhatre, Christian Theobalt, Libin Liu, Stefan Kopp, Rachel McDonnell, Michael Neff, Taras Kucherenko, Youngwoo Yoon, Gustav Eje Henter*
210 SuperGlasses: Benchmarking Vision Language Models as Intelligent Agents for AI Smart Glasses, *Zhuohang Jiang, Xu Yuan, Haohao Qu, Shanru Lin, Kanglong Liu, Wenqi Fan, Li Qing*
211 InEdit-Bench: Benchmarking Intermediate Logical Pathways for Intelligent Image Editing Models, *Zhiqiang Sheng, Xumeng Han, Zhiwei Zhang, Zenghui Xiong, Yifan Ding, Aoxiang Ping, Xiang Li, Tong Guo, Yao Mao*
212 VEBench: Benchmarking Large Multimodal Models for Real-world Video Editing, *Andong Deng, Dawei Du, Zhenfang Chen, Wen Zhong, Fan Chen, Guang Chen, Chia-Wen Kuo, Longyin Wen, Chen Chen, Sijie Zhu*
213 CrowdVerse: A Bidirectional Reality-Calibrated Benchmark for Crowd Understanding and Simulation, *Pingrui Lai, Yanshan Zhou, Zihao Xie, Hua Yang*
214 Can Language Models Understand mmWave Data? Benchmarking Large Language Models for mmWave Radar-Based Human Understanding, *Jeongwan Shin, Jaehyeon Kim, Donguk Ko, Jaeho Choi*
215 From Static Snapshots to Dynamic Trajectories: Evaluating and Enhancing the Learning Pathways of Multimodal Large Language Models, *Yukang Feng, Wenxiao Wu, Jianwen Sun, Chuanhao Li, Fanrui Zhang, Zizhen Li, Jiaxin Ai, Sizhuo Zhou, Yifan Chang, Changxin Gao, Shenglin Zhang, Kaipeng Zhang*
216 Evaluating Dataset Watermarking for Fine-Tuning Traceability of Customized Diffusion Models: A Comprehensive Benchmark and Removal Approach, *Xincheng Wang, Hanchi Sun, Wenjun Sun, Kejun Xue, Wangqiu Zhou, Jianbo Zhang, Wei Sun, Dandan Zhu, Xiongkuo Min, Jun Jia, Zhijun Fang*
217 BMD-45: A Large-Scale CCTV Vehicle Detection Dataset for Urban Traffic in Developing Cities, *Akash Sharma, Chinmay Mhatre, Sankalp Gawali, Ruthvik Bokkasam, Brij Sharma, Vishwajeet Pattanaik, Punit Rathore, Raghu Krishnapuram, Vijay Gopal Kovvali, Anirban Chakraborty, Yogesh Simmhan*
218 SciGA: A Comprehensive Dataset for Designing Graphical Abstracts in Academic Papers, *Takuro Kawada, Shunsuke Kitada, Sota Nemoto, Hitoshi Iyatomi*
219 Splatwizard: A Benchmark Toolkit for 3D Gaussian Splatting Compression, *Xiang Liu, Yimin Zhou, Jinxiang Wang, Yujun Huang, Shuzhao Xie, Shiyu Qin, Mingyao Hong, Jiawei Li, Yaowei Wang, Zhi Wang, Shu-Tao Xia, Bin Chen*
220 See, Hear, and Understand: Benchmarking Audiovisual Human Speech Understanding in Multimodal Large Language Models, *Le Thien Phuc Nguyen, Zhuoran Yu, Samuel Low Yu Hang, Subin An, Jeongik Lee, Yohan Ban, SeungEun Chung, Thanh-Huy Nguyen, JuWan Maeng, Soochahn Lee, Yong Jae Lee*
221 Unlocking ImageNet's Multi-Object Nature: Automated Large-Scale Multilabel Annotation, *Junyu Chen, Md Yousuf Harun, Christopher Kanan*

222 CATS-V2V: A Real-World Vehicle-to-Vehicle Cooperative Perception Dataset with Complex Adverse Traffic Scenarios, *Hangyu Li, Bofeng Cao, Zhaohui Liang, Wuzhen Li, Juyoung Oh, Yuxuan Chen, Shixiao Liang, Hang Zhou, Chengyuan Ma, Jiaxi Liu, Zheng Li, Peng Zhang, Keke Long, Maolin Liu, Jackson Jiang, Chunlei Yu, Shengxiang Liu, Hongkai Yu, Xiaopeng Li*

223 Seeing the Abstract: A Benchmark for Visual-Only Metaphor Understanding in Multimodal Large Language Models, *Shan Zhao, Zhao Yang, Tianwei Yan, Yusong Gong, Qian Wan, Shizhao Chen, Shezheng Song, Chengyu Wang, Meng Wang*

224 Cross-Dimensional Forgery Pattern Extraction for Generalizable Forgery Localization Framework, *Yilin Wang, Dawei Luo, Shuai Chen, Feng Xu, Jiachi Wang, Zunlei Feng, Yijun Bei*

225 Reliable Test-time Adaptation Via Evidential Uncertainty Modeling in Vision–Language Models, *Yiwei You, Zan Chen, Bo Wang, Xiaofei Zhou*

226 SPOT: Sparsification with Attention Dynamics via Token Relevance in Vision Transformers, *Oded Schlesinger, Amirhossein Farzam, J. Matias Di Martino, Guillermo Sapiro*

227 Do LLMs and VLMs Share Reasoning Neurons? Evidence and Mechanisms of Cross-Modal Transfer, *Chenhang Cui, An Zhang, Yuxin Chen, Gelei Deng, Jingnan Zheng, Zhenkai Liang, Xiang Wang, Tat-Seng Chua*

228 Debiased One-Shot NAS Via Density-Aware Sampling, *Mehraveh Javan Roshtkhari, Matthew Toews, Marco Pedersoli*

229 PSLIF: A Primary-Supplementary LIF Neuron for Spiking Neural Networks, *Jie Guo, JunXiang Wu, Nan An, Zhen Zhang, Shuiying Xiang, Mingjin Zhang, Yunsong Li, Yu'e Gao*

230 SHIELD: Secure Hypernetworks for Incremental Expansion Learning Defense, *Patryk Krukowski, Łukasz Gorczyca, Piotr Helm, Kamil Książek, Przemysław Spurek*

231 Eigen-Value: Efficient Domain-Robust Data Valuation Via Eigenvalue-Based Approach, *Youngjun Choi, Joonseong Kang, Sungjun Lim, Kyungwoo Song*

232 In2CLR: Joint Intra-Inter Curriculum Learning with Review for Degraded Fake Image Detection, *Yunxuan Li, Bohao Liu, Yanxia Wu, Rongsheng Li*

233 HAMSA: Scanning-Free Vision State Space Models via SpectralPulseNet, *Badri N Patro, Vijay S Agneeswaran*

234 Latent Domain Modeling Improves Robustness to Geographic Shifts, *Ruth Crasto, Esther Rolf*

235 Any-Class Presence Likelihood for Robust Multi-Label Classification with Abundant Negative Data, *Dumindu Tissera, Omar Awadallah, Muhammad Umair Danish, Ayan Sadhu, Katarina Grolinger*

236 VideoMatGen: PBR Materials through Joint Generative Modeling, *Jon Hasselgren, Miloš Hašan, Zheng Zeng, Jacob Munkberg*

237 TransKV: A Data-Driven Pruning Method for Large Foundation Models, *Guangning Xu, Fanxu Meng, Ruijie Zhou, Michael K Ng, Wenjie Pei, Muhan Zhang*

238 Rich Feature Learning via Diversification, *Xi Leng, Yongqiang Chen, Xiaoying Tang, Yatao Bian*

239 Image Classification Using CNN-QNN Hybrid Model with Optimized Correlated Features, *Minseo Seong, Youngwook Kim*

240 Dual Strategies for Test-Time Adaptation, *Nam Nguyen Phuong, Duc Nguyen The Minh, Phi Le Nguyen, Ehsan Abbasnejad, Minh Hoai*

241 CPUBone: Efficient Vision Backbone Design for Devices with Low Parallelization Capabilities, *Moritz Nottebaum, Matteo Dunnhofer, Christian Micheloni*

242 FLToM: Robust Federated Learning with Theory-of-Mind Structure, *Tianshu Xiao, Liu Yang, Sichang Guo, Qilong Wang, Qinghua Hu*

243 FedCVC: Federated Primal-Dual Learning with Client-Driven Virtual Compensation for Mitigating Dual Drift, *Jinshan Lai, Tingxuan Huang, Baoyang Jiang, Liuyu Xiang, Qiang Ma, Jianwei Hu*

244 Q-MambaIR: Accurate Quantized Mamba for Efficient Image Restoration, *Yujie Chen, Haotong Qin, Zhang Zhang, Michele Magno, Luca Benini, Yawei Li*

245 PHATE-Net: Differentiable Pseudotime Learning for Trustworthy Disease Trajectories in PET, *Yixin Chen, Yan Wang, Wenrui Shao, Zhaoheng Xie*

246 PoM: A Linear-Time Replacement for Attention with the Polynomial Mixer, *David Picard, Nicolas Dufour, Lucas Degeorge, Arijit Ghosh, Davide Allegro, Tom Ravaud, Yohann Perron, Corentin Sautier, Zeynep Sonat Baltaci, Fei Meng, Syrine Kalleli, Marta López-Rauhut, Thibaut Loiseau, Ségolène Albouy, Raphael Baena, Elliot Vincent, Loic Landrieu*

247 Qinling-GFFE: A Novel Station-based Benchmark and Graph-Frequency Fusion Enhancer for Precipitation Forecasting, *Zhenhe Liang, Congqi Cao, Lanshu Hu, Liujie Pan*

248 Deep Feedback ConvNets by Embedding the Working Memory Module for Image Classification, *Lulu Fang, Jiaxiang Qin, Ruiheng Yan, Ning Pan, Haihua Liu, Xinxin Chen*

249 Channel Correlation Loss for Binary Neural Networks, *Xindi Zuo, Wei Zhang, Hai Yu, Zhiliang Zhu*

250 MegAD: An Expert in Meta-Learning Guided Few-Shot Anomaly Detection, *Xinying Li, Junfeng Jing, Tong Wu, Tian Gao, Zhihong Sheng*

251 SGST-Transformer: A Spherical Geometry-Aware Spatio-Temporal Transformer for 360° Video Saliency Prediction, *Kao Zhang, Tao Song, Zhihua Hu, Ming Li, Xin Ding*

252 From Drops to Grid: Noise-Aware Spatio-Temporal Neural Process for Rainfall Estimation, *Rafael Pablos Sarabia, Joachim Nyborg, Morten Birk, Ira Assent*

253 AdaPerceiver: Transformers with Adaptive Width, Depth, and Tokens, *Purvish Jajal, Nicholas John Eliopoulos, Benjamin Shiue-Hal Chou, George K Thiruvathukal, Yung-Hsiang Lu, James C. Davis*

254 Texture-Guided Multiscale Cross-Modal Fusion for AI-Generated Image Quality Assessment, *Qinlin Hu, Mingliang Zhou, Xingran Liao*

255 Res2SPDNet: Multi-Granularity SPD Matrix Residual Learning for Signal Classification, *Shenghui Yue, Rui Wang, Tianyang Xu, Tao Zhou, Xiao-Jun Wu, Josef Kittler*

256 From Navigation to Refinement: Revealing the Two-Stage Nature of Flow-based Diffusion Models through Oracle Velocity, *Haoming Liu, Jinnuo Liu, Yanhao Li, Liuyang Bai, Yunkai Ji, Yuanhe Guo, Shenji Wan, Hongyi Wen*

257 MambaEye: A Size-Agnostic Visual Encoder with Causal Sequential Processing, *Changho Choi, Minho Kim, Jinkyu Kim*

258 Spectral-Aware Adaptive Convolution for Fine-Grained Cross-View Visual Localization, *Linsi Wu, Gang Shen, Xuefei Lv, Chenglong Wu, Yuru Pei*

259 Context-Aware Semantic Segmentation via Stage-Wise Attention, *Antoine Carreaud, Elias Naha, Arthur Chansel, Nina Lahellec, Jan Skaloud, Adrien Gressin*

260 MFI-ResNet: Efficient ResNet Architecture Optimization via MeanFlow Compression and Selective Incubation, *Nuolin Sun, Linyuan Wang, Haonan Wei, Lei Li, Bin Yan*

261 AlphaMerging: Orthogonal Subspace Projection of Task Vectors to Reduce Task Interference for Multi-Task Model Merging, *Zuchi Bazarvaani, Seung-Ho Lee, Jeongmin Ahn, Donghyeon Jeon, Inho Kang, Seung-Hoon Na*

262 Rethinking Compact (<1M) Vision Models: Balancing Accuracy and Speed through Multi-Path Atrous Convolutions, *Christos Kyrkou*

263 Hi3Doc: Hierarchical Tri-Level Representations for Multimodal Long-Document Understanding, *Wanying Zhou, Zhuo Chen, Jianzhi Lu, Chenxi Ma, Weimin Tan, Bo Yan*

264 LongDocSpan: Extending LVLMs for Long Document Understanding, *Junwei Liu, Xiong Wang, Junchao Wu, Yefeng Liu, Congyun Jin, Jiangwei Lao, Junjie Wang, Derek F. Wong, Zhihong Lu, Jian Wang, Ping Wang*

265 M-DocSum: Do LVLMs Genuinely Comprehend Interleaved Image-Text in Document Summarization?, *Haolong Yan, Kaijun Tan, Yeqing Shen, Xin Huang, Jia Wang, Zheng Ge, Xiangyu Zhang, Si Li, Daxin Jiang*

266 InstructTable: Improving Table Structure Recognition Through Instruction, *Boming Chen, Zining Wang, Zhentao Guo, Jianqiang Liu, Chen Duan, Yu Gu, Kai zhou, Pengfei Yan*

267 SciPostLayoutTree: A Dataset for Structural Analysis of Scientific Posters, *Shohei Tanaka, Atsushi Hashimoto, Yoshitaka Ushiku*

268 Efficient Document Parsing via Parallel Token Prediction, *Lei Li, Ze Zhao, Meng Li, Zhongwang Lun, Yi Yuan, Xingjing Lu, Zheng Wei, Jiang Bian, Zang Li*

269 ChartAgent: A Chart Understanding Framework with Tool Integrated Reasoning, *Boran Wang, Xinming Wang, Yi Chen, Xiang Li, Jian Xu, Jing Yuan, Cheng-Lin Liu*
270 RedVTP: Training-Free Acceleration of Diffusion Vision-Language Models Inference via Masked Token-Guided Visual Token Pruning, *Jingqi Xu, Jingxi Lu, Chenghao Li, Sreetama Sarkar, Souvik Kundu, Peter A Beerel*
271 FREE-Switch: Frequency-Based Dynamic LoRA Switch for Style Transfer, *Shenghe Zheng, Minyu Zhang, Tianhao Liu, Hongzhi Wang*
272 FedVG: Gradient-Guided Aggregation for Enhanced Federated Learning, *Alina Devkota, Jacob Thrasher, Donald Adjeroh, Binod Bhattarai, Prashnna k. Gyawali*
273 What and Where to Adapt: Structure–Semantics Co-Tuning for Machine Vision Compression via Synergistic Adapters, *Shaobo Liu, Haobo Xiong, Kai Liu, Yuna Lin*
274 A Comprehensive Study on Visual Token Redundancy for Discrete Diffusion-based Multimodal Large Language Models, *Duo Li, Zuhao Yang, Xiaoqin Zhang, Ling Shao, Shijian Lu*
275 GM-Skip: Metric-Guided Transformer Block Skipping for Efficient Vision-Language Models, *Lianming Huang, Haibo Hu, Qiao Li, Xin He, Nan Guan, Chun Jason Xue*
276 Dyna-ViT: Parameter-Free Pre-Encoder Token Pruning for Efficient Vision Transformers, *Syeda Fiza Rubab, Arslan Abdul Ghaffar, Malik Junaid Jami Gul, Sheriff Murtala, Ingyu Lee, Gyu Sang Choi*
277 Efficient Long-Context Modeling in Diffusion Language Models via Block Approximate Sparse Attention, *Wenhu Zhang, Yiming Wu, Huanyu Wang, YaoYang Liu, Huanzhang Dou, Senqiao Yang, Sitong Wu, Hanbin Zhao, Jiaya Jia*
278 MaMe: Matrix-Based Token Merging, *Simin Huo, Ning Li*
279 Tiny Inference-Time Scaling with Latent Verifiers, *Davide Bucciarelli, Evelyn Turri, Lorenzo Baraldi, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara*
280 DaMN: Deleting and Migrating Normalization Layers from Transformers, *Alexey Ryabykin, Irina Zhelavskaya, Egor Shvetsov, Alexey Rukhovich, Nikita Okhotnikov, Artem Khrapov, Evgeny Burnaev, Vladimir Mikhailovich Kryzhanovskiy*
281 Enriching Knowledge Distillation with Cross-Modal Teacher Fusion, *Amir M. Mansourian, Amir Mohammad Babaei, Shohreh Kasaei*
282 UNIFORM: Unifying Knowledge from Large-scale and Diverse Pre-trained Models, *Yimu Wang, Weiming Zhuang, Chen Chen, Jiabo Huang, Jingtao Li, Lingjuan Lyu*
283 ProGIC: Progressive and Lightweight Generative Image Compression with Residual Vector Quantization, *Hao Cao, Chengbin Liang, Wenqi Guo, Zhijin Qin, Jungong Han*
284 MipKV: A Sparsify-then-Recover Paradigm for Accelerating Large Vision-Language Model Pre-Filling, *Junming Zhang, Yifei Ji, Yongxuan Han, Zhenzhe Zheng*
285 Mix-to-Max: Optimizing Data Mixtures for Peak Vision-Language Efficiency, *Erwei Zhao, Haijin Zeng, Weiwei Xiao, Shijie Cao, Qiben Shan, Shaocong Wu, Jingyong Su, Jie Liu*
286 INTERLACE: Interleaved Layer Pruning and Efficient Adaptation in Large Vision-Language Models, *Parsa Madinei, Ryan Solgi, Ziqi Wen, Jonathan Skaza, Miguel Eckstein, Ramtin Pedarsani*
287 JetViT: Efficient High-Resolution Vision Transformer with Post-Training Attention Search, *Dongyun Zou, Zhuoyang Zhang, Junyu Chen, Wenkun He, Qinhe Peng, Hanrong Ye, Yao Lu, Hongxu Yin, Yu Wang, Song Han, Han Cai*
288 SLAD : Shared LoRA Adapters for Task Specific Distillation, *Reda Bensaid, Yassir Bendou, Vincent Gripon, François Leduc-Primeau*
289 D4C: Data-Free Quantization for Contrastive Language-Image Pre-Training Models, *Wenlun Zhang, Yunshan Zhong, Zihao Ding, Xinyu Li, Kentaro Yoshioka*
290 ELSA: Exact Linear-Scan Attention for Fast and Memory-Light Vision Transformers, *Chih-Chung Hsu, Xin-Di Ma, Wo-Ting Liao, Chia-Ming Lee*
291 MPM: Mutual Pair Merging for Efficient Vision Transformers, *Simon Ravé, Pejman Rasti, David Rousseau*
292 Generalized Neighborhood Attention: Multi-dimensional Sparse Attention at the Speed of Light, *Ali Hassani, Fengzhe Zhou, Aditya Kane, Jiannan Huang, Chieh-Yun Chen, Min Shi, Steven Walton, Markus Hoehnerbach, Vijay Thakkar, Mikhail Isaev, Qinsheng Zhang, Bing Xu, Haicheng Wu, Wen-mei Hwu, Ming-Yu Liu, Humphrey Shi*
293 AlignFL: Adaptive Learning and Intelligent Generation of Networks for Federated Learning, *Qilin Xiang, Qilin Fan, Xinrui Li, Tianfu Wang, Shuting Qiu, Yue Niu*
294 Beyond Loss Values: Robust Dynamic Pruning via Loss Trajectory Alignment, *Huaiyuan Qin, Muli Yang, Gabriel James Goenawan, Kai Wang, Zheng Wang, Peng Hu, Xi Peng, Hongyuan Zhu*
295 Positive Divide and Negative Discrepancy: A New Perspective on Multi-Label Logit Distillation, *Cong Li, Gong Cheng*
296 Beyond Accuracy: An Empirical Study of Perception Stability in Multimodal Large Language Models, *Feng Chen, Chenhui Gou, Yefei He, Yang Yang, Bohan Zhuang, Qi Wu*
297 Event-VStream: Event-Driven Real-Time Understanding for Long Video Streams, *Zhenghui Guo, Yuanbin Man, Junyuan Sheng, Bowen Lin, Ahmed Ahmed, Bo Jiang, Boyuan Zhang, Miao Yin, Sian Jin, Omprakash Gnawali, Chengming Zhang*
298 M^3A Policy: Mutable Material Manipulation Augmentation Policy through Photometric Re-rendering, *Jiayi Li, Yuxuan Hu, Haoran Geng, Xiangyu Chen, Chuhao Zhou, Ziteng Cui, Jianfei Yang*
299 AOMGen: Photoreal, Physics-Consistent Demonstration Generation for Articulated Object Manipulation, *Yulu Wu, Jiujun Cheng, Haowen Wang, Dengyang Suo, Pei Ren, Qichao Mao, Shangce Gao, Yakun Huang*
300 Environmental Understanding Vision-language Model for Embodied Agent, *Jinsik Bang, Jaeyeon Bae, Donggyu Lee, Siyeol Jung, Taehwan Kim*
301 DEGround: An Effective Baseline for Ego-centric 3D Visual Grounding With a Homogeneous Framework, *Yani Zhang, Dongming Wu, Hao Shi, Yingfei Liu, Tiancai Wang, Xingping Dong*
302 Prune-Then-Plan: Step-Level Calibration for Stable Frontier Exploration in Embodied Question Answering, *Noah Frahm, Prakrut Patel, Yue Zhang, Shoubin Yu, Mohit Bansal, Roni Sengupta*
303 Think Twice, Act Once: Verifier-Guided Action Selection For Embodied Agents, *Nishad Singhi, Christian Bialas, Snehal Jauhri, Vignesh Prasad, Georgia Chalvatzaki, Marcus Rohrbach, Anna Rohrbach*
304 Plug-and-Think: Structured Reasoning for Vision–Language–Action Models, *Kaikai Wei, Di wen, Xinhai Li, Senwei Xiang*
305 World Model Robustness via Surprise Recognition, *Geigh Zollicoffer, Tanush Chopra, Mingkuan Yan, Xiaoxu Ma, Kenneth Eaton, Mark Riedl*
306 PlanGS: Active 3D Gaussian Reconstruction with Real-Time Planning, *Wenxiang Xie, Anpei Chen, Haoming Yu, Yujun Shen, Weiwei Xu*
307 A Simple Framework for Visual Navigation, *Faith Johnson, Bryan Bo Cao, Shubham Jain, Ashwin Ashok, Kristin Dana*
308 Event-Based Optical Flow Leveraging Precise Event Timing, *Hugh Greatorex, Elisabetta Chicca*
309 Generative Event Pretraining with Foundation Model Alignment, *Jianwen Cao, Jiaxu Xing, Nico Messikommer, Davide Scaramuzza*
310 HelixTrack: Event-Based Tracking and RPM Estimation of Propeller-like Objects, *Radim Spetlik, Michal Pliska, Vojtěch Vrba, Jiří Matas*
311 PEPR: Privileged Event-based Predictive Regularization for Domain Generalization, *Gabriele Magrini, Federico Becattini, Niccolò Biondi, Pietro Pala*
312 Unleashing the Potential of Event-Based Stereo Via Coarse-to-Fine Bio-Inspired Regression, *Haihao Zhang, Siwei Dong, Jianing Li, Rui Zhao, Yunjian Zhang, Geng Qin, Lin Zhu*
313 Metric-Guided Feature Fusion of Visual Foundation Models for Segmentation Tasks, *Yachan Guo, Jose Lu Gómez, Danna Xue, Yi Xiao, Antonio M. López*
314 An Interpretable Alzheimer's Disease Diagnosis Model via Gray Matter Attention Guided Counterfactual Reasoning, *Pengzhou Chen, Qiling Tang, XinYu Chai, Rong Liu, Zhi Li, Liman Liu*
315 Beyond Semantics: Disentangling Information Scope in Sparse Autoencoders for CLIP, *Yusung Ro, Jaehyun Choi, Junmo Kim*
316 CLIP-Free, Label Free, Unsupervised Concept Bottleneck Models, *Fawaz Sammani, Jonas Fischer, Nikos Deligiannis*
317 Concept-wise Attention for Fine-grained Concept Bottleneck Models, *Minghong Zhong, Guoshuai Zou, Kanghao Chen, Dexia Chen, Ruixuan Wang*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

9:15 - 10:30 Oral Session 1A: Multimodal Vision (Bluebird Ballroom)
🏆 - Award candidate paper

1 A Style is Worth One Code: Unlocking Code-to-Style Image Generation with Discrete Style Space, *Huijie Liu, Shuhao Cui, Haoxiang Cao, Shuai Ma, Kai Wu, Guoliang Kang* 🏆
2 Adversarial Style Optimization: Enhancing VLM Jailbreaks by GRPO-based Stylistic Triggers Optimization, *Bingjun Luo, Jialin Guo, Yue Yao, Xinpeng Ding* 🏆
3 ANTS: Adaptive Negative Textual Space Shaping for OOD Detection via Test-Time MLLM Understanding and Reasoning, *Wenjie Zhu, Yabin Zhang, Xin Jin, Wenjun Zeng, Lei Zhang,*
4 ARGUS: Defending Against Multimodal Indirect Prompt Injection via Steering Instruction-Following Behavior, *Weikai Lu, Ziqian Zeng, Kehua Zhang, Haoran Li, Huiping Zhuang, Ruidong Wang, Cen Chen, Hao Peng,*
5 TEAR: Temporal-aware Automated Red-teaming for Text-to-Video Models, *Jiaming He, Guanyu Hou, Hongwei Li, Zhicong Huang, Kangjie Chen, Yi Yu, Wenbo Jiang, Guowen Xu, Tianwei Zhang*
6 ViT^3: Unlocking Test-Time Training in Vision, *Dongchen Han, Yining Li, Tianyu Li, Zixuan Cao, Ziming Wang, Jun Song, Yu Cheng, Bo Zheng, Gao Huang* 🏆

9:15 - 10:30 Oral Session 1B: Visual Security
(Four Seasons Ballroom)

1 Black-box Membership Inference Attacks on the Pre-training Data of Image-generation Models, *Tao Qi, Huili Wang, Yuanhong Huang, Wendan Wang, Lianchao Zhao, Jinrui Wang, Zichen Qin, Shangguang Wang, Yongfeng Huang* 🏆
2 Data Leakage Detection and De-duplication in Large Scale Geospatial Image Datasets, *Yeshwanth Kumar Adimoolam, Charalambos Poullis, Melinos Averkiou* 🏆
3 RAVEN: Erasing Invisible Watermarks via Novel View Synthesis, *Fahad Shamshad, Nils Lukas, Karthik Nandakumar*
4 LDP-Slicing: Local Differential Privacy for Images via Randomized Bit-Plane Slicing, *Yuanming Cao, Chengqi Li, Wenbo He* 🏆
5 NOWA: Null-space Optical Watermark for Invisible Capture Fingerprinting and Tamper Localization, *Edwin Vargas, Jhon Lopez, Henry Arguello, Ashok Veeraraghavan*
6 Revisiting Geometric Obfuscation with Dual Convergent Lines for Privacy-Preserving Image Queries in Visual Localization, *Jeonggon Kim, Heejoon Moon, Je Hyeong Hong* 🏆

9:15 - 10:30 Oral Session 1C: Efficient Reasoning
(Mile High Ballroom 1A - 2A)

1 Advancing Image Classification with Discrete Diffusion Classification Modeling, *Omer Belhasin, Shelly Golan, Ran El-Yaniv, Michael Elad*
2 Does YOLO Really Need to See Every Training Image in Every Epoch?, *Xingxing Xie, Jiahua Dong, Junwei Han, Gong Cheng*
3 Fine-grained Image Aesthetic Assessment: Learning Discriminative Scores from Relative Ranks, *Zhichao Yang, Jianjie Wang, Zhixianhe Zhang, Pangu Xie, Xiangfei Sheng, Pengfei Chen, Leida Li*
4 NuWa: Deriving Lightweight Class-Specific Vision Transformers for Edge Devices, *Ziteng Wei, Qiang He, Bing Li, Feifei Chen, Hai Jin, Yun Yang* 🏆
5 Plant Taxonomy Meets Plant Counting: A Fine-Grained, Taxonomic Dataset for Counting Hundreds of Plant Species, *Jinyu Xu, Tianqi Hu, Xiaonan Hu, Letian Zhou, Songliang Cao, Meng Zhang, Hao Lu* 🏆
6 Rethinking Dataset Distillation: Hard Truths about Soft Labels, *Priyam Dey, Aditya Sahdev, Sunny Bhati, Konda Reddy Mopuri, Venkatesh Babu Radhakrishnan* 🏆

9:15 - 10:30 Oral Session 1D: Computational Imaging
(Mile High Ballroom 3A - 4A)

1 Customized Fusion: A Closed-Loop Dynamic Network for Adaptive Multi-Task-Aware Infrared-Visible Image Fusion, *Zengyi Yang, Yu Liu, Juan Cheng, Zhiqin Zhu, Yafei Zhang, Huafeng Li* 🏆
2 Dual Band Thermal Videography: Separating Time-Varying Reflection and Emission Near Ambient Conditions, *Sriram Narayanan, Mani Ramanagopal, Srinivasa Narasimhan* 🏆
3 MetaSpectra+: A Compact Broadband Metasurface Camera for Snapshot Hyperspectral+ Imaging, *Yuxuan Liu, Wei Xu, Qi Guo* 🏆
4 Spectrum from Defocus: Fast Spectral Imaging with Chromatic Focal Stack, *M. Kerem Aydin, Yi-Chun Hung, Jaclyn Pytlarz, Qi Guo, Emma Alexander*
5 Towards Photorealistic and Efficient Bokeh Rendering via Diffusion Framework, *Linxiao Shi, Siming Zheng, Zerong Wang, Hao Zhang, Jinwei Chen, Bo Li, Shifeng Chen, Peng-Tao Jiang* 🏆
6 UnReflectAnything: RGB-Only Highlight Removal by Rendering Synthetic Specular Supervision, *Alberto Rota, Mert Kiray, Mert Asim Karaoglu, Patrick Ruhkamp, Elena De Momi, Nassir Navab, Benjamin Busam* 🏆

10:45 - 12:45 Poster Session 1 and Exhibit Hall (ExHall A-F)
❊ - Highlight paper 🏆 - Award candidate paper

1 A Style is Worth One Code: Unlocking Code-to-Style Image Generation with Discrete Style Space, *Huijie Liu, Shuhao Cui, Haoxiang Cao, Shuai Ma, Kai Wu, Guoliang Kang* 🏆
2 Adversarial Style Optimization: Enhancing VLM Jailbreaks by GRPO-based Stylistic Triggers Optimization, *Bingjun Luo, Jialin Guo, Yue Yao, Xinpeng Ding* 🏆
3 ANTS: Adaptive Negative Textual Space Shaping for OOD Detection via Test-Time MLLM Understanding and Reasoning, *Wenjie Zhu, Yabin Zhang, Xin Jin, Wenjun Zeng, Lei Zhang*
4 ARGUS: Defending Against Multimodal Indirect Prompt Injection via Steering Instruction-Following Behavior, *Weikai Lu, Ziqian Zeng, Kehua Zhang, Haoran Li, Huiping Zhuang, Ruidong Wang, Cen Chen, Hao Peng*
5 TEAR: Temporal-aware Automated Red-teaming for Text-to-Video Models, *Jiaming He, Guanyu Hou, Hongwei Li, Zhicong Huang, Kangjie Chen, Yi Yu, Wenbo Jiang, Guowen Xu, Tianwei Zhang*
6 ViT3: Unlocking Test-Time Training in Vision, *Dongchen Han, Yining Li, Tianyu Li, Zixuan Cao, Ziming Wang, Jun Song, Yu Cheng, Bo Zheng, Gao Huang* 🏆
7 Black-box Membership Inference Attacks on the Pre-training Data of Image-generation Models, *Tao Qi, Huili Wang, Yuanhong Huang, Wendan Wang, Lianchao Zhao, Jinrui Wang, Zichen Qin, Shangguang Wang, Yongfeng Huang* 🏆
8 Data Leakage Detection and De-duplication in Large Scale Geospatial Image Datasets, *Yeshwanth Kumar Adimoolam, Charalambos Poullis, Melinos Averkiou* 🏆
9 RAVEN: Erasing Invisible Watermarks via Novel View Synthesis, *Fahad Shamshad, Nils Lukas, Karthik Nandakumar*
10 LDP-Slicing: Local Differential Privacy for Images via Randomized Bit-Plane Slicing, *Yuanming Cao, Chengqi Li, Wenbo He* 🏆
11 NOWA: Null-space Optical Watermark for Invisible Capture Fingerprinting and Tamper Localization, *Edwin Vargas, Jhon Lopez, Henry Arguello, Ashok Veeraraghavan*
12 Revisiting Geometric Obfuscation with Dual Convergent Lines for Privacy-Preserving Image Queries in Visual Localization, *Jeonggon Kim, Heejoon Moon, Je Hyeong Hong* 🏆
13 Advancing Image Classification with Discrete Diffusion Classification Modeling, *Omer Belhasin, Shelly Golan, Ran El-Yaniv, Michael Elad*
14 Does YOLO Really Need to See Every Training Image in Every Epoch?, *Xingxing Xie, Jiahua Dong, Junwei Han, Gong Cheng*
15 Fine-grained Image Aesthetic Assessment: Learning Discriminative Scores from Relative Ranks, *Zhichao Yang, Jianjie Wang, Zhixianhe Zhang, Pangu Xie, Xiangfei Sheng, Pengfei Chen, Leida Li*
16 NuWa: Deriving Lightweight Class-Specific Vision Transformers for Edge Devices, *Ziteng Wei, Qiang He, Bing Li, Feifei Chen, Hai Jin, Yun Yang* 🏆
17 Plant Taxonomy Meets Plant Counting: A Fine-Grained, Taxonomic Dataset for Counting Hundreds of Plant Species, *Jinyu Xu, Tianqi Hu, Xiaonan Hu, Letian Zhou, Songliang Cao, Meng Zhang, Hao Lu* 🏆
18 Rethinking Dataset Distillation: Hard Truths about Soft Labels, *Priyam Dey, Aditya Sahdev, Sunny Bhati, Konda Reddy Mopuri, Venkatesh Babu Radhakrishnan* 🏆
19 Customized Fusion: A Closed-Loop Dynamic Network for Adaptive

- 19 Multi-Task-Aware Infrared-Visible Image Fusion, *Zengyi Yang, Yu Liu, Juan Cheng, Zhiqin Zhu, Yafei Zhang, Huafeng Li*
- 20 Dual Band Thermal Videography: Separating Time-Varying Reflection and Emission Near Ambient Conditions, *Sriram Narayanan, Mani Ramanagopal, Srinivasa Narasimhan*
- 21 MetaSpectra+: A Compact Broadband Metasurface Camera for Snapshot Hyperspectral+ Imaging, *Yuxuan Liu, Wei Xu, Qi Guo*
- 22 Spectrum from Defocus: Fast Spectral Imaging with Chromatic Focal Stack, *M. Kerem Aydin, Yi-Chun Hung, Jaclyn Pytlarz, Qi Guo, Emma Alexander*
- 23 Towards Photorealistic and Efficient Bokeh Rendering via Diffusion Framework, *Linxiao Shi, Siming Zheng, Zerong Wang, Hao Zhang, Jinwei Chen, Bo Li, Shifeng Chen, Peng-Tao Jiang*
- 24 UnReflectAnything: RGB-Only Highlight Removal by Rendering Synthetic Specular Supervision, *Alberto Rota, Mert Kiray, Mert Asim Karaoglu, Patrick Ruhkamp, Elena De Momi, Nassir Navab, Benjamin Busam*
- 25 AVGGT: Rethinking Global Attention for Accelerating VGGT, *Xianbing Sun, Zhikai Zhu, Zhengyu Lou, Bo Yang, Jinyang Tang, Liqing Zhang, He Wang, Jianfu Zhang*
- 26 ManifoldNeuS: Manifold-aware View Optimizability for Pose-Free Neural Surface Reconstruction, *Xinxin Liu, Xue Wang, Guoqing Zhou, Qing Wang*
- 27 LongStream: Long-Sequence Streaming Autoregressive Visual Geometry, *Chong Cheng, Xianda Chen, Tao Xie, Wei Yin, Weiqiang Ren, Qian Zhang, Xiaoyang Guo, Hao Wang*
- 28 RPGFusion: 4D Radar Prior-Guided Multi-Modal Fusion for 3D Detection, *Xin Qiu, Wenjie Liu*
- 29 MoVieS: Motion-Aware 4D Dynamic View Synthesis in One Second, *Chenguo Lin, Yuchen Lin, Panwang Pan, Yifan Yu, Tao Hu, Honglei Yan, Katerina Fragkiadaki, Yadong Mu*
- 30 JRM: Joint Reconstruction Model for Multiple Objects without Alignment, *Qirui Wu, Yawar Siddiqui, Duncan Frost, Samir Aroudj, Armen Avetisyan, Richard Newcombe, Angel X. Chang, Jakob Engel, Henry Howard-Jenkins*
- 31 Inferring Compositional 4D Scenes without Ever Seeing One, *Ahmet Berke Gökmen, Ajad Chhatkuli, Luc Van Gool, Danda Pani Paudel*
- 32 FreeScale: Scaling 3D Scenes via Certainty-Aware Free-View Generation, *Chenhan Jiang, Yu Chen, Qingwen Zhang, Jifei Song, Songcen Xu, Dit-Yan Yeung, Jiankang Deng*
- 33 Complet4R: Geometric Complete 4D Reconstruction, *Weibang Wang, Kenan Li, Zhuoguang Chen, Yijun Yuan, Hang Zhao*
- 34 Unblur-SLAM: Dense Neural SLAM for Blurry Inputs, *Qi Zhang, Denis Rozumny, Francesco Girlanda, Sezer Karaoglu, Marc Pollefeys, Theo Gevers, Martin R. Oswald*
- 35 Learning Compact 3D Representations from Feed-Forward Novel View Synthesis, *Honggyu An, Jaewoo Jung, Mungyeom Kim, Chaehyun Kim, Minkyeong Jeon, Jisang Han, Kazumi Fukuda, Takuya Narihira, Hyunah Ko, Junsu Kim, Sunghwan Hong, Yuki Mitsufuji, Seungryong Kim*
- 36 Fast Spatial Tracking with Visual Geometry Transformer, *Chengjie Huang, Guile Wu, Dongfeng Bai, Bingbing Liu*
- 37 How Much 3D Do Video Foundation Models Encode?, *Zixuan Huang, Xiang Li, Zhaoyang Lv, James M. Rehg*
- 38 MetroGS: Efficient and Stable Reconstruction of Geometrically Accurate High-Fidelity Large-Scale Scenes, *Kehua Chen, Tianlu Mao, Xinzhu Ma, Hao Jiang, Zehao Li, Zihan Liu, Shuqin Gao, Honglong Zhao, Feng Dai, Yucheng Zhang, Zhaoqi Wang*
- 39 RnG: A Unified Transformer for Complete 3D Modeling from Partial Observations, *Mochu Xiang, Zhelun Shen, Xuesong Li, Jiahui Ren, Jing Zhang, Chen Zhao, Shanshan Liu, Haocheng Feng, Jingdong Wang, Yuchao Dai*
- 40 Long-Tail Internet Photo Reconstruction, *Yuan Li, Yuanbo Xiangli, Hadar Averbuch-Elor, Noah Snavely, Ruojin Cai*
- 41 Emergent Outlier View Rejection in Visual Geometry Grounded Transformers, *Jisang Han, Sunghwan Hong, Jaewoo Jung, Wooseok Jang, Honggyu An, Qianqian Wang, Seungryong Kim, Chen Feng*
- 42 Flow3r: Factored Flow Prediction for Scalable Visual Geometry Learning, *Zhongxiao Cong, Qitao Zhao, Minsik Jeon, Shubham Tulsiani*
- 43 MultiBanana: A Challenging Benchmark for Multi-Reference Text-to-Image Generation, *Yuta Oshima, Daiki Miyake, Kohsei Matsutani, Yusuke Iwasawa, Masahiro Suzuki, Yutaka Matsuo, Hiroki Furuta*
- 44 HoloCine: Holistic Generation of Cinematic Multi-Shot Long Video Narratives, *Yihao Meng, Hao Ouyang, Yue Yu, Qiuyu Wang, Wen Wang, Ka Leong Cheng, Hanlin Wang, Shuailei Ma, Yixuan Li, Cheng Chen, Yanhong Zeng, Xing Zhu, Yujun Shen, Huamin Qu*
- 45 Design Your Ad: Personalized Advertising Image and Text Generation with Unified Autoregressive Models, *Yexing Xu, Wei Feng, Shen Zhang, Haohan Wang, Yuxin Qin, Yaoyu Li, Ao Ma, Yuhao Luo, Lu Wang, Xudong Ren, Haoran Wang, Run Ling, Zheng Zhang, Jingjing Lv, Junjie Shen, Ching Law, Longguang Wang, Yulan Guo*
- 46 SketchDeco: Training-Free Latent Composition for Precise Sketch Colourisation, *Chaitat Utintu, Yi-Zhe Song*
- 47 ConsistCompose: Unified Multimodal Layout Control for Image Composition, *Xuanke Shi, Boxuan Li, Xiaoyang Han, Zhongang Cai, Lei Yang, Quan Wang, Dahua Lin*
- 48 A Training-Free Style-Personalization via SVD-Based Feature Decomposition, *Kyoungmin Lee, Jihun Park, Jongmin Gim, Wonhyeok Choi, Kyumin Hwang, Jaeyeul Kim, Sunghoon Im*
- 49 Beyond Patches: Global-aware Autoregressive Model for Multimodal Few-Shot Font Generation, *Haonan Cai, Yuxuan Luo, Zhouhui Lian*
- 50 ImageRAGTurbo: Towards One-step Text-to-Image Generation with Retrieval-Augmented Diffusion Models, *Peijie Qiu, Hariharan Ramshankar, Arnau Ramisa, Amit Kumar K C, René Vidal, Vamsi Salaka, Rahul Bhagat*
- 51 OmniSonic: Towards Universal and Holistic Audio Generation from Video and Text, *Weiguo Pian, Saksham Singh Kushwaha, Zhimin Chen, Shijian Deng, Kai Wang, Yunhui Guo, Yapeng Tian*
- 52 Ar2Can: An Architect and an Artist Leveraging a Canvas for Multi-Human Generation, *Shubhankar Borse, Phuc Pham, Farzad Farhadzadeh, Seokeon Choi, Phong Nguyen, Anh Tran, Sungrack Yun, Munawar Hayat, Fatih Porikli*
- 53 Curriculum Group Policy Optimization: Adaptive Sampling for Unleashing the Potential of Text-to-Image Generation, *Baoteng Li, Xianghao Zang, Xinran Wang, Xiangyu Na, Zhixiang He, Hao Sun, Chi Zhang, Zhongjiang He, Tianwei Cao, Kongming Liang, Zhanyu Ma*
- 54 SplitFlux: Learning to Decouple Content and Style from a Single Image, *Yitong Yang, Yinglin Wang, Changshuo Wang, Yongjun Zhang, Ziyang Chen, Shuting He*
- 55 FontCrafter: High-Fidelity Element-Driven Artistic Font Creation with Visual In-Context Generation, *Wuyang Luo, Chengkai Tan, Chang Ge, Binye Hong, Su Yang, Yongjiu Ma*
- 56 EmoStyle: Emotion-Driven Image Stylization, *Jingyuan Yang, Zihuan Bai, Hui Huang*
- 57 Text-Image Conditioned 3D Generation, *Jiazhong Cen, Jiemin Fang, Sikuang Li, Guanjun Wu, Chen Yang, Taoran Yi, Zanwei Zhou, Zhikuan Bao, Lingxi Xie, Wei Shen, Qi Tian*
- 58 IntroSVG: Learning from Rendering Feedback for Text-to-SVG Generation via an Introspective Generator–Critic Framework, *Feiyu Wang, Jiayuan Yang, Zhiyuan Zhao, Da Zhang, Bingyu Li, Peng Liu, Junyu Gao*
- 59 AnyDoc: Enhancing Document Generation via Large-Scale HTML/CSS Data Synthesis and Height-Aware Reinforcement Optimization, *Jiawei Lin, Wanrong Zhu, Vlad I Morariu, Christopher Tensmeyer*
- 60 Reasoning Diffusion for Unpaired Test Time Out-of-distribution Text-Image to Video Generation, *Zirui Pan, Xin Wang, Yipeng Zhang, Hong Chen, Kecheng Zheng, Wenwu Zhu*
- 61 SpatialReward: Verifiable Spatial Reward Modeling for Fine-Grained Spatial Consistency in Text-to-Image Generation, *Sashuai Zhou, Qiang Zhou, Junpeng Ma, Yue Cao, Ruofan Hu, Ziang Zhang, Xiaoda Yang, Zhibin Wang, Jun Song, Cheng Yu, Bo Zheng, Zhou Zhao*
- 62 STAGE: Storyboard-Anchored Generation for Cinematic Multi-shot Narrative, *Peixuan Zhang, Zijian Jia, Kaiqi Liu, Shuchen Weng, Si Li, Boxin Shi*
- 63 MTA: Multimodal Task Alignment for BEV Perception and Captioning, *Yunsheng Ma, Burhaneddin Yaman, Xin Ye, Jingru Luo, Feng Tao, Abhirup Mallik, Ziran Wang, Liu Ren*
- 64 β-CLIP: Text-Conditioned Contrastive Learning for Multi-Granular Vision-Language Alignment, *Fatimah Zohra, Chen Zhao, Hani Itani, Bernard Ghanem*

65 SafeRoPE: Risk-specific Head-wise Embedding Rotation for Safe Generation in Rectified Flow Transformers, *Xiang Yang, Feifei Li, Mi Zhang, Geng Hong, Xiaoyu You, Min Yang*
66 FALCON: False-Negative Aware Learning of Contrastive Negatives in Vision-Language Alignment, *Myunsoo Kim, Seongwoong Shim, Byung-Jun Lee*
67 Spatial-Aware VLA Pretraining through Visual-Physical Alignment from Human Videos, *Yicheng Feng, Wanpeng Zhang, Ye Wang, Hao Luo, Haoqi Yuan, Sipeng Zheng, Zongqing Lu*
68 ❊ Training One Model to Master Cross-Level Agentic Actions via Reinforcement Learning, *Kaichen He, Zihao Wang, Muyao Li, Anji Liu, Yitao Liang*
69 Graph2Eval: Automatic Multimodal Task Generation for Agents via Knowledge Graphs, *Yurun Chen, Xueyu Hu, Yuhan Liu, Ziqi Wang, Zeyi Liao, Lin Chen, Feng Wei, Yuxi Qian, Bo Zheng, Keting Yin, Shengyu Zhang*
70 ❊ EMO-R3: Reflective Reinforcement Learning for Emotional Reasoning in Multimodal Large Language Models, *Yiyang Fang, Wenke Huang, Pei Fu, Yihao Yang, Kehua Su, Zhenbo Luo, Jian Luan, Mang Ye*
71 EvoGraph-R1: Self-Evolving Multimodal Knowledge Hypergraphs for Agentic Retrieval, *Jiashi Lin, Changhong Jiang, Xiangru Lin, Ruifei Zhang, Xinyi Zhu, Jiyao Liu, Cheng Tang, Ye Du, Shujian Gao, Junzhi Ning, Lihao Liu, Ziyan Huang, Tianbin Li, Jin Ye, Junjun He*
72 Cross-modal Identity Mapping: Minimizing Information Loss in Modality Conversion via Reinforcement Learning, *Haonan Jia, Shichao Dong, Xin Dong, Zenghui Sun, Jin Wang, Jinsong Lan, Xiaoyong Zhu, Bo Zheng, Kaifu Zhang*
73 Downscaling Intelligence: Exploring Perception and Reasoning Bottlenecks in Small Multimodal Models, *Mark Endo, Serena Yeung-Levy*
74 ❊ Stabilizing Feature Geometry in Noisy Pretrained Models for Robust Downstream Tasks, *Quanyu Zhang, Zhongyi Han, Hao Sun, Yongshun Gong, Xiaoyan Wang, Yilong Yin, Shuo Li*
75 Black-Box Domain Adaptation for Object Detection with Retention-Driven Knowledge Compression, *Yuwu Lu, Chunzhi Liu*
76 Decoupled and Reusable Adaptation for Efficient Cross-Modal Transfer, *Yajing Liu, Yumeng Zhang, Yue Si, Baojie Fan, Jiandong Tian*
77 Preference-Aligned LoRA Merging: Preserving Subspace Coverage and Addressing Directional Anisotropy, *Wooseong Jeong, Wonyoung Lee, Kuk-Jin Yoon*
78 Curvature-Aware Zeroth-Order Optimization for Memory-Efficient Test-Time Adaptation, *Junming Zhang, Shuyu Yin, Peilin Liu, Rendong Ying, Fei Wen*
79 Label-Free Cross-Task LoRA Merging with Null-Space Compression, *Wonyoung Lee, Wooseong Jeong, Kuk-Jin Yoon*
80 Basis-Oriented Low-rank Transfer for Few-Shot and Test-Time Adaptation, *Junghwan Park, Woojin Cho, Junhyuk Heo, Darongsae Kwon, Kookjin Lee*
81 GeCo: Geometry-Consistent Regularization for Domain Generalized Semantic Segmentation, *Qi Zang, Dong Zhao, Nan Pu, Wenjing Li, Zhun Zhong, Meng Wang*
82 ❊ Event-based Motion Deblurring with Unpaired Data, *Hoonhee Cho, Yuhwan Jeong, Kuk-Jin Yoon*
83 Stable Spike: Dual Consistency Optimization via Bitwise AND Operations for Spiking Neural Networks, *Yongqi Ding, Kunshan Yang, Linze Li, Yiyang Zhang, Mengmeng Jing, Lin Zuo*
84 ❊ Event-based Visual Deformation Measurement, *Yuliang Wu, Wei Zhai, Yuxin Cui, Tiesong Zhao, Yang Cao, Zheng-Jun Zha*
85 Bidirectional Cross-Modal Prompting for Event-Frame Asymmetric Stereo, *Ninghui Xu, Fabio Tosi, Lihui Wang, Jiawei Han, Luca Bartolomei, Zhiting Yao, Matteo Poggi, Stefano Mattoccia*
86 SpikeTrack: High-performance and Energy-efficient Event-Based Object Tracking with Spiking Neural Network, *Yang Wang, Jiqing Zhang, Chuanyu Sun, Qianhui Liu, Huilin Ge, Ziqi Wei, Xin Yang*
87 Event Structural Valley: A Unified Theoretical and Practical Framework for Event Camera Autofocus, *Xijie Xiang, Lin Zhu, Wei Zhang, Yonghong Tian*
88 Adaptive Spatial-Temporal Window: Unlocking the Potential of Event Cameras in Heterogeneous Velocity Scenarios, *Zhipeng Sui, Haiqing Hao, Weihua He, Seng-Hong Lee, Wenhui Wang*
89 ❊ Do You Have Freestyle? Expressive Humanoid Locomotion via Audio Control, *Zhe Li, Cheng Chi, Yangyang Wei, Boan Zhu, Tao Huang, Zhenguo Sun, Yibo Peng, Pengwei Wang, Zhongyuan Wang, Fangzhou Liu, Chang Xu, Shanghang Zhang*
90 ❊ CLaD: Planning with Grounded Foresight via Cross-Modal Latent Dynamics, *Andrew Jeong, Jaemin Kim, Sebin Lee, Sung-Eui Yoon*
91 InternData-A1: Pioneering High-Fidelity Synthetic Data for Pre-training Generalist Policy, *Yang Tian, Yuyin Yang, Yiman Xie, Zetao Cai, Xu Shi, Ning Gao, Hangxu Liu, Xuekun Jiang, Zherui Qiu, Feng Yuan, Yaping Li, Ping Wang, Junhao Cai, Jia Zeng, Hao Dong, Jiangmiao Pang*
92 DemoFunGrasp: Universal Dexterous Functional Grasping via Demonstration-Editing Reinforcement Learning, *Chuan Mao, Haoqi Yuan, Ziye Huang, Chaoyi Xu, Kai Ma, Zongqing Lu*
93 ❊ GeniNav: Generative Model Driven Image-Goal Navigation via Imagination-Guided Consistency Flow Matching, *Yuqi Chen, Junjie Gao, Yongzhou Pan, Siyuan Song, Zixuan Zhang, Jiaping Xiao, Mir Feroskhan*
94 Cross from Left to Right Brain: Adaptive Text Dreamer for Vision-and-Language Navigation, *Pingrui Zhang, Yifei Su, Pengyuan Wu, Dong An, Li Zhang, Zhigang Wang, Dong Wang, Bin Zhao*
95 DRAMA: Next-Gen Dynamic Orchestration for Resilient Multi-Agent Ecosystems in Flux, *Xinkui Zhao, Yifan Zhang, Sai Liu, Naibo Wang, Guanjie Cheng, Yueshen Xu, Chang Liu, Shuiguang Deng, Jianwei Yin*
96 Arcadia: Toward a Full-Lifecycle Framework for Embodied Lifelong Learning, *Minghe Gao, Juncheng Li, Yuze Lin, Xuqi Liu, Jiaming Ji, Xiaoran Pan, Zihan Xu, Xian Li, Mingjie Li, Wei Ji, Rong Wei, Rui Tang, Qizhou Wang, Kai Shen, Jun Xiao, Qi Wu, Siliang Tang, Yueting Zhuang*
97 ❊ Wanderland: Geometrically Grounded Simulation for Open-World Embodied AI, *Xinhao Liu, Jiaqi Li, Youming Deng, Ruxin Chen, Yingjia Zhang, Yifei Ma, Li Guo, Yiming Li, Jing Zhang, Chen Feng*
98 ORV: 4D Occupancy-centric Robot Video Generation, *Xiuyu Yang, Bohan Li, Shaocong Xu, Nan Wang, Chongjie Ye, Zhaoxi Chen, Minghan Qin, Yikang Ding, Zheng Zhu, Xin Jin, Hang Zhao, Hao Zhao*
99 DextER: Language-driven Dexterous Grasp Generation with Embodied Reasoning, *Junha Lee, Eunha Park, Minsu Cho*
100 Language-Free Generative Editing from One Visual Example, *Omar Elezabi, Eduard Zamfir, Zongwei Wu, Radu Timofte*
101 Omni IIE Bench: Benchmarking the Practical Capabilities of Image Editing Models, *Yujia Yang, Yuanxiang Wang, Zhenyu Guan, Tiankun Yang, Chenxi Bao, Haopeng Jin, Jinwen Luo, Xinyu Zuo, Lisheng Duan, Haijin Liang, Jin Ma, Xinming Wang, Ruiwen Tao, Hongzhu Yi*
102 LuxRemix: Lighting Decomposition and Remixing for Indoor Scenes, *Ruofan Liang, Norman Müller, Ethan Weber, Duncan Zauss, Nandita Vijaykumar, Peter Kontschieder, Christian Richardt*
103 CompBench: Benchmarking Complex Instruction-guided Image Editing, *Bohan Jia, Wenxuan Huang, Yuntian Tang, Junbo Qiao, Jincheng Liao, Shaosheng Cao, Fei Zhao, Zhaopeng Feng, Zhouhong Gu, Zhenfei Yin, Lei Bai, Wanli Ouyang, Lin Chen, Fei Zhao, Zihan Wang, Yuan Xie, Shaohui Lin*
104 Garments2Look: A Multi-Reference Dataset for High-Fidelity Outfit-Level Virtual Try-On with Clothing and Accessories, *Junyao Hu, Zhongwei Cheng, Waikeung Wong, Xingxing Zou*
105 Learning Personalized Photographic Style from Pairwise User Preferences, *Jinwoo Kim, Jihye Yoo, Seon Joo Kim*
106 CogniEdit: Dense Gradient Flow Optimization for Fine-Grained Image Editing, *Yan Li, Lin Liu, Xiaopeng Zhang, Wei Xue, Wenhan Luo, Yike Guo, Qi Tian*
107 Efficient Weighted Sampling via Score-based Generative Models, *Heasung Kim, Taekyun Lee, Hyeji Kim, Gustavo De Veciana*
108 MOSAIC-GS: Monocular Scene Reconstruction via Advanced Initialization for Complex Dynamic Environments, *Svitlana Morkva, Vaishakh Patil, Alessio Tonioni, Michael Oechsle, Maximum Wilder-Smith, Marco Hutter*
109 REArtGS++: Generalizable Articulation Reconstruction with Temporal Geometry Constraint via Planar Gaussian Splatting, *Di Wu, Liu Liu, Anran Huang, Yuyan Liu, Qiaojun Yu, Shaofan Liu, Liangtu Song, Cewu Lu*
110 Dynamic-eDiTor: Training-Free Text-Driven 4D Scene Editing with Multimodal Diffusion Transformer, *Dong In Lee, Hyungjun Doh, Seunggeun Chi, Runlin Duan, Sangpil Kim, Karthik Ramani*

111 FaithFusion: Harmonizing Reconstruction and Generation via Pixel-wise Information Gain, *YuAn Wang, Xiaofan Li, Chi Huang, Wenhao Zhang, Hao Li, Bosheng Wang, Xun Sun, Jun Wang*
112 IR-HGP: Physically-Aware Gaussian Inverse Rendering for High-❊ Illumination Scenes via Generative Priors, *Qingan Zhang, Wensheng Li, Chengying Gao*
113 Seeing through boxes: Non-Line-of-Sight 3D Reconstruction from Radar Signals, *Jiachen Lu, Hailan Shanbhag, Haitham Al Hassanieh*
114 Speeding Up the Learning of 3D Gaussians with Much Shorter Gaussian Lists, *Jiaqi Liu, Zhizhong Han*
115 DynamicTree: Interactive Real Tree Animation via Sparse Voxel Spectrum, *Yaokun Li, Lihe Ding, Xiao Chen, Guang Tan, Tianfan Xue*
116 WildRayZer: Self-supervised Large View Synthesis in Dynamic ❊ Environments, *Xuweiyi Chen, Wentao Zhou, Zezhou Cheng*
117 DGGT: Feedforward 4D Reconstruction of Dynamic Driving Scenes using Unposed Images, *Xiaoxue Chen, Ziyi Xiong, Yuantao Chen, Gen Li, Nan Wang, Hongcheng Luo, Long Chen, Haiyang Sun, Bing Wang, Guang Chen, Hongyang Li, Ya-Qin Zhang, Hangjun Ye, Hao Zhao*
118 Retrieve-to-Restore: Efficient All-in-One Image Restoration with a Retrieval-Based Degradation Bank, *Chenxu Wang, Kai Zhang, Jian Yang*
119 MRI Contrast Enhancement Kinetics World Model, *Jindi Kong, Yuting He, Cong Xia, Rongjun Ge, Shuo Li*
120 ReflexSplit: Single Image Reflection Separation via Layer Fusion-Separation, *Chia-Ming Lee, Yu-Fan Lin, Jin-Hui Jiang, Yu-Jou Hsiao, Chih-Chung Hsu, Yu-Lun Liu*
121 Rethinking Knowledge Transfer in Image Quality Assessment: A Perceptual Preference Structure Alignment Perspective, *Aobo Li, Jinjian Wu, Yongxu Liu, Jupo Ma, Weisheng Dong*
122 ZeroIDIR: Zero-Reference Illumination Degradation Image Restoration with Perturbed Consistency Diffusion Models, *Hai Jiang, Zhen Liu, Yinjie Lei, Songchen Han, Bing Zeng, Shuaicheng Liu*
123 White-Balance First, Adjust Later: Cross-Camera Color Constancy via Vision-Language Evaluation, *Shuwei Li, Lei Tan, Robby T. Tan*
124 Unpaired Image Deraining Using Reward-Guided Self-Reinforcement Strategy, *Yinghao Chen, Yeying Jin, Xiang Chen, Yanyan Wei, Ziyang Yan, Yaowen Fu*
125 LF-BVN: Blind-View Network for Self-Supervised Light Field ❊ Denoising, *Longzhao Guo, Shuo Zhang, Chen Gao, Qian Tian, Youfang Lin*
126 rPPG-VQA: A Video Quality Assessment Framework for Unsupervised rPPG Training, *Tianyang Dai, Ming Chang, Yan Chen, Yang Hu*
127 Efficient Real-Time Raw-to-Raw Denoising for Extreme Low-Light Ultra HD Video on Mobile Devices, *Charantej Pochimireddy, Subhasmita Sahoo, Apoorva Verma, Palavalli Shyam, Swapnil Malviya, Sarvesh Sarvesh, Raj Gadde*
128 Towards Generalized Representations for Low-Light Understanding: When Signal Constancy Meets Semantic Enrichment, *Yifan Li, Haofeng Huang, Wenhan Yang, Jiaying Liu*
129 Synergistic Bleeding Region and Point Detection in Laparoscopic Surgical Videos, *Jialun Pei, Zhangjun Zhou, Diandian Guo, Zhixi Li, Jing Qin, Bo Du, Pheng-Ann Heng*
130 MedCLIPSeg: Probabilistic Vision-Language Adaptation for Data-Efficient and Generalizable Medical Image Segmentation, *Taha Koleilat, Hojat Asgariandehkordi, Omid Nejatimanzari, Berardino Barile, Yiming Xiao, Hassan Rivaz*
131 AD-GBC: Anisotropic Granular-Ball Skip-Connection Refiner for UNet-❊ Based Medical Image Segmentation, *Xiya Shen, Qinglin Zhao, Li Feng*
132 OSA: Echocardiography Video Segmentation via Orthogonalized ❊ State Update and Anatomical Prior-aware Feature Enhancement, *Rui Wang, Huisi Wu, Jing Qin*
133 VesMamba: 3D Pulmonary Vessel Segmentation from CT images via Mamba with Structural Perception and Scale-aware Filtering, *Zhipeng Liu, Guilian Chen, Zheng Jiang, Huisi Wu, Jing Qin*
134 SemiGDA: Generative Dual-distribution Alignment for Semi-Supervised Medical Image Segmentation, *Kaiwen Huang, Yi Zhou, Yizhe Zhang, Jingxiong Li, Tao Zhou*
135 Diffusion-Based Native Adversarial Synthesis for Enhanced Medical ❊ Segmentation Generalization, *Hongyu Zhang, Haipeng Chen, Zhimin Xu, Chengxin Yang, Yingda Lyu*
136 CG-Reasoner: Centroid-Guided Positional Reasoning Segmentation for Medical Imaging with a Robust Visual-Text Consistency Metric, *Lakshmikar Reddy Polamreddy, Ming Ma*
137 Instruction-Guided Lesion Segmentation for Chest X-rays with Automatically Generated Large-Scale Dataset, *Geon Choi, Hangyul Yoon, Hyunju Shin, Hyunki Park, Sang Hoon Seo, Eunho Yang, Edward Choi*
138 Towards Highly Transferable Vision-Language Attack via Semantic-Augmented Dynamic Contrastive Interaction, *Yuanbo Li, Tianyang Xu, Cong Hu, Tao Zhou, Xiaojun Wu, Josef Kittler*
139 Towards Human-Imperceptible Backdoor Attacks on Text-to-Image Diffusion Models, *Yiming Wu, Chenghao Chen, Changkun Wu, Chong Fu, Biru Zhu, Zhenyu Wen, Zhen Hong*
140 TTP: Test-Time Padding for Adversarial Detection and Robust Adaptation on Vision-Language Models, *Zhiwei Li, Yitian Pang, Weining Wang, Zhenan Sun, Qi Li*
141 DualMirage: Hunting Stealthy Multimodal LLM Agents via CAPTCHAs with Contour and Adversarial Illusions, *Bei Chen, Gaolei Li, Jun Wu, Jianhua Li*
142 Models as Lego Builders: Assembling Malice from Benign Blocks via Semantic Blueprints, *Chenxi Li, Xianggan Liu, Dake Shen, Yaosong Du, Zhibo Yao, Hao Jiang, Linyi Jiang, Chengwei Cao, Jingzhe Zhang, RanYi Peng, Peiling Bai, Xiande Huang*
143 Source Models Leak What They Shouldn't: Unlearning Zero-Shot Transfer in Domain Adaptation Through Adversarial Optimization, *Arnav Devalapally, Poornima Jain, Kartik Srinivas, Vineeth N. Balasubramanian*
144 A Unified Perspective on Adversarial Membership Manipulation in Vision Models, *Ruize Gao, Kaiwen Zhou, Yongqiang Chen, Feng Liu*
145 Shedding Light on VLN Robustness: A Black-box Framework for Indoor Lighting-based Adversarial Attack, *Chenyang Li, Wenbing Tang, Yihao Huang, Sinong Simon Zhan, Ming Hu, Xiaojun Jia, Yang Liu*
146 OddGridBench: Exposing the Lack of Fine-Grained Visual ❊ Discrepancy Sensitivity in Multimodal Large Language Models, *Tengjin Weng, Wenhao Jiang, Jingyi Wang, Ming Li, Lin Ma, Zhong Ming*
147 Beyond What's Shared: Recovering Lost Unique Information from Intermediate Layers to Boost Multimodal Geo-Foundation Models, *JangHyeon Lee, Philipe Ambrozio Dias, Yao-Yi Chiang, Dalton Lunga*
148 WikiCLIP: An Efficient Contrastive Baseline for Open-domain Visual ❊ Entity Recognition, *Shan Ning, Longtian Qiu, Jiaxuan Sun, Xuming He*
149 CLCR: Cross-Level Semantic Collaborative Representation for Multimodal Learning, *Chunlei Meng, Guanhong Huang, Rong Fu, Runmin Jian, Zhongxue Gan, Chun Ouyang*
150 Learning Anchor in Dual Orthogonal Space for Fast Multi-view ❊ Clustering, *Yalan Qin, Hanzhou Wu*
151 Bootstrapping Multi-view Learning for Test-time Noisy Correspondence, *Changhao He, Di Xue, Shuxian Li, Yanji Hao, Xi Peng, Peng Hu*
152 Differences That Matter: Auditing Models for Capability Gap ❊ Discovery and Rectification, *Qihao Liu, Chengzhi Mao, Yaojie Liu, Alan Yuille, Wen-Sheng Chu*
153 FAVE: A Structured Benchmark for Fine-Grained Audio-Visual Temporal Evaluation in Multimodal LLMs, *Weiheng Lu, An Yu, Jian Li, Zhenfei Zhang, Felix X.-F. Ye, Ming-Ching Chang*
154 Omni2Sound: Towards Unified Video-Text-to-Audio Generation, ❊ *Yusheng Dai, Zehua Chen, Yuxuan Jiang, Qiuhong Ke, Jianfei Cai, Jun Zhu*
155 EmoThinker: Advancing Visual-Acoustic Emotion Analysis via Structural Token Selection and Chain-of-Thought Reasoning, *Qinfu Xu, Liyuan Pan, Yiwei Wei, Shaozu Yuan, Jiaqi Chen, Tianyu Liu*
156 Enhancing Descriptive Captions with Visual Attributes for Multimodal Perception, *Yanpeng Sun, Jing Hao, Ke Zhu, Jiang-Jiang Liu, Xiaofan Li, Na Zhao, Zechao Li, Jingdong Wang*
157 DiG: Differential Grounding for Enhancing Fine-Grained Perception in Multimodal Large Language Models, *Zhou Tao, Shida Wang, YongXiang Hua, Haoyu Cao, Linli Xu*
158 Vision-Speech Models: Teaching Speech Models to Converse about Images, *Amélie Royer, Moritz Böhle, Laurent Mazaré, Neil Zeghidour, Alexandre Défossez, Patrick Pérez*

159 EMMA: Extracting Multiple physical parameters from Multimodal Data, *Farhat Shaikh, Ayan Banerjee, Sandeep Gupta*
160 MMGait: Towards Multi-Modal Gait Recognition, *Chenye Wang, Qingyuan Cai, Saihui Hou, Aoqi Li, Yongzhen Huang*
161 OSMO: Open-vocabulary Self-eMOtion Tracking, *Mohamed Abdelfattah, Bugra Tekin, Fadime Sener, Necati Cihan Camgoz, Eric Sauser, Shugao Ma, Alexandre Alahi, Edoardo Remelli*
162 MuCo: Multi-turn Contrastive Learning for Multimodal Embedding Model, *Geonmo Gu, Byeongho Heo, Jaemyung Yu, Jaehui Hwang, Taekyung Kim, Sangmin Lee, HeeJae Jun, Yoohoon Kang, Sangdoo Yun, Dongyoon Han*
163 Cross-Modal Emotion Transfer for Emotion Editing in Talking Face Video, *Chanhyuk Choi, Taesoo Kim, Donggyu Lee, Siyeol Jung, Taehwan Kim*
164 Unleashing the Intrinsic Visual Representation Capability of Multimodal Large Language Models, *Hengzhuang Li, Xinsong Zhang, Qiming Peng, Bin Luo, Han Hu, Dengyang Jiang, Han-Jia Ye, Teng Zhang, Hai Jin*
165 Active Perceptual Inference: A Corticothalamic-Inspired Dynamic Nested Recurrent Network for Multimodal Sentiment Analysis with Incomplete Data, *Yujuan Zhang, Qing Li, Ziyu Li, Xiuxing Li, Zhuo Wang, Mengrui Xu, Xia Wu*
166 Scalable Trajectory Generation for Whole-Body Mobile Manipulation, ❊ *Yida Niu, Xinhai Chang, Xin Liu, Ziyuan Jiao, Yixin Zhu*
167 Breaking the 3D Dataset Bottleneck: Fast Scalable Generation of Aligned 3D Assets from Scratch for Category 6D Pose Estimation and Robotic Grasping, *Duret Guillaume, Danylo Mazurak, Florence Zara, Jan Peters, Liming Chen*
168 Real-Time Multimodal Fingertip Contact Detection via Depth and Motion Fusion for Vision-Based Human–Computer Interaction, *Mukhiddin Toshpulatov, Wookey Lee, Suan Lee, Geehyuk Lee*
169 Glove2Hand: Synthesizing Natural Hand-Object Interaction from ❊ Multi-Modal Sensing Gloves, *Xinyu Zhang, Ziyi Kou, Chuan Qin, Mia Huang, Ergys Ristani, Ankit Kumar, Lele Chen, Kun He, Abdeslam Boularias, Li Guan*
170 UniDex: A Robot Foundation Suite for Universal Dexterous Hand Control from Egocentric Human Videos, *Gu Zhang, Qicheng Xu, Haozhe Zhang, Jianhan Ma, Long He, Yiming Bao, Zeyu Ping, Zhecheng Yuan, Chenhao Lu, Chengbo Yuan, Tianhai Liang, Xiaoyu Tian, Maanping Shao, Feihong Zhang, Mingyu Ding, Yang Gao, Hao Zhao, Hang Zhao, Huazhe Xu*
171 ConsID-Gen: View-Consistent and Identity-Preserving Image-to-Video Generation, *Mingyang Wu, Ashirbad Mishra, Soumik Dey, Shuo Xing, Naveen Ravipati, Hansi Wu, Binbin Li, Zhengzhong Tu*
172 DiverseGRPO: Mitigating Mode Collapse in Image Generation via ❊ Diversity-Aware GRPO, *Henglin Liu, Huijuan Huang, Jing Wang, Chang Liu, Xiu Li, Xiangyang Ji*
173 VAR RL Done Right: Tackling Asynchronous Policy Conflicts in Visual Autoregressive Generation, *Shikun Sun, Liao Qu, Huichao Zhang, Yiheng Liu, Yangyang Song, Xian Li, Yi Jiang, Xu Wang, Jia Jia, Daniel K. Du, Xinglong Wu*
174 Video Generation with Stable Transparency via Shiftable RGB-A ❊ Distribution Learner, *Haotian Dong, Wenjing Wang, Chen Li, Jing Lyu, Di Lin*
175 MOFA-VTON: More Fashion Possibilities with Fine-Grained ❊ Adaptations in Virtual Try-On, *Xiaoyu Han, Chenyang Wang, Jing Wang, Shunyuan Zheng, Quanling Meng, Shengping Zhang*
176 Scaling Multi-Identity Consistency for Image Customization via Multi-to-Multi Matching Paradigm, *Yufeng Cheng, Wenxu Wu, Shaojin Wu, Mengqi Huang, Fei Ding, Qian He*
177 NOVA: Sparse Control, Dense Synthesis for Pair-Free Video Editing, *Tianlin Pan, Jiayi Dai, Chenpu Yuan, Zhengyao Lv, Binxin Yang, Hubery Yin, Chen Li, Jing Lyu, Caifeng Shan, Chenyang Si*
178 Functional Mean Flow in Hilbert Space, *Zhiqi Li, Yuchen Sun, Greg Turk, Bo Zhu*
179 Benchmarking Single-Factor Physical Video-to-Audio Generation, *Tingle Li, Siddharth Gururani, Kevin J. Shih, Gantavya Bhatt, Sang-gil Lee, Zhifeng Kong, Arushi Goel, Gopala Anumanchipalli, Ming-Yu Liu*
180 UniAVGen: Unified Audio and Video Generation with Asymmetric Cross-Modal Interactions, *Guozhen Zhang, Zixiang Zhou, Teng Hu, Ziqiao Peng, Youliang Zhang, Yi Chen, Yuan Zhou, Qinglin Lu, Limin Wang*
181 Refaçade: Editing Object with Given Reference Texture, *Youze Huang, Penghui Ruan, Bojia Zi, Xianbiao Qi, Jianan Wang, Rong Xiao*
182 Free-Lunch Long Video Generation via Layer-Adaptive O.O.D Correction, *Jiahao Tian, Chenxi Song, Wei Cheng, Chi Zhang*
183 Not All Birds Look The Same: Identity-Preserving Generation For ❊ Birds, *Aaron Sun, Oindrila Saha, Subhransu Maji*
184 HiFi-Inpaint: Towards High-Fidelity Reference-Based Inpainting for Generating Detail-Preserving Human-Product Images, *Yichen Liu, Donghao Zhou, Jie Wang, Xin Gao, Guisheng Liu, Jiatong Li, Quanwei Zhang, Qiang Lyu, Lanqing Guo, Shilei Wen, Weiqiang Wang, Pheng-Ann Heng*
185 EffectErase: Joint Video Object Removal and Insertion for High- ❊ *Quality Effect Erasing, Yang Fu, Yike Zheng, Ziyun Dai, Henghui Ding*
186 Clothe and Pose, *Nakul Sharma, Aayush Bansal, Minh Vo*
187 FlowPortal: Residual-Corrected Flow for Training-Free Video Relighting and Background Replacement, *Wenshuo Gao, Junyi Fan, Jiangyue Zeng, Shuai Yang*
188 The Consistency Critic: Correcting Inconsistencies in Generated Images via Reference-Guided Attentive Alignment, *Ziheng Ouyang, Yiren Song, Yaoli Liu, Shihao Zhu, Qibin Hou, Ming-Ming Cheng, Mike Zheng Shou*
189 Rethinking UMM Visual Generation: Masked Modeling for Efficient Image-Only Pre-training, *Peng Sun, Jun Xie, Tao Lin*
190 VibeToken: Scaling 1D Image Tokenizers and Autoregressive Models for Dynamic Resolution Generations, *Maitreya Patel, Jingtao Li, Weiming Zhuang, Yezhou Yang, Lingjuan Lv*
191 Bidirectional Normalizing Flow: From Data to Noise and Back, ❊ *Yiyang Lu, Qiao Sun, Xianbang Wang, Zhicheng Jiang, Hanhong Zhao, Kaiming He*
192 ShotDirector: Directorially Controllable Multi-Shot Video Generation with Cinematographic Transitions, *Xiaoxue Wu, Xinyuan Chen, Yaohui Wang, Yu Qiao*
193 Are Image-to-Video Models Good Zero-Shot Image Editors?, *Zechuan Zhang, Zhenyuan Chen, Zongxin Yang, Yi Yang*
194 FastLightGen: Fast and Light Video Generation with Fewer Steps and Parameters, *Shitong Shao, Yufei Gu, Zeke Xie*
195 Unified Latent Space for Understanding and Generation via Semantic Auto-encoder, *Xiaojie Li, Yang Zhao, Ming Li, Yancheng Zhang, Zonglin Lyu, Yunpeng Chen, Rui Wang, Daquan Zhou*
196 AHS: Adaptive Head Synthesis via Synthetic Data Augmentations, *Taewoong Kang, Hyojin Jang, Sohyun Jeong, Seunggi Moon, Gihwi Kim, Hoon Jin Jung, Jaegul Choo*
197 CASR: A Robust Cyclic Framework for Arbitrary Large-Scale Super-Resolution with Distribution Alignment and Self-Similarity Awareness, *Wenhao Guo, Zhaoran Zhao, Peng Lu, Sheng Li, Qian Qiao, RuiDe Li*
198 Thermal Diffusion Matters: Infrared Spatial-Temporal Video Super- ❊ *Resolution through Heat Conduction Priors, Mingxuan Zhou, Shuang Li, Yutang Zhang, Jing Geng, Yirui Shen, Jingxuan Kang, Fuzhen Zhuang, Shuigen Wang*
199 TextOVSR: Text-Guided Real-World Opera Video Super-Resolution, *Hua Chang, Xin Xu, Wei Liu, Jiayi Wu, Kui Jiang, Fei Ma, Qi Tian*
200 VoDaSuRe: A Large-Scale Dataset Revealing Domain Shift in Volumetric Super-Resolution, *August Leander Høeg, Sophia Wiinberg Bardenfleth, Hans Martin Kjer, Tim Bjørn Dyrby, Vedrana Andersen Dahl, Anders Bjorholm Dahl*
201 GDPO-SR: Group Direct Preference Optimization for One-Step Generative Image Super-Resolution, *Qiaosi Yi, Shuai Li, Rongyuan Wu, Lingchen Sun, Zhengqiang Zhang, Lei Zhang*
202 Adaptive Anisotropic Gaussian Splatting for Multi-contrast MRI Arbitrary-Scale Super-Resolution with Anatomy Guidance, *Qiuhai Yan, Kang Chen, Zhengjie Lu, Tingting Wang, Faming Fang, Guixu Zhang*
203 SignPR: A Progressive Vector-Quantized Diffusion Framework for Sign Language Production, *Xiao Liu, Shiwei Gan, Yafeng Yin, Bowen Guo, Zhiwei Jiang, Shunmei Meng, Lei Xie, Sanglu Lu*
204 LLaMo: Scaling Pretrained Language Models for Unified Motion Understanding and Generation with Continuous Autoregressive Tokens, *Zekun Li, Sizhe An, Chengcheng Tang, Chuan Guo, Ivan Shugurov, Linguang Zhang, Amy Zhao, Srinath Sridhar, Lingling Tao, Abhay Mittal*
205 FlashCap: Millisecond-Accurate Human Motion Capture via Flashing LEDs and Event-Based Vision, *Zekai Wu, Shuqi Fan, Mengyin Liu, Yuhua Luo, Xincheng Lin, Ming Yan, Junhao Wu, Xiuhong Lin,*

Yuexin Ma, Chenglu Wen, Lan Xu, Siqi Shen, Cheng Wang

206 Geometric Neural Distance Fields for Learning Human Motion Priors, ❊ *Zhengdi Yu, Simone Foti, Linguang Zhang, g921@gmail.com Meta Reality Labs, Amy Zhao, Cem Keskin, Stefanos Zafeiriou, Tolga Birdal*

207 3D-Aware Implicit Motion Control for View-Adaptive Human Video Generation, *Zhixue Fang, Xu He, Songlin Tang, Haoxian Zhang, Qingfeng Li, Xiaoqiang Liu, Pengfei Wan, Kun Gai*

208 Decoupled Generative Modeling for Human-Object Interaction Synthesis, *Hwanhee Jung, Seunggwan Lee, Jeongyoon Yoon, SeungHyeon Kim, Giljoo Nam, Qixing Huang, Sangpil Kim*

209 LiveGesture: Streamable Co-Speech Gesture Generation Model, *Muhammad Usama Saleem, Mayur Jagdishbhai Patel, Ekkasit Pinyoanuntapong, Zhongxing Qin, Li Yang, Hongfei Xue, Ahmed Helmy, Chen Chen, Pu Wang*

210 HandX: Scaling Bimanual Motion and Interaction Generation, *Zimu Zhang, Yucheng Zhang, Xiyan Xu, Ziyin Wang, Sirui Xu, Kai Zhou, Bing Zhou, Chuan Guo, Jian Wang, Yu-Xiong Wang, Liang-Yan Gui*

211 MaskAdapt: Learning Flexible Motion Adaptation via Mask-Invariant ❊ Prior for Physics-Based Characters, *Soomin Park, Eunseong Lee, Kwang Bin Lee, Sung-Hee Lee*

212 FloodDiffusion: Tailored Diffusion Forcing for Streaming Motion ❊ Generation, *Yiyi Cai, Yuhan Wu, Kunhang Li, You Zhou, Bo Zheng, Haiyang Liu*

213 ProjFlow: Projection Sampling with Flow Matching for Zero-Shot Exact Spatial Motion Control, *Akihisa Watanabe, Qing Yu, Edgar Simo-Serra, Kent Fujiwara*

214 Correspondence-Attention Alignment for Multi-View Diffusion Models, *Minkyung Kwon, Jinhyeok Choi, Jiho Park, Seonghu Jeon, Jinhyuk Jang, Junyoung Seo, Minseop Kwak, Jin-Hwa Kim, Seungryong Kim*

215 GenErase: Generalizable and Semantically-Aware Concept Erasure in ❊ Diffusion Models, *Korada Sri Vardhana, Soma Biswas*

216 MatMart: Material Reconstruction of 3D Objects via Diffusion, *Xiuchao Wu, Pengfei Zhu, Jiangjing Lyu, Xinguo Liu, Jie Guo, Yanwen Guo, Weiwei Xu, Chengfei Lyu*

217 Region-Adaptive Sampling for Diffusion Transformers, *Ziming Liu, Yifan Yang, Chengruidong Zhang, Yiqi Zhang, Lili Qiu, Yang You, Yuqing Yang*

218 Diffusion Guided Chain-of-Vision for Large Autoregressive Vision Models, *Xinyang Wang, Kecheng Zheng, Minfeng Zhu, Wei Wu, Fan Lu, Wei Zhai, Wei Chen*

219 Guiding Diffusion-based Reconstruction with Contrastive Signals for Balanced Visual Representation, *Boyu Han, Qianqian Xu, Shilong Bao, Zhiyong Yang, Ruochen Cui, Xilin Zhao, Qingming Huang*

220 ConceptPrism: Concept Disentanglement in Personalized Diffusion Models via Residual Token Optimization, *Minseo Kim, Minchan Kwon, Dongyeun Lee, Yunho Jeon, Junmo Kim*

221 Heterogeneous Decentralized Diffusion Models, *Zhiying Jiang, Raihan Seraj, Marcos Villagra, Bidhan Roy*

222 Refining Few-Step Text-to-Multiview Diffusion via Reinforcement Learning, *Ziyi Zhang, Li Shen, Deheng Ye, Yong Luo, Huangxuan Zhao, Meng Liu, Wei Yu, Lefei Zhang*

223 GroundingME: Exposing the Visual Grounding Gap in MLLMs through Multi-Dimensional Evaluation, *Rang Li, Lei Li, Shuhuai Ren, Hao Tian, Shuhao Gu, Shicheng Li, Zihao Yue, Yudong Wang, Wenhan Ma, Zhe Yang, Jingyuan Ma, Zhifang Sui, Fuli Luo*

224 ENC-Bench: A Benchmark for Evaluating Multimodal Large Language ❊ Models in Electronic Navigational Chart Understanding, *Ao Cheng, Xingming Li, Xuanyu Ji, Xixiang He, Qiyao Sun, Chunping Qiu, Runke Huang, Qingyong Hu*

225 Nonparametric Deep Fine-grained Clustering with Low-Rank Guided Vision-Language Model, *Xulun Ye, Benyu Wu, Jie Hong, Kun Zhou*

226 RealBirdID: Benchmarking Bird Species Identification in the Era of MLLMs, *Logan Lawrence, Oindrila Saha, Rangel Daroya, Mustafa Chasmai, Wuao Liu, Max Hamilton, Aaron Sun, Seoyun Jeong, Fabien Delattre, Subhransu Maji, Grant Van Horn*

227 Fast SceneScript: Fast and Accurate Language-Based 3D Scene Understanding via Multi-Token Prediction, *Ruihong Yin, Xuepeng Shi, Oleksandr Bailo, Marco Manfredi, Theo Gevers*

228 PP-OCRv5: A Specialized 5M-Parameter Model Rivaling Billion-Parameter Vision-Language Models on OCR Tasks, *Cheng Cui, Yubo Zhang, Ting Sun, Xueqing Wang, Hongen Liu, Manhui Lin, Yue Zhang, Tingquan Gao, Changda Zhou, Jiaxuan Liu, Zelun Zhang, Jing Zhang, Jun Zhang, Yi Liu*

229 World in a Frame: Understanding Culture Mixing as a New Challenge for Vision-Language Models, *Eunsu Kim, Junyeong Park, Na Min An, Junseong Kim, Hitesh Laxmichand Patel, Jiho Jin, Julia Kruk, Amit Agarwal, Srikant Panda, Fenal Ashokbhai Ilasariya, Hyunjung Shim, Alice Oh*

230 Gastric-X: A Multimodal Multi-Phase Benchmark Dataset for ❊ Advancing Vision-Language Models in Gastric Cancer Analysis, *Yuanzhe Li, Hao Chen, Rui Yin, Juyan Ba, Yu Zhang, Sheng Lu*

231 HiSpatial: Taming Hierarchical 3D Spatial Understanding in Vision-Language Models, *Huizhi Liang, Yichao Shen, Yu Deng, Sicheng Xu, ZhiYuan Feng, Tong Zhang, Yaobo Liang, Jiaolong Yang*

232 HandVQA: Diagnosing and Improving Fine-Grained Spatial Reasoning about Hands in Vision-Language Models, *MD Khalequzzaman Chowdhury Sayem, Mubarrat Tajoar Chowdhury, Yihalem Yimolal Tiruneh, Muneeb A. Khan, Muhammad Salman Ali, Binod Bhattarai, Seungryul Baek*

233 Probing and Bridging Geometry–Interaction Cues for Affordance Reasoning in Vision Foundation Models, *Qing Zhang, Xuesong Li, Jing Zhang*

234 ARC Is a Vision Problem!, *Keya Hu, Ali Cy, Linlu Qiu, Xiaoman Delores Ding, Runqian Wang, Yeyin Eva Zhu, Jacob Andreas, Kaiming He*

235 Geoint-R1: Formalizing Multimodal Geometric Reasoning with Dynamic Auxiliary Constructions, *Jingxuan Wei, Caijun Jia, Qi Chen, Honghao He, Linzhuang Sun, Conghui He, Lijun Wu, Bihui Yu, Cheng Tan*

236 S^2-MLLM: Boosting Spatial Reasoning Capability of MLLMs for 3D Visual Grounding with Structural Guidance, *Beining Xu, Siting Zhu, Zhao Jin, Junxian Li, Hesheng Wang*

237 Learning Multi-View Spatial Reasoning from Cross-View Relations, *Suchae Jeong, Jaehwi Song, Haeone Lee, Hanna Kim, Jian Kim, Dongjun Lee, Dong Kyu Shin, Changyeon Kim, Dongyoon Hahm, Woogyeol Jin, Juheon Choi, Kimin Lee*

238 Exploring Spatial Intelligence from a Generative Perspective, *Muzhi Zhu, Shunyao Jiang, Huanyi Zheng, Zekai Luo, Hao Zhong, Anzhou Li, Kaijun Wang, Jintao Rong, Yang Liu, Hao Chen, Tao Lin, Chunhua Shen*

239 Physical Object Understanding with a Physically Controllable World ❊ Model, *Rahul Venkatesh, Klemen Kotar, Lilian Naing Chen, Wanhee Lee, Gia Ancone, Seungwoo Kim, Luca Thomas Wheeler, Jared Watrous, Honglin Chen, Daniel Bear, Stefan Stojanov, Daniel LK Yamins*

240 QueryMe: Query-Driven Open-Vocabulary 3D Object Affordances Grounding from Multimodal Evidence, *Weiyu Zhao, Ru Li, Jiaqi Liu, Sizhe Zhao, Qinglin Liu, Shengping Zhang*

241 Think with 3D: Geometric Imagination Grounded Spatial Reasoning from Limited Views, *Zhangquan Chen, Manyuan Zhang, Xinlei Yu, Xufang Luo, Mingze Sun, Zihao Pan, Xiang An, Yan Feng, Peng Pei, Xunliang Cai, Ruqi Huang*

242 EG-3DVG: Expression and Geometry Aware Grounding Decoder for 3D Visual Grounding, *GwangWook Park, Hyo-Jun Lee, Jong-Hyeon Baek, Hanul Kim, Yeong Jun Koh*

243 AffordMatcher: Affordance Learning in 3D Scenes from Visual Signifiers, *Nghia Vu, Tuong Do, Khang Nguyen, Baoru Huang, Nhat Le, Binh Xuan Nguyen, Erman Tjiputra, Quang D. Tran, Ravi Prakash, Te-Chuan Chiu, Anh Nguyen*

244 SpatiaLQA: A Benchmark for Evaluating Spatial Logical Reasoning in Vision-Language Models, *Yuechen Xie, Xiaoyan Zhang, Yicheng Shan, Zhu Hao, Rui Tang, Rong Wei, Mingli Song, Yuanyu Wan, Jie Song*

245 Air-Know: Arbiter-Calibrated Knowledge-Internalizing Robust Network for Composed Image Retrieval, *Zhiheng Fu, Yupeng Hu, Qianyun Yang, Shiqi Zhang, Zhiwei Chen, Zixu Li*

246 Intra-class Distribution-guided Generative Hashing with Neighbor ❊ Refinement for Cross-modal Retrieval, *Hao Sun, Yadong Huo, Qibing Qin, Wenfeng Zhang, Lei Huangs*

247 Language-driven Fine-grained Retrieval, *Shijie Wang, Xin Yu, Yadan Luo, Zijian Wang, Pengfei Zhang, Zi Huang*

248 MRD: Multi-resolution Retrieval-Detection Fusion for High-Resolution Image Understanding, *Fan Yang, Xingping Dong, Xin Yu, Wenhan Luo, Wei Liu, Kaihao Zhang*

249 RetFormer: Multimodal Retrieval for Enhancing Image Recognition, *Tianrui Yu, Xiubo Liang, Hongzhi Wang*
250 DREAM: Document Recognition with Explicit Adaptive Memory, *Tianqi Zhao, Di Wu, Liangrui Peng, Yifan Huang, Kemeng Zhao, Shuo Li, Zhiyu Li, Yizhu Wang, Borui Jiang, Yuyang Li*
251 ❊ RMIR: A Benchmark Dataset for Reasoning-Intensive Multimodal Image Retrieval, *Yijiang Li, Kunal Kotian, Ali Marjaninejad, Meir Friedenberg, Kaushik Pavani, Sunny Dasgupta*
252 POGA: Paraphrased and Oppositional Graph Alignment for Fine-Grained Cross-Modal Retrieval, *Junfeng Zhang, Zhe Xue, Yuankai Qi, Junping Du, Xiangyang Kong, Yishuo Yan, Amin Beheshti, Jian Yang, Anton van den Hengel, Ming-Hsuan Yang*
253 Chain-of-Frames: Advancing Video Understanding in Multimodal LLMs via Frame-Aware Reasoning, *Sara Ghazanfari, Francesco Croce, Nicolas Flammarion, Prashanth Krishnamurthy, Farshad Khorrami, Siddharth Garg*
254 TempR1: Improving Temporal Understanding of MLLMs via Temporal-Aware Multi-Task Reinforcement Learning, *Tao Wu, Li Yang, Gen Zhan, Yabin Zhang, Yiting Liao, Junlin Li, Deliang Fu, Li Zhang, Limin Wang*
255 ❊ RiskProp: Collision-Anchored Self-Supervised Risk Propagation For Early Accident Anticipation, *Yiyang Zou, Tianhao Zhao, Peilun Xiao, Hongyu Jin, Longyu Qi, Yuxuan Li, Liyin Liang, Yifeng Qian, Chunbo Lai, Yutian Lin, Zhihui Li, Yu Wu*
256 MotionEnhancer: Leveraging Video Diffusion for Motion-Enhanced Vision-Language Models, *Yifan Xu, Chao Zhang, Ruifei Ma, Fei Gao, Zhifei Yang, Jiaxing Qi, Zhipeng Chen*
257 MedGRPO: Multi-Task Reinforcement Learning for Heterogeneous Medical Video Understanding, *Yuhao Su, Anwesa Choudhuri, Zhongpai Gao, Benjamin Planche, Van Nguyen Nguyen, Meng Zheng, Yuhan Shen, Arun Innanje, Terrence Chen, Ehsan Elhamifar, Ziyan Wu*
258 Asynchronous Temporal Modeling with Two-Agent Framework for Streaming Dense Video Captioning, *Yolo Y. Tang, Chao Huang, Susan Liang, Jing Bi, Yicheng Wang, Daiki Shimada, Chenliang Xu*
259 TRCoRSurg: Temporal-Relational Co-Reasoning for Surgical Video Triplet Recognition, *Fang Li, Shihao Zou, Weixin Si, Yang Gao, Shuai Li, Aimin Hao*
260 OASIS: On-Demand Hierarchical Event Memory for Streaming Video Reasoning, *Zhijia Liang, Jiaming Li, Weikai Chen, Yanhao Zhang, Haonan Lu, Guanbin Li*
261 ❊ One-Shot Flow, Any-Time Frame: A Bidirectional Warping Framework for Event-Based Video Frame Interpolation, *Linghui Fu, Yuhan Liu, Hao Chen, Zhen Yang, Yongjian Deng*
262 ❊ TF-CADE: Foreground-Concentrated Text-Video Alignment for Zero-Shot Temporal Action Detection, *Yearang Lee, Ho-Joong Kim, Seong-Whan Lee*
263 ❊ PRISM: Prototype-based Reasoning with Inter-modal Semantic Mining for Interpretable Image Recognition, *Anni Yu, Yu-Bin Yan*
264 Concept Regions Matter: Benchmarking CLIP with a New Cluster-Importance Approach, *Aishwarya Agarwal, Srikrishna Karanam, Vineet Gandhi*
265 PhaseWin Search Framework Enable Efficient Object-Level Interpretation, *Zihan Gu, Ruoyu Chen, Junchi Zhang, Yue Hu, Hua Zhang, Xiaochun Cao*
266 Beyond Top Activations: Efficient and Reliable Crowdsourced Evaluation of Automated Interpretability, *Tuomas Oikarinen, Ge Yan, Akshay Kulkarni, Tsui-Wei Weng*
267 From Weights to Concepts: Data-Free Interpretability of CLIP via Singular Vector Decomposition, *Francesco Gentile, Nicola Dall'Asen, Francesco Tonini, Massimiliano Mancini, Lorenzo Vaquero, Elisa Ricci*
268 Hierarchical Concept Embedding & Pursuit for Interpretable Image Classification, *Nghia Nguyen, Tianjiao Ding, René Vidal*
269 Interpretable and Steerable Concept Bottleneck Sparse Autoencoders, *Akshay Kulkarni, Tsui-Wei Weng, Vivek Narayanaswamy, Shusen Liu, Wesam A. Sakla, Kowshik Thopalli*
270 C-LaV: Conditional Latent Velocity Field Denoising for Weather-Robust LiDAR Place Recognition, *Xuewei Cao, Jiayue Yang, Zhiwen Zeng, Yanyong Zhang, Yan Xia*
271 Towards Foundation Models for 3D Scene Understanding: Instance-Aware Self-Supervised Learning for Point Clouds, *Bin Yang, Mohamed Abdelsamad, Miao Zhang, Alexandru Paul Condurache*
272 ❊ Generalized-CVO: Fast and Correspondence-Free Local Point Cloud Registration with Second Order Riemannian Optimization, *Ray Zhang, Marcus Greiff, Thomas Lew, John Subosits*
273 LiDeRe: A Lightweight Readout for Fast and Data-Efficient Dense Prediction, *Timo Lüddecke, Jan Frederik Meier, Jan van Delden, Alexander Ecker*
274 AnyPcc: Compressing Any Point Cloud with a Single Universal Model, *Kangli Wang, Qianxi Yi, Yuqi Ye, Shihao Li, Wei Gao*
275 CoLC: Communication-Efficient Collaborative Perception with LiDAR Completion, *Yushan Han, Hui Zhang, Qiming Xia, Yi Jin, Yidong Li*
276 Spectral-Geometric Neural Fields for Pose-Free LiDAR View Synthesis, *Yinuo Jiang, Jun Cheng, Yiran Wang, Cheng Cheng*
277 C-GenReg: Training-Free 3D Point Cloud Registration by Multi-View-Consistent Geometry-to-Image Generation with Probabilistic Modalities Fusion, *Yuval Haitman, Amit Efraim, Joseph M. Francos*
278 PatchAlign3D: Local Feature Alignment for Dense 3D Shape Understanding, *Souhail Hadgi, Bingchen Gong, Ramana Sundararaman, Emery Pierson, Lei Li, Peter Wonka, Maks Ovsjanikov*
279 FoV-Net: Rotation-Invariant CAD B-rep Learning via Field-of-View Ray Casting, *Matteo Ballegeer, Dries F. Benoit*
280 Neural Distribution Prior for LiDAR Out-of-Distribution Detection, *Zizhao Li, Zhengkang Xiang, Jiayang Ao, Feng Liu, Joseph West, Kourosh Khoshelham*
281 ❊ DENALI: A Dataset Enabling Non-Line-of-Sight Spatial Reasoning with Low-Cost LiDARs, *Nikhil Behari, Diego Rivero, Luke Apostolides, Suman Ghosh, Paul Pu Liang, Ramesh Raskar*
282 Concept-Aware Batch Sampling Improves Language-Image Pretraining, *Adhiraj Ghosh, Vishaal Udandarao, Thao Nguyen, Matteo Farina, Mehdi Cherti, Jenia Jitsev, Sewoong Oh, Elisa Ricci, Ludwig Schmidt, Matthias Bethge*
283 HiFICL: High-Fidelity In-Context Learning for Multimodal Tasks, *Xiaoyu Li, Yuhang Liu, Xuanshuo Kang, Zheng Luo, Fangqi Lou, Xiaohua Wu, Zihan Xiong*
284 InstAP: Instance-Aware Vision-Language Pre-Train for Spatial-Temporal Understanding, *Ashutosh Kumar, Rajat Saini, Jingjing Pan, Mustafa Erdogan, Mingfang Zhang, Betty Le Dem, Norimasa Kobori, Quan Kong*
285 ❊ Vocabulary Scaling Law: Tuning Open-vocabulary Predictors for Their Openness, *Ziliang Chen, Yulu Li, Liangda Fang, Jusheng Zhang, Yongsen Zheng, Quanlong Guan, Xipeng Chen*
286 Render-to-Adapt: Unsupervised Personal Adaptation for Gaze Estimation, *Yangshi Ge, Zheng Liu, Feng Lu*
287 ViTPrompt: Training-Free Prompt Refinement with Visual Tokens for Open-Vocabulary Detection, *Yitong Qin, Lihua Zhou, Jiwei Wei, Ran Ran, Shiyuan He, Zeyu Ma, Shuaifeng Li, Nianxin Li, Heng Tao Shen*
288 Cluster-Aware Neural Collapse Prompt Tuning for Long-Tailed Generalization of Vision-Language Models, *Boyang Guo, Liang Li, Lin Peng, Yuhan Gao, Xichun Sheng, Chenggang Yan*
289 ❊ LLMind: Bio-inspired Training-free Adaptive Visual Representations for Vision-Language Models, *Soumyaratna Debnath, Bui Duc Manh, Zinan Liu, Lin Wang*
290 Dynamic Logits Adjustment and Exploration for Test-Time Adaptation in Vision Language Models, *Haoyan Wu, Yahao Liu, Yinjie Lei, Lixin Duan, Wen Li*
291 CAPT: Confusion-Aware Prompt Tuning for Reducing Vision-Language Misalignment, *Maoyuan Shao, Yutong Gao, Xinyang Huang, Lijuan Sun, Guoshun Nan, Chuang Zhu*
292 GenMatter: Perceiving Physical Objects with Generative Matter Models, *Eric Li, Arijit Dasgupta, Yoni Friedman, Mathieu Huot, Vikash Mansinghka, Thomas O'Connell, William T. Freeman, Joshua B. Tenenbaum*
293 Bidirectional Query-Driven Generation of Parametric CAD Sketch, *Yang Liu, Daxuan Ren, Yijie Ding, Jianmin Zheng, Fang Deng*
294 The Missing GAP: From Solving Square Jigsaw Puzzles to Handling Real World Archaeological Fragments, *Ofir Itzhak Shahar, Gur Elkin, Ohad Ben-Shahar*

295 Are We Ready for RL in Text-to-3D Generation? A Progressive Investigation, *Yiwen Tang, Zoey Guo, Kaixin Zhu, Ray Zhang, Qizhi Chen, Dongzhi Jiang, Junli Liu, Bohan Zeng, Haoming Song, Delin Qu, Tianyi Bai, Dan Xu, Wentao Zhang, Bin Zhao*
296 OmniDocLayout: Towards Diverse Document Layout Generation
❊ via Coarse-to-Fine LLM Learning, *Hengrui Kang, Zhuangcheng Gu, Zhiyuan Zhao, Zichen Wen, Bin Wang, Weijia Li, Conghui He*
297 Yo'City: Personalized and Boundless 3D Realistic City Scene Generation via Self-Critic Expansion, *Keyang Lu, Sifan Zhou, Hongbin Xu, Gang Xu, Zhifei Yang, Yikai Wang, Zhen Xiao, Jieyi Long, Ming Li*
298 Repurposing 3D Generative Model for Autoregressive Layout Generation, *Haoran Feng, Yifan Niu, Zehuan Huang, Yang-Tian Sun, Chunchao Guo, Yuxin Peng, Lu Sheng*
299 CAD-Refiner: A Unified Framework for CAD Generation and Iterative Editing, *Meng Yuan, Dawei Lin, Hongxia Xie, Tieru Wu, Rui Ma*
300 A Debiased Reconstruction-based Framework for Training-
❊ *Free Detection of AI-Generated Images, Sungik Choi, Hankook Lee, Jaehoon Lee, Robin Kim, Stanley Jungkyu Choi, Moontae Lee*
301 Global Information Thresholding for Sufficient and Necessary Circuits, *Jegyeong Cho*
302 PrivateEyes: Gaze-Preserving Anonymization for Data Sharing, *Surabhi Gupta, Dinesh Prabhu Muthumariappan, Biplab Das, Anoop Kolar Rajagopal, Kiran Nanjunda Iyer, Donghwan Seo*
303 From Measurement to Mitigation: Quantifying and Reducing Identity Leakage in Image Representation Encoders with Linear Subspace Removal, *Daniel George, Charles Yeh, Daniel Lee, Yifei Zhang*
304 Bias In, Bias Out? Finding Unbiased Subnetworks in Vanilla Models, *Ivan Luiz De Moura Matos, Abdel Djalil Sad Saoud, Ekaterina Iakovleva, Vito Paolo Pastore, Enzo Tartaglione*
305 pH-Strips for Selective Forgetting: A Blunt but Fast Diagnostic Baseline for Machine Unlearning, *Chengyao Qian, Jing Wu, Trung Le, Dinh Phung, Mehrtash Harandi*
306 Decoupling Defense Strategies for Robust Image Watermarking, *Jiahui Chen, Zehang Deng, Zeyu Zhang, Chaoyang Li, Lianchen Jia, Lifeng Sun*
307 Unsafe2Safe: Controllable Image Anonymization for Downstream
❊ Utility, *Minh Dinh, SouYoung Jin*
308 Rel-Zero: Harnessing Patch-Pair Invariance for Robust Zero-Watermarking Against AI Editing, *Pengzhen Chen, Yanwei Liu, Xiaoyan Gu, Xiaojun Chen, Wu Liu, Weiping Wang*
309 Computation and Communication Efficient Federated Unlearning via On-server Gradient Conflict Mitigation and Expression, *Minh-Duong Nguyen, Senura Wanasekara, Le-Tuan Nguyen, Quoc-Viet Pham, Ken-Tye Yong, Nguyen H. Tran, Dung D. Le*
310 DP-FedAdamW: An Efficient Optimizer for Differentially Private Federated Large Models, *Jin Liu, Ning Xi, Yinbin Miao, Junkang Liu*
311 Submodel Extraction for Efficient and Personalized Federated Learning via Optimal Transport, *Zheng Jiang, Nan He, Yiming Chen, Lifeng Sun*
312 FedSDR: Federated Graph Learning with Structural Noise Detection
❊ and Reconstruction, *Jiaqi Liu, Zihan Tan, Guancheng Wan, Wenke Huang, He Li, Mang Ye*
313 FedDAP: Domain-Aware Prototype Learning for Federated Learning under Domain Shift, *Huy Q. Le, Loc X. Nguyen, Yu Qiao, Seong Tae Kim, Eui-Nam Huh, Choong Seon Hong*
314 FedAFD: Multimodal Federated Learning via Adversarial Fusion and Distillation, *Min Tan, Junchao Ma, Yinfu Feng, Jiajun Ding, Wenwen Pan, Tingting Han, Qian Zheng, Zhenzhong Kuang, Zhou Yu*
315 VIRST: Video-Instructed Reasoning Assistant for SpatioTemporal Segmentation, *Jihwan Hong, Jaeyoung Do*
316 AXG-Reasoner: Error Detection and Explanation in Long Task Videos with Vision–Language Models, *Shih-Po Lee, Ehsan Elhamifar*
317 Stay in your Lane: Role Specific Queries with Overlap Suppression Loss for Dense Video Captioning, *Seung Hyup Baek, Jimin Lee, Hyeongkeun Lee, Jae Won Cho*
318 T2SGrid: Temporal-to-Spatial Gridification for Video Temporal Grounding, *Chaohong Guo, Yihan He, Yongwei Nie, Fei Ma, Xuemiao Xu, Chengjiang Long*
319 HanDyVQA: A Video QA Benchmark for Fine-Grained Hand-Object
❊ Interaction Dynamics, *Masatoshi Tateno, Gido Kato, Hirokatsu Kataoka, Yoichi Sato, Takuma Yagi*
320 SAIL: Similarity-Aware Guidance and Inter-Caption Augmentation-based Learning for Weakly-Supervised Dense Video Captioning, *Ye-Chan Kim, SeungJu Cha, Si-Woo Kim, Minju Jeon, Hyungee Kim, Dong-Jin Kim*
321 Token Warping Helps MLLMs Look from Nearby Viewpoints, *Phillip Y. Lee, Chanho Park, Mingue Park, Seungwoo Yoo, Juil Koo, Minhyuk Sung*
322 Variation-aware Vision Token Dropping for Faster Large Vision-Language Models, *Junjie Chen, Xuyang Liu, Zichen Wen, Yiyu Wang, Siteng Huang, Honggang Chen*
323 Fine-Grained Post-Training Quantization for Large Vision Language Models with Quantization-Aware Integrated Gradients, *Ziwei Xiang, Fanhu Zeng, Hongjian Fang, Rui-Qi Wang, Renxing Chen, Yanan Zhu, Yi Chen, Peipei Yang, Xu-Yao Zhang*
324 Blink: Dynamic Visual Token Resolution for Enhanced Multimodal Understanding, *Yuchen Feng, Zhenyu Zhang, Naibin Gu, Yilong Chen, Peng Fu, Zheng Lin, Shuohuan Wang, Yu Sun, Hua Wu, Weiping Wang, Haifeng Wang*
325 IF-Prune: Information-Flow Guided Token Pruning for Efficient Vision-Language Models, *Guohao Sun, Yufei Wang, Sizhuo Ma, Yuege Xie, Yuting Cheng, Zhiqiang Tao, Jian Wang*
326 EvoComp: Learning Visual Token Compression for Multimodal Large Language Models via Semantic-Guided Evolutionary Labeling, *Jiafei Song, Fengwei Zhou, Jin Qu, Wenjin Jason Li, Tong Wu, Gengjian Xue, Zhikang Zhao, Daomin Wei, Yichao Lu, Bailin Na*
327 DocPrune: Efficient Document Question Answering via Background, Question, and Comprehension-aware Token Pruning, *Joonmyung Choi, Sanghyeok Lee, Jongha Kim, Sehyung Kim, Dohwan Ko, Jihyung Kil, Hyunwoo J. Kim*
328 QuietPrune: Query-Guided Early Token Pruning for Vision-Language Models, *Tianxiao Gao, Shanwei Zhao, Shuo Fang, Shiai Zhu, Chenguang Ma*
329 The Devil Is in Gradient Entanglement: Energy-Aware Gradient
❊ Coordinator for Robust Generalized Category Discovery, *Haiyang Zheng, Nan Pu, Yaqi Cai, Teng Long, Wenjing Li, Nicu Sebe, Zhun Zhong*
330 LLM-Guided Probabilistic Fusion for Label-Efficient Document Layout Analysis, *Ibne Farabi Shihab, Sanjeda Akter, Anuj Sharma*
331 Coordinate Denoising for Non-Equilibrium Molecular Representation Learning, *Qianwei Tang, Baile Xu, Jian Zhao, Furao Shen*
332 Plug-and-Play Incomplete Multi-View Clustering via Janus-Faced
❊ Affinity Learning with Topology Harmonization, *Shengju Yu, Suyuan Liu, Wenhao Shao, Siwei Wang, Ke Liang, Xihong Yang, Tiejun Li, Xinwang Liu*
333 Meta-Learning In-Context Enables Training-Free Cross Subject Brain Decoding, *Mu Nan, Muquan Yu, Weijian Mai, Jacob S. Prince, Hossein Adeli, Rui Zhang, Jiahang Cao, Benjamin Becker, John A. Pyles, Margaret M. Henderson, Chunfeng Song, Nikolaus Kriegeskorte, Michael J. Tarr, Xiaoqing Hu, Andrew F. Luo*
334 Measure The Feature Universe: Topology-based Pseudo Labeling and Gravity Consistency for Source-Free Domain Adaptation, *Jae Yun Lee, Hyeok Nam, Sung In Cho*
335 Conditional Factuality Controlled LLMs with Generalization Certificates via Conformal Sampling, *Kai Ye, Qingtao Pan, Shuo Li*
336 Harnessing the Power of Foundation Models for Accurate Material Classification, *Qingran Lin, Fengwei Yang, Chaolun Zhu*
337 Content-Aware Frequency Encoding for Implicit Neural
❊ Representations with Fourier-Chebyshev Features, *Junbo Ke, Yangyang Xu, Chao Wang, You-Wei Wen*
338 ActiveAD: Planning-Oriented Active Learning for End-to-End
❊ Autonomous Driving, *Han Lu, Xiaosong Jia, Yichen Xie, Siyu Sun, Wenlong Liao, Xiaokang Yang, Junchi Yan*
339 TeFlow: Enabling Multi-frame Supervision for Self-Supervised
❊ Feed-forward Scene Flow Estimation, *Qingwen Zhang, Chenhan Jiang, Xiaomeng Zhu, Yunqi Miao, Yushan Zhang, Olov Andersson, Patric Jensfelt*
340 Think Before You Drive: World Model-Inspired Multimodal Grounding, *Haicheng Liao, Huanming Shen, Bonan Wang, Yongkang Li, Yihong Tang, Chengyue Wang, Dingyi Zhuang, Kehua Chen, Hai Yang, Chengzhong Xu, Zhenning Li*

341 DrivePI: Spatial-aware 4D MLLM for Unified Autonomous Driving Understanding, Perception, Prediction and Planning, *Zhe Liu, Runhui Huang, Rui Yang, Siming Yan, Zining Wang, Lu Hou, Di Lin, Xiang Bai, Hengshuang Zhao*
342 DrivePTS: A Progressive Learning Framework with Textual and Structural Enhancement for Driving Scene Generation, *Zhechao Wang, Yiming Zeng, Lufan Ma, Zeqing Fu, Chen Bai, Dongshuo Yin, Ziyao Lin, Cheng Lu*
343 WOD-E2E: Waymo Open Dataset for End-to-End Driving in Challenging Long-tail Scenarios, *Runsheng Xu, Hubert Lin, Wonseok Jeon, Hao Feng, Yuliang Zou, Liting Sun, John Gorman, Kate Tolstaya, Sarah Tang, Brandyn White, Ben Sapp, Mingxing Tan, Jyh-Jing Hwang, Dragomir Anguelov*
344 ❊ GuideFlow: Constraint-Guided Flow Matching for Planning in End-to-End Autonomous Driving, *Lin Liu, Caiyan Jia, Guanyi Yu, Ziying Song, Junqiao Li, Feiyang Jia, Peiliang Wu, Xiaoshuai Hao, Yadan Luo*
345 ResAD: Normalized Residual Trajectory Modeling for End-to-End Autonomous Driving, *Zhiyu Zheng, Shaoyu Chen, Haoran Yin, Xinbang Zhang, Jialv Zou, Xinggang Wang, Qian Zhang, Lefei Zhang*
346 KnowVal: A Knowledge-Augmented and Value-Guided Autonomous Driving System, *Zhongyu Xia, Wenhao Chen, Yongtao Wang, Ming-Hsuan Yang*
347 FoSS: Modeling Long-Range Dependencies and Multimodal Uncertainty in Trajectory Prediction via Fourier–State Space Integration, *Yizhou Huang, Genze Jiang, Yihua Cheng, Kezhi Wang*
348 NexusFlow: Unifying Disparate Tasks under Partial Supervision via Invertible Flow Networks, *Fangzhou Lin, Yuping Wang, Yuliang Guo, Zixun Huang, Xinyu Huang, Haichong Zhang, Kazunori Yamada, Zhengzhong Tu, Liu Ren, Ziming Zhang*
349 Visual Prototype Conditioned Focal Region Generation for UAV-Based Object Detection, *Wenhao Li, Zimeng Wu, Yu Wu, Zehua Fu, Jiaxin Chen*
350 Consistent Instance Field for Dynamic Scene Understanding, *Junyi Wu, Van Nguyen Nguyen, Benjamin Planche, Jiachen Tao, Changchang Sun, Zhongpai Gao, Zhenghao Zhao, Anwesa Choudhuri, Gengyu Zhang, Meng Zheng, Feiran Wang, Terrence Chen, Yan Yan, Ziyan Wu*
351 CLP: A Real-World Dataset of Contaminated Lens Protectors for Robust Semantic Segmentation, *Sungyong Park, Sooyoung Choi, Hyunsuh Koh, Youngjae Choi, Heewon Kim*
352 ReSAM: Refine, Requery, and Reinforce: Self-Prompting Point-Supervised Segmentation for Remote Sensing Images, *Muhammad Naseer Subhani*
353 ❊ Heuristic Self-Paced Learning for Domain Adaptive Semantic Segmentation under Adverse Conditions, *Shiqin Wang, Haoyang Chen, Huaizhou Huang, Yinkan He, Dongfang Sun, Xiaoqing Chen, Xingyu Liu, Zheng Wang, Kaiyan Zhao*
354 SAM2Text: Towards Prompt-Free and Multi-Resolution Video Scene Text Segmentation, *Jing-Yao Zhang, Heng Zhang, Mingsen Zhang, Binbin Yang, Fei Yin*
355 Reinforcing Video Reasoning Segmentation to Think Before It Segments, *Sitong Gong, Yunzhi Zhuge, Lu Zhang, Jiazuo Yu, Pingping Zhang, Xu Jia, Huchuan Lu*
356 VideoMaMa: Mask-Guided Video Matting via Generative Prior, *Sangbeom Lim, Seoung Wug Oh, Jiahui Huang, Heeji Yoon, Seungryong Kim, Joon-Young Lee*
357 Quantized Residuals to Continuous Prompts for Few-Shot Class Incremental Learning in Vision-Language Models, *Abhishek Kumar Sinha, Nitant Dube, Soma Biswas*
358 The Golden Subspace: Where Efficiency Meets Generalization in Continual Test-Time Adaptation, *Guannan Lai, Da-Wei Zhou, Zhenguo Li, Han-Jia Ye*
359 SAIDO: Generalizable Detection of AI-Generated Images via Scene-Aware and Importance-Guided Dynamic Optimization in Continual Learning, *Yongkang Hu, Yu Cheng, Yushuo Zhang, Yuan Xie, Zhaoxia Yin*
360 Is Parameter Isolation Better for Prompt-Based Continual Learning?, *Jiangyang Li, Chenhao Ding, SongLin Dong, Qiang Wang, Jianchao Zhao, Yuhang He, Yihong Gong*
361 Octopus: History-Free Gradient Orthogonalization for Continual Learning in Multimodal Large Language Models, *Yuehao Liu, Shanyan Guan, Weijia Zhang, Xuanming Shang, Yanhao Ge, Wei Li, Chao Ma*
362 Affordance-First Decomposition for Continual Learning in Video–Language Understanding, *Mengzhu xu, Hanzhi Liu, Ningkang Peng, Qianyu Chen, Canran Xiao*
363 Quantum-Gated Task-interaction Knowledge Distillation for Pre-trained Model-based Class-Incremental Learning, *Linjie Li, Huiyu Xiao, Jiarui Cao, Zhenyu Wu, Yang Ji*
364 Elastic Weight Consolidation Done Right for Continual Learning, *Xuan Liu, Xiaobin Chang*
365 On Token's Dilemma: Dynamic MoE with Drift-Aware Token Assignment for Continual Learning of Large Vision Language Models, *Chongyang Zhao, Mingsong Li, Haodong Lu, Dong Gong*
366 Soul: Breathe Life into Digital Human for High-fidelity Long-term Multimodal Animation, *Jiangning Zhang, Junwei Zhu, Zhenye Gan, Donghao Luo, Chuming Lin, FeiFan Xu, Xu Peng, Jianlong Hu, Yuansen Liu, Yijia Hong, Weijian Cao, Han Feng, Xu Chen, Chencan Fu, Keke He, Xiaobin Hu, Chengjie Wang*
367 Talking Together: Synthesizing Co-Located 3D Conversations from Audio, *Mengyi Shan, Shouchieh Chang, Ziqian Bai, Shichen Liu, Yinda Zhang, Luchuan Song, Rohit Pandey, Sean Fanello, Zeng Huang*
368 InfinityHuman: Towards Long-Term Audio-Driven Human Animation, *Xiaodi Li, Pan Xie, Yi Ren, Qijun Gan, Chen Zhang, Fangyuan Kong, Xiang Yin, Zehuan Yuan, Bingyue Peng*
369 ❊ Vanast: Virtual Try-On with Human Image Animation via Synthetic Triplet Supervision, *Hyunsoo Cha, Wonjung Woo, Byungjun Kim, Hanbyul Joo*
370 AudioAvatar: Personalized Audio-driven Whole-body Talking Avatars, *Seungeun Lee, SeungJun Moon, Hah Min Lew, Ji-Su Kang, Gyeong-Moon Park*
371 One-to-All Animation: Alignment-Free Character Animation and Image Pose Transfer, *Shijun Shi, Jing Xu, Zhihang Li, Chunli Peng, Xiaoda Yang, Lijing Lu, Kai Hu, Jiangning Zhang*
372 ❊ Counterfactual VLA: Self-Reflective Vision-Language-Action Model with Adaptive Reasoning, *Zhenghao Peng, Wenhao Ding, Yurong You, Yuxiao Chen, Wenjie Luo, Thomas Tian, Yulong Cao, Apoorva Sharma, Danfei Xu, Boris Ivanovic, Boyi Li, Yan Wang, Marco Pavone*
373 SGDrive: Scene-to-Goal Hierarchical World Cognition for Autonomous Driving, *Jingyu Li, Junjie Wu, Dongnan Hu, Xiangkai Huang, Bin Sun, Zhihui Hao, Xianpeng Lang, Xiatian Zhu, Li Zhang*
374 CapNav: Benchmarking Vision Language Models on Capability-conditioned Indoor Navigation, *Xia Su, Ruiqi Chen, Benlin Liu, Jingwei Ma, Zonglin Di, Ranjay Krishna, Jon Froehlich*
375 AutoTraces: Autoregressive Trajectory Forecasting via Multimodal Large Language Models, *Teng Wang, Yanting Lu, Ruize Wang*
376 AwareVLN: Reasoning with Self-awareness for Vision-Language Navigation, *Wenxuan Guo, Xiuwei Xu, Yichen Liu, Xiangyu Li, Hang Yin, Huangxing Chen, Wenzhao Zheng, Jianjiang Feng, Jie Zhou, Jiwen Lu*
377 Progress-Think: Semantic Progress Reasoning for Vision-Language Navigation, *Shuo Wang, Yucheng Wang, Guoxin Lian, Yongcai Wang, Maiyue Chen, Kaihui Wang, Bo Zhang, Zhizhong Su, Yutian Zhou, Wanting Li, Deying Li, Zhaoxin Fan*
378 Tavatar: Topology-Aware Gaussian Attribute Derivation for Animatable Human Avatars, *Hailin Luo, Yifan Yang, Jiazhi Shu, Zixiong Huang, Qi Chen, Qing Du, Mingkui Tan*
379 ❊ PercHead: Perceptual Head Model for Single-Image 3D Head Reconstruction & Editing, *Antonio Oroz, Matthias Nießner, Tobias Kirschstein*
380 PhysHead: Simulation-Ready Gaussian Head Avatars, *Berna Kabadayi, Vanessa Sklyarova, Wojciech Zielonka, Justus Thies, Gerard Pons-Moll*
381 ReWeaver: Towards Simulation-Ready and Topology-Accurate Garment Reconstruction, *Ming Li, Hui Shan, Kai Zheng, Chentao Shen, Siyu Liu, Yanwei Fu, Zhen Chen, Xiangru Huang*
382 FHAvatar: Fast and High-Fidelity Reconstruction of Face-and-Hair Composable 3D Head Avatar from Few Casual Captures, *Yujie Sun, Zhuoqiang Cai, Chaoyue Niu, Jianchuan Chen, Zhiwen Chen, Chengfei Lv, Fan Wu*
383 Feed-Forward One-Shot Animatable Textured Mesh Avatar Reconstruction, *Yisheng He*
384 Reallocating Attention Across Layers to Reduce Multimodal

Hallucination, *Haolang Lu, Bolun Chu, WeiYe Fu, Guoshun Nan, Junning Liu, Minghui Pan, Qiankun Li, Yi Yu, Hua Wang, Kun Wang*

385 VES-RFT: Rewarding Visual Evidence Sensitivity to Mitigate Hallucinations in Large Vision–Language Models, *Xuehe Hou, Wenshuo Li, Yali Li, Han Shu, Yuan Wang, Xinghao Chen, Shengjin Wang*

386 Fighting Hallucinations with Counterfactuals: Diffusion-Guided Perturbations for LVLM Hallucination Suppression, *Hamidreza Dastmalchi, Aijun An, Ali Cheraghian, Hamed Barzamini*

387 ❊ Unstitching the Chimera: Frame-Level Risk and Train-Free Mitigation for Video Hallucination, *Songyuan Yang, Guijian Tang, Kun Hu, Haotian Wang, Shixuan Liu, Wenjing Yang, Long Lan, Huibin Tan*

388 CausalLens: Sensitivity-Guided Multi-Head Causal Intervention for Hallucination Mitigation in Large Vision-Language Models, *Junyang Ji, Qifan Liu, Wenming Yang, Zhihai He*

389 Breaking the Illusion: When Positive Meets Negative in Multimodal Decoding, *Yubo Jiang, Yitong An, Xin Yang, Abudukelimu Wuerkaixi, Xuxin Cheng, Fengying Xie, Zhiguo Jiang, Cao Liu, Ke Zeng, Haopeng Zhang*

390 FlexTraj: Image-to-Video Generation with Flexible Point Trajectory Control, *Zhiyuan Zhang, Can Wang, Dongdong Chen, Jing Liao*

391 Diff4Splat: Repurposing Video Diffusion Models for Dynamic Scene Generation, *Panwang Pan, Chenguo Lin, Chenxin Li, Jingjing Zhao, Yuchen Lin, Haopeng Li, Yunlong Lin, Kairun Wen, Yixuan Yuan, Yadong MU*

392 Spatia: Video Generation with Updatable Spatial Memory, *Jinjing Zhao, Fangyun Wei, Zhening Liu, Hongyang Zhang, Chang Xu, Yan Lu*

393 Geometry-as-context: Modulating Explicit 3D in Scene-consistent Video Generation to Geometry Context, *JiaKui Hu, Jialun Liu, Liying Yang, Xinliang Zhang, Kaiwen Li, Shuang Zeng, Yuanwei Li, Haibin Huang, Chi Zhang, Yanye Lu*

394 EgoControl: Controllable Egocentric Video Generation via 3D Full-Body Poses, *Enrico Pallotta, Sina Mokhtarzadeh Azar, Lars Doorenbos, Serdar Ozsoy, Umar Iqbal, Juergen Gall*

395 CustomTex: High-fidelity Indoor Scene Texturing via Multi-Reference Customization, *Weilin Chen, Jiahao Rao, Wenhao Wang, Xinyang Li, Xuan Cheng, Liujuan Cao*

396 ❊ FoleyDesigner: Immersive Stereo Foley Generation with Precise Spatio-Temporal Alignment for Film Clips, *Mengtian Li, Kunyan Dai, Yi Ding, Ruobing Ni, Ying Zhang, Wenwu Wang, Zhifeng Xie*

397 Physical Simulator In-the-Loop Video Generation, *Lin Geng Foo, Mark He Huang, Alexandros Lattas, Stylianos Moschoglou, Thabo Beeler, Christian Theobalt*

398 ❊ Refracting Reality: Generating Images with Realistic Transparent Objects, *Yue Yin, Enze Tao, Dylan Campbell*

399 Generating Humanless Environment Walkthroughs from Egocentric Walking Tour Videos, *Yujin Ham, Junho Kim, Vivek Boominathan, Guha Balakrishnan*

400 EgoFlow: Gradient-Guided Flow Matching for Egocentric 6DoF Object Motion Generation, *Abhishek Saroha, Huajian Zeng, Xingxing Zuo, Daniel Cremers, Xi Wang*

401 Spatial-Frequency Collaborative Learning for Occluded Visible-Infrared Person Re-Identification, *Jian Yu, Yujian Feng, Shuai You, Zhongkai Zhou, Fei Wu, Zhengjun Jing, Yimu Ji*

402 Mind the Gap: Transferring Labels to Align Object Detection Datasets, *Mikhail Kennerley, Angelica I. Aviles-Rivero, Carola-Bibiane Schönlieb, Robby T. Tan*

403 ❊ SSM-Aware Token-Efficient VMamba via Adaptive Patch Pruning and Merging for Person Re-Identification, *Huiyuan Huang, Sang Min Yoon*

404 Tri-Modal Fusion Transformers for UAV-based Object Detection, *Craig Iaboni, Pramod Abichandani*

405 View-Aware Semantic Alignment for Aerial-Ground Person Re-Identification, *Quan Zhang, Zeqiang Cai, Peiming Zhao, Jingze Wu, Cailun Wu, Hongbo Chen, Jianhuang Lai*

406 RHCNet: Residual-Guided Hierarchical Calibration Network for Robust Underwater Object Detection, *Yueying Wang, Yiteng Guo, Weidong Zhang, Jie Wen, Liquan Shen, Huaicheng Yan, Xin Xu*

407 X-AVDT: Audio-Visual Cross-Attention for Robust Deepfake Detection, *Youngseo Kim, Kwan Yun, Seokhyeon Hong, Sihun Cha, Colette Suhjung Koo, Junyong Noh*

408 Beyond Duality: A Hybrid Framework of Leveraging Shared and Private Features for RGB-Event Object Detection, *Keyao Wang, Shuai Liu, Hengda Shi, Lukui Shi, Haiyong Chen*

409 FVBench: Benchmarking Deepfake Video Detection Capability of Large Multimodal Models, *Jiarui Wang, Huiyu Duan, Juntong Wang, Xiongkuo Min*

410 AKCMamba-YOLO: Selective State Space Models For Real-Time Object Detection, *Long Chen, Hui Wang, Man Xu, Zexuan Li, Zizhu Fan*

411 When AVSR Meets Video Conferencing: Dataset, Degradation, and the Hidden Mechanism Behind Performance Collapse, *Yihuan Huang, Jun Xue, Liu Jiajun, Daixian Li, Tong Zhang, Zhuolin Yi, Yanzhen Ren, Kai Li*

412 Your One-Stop Solution for AI-Generated Video Detection, *Long Ma, Zihao Xue, Yan Wang, Zhiyuan Yan, Jin Xu, Xiaorui Jiang, Haiyang Yu, Yong Liao, Zhen Bi*

413 UnityVideo: Unified Multi-Modal Multi-Task Learning for Enhancing World-Aware Video Generation, *Jiehui Huang, Yuechen Zhang, Xu He, Yuan Gao, Zhi Cen, Bin Xia, Yan Zhou, Xin Tao, Pengfei Wan, Jiaya Jia*

414 Skyra: AI-Generated Video Detection via Grounded Artifact Reasoning, *Yifei Li, Wenzhao Zheng, Yanran Zhang, Runze Sun, Yu Zheng, Lei Chen, Jie Zhou, Jiwen Lu*

415 HumanVBench: Probing Human-Centric Video Understanding in MLLMs with Automatically Synthesized Benchmarks, *Ting Zhou, Daoyuan Chen, Qirui Jiao, Bolin Ding, Yaliang Li, Ying Shen*

416 HERBench: A Benchmark for Multi-Evidence Integration in Video Question Answering, *Dan Ben Ami, Gabriele Serussi, Kobi Cohen, Chaim Baskin*

417 ❊ Seeing the Scene Matters: Revealing Forgetting in Video Understanding Models with a Scene-Aware Long-Video Benchmark, *Seng Nam Chen, Hao Chen, Chenglam Ho, Xinyu Mao, Jinping Wang, Yu Zhang, Chao Li*

418 Thinking with Frames: Generative Video Distortion Evaluation via Frame Reward Model, *Yuan Wang, Borui Liao, Huijuan Huang, Jinda Lu, Ouxiang Li, Kuien Liu, Meng Wang, Xiang Wang*

419 MovieRecapsQA: A Multimodal Open-Ended Video Question-Answering Benchmark, *Shaden Shaar, Bradon Thymes, Sirawut Chaixanien, Claire Cardie, Bharath Hariharan*

420 Training-free, Perceptually Consistent Low-Resolution Previews with High-Resolution Image for Efficient Workflows of Diffusion Models, *Wongi Jeong, Hoigi Seo, Se Young Chun*

421 One Model, Many Budgets: Elastic Latent Interfaces for Diffusion Transformers, *Moayed Haji-Ali, Willi Menapace, Ivan Skorokhodov, Dogyun Park, Anil Kag, Michael Vasilkovsky, Sergey Tulyakov, Vicente Ordonez, Aliaksandr Siarohin*

422 Reflection Separation from a Single Image via Joint Latent Diffusion, *Zheng-Hui Huang, Zhixiang Wang, Yu-Lun Liu, Yung-Yu Chuang*

423 MMFace-DiT: A Dual-Stream Diffusion Transformer for High-Fidelity Multimodal Face Generation, *Bharath Krishnamurthy, Ajita Rattani*

424 DisCa: Accelerating Video Diffusion Transformers with Distillation-Compatible Learnable Feature Caching, *Chang Zou, Changlin Li, Songtao Liu, Zhao Zhong, Kailin Huang, Linfeng Zhang*

425 ❊ MatLat: Material Latent Space for PBR Texture Generation, *Kyeongmin Yeo, Yunhong Min, Jaihoon Kim, Minhyuk Sung*

426 ❊ VMonarch: Efficient Video Diffusion Transformers with Structured Attention, *Cheng Liang, Haoxian Chen, Liang Hou, Qi Fan, Gangshan Wu, Xin Tao, Limin Wang*

427 DiffDecompose: Layer-Wise Decomposition of Alpha-Composited Images via Diffusion Transformers, *Zitong Wang, Hang Zhao, Qianyu Zhou, Xuequan Lu, Xiangtai Li, Hao Yang, Bo Yang, Yiren Song*

428 Calibri: Enhancing Diffusion Transformers via Parameter-Efficient Calibration, *Danil Tokhchukov, Aysel Mirzoeva, Andrey Kuznetsov, Konstantin Sobolev*

429 Transition Matching Distillation for Fast Video Generation, *Weili Nie, Julius Berner, Nanye Ma, Chao Liu, Saining Xie, Arash Vahdat*

430 Diffusion-Based Makeup Transfer with Facial Region-Aware Makeup Features, *Zheng Gao, Debin Meng, Yunqi Miao, Zhensong Zhang, Songcen Xu, Ioannis Patras, Jifei Song*

431 UniPR: Unified Object-level Real-to-Sim Perception and Reconstruction from a Single Stereo Pair, *Chuanrui Zhang, Yingshuang Zou, ZhengXian Wu, Yonggen Ling, Yuxiao Yang, Ziwei Wang*

432 Query2Uncertainty: Robust Uncertainty Quantification and Calibration for 3D Object Detection under Distribution Shift, *Till Beemelmanns, Alexey Nekrasov, Stefan Vilceanu, Jonas Steinhaus, Timo Woopen, Bastian Leibe, Lutz Eckstein*
433 DICArt: Advancing Category-level Articulated Object Pose Estimation in Discrete State-Spaces, *Li Zhang, Mingyu Mei, Ailing Wang, Xianhui Meng, Yan Zhong, Xinyuan Song, Liu Liu, Rujing Wang, Zaixing He, Cewu Lu*
434 PoseGaussian: 6D Pose Estimation for Unseen Objects via Sparse-View Object-Level 3D Gaussian Splatting, *Wubin Shi, Shaoyan Gai, Feipeng Da*
435 VGGT-Det: Mining VGGT Internal Priors for Sensor-Geometry-Free Multi-View Indoor 3D Object Detection, *Yang Cao, Feize Wu, Dave Zhenyu Chen, Yingji Zhong, Lanqing Hong, Dan Xu*
436 MonoSAOD: Monocular 3D Object Detection with Sparsely Annotated Label, *Junyoung Jung, Seokwon Kim, Jung Uk Kim*
437 V2U4Real: A Real-world Large-scale Dataset for Vehicle-to-UAV Cooperative Perception, *Weijia Li, Haoen Xiang, Tianxu Wang, Shuaibing Wu, Qiming Xia, Cheng Wang, Chenglu Wen*
438 SketchVL: Policy Optimization via Fine-Grained Credit Assignment for Chart Understanding and More, *Muye Huang, Lingling Zhang, Yifei Li, Yaqiang Wu, Jun Liu*
439 A Causal Marriage between VLM and IRM from Understanding to Reasoning, *Ziliang Chen, Tianang Xiao, Jusheng Zhang, Yongsen Zheng, Yang Liu, Zhao-rong Lai, Liang Lin*
440 Why Does RL Generalize Better Than SFT? A Data-Centric Perspective on VLM Post-Training, *Aojun Lu, Tao Feng, Hangjie Yuan, Wei Li, Yanan Sun*
441 SoC: Semantic Orthogonal Calibration for Test-Time Prompt Tuning, *Leo Fillioux, Omprakash Chakraborty, Ismail Ben Ayed, Paul-Henry Cournède, Stergios Christodoulidis, Maria Vakalopoulou, Jose Dolz*
442 Learning to Select Visual Tools from Experience, *Zeyi Huang, Yuyang Ji, Anirudh Sundara Rajan, Zefan Cai, Wen Xiao, Haohan Wang, Junjie Hu, Yong Jae Lee*
443 ❊ Agile Deliberation: Concept Deliberation for Subjective Visual Classification, *Leijie Wang, Otilia Stretcu, Wei Qiao, Thomas Denby, Krishnamurthy Viswanathan, Enming Luo, Chun-Ta Lu, Tushar Dogra, Ranjay Krishna, Ariel Fuxman*
444 Tea-Adapter: Teacher Adapter for Efficient Conditional Generation, *Yinhan Zhang, Yue Ma, Fangqiu Yi, Chenyang Qi, Chi Zhang, Kunyu Feng, Zeyu Wang*
445 From Failure to Feedback: Group Revision Unlocks Hard Cases in Object-Level Grounding, *Yuyuan Liu, Yiping Ji, Anjie Le, Jiayuan Zhu, Jiazhen Pan, Can Peng, Jiajun Deng, Fengbei Liu, Junde Wu*
446 Perception Characteristics Distance: Measuring Stability and Robustness of Perception System in Dynamic Conditions under a Certain Decision Rule, *Boyu Jiang, Liang Shi, Zhengzhi Lin, Lanxin Xiang, Loren Stowe, Feng Guo*
447 FinPercep-RM: A Fine-grained Reward Model and Co-evolutionary Curriculum for RL-based Real-world Super-Resolution, *Yidi Liu, Zihao Fan, Jie Huang, Jie Xiao, Dong Li, Wenlong Zhang, LEI BAI, Xueyang Fu, Zheng-Jun Zha*
448 Twin-T & TwintVQA: A Reliable Structure–Detail Separating VLM and a Comprehensive Benchmark for Chart and Table Tasks, *Jiahua Bao, Siyao Cheng, Jiaxing Du, Qingtao Xia, Changjiang He, Zeming Lang, Jie Liu*
449 SDGS: Spatial Difference Guided Gaussian Splatting for Simultaneous Localization and 3D Reconstruction, *Yijian Tian, Mingtao Ou, Zijian Pan, Xinglong Ji*
450 ❊ RT-Splatting: Joint Reflection-Transmission Modeling with Gaussian Splatting, *Ji Shi, Xianghua Ying, Bowei Xing, Ruohao Guo, Wenzhen Yue*
451 Pose-Free Omnidirectional Gaussian Splatting for 360-Degree Videos with Consistent Depth Priors, *Chuanqing Zhuang, Xin Lu, Zehui Deng, Zhengda Lu, Yiqun Wang, Junqi Diao, Jun Xiao*
452 Distilling Unsigned Distance Function for Surface Reconstruction from 3D Gaussian Splatting, *Qian Li, Rao Fu, Jiangtao Li, Fan Liu*
453 Exact-GS: Mathematically Rigorous and Accurate 3D Gaussian Splatting for 3D X-ray Reconstruction, *Guangpu Yang, Steffen Kieß, Hanxiang Luo, Xingyu Liu, Sven Simon*
454 DualSplat: Robust 3D Gaussian Splatting via Pseudo-Mask Bootstrapping from Reconstruction Failures, *Xu Wang, Zhiru Wang, Shiyun Xie, Chengwei Pan, Yisong Chen*
455 ❊ E2EGS: Event-to-Edge Gaussian Splatting for Pose-Free 3D Reconstruction, *Yunsoo Kim, Changki Sung, Dasol Hong, Hyun Myung*
456 Neural Gabor Splatting: Enhanced Gaussian Splatting with Neural Gabor for High-frequency Surface Reconstruction, *Haato Watanabe, Nobuyuki Umetani*
457 ❊ DirectFisheye-GS: Enabling Native Fisheye Input in Gaussian Splatting with Cross-View Joint Optimization, *Zhengxian Yang, Fei Xie, Xutao Xue, Rui Zhang, Taicheng Huang, Yang Liu, Mengqi Ji, Tao Yu*
458 VAD-GS: Visibility-Aware Densification for 3D Gaussian Splatting in Dynamic Urban Scenes, *Yikang Zhang, Rui Fan*
459 GauMVC: Generative Decoupled Gaussian Representation for Human-centric Multi-view Video Compression, *Ruoke Yan, Mingjia Yang, Xinfeng Zhang, Haocheng Tang, Qian Yin, Zhipin Deng, Kai Zhang, Li Zhang, Siwei Ma*
460 A Geometric Algebra-Informed 3DGS Framework for Wireless Channel Prediction, *Jingzhou Shen, Tianya Zhao, Xuyu Wang*
461 RaGS: Unleashing 3D Gaussian Splatting from 4D Radar and Monocular Cue for 3D Object Detection, *Xiaokai Bai, Chenxu Zhou, Lianqing Zheng, Jianan Liu, Si-Yuan Cao, Xiaohan Zhang, Yiming Li, Zhengzhuang Zhang, Hui-Liang Shen*
462 Cross-Instance Gaussian Splatting Registration via Geometry-Aware Feature-Guided Alignment, *Roy Amoyal, Oren Freifeld, Chaim Baskin*
463 ActivePolicy: Active Gaussian Reconstruction and Optimization Strategy Based on Global-Local Information Gain, *Yingzhao Li, Yanjie Liu, Lijun Zhao*
464 Uncertainty-driven 3D Gaussian Splatting Active Mapping via Anisotropic Visibility Field, *Shangjie Xue, Jesse Dill, Dhruv Ahuja, Frank Dellaert, Panagiotis Tsiotras, Danfei Xu*
465 SV-GS: Sparse View 4D Reconstruction with Skeleton-Driven Gaussian Splatting, *Jun-Jee Chao, Volkan Isler*
466 NimbusGS: Unified 3D Scene Reconstruction under Hybrid Weather, *Yanying Li, Jinyang Li, Shengfeng He, Yangyang Xu, Junyu Dong, Yong Du*
467 SparseSplat: Towards Applicable Feed-Forward 3D Gaussian Splatting with Pixel-Unaligned Prediction, *Zicheng Zhang, Xiangting Meng, Ke Wu, Wenchao Ding*
468 REVISOR: Beyond Textual Reflection, Towards Multimodal Introspective Reasoning in Long-Form Video Understanding, *Jiaze Li, Hao Yin, Wenhui Tan, Jingyang Chen, Boshen Xu, Yuxun Qu, Yijing Chen, Jianzhong Ju, Zhenbo Luo, Jian Luan*
469 Fast-ThinkAct: Efficient Vision-Language-Action Reasoning via Verbalizable Latent Planning, *Chi-Pin Huang, Yunze Man, Zhiding Yu, Min-Hung Chen, Jan Kautz, Yu-Chiang Frank Wang, Fu-En Yang*
470 Unlocking Token Rewards via Training-Free Reward Attribution, *Sitong Wu, Haoru Tan, Bin Xia, Xichen Zhang, Jingyao Li, Shaofeng Zhang, Xiaojuan Qi, Bei Yu, Jiaya Jia*
471 MedMO: Grounding and Understanding Multimodal Large Language Model for Medical Images, *Ankan Deria, Komal Kumar, Adinath Madhavrao Dukre, Eran Segal, Salman Khan, Imran Razzak*
472 When to Think and When to Look: Uncertainty-Guided Lookback, *Jing Bi, Filippos Bellos, JunJia Guo, Yayuan Li, Chao Huang, Yolo Tang, Luchuan Song, Susan Liang, Zhongfei Zhang, Jason J Corso, Chenliang Xu*
473 StaR-KVQA: Structured Reasoning Traces for Implicit-Knowledge Visual Question Answering, *Zhihao Wen, Wenkang Wei, Yuan Fang, Xingtong Yu, Hui Zhang, Weicheng Zhu, Xin Zhang*
474 Understanding Counting Mechanisms in Large Language and Vision-Language Models, *Hosein Hasani, Amirmohammad Izadi, Fatemeh Askari, Mobin Bagherian, Sadegh Mohammadian, Mohammad Izadi, Mahdieh Soleymani Baghshah*
475 CLiViS: Unleashing Cognitive Map through Linguistic-Visual Synergy for Embodied Visual Reasoning, *Kailing Li, Qi'ao Xu, Tianwen Qian, Yuqian Fu, Yang Jiao, Xiaoling Wang*
476 Proof-of-Perception: Certified Tool-Using Multimodal Reasoning with Compositional Conformal Guarantees, *Arya Fayyazi, Haleh Akrami*
477 Thinking Diffusion: Penalize and Guide Visual-Grounded Reasoning in Diffusion Multimodal Language Models, *Keuntae Kim, Mingyu Kang, Yong Suk Choi*

478 Don't Show Pixels, Show Cues: Unlocking Visual Tool Reasoning in Language Models via Perception Programs, *Muhammad Kamran Janjua, Hugo Silva, Di Niu, Bahador Rashidi*
479 Hugging Visual Prompt and Segmentation Tokens: Consistency Learning for Fine-Grained Visual Understanding in MLLMs, *Jing Yang, Sen Yang, Boqiang Duan, Ming Dai, Wei Zhang, Xiao Tan, Kunbin Chen, Wei He, Jingdong Wang, Hanli Wang*
480 VisionLeaf: Entropy-Guided Leaf-First Reasoning for Efficient and Accurate Think-with-Image, *Haokun Gui, Senqiao Yang, Mingkang Zhu, Meng Chu, Sitong Wu, Changsheng Lu, Zihao Wang, Zhuotao Tian, Jiaya Jia*
481 GGBench: A Geometric Generative Reasoning Benchmark for Unified Multimodal Models, *Jingxuan Wei, Caijun Jia, Xi Bai, Xinglong Xu, Siyuan Li, Linzhuang Sun, Bihui Yu, Conghui He, Lijun Wu, Cheng Tan*
482 Beyond Depth: Evaluating the Width-centric Reasoning Capability of MLLMs, *Mingrui Chen, Hexiong Yang, Haogeng Liu, Huaibo Huang, Ran He*
483 GenSplat: Bridging the Generalization Gap in 3DGS Language Comprehension, *Fang Liu, Yuhao Liu, Ke Xu, Gerhard Petrus Hancke, Rynson W. H. Lau*
484 CC-VQA: Conflict- and Correlation-Aware Method for Mitigating Knowledge Conflict in Knowledge-Based Visual Question Answering, *Yuyang Hong, Jiaqi Gu, Yujing Lou, Lubin Fan, Qi Yang, Ying Wang, Kun Ding, Yue Wu, Shiming Xiang, Jieping Ye*
485 LoPrune: Efficient Data Pruning for LoRA-Based Fine-Tuning of Vision Transformer, *Qiang He, Yaozong Yang, Kaibin Wang, Ziteng Wei, Feifei Chen, Caslon Chua, Yun Yang*
486 Multi-Scale Local Speculative Decoding for Image Generation, *Elia Peruzzo, Guillaume Sautière, Amirhossein Habibian*
487 Globscope: Toward a Global View of the Loss Landscape, *Mashiat Mustaq, Xavier M. Tricoche*
488 RADAR: VQ-VAE Decoder of VAR is a Good Student for Restoring Against Degradation by Acceleration, *Ziyang Wang, Yue Zhang, Mingdao Wang, Yasen Zhang, Teer Song, Yu Tian, Xueming Li*
489 ❊ Beyond Single Solution: Multi-Hypothesis Deep Unfolding Network for Image Compressive Sensing, *Wenxue Cui, Hualin Li, Yuhang Qin, Yifu Xu, Xiaopeng Fan, Debin Zhao*
490 FlashDecoder: Real-Time Latent-to-Pixel Streaming Decoder with Transformers, *Minguk Kang, Suha Kwak*
491 MambaSIC: Mamba-based Stereo Image Compression with Bi-directional Multi-reference Entropy Model, *Shiyu Qin, Xinjie Zhang, Zhening Liu, Jinpeng Wang, Bin Chen, Jiawei Li, Yifan Ren, Shu-Tao Xia, Jun Zhang*
492 Neural Dynamic GI: Random-Access Neural Compression for Temporal Lightmaps in Dynamic Lighting Environments, *Jianhui Wu, Jian Zhou, Zhi Zhou, Zhangjin Huang, Chao Li*
493 Discovering Adaptive Task Dependencies for Efficient Multi-Task Representation Compression, *Zhimeng Huang, Rongao Yuan, Junlong Gao, Qi Mao, Siwei Ma, Wen Gao, Chuanmin Jia*
494 OmniZip: Learning a Unified and Lightweight Lossless Compressor for Multi-Modal Data, *Yan Zhao, Zhengxue Cheng, Junxuan Zhang, Dajiang Zhou, Qunshan Gu, Qi Wang, Li Song*
495 Perceptual Neural Video Compression with Color Separation and Rank Chain, *Xiongzhuang Liang, Chuanbo Tang, Zhuoyuan Li, Li Li, Dong Liu*
496 Beyond Matching to Tiles: Bridging Unaligned Aerial and Satellite Views for Vision-Only UAV Navigation, *Kejia Liu, Haoyang Zhou, Ruoyu Xu, Peicheng Wang, Mingli Song, Haofei Zhang*
497 GeoFlow: Real-Time Fine-Grained Cross-View Geolocalization via Iterative Flow Prediction, *Ayesh Abu Lehyeh, Xiaohan Zhang, Ahmad Arrabi, Waqas Sultani, Chen Chen, Safwan Wshah*
498 ❊ PiLoT: Neural Pixel-to-3D Registration for UAV-based Ego and Target Geo-localization, *Xiaoya Cheng, Long Wang, Yan Liu, Xinyi Liu, Hanlin Tan, Yu Liu, Maojun Zhang, Shen Yan*
499 PAUL: Uncertainty-Guided Partition and Augmentation for Robust Cross-View Geo-Localization under Noisy Correspondence, *Zheng Li, Xueyi Zhang, Yanming Guo, Yuxiang Xie, Zhaoyun Ding, Siqi Cai, Haizhou Li, Mingrui Lao*
500 UniGeoRS: A Unified Benchmark for Tri-view Geo-Localization, *Xiao Liang, Huaizhi Tang, Feiyang Zhang, Shiji Yuan, Chun Hu, Dezhi Zheng, Kang Ma*
501 VGA: Empowering Aerial-Ground Localization by Visual Geometry Alignment, *Tao Jun Lin, Yujiao Shi, Hongdong Li*
502 Watch and Learn: Learning to Use Computers from Online Videos, *Chan Hee Song, Yiwen Song, Palash Goyal, Yu Su, Oriana Riva, Hamid Palangi, Tomas Pfister*
503 OneThinker: All-in-one Reasoning Model for Image and Video, *Kaituo Feng, Manyuan Zhang, Hongyu Li, Kaixuan Fan, Shuang Chen, Yilei Jiang, Dian Zheng, Peiwen Sun, Yiyuan Zhang, Haoze Sun, Yan Feng, Peng Pei, Xunliang Cai, Xiangyu Yue*
504 Incentivizing Versatile Video Reasoning in MLLMs via Data-Efficient Reinforcement Learning, *Xiaodong Wang, Zhirong Wu, Langling Huang, Yuxi Zheng, Peixi Peng*
505 Act2See: Emergent Active Visual Perception for Video Reasoning, *Martin Q. Ma, Yuxiao Qu, Aditya Agrawal, Willis Guo, Paul Pu Liang, Ruslan Salakhutdinov, Louis-Philippe Morency*
506 VideoSeek: Long-Horizon Video Agent with Tool-Guided Seeking, *Jingyang Lin, Jialian Wu, Jiang Liu, Ximeng Sun, Ze Wang, Xiaodong Yu, Jiebo Luo, Zicheng Liu, Emad Barsoum*
507 ViLoMem: Agentic Learner with Grow-and-Refine Multimodal Semantic Memory, *Weihao Bo, Shan Zhang, Yanpeng Sun, Jingjing Wu, Qunyi Xie, Xiao Tan, Kunbin Chen, Wei He, Xiaofan Li, Na Zhao, Jingdong Wang, Zechao Li*
508 ❊ ReMoT: Reinforcement Learning with Motion Contrast Triplets, *Cong Wan, Zeyu Guo, Jiangyang Li, Songlin Dong, Yifan Bai, Lin Peng, Zhiheng Ma, Yihong Gong*
509 Incentivizing Generative Zero-Shot Learning via Outcome-Reward Reinforcement Learning with Visual Cues, *Wenjin Hou, Xiaoxiao Sun, Hehe Fan*
510 Semantic-Guided Global-Local Collaborative Prompt Learning for Few-Shot Class Incremental Learning, *Yongxin Yan, Weisen Chen, Xingye Chen, Yuanjie Shao, Zhengrong Zuo, Wenming Tan, Wenqi Ren, Changxin Gao, Nong Sang*
511 Beyond Heuristic Prompting: A Concept-Guided Bayesian Framework for Zero-Shot Image Recognition, *Hui Liu, Kecheng Chen, Jialiang Wang, Xianming Liu, Wenya Wang, Haoliang Li*
512 One Patch to Caption Them All: A Unified Zero-Shot Captioning Framework, *Lorenzo Bianchi, Giacomo Pacini, Fabio Carrara, Nicola Messina, Giuseppe Amato, Fabrizio Falchi*
513 Data-Centric Meta-Learning for Robust Few-Shot Generalization, *Jongmin Lim, Soobin Cha, Jaehun Park, Inho Oh, Minho Park, Kwangsu Kim*
514 Bridging the Modality Gap in Compositional Zero-Shot Learning via Sparse Alignment and Unimodal Memory Bank, *Yang Zhang, Zhixiang Chi, Xudong Yan, Yang Wang, Songhe Feng*
515 LIFT and PLACE: A Simple, Stable, and Effective Knowledge Distillation Framework for Lightweight Diffusion Models, *Hyunsoo Han, Sangyeop Yeo, Jaejun Yoo*
516 WaDi: Weight Direction-aware Distillation for One-step Image Synthesis, *Lei Wang, Yang Cheng, Senmao Li, Ge Wu, Yaxing Wang, Jian Yang*
517 Uncertainty-Aware Knowledge Distillation for Multimodal Large Language Models, *Jingchen Sun, Shaobo Han, Deep Patel, Wataru Kohno, Can Jin, Changyou Chen*
518 Beyond Soft Label: Dataset Distillation via Orthogonal Gradient Matching, *Deyu Bo, Xinchao Wang*
519 BHCast: Unlocking Black Hole Plasma Dynamics from a Single Blurry Image with Long-Term Forecasting, *Renbo Tu, Ali SaraerToosi, Nicholas S. Conroy, Gennady Pekhimenko, Aviad Levis*
520 ❊ RawMetaDiff: Unlocking Extreme Darkness from Dual-Exposure RAW with Meta-Guided Diffusion, *Panjun Liu, Jiyuan Xia, Yuanshen Guan, Yong Li, Zhiqiang Lang, Ruikang Xu, Chang Chen, Dehua Song, Fenglong Song, Zhiwei Xiong*
521 Prospective Dynamic 3D MRI Reconstruction via Latent-Space Motion Tracking from Single Measurement, *Lixuan Chen, Zhongnan Liu, Jesse Hamilton, James M. Balter, Jeong Joon Park, Liyue Shen*
522 Lens Component Deletion based on Differentiable Ray Tracing, *Wenguan Zhang, Qirun Zhang, Tuo Sun, Jiajian He, Jiahui Xu, Huajun Feng, Qi Li*
523 X-band Radar Non-Line-of-Sight Imaging, *Dongyu Du, Mingkun Zhao, Yutong Yang, Dominik Scheuble, Xiaolong Huang, Zijian Shao, Mario Bijelic, Kaushik Sengupta, Felix Heide*

524 3M-TI: High-Quality Mobile Thermal Imaging via Calibration-free Multi-Camera Cross-Modal Diffusion, *Minchong Chen, Xiaoyun Yuan, Junzhe Wan, Jianing Zhang, Jun Zhang*
525 UAVLight: A Benchmark for Illumination-Robust 3D Reconstruction in Unmanned Aerial Vehicle (UAV) Scenes, *Kang Du, Xue Liao, Junpeng Xia, Chaozheng Guo, Yi Gu, Yirui Guan, Duotun Wang, Sheng Huang, Zeyu Wang*
526 Polarization State Tracing for Reflection Removal and Color-Consistent Reconstruction, *Dongyue Wang, Yang Lu, Jiandong Tian*
527 GFRRN: Explore the Gaps in Single Image Reflection Removal, *Yu Chen, Zewei He, Xingyu Liu, Zixuan Chen, Zhe-Ming Lu*
528 Efficient All-Pairs Correlation Volume Sampling for Optical Flow Estimation, *Karlis Martins Briedis, Markus Gross, Christopher Schroers*
529 Cross-Slice Knowledge Transfer via Masked Multi-Modal Heterogeneous Graph Contrastive Learning for Spatial Gene Expression Inference, *Zhiceng Shi, Changmiao Wang, Jun Wan, Wenwen Min*
530 Adapting a Pre-trained Single-Cell Foundation Model to Spatial Gene Expression Generation from Histology Images, *Donghai Fang, Yongheng Li, Zhen Wang, Yuansong Zeng, Wenwen Min*
531 HyperST: Hierarchical Hyperbolic Learning for Spatial Transcriptomics
❊ Prediction, *Chen Zhang, Yilu An, Ying Chen, Hao Li, Xitong Ling, Lihao Liu, Junjun He, Yuxiang Lin, Zihui Wang, Rongshan Yu*
532 SO(3)-Equivariant ViT-Adapter for Data-Efficient Zero-Shot Sim-to-Real Indoor Panoramic Depth Estimation, *Ziyan He, Qiudan Zhang, Lin Ma, Xu Wang*
533 Sparsity-Aware Voxel Attention and Foreground Modulation for 3D Semantic Scene Completion, *Yu Xue, Longjun Gao, Yuanqi Su, HaoAng Lu, Xiaoning Zhang*
534 XPaintNet: An eXtreme Lightweight Framework for Stereoscopic Conversion without Inpainting Network, *Kihwan Yoon, Juyeon Shin, Jungheum Kang, Sijung Kim, Minyong Jeon*
535 MD2E: Modeling Depth-to-Edge Cues for Monocular Metric Depth Estimation, *Chao Ning, Minghe Shen, Naoto Yokoya*
536 LiteSense: Lifting Lightweight ToF with RGB for High-Resolution
❊ Metric Depth Estimation, *Yusheng Li, Lizhi Lou, Yan Tang, Zekai Miao, Shaoming Zhang, Jianmei Wang*
537 3D-Aware Multi-Task Learning with Cross-View Correlations for Dense Scene Understanding, *Xiaoye Wang, Chen Tang, Xiangyu Yue, Wei-Hong Li*
538 The Midas Touch for Metric Depth, *Yu Ma, Zizhan Guo, Zuyi Xiong,*
❊ *Haoran Zhang, Yi Feng, Hongbo Zhao, Hanli Wang, Rui Fan*
539 Lifting Unlabeled Internet-level Data for 3D Scene Understanding, *Yixin Chen, Yaowei Zhang, Huangyue Yu, Junchao He, Yan Wang, Jiangyong Huang, Hongyu Shen, Junfeng Ni, Shaofei Wang, Baoxiong Jia, Song-Chun Zhu, Siyuan Huang*
540 ObjectMorpher: 3D-Aware Image Editing via Deformable 3DGS, *Yuhuan Xie, Aoxuan Pan, Yi-Hua Huang, Chirui Chang, Peng Dai, Xin Yu, Xiaojuan Qi*
541 PhysX-Anything: Simulation-Ready Physical 3D Assets from Single Image, *Ziang Cao, Fangzhou Hong, Zhaoxi Chen, Liang Pan, Ziwei Liu*
542 MeshFlow: Efficient Artistic Mesh Generation via MeshVAE and Flow-
❊ *based DiTs, Weiyu Li, Antoine Toisoul, Tom Monnier, Roman Shapovalov, Rakesh Ranjan, Ping Tan, Andrea Vedaldi*
543 WonderZoom: Multi-Scale 3D World Generation, *Jin Cao, Hong-Xing Yu, Jiajun Wu*
544 SceneTok: A Compressed, Diffusable Token Space for 3D Scenes, *Mohammad Asim, Christopher Wewer, Jan Eric Lenssen*
545 PixARMesh: Autoregressive Mesh-Native Single-View Scene Reconstruction, *Xiang Zhang, Sohyun Yoo, Hongrui Wu, Chuan Li, Jianwen Xie, Zhuowen Tu*
546 Extend3D: Town-Scale 3D Generation, *Seungwoo Yoon, Jinmo Kim, Jaesik Park*
547 Pano3DComposer: Feed-Forward Compositional 3D Scene Generation from Single Panoramic Image, *Zidian Qiu, Ancong Wu*
548 MeshWeaver: Sparse-Voxel-Guided Surface Weaving for Autoregressive Mesh Generation, *Jiale Xu, Wang Zhao, Ying Shan*
549 CaliTex: Geometry-Calibrated Attention for View-Coherent 3D Texture Generation, *Chenyu Liu, Hongze Chen, Jingzhi Bao, Lingting Zhu, Runze Zhang, Weikai Chen, Zeyu Hu, Yingda Yin, Keyang Luo, Xin Wang*
550 CraftMesh: High-Fidelity Generative Mesh Manipulation via Poisson Seamless Fusion, *James Jincheng Hu, Yuxiao Wu, Youcheng Cai, Ligang Liu*
551 LoG3D: Ultra-High-Resolution 3D Shape Modeling via Local-to-Global Partitioning, *Xinran Yang, Shuichang Lai, Jiangjing Lyu, Hongjie Li, Bowen Pan, Yuanqi Li, Jie Guo, Zhengkang Zhou, Yanwen Guo*
552 MaskFocus: Focusing Policy Optimization on Critical Steps for Masked Image Generation, *Guohui Zhang, Hu Yu, Xiaoxiao Ma, Yaning Pan, Hang Xu, Jie Huang, Feng Zhao*
553 Efficient Training for Human Video Generation with Entropy-Guided Prioritized Progressive Learning, *Changlin Li, Jiawei Zhang, Shuhao Liu, Sihao Lin, Zeyi Shi, Zhihui Li, Xiaojun Chang*
554 PosterOmni: Generalized Artistic Poster Creation via Task Distillation and Unified Reward Feedback, *Sixiang Chen, Jianyu Lai, Jialin Gao, Hengyu Shi, Zhongying Liu, Tian Ye, Junfeng Luo, Xiaoming Wei, Lei Zhu*
555 GRPO-Guard: Mitigating Implicit Over-Optimization in Flow Matching
❊ via Regulated Clipping, *Jing Wang, Jiajun Liang, Jie Liu, Henglin Liu, Gongye Liu, Jun Zheng, Wanyuan Pang, Ao Ma, Zhenyu Xie, Xintao Wang, Meng Wang, Pengfei Wan, Xiaodan Liang*
556 The Image as Its Own Reward: Reinforcement Learning with Adversarial Reward for Image Generation, *Weijia Mao, Hao Chen, Zhenheng Yang, Mike Zheng Shou*
557 Flash-DMD: Towards High-Fidelity Few-Step Image Generation with Efficient Distillation and Joint Reinforcement Learning, *Guanjie Chen, ShiruiHuang ShiruiHuang, Yifu Sun, Kai Liu, Jianchen Zhu, Xiaoye Qu, Yu Cheng, Peng Chen*
558 VISTA: A Test-Time Self-Improving Video Generation Agent, *Do Xuan Long, Xingchen Wan, Hootan Nakhost, Chen-Yu Lee, Tomas Pfister, Sercan Ö. Arik*
559 Neighbor GRPO: Contrastive ODE Policy Optimization Aligns Flow
❊ Models, *Dailan He, Guanlin Feng, Xingtong Ge, Yazhe Niu, Yi Zhang, Bingqi Ma, Guanglu Song, Yu Liu, Hongsheng Li*
560 SMV-EAR: Bring Spatiotemporal Multi-View Representation Learning into Efficient Event-Based Action Recognition, *Rui Fan, Weidong Hao, Juntao Guan, Lai Rui, Tong Wu, Fanhong Zeng, Lin Gu*
561 Hierarchical Action Learning for Weakly-Supervised Action Segmentation, *Junxian Huang, Ruichu Cai, Juntao Fang, Hao Zhu, Boyan Xu, Weilin Chen, Zijian Li, Shenghua Gao*
562 Gamba: Mamba-based graph convolutional network with dynamic graph topology learning for action recognition, *Rouyi Zhou, Yangzhi Wu, Jiajun Wen, Can Gao, Feng Liu, Zhihui Lai, Linlin Shen*
563 Beyond Binary Contrast: Modeling Continuous Skeleton Action Spaces with Transitional Anchors, *Yingjie Feng, Yi Wang, Jiaze Wang, Anfeng Liu, Zhuotao Tian*
564 PRISM: Learning a Shared Primitive Space for Transferable Skeleton Action Representation, *Di Yang, Yaohui Wang, Shuai Shao, François Brémond, Jiangtao Wang*
565 TWEO: Transformers Without Extreme Outliers Enables FP8 Training And Quantization For Dummies, *Guang Liang, Jie Shao, Ningyuan Tang, Xinyao Liu, Jianxin Wu*
566 Unified Spherical Frontend: Learning Rotation-Equivariant Representations of Spherical Images from Any Camera, *Mukai Yu, Mosam Dabhi, Liuyue Xie, Sebastian Scherer, László A. Jeni*
567 The Surprising Effectiveness of Noise Pretraining for Implicit Neural Representations, *Kushal Vyas, Alper Kayabasi, Daniel Kim, Vishwanath Saragadam, Ashok Veeraraghavan, Guha Balakrishnan*
568 DABO: Difficulty-Aware Bayesian Optimization with Diffusion-Learned
❊ Priors, *Mengyang Li, Pinlong Zhao*
569 Towards Knowledge-augmented Bayesian Deep Learning For Computer
❊ Vision, *Wang Ma, Hanjing Wang, Yufei Zhang, Darsha Udayanga, Qiang Ji*
570 NESTOR: A Nested MOE-based Neural Operator for Large-Scale PDE Pre-Training, *Dengdi Sun, Xiaoya Zhou, Xiao Wang, Hao Si, Wanli Lyu, Jin Tang, Bin Luo*
571 Evidential Transformation Network: Turning Pretrained Models into
❊ Evidential Models for Post-hoc Uncertainty Estimation, *Yongchan Chun, Chanhee Park, Jeongho Yoon, Jaehyung Seo, Heuiseok Lim*
572 Beyond Euclidean Gossip: KL-Barycentric Consensus on Heterogeneous and Imbalanced Images, *Lu Xu, Guosheng Yin*
573 Prime Once, then Reprogram Locally: An Efficient Alternative to
❊ Black-Box Service Model Adaptation, *Yunbei Zhang, Chengyi Cai, Feng Liu, Jihun Hamm*
574 Batch Loss Score for Dynamic Data Pruning, *Qing Zhou,*
❊ *Bingxuan Zhao, Tao Yang, Hongyuan Zhang, Junyu Gao, Qi Wang*

575 Teacher-Guided Routing for Sparse Vision Mixture-of-Experts, *Masahiro Kada, Ryota Yoshihashi, Satoshi Ikehata, Rei Kawakami, Ikuro Sato*
576 WebChain: A Large-Scale Human-Annotated Dataset of Real-World Web Interaction Traces, *Sicheng Fan, Rui Wan, Yifei Leng, Gaoning Liang, Li Ling, Yanyi Shang, Dehan Kong*
577 MangoBench: A Benchmark for Multi-Agent Goal-Conditioned Offline Reinforcement Learning, *Yi Wang, Ningze Zhong, Zhiheng Fu, Longguang Wang, Ye Zhang, Yulan Guo*
578 iSHIFT: Lightweight Slow-Fast GUI Agent with Adaptive Perception, *Sarthak Mehrotra, Sairam VC Rebbapragada, Mani Bonthu, Vineeth N. Balasubramanian*
579 MMBench-GUI: A Unified Hierarchical Evaluation Framework for Multi-Platform GUI Agents, *Xuehui Wang, Zhenyu Wu, JingJing Xie, Zichen Ding, Bowen Yang, Zehao Li, Zhaoyang Liu, Qingyun Li, Xuan Dong, Zhe Chen, Weiyun Wang, Xiangyu Zhao, Jixuan Chen, Haodong Duan, Tianbao Xie, Chenyu Yang, Shiqian Su, Yue Yu, Yanting Zhang, Xiangyu Yue, Weijie Su, Xizhou Zhu, Wei Shen, Jifeng Dai, Wenhai Wang*
580 ❊ Boosting Vision-Language Models Towards Cross-Domain Incremental Object Detection, *Xu Wang, Zihan Lin, Yixin Zhang, Zilei Wang*
581 UniSpector: Towards Universal Open-set Defect Recognition via Spectral-Contrastive Visual Prompting, *Geonuk Kim, Minhoi Kim, Kangil Lee, Minsu Kim, Hyeonseong Jeon, Jeonghoon Han, Hyoungjoon Lim, Junho Yim*
582 Unlearning without Forgetting: Securely Removing Targeted Concepts from Large-Scale Vision-Language Open-Vocabulary Detectors, *Zhongze Wu, Xiu Su, Feng Yang, Dan Niu, Shan You, Yueyi Luo, Jun Long*
583 UNI-OOD: Unified Object- and Image-level Out-of-Distribution Detection via Cross-Context Attentive Vision-Language Modeling, *Yuchuan Li, Azadeh Motamedi, Hyock Ju Kwon, Chul B Park, Il-Min Kim*
584 S2C2Seg: Semantic-Spatial Consistency and Category Optimization for Open-Vocabulary Segmentation, *Yuhao Qing, Yueying Wang, Chaoyang Chen, Weidong Zhang, Jie Wen, Xin Xu*
585 NoOVD: Novel Category Discovery and Embedding for Open-Vocabulary Object Detection, *Yupeng Zhang, Ruize Han, Zhiwei Chen, Wei Feng, Liang Wan*
586 The Missing Point in Vision Transformers for Universal Image Segmentation, *Sajjad Shahabodini, Mobina Mansoori, Farnoush Bayatmakou, Jamshid Abouei, Konstantinos Plataniotis, Arash Mohammadi*
587 ❊ PromptMoE: A Segmentation Refinement Framework Leveraging Mixture of Experts for Improved Prompting, *Stephen Price, Danielle L. Cote, Elke A. Rundensteiner*
588 The Power of Prior: Training-Free Open-Vocabulary Semantic Segmentation with LLaVA, *Bingfeng Zhang, Siyue Yu, Hui Li, Jiahua Lin, Wenwu Wang, Jimin Xiao*
589 Beyond Text: Visual Description Assembly by Probabilistic Model for CLIP-based Weakly Supervised Semantic Segmentation, *Xianglin Qiu, Jian Wang, Xiaolei Wang, Zhen Zhang, Jimin Xiao*
590 High-Precision Dichotomous Image Segmentation via Depth Integrity-Prior and Fine-Grained Patch Strategy, *Xianjie Liu, Keren Fu, Qijun Zhao*
591 GeoSAM2: Unleashing the Power of SAM2 for 3D Part Segmentation, *Ken Deng, Yunhan Yang, Jingxiang Sun, Xihui Liu, Yebin Liu, Ding Liang, Yan-Pei Cao*
592 Material Magic Wand: Material-Aware Grouping of 3D Parts in Untextured Meshes, *Umangi Jain, Vladimir Kim, Matheus Gadelha, Igor Gilitschenski, Zhiqin Chen*
593 Synthetic Object Compositions for Scalable and Accurate Learning in Detection, Segmentation, and Grounding, *Weikai Huang, Jieyu Zhang, Taoyang Jia, Chenhao Zheng, Ziqi Gao, Jae Sung Park, Ranjay Krishna*
594 Unlocking 3D Affordance Segmentation with 2D Semantic Knowledge, *Yu Huang, Zelin Peng, Changsong Wen, Xiaokang Yang, Wei Shen*
595 HySeg: Learning Generative Priors for Structure-Aware Remote Sensing Segmentation, *Jie Qiu, Xin Li, Fan Yang, Yan Wang, Dong Yu, Changying Wang, Linwei Dai, Yongxiang Chen, Youqin Chen, Jianzhang Chen*
596 Real-Time Long Horizon Air Quality Forecasting via Group-Relative Policy Optimization, *Inha Kang, Eunki Kim, Wonjeong Ryu, Jaeyo Shin, Seungjun Yu, Yoon-Hee Kang, Seongeun Jeong, Eunhye Kim, Soontae Kim, Hyunjung Shim*
597 MMVIP: A Visible-infrared Paired Dataset for Multi-weather Marine Vision, *Yunpeng Yin, Lihan Wang, Zhaoshen He, Xinqiang He, Xingming Liao, Zhuowei Wang, Lianglun Cheng*
598 Beyond Tie Points: Satellite Image Block Adjustment based on Dense Feature Consistency, *Yi Liu, Yi Wan, Lei Yu, Panwang Xia, Qiong Wu, Yingying Pei, Xuejun Huang, Junjian Zhang, Xiangyuan Cai, Hongwei Hu, Yongjun Zhang*
599 Spectrally Distilled Representations Aligned with Instruction-Augmented LLMs for Satellite Imagery, *Minh Kha Do, Wei Xiang, Kang Han, Di Wu, Khoa Phan, Yi-Ping Phoebe Chen, Gaowen Liu, Ramana Rao Kompella*
600 Global Underwater Geolocation from Time-Lapse Polarization Imagery, *Sara Aghajanzadeh, Xiaoyang Bai, Zhongmin Zhu, David Forsyth, Viktor Gruev*
601 Olbedo: An Albedo and Shading Aerial Dataset for Large-Scale Outdoor Environments, *Shuang Song, Debao Huang, Deyan Deng, Haolin Xiong, Yang Tang, Yajie Zhao, Rongjun Qin*
602 PRUE: A Practical Recipe for Field Boundary Segmentation at Scale, *Gedeon Muhawenayo, Caleb Robinson, Subash Khanal, Zhanpei Fang, Isaac Corley, Alexander Wollam, Tianyi Gao, Leonard Strnad, Ryan Avery, Lyndon Estes, Ana Tárano, Nathan Jacobs, Hannah Kerner*
603 SARMAE: Masked Autoencoder for SAR Representation Learning, *Danxu Liu, Di Wang, Hebaixu Wang, Haoyang Chen, Wentao Jiang, Yilin Cheng, Haonan Guo, Wei Cui, Jing Zhang*
604 LNEM: Lunar Neural Elevation Model, *Suwan Lee, Jo Ryeong Yim, Kibaek Park, Dong-Gyu Kim, Eunhyeuk Kim, Minsup Jeong, Chae Kyung Sim, Seokju Lee*
605 A Polarized Reflection and Material Dataset of Real World Objects, *Jing Yang, Krithika Dharanikota, Emily Jia, Haiwei Chen, Yajie Zhao*
606 LaSM: Layer-wise Scaling Mechanism for Defending Pop-up Attack on GUI Agents, *Zihe Yan, Zhuosheng Zhang, Jiaping Gui, Gongshen Liu*
607 RaPA: Enhancing Transferable Targeted Attacks via Random Parameter Pruning, *Tongrui Su, Qingbin Li, Shengyu Zhu, Wei Chen, Xueqi Cheng*
608 All Vehicles Can Lie: Efficient Adversarial Defense in Fully Untrusted-Vehicle Collaborative Perception via Pseudo-Random Bayesian Inference, *Yi Yu, Libing Wu, Zhuangzhuang Zhang, Jing Qiu, Lijuan Huo, Jiaqi Feng*
609 A Combination of Noise and Bilateral Filters Achieve Supralinear and Scalable Adversarial Robustness in CNNs, *Nicolas Stalder, Benjamin F. Grewe, Matteo Saponati, Pau Vilimelis Aceituno*
610 DeepProtect: Proactive Face-Swapping Defense using Identity Blending and Attribute Distortion, *Eungi Lee, Seung-hyeok Back, Hyung-Il Kim, Seok Bong Yoo*
611 Write Where It Matters: Policy-Guided Watermarks for 3D Gaussian Splatting, *Nan Li, Yike Zeng, Qian Zhang, Qi Zhang, Zhiyi Pan, Wei Feng, Liang Wan*
612 Attack for Defense: Adversarial Agents for Point Prompt Optimization Empowering Segment Anything Model, *Xueyu Liu, Xiaoyi Zhang, Meilin Liu, Guangze Shi, Jia Shen, Yujie Wang, Cai Zhao, Ziyuan He, Yongfei Wu, Mingqiang Wei, Yongle Chen*
613 RevINN: An End-to-End Invertible Neural Network for Reversible Adversarial Examples Generation, *Jielun Huang, Chi-Man Pun, Guoheng Huang*
614 CamPI: Physical Adversarial Examples through Camera Power Signal Injection, *Yanze Ren, Mingyuan Lv, Qinhong Jiang, Yan Jiang, Chen Yan, Xiaoyu Ji, Wenyuan Xu*
615 ❊ Authorize-on-Demand: Dynamic Authorization with Legality-Aware Intellectual Property Protection for VLMs, *Lianyu Wang, Meng Wang, Huazhu Fu, Daoqiang Zhang*
616 ❊ GraspALL: Adaptive Structural Compensation from Illumination Variation for Robotic Garment Grasping in Any Low-Light Conditions, *Haifeng Zhong, Wenshuo Han, Zhouyu Wang, Runyang Feng, Fan Tang, Tong-Yee Lee, Zipei Fan, Ruihai Wu, Yuran Wang, Hao Dong, Hechang Chen, Hyung Jin Chang, Yixing Gao*
617 Opening the Sim-to-Real Door for Humanoid Pixel-to-Action Policy Transfer, *Haoru Xue, Tairan He, Zi Wang, Qingwei Ben, Wenli Xiao, Zhengyi Luo, Xingye Da, Fernando Castañeda, Guanya Shi, Shankar Sastry, Linxi Fan, Yuke Zhu*

618 Learning Cross-View Object Correspondence via Cycle-Consistent Mask Prediction, *Shannan Yan, Leqi Zheng, Keyu Lv, Jingchen Ni, Hongyang Wei, Jiajun Zhang, Guangting Wang, Jing LYU, Chun Yuan, Fengyun Rao*
619 RoboWheel: A Data Engine from Real-World Human Demonstrations for Cross-Embodiment Robotic Learning, *Yuhong Zhang, Zihan Gao, Shengpeng Li, Ling-Hao Chen, Kaisheng Liu, Runqing Cheng, Xiao Lin, Junjia Liu, Zhuoheng Li, Jingyi Feng, Ziyan He, Jintian Lin, Zheyan Huang, Zhifang Liu, Haoqian Wang*
620 Chain of World: World Model Thinking in Latent Motion, *Fuxiang Yang, Donglin Di, Lulu Tang, Xuancheng Zhang, Lei Fan, Hao Li, Wei Chen, Tonghua Su, Baorui Ma*
621 ❊ Scalable Feature Matching via State Space Modeling and Sparse Correlation, *Sin Wai Choo, Bo Li*
622 Video2Robo: 3DGS-based Synthetic Data from One Video Enables Scalable Robot Learning, *Yinan Deng, Kejia Hu, Ye Chen, Jianyu Dou, Jiahui Wang, Jingyu Zhao, Haojia Ao, Yi Yang, Yufeng Yue*
623 ConsisVLA-4D: Advancing Spatiotemporal Consistency in Efficient 3D-Perception and 4D-Reasoning for Robotic Manipulation, *Wei Li, Jizhihui Liu, Li Yixing, Junwen Tong, Rui Shao, Liqiang Nie*
624 SRPO: Self-Referential Policy Optimization for Vision-Language-Action Models, *Senyu Fei, Siyin Wang, Li Ji, Ao Li, Shiduo Zhang, Liming Liu, Jinlong Hou, Jingjing Gong, Xianzhong Zhao, Xipeng Qiu*
625 GeoDexGrasp: Geometry-aware Generation for Data-efficient and Physics-plausible Dexterous Grasping, *Bing Han, Weiyuan Liu, Changlong Zhang, Chenxi Wang, Zhibin Zhao, Zhi Zhai*
626 Lifelong Imitation Learning with Multimodal Latent Replay and Incremental Adjustment, *Fanqi Yu, Matteo Tiezzi, Tommaso Apicella, Cigdem Beyan, Vittorio Murino*
627 From Observation to Action: Latent Action-based Primitive Segmentation for VLA Pre-training in Industrial Settings, *Jiajie Zhang, Sören Schwertfeger, Alexander Kleiner*
628 AGiLe: Learning Robust Long-Horizon Manipulation via Affordance-Grounded Bidirectional Latent Planning, *Zixuan Chen, Xiangrong Feng, Jieqi Shi, Lin Shao, Jing Huo, Yang Gao*
629 Language-Grounded Decoupled Action Representation for Robotic Manipulation, *Wuding Weng, Tongshu Wu, Liucheng Chen, Siyu Xie, Zheng Wang, Xing Xu, Jingkuan Song, Heng Tao Shen*
630 Learning to Act Robustly with View-Invariant Latent Actions, *Youngjoon Jeong, Junha Chun, Taesup Kim*
631 ORBIT: Benchmarking SfM in the Wild with 360° Video, *Sara Sabour, Richard Tucker, Marcus Brubaker, Saurabh Saxena, Junhwa Hur, Andrea Tagliasacchi, Deqing Sun, David J. Fleet, Richard Szeliski, Noah Snavely*
632 SpikeTrack: A Spike-driven Framework for Efficient Visual Tracking, *Qiuyang Zhang, Jiujun Cheng, Qichao Mao, Cong Liu, Yu Fang, Yuhong Li, Mengying Ge, Shangce Gao*
633 Time Without Time: Pseudo-Temporal Representation for Space-Time Super-Resolution, *Hee Min Choi, Hyoa Kang, Suji Kim, Dokwan Oh, Nam Ik Cho*
634 Envisioning the Future, One Step at a Time, *Stefan Andreas Baumann, Jannik Wiese, Tommaso Martorella, Mahdi M. Kalayeh, Björn Ommer*
635 FlowFM: Advancing Dark Optical Flow Estimation with Flow Matching, *Fengyuan Zuo, Haiyan Jin, Yuanlin Zhang, Zhaolin Xiao, Bin Wang, Yuerong Mu*
636 Drift-Resilient Temporal Priors for Visual Tracking, *Yuqing Huang, Liting Lin, Weijun Zhuang, Zhenyu He, Xin Li*
637 ❊ An Efficient Token Compression Framework for Visual Object Tracking, *Weijing Wu, Qihua Liang, Bineng Zhong, Haiying Xia, Zhiyi Mo, Shuxiang Song*
638 No Labels, No Look-Ahead: Unsupervised Online Video Stabilization with Classical Priors, *Tao Liu, Kan Ren, Gang Wan, Shibo Wen*
639 From Detection to Association: Learning Discriminative Object Embeddings for Multi-Object Tracking, *Yuqing Shao, Yuchen Yang, Rui Yu, Weilong Li, Xu Guo, Huaicheng Yan, Wei Wang, Xiao Sun*
640 Momentum Memory for Knowledge Distillation in Computational Pathology, *Yongxin Guo, Hao Lu, Onur C. Koyun, Zhengjie Zhu, Muhammet F. Demir, Metin N. Gurcan*
641 Modeling the Brain's Grammar: ROI-Guided fMRI Pretraining for Transferable and Interpretable Vision Decoding, *Yulong Liu, Hua Xu, Yiyang Cai, Chunyang Jiang, Sirui Han, Yike Guo*
642 Joint Spectral Image Reconstruction and Semantic Segmentation with Cooperative Unfolding, *Zijun He, Ping Wang, Xiaodong Wang, Chang Chen, Xin Yuan*
643 X-WIN: Building Chest Radiograph World Model via Predictive Sensing, *Zefan Yang, Ge Wang, James Hendler, Mannudeep K. Kalra, Pingkun Yan*
644 fMRI-LM: Towards a Universal Foundation Model for Language-Aligned fMRI Understanding, *Yuxiang Wei, Yanteng Zhang, Xi Xiao, Chengxuan Qian, Tianyang Wang, Vince D. Calhoun*
645 Tell2Adapt: A Unified Framework for Source Free Unsupervised Domain Adaptation via Vision Foundation Model, *Yulong Shi, Shijie Li, Ziyi Li, Lin Qi*
646 TIM: Temporal Decoupling with Iterative Mutual-Refinement Model for Longitudinal Radiology Report Generation, *Yiheng Dong, Yi Lin, Shilong Huang, Xiyan Yang, Xin Yang*
647 Ultrasound-CLIP: Semantic-Aware Contrastive Pre-training for Ultrasound Image-Text Understanding, *Jiayun Jin, Haolong Chai, Xueying Huang, Xiaoqing Guo, Zengwei Zheng, Zhan Zhou, Junmei Wang, Xinyu Wang, Jie Liu, Binbin Zhou*
648 Act Like a Pathologist: Tissue-Aware Whole Slide Image Reasoning, *Wentao Huang, Weimin Lyu, Peiliang Lou, Qingqiao Hu, Xiaoling Hu, Shahira Abousamra, Wenchao Han, Ruifeng Guo, Jiawei Zhou, Chao Chen, Chen Wang*
649 BiGMINT: Biologically-guided Hierarchical Multimodal Integration for Modeling Multiple Compound Activities in Drug Discovery, *Pushpak Pati, Bo Li, Abbas Rayabat Khan, Tomé Albuquerque, Steffen Jaensch, Amina Mollaysa, Walid M. Abdelmoula, Samantha J. Allen, Joke Reumers, Helai P. Mohammad, Scott Oloff, Tommaso Mansi, Rui Liao, Dmytro S. Lituiev, Zhoubing Xu*
650 Modeling Spatiotemporal Neural Frames for High Resolution Brain Dynamic, *Wanying Qu, Jianxiong Gao, Wei Wang, Yanwei Fu*
651 ❊ CMR-RD: Long-Tailed Adaptive VLM for Explainable CMR Diagnosis, *Yansong Li, Zhongxi Qiu, Yun Tian, Zheng Jinyu, Shuo Li*
652 Clinically-Grounded Counterfactual Reasoning for Medical Video Diagnosis, *Jianzhe Gao, Churan Wang, Weiyi Zhang, Jianghua Li, Li-An Li, Wenguan Wang, Yixin Zhu, Yizhou Wang*
653 FBTA: Enabling Single-GPU End-to-End Gigapixel WSI Classification with Feature Bridging and Translation Alignment, *Jiuyang Dong, Jiahan Li, Junjun Jiang, Yongbing Zhang*
654 Ultra Diffusion Poser: Diffusion-Based Human Motion Tracking from Sparse Inertial Sensors and Ranging-based Between-sensor Distances, *Dominik Hollidt, Tommaso Bendinelli, Christian Holz*
655 Egocentric Visibility-Aware Human Pose Estimation, *Peng Dai, Yu Zhang, Feng Yiqiang, Zhen Fan, Yang Zhang*
656 Shoe Style-Invariant and Ground-Aware Learning for Dense Foot Contact Estimation, *Daniel Sungho Jung, Kyoung Mu Lee*
657 ❊ OMG-Bench: A New Challenging Benchmark for Skeleton-based Online Micro Hand Gesture Recognition, *Haochen Chang, Pengfei Ren, Buyuan Zhang, Da Li, Tianhao Han, Haoyang Zhang, Liang Xie, Hongbo Chen, Erwei Yin*
658 ❊ Recovering Physically Plausible Human-Object Interactions from Monocular Videos, *Dingbang Huang, Etienne Vouga, Qixing Huang, Georgios Pavlakos*
659 MoCapAnything: Unified 3D Motion Capture for Arbitrary Skeletons from Monocular Videos, *Kehong Gong, Zhengyu Wen, Weixia He, Mingxi Xu, Qi Wang, Ning Zhang, Zhengyu Li, Dongze Lian, Wei Zhao, Xiaoyu He, Mingyuan Zhang*
660 ❊ TeHOR: Text-Guided 3D Human and Object Reconstruction with Textures, *Hyeongjin Nam, Daniel Sungho Jung, Kyoung Mu Lee*
661 SHOW3D: Capturing Scenes of 3D Hands and Objects in the Wild, *Patrick Rim, Kevin Harris, Braden Copple, Shangchen Han, Xu Xie, Ivan Shugurov, Sizhe An, He Wen, Alex Wong, Tomas Hodan, Kun He*
662 CrossHOI: Learning Cross-View Representations for Monocular 3D Human-Object Interaction Reconstruction, *Pei Geng, Shanshan Zhang, Jian Yang*
663 Gaussian-Mixture Latent Flow for Stochastic 3D Human Motion Prediction, *Yue Ma, Frederick W. B. Li, Xiaohui Liang*
664 SGSoft: Learning Fused Semantic-Geometric Features for 3D Shape Correspondence via Template-Guided Soft Signals, *Soyeon Yoon, Chang Wook Seo, Hyunjung Shim*
665 Beyond Single-View Sufficiency: CVBench for Cross-View Human Understanding, *Tianchen Guo, Chen Liu, Xin Yu*

666 Breaking Spurious Correlations: Uncertainty-Driven Causal Transformers for AU Detection, *Yuru Wang, Yue Zhou*

10:45 - 12:45 DEMOS (ExHall F)

1 Agentic Gaze Analysis with AR Interaction, *Houze Yang, Xu Cao, Chen Fang, Inki Kim, James M. Rehg*
2 PISCO: Precise Video Instance Insertion with Sparse Control, *Xiangbo Gao, Renjie Li, Xinghao Chen, Yuheng Wu, Suofei Feng, Qing Yin, Zhengzhong Tu*
3 LangFlash Demo, *Yilong Liu, Wanhua Li, Chen Zhu-Tian, Hanspeter Pfister*
4 Pixels-to-Layers: Turning Generated Images into Editable Assets, *Abhay Bhandarkar, Vineeth N Balasubramanian*
5 Computational Speckle Pattern Interferometry, *Shengxi Wu, Sophia Yang, Dorian Chan, Matthew O'Toole*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

13:00 - 14:15 Oral Session 2A: 3D Reconstruction (Bluebird Ballroom)
🏆 - Award candidate paper

1 🏆 MAMMA: Markerless Accurate Multi-person Motion Acquisition, *Hanz Cuevas Velasquez, Anastasios Yiannakidis, Soyong Shin, Giorgio Becherini, Markus Höschle, Joachim Tesch, Taylor Obersat, Tsvetelina Alexiadis, Eni Halilaj, Michael J. Black*
2 🏆 Natural Human Motion Recovery by Aligning High-Order Temporal Dynamics from Monocular Videos, *Dingkun Wei, Zehong Shen, Yan Xia, Georgios Pavlakos, Yujun Shen, Xiaowei Zhou*
3 PoseGAM: Robust Unseen Object Pose Estimation via Geometry-Aware Multi-View Reasoning, *Jianqi Chen, Biao Zhang, Xiangjun Tang, Peter Wonka*
4 🏆 SAM 3D Body: Robust Full-Body Human Mesh Recovery, *Xitong Yang, Devansh Kukreja, Don Pinkus, Taosha Fan, Jinhyung Park, Soyong Shin, Jinkun Cao, Jia-Wei Liu, Nicolás Ugrinovic, Anushka Sagar, Jitendra Malik, Matt Feiszli, Piotr Dollár, Kris Kitani*
5 🏆 SAM 3D: 3Dfy Anything in Images, *Xingyu Chen, FU-JEN CHU, Pierre Gleize, Kevin J Liang, Alexander Sax, Hao Tang, Weiyao Wang, Michelle Guo, Thibaut Hardin, Xiang Li, Aohan Lin, Jia-Wei Liu, Ziqi Ma, Anushka Sagar, Bowen Song, Xiaodong Wang, Jianing Yang, Bowen Zhang, Piotr Dollár, Georgia Gkioxari, Matt Feiszli, Jitendra Malik*
6 SPARK: Sim-ready Part-level Articulated Reconstruction with VLM Knowledge, *Yumeng He, Ying Jiang, Jiayin Lu, Yin Yang, Chenfanfu Jiang*

13:00 - 14:15 Oral Session 2B: Materials & Lighting (Four Seasons Ballroom)

1 🏆 3DReflecNet: A Large-Scale Dataset for 3D Reconstruction of Reflective, Transparent, and Low-Texture Objects, *Zhicheng Liang, Haoyi Yu, Boyan Li, Dayou Zhang, Zijian Cao, Tianyi Gong, Junhua Liu, Shuguang Cui, Fangxin Wang*
2 🏆 GLINT: Modeling Scene-Scale Transparency via Gaussian Radiance Transport, *Youngju Na, Jaeseong Yun, Soohyun Ryu, Hyunsu Kim, Sung-Eui Yoon, Suyong Yeon*
3 Neural Field-Based 3D Surface Reconstruction of Microstructures from Multi-Detector Signals in Scanning Electron Microscopy, *Shuo Chen, Yijin Li, Xi Zheng, Guofeng Zhang*
4 🏆 PhyGaP: Physically-Grounded Gaussians with Polarization Cues, *Jiale Wu, Xiaoyang Bai, Zongqi He, Weiwei Xu, Yifan Peng*
5 PPISP: Physically-Plausible Compensation and Control of Photometric Variations in Radiance Field Reconstruction, *Isaac Deutsch, Nicolas Moënne-Loccoz, Gavriel State, Zan Gojcic*
6 🏆 SeeGroup: Multi-Layer Depth Estimation of Transparent Surfaces via Self-Determined Grouping, *Hongyu Wen, Jia Deng*

13:00 - 14:15 Oral Session 2C: Gaussian Splatting & Reconstruction, (Mile High Ballroom 1A - 2A)

1 Energy-GS: Image Energy-guided Pose Alignment Gaussian Splatting with redesigned pose gradient flow, *Yu Gao, Lutong Su, Ruixiang Huang, Tianji Jiang, Jiadong Tang, Yufeng Yue, Yi Yang*
2 MeshSplatting: Differentiable Rendering with Opaque Meshes, *Jan Held, Sanghyun Son, Renaud Vandeghen, Daniel Rebain, Matheus Gadelha, Yi Zhou, Anthony Cioppa, Ming C. Lin, Marc Van Droogenbroeck, Andrea Tagliasacchi*
3 🏆 Proxy-GS: Unified Occlusion Priors for Training and Inference in Structured 3D Gaussian Splatting, *Yuanyuan Gao, Yuning Gong, Yifei Liu, Jingfeng Li, Dan Xu, Yanci Zhang, Dingwen Zhang, Xiao Sun, Zhihang Zhong*
4 RetimeGS: Continuous-Time Reconstruction of 4D Gaussian Splatting, *Xuezhen Wang, Li Ma, Yulin Shen, Zeyu Wang, Pedro V. Sander*
5 Selfi: Self-improving Reconstruction Engine via 3D Geometric Feature Alignment, *Youming Deng, Songyou Peng, Junyi Zhang, Kathryn Heal, Tiancheng Sun, John Flynn, Steve Marschner, Lucy Chai*
6 Z-Order Transformer for Feed-Forward Gaussian Splatting, *Can Wang, Lei Liu, Wei Jiang, Dong Xu*

13:00 - 14:15 Oral Session 2D: Spatio-Temporal Reconstruction, (Mile High Ballroom 3A - 4A)

1 4D Primitive-Mâché: Glueing Primitives for Persistent 4D Scene Reconstruction, *Kirill Mazur, Marwan Taher, Andrew J. Davison*
2 🏆 Efficiently Reconstructing Dynamic Scenes One D4RT at a Time, *Chuhan Zhang, Guillaume Le Moing, Skanda Koppula, Ignacio Rocco, Liliane Momeni, Junyu Xie, Shuyang Sun, Rahul Sukthankar, Joëlle K. Barral, Raia Hadsell, Zoubin Ghahramani, Andrew Zisserman, Junlin Zhang, Mehdi S. M. Sajjadi*
3 🏆 FUSER: Feed-Forward Multiview 3D Registration Transformer and SE(3)^N Diffusion Refinement, *Haobo Jiang, Jin Xie, Jian Yang, Liang Yu, Jianmin Zheng*
4 🏆 Residual Primitive Fitting of 3D Shapes with SuperFrusta, *Aditya Ganeshan, Matheus Gadelha, Thibault Groueix, Zhiqin Chen, Siddhartha Chaudhuri, Vladimir Kim, Wang Yifan, Daniel Ritchie*
5 🏆 SmokeSVD: Smoke Reconstruction from A Single View via Progressive Novel View Synthesis and Refinement with Diffusion Models, *Chen Li, Shanshan Dong, Sheng Qiu, Jianmin Han, Yibo Zhao, Zan Gao, Taku Komura, Kemeng Huang*
6 🏆 SparseWorld-TC: Trajectory-Conditioned Sparse Occupancy World Model, *Jiayuan Du, Yiming Zhao, Zhenglong Guo, Yong Pan, Wenbo Hou, Zhihui Hao, Kun Zhan, Qijun Chen*

16:00 - 18:00 Poster Session 2 & Exhibit Hall w/ Coffee Break (Exhall A & F)
❊ - Highlight paper 🏆 - Award candidate paper

1 🏆 MAMMA: Markerless Accurate Multi-person Motion Acquisition, *Hanz Cuevas Velasquez, Anastasios Yiannakidis, Soyong Shin, Giorgio Becherini, Markus Höschle, Joachim Tesch, Taylor Obersat, Tsvetelina Alexiadis, Eni Halilaj, Michael J. Black*
2 🏆 Natural Human Motion Recovery by Aligning High-Order Temporal Dynamics from Monocular Videos, *Dingkun Wei, Zehong Shen, Yan Xia, Georgios Pavlakos, Yujun Shen, Xiaowei Zhou*
3 PoseGAM: Robust Unseen Object Pose Estimation via Geometry-Aware Multi-View Reasoning, *Jianqi Chen, Biao Zhang, Xiangjun Tang, Peter Wonka*
4 🏆 SAM 3D Body: Robust Full-Body Human Mesh Recovery, *Xitong Yang, Devansh Kukreja, Don Pinkus, Taosha Fan, Jinhyung Park, Soyong Shin, Jinkun Cao, Jia-Wei Liu, Nicolás Ugrinovic, Anushka Sagar, Jitendra Malik, Matt Feiszli, Piotr Dollár, Kris Kitani*
5 🏆 SAM 3D: 3Dfy Anything in Images, *Xingyu Chen, Fu-Jen Chu, Pierre Gleize, Kevin J Liang, Alexander Sax, Hao Tang, Weiyao Wang, Michelle Guo, Thibaut Hardin, Xiang Li, Aohan Lin, Jia-Wei Liu, Ziqi Ma, Anushka Sagar, Bowen Song, Xiaodong Wang, Jianing Yang, Bowen Zhang, Piotr Dollár, Georgia Gkioxari, Matt Feiszli, Jitendra Malik*
6 SPARK: Sim-ready Part-level Articulated Reconstruction with VLM Knowledge, *Yumeng He, Ying Jiang, Jiayin Lu, Yin Yang, Chenfanfu Jiang*
7 🏆 3DReflecNet: A Large-Scale Dataset for 3D Reconstruction of Reflective, Transparent, and Low-Texture Objects, *Zhicheng Liang, Haoyi Yu, Boyan Li, Dayou Zhang, Zijian Cao, Tianyi Gong, Junhua Liu, Shuguang Cui, Fangxin Wang*
8 🏆 GLINT: Modeling Scene-Scale Transparency via Gaussian Radiance Transport, *Youngju Na, Jaeseong Yun, Soohyun Ryu, Hyunsu Kim, Sung-Eui Yoon, Suyong Yeon*
9 Neural Field-Based 3D Surface Reconstruction of Microstructures from Multi-Detector Signals in Scanning Electron Microscopy, *Shuo Chen, Yijin Li, Xi Zheng, Guofeng Zhang*

10 PhyGaP: Physically-Grounded Gaussians with Polarization Cues, 🏆 *Jiale Wu, Xiaoyang Bai, Zongqi He, Weiwei Xu, Yifan Peng*
11 PPISP: Physically-Plausible Compensation and Control of Photometric Variations in Radiance Field Reconstruction, *Isaac Deutsch, Nicolas Moënne-Loccoz, Gavriel State, Zan Gojcic*
12 SeeGroup: Multi-Layer Depth Estimation of Transparent Surfaces via 🏆 Self-Determined Grouping, *Hongyu Wen, Jia Deng*
13 Energy-GS: Image Energy-guided Pose Alignment Gaussian Splatting with redesigned pose gradient flow, *Yu Gao, Lutong Su, Ruixiang Huang, Tianji Jiang, Jiadong Tang, Yufeng Yue, Yi Yang*
14 MeshSplatting: Differentiable Rendering with Opaque Meshes, *Jan Held, Sanghyun Son, Renaud Vandeghen, Daniel Rebain, Matheus Gadelha, Yi Zhou, Anthony Cioppa, Ming C. Lin, Marc Van Droogenbroeck, Andrea Tagliasacchi*
15 Proxy-GS: Unified Occlusion Priors for Training and Inference in Structured 🏆 3D Gaussian Splatting, *Yuanyuan Gao, Yuning Gong, Yifei Liu, Jingfeng Li, Dan Xu, Yanci Zhang, Dingwen Zhang, Xiao Sun, Zhihang Zhong*
16 RetimeGS: Continuous-Time Reconstruction of 4D Gaussian Splatting, *Xuezhen Wang, Li Ma, Yulin Shen, Zeyu Wang, Pedro V. Sander*
17 Selfi: Self-improving Reconstruction Engine via 3D Geometric Feature Alignment, *Youming Deng, Songyou Peng, Junyi Zhang, Kathryn Heal, Tiancheng Sun, John Flynn, Steve Marschner, Lucy Chai*
18 Z-Order Transformer for Feed-Forward Gaussian Splatting, *Can Wang, Lei Liu, Wei Jiang, Dong Xu*
19 4D Primitive-Mâché: Glueing Primitives for Persistent 4D Scene Reconstruction, *Kirill Mazur, Marwan Taher, Andrew J. Davison*
20 Efficiently Reconstructing Dynamic Scenes One D4RT at a Time, 🏆 *Chuhan Zhang, Guillaume Le Moing, Skanda Koppula, Ignacio Rocco, Liliane Momeni, Junyu Xie, Shuyang Sun, Rahul Sukthankar, Joëlle K. Barral, Raia Hadsell, Zoubin Ghahramani, Andrew Zisserman, Junlin Zhang, Mehdi S. M. Sajjadi*
21 FUSER: Feed-Forward Multiview 3D Registration Transformer and 🏆 SE(3)^N Diffusion Refinement, *Haobo Jiang, Jin Xie, Jian Yang, Liang Yu, Jianmin Zheng*
22 Residual Primitive Fitting of 3D Shapes with SuperFrusta, 🏆 *Aditya Ganeshan, Matheus Gadelha, Thibault Groueix, Zhiqin Chen, Siddhartha Chaudhuri, Vladimir Kim, Wang Yifan, Daniel Ritchie*
23 SmokeSVD: Smoke Reconstruction from A Single View via 🏆 Progressive Novel View Synthesis and Refinement with Diffusion Models, *Chen Li, Shanshan Dong, Sheng Qiu, Jianmin Han, Yibo Zhao, Zan Gao, Taku Komura, Kemeng Huang*
24 SparseWorld-TC: Trajectory-Conditioned Sparse Occupancy World 🏆 Model, *Jiayuan Du, Yiming Zhao, Zhenglong Guo, Yong Pan, Wenbo Hou, Zhihui Hao, Kun Zhan, Qijun Chen*
25 Affostruction: 3D Affordance Grounding with Generative Reconstruction, *Chunghyun Park, Seunghyeon Lee, Minsu Cho*
26 MV-RoMa: From Pairwise Matching into Multi-View Track ❊ Reconstruction, *Jongmin Lee, Seungyeop Kang, Sungjoo Yoo*
27 Unified Primitive Proxies for Structured Shape Completion, *Zhaiyu Chen, Yuqing Wang, Xiao Xiang Zhu*
28 ART: Articulated Reconstruction Transformer, *Zizhang Li, Cheng Zhang, Zhengqin Li, Henry Howard-Jenkins, Zhaoyang Lv, Chen Geng, Jiajun Wu, Richard Newcombe, Jakob Engel, Zhao Dong*
29 SCE-SLAM: Scale-Consistent Monocular SLAM via Scene Coordinate Embeddings, *Yuchen Wu, Jiahe Li, Xiaohan Yu, Lina Yu, Jin Zheng, Xiao Bai*
30 S2D: Sparse to Dense Lifting for 3D Reconstruction with Minimal Inputs, *Yuzhou Ji, Qijian Tian, He Zhu, Xiaoqi Jiang, Guangzhi Cao, Lizhuang Ma, Yuan Xie, Xin Tan*
31 Pip-Stereo: Progressive Iterations Pruner for Iterative Optimization based Stereo Matching, *Jintu Zheng, Qizhe Liu, Huangxin Xu, Zhuojie Chen*
32 Fast-FoundationStereo: Real-Time Zero-Shot Stereo Matching, *Bowen Wen, Shaurya Dewan, Stan Birchfield*
33 E-RayZer: Self-supervised 3D Reconstruction as Spatial Visual ❊ Pre-training, *Qitao Zhao, Hao Tan, Qianqian Wang, Sai Bi, Kai Zhang, Kalyan Sunkavalli, Shubham Tulsiani, Hanwen Jiang*
34 QVGGT: Post-Training Quantized Visual Geometry Grounded Transformer, *Zhizhen Pan, Hesong Wang, Huan Wang*
35 SRGCD: Stability-Driven Region Growth Framework for 3D Change Detection, *Yue Wu, Tao Peng, Yongzhe Yuan, Kaiyuan Feng, Hao Li, Maoguo Gong, Qiguang Miao, Wenping Ma*
36 D-Prism: Differentiable Primitives for Structured Dynamic Modeling, *Xingyuan Yu, Yijin Li, Chong Zeng, Yuhang Ming, Hujun Bao, Guofeng Zhang*
37 STAC: Plug-and-Play Spatio-Temporal Aware Cache Compression for ❊ Streaming 3D Reconstruction, *Runze Wang, Yuxuan Song, Youcheng Cai, Ligang Liu*
38 Stabilizing Streaming Video Geometry via Dynamic Feature Normalization, *Xiaoyang Lyu, Muxin Liu, Xiaoshan Wu, Ruicheng Wang, Yi-Hua Huang, Yang-Tian Sun, Shaoshuai Shi, Xiaojuan Qi*
39 LaS-Comp: Zero-shot 3D Completion with Latent–Spatial Consistency, *Weilong Yan, Haipeng Li, Hao Xu, Nianjin Ye, Yihao Ai, Shuaicheng Liu, Jingyu Hu*
40 Pano360: Perspective to Panoramic Vision with Geometric Consistency, *Zhengdong Zhu, Weiyi Xue, Zuyuan Yang, Wenlve Zhou, Zhiheng Zhou*
41 EfficientMonoHair: Fast Strand-Level Reconstruction from Monocular Video via Multi-View Direction Fusion, *Da Li, Dominik Engel, Deng Luo, Ivan Viola*
42 OSPO: Object-Centric Self-Improving Preference Optimization for Text-to-Image Generation, *Yoonjin Oh, Yongjin Kim, Hyomin Kim, Donghwan Chi, Sungwoong Kim*
43 MoReGen: Multi-Agent Motion-Reasoning Engine for Code-based Text-to-Video Synthesis, *Xiangyu Bai, He Liang, Bishoy Galoaa, Utsav Nandi, Shayda Moezzi, Yuhang He, Sarah Ostadabbas*
44 StyleTextGen: Style-Conditioned Multilingual Scene Text Generation, *Zeyu Chen, Fangmin Zhao, Yan Shu, Yichao Liu, Liu Yu, Yu Zhou*
45 CRAFT-LoRA: Content-Style Personalization via Rank-Constrained Adaptation and Training-Free Fusion, *Yu Li, Yujun Cai, Chi Zhang*
46 OneHOI: Unifying Human-Object Interaction Generation and Editing, *Jiun Tian Hoe, Weipeng Hu, Xudong Jiang, Yap-Peng Tan, Chee Seng Chan*
47 GlyphPrinter: Region-Grouped Direct Preference Optimization for ❊ Glyph-Accurate Visual Text Rendering, *Xincheng Shuai, Ziye Li, Henghui Ding, Dacheng Tao*
48 Self-Paced and Self-Corrective Masked Prediction for Movie Trailer ❊ Generation, *Sidan Zhu, Hongteng Xu, Dixin Luo*
49 TV2TV: A Unified Framework for Interleaved Language and Video Generation, *Xiaochuang Han, Youssef Emad, Melissa Hall, John Nguyen, Karthik Padthe, Liam Robbins, Amir Bar, Delong Chen, Michal Drozdzal, Maha Elbayad, Yushi Hu, Shang-Wen Li, Jakob Verbeek, XuDong Wang, Marjan Ghazvininejad, Luke Zettlemoyer, Emily Dinan*
50 Narrative Weaver: Towards Controllable Long-Range Visual Consistency with Multi-Modal Conditioning, *Zhengjian Yao, Yongzhi Li, Xinyuan Gao, Quan Chen, Peng Jiang, Yanye Lu*
51 Ref4D-VideoBench: Four-Dimensional Reference-Based Evaluation of Text-to-Video Generative Models, *Jiajia Wei, Yujia He, Yuhan Hou, Hang Qi, Sihua Wang, Jincheng Shi, Kwok Fung Li, Zibin Zheng, Weibin Wu*
52 PureCC: Pure Learning for Text-to-Image Concept Customization, *Zhichao Liao, Xiaole Xian, Qingyu Li, Wenyu Qin, Meng Wang, Weicheng Xie, Siyang Song, Pingfa Feng, Long Zeng, Liang Pan*
53 Disentangling to Re-couple: Resolving the Similarity-Controllability Paradox in Subject-Driven Text-to-Image Generation, *Shuang Li, Chao Deng, Hang Chen, Liqun Liu, Zhenyu Hu, Te Cao, Mengge Xue, Yuan Chen, Peng Shu, Huan Yu, Jie Jiang*
54 Yume1.5: A Text-Controlled Interactive World Generation Model, *Xiaofeng Mao, Zhen Li, Chuanhao Li, Xiaojie Xu, Kaining Ying, Kaipeng Zhang*
55 PosterReward: Unlocking Accurate Evaluation for High-Quality Graphic Design Generation, *Jianyu Lai, Sixiang Chen, Jialin Gao, Hengyu Shi, Zhongying Liu, Fuxiang Zhai, Junfeng Luo, Xiaoming Wei, Lujia Wang, Lei Zhu*
56 Scone: Bridging Composition and Distinction in Subject-Driven Image ❊ Generation via Unified Understanding-Generation Modeling, *Yuran Wang, Bohan Zeng, Chengzhuo Tong, Wenxuan Liu, Yang Shi, Xiaochen Ma, Hao Liang, Yuanxing Zhang, Wentao Zhang*
57 SLVMEval: Synthetic Meta Evaluation Benchmark for Text-to-Long Video Generation, *Ryosuke Matsuda, Keito Kudo, Haruto Yoshida, Nobuyuki Shimizu, Jun Suzuki*
58 PROMPTMINER: Black-Box Prompt Stealing against Text-to-Image Generative Models via Reinforcement Learning and VLM-Guided Optimization, *Mingzhe Li, Renhao Zhang, Zhiyang Wen, Siqi Pan, Bruno Castro da Silva, Juan Zhai, Shiqing Ma*

59 FlowDirector: Training-Free Flow Steering for Precise Text-to-Video Editing, *Guangzhao Li, Yanming Yang, Chenxi Song, Xiaohong Liu, Chi Zhang*
60 Self-Evaluation Unlocks Any-Step Text-to-Image Generation, *Xin Yu, Xiaojuan Qi, Zhengqi Li, Kai Zhang, Richard Zhang, Zhe Lin, Eli Shechtman, Tianyu Wang, Yotam Nitzan*
61 Say Cheese! Detail-Preserving Portrait Collection Generation via Natural Language Edits, *Zelong Sun, Jiahui Wu, Ying Ba, Dong Jing, Zhiwu Lu*
62 LVLM-Aided Alignment of Task-Specific Vision Models, *Alexander Koebler, Lukas Kuhn, Ingo Thon, Florian Buettner*
63 DeepAlign: Mitigating Modality Conflict through Modality-Specific Alignment, *Shuo Li, Bingchen Miao, Wendong Bu, Juncheng Li, Hanwang Zhang, Fei Wu*
64 PG-VTON: Single-Pass Training-Free Virtual Try-On via Patch-Guided Reference Alignment, *Guohao Zhao, Yuxin Peng*
65 Linguistic Priors for Visual Decoupling: Towards Symmetric Vision-Brain Alignment, *Dongjun Liu, Weichen Dai, Jingsheng Qian, Honggang Liu, Hangjie Yi, Wanzeng Kong*
66 Scaling Spatial Intelligence with Multimodal Foundation Models, *Zhongang Cai, Ruisi Wang, Chenyang Gu, Fanyi Pu, Junxiang Xu, Yubo Wang, Wanqi Yin, Zhitao Yang, Chen Wei, Tongxi Zhou, Qingping Sun, Hui En Pang, Jiaqi Li, Oscar Qian, Zhiqian Lin, Xuanke Shi, Kewang Deng, Xiaoyang Han, Zukai Chen, Xiangyu Fan, Hanming Deng, Lewei Lu, Liang Pan, Bo Li, Ziwei Liu, Quan Wang, Dahua Lin, Lei Yang*
67 R-4B: Incentivizing General-Purpose Auto-Thinking Capability in MLLMs via Bi-Mode Annealing and Reinforce Learning, *Qi Yang, Bolin Ni, Shiming Xiang, Houwen Peng*
68 SafeGRPO: Self-Rewarded Multimodal Safety Alignment via Rule-Governed Policy Optimization, *Xuankun Rong, Wenke Huang, Tingfeng Wang, Daiguo Zhou, Bo Du, Mang Ye*
69 AVATAR: Reinforcement Learning to See, Hear, and Reason Over Video, *Yogesh Kulkarni, Pooyan Fazli*
70 CogniVerse: Revolutionizing Multi-Modal Retrieval-Augmented Generation with Cognitive Reflection and Geometric Reasoning, *Xiang Fang, Wanlong Fang, Changshuo Wang*
71 FOZO: Forward-Only Zeroth-Order Prompt Optimization for Test-Time Adaptation, *Xingyu Wang, Tao Wang*
72 Language Does Matter for Cross-Domain Few-Shot Visual Feature Enhancement, *Fei Zhou, Xiwen Zhang, Qingqing Qiu, Lei Zhang, Wei Wei, Chen Ding, Yi Zhang, Liang Li, Xiangyu Yue, Yanning Zhang*
73 Back to Source: Open-Set Continual Test-Time Adaptation via Domain Compensation, *Yingkai Yang, Chaoqi Chen, Hui Huang*
74 Bridging Domain Expertise and Generalization for Performance Estimation, *Shuxuan Li, Zhilin Zhao, Quyu Kong, Wei-Shi Zheng*
75 Adaptive Data Augmentation with Multi-armed Bandit: Sample-Efficient Embedding Calibration for Implicit Pattern Recognition, *Minxue Tang, Yangyang Yu, Aolin Ding, Maziyar Baran Pouyan, Taha Belkhouja, Yujia Bao*
76 Bridging Domains through Subspace-Aware Model Merging, *Levy Chaves, Chao Zhou, Rebekka Burkholz, Eduardo Valle, Sandra Avila*
77 DA-Mamba: Learning Domain-Aware State Space Model for Global-
❊ Local Alignment in Domain Adaptive Object Detection, *Haochen Li, Rui Zhang, Hantao Yao, Xin Zhang, Yifan Hao, Shaohui Peng, Yongwei Zhao, Ling Li*
78 Scaling Dense Event-Stream Pretraining from Visual Foundation
❊ Models, *Zhiwen Chen, Junhui Hou, Zhiyu Zhu, Jinjian Wu, Guangming Shi*
79 Event Stream Filtering via Probability Flux Estimation, *Jinze Chen, Wei Zhai, Yang Cao, Bin Li, Zheng-Jun Zha*
80 AIMDepth: Asymmetric Image-Event Mamba for Monocular Depth Estimation, *Luoxi Jing, Dianxi Shi, Yushe Cao, Yuanze Wang, Junze Zhang, Yuning Cui, Mengzhu Wang*
81 Time-Specialized Event-Image Alignment for Blur-to-Video Decomposition, *Zhijing Sun, Senyan Xu, Ruixuan Jiang, Kean Liu, Runze Tian, Xueyang Fu, Zheng-Jun Zha*
82 eRetinexGS: Retinex Modeling for Low-Light Scene Enhancement via Event Streams and 3D Gaussian Splatting, *Haojie Yan, Zehao Chen, Yan Liu, Shi Gu, Peng Lin, De Ma, Huajin Tang, Qian Zheng, Gang Pan*
83 Unsupervised 3d Motion Estimation Using Event Camera, *Han Han, Wei Zhai, Tiesong Zhao, Bin Li, Yang Cao, Zheng-jun Zha*
84 Goal-Driven Reward by Video Diffusion Models for Reinforcement Learning, *Qi Wang, Mian Wu, Yuyang Zhang, Mingqi Yuan, Wenyao Zhang, Haoxiang You, Yunbo Wang, Xin Jin, Xiaokang Yang, Wenjun Zeng*
85 ModularAgent: A Task-Aware Modular Framework for Joint Optimization of Multimodal Large Language Models and World Models, *Yu-Wei Zhan, Xin Wang, Pengzhe Mao, Tongtong Feng, Ren Wang, Wenwu Zhu*
86 AstraNav-Memory: Contexts Compression for Long Memory, *Junjun Hu, Xinda Xue, Botao Ren, Minghua Luo, Jintao Chen, Haochen Bai, Liangliang You, Mu Xu*
87 Test-Time Perturbation Learning with Delayed Feedback for Vision-Language-Action Models, *Zehua Zang, Xi Wang, Fuchun Sun, Xiao Xu, Lixiang Liu, Jiahuan Zhou, Jiangmeng Li*
88 OVSegDT: Segmenting Transformer for Open-Vocabulary Object Goal Navigation, *Tatiana Zemskova, Aleksei Staroverov, Dmitry Yudin, Aleksandr Panov*
89 ShowUI-π: Flow-based Generative Models as GUI Dexterous Hands, *Siyuan Hu, Kevin Qinghong Lin, Mike Zheng Shou*
90 ActiveVLA: Injecting Active Perception into Vision-Language-Action Models for Precise 3D Robotic Manipulation, *Zhenyang Liu, Yongchong Gu, Yikai Wang, Xiangyang Xue, Yanwei Fu*
91 ACoT-VLA: Action Chain-of-Thought for Vision-Language-Action Models, *Linqing Zhong, Yi Liu, Yifei Wei, Ziyu Xiong, Si Liu, Guanghui Ren*
92 BridgeEQA: Virtual Embodied Agents for Real Bridge Inspections, *Subin Varghese, Joshua Gao, Asad Ur Rahman, Vedhus Hoskere*
93 SyncMos: Scalable Motion Synchronisation for Multi-Agent Scene Interaction, *Lingxiao Li, Dongwon Kim, Lingyan Ruan, Bin Chen, Taesoo Kwon, Taehyun Rhee*
94 Planning in 8 Tokens: A Compact Discrete Tokenizer for Latent World Model, *Dongwon Kim, Gawon Seo, Jinsung Lee, Minsu Cho, Suha Kwak*
95 Omni-Attribute: Open-vocabulary Attribute Encoder for Visual Concept Personalization, *Tsai-Shien Chen, Aliaksandr Siarohin, Gordon Guocheng Qian, Kuan-Chieh Jackson Wang, Egor Nemchinov, Moayed Haji-Ali, Riza Alp Guler, Willi Menapace, Ivan Skorokhodov, Anil Kag, Jun-Yan Zhu, Sergey Tulyakov*
96 IF-Bench: Benchmarking and Enhancing MLLMs for Infrared Images with Generative Visual Prompting, *Tao Zhang, Yuyang Hong, Yang Xia, Kun Ding, Zeyu Zhang, Ying Wang, Shiming Xiang, Chunhong Pan*
97 InstantRetouch: Efficient and High-Fidelity Instruction-Guided Image Retouching with Bilateral Space, *Jiarui Wu, Yujin Wang, Ruikang Li, Fan Zhang, Mingde Yao, Tianfan Xue*
98 MICON-Bench: Benchmarking and Enhancing Multi-Image Context Image Generation in Unified Multimodal Models, *Mingrui Wu, Hang Liu, Jiayi Ji, Xiaoshuai Sun, Rongrong Ji*
99 The Devil is in Attention Sharing: Improving Complex Non-rigid Image Editing Faithfulness via Attention Synergy, *Zhuo Chen, Fanyue Wei, Runze Xu, Jingjing Li, Lixin Duan, Angela Yao, Wen Li*
100 ShreddingNet: Coarse-to-Fine Restoration for Multi-Source Shredded Manuscripts, *Haoyang Cui, Hao Jiang, Yadong Mu*
101 Image Guides Images: Consistent Video Amodal Completion with Rectified In-Context Exemplar Guidance, *Xiaoyu Kong, Ketong Ren, Dongyu She, Weiming Dong, Miao Wang*
102 Radiance Meshes for Volumetric Reconstruction, *Alexander Mai,*
❊ *Trevor Hedstrom, George Kopanas, Janne Kontkanen, Falko Kuester, Jonathan T. Barron*
103 Aesthetic Camera Viewpoint Suggestion with 3D Aesthetic Field, *Sheyang Tang, Armin Shafiee Sarvestani, Jialu Xu, Xiaoyu Xu, Zhou Wang*
104 CoRoGS: Contextual Gaussian Splatting for Robust Large-Deviation
❊ View Synthesis, *Xin Ma, Peng Lu, Yisong Chen, Chengwei Pan, Sheng Li*
105 ChronoGS: Disentangling Invariants and Changes in Multi-Period
❊ Scenes, *Zhongtao Wang, Jiaqi Dai, Qingtian Zhu, Yilong Li, Mai Su, Fei Zhu, Meng Gai, Shaorong Wang, Chengwei Pan, Yisong Chen, Guoping Wang*
106 Real-Time Dynamic Scene Rendering with Controlled Compressibility and Contact Awareness, *Boya Shi, Naiyang Guan, Xiaodong Yi*
107 Splatent: Splatting Diffusion Latents for Novel View Synthesis, *Or Hirschorn, Omer Sela, Inbar Huberman-Spiegelglas, Netalee Efrat, Eli Alshan, Ianir Ideses, Frederic Devernay, Yochai Zvik, Lior Fritz*
108 ParticleGS: Learning Neural Gaussian Particle Dynamics from Videos
❊ for Prior-free Physical Motion Extrapolation, *Jinsheng Quan, Qiaowei Miao, Yichao Xu, Zizhuo Lin, Ying Li, Wei Yang, Zhihui Li, Yawei Luo*

109 Dynamic-Static Decomposition for Novel View Synthesis of Dynamic Scenes with Spiking Neurons, *Lingyun Dai, Zehao Chen, Yan Liu, Shi Gu, Peng Lin, De Ma, Huajin Tang, Qian Zheng, Gang Pan*
110 DiffSoup: Direct Differentiable Rasterization of Triangle Soup for Extreme Radiance Field Simplification, *Kenji Tojo, Bernd Bickel, Nobuyuki Umetani*
111 Gyro-based Deep Video Deblurring, *Jaesung Rim, Woohyeok Kim, Haeyun Lee, Heemin Yang, Ke Wang, Sunghyun Cho*
112 ❊ Residual Diffusion Bridge Model for Image Restoration, *Hebaixu Wang, Jing Zhang, Haoyang Chen, Haonan Guo, Di Wang, Jiayi Ma, Bo Du*
113 MMDIR: Multimodal Instruction-Driven Framework for Mixed-Degradation Document Image Restoration, *Heng Li, Xingyuan Wang, Yang Fan, Yunan Zhang, Xiangping Wu, Qingcai Chen*
114 Rectifying Latent Space for Generative Single-Image Reflection Removal, *Mingjia Li, Jin Hu, Hainuo Wang, Qiming Hu, Jiarui Wang, Xiaojie Guo*
115 ❊ Towards Generalized Multimodal Homography Estimation, *Jinkun You, Jiaxin Cheng, Jie Zhang, Yicong Zhou*
116 Edit-aware RAW reconstruction, *Abhijith Punnappurath, Luxi Zhao, Ke Zhao, Hue Nguyen, Radek Grzeszczuk, Michael S. Brown*
117 Face2Scene: Using Facial Degradation as an Oracle for Diffusion-Based Scene Restoration, *Amirhossein Kazerouni, Maitreya Suin, Tristan Aumentado-Armstrong, Sina Honari, Amanpreet Walia, Iqbal Mohomed, Konstantinos G. Derpanis, Babak Taati, Alex Levinshtein*
118 ❊ HG-Lane: High-Fidelity Generation of Lane Scenes under Adverse Weather and Lighting Conditions without Re-annotation, *Daichao Zhao, Qiupu Chen, Feng He, Xin Ning, Qiankun Li*
119 NanoSD: Edge Efficient Foundation Model for Real Time Image Restoration, *Subhajit Sanyal, Srinivas Soumitri Miriyala, Akshay Janardan Bankar, Manjunath Arveti, Sowmya Vajrala, Shreyas Pandith, Sravanth Kodavanti, Abhishek Ameta, Harshit Harshit, Amit Satish Unde*
120 MR. Illuminate: Zero-Shot Low-Light Image Enhancement with Diffusion Prior, *Joshua Cho, Sara Aghajanzadeh, Zhen Zhu, David Forsyth*
121 FoundIR-v2: Optimizing Pre-Training Data Mixtures for Image Restoration Foundation Model, *Xiang Chen, Jinshan Pan, Jiangxin Dong, Jian Yang, Jinhui Tang*
122 SPEGC: Continual Test-Time Adaptation via Semantic-Prompt-Enhanced Graph Clustering for Medical Image Segmentation, *Xiaogang Du, Jiawei Zhang, Tongfei Liu, Tao Lei, Yingbo Wang*
123 BackSplit: The Importance of Sub-dividing the Background in Biomedical Lesion Segmentation, *Rachit Saluja, Asli Cihangir, Ruining Deng, Johannes C. Paetzold, Fengbei Liu, Mert R. Sabuncu*
124 Divide, Conquer, and Aggregate: Asymmetric Experts for Class-Imbalanced Semi-Supervised Medical Image Segmentation, *Yajun Liu*
125 ❊ CROWn: A Unified Framework for Anti-Aliased Downsampling and Phase-Calibrated Fusion in 3D Medical Segmentation, *Xingru Huang, Shuanghua Ye, Zhao Huang, Wenwen Tang, Huiyu Zhou, Zhiwen Zheng, Jin Liu, Xiaoshuai Zhang*
126 Rethinking Box Supervision: Bias-Free Weakly Supervised Medical Segmentation, *Jun Wei, Hui Huang*
127 Semi-supervised Echocardiography Video Segmentation via Anchor Semantic Awareness and Continuous Pseudo-label Reforging, *Yunpeng Fang, Yimu Sun, Jingxing Guo, Huisi Wu, Jing Qin*
128 TANGO: Learning Distribution-wise Foundation Prior Consistency and Instance-wise Style Calibration for Medical Image Generalization, *Chuang Liu, Yichao Cao, Xiu Su, Haogang Zhu*
129 MambaLiteUNet: Cross-Gated Adaptive Feature Fusion for Robust Skin Lesion Segmentation, *Md Maklachur Rahman, Soon Ki Jung, Tracy Hammond*
130 Breaking Multimodal LLM Safety via Video-Driven Prompting, *Dong Wang, Xiangyu He, Xinqi Lyu, Bin Xiao*
131 When LoRA Betrays: Backdooring Text-to-Image Models by Masquerading as Benign Adapters, *Liangwei Lyu, Jiaqi Xu, Jianwei Ding, Qiyao Deng*
132 RecoverMark: Robust Watermarking for Localization and Recovery of Manipulated Faces, *Haonan An, Xiaohui Ye, Guang Hua, Yihang Tao, Hangcheng Cao, Xiangyu Yu, Yuguang Fang*
133 A Provable Energy-Guided Test-Time Defense Boosting Adversarial Robustness of Large Vision-Language Models, *Mujtaba Hussain Mirza, Antonio D'Orazio, Odelia Melamed, Iacopo Masi*
134 FORCE: Transferable Visual Jailbreaking Attacks via Feature Over-Reliance CorrEction, *Runqi Lin, Alasdair Paren, Suqin Yuan, Muyang Li, Philip Torr, Adel Bibi, Tongliang Liu*
135 PureProof: Diffusion-Resistant Black-box Targeted Attack on Large Vision-Language Models, *Yiming Cao, Dong Wang, Xinqi Lyu, Bin Xiao*
136 UniDef: Universal Defense Against Unauthorized Image Manipulation, *Mingwen Shao, Lingzhuang Meng, Xiang Lv, Mengyao Wu, Xinyuan Chen, Qiao Zhang, Chang Liu, Yuanjian Qiao, Chao Dong*
137 ❊ Multi-Crit: Benchmarking Multimodal Judges on Pluralistic Criteria-Following, *Tianyi Xiong, Yi Ge, Ming Li, Zuolong Zhang, Pranav Kulkarni, Kaishen Wang, Qi He, Zeying Zhu, Chenxi Liu, Ruibo Chen, Tong Zheng, Yanshuo Chen, Xiyao Wang, Renrui Zhang, Wenhu Chen, Heng Huang*
138 MERLIN: Building Low-SNR Robust Multimodal LLMs for Electromagnetic Signals, *Junyu Shen, Zhendong She, Chenghanyu Zhang, Yuchuang Sun, Luqing Luo, Dingwei Tan, Zonghao Guo, Bo Guo, Zehua Han, Wupeng Xie, Yaxin Mu, Peng Zhang, Peipei Li, Fengxiang Wang, Yangang Sun, Maosong Sun*
139 Rethinking Cross-Modal Anchor Alignment for Mitigating Error Accumulation, *Bin Liu, Wei Sun, Qianqian Wang, Wei Feng, Yijie Chen, Haixi Zhang*
140 SOUPLE: Enhancing Audio-Visual Localization and Segmentation with Learnable Prompt Contexts, *Khanh Binh Nguyen, Chae Jung Park*
141 Omni-MMSI: Toward Identity-attributed Social Interaction Understanding, *Xinpeng Li, Bolin Lai, Hardy Chen, Shijian Deng, Cihang Xie, Yuyin Zhou, James M. Rehg, Yapeng Tian*
142 Inconsistency-aware Multimodal Schrödinger Bridge for Deepfake Localization, *Jiayu Xiong, Jing Wang, Qi Zhang, Wanlong Wang, Jun Xue*
143 MASQuant: Modality-Aware Smoothing Quantization for Multimodal Large Language Models, *Lulu Hu, Wenhu Xiao, Xin Chen, Xinhua Xu, Bowen Xu, Kun Li, Yongliang Tao*
144 Seeing Through Touch: Tactile-Driven Visual Localization of Material Regions, *Seongyu Kim, Seungwoo Lee, Hyeonggon Ryu, Joon Son Chung, Arda Senocak*
145 Seeing What Matters: A Training-Free Self-Guided Framework for Multimodal Detail Perception and Reasoning, *Mingjie Ma, yichao ma, Zhong Yang, Guohui Li*
146 Illuminating Visual Identity in Universal Multimodal Embeddings, *Jiawei Cao, Junyi Feng, Jiashen Hua, Ziheng Huang, Bing Deng, Kaijie Wu, Chaochen Gu, Jieping Ye*
147 Anti-Degradation Lifelong Multi-View Clustering, *Xingfeng Li, Hao Pan, Honglin Yuan, Yuan Sun, Xujian Zhao, Jiaqi Lin, Zhenwen Ren*
148 The Coherence Trap: When MLLM-Crafted Narratives Exploit Manipulated Visual Contexts, *Yuchen Zhang, Yaxiong Wang, Yujiao Wu, Lianwei Wu, Li Zhu, Zhedong Zheng*
149 Efficient and High-Fidelity Omni Modality Retrieval, *Chuong Huynh, Manh Luong, Abhinav Shrivastava*
150 Same Content, Different Answers: Cross-Modal Inconsistency in MLLMs, *Angela van Sprang, Laurens Samson, Ana Lucic, Erman Acar, Sennay Ghebreab, Yuki M. Asano*
151 Tri-Subspaces Disentanglement for Multimodal Sentiment Analysis, *Chunlei Meng, Jiabin Luo, Zhenglin Yan, Zhenyu Yu, Rong Fu, Zhongxue Gan, Chun Ouyang*
152 HAVE-Bench: Hierarchical Audio-Visual Evaluation from Perception to Interaction, *Muyan Zhong, Erfei Cui, Sen Xing, Weiyun Wang, Wen Wu, Yuchen Hu, Yanting Zhang, Xiaowei Hu, Wenhai Wang, Chao Zhang, Jifeng Dai*
153 Predictive Regularization Against Visual Representation Degradation in Multimodal Large Language Models, *Enguang Wang, Qiang Wang, Yuanchen Wu, Ke Yan, Xinbin Yuan, Shouhong Ding, Xialei Liu, Ming-Ming Cheng*
154 ❊ The More, The Merrier: Contrastive Fusion For Higher-Order Multimodal Alignment, *Stefanos Koutoupis, Michaela Areti Zervou, Konstantinos Kontras, Maarten De Vos, Panagiotis Tsakalides, Grigorios Tsagkatakis*
155 CineSRD: Leveraging Visual, Acoustic, and Linguistic Cues for Open-World Visual Media Speaker Diarization, *Liangbin Huang, Xiaohua Liao, Chaoqun Cui, Shijing Wang, Zhaolong Huang, Yanlong Du, Wenji Mao*
156 HandDreamer: Zero-Shot Text to 3D Hand Model Generation using Corrective Hand Shape Guidance, *Green Rosh, Prateek Kukreja, Vishakha SR, Pawan Prasad B H*

157 UST-Hand: An Uncertainty-aware Spatiotemporal Point Cloud Interaction Network for 3D Self-supervised Hand Pose Estimation, *Tianhao Han, Haoyang Zhang, Liang Xie, Haochen Chang, Kun Gao, Yuan Cheng, Pengfei Ren, Erwei Yin*
158 ForeHOI: Feed-forward 3D Object Reconstruction from Daily Hand-Object Interaction Videos, *Yuantao Chen, Jiahao Chang, Chongjie Ye, Chaoran Zhang, Zhaojie Fang, Chenghong Li, Xiaoguang Han*
159 Hoi! - A Multimodal Dataset for Force-Grounded, Cross-View ❊ Articulated Manipulation, *Tim Engelbracht, René Zurbrügg, Matteo Wohlrapp, Martin Büchner, Abhinav Valada, Marc Pollefeys, Hermann Blum, Zuria Bauer*
160 Enhancing Hands in 3D Whole-Body Pose Estimation with Conditional Hands Modulator, *Gyeongsik Moon*
161 TouchDream: 3D Object Completion through Imagined Touch, ❊ *Yuanbo Wang, Xinning Wang, Zhaoxuan Zhang, Changlong Wang, Qianchen Xia, Xiaopeng Wei, Xin Yang*
162 ForceVLA2: Unleashing Hybrid Force-Position Control with Force Awareness for Contact-Rich Manipulation, *Yang Li, Zhaxizhuoma Zhaxizhuoma, Hongru Jiang, Junjie Xia, Hongquan Zhang, Jinda Du, Yunsong Zhou, Jia Zeng, Ce Hao, Jieji Ren, Qiaojun Yu, Cewu Lu, Yu Qiao, Jiangmiao Pang*
163 TokenHand: Discrete Token Representation for Efficient Hand Mesh Reconstruction, *Xinguo He, Yixin Shen, Rahul Chaudhari*
164 Artiverse: A Diverse and Physically Grounded Dataset for Articulated Objects, *Denys Iliash, Jiayi Liu, Egor Fokin, Qirui Wu, Ali Mahdavi Amiri, Manolis Savva, Angel X. Chang*
165 MatPedia: A Universal Generative Foundation for High-Fidelity Material Synthesis, *Di Luo, Shuhui Yang, Mingxin Yang, Jiawei Lu, Yixuan Tang, Xintong Han, Zhuo Chen, Beibei Wang, Chunchao Guo*
166 LogCD: Local-to-global Consistency Distillation for Few-step Image Generation, *Qingsong Xie, Zhenyi Liao, Chen Chen, Zhijie Deng, Haonan Lu*
167 EditCtrl: Disentangled Local and Global Control for Real-Time Generative Video Editing, *Yehonathan Litman, Shikun Liu, Dario Seyb, Nicholas Milef, Yang Zhou, Carl Marshall, Shubham Tulsiani, Caleb Leak*
168 Anchoring and Rescaling Attention for Semantically Coherent ❊ Inbetweening, *Tae Eun Choi, Sumin Shim, Junhyeok Kim, Seong Jae Hwang*
169 FlashMotion: Few-Step Controllable Video Generation with Trajectory Guidance, *Quanhao Li, Zhen Xing, Rui Wang, Haidong Cao, Qi Dai, Daoguo Dong, Zuxuan Wu*
170 LightMover: Generative Light Movement with Color and Intensity Controls, *Gengze Zhou, Tianyu Wang, Soo Ye Kim, Zhixin Shu, Xin Yu, Yannick Hold-Geoffroy, Sumit Chaturvedi, Qi Wu, Zhe Lin, Scott Cohen*
171 Parallel Jacobi Decoding for Fast Autoregressive Image Generation, *Boya Liao, Ying Li, Siyong Jian, Huan Wang*
172 CARE-Edit: Condition-Aware Routing of Experts for Contextual Image Editing, *Yucheng Wang, Zedong Wang, Yuetong Wu, Yue Ma, Dan Xu*
173 CREval: An Automated Interpretable Evaluation for Creative Image Manipulation under Complex Instructions, *Chonghuinan Wang, Zihan Chen, Yuxiang Wei, Tianyi Jiang, Xiaohe Wu, Fan Li, Wangmeng Zuo, Hongxun Yao*
174 EchoVDiff: Cardiac-Cycle Echocardiography Video Generation from Arbitrary Frame, *Jiansong Zhang, Xiaying Yang, Xiaoling Luo, Linlin Shen*
175 Re-Align: Structured Reasoning-guided Alignment for In-Context Image Generation and Editing, *Runze He, Yiji Cheng, Tiankai Hang, Zhimin Li, Yu Xu, Zijin Yin, Shiyi Zhang, Wenxun Dai, Penghui Du, Ao Ma, Chunyu Wang, Qinglin Lu, Jizhong Han, Jiao Dai*
176 ChimeraLoRA: Multi-Head LoRA-Guided Synthetic Datasets, *Hoyoung Kim, Minwoo Jang, Jabin Koo, Sangdoo Yun, Jungseul Ok*
177 Frequency-Aware Flow Matching for High-Quality Image Generation, *Sucheng Ren, Qihang Yu, Ju He, Xiaohui Shen, Liang-Chieh Chen*
178 STARFlow-V: End-to-End Video Generative Modeling with Autoregressive ❊ Normalizing Flows, *Jiatao Gu, Ying Shen, Tianrong Chen, Laurent Dinh, Yuyang Wang, Miguel Ángel Bautista, David Berthelot, Josh Susskind, Shuangfei Zhai*
179 MixFlow Training: Alleviating Exposure Bias with Slowed Interpolation Mixture, *Hui Li, Jiayue Lyu, Fu-Yun Wang, Kaihui Cheng, Siyu Zhu, Jingdong Wang*
180 Improving Controllable Generation: Faster Training and Better Performance via x0-Supervision, *Amadou S. Sangare, Adrien Maglo, Mohamed Chaouch, Bertrand Luvison*
181 Visual-Aware CoT: Achieving High-Fidelity Visual Consistency in Unified Models, *Zixuan Ye, Quande Liu, Cong Wei, Yuanxing Zhang, Xintao Wang, Pengfei Wan, Kun Gai, Wenhan Luo*
182 OrionEdit: Bridging Reference and Source Images for Generalized Cross-Image Editing, *Zeyu Jiang, Lai Man Po, Xuyuan Xu, Yexin Wang, Guoping Gong, Haoxuan Wu, Chenbo Yan, Kun Li, Yuyang Liu*
183 PositionIC: Unified Position and Identity Consistency for Image Customization, *Junjie Hu, Tianyang Han, Kai Ma, Jialin Gao, Yang Song, Xianhua He, Junfeng Luo, Xiaoming Wei, Wenqiang Zhang*
184 P-Flow: Prompting Visual Effects Generation, *Rui Zhao, Mike Zheng Shou*
185 Clair Obscur: an Illumination-Aware Method for Real-World Image ❊ Vectorization, *Xingyue Lin, Shuai Peng, Xiangyu Xie, Jianhua Zhu, Yuxuan Zhou, Liangcai Gao*
186 SURF: Signature-Retained Fast Video Generation, *Kaixin Ding, Xi Chen, Sihui Ji, Yuan Gao, Liang Hou, Xin Tao, Hengshuang Zhao*
187 The devil is in the details: Enhancing Video Virtual Try-On via Keyframe-Driven Details Injection, *Qingdong He, Xueqin Chen, Yanjie Pan, Peng Tang, Pengcheng Xu, Zhenye Gan, Chengjie Wang, Xiaobin Hu, Jiangning Zhang, Yabiao Wang*
188 Lynx: Towards High-Fidelity Personalized Video Generation, *Shen Sang, Tiancheng Zhi, Tianpei Gu, Jing Liu, Linjie Luo*
189 VisionDirector: Vision-Language Guided Closed-Loop Refinement for Generative Image Synthesis, *Meng Chu, Senqiao Yang, Haoxuan Che, Suiyun Zhang, Xichen Zhang, Shaozuo Yu, Haokun Gui, Zhefan Rao, Dandan Tu, Rui Liu, Jiaya Jia*
190 ClusterMark: Towards Robust Watermarking for Autoregressive Image Generators with Visual Token Clustering, *Denis Lukovnikov, Andreas Müller, Erwin Quiring, Asja Fischer*
191 Stable Mean Flow: Lyapunov-Inspired One-Step Flow Matching, ❊ *Guangxun Zhang, Mason Haberle, Davi Geiger*
192 OPRO: Orthogonal Panel-Relative Operators for Panel-Aware In-Context Image Generation, *Sanghyeon Lee, Minwoo Lee, Euijin Shin, Kangyeol Kim, Seunghwan Choi, Jaegul Choo*
193 First Frame Is the Place to Go for Video Content Customization, *Jingxi Chen, Zongxia Li, Zhichao Liu, Guangyao Shi, Xiyang Wu, Fuxiao Liu, Cornelia Fermüller, Brandon Y. Feng, Yiannis Aloimonos*
194 Scaling Zero-Shot Reference-to-Video Generation, *Zijian Zhou, Shikun Liu, Haozhe Liu, Haonan Qiu, Zhaochong An, Weiming Ren, Zhiheng Liu, Xiaoke Huang, Kam-Woh Ng, Tian Xie, Xiao Han, Yuren Cong, Hang Li, Chuyan Zhu, Aditya Patel, Tao Xiang, Sen He*
195 MotionEdit: Benchmarking and Learning Motion-Centric Image Editing, *Yixin Wan, Lei Ke, Wenhao Yu, Kai-Wei Chang, Dong Yu*
196 VDOT: Efficient Unified Video Creation via Optimal Transport Distillation, *Yutong Wang, Haiyu Zhang, Tianfan Xue, Yu Qiao, Yaohui Wang, Chang Xu, Xinyuan Chen*
197 Real-Time Generation of Streamable Talking Portrait Video with ❊ Reference-Guided Deep Compression VAEs, *Sicheng Xu, Yu Deng, Shoukang Hu, Yichuan Wang, Yizhong Zhang, Zhan Chen, Jiaolong Yang, Baining Guo*
198 RunawayEvil: Jailbreaking the Image-to-Video Generative Models, *Songping Wang, Rufan Qian, Yueming Lyu, Qinglong Liu, Linzhuang Zou, Jie Qin, Songhua Liu, Caifeng Shan*
199 MultiAnimate: Pose-Guided Image Animation Made Extensible, *Yingcheng Hu, Haowen Gong, Chuanguang Yang, Zhulin An, Yongjun Xu, Songhua Liu*
200 Translating Signals to Languages for sEMG-Based Activity Recognition, *Ming Wang, Haoxuan Qu, Qiuhong Ke, Wei Zhou, Hossein Rahmani, Jun Liu*
201 Open the Motion Door: Atomic Motion Decomposition and Recomposition for Open-Vocabulary Motion Generation, *Ke Fan, Jiangning Zhang, Ran Yi, Jingyu Gong, Yabiao Wang, Yating Wang, Xin Tan, Chengjie Wang, Lizhuang Ma*
202 Multi-level Causal LLM-based Text-to-Motion Generation with Human Alignment, *Xiaodong Chen, Qian Bao, Xudong Liu, Jianping Fang, Jintao Fang, Yongdong Zhang, Tao Mei, Wu Liu*
203 MotionHiFlow: Text-to-Motion via Hierarchical Flow Matching, *Heng Li, Xiaotong Lin, Ling-An Zeng, Yulei Kang, Shuai Li, Jian-Fang Hu*
204 LaMoGen: Language to Motion Generation Through LLM-Guided Symbolic Inference, *Junkun Jiang, Ho Yin Au, Jingyu Xiang, Jie Chen*
205 Accelerating Diffusion via Hybrid Data-Pipeline Parallelism Based on Conditional Guidance Scheduling, *Euisoo Jung, Byunghyun Kim, Hyunjin Kim, Seonghye Cho, Jae-Gil Lee*

206 GVIS: Generative Vector Image Steganography, *Zihao Xu, Dawei Xu, Zihan Li, Xixi Zheng, Chuan Zhang*
207 MaxMark: High-Capacity Diffusion-Native Watermarking via Robust and Invertible Latent Embedding, *Xuanhang Chang, Zhonghao Yang, Cheng Zhuo, Yu Li*
208 GeoRK2: Geometry-Guided Runge–Kutta Integration for Diffusion
❊ Transformer Acceleration, *Chaoqun Sun, Zongjing Fu, Powei Chang, Jinpeng Zhang, Jianxiang Xiang, Yukang Gao, Chenyu Wang*
209 Test-time Sparsity for Extreme Fast Action Diffusion, *Kangye Ji, Yuan Meng, Jianbo Zhou, Ye Li, Chen Tang, Zhi Wang*
210 Trainable Log-linear Sparse Attention for Efficient Diffusion Transformers,
❊ *Yifan Zhou, Zeqi Xiao, Tianyi Wei, Shuai Yang, Xingang Pan*
211 A Self-Conditioned Representation Guided Diffusion Model for Realistic Text-to-LiDAR Scene Generation, *Wentao Qu, Guofeng Mei, Yang Wu, YongShun Gong, Xiaoshui Huang, Liang Xiao*
212 When Local Rules Create Global Order: Self-Organized Representation Learning for Latent Diffusion Models, *Junrong Lian, Weijian Deng, Pengxu Wei, Yaqin Chen, Qixiang Ye, Liang Lin*
213 ViStoryBench: Comprehensive Benchmark Suite for Story Visualization, *Cailin Zhuang, Ailin Huang, Yaoqi Hu, Jingwei Wu, Wei Cheng, Jiaqi Liao, Hongyuan Wang, Xinyao Liao, Weiwei Cai, Hengyuan Xu, Xuanyang Zhang, Xianfang Zeng, Zhewei Huang, Gang Yu, Chi Zhang*
214 R4-CGQA: Retrieval-based Vision Language Models for Computer Graphics Image Quality Assessment, *Zhuangzi Li, Jian Jin, Shilv Cai, Weisi Lin*
215 A³: Towards Advertising Aesthetic Assessment, *Kaiyuan Ji, Yixuan Gao, Lu Sun, Yushuo Zheng, Zijian Chen, Jianbo Zhang, Xiangyang Zhu, Yuan Tian, Zicheng Zhang, Guangtao Zhai*
216 GraphVLM: Benchmarking Vision Language Models for Multimodal Graph Learning, *Jiajin Liu, Dongzhe Fan, Chuanhao Ji, Daochen Zha, Qiaoyu Tan*
217 Phrase-Grounding-Aware Supervised Fine-Tuning for Chart Recognition via Side-Masked Attention, *Koichiro Ito*
218 VL-RouterBench: A Benchmark for Vision–Language Model Routing,
❊ *Zhehao Huang, Baijiong Lin, Jingyuan Zhang, Jingying Wang, Yuhang Liu, Ning Lu, Tao Li, Xiaolin Huang*
219 CLIP Is Shortsighted: Paying Attention Beyond the First Sentence,
❊ *Marc-Antoine Lavoie, Anas Mahmoud, Aldo Zaimi, Arsene Fansi Tchango, Steven L. Waslander*
220 G^2VLM: Geometry Grounded Vision Language Model with Unified 3D Reconstruction and Spatial Reasoning, *Wenbo Hu, Jingli Lin, Yilin Long, Yunlong Ran, Lihan Jiang, Yifan Wang, Chenming Zhu, Runsen Xu, Tai Wang, Jiangmiao Pang*
221 UZ3DVG: Unaided Zero-Shot 3D Visual Grounding with Generated Language Conditions, *Wenbin Tan, Jiawen Lin, Yuan Xie, Yachao Zhang, Yanyun Qu*
222 LangField4D: Learning Identity-Adaptive and Spatio-Temporal Continuous 4D Language Fields for Dynamic Scenes, *Yichao Xu, Qiaowei Miao, Jinsheng Quan, Wei Yang, Zhihui Li, Yawei Luo*
223 Spatial-SSRL: Enhancing Spatial Understanding via Self-Supervised Reinforcement Learning, *Yuhong Liu, Beichen Zhang, Yuhang Zang, Yuhang Cao, Long Xing, Xiaoyi Dong, Haodong Duan, Dahua Lin, Jiaqi Wang*
224 CLIPoint3D: Language-Grounded Few-Shot Unsupervised 3D Point Cloud Domain Adaptation, *Mainak Singha, Sarthak Mehrotra, Paolo Casari, Subhasis Chaudhuri, Elisa Ricci, Biplab Banerjee*
225 GeoTikzBridge: Advancing Multimodal Code Generation for Geometric Perception and Reasoning, *Jiayin Sun, Caixia Sun, Boyu Yang, Hailin Li, Xiao Chen, Yi Zhang, Errui Ding, Liang Li, Chao Deng, Junlan Feng*
226 Keep it SymPL: Symbolic Projective Layout for Allocentric Spatial Reasoning in Vision-Language Models, *Jaeyun Jang, Seunghui Shin, Taeho Park, Hyoseok Hwang*
227 Geometry-Guided 3D Visual Token Pruning for Video-Language Models, *Han Li, Zehao Huang, Jiahui Fu, Naiyan Wang, Si Liu*
228 Context-Nav: Context-Driven Exploration and Viewpoint-Aware 3D Spatial Reasoning for Instance Navigation, *Won Shik Jang, Ue-Hwan Kim*
229 Learning to Reason in 4D: Dynamic Spatial Understanding for Vision Language Models, *Shengchao Zhou, Yuxin Chen, Yuying Ge, Wei Huang, Jiehong Lin, Ying Shan, Xiaojuan Qi*
230 PanoEnv: Exploring 3D Spatial Intelligence in Panoramic Environments
❊ with Reinforcement Learning, *Zekai Lin, Xu Zheng*
231 Hilbert-Geo: Solving Solid Geometric Problems by Neural-Symbolic Reasoning, *Ruoran Xu, Haoyu Cheng, Bin Dong, Qiufeng Wang*
232 Direction-aware 3D Large Multimodal Models, *Quan Liu, Weihao Xuan, Junjue Wang, Naoto Yokoya, Ling Shao, Shijian Lu*
233 CLAY: Conditional Visual Similarity Modulation in Vision-Language Embedding Space, *Sohwi Lim, Lee Hyoseok, Jungjoon Park, Tae-Hyun Oh*
234 Tackling Alignment Ambiguity in Person Retrieval through Conversational Attribute Mining, *Hao Zou, Runqing Zhang, Jin Ding, Xue Zhou, Jianxiao Zou, Mingzhu Cai*
235 Beyond Global Similarity: Multi-Conditional Retrieval for Fine-Grained Cross-Modal Understanding, *Xuan Lu, Kangle Li, Haohang Huang, Rui Meng, Wenjun Zeng, Xiaoyu Shen*
236 Imagine Before Concentration: Diffusion-Guided Registers Enhance Partially Relevant Video Retrieval, *Jun Li, Xuhang Lou, Jinpeng Wang, Yuting Wang, Yaowei Wang, Shu-Tao Xia, Bin Chen*
237 What Is the Optimal Ranking Score Between Precision and Recall? We Can Always Find It and It Is Rarely F1, *Sébastien Piérard, Adrien Deliège, Marc Van Droogenbroeck*
238 Robust Remote Sensing Image–Text Retrieval with Noisy Correspondence, *Qiya Song, Yiqiang Xie, Yuan Sun, Renwei Dian, Xudong Kang*
239 PinPoint: Evaluation of Composed Image Retrieval with Explicit Negatives, Multi-Image Queries, and Paraphrase Testing, *Rohan Mahadev, Joyce Yuan, Patrick Poirson, David Xue, Hao-Yu Wu, Dmitry Kislyuk*
240 Single-step Diffusion-based Video Coding with Semantic-Temporal Guidance, *Naifu Xue, Zhaoyang Jia, Jiahao Li, Bin Li, Zihan Zheng, Yuan Zhang, Yan Lu*
241 Memory Matters: Boosting Training-Free Zero-Shot Temporal Action Localization with a Learnable Lookup Table, *Han Jiang, Haoyu Tang, Xiaoxuan Mu, Chen Li, Jihua Zhu*
242 TVHighlights: LLM-Guided Human-Free Collaborative Training for Video Highlight Detection in Movies and TV Dramas, *Qi Qiu, Xuan Wu, Jiawei Peng, Yuan Miao, Xu Yang, Yanlong Du*
243 Color When It Counts: Grayscale-Guided Online Triggering for Always-On Streaming Video Sensing, *Weitong Cai, Hang Zhang, Yukai Huang, Shitong Sun, Jiankang Deng, Songcen Xu, Jifei Song, Zhensong Zhang*
244 Reinforcing Structured Chain-of-Thought for Video Understanding, *Peiyao Wang, Haotian Xu, Noranart Vesdapunt, Rui Hou, Jingyi Zhang, Haibin Ling, Oleksandr Obiednikov, Ning Zhou, Kah Kuen Fu*
245 FlexiVideo: Variation-Aware Temporal Dynamics Modeling for Efficient Video Understanding, *Da Peng, Xuesong Yang, Zonghao Guo, Yichen Zhang, Chi Chen, Yidan Zhang, Yuan Yao, Fang Wan, Wei Ke, Maosong Sun*
246 MS-Temba: Multi-Scale Temporal Mamba for Understanding Long Untrimmed Videos, *Arkaprava Sinha, Monish Soundar Raj, Pu Wang, Ahmed Helmy, Hieu Le, Srijan Das*
247 Learning Effective Sign Features without Text for Gloss-free Sign
❊ Language Translation, *Shiwei Gan, Xiao Liu, Yafeng Yin, Nan Liu, Kuizhuang Liu, Desibieer Tuerdaken, Zhiwei Jiang, Lei Xie, Sanglu Lu, Hongkai Wen*
248 META: Meta Evolution of Tool Trajectory Adaptation for Long-Video Understanding, *Jing Huang, Luyuan Chen, Zhijie Xu, Yadong Li, Xingzhong Xu, Siye Chen, Jie Liu, Ming Kong, Qiang Zhu*
249 GT-SVJ: Generative-Transformer-Based Self-Supervised Video Judge For Efficient Video Reward Modeling, *Shivanshu Shekhar, Uttaran Bhattacharya, Raghavendra Addanki, Mehrab Tanjim, Somdeb Sarkhel, Tong Zhang*
250 Local Motion Matters: A Deconstruct–Recompose Paradigm for Reinforcement Learning Pre-training from Videos, *Jinwen Wang, Youfang Lin, Xiaobo Hu, Shuo Wang, Kai Lv*
251 Align Once to Explain: Feature Alignment for Scalable B-cosification of Foundational Vision Transformers, *Raphael Maser, Siddhartha Gairola, Sukrut Rao, Bernt Schiele*
252 Rounded or Streamlined Head? Bridging Concept Bottleneck Models
❊ and Attribute-Described Object Parts, *Yang Liu, Jiajin Zhang, Yaojun Hu, Bingguang Hao, Xin Cao, Yingda Xia, Danyang Tu, Shi Gu, Ling Zhang*

253 CIGMA: Causal Information-Gain Mechanistic Attribution of Attention Heads in Vision Transformers, *Maisha Maliha, Dean F. Hougen*
254 Rethinking Concept Bottleneck Models: From Pitfalls to Solutions, *Merve Tapli, Quentin Bouniot, Wolfgang Stammer, Zeynep Akata, Emre Akbas*
255 Make it SING: Analyzing Semantic Invariants in Classifiers, *Harel Yadid, Meir Yossef Levi, Roy Betser, Guy Gilboa*
256 Back to the Feature: Explaining Video Classifiers with Video Counterfactual Explanations, *Chao Wang, Chengan Che, Xinyue Chen, Sophia Tsoka, Luis C. Garcia-Peraza-Herrera*
257 ❊ LEADER: Learning Reliable Local-to-Global Correspondences for LiDAR Relocalization, *Jianshi Wu, Minghang Zhu, Dunqiang Liu, Wen Li, Sheng Ao, Siqi Shen, Chenglu Wen, Cheng Wang*
258 UniCorrn: Unified Correspondence Transformer Across 2D and 3D, *Prajnan Goswami, Tianye Ding, Feng Liu, Huaizu Jiang*
259 Probabilistic Discrepancy Learning for Roadside LiDAR Scene Completion, *Xiaogang Wu, Jinchao Hu, Zixian Wang, Dun Liu, BoXiang Cheng, Yiqiang Wu*
260 TACO: Task-Aware Contrastive Learning for Joint LiDAR Localization and 3D Object Detection, *Leyuan Xing, Huanjia Zhang, Dongyu Pan, Hai Wu, Qiming Xia, Kezheng Xiong, Wen Li, Chenglu Wen, Cheng Wang*
261 Adapting Point Cloud Analysis via Multimodal Bayesian Distribution Learning, *Xingyu Zhu, Liang Yi, Shuo Wang, Wenbo Zhu, Yongliang Wu, Beier Zhu, Hanwang Zhang*
262 Learning Coordinate-based Convolutional Kernels for Continuous SE(3) Equivariant and Efficient Point Cloud Analysis, *Jaein Kim, Hee Bin Yoo, Dong-Sig Han, Byoung-Tak Zhang*
263 R3-PCQA: Ray-Reprojection-Reinforcement for No-Reference 3D Point Cloud Quality Assessment, *Junhyuk Seo, Sanghyuk Seo, Dawoon Kim, Heeseok Oh*
264 Geometric-Aware Hypergraph Reasoning for Novel Class Discovery in Point Cloud Segmentation, *Zihao Zhang, Aming Wu, Yang Li, Yahong Han, Jialie Shen*
265 PointCSP: Cross-Sample Semantic Propagation and Stability Preservation in Self-Supervised Point Cloud Learning, *Xinxing Yu, Ajian Liu, Sunyuan Qiang, Hui Ma, Liying Yang, Yuzhong Wang, Zhi Rao, Yanyan Liang*
266 ❊ U4D: Uncertainty-Aware 4D World Modeling from LiDAR Sequences, *Xiang Xu, Alan Liang, Youquan Liu, Linfeng Li, Lingdong Kong, Ziwei Liu, Qingshan Liu*
267 TerraSeg: Self-Supervised Ground Segmentation for Any LiDAR, *T ed Lentsch, Santiago Montiel-Marín, Holger Caesar, Dariu M. Gavrila*
268 Where Does Vision Meet Language? Understanding and Refining Visual Fusion in MLLMs via Contrastive Attention, *Shezheng Song, Shasha Li, Shan Zhao, Xiaopeng Li, Qian Wan, Chengyu Wang, Tianwei Yan, Ma Jun, Jie Yu*
269 UniRefiner: Teaching Pre-trained ViTs to Self-Dispose Dross via Contrastive Register, *Congpei Qiu, Zhaoyu Hu, Wei Ke, Zhuotao Tian, Yanhao Wu, Tong Zhang*
270 ❊ SigLino: Efficient Multi-Teacher Distillation for Agglomerative Vision Foundation Models, *Sofian Chaybouti, Sanath Narayan, Yasser Dahou, Phúc H. Lê Khắc, Ankit Singh, Ngoc Huynh, Wamiq Reyaz Para, Hilde Kuehne, Hakim Hacid*
271 Heuristic-inspired Reasoning Priors Facilitate Data-Efficient Referring Object Detection, *Xu Zhang, Zhe Chen, Jing Zhang, Dacheng Tao*
272 LLaDA-V: Large Language Diffusion Models with Visual Instruction Tuning, *Zebin You, Shen Nie, Xiaolu Zhang, JUN ZHOU, Zhiwu Lu, Ji-Rong Wen, Chongxuan Li*
273 AVION: Aerial Vision–Language Instruction from Offline Teacher to Prompt-Tuned Network, *Yu Hu, Jianyang Gu, Hao Liu, Yue Cao, Jozsef Hamari, Zheng Liu, Mohsen Zardadi*
274 CrossVL: Complexity-Aware Feature Routing and Paired Curriculum for Cross-View Vision-Language Detection, *Zhipeng Liu, Chunbo Luo*
275 Masking Teacher and Reinforcing Student for Distilling Vision-Language Models, *Byung-Kwan Lee, Yu-Chiang Frank Wang, Ryo Hachiuma*
276 Role-SynthCLIP: A Role-Play Driven Diverse Synthetic Data Approach, *Yuanxiang Huangfu, Chaochao Wang, Weilei Wang*
277 ❊ BiMotion: B-spline Motion for Text-guided Dynamic 3D Character Generation, *Miaowei Wang, Qingxuan Yan, Zhi Cao, Yayuan Li, Oisin Mac Aodha, Jason J Corso, Amir Vaxman*
278 PSDesigner: Automated Graphic Design with a Human-Like Creative Workflow, *Xincheng Shuai, Song Tang, Yutong Huang, Henghui Ding, Dacheng Tao*
279 CADFS: A Big CAD Program Dataset and Framework for Computer-Aided Design with Large Language Models, *Vladislav Pyatov, Gleb Bobrovskikh, Saveliy Galochkin, Nikita Boldyrev, Oleg Voynov, Alexander Filippov, Gonzalo Ferrer, Peter Wonka, Evgeny Burnaev*
280 MapRoute:Precise-Concept Erasing Mappers via Semantic Routing, *Sihao Li, Baixi Liang, Shuohong Xia, Yunyun Yang*
281 PhotoFramer: Multi-modal Image Composition Instruction, *Zhiyuan You, Ke Wang, He Zhang, Xin Cai, Jinjin Gu, Tianfan Xue, Chao Dong, Zhoutong Zhang*
282 Can We Build Scene Graphs, Not Classify Them? FlowSG: Progressive Image-Conditioned Scene Graph Generation with Flow Matching, *Xin Hu, Ke Qin, Wen Yin, Yuan-Fang Li, Ming Li, Tao He*
283 DuetSVG: Unified Multimodal SVG Generation with Internal Visual Guidance, *Peiying Zhang, Nanxuan Zhao, Matthew Fisher, Yiran Xu, Jing Liao, Difan Liu*
284 Bias Is a Subspace, Not a Coordinate: A Geometric Rethinking of Post-hoc Debiasing in Vision-Language Models, *Dachuan Zhao, Weiyue Li, Zhenda Shen, Yushu Qiu, Bowen Xu, Haoyu Chen, Yongchao Chen*
285 ❊ Frequency-domain Manipulation for Face Obfuscation, *Jintae Kim, Keunsoo Ko, Chang-Su Kim*
286 Towards Reasoning-Preserving Unlearning in Multimodal Large Language Models, *Hongji Li, Manjiang Yu, Junchi Yao, Priyanka Singh, Xue Li, Di Wang, Lijie Hu*
287 Erasing Thousands of Concepts: Towards Scalable and Practical Concept Erasure for Text-to-Image Diffusion Models, *Hoigi Seo, Byung Hyun Lee, Jaehyun Cho, Sungjin Lim, Se Young Chun*
288 POUR: A Provably Optimal Method for Unlearning Representation via Neural Collapse, *Anjie Le, Can Peng, Yuyuan Liu, J. Alison Noble*
289 Do Vision-Language Models Leak What They Learn? Adaptive Token-Weighted Model Inversion Attacks, *Ngoc-Bao Nguyen, Sy-Tuyen Ho, Koh Jun Hao, Ngai-Man Cheung*
290 Protego: User-Centric Pose-Invariant Privacy Protection Against Face Recognition-Induced Digital Footprint Exposure, *Ziling Wang, Shuya Yang, Jialin Lu, Ka-Ho Chow*
291 ❊ SPDMark: Selective Parameter Displacement for Robust Video Watermarking, *Samar Fares, Nurbek Tastan, Karthik Nandakumar*
292 ❊ Enhancing Visual Representation with Textual Semantics: Textual Semantics-Powered Prototypes for Heterogeneous Federated Learning, *Xinghao Wu, Jianwei Niu, Xuefeng Liu, Guogang Zhu, Jiayuan Zhang, Shaojie Tang, Wei Chen*
293 ❊ FedHarmony: Harmonizing Heterogeneous Label Correlations in Federated Multi-Label Learning, *Zhiqiang Kou, Junxiang Wu, Wenke Huang, Wenwen He, Ming-Kun Xie, Changwei Wang, Yuheng Jia, Di Jiang, Yang Liu, Xin Geng, Qiang Yang*
294 FedSST: Rethinking Fair Federated Graph Learning under Structural Shift, *Dingyi Zhao*
295 GDFA: Geometry-Driven Federated Unlearning with Directional Task Vector Alignment, *Xiuting Weng, Ruizhi Pu, Yuanhang Yao, Kun Yue, Zhiwen Tang, Lixing Yu*
296 FedARA: Resource-adaptive Low-rank Personalized Federated Learning via Anchor-driven Representation Alignment on Heterogeneous Edge Devices, *Ruonan Zhao, Zheng Wang, Debin Liu, Shijie Lv, Laurence Tianruo Yang*
297 InterRVOS: Interaction-Aware Referring Video Object Segmentation, *Woojeong Jin, Seongchan Kim, Jaeho Lee, Seungryong Kim*
298 RE-VLM: Event-Augmented Vision-Language Model for Scene Understanding, *Hanqing Liu, Mingjie Liu, Luoping Cui, Endian Lin, Donghong Jiang, Chuang Zhu*
299 RegFormer: Transferable Relational Grounding for Efficient Weakly-Supervised Human-Object Interaction Detection, *Jihwan Park, Chanhyeong Yang, Jinyoung Park, Taehoon Song, Hyunwoo J. Kim*
300 Learning to Refuse: Refusal-Aware Reinforcement Fine-Tuning for Hard-Irrelevant Queries in Video Temporal Grounding, *Jin-Seop Lee, SungJoon Lee, SeongJun Jung, Boyang Li, Jee-Hyong Lee*
301 GroundVTS: Visual Token Sampling in Multimodal Large Language Models for Video Temporal Grounding, *Rong Fan, Kaiyan Xiao, Minghao Zhu, Liuyi Wang, Kai Dai, Zhao Yang*

302 TimeLens: Rethinking Video Temporal Grounding with Multimodal LLMs, *Jun Zhang, Teng Wang, Yuying Ge, Yixiao Ge, Xinhao Li, Limin Wang*
303 Tokenization Allows Multimodal Large Language Models to Understand, Generate and Edit Architectural Floor Plans, *Sizhong Qin, Ramon Elias Weber, Xinzheng Lu*
304 MeToM: Metadata-Guided Token Merging for Efficient Video LLMs, *Zhuojie Wu, Shijie Wang, Xin Yu*
305 Token Reduction via Local and Global Contexts Optimization for Efficient Video Large Language Models, *Jinlong Li, Liyuan Jiang, Haonan Zhang, Nicu Sebe*
306 VLIC: Vision-Language Models As Perceptual Judges for Human-Aligned Image Compression, *Kyle Sargent, Ruiqi Gao, Philipp Henzler, Charles Herrmann, Aleksander Holynski, Li Fei-Fei, Jiajun Wu, Jason Y. Zhang*
307 Mostly Text, Smart Visuals: Asymmetric Text-Visual Pruning for Large Vision-Language Models, *Sijie Li, Biao Qian, Jungong Han*
308 Attention-aware Inference Optimizations for Large Vision-Language Models with Memory-efficient Decoding, *Fatih Ilhan, Gaowen Liu, Ramana Rao Kompella, Selim Furkan Tekin, Tiansheng Huang, Zachary Yahn, Yichang Xu, Ling Liu*
309 ❊ CoIn: Coverage and Informativeness-Guided Token Reduction for Efficient Large Multimodal Models, *Chenxi Du, Yongheng Deng, Jiani Liu, Yujia Zhang, Xi Chen, Ju Ren*
310 TAMER: A Tri-Modal Contrastive Alignment and Multi-Scale Embedding Refinement Framework for Zero-Shot ECG Diagnosis, *Xuewei Zhou, Yajie Meng, Pan Zeng, Xianfang Tang, Feifei Cui, Qiangguo Jin, Jialiang Yang, Junlin Xu*
311 Your Dissimilarities Define You: Complementary Learning Exploiting Class Diversities, *Dimitrios Katsikas, Nikolaos Passalis, Anastasios Tefas*
312 CGU-Bayes: Causal Graph Uncertainty-Guided Bayesian Inference for Domain Generalization, *Naiyu Yin, Hanjing Wang, Yue Yu, Tian Gao, Amit Dhurandhar, Chung-Hao Lee, Qiang Ji*
313 Franca: Nested Matryoshka Clustering for Scalable Visual Representation Learning, *Shashanka Venkataramanan, Valentinos Pariza, Mohammadreza Salehi, Lukas Knobel, Elias Ramzi, Spyros Gidaris, Andrei Bursuc, Yuki M Asano*
314 Towards Stable Self-Supervised Object Representations in Unconstrained Egocentric Video, *Yuting Tan, Xilong Cheng, Yunxiao Qin, Zhengnan Li, Jingjing Zhang*
315 LRDUN: A Low-Rank Deep Unfolding Network for Efficient Spectral Compressive Imaging, *He Huang, Yujun Guo, Wei He*
316 Neural Collapse in Test-Time Adaptation, *Xiao Chen, Zhongjing Du, Jiazhen Huang, Xu Jiang, Li Lu, Jingyan Jiang, Zhi Wang*
317 CLEX: Complementary Label Exchange Learning for Noisy Facial Expression Recognition, *Lin Wang, Fang Liu, Xiaofen Xing, Kailing Guo, Xiangmin Xu*
318 TruckDrive: Long-Range Autonomous Highway Driving Dataset, *Filippo Ghilotti, Edoardo Palladin, Samuel Brucker, Adam Sigal, Mario Bijelic, Felix Heide*
319 Neuro-Cognitive Reward Modeling for Human-Centered Autonomous Vehicle Control, *Zhuoli Zhuang, Yu-Cheng Chang, Yu-Kai Wang, Thomas Do, Chin-Teng Lin*
320 E3AD: An Emotion-Aware Vision-Language-Action Model for Human-Centric End-to-End Autonomous Driving, *Yihong Tang, Haicheng Liao, Tong Nie, Junlin He, Ao Qu, Kehua Chen, Wei Ma, Zhenning Li, Lijun Sun, Chengzhong Xu*
321 The Blind Spot of Adaptation: Quantifying and Mitigating Forgetting in Fine-tuned Driving Models, *Runhao Mao, Hanshi Wang, Yixiang Yang, Qianli Ma, Jingmeng Zhou, Zhipeng Zhang*
322 Den-TP: A Density-Balanced Data Curation and Evaluation Framework for Trajectory Prediction, *Ruining Yang, Yi Xu, Yun Fu, Lili Su*
323 Percept-WAM: Perception-Enhanced World-Awareness-Action Model for Robust End-to-End Autonomous Driving, *Jianhua Han, Meng Tian, Jiangtong Zhu, Fan He, Huixin Zhang, Sitong Guo, Dechang Zhu, Hao Tang, Pei Xu, Yuze Guo, Minzhe Niu, Haojie Zhu, Qichao Dong, Xuechao Yan, Siyuan Dong, Lu Hou, Qingqiu Huang, Xiaosong Jia, Hang Xu*
324 GaussianDWM: 3D Gaussian Driving World Model for Unified Scene Understanding and Multi-Modal Generation, *Tianchen Deng, Xuefeng Chen, Yi Chen, Qu Chen, Yuyao Xu, Lijin Yang, Le Xu, Yu Zhang, Bo Zhang, Wuxiong Huang, Hesheng Wang*
325 Mind the Hitch: Dynamic Calibration and Articulated Perception for Autonomous Trucks, *Morui Zhu, Yongqi Zhu, Song Fu, Qing Yang*
326 DriveMoE: Mixture-of-Experts for Vision-Language-Action Model in End-to-End Autonomous Driving, *Zhenjie Yang, Yilin Chai, Xiaosong Jia, Qifeng Li, Yuqian Shao, Xuekai Zhu, Haisheng Su, Junchi Yan*
327 Beyond Rule-Based Agents: Active Markov Games for Realistic Multi-Agent Interaction in Autonomous Driving, *Yuan Gui, Hongchen Luo, Jiao Wang, Liqi Qu*
328 ❊ Test-Time Multi-Prompt Adaptation for Open-Vocabulary Remote Sensing Image Segmentation, *Ting Yang, Qilong Wang, Qibin Hou, Qinghua Hu*
329 ReScene4D: Temporally Consistent Semantic Instance Segmentation of Evolving Indoor 3D Scenes, *Emily Steiner, Jianhao Zheng, Henry Howard-Jenkins, Chris Xie, Iro Armeni*
330 CrackSSM: Reviving SSMs for Crack Segmentation via Dynamic Scanning, *Yubin Gu, Boyang Hou, Yuan Meng, Wenting Luo, Jiayi Ji, Xiaoshuai Sun*
331 BiPA: Bilevel Prompt Adaptation for Underwater Instance Segmentation, *Long Ma, Haoze Zheng, Yuhang Mao, Jinyuan Liu, Chengpei Xu, Xinwei Xue, Yi Wang, Xiangjian He, Weimin Wang*
332 RS-SSM: Refining Forgotten Specifics in State Space Model for Video Semantic Segmentation, *Kai Zhu, Zhenyu Cui, Zehua Zang, Jiahuan Zhou*
333 Scene-Centric Unsupervised Video Panoptic Segmentation, *Christoph Reich, Oliver Hahn, Nikita Araslanov, Laura Leal-Taixé, Christian Rupprecht, Daniel Cremers, Stefan Roth*
334 Bootstrapping Video Semantic Segmentation Model via Distillation-assisted Test-Time Adaptation, *Jihun Kim, Hoyong Kwon, Hyeokjun Kweon, Kuk-Jin Yoon*
335 GeoFree-CoSeg: Unsupervised Point Cloud-Image Cross-Modal Co-Segmentation Without Geometric Alignment, *Xin Duan, Xiabi Liu, Liyuan Pan*
336 Parameter-efficient Continual Learning for Enhancing Plasticity without Forgetting under Limited Model Capacity, *Yitian Chen, Shigeng Zhang, Xuan Liu, Mingming Lu, Kai Chen, Hongye Zhu, Xinning Chen*
337 Dual-Estimator: Decoupling Global and Local Semantic Shift for Drift Compensation in Class-Incremental Learning, *Fankang Xu, Lu Jin, Yanpeng Sun, Shiyu Xuan, Zechao Li*
338 Continual Distillation of Teachers from Different Domains, *Nicolas Michel, Maorong Wang, Jiangpeng He, Toshihiko Yamasaki*
339 Multimodal Continual Instruction Tuning with Dynamic Gradient Guidance, *Songze Li, Mingyu Gao, Tonghua Su, Xu-Yao Zhang, Zhongjie Wang*
340 Learning from Itself: Mining Internal Knowledge from Vision Language Models for Continual Learning, *Yizheng Gong, Siyue Yu, Waleed Al-Nuaimy, Jimin Xiao*
341 ❊ AdaPrior: Bayesian-Inspired Adaptive Prior Correction for Long-Tailed Continual Learning, *S Divakar Bhat, Amit Popat More, Mudit Soni, Bhuvan Aggarwal*
342 An Optimal Transport-driven Approach for Cultivating Latent Space in Online Incremental Learning, *Quyen Tran, Hai Nguyen, Quan Dao, Hoang Phan, Linh Van, Khoat Than, Dinh Phung, Dimitris Metaxas, Trung Le*
343 HAD: Heterogeneity-Aware Distillation for Lifelong Heterogeneous Learning, *Xuerui Zhang, Xuehao Wang, Zhan Zhuang, Linglan Zhao, Ziyue Li, Xinmin Zhang, Zhihuan Song, Yu Zhang*
344 U-Mind: A Unified Framework for Real-Time Multimodal Interaction with Audiovisual Generation, *Xiang Deng, Feng Gao, Yong Zhang, Youxin Pang, Xu Xiaoming, Zhuoliang Kang, Xiaoming Wei, Yebin Liu*
345 StreamAvatar: Streaming Diffusion Models for Real-Time Interactive Human Avatars, *Zhiyao Sun, Ziqiao Peng, Yifeng Ma, Yi Chen, Zhengguang Zhou, Zixiang Zhou, Guozhen Zhang, Youliang Zhang, Yuan Zhou, Qinglin Lu, Yong-Jin Liu*
346 FlashLips: 100-FPS Mask-Free Latent Lip-Sync using Reconstruction Instead of Diffusion or GANs, *Andreas Zinonos, Michał Stypułkowski, Antoni Bigata, Stavros Petridis, Maja Pantic, Nikita Drobyshev*
347 WildCap: Facial Albedo Capture in the Wild via Hybrid Inverse Rendering, *Yuxuan Han, Xin Ming, Tianxiao Li, Zhuofan Shen,*

Qixuan Zhang, Lan Xu, Feng Xu

348 EmoTaG: Emotion-Aware Talking Head Synthesis on Gaussian Splatting with Few-Shot Personalization, *Haolan Xu, Keli Cheng, Lei Wang, Ning Bi, Xiaoming Liu*

349 DyaDiT: A Multi-Modal Diffusion Transformer for Socially Favorable Dyadic Gesture Generation, *Yichen Peng, Jyun-Ting Song, Siyeol Jung, Ulsan National Institute of Science & Technology blank, Ruofan Liu, Haiyang Liu, Xuangeng Chu, Ruicong Liu, Erwin Wu, Hideki Koike, Kris Kitani*

350 TRM-VLA: Temporal-Aware Chain-of-Thought Reasoning and Memorization for Vision-Language-Action Models, *Xiang Li, Ya-Li Li, Yuan Wang, Shengjin Wang*

351 VGGDrive: Empowering Vision-Language Models with Cross-View Geometric Grounding for Autonomous Driving, *Jie Wang, Guang Li, Zhijian Huang, Chenxu Dang, Hangjun Ye, Yahong Han, Long Chen*

352 NoRD: A Data-Efficient Vision-Language-Action Model that Drives without Reasoning, *Ishaan Rawal, Shubh Gupta, Yihan Hu, Wei Zhan*

353 ❊ HTNav: A Hybrid Navigation Framework with Tiered Structure for Urban Aerial Vision-and-Language Navigation, *Chengjie Fan, Cong Pan, Zijian Liu, Ningzhong Liu, Jie Qin*

354 CycleBEV: Regularizing View Transformation Networks via View Cycle Consistency for Bird's-Eye-View Semantic Segmentation, *Jeongbin Hong, Dooseop Choi, Taeg-Hyun An, Kyounghwan An, Kyoung-Wook Min*

355 STAvatar: Soft Binding and Temporal Density Control for Monocular 3D Head Avatars Reconstruction, *Jiankuo Zhao, Xiangyu Zhu, Zidu Wang, Zhen Lei*

356 CrowdGaussian: Reconstructing High-Fidelity 3D Gaussians for Human Crowd from a Single Image, *Yizheng Song, Yiyu Zhuang, Qipeng Xu, Haixiang Wang, Jiahe Zhu, Jing Tian, Siyu Zhu, Hao Zhu*

357 OMG-Avatar: One-shot Multi-LOD Gaussian Head Avatar, *Jianqiang Ren, Lin Liu, Steven Hoi*

358 ❊ Globally Optimal Pose from Orthographic Silhouettes, *Agniva Sengupta, Dilara Kus, Jianning Li, Stefan Zachow*

359 AvatarPointillist: AutoRegressive 4D Gaussian Avatarization, *Hongyu Liu, Xuan Wang, Zijian Wu, Yating Wang, Ziyu Wan, Yue Ma, Runtao Liu, Boyao Zhou, Yujun Shen, Qifeng Chen*

360 COPO: Causal-Oriented Policy Optimization for Hallucinations of MLLMs, *Peizheng Guo, Jingyao Wang, Wenwen Qiang, Jiahuan Zhou, Changwen Zheng, Gang Hua*

361 ❊ Thinking in Uncertainty: Mitigating Hallucinations in MLRMs with Latent Entropy-Aware Decoding, *Zhongxing Xu, Zhonghua Wang, Zhe Qian, Dachuan Shi, Feilong Tang, Ming Hu, Shiyan Su, Xiaocheng Zou, Wei Feng, Dwarikanath Mahapatra, Yifan Peng, Minquan Lin, Zongyuan Ge*

362 AdaIAT: Adaptively Increasing Attention to Generated Text to Alleviate Hallucinations in LVLM, *Li'an Zhong, Ziqiang He, Jibin Zheng, Jin Li, Z. Jane Wang, Xiangui Kang*

363 HulluEdit: Single-Pass Evidence-Consistent Subspace Editing for Mitigating Hallucinations in Large Vision-Language Models, *Yangguang Lin, Quan Fang, Yufei Li, Jiachen Sun, Junyu Gao, Jitao Sang*

364 ❊ SEASON: Mitigating Temporal Hallucination in Video Large Language Models via Self-Diagnostic Contrastive Decoding, *Chang-Hsun Wu, Kai-Po Chang, Yu-Yang Sheng, Hung-Kai Chung, Kuei-Chun Wang, Yu-Chiang Frank Wang*

365 One Token, Two Fates: A Unified Framework via Vision Token Manipulation Against MLLMs Hallucination, *Zhan Fa, Yue Duan, Jian Zhang, Lei Qi, Yinghuan Shi*

366 EgoX: Egocentric Video Generation from a Single Exocentric Video, *Taewoong Kang, Kinam Kim, Dohyeon Kim, Minho Park, Junha Hyung, Jaegul Choo*

367 SymphoMotion: Joint Control of Camera Motion and Object Dynamics for Coherent Video Generation, *Guiyu Zhang, Yabo Chen, Xunzhi Xiang, Junchao Huang, Zhongyu Wang, Li Jiang*

368 Pantheon360: Taming Digital Twin Generation via 3D-Aware 360° Video Diffusion, *Ting-Hsuan Chen, Ying-Huan Chen, Tao Tu, Jie-Ying Lee, Cho-Ying Wu, Fangzhou Lin, Hengyuan Zhang, David Paz, Xinyu Huang, Yuliang Guo, Yu-Lun Liu, Yue Wang, Liu Ren*

369 SeeU: Seeing the Unseen World via 4D Dynamics-aware Generation, *Yu Yuan, Tharindu Wickremasinghe, Zeeshan Nadir, Xijun Wang, Yiheng Chi, Stanley H. Chan*

370 ReDirector: Creating Any-Length Video Retakes with Rotary Camera Encoding, *Byeongjun Park, Byung-Hoon Kim, Hyungjin Chung, Jong Chul Ye*

371 Scaling4D: Pushing the Frontier of Video Novel View Synthesis through Large-Scale Monocular Videos, *Hongrui Cai, Junjie Luo, Zhihong Fu, Shengnan Zhu, Jiawei Wen, Wanquan Feng, Songtao Zhao, Qian He*

372 PHANTOM: Physics-Infused Video Generation via Joint Modeling of Visual and Latent Physical Dynamics, *Ying Shen, Jerry Xiong, Tianjiao Yu, Ismini Lourentzou*

373 WorldReel: 4D Video Generation with Consistent Geometry and Motion Modeling, *Shaoheng Fang, Hanwen Jiang, Yunpeng Bai, Niloy J. Mitra, Qixing Huang*

374 Let Your Image Move with Your Motion! -- Implicit Multi-Object Multi-Motion Transfer, *Yuze Li, Dong Gong, Xiao Cao, Junchao Yuan, Dongsheng Li, Lei Zhou, Yun Sing Koh, Cheng Yan, Xinyu Zhang*

375 SpaceTimePilot: Generative Rendering of Dynamic Scenes Across Space and Time, *Zhening Huang, Hyeonho Jeong, Xuelin Chen, Yulia Gryaditskaya, Tuanfeng Y. Wang, Joan Lasenby, Chun-Hao Huang*

376 D2FANet: Enhancing Video Object Detection with Dual-Domain Feature Aggregation Network, *Qiang Qi, Wenqi Shang, Meifang Wang, Xiao Wang*

377 HierUQ: Hierarchical Uncertainty Quantification with Adaptive Granularity Reconciliation for Degraded Image Classification, *Yang Chu, Xiaomeng Yang, Keli Deng, Yuntao Qian*

378 ID-Sim: An Identity-Focused Similarity Metric, *Julia Chae, Nicholas Kolkin, Jui-Hsien Wang, Richard Zhang, Sara Beery, Cusuh Ham*

379 Hier-COS: Making Deep Features Hierarchy-aware via Composition of Orthogonal Subspaces, *Depanshu Sani, Saket Anand*

380 Towards Cross-Modal Preservation, Consistency and Alignment for Privacy-Preserving Visible-Infrared Person Re-Identification, *Yudi Xie, Zhongao Zhou, Bin Yang, Zhenghan Chen, Mang Ye*

381 Enhancing Mixture-of-Experts Specialization via Cluster-Aware Upcycling, *Sanghyeok Chu, Pyunghwan Ahn, Gwangmo Song, Seung Hwan Kim, Honglak Lee, Bohyung Han*

382 COPE: Consistent Occlusion and Prompt Enhancement Network for Occluded Person Re-identification, *Siyi Sun, Jinliang Lin, Juanjuan Weng, Zhihui Liu, Shaozi Li, Zhiming Luo*

383 Assignment-Driven Hash Learning in a Hyper-Semantic Space for On-the-Fly Category Discovery, *Kaibing Yang, Yucheng Wang, Tingzhang Luo*

384 DyFCLT: Dynamic Frequency-Decoupled Cross-Modal Learning Transformer for Multimodal Tiny Object Detection, *Chaolang Li, Pengwen Dai, Jingyu Li, Siyuan Yao, Yuchen Jiang, Zhuoran Zheng*

385 EW-DETR: Evolving World Object Detection via Incremental Low-Rank DEtection TRansformer, *Munish Monga, Vishal Chudasama, Pankaj Wasnik, C.V. Jawahar*

386 ❊ Building a Precise Video Language with Human–AI Oversight, *Zhiqiu Lin, Siyuan Cen, Chancharik Mitra, Isaac Li, Yuhan Huang, Yu Tong Tiffany Ling, Hewei Wang, Irene Pi, Shihang Zhu, Yili Han, Yilun Du, Deva Ramanan*

387 CoCoVideo: The High-Quality Commercial-Model-Based Contrastive Benchmark for AI-Generated Video Detection, *Huidong Feng, Wentao Chen, Jie Chen, Xinqi Cai, Ruolong Ma, Yinglin Zheng, Yuxin Lin, Ming Zeng*

388 Towards Sparse Video Understanding and Reasoning, *Chenwei Xu, Zhen Ye, Shang Wu, Weijian Li, Zihan Wang, Zhuofan Xia, Lie Lu, Pranav Maneriker, Fan Du, Manling Li, Han Liu*

389 Divide, then Ground: Adapting Frame Selection to Query Types for Long-Form Video Understanding, *Jialuo Li, Bin Li, Jiahao Li, Yan Lu*

390 MuKV: Multi-Grained KV Cache Compression for Long Streaming Video Question-Answering, *Junbin Xiao, Jiajun Chen, Tianxiang Sun, Xun Yang, Angela Yao*

391 ParallelVLM: Lossless Video-LLM Acceleration with Visual Alignment Aware Parallel Speculative Decoding, *Quan Kong, Yuhao Shen, Yicheng Ji, Huan Li, Cong Wang*

392 TiViBench: Benchmarking Think-in-Video Reasoning for Video Generation, *Harold Haodong Chen, Disen Lan, Wen-Jie Shu, Qingyang Liu, Zihan Wang, Sirui Chen, Wenkai Cheng, Kanghao Chen, Hongfei Zhang, Zixin Zhang, Rongjin Guo, Yu Cheng, Ying-Cong Chen*

393 What Are You Doing? A Closer Look at Controllable Human Video Generation, *Emanuele Bugliarello, Anurag Arnab, Roni Paiss, Christy Koh, Pieter-Jan Kindermans, Cordelia Schmid*
394 Score2Instruct: Scaling Up Video Quality-Centric Instructions via Automated Dimension Scoring, *Qizhi Xie, Kun Yuan, Yunpeng Qu, Jiachao Gong, Mingda Wu, Ming Sun, Chao Zhou, Jihong Zhu*
395 ❊ CFG-Ctrl: Control-Based Classifier-Free Diffusion Guidance, *Hanyang Wang, Yiyang Liu, Jiawei Chi, Fangfu Liu, Ran Xue, Yueqi Duan*
396 Towards Holistic Modeling for Video Frame Interpolation with Auto-regressive Diffusion Transformers, *Xinyu Peng, Han Li, Yuyang Huang, Ziyang Zheng, Yaoming Wang, Xin Chen, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong*
397 ❊ DDiT: Dynamic Patch Scheduling for Efficient Diffusion Transformers, *Dahye Kim, Deepti Ghadiyaram, Raghudeep Gadde*
398 ❊ Towards High-resolution and Disentangled Reference-based Sketch Colorization, *Dingkun Yan, Xinrui Wang, Ru Wang, Zhuoru Li, Jinze Yu, Yusuke Iwasawa, Yutaka Matsuo, Jiaxian Guo*
399 MakeAnything: Harnessing Diffusion Transformers for Multi-Domain Procedural Sequence Generation, *Yiren Song, Cheng Liu, Mike Zheng Shou*
400 Layer-wise Instance Binding for Regional and Occlusion Control in Text-to-Image Diffusion Transformers, *Ruidong Chen, Yancheng Bai, Xuanpu Zhang, Jianhao Zeng, Lanjun Wang, Dan Song, Lei Sun, Xiangxiang Chu, Anan Liu*
401 Memory-Efficient Fine-Tuning Diffusion Transformers via Dynamic Patch Sampling and Block Skipping, *Sunghyun Park, Jeongho Kim, Hyoungwoo Park, Debasmit Das, Sungrack Yun, Munawar Hayat, Jaegul Choo, Fatih Porikli, Seokeon Choi*
402 COT-FM: Cluster-wise Optimal Transport Flow Matching, *Chiensheng Chiang, Kuan-Hsun Tu, Jia-Wei Liao, Cheng-Fu Chou, Tsung-Wei Ke*
403 ❊ Interpretable Motion-Attentive Maps: Spatio-Temporally Localizing Concepts in Video Diffusion Transformers, *Youngjun Jun, Seil Kang, Woojung Han, Seong Jae Hwang*
404 ❊ Guiding a Diffusion Transformer with the Internal Dynamics of Itself, *Xingyu Zhou, Qifan Li, Xiaobin Hu, Hai Chen, Shuhang Gu*
405 CoopDiff: A Diffusion-Guided Approach for Cooperation under Corruptions, *Gong Chen, Chaokun Zhang, Pengcheng Lv*
406 ❊ RARE: Learn to RAnk and REtrieve for Monocular 3D Object Detection, *Hyeonjeong Park, Peixi Xiong, Xiaoqian Ruan, Dian Jia, Pei Yu, Wei Tang*
407 COG: Confidence-aware Optimal Geometric Correspondence for Unsupervised Single-reference Novel Object Pose Estimation, *Yuchen Che, Jingtu Wu, Hao Zheng, Asako Kanezaki*
408 Learnability-Driven Submodular Optimization for Active Roadside 3D Detection, *Ruiyu Mao, Baoming Zhang, Nicholas Ruozzi, Yunhui Guo*
409 Look Before You Fuse: 2D-Guided Cross-Modal Alignment for Robust 3D Detection, *Xiang Li, Zhangchi Hu, Xu Xiao, Bin Kong*
410 Long-SCOPE: Fully Sparse Long-Range Cooperative 3D Perception, *Jiahao Wang, Zikun Xu, Yuner Zhang, Zhongwei Jiang, Chenyang Lu, Shuocheng Yang, Yuxuan Wang, Jiaru Zhong, Chuang Zhang, Shaobing Xu, Jianqiang Wang*
411 Dynamics-Aware Preference Optimization for Vision-Language Models, *Jusheng Zhang, Kaitong Cai, Jing Yang, Jian Wang, Keze Wang*
412 ❊ Selection-as-Nonlinearity: Bridging Attention and Activation via a Joint Game–Decision Lens for Interpretable, Discriminative Visual Representations, *Sudong Cai, Shuai Yuan, Bingzhi Chen, Rui Mao, Bing Wang*
413 Learning What Helps: Task-Aligned Context Selection for Vision Tasks, *Jingyu Guo, Emir Konuk, Fredrik Strand, Christos Matsoukas, Kevin Smith*
414 Consensus Entropy: Harnessing Multi-VLM Agreement for Self-Verifying and Self-Improving OCR, *Yulong Zhang, Tianyi Liang, Erfei Cui, Guoqing Wang, Xu Guo, Chenhui Li, Gongshen Liu*
415 NeuroRule: Bridging Vision and Logic with Differentiable Rule Induction, *Muhammad Zarar, Mingzheng Zhang, Xiaowang Zhang, Zhiyong Feng*
416 Beyond Graph Model: Reliable VLM Fine-Tuning via Random Graph Adapter, *Bo Jiang, Xueyang Ze, Beibei Wang, Xixi Wang, Xixi Wan, Bin Luo*
417 Ego: Embedding-Guided Personalization of Vision-Language Models, *Soroush Seifi, Simon Gardier, Vaggelis Dorovatas, Daniel Olmeda Reino, Rahaf Aljundi*
418 JoPPO: Hierarchical Photography Assessment via Contrastive Joint Conditional Probabilistic Reinforcement Learning, *Yifan Yang, Juntuo Wang, Yuming Qiao, Xudong Zhang, Chunyang Yu, Yan Li, Xiao Lin, Liang Luo, Dan Meng*
419 AeroAgent: A Vision–Physics–Decision Framework for Aerodynamic Vehicle Design, *Ye Liu, Shouyi Liu, Huiyu Yang, Jianghang Gu, Wenhao Fan, Zhongxin Yang, Ding Wang, Simeng Chen, Zirun Jiang, Yuanwei Bin, Shiyi Chen, Yuntian Chen*
420 MiniCPM-V 4.5: Cooking Efficient MLLMs via Architecture, Data, and Training Recipe, *Tianyu Yu, Zefan Wang, Chongyi Wang, Fuwei Huang, Wenshuo Ma, Zhihui He, Tianchi Cai, Weize Chen, Yuxiang Huang, Ranchi Zhao, Bokai Xu, Junbo Cui, Yingjing Xu, Liqing Ruan, Luoyuan Zhang, Hanyu Liu, Jingkun Tang, Hongyuan Liu, Qining Guo, Wenhao Hu, Bingxiang He, Jie Zhou, Jie Cai, Ji Qi, Zonghao Guo, Chi Chen, Guoyang Zeng, Yuxuan Li, Ganqu Cui, Ning Ding, Xu Han, Yuan Yao, Zhiyuan Liu, Maosong Sun*
421 Prune Wisely, Reconstruct Sharply: Compact 3D Gaussian Splatting via Adaptive Pruning and Difference-of-Gaussian Primitives, *Haoran Wang, Guoxi Huang, Fan Zhang, David Bull, Nantheera Anantrasirichai*
422 MSCD-GS: Motion-Separated Cooperative Deblurring Dynamic Reconstruction via Gaussian Splatting, *Yongjian Liao, Xu Zou, Wenjun Chen, Huixuan Li, Xiaoen Xie, Chunxi Li, Shixiang Huang, Gang Zhang, Jiahuan Zhou, Sheng Zhong, Luxin Yan*
423 P2GS: Physical Prior-guided Gaussian Splatting for Photometrically Consistent Urban Reconstruction, *Kota Shimomura, Hidehisa Arai, Tsubasa Takahashi, Takayoshi Yamashita, Hironobu Fujiyoshi*
424 iSplat: Iterative Learning for Fine-Grained Gaussian Splatting, *Haifeng Wu, Wei Long, Shuhang Gu, Lixin Duan, Wen Li*
425 Off The Grid: Detection of Primitives for Feed-Forward 3D Gaussian Splatting, *Arthur Moreau, Richard Shaw, Michal Nazarczuk, Jisu Shin, Thomas Tanay, Zhensong Zhang, Songcen Xu, Eduardo Pérez-Pellitero*
426 MAPo: Motion-Aware Partitioning of Deformable 3D Gaussian Splatting for High-Fidelity Dynamic Scene Reconstruction, *Han Jiao, Jiakai Sun, Yexing Xu, Lei Zhao, Wei Xing, Huaizhong Lin*
427 FreeArtGS: Articulated Gaussian Splatting Under Free-moving Scenario, *Hang Dai, Hongwei Fan, Han Zhang, Duojin Wu, Jiyao Zhang, Hao Dong*
428 HeroGS: Hierarchical Guidance for Robust 3D Gaussian Splatting under Sparse Views, *Jiashu Li, Xumeng Han, Zhaoyang Wei, Zipeng Wang, Kuiran Wang, Guorong Li, Zhenjun Han, Jianbin Jiao*
429 SharpTimeGS: Sharp and Stable Dynamic Gaussian Splatting via Lifespan Modulation, *Zhanfeng Liao, Jiajun Zhang, Hanzhang Tu, Zhixi Wang, Yunqi Gao, Hongwen Zhang, Yebin Liu*
430 Physically Inspired Gaussian Splatting for HDR Novel View Synthesis, *Huimin Zeng, Yue Bai, Hailing Wang, Yun Fu*
431 PhysIR-Splat: Physically Consistent Thermal Infrared Radiative Transfer in 3D Gaussian Splatting, *Jingyuan Gao, Yumeng Hu, Fei Gao, Mingjin Zhang*
432 4C4D: 4 Camera 4D Gaussian Splatting, *Junsheng Zhou, Zhifan Yang, Liang Han, Wenyuan Zhang, Kanle Shi, Shenkun Xu, Yu-Shen Liu*
433 SplatSuRe: Selective Super-Resolution for Multi-view Consistent 3D Gaussian Splatting, *Pranav Asthana, Alex Hanson, Allen Tu, Tom Goldstein, Matthias Zwicker, Amitabh Varshney*
434 GaussianZoom: Progressive Zoom-in Generative 3D Gaussian Splatting with Geometric and Semantic Guidance, *Jiale Shi, Jiarui Hu, Zesong Yang, Kaixuan Luan, Hujun Bao, Zhaopeng Cui*
435 MotionScale: Reconstructing Appearance, Geometry, and Motion of Dynamic Scenes with Scalable 4D Gaussian Splatting, *Haoran Zhou, Gim Hee Lee*
436 PRIMU: Uncertainty Estimation for Novel Views in Gaussian Splatting from Primitive-Based Representations of Error and Coverage, *Thomas Gottwald, Edgar Heinert, Peter Stehr, Chamuditha Jayanga Galappaththige, Matthias Rottmann*
437 TGSFormer: Scalable Temporal Gaussian Splatting for Embodied Semantic Scene Completion, *Rui Qian, Haozhi Cao, Tianchen Deng, Tianxin Hu, Weixiang Guo, Shenghai Yuan, Lihua Xie*
438 Disco-GS: Gaussian Splatting in Dynamic Color Lighting, *Ashish Kumar, A. N. Rajagopalan*
439 ❊ ReAG: Reasoning-Augmented Generation for Knowledge-based Visual Question Answering, *Alberto Compagnoni, Marco Morini, Sara Sarto, Federico Cocchi, Davide Caffagni, Marcella Cornia,*

Lorenzo Baraldi, Rita Cucchiara

440 GuardTrace-VL: Detecting Unsafe Multimodel Reasoning via Iterative Safety Supervision, *Yuxiao Xiang, Junchi Chen, Zhenchao Jin, Changtao Miao, Haojie Yuan, Qi Chu, Tao Gong, Nenghai Yu*

441 ❊ AdaptVision: Efficient Vision-Language Models via Adaptive Visual Acquisition, *Zichuan Lin, Yicheng Liu, Yang Yang, Lvfang Tao, Deheng Ye*

442 See It, Say It, Sorted: An Iterative Training-Free Framework for Visually-Grounded Multimodal Reasoning in LVLMs, *Yongchang Zhang, Oliver Ma, Tianyi Liu, Guangquan Zhou, Yang Chen*

443 Will Multimodal Models Be Dazzled by Multi-Image Visual Puzzles?, *Zhi Zhu, YaoQi Fan, Zhe Chen, Yue Cao, Yangzhou Liu, Tong Lu*

444 GThinker: Towards General Multimodal Reasoning via Cue-Guided Rethinking, *Yufei Zhan, Ziheng Wu, Yousong Zhu, Rongkun Xue, Guanghao Zhou, Ruipu Luo, Zhenghao Chen, Can Zhang, Yifan Li, Zhentao He, Zheming Yang, Ming Tang, Minghui Qiu, Jinqiao Wang*

445 Visual Grounding for Object Questions, *Martin Nicolas Everaert, Xiruo Liu, Hiroyuki Takeda, Raja Bala, Vivek Yadav, Vidya Narayanan*

446 CARE What Fails: Contrastive Anchored-REflection for Verifiable Multimodal Reasoning, *Yongxin Wang, Zhicheng Yang, Meng Cao, Mingfei Han, Haokun Lin, Yingying Zhu, Xiaojun Chang, Xiaodan Liang*

447 ❊ What Do Visual Tokens Really Encode? Uncovering Sparsity and Redundancy in Multimodal Large Language Models, *Yingqi Fan, Junlong Tong, Anhao Zhao, Xiaoyu Shen*

448 Think-as-You-See: Streaming Chain-of-Thought Reasoning for Large Vision-Language Models, *Jialiang Zhang, Junlong Tong, Junyan Lin, Hao Wu, Yirong Sun, Yunpu Ma, Xiaoyu Shen*

449 Stable and Efficient Single-Rollout RL for Multimodal Reasoning, *Rui Liu, Dian Yu, Lei Ke, Haolin Liu, Yujun Zhou, Zhenwen Liang, Haitao Mi, Pratap Tokekar, Dong Yu*

450 Revisiting the Necessity of Lengthy Chain-of-Thought in Vision-centric Reasoning Generalization, *Yifan Du, Kun Zhou, Yingqian Min, Yue Ling, Wayne Xin Zhao, Youbin Wu, Ji-Rong Wen*

451 Monet: Reasoning in Latent Visual Space Beyond Image and Language, *Qixun Wang, Yang Shi, Yifei Wang, Yuanxing Zhang, Pengfei Wan, Kun Gai, Xianghua Ying, Yisen Wang*

452 STAR-R1: Multi-View Spatial TrAnsformation Reasoning by Reinforcing Multimodal LLMs, *Zongzhao Li, Zongyang Ma, Mingze Li, Songyou Li, Yu Rong, Tingyang Xu, Ziqi Zhang, Deli Zhao, Wenbing Huang*

453 From Where Things Are to What They Are For: Benchmarking Spatial–Functional Intelligence in Multimodal LLMs, *Le Zhang, Jihan Yang, Soundarya Krishnan, Jimit Majmudar, Xiou Ge, Prasoon Puri, Prathamesh Saraf, Shruti Bhargava, Dhivya Piraviperumal, Yinan Ling, Cindy Pan, Hong Yu, Aishwarya Agrawal, Bo-Hsiang Tseng*

454 Deeper Thought, Weaker Aim: Understanding and Mitigating Perceptual Impairment during Reasoning in Multimodal Large Language Models, *Ruiying Peng, Xueyu Wu, Jing Lei, Lu Hou, Yuanzheng Ma, Xiao-Hui Li*

455 S2D: Selective Spectral Decay for Quantization-Friendly Conditioning of Neural Activations, *Arnav Chavan, Nahush Lele, Udbhav Bamba, Sankalp Dayal, Aditi Raghunathan, Deepak Gupta*

456 OneSparse: A Unified Framework for Sparse Activation Layers in Vision Models, *Xingkui Zhu, Dingkang Liang, Cheng Chen, Daoxin Zhang, Iv Hanxiang, Zhe Xu, Yao Hu, Xiang Bai*

457 ❊ What Matters in Practical Learned Image Compression, *Kedar Tatwawadi, Parisa Rahimzadeh, Zhanghao Sun, Zhiqi Chen, Ziyun Yang, Sanjay Nair, Divija Hasteer, Oren Rippel*

458 BinaryAttention: One-Bit QK-Attention for Vision and Diffusion Transformers, *Chaodong Xiao, Zhengqiang Zhang, Lei Zhang*

459 Ultra-Low Bitrate Perceptual Image Compression with Shallow Encoder, *Tianyu Zhang, Dong Liu, Chang Wen Chen*

460 LazyVAR: Accelerating Visual Autoregressive Models via Scale-wise Token Pruning and Parallel Group Decoding, *Rongge Mao, Chengqi Dong, S Kevin Zhou*

461 Spk2VidNet: A Hierarchical Recurrent Architecture for High-Fidelity Video Reconstruction from Long Spike-Camera Streams, *Yuanlin Wang, Ruiqin Xiong, Jiyu Xie, Zhenkun Zhu, Zhaofei Yu, Xiaopeng Fan, Tiejun Huang*

462 Adaptive Learned Image Compression with Graph Neural Networks, *Yunuo Chen, Bing He, Zezheng Lyu, Hongwei Hu, Qunshan Gu, Yuan Tian, Guo Lu*

463 ❊ SGI: Structured 2D Gaussians for Efficient and Compact Large Image Representation, *Zixuan Pan, Kaiyuan Tang, Jun Xia, Yifan Qin, Lin Gu, Chaoli Wang, Jianxu Chen, Yiyu Shi*

464 VVS: Accelerating Speculative Decoding for Visual Autoregressive Generation via Partial Verification Skipping, *Haotian Dong, Ye Li, Rongwei Lu, Chen Tang, Shu-Tao Xia, Zhi Wang*

465 HypeVPR: Exploring Hyperbolic Space for Perspective to Equirectangular Visual Place Recognition, *Suhan Woo, Seongwon Lee, Jinwoo Jang, Euntai Kim*

466 LoD-Loc v3: Generalized Aerial Localization in Dense Cities using Instance Silhouette Alignment, *Shuaibang Peng, Juelin Zhu, Xia Li, Kun Yang, Yu Liu, Maojun Zhang, Shen Yan*

467 ❊ CoLoR: The Devil is in Scene Coordinate Regression for Large-Scale Visual Localization, *Xindong Mao, Hang Li, Yuchen Wu, Jiahe Li, Xiao Bai, Jin Zheng*

468 ❊ Affine Perspective-Three-Point Problem, *Gaku Nakano*

469 Sky2Ground: A Benchmark for Site Modeling under Varying Altitude, *Zengyan Wang, Sirshapan Mitra, Rajat Modi, Hui Lim, Yogesh Rawat*

470 SemanticVLA: Towards Semantic Reasoning over Action Memorization via Synergistic Explicit Trace and Latent Action Planning, *Fei Ni, Zhuo Chen, Yifu Yuan, Zibin Dong, Xianze Yao, Shan Luo, Jianye Hao, Jiankang Deng, Stefanos Zafeiriou*

471 WebGym: Scaling Training Environments for Long-Horizon Visual Web Agents with Realistic Tasks, *Hao Bai, Alexey Taymanov, Tong Zhang, Aviral Kumar, Spencer Whitehead*

472 Beyond Perceptual Shortcuts: Causal-Inspired Debiasing Optimization for Generalizable Video Reasoning in Lightweight MLLMs, *Jingze Wu, Quan Zhang, Hongfei Suo, Zeqiang Cai, Hongbo Chen*

473 APPO: Attention-guided Perception Policy Optimization for Video Reasoning, *Henghui Du, Chang Zhou, Xi Chen, Di Hu*

474 RetouchIQ: MLLM Agents for Instruction-Based Image Retouching with Generalist Reward, *Qiucheng Wu, Jing Shi, Simon Jenni, Kushal Kafle, Tianyu Wang, Shiyu Chang, Handong Zhao*

475 EVA: Efficient Reinforcement Learning for End-to-End Video Agent, *Yaolun Zhang, Ruohui Wang, Jiahao Wang, Yepeng Tang, Xuanyu Zheng, Haonan Duan, Hao Lu, Hanming Deng, Lewei Lu*

476 ❊ Visual Document Understanding and Reasoning: A Multi-Agent Collaboration Framework with Agent-Wise Adaptive Test-Time Scaling, *Xinlei Yu, Chengming Xu, Zhangquan Chen, Yudong Zhang, Shilin Lu, Cheng Yang, Jiangning Zhang, Shuicheng Yan, Xiaobin Hu*

477 GazeOnce360: Fisheye-Based 360° Multi-Person Gaze Estimation with Global–Local Feature Fusion, *Zhuojiang Cai, Zhenghui Sun, Feng Lu*

478 Bridging Human Evaluation to Infrared and Visible Image Fusion, *Jinyuan Liu, Xingyuan Li, Qingyun Mei, Haoyuan Xu, Zhiying Jiang, Long Ma, Risheng Liu, Xin Fan*

479 ❊ Beyond Strict Pairing: Arbitrarily Paired Training for High-Performance Infrared and Visible Image Fusion, *Yanglin Deng, Tianyang Xu, Chunyang Cheng, Hui Li, Xiaojun Wu, Josef Kittler*

480 ❊ Semantic-Adaptive Diffusion for Dynamic Spatiotemporal Fusion, *Jinsong Zhang, Ying Qu, Yuan Liao, Hairong Qi, Zhenzhou Shao*

481 Bayesian Decomposition and Semantic Completion for Few-shot Semantic Segmentation, *Guangchen Shi, Yirui Wu, Wei Zhu, Tao Wang, Hao Zhang, Bo Li, Tong Lu*

482 From Few-way to Many-way: Rethinking Few-shot Fine-grained Image Classification, *Li-Jun Zhao, Zhen-Duo Chen, Xin Luo, Xin-Shun Xu*

483 STiTch: Semantic Transition and Transportation in Collaboration for Training-Free Zero-Shot Composed Image Retrieval, *Miaoge Li, Dongsheng Wang, Zening Sun, Jinsen Zhang, Wenhan Luo, Jingcai Guo*

484 Selective, Regularized, and Calibrated: Harnessing Vision Foundation Models for Cross-Domain Few-Shot Semantic Segmentation, *Junyuan Ma, Xunzhi Xiang, Wenbin Li, Qi Fan, Yang Gao*

485 FlowComposer: Composable Flows for Compositional Zero-Shot Learning, *Zhenqi He, Lin Li, Long Chen*

486 ManifoldGD: Training-Free Hierarchical Manifold Guidance for Diffusion-Based Dataset Distillation, *Ayush Roy, Wei-Yang Alex Lee, Rudrasis Chakraborty, Vishnu Suresh Lokhande*

487 DMGD: Train-Free Dataset Distillation with Semantic-Distribution Matching in Diffusion Models, *Qichao Wang, Yunhong Lu, Hengyuan Cao, Junyi Zhang, Min Zhang*

488 UniRain: Unified Image Deraining with RAG-based Dataset Distillation and Multi-objective Reweighted Optimization, *Qianfeng Yang, Qiyuan Guan, Xiang Chen, Jiyu Jin, Guiyue Jin, Jiangxin Dong*

489 Leveraging Multispectral Sensors for Color Correction in Mobile Cameras, *Luca Cogo, Marco Buzzelli, Simone Bianco, Javier Vazquez-Corral, Raimondo Schettini*

490 Differentiable Adaptive 4D Structured Illumination for Joint Capture of Shape and Reflectance, *Huakeng Ding, Yaowen Chen, Kun Zhou, Hongzhi Wu*

491 Optical Diffraction-based Convolution for Semiconductor Lithography, *Young-Han Son, Dong-Hee Shin, Deok-Joong Lee, Hyun Jung Lee, Tae-Eui Kam*

492 GSNR: Graph Smooth Null-Space Representation for Inverse Problems, *Romario Gualdrón-Hurtado, Roman Jacome, Rafael S. Suárez, Henry Arguello*

493 MatE: Material Extraction from Single-Image via Geometric Prior, *Zeyu Zhang, Wei Zhai, Jian Yang, Yang Cao*

494 αMatte4K & μMatting: Dataset and Model for Ultra-Micro Precision Alpha Video Matting, *Xinyi Chen, Hang Dong, Baowei Jiang, Shenkun Xu, Youqi Guan, Kanle Shi, Kun Gai, Haichuan Song*

495 Revisiting Optimal Coding for I-ToF under Practical Sensor Constraints, *Wenbin Luo, Takafumi Iwaguchi, Ryusuke Sagawa, Hiroshi Kawasaki*

496 ❊ Dynamic Black-hole Emission Tomography with Physics-informed Neural Fields, *Berthy T. Feng, Andrew A. Chael, David Bromley, Aviad Levis, William T. Freeman, Katherine L. Bouman*

497 Exploring Spatiotemporal Feature Propagation for Video-Level Compressive Spectral Reconstruction: Dataset, Model and Benchmark, *Lijing Cai, Zhan Shi, Chenglong Huang, Jinyao Wu, Qiping Li, Zikang Huo, Linsen Chen, Chongde Zi, Xun Cao*

498 Generalizable Radio-Frequency Radiance Fields for Spatial Spectrum Synthesis, *Kang Yang, Yuning Chen, Wan Du*

499 SAR2Net: Learning Spatially Anchored Representations for Retrieval-Guided Cross-Stain Alignment, *Tianle Shen, Fang Yan, Xiaofan Zhang*

500 Advancing Cancer Prognosis with Hierarchical Fusion of Genomic, Proteomic and Pathology Imaging Data from a Systems Biology Perspective, *Junjie Zhou, Bao Xue, Meiling Wang, Wei Shao, Daoqiang Zhang*

501 PromptStereo: Zero-Shot Stereo Matching via Structure and Motion Prompts, *Xianqi Wang, Hao Yang, Hangtian Wang, Junda Cheng, Gangwei Xu, Min Lin, Xin Yang*

502 Any Resolution Any Geometry: From Multi-View To Multi-Patch, *Wenqing Cui, Zhenyu Li, Mykola Lavreniuk, Jian Shi, Ramzi Idoughi, Xiangjun Tang, Peter Wonka*

503 Paparazzo: Active Mapping of Moving 3D Objects, *Davide Allegro, Shiyao Li, Stefano Ghidoni, Vincent Lepetit*

504 DepthFocus: Controllable Depth Estimation for See-Through Scenes, *Junhong Min, Jimin Kim, Minwook Kim, Cheol-Hui Min, Youngpil Jeon, Minyong Choi*

505 ❊ OVI-MAP: Open-Vocabulary Instance-Semantic Mapping, *Zilong Deng, Federico Tombari, Marc Pollefeys, Johanna Wald, Daniel Barath*

506 PTC-Depth: Pose-Refined Monocular Depth Estimation with Temporal Consistency, *Leezy Han, Seunggyu Kim, Dongseok Shim, Hyeonbeom Lee*

507 ❊ SceneScribe-1M: A Large-Scale Video Dataset with Comprehensive Geometric and Semantic Annotations, *Yunnan Wang, Kecheng Zheng, Jianyuan Wang, Minghao Chen, David Novotny, Christian Rupprecht, Yinghao Xu, Xing Zhu, Wenjun Zeng, Xin Jin, Yujun Shen*

508 ❊ Omni-3DEdit: Generalized Versatile 3D Editing in One-Pass, *Liyi Chen, Pengfei Wang, Guowen Zhang, Zhiyuan Ma, Lei Zhang*

509 Ani3DHuman: Photorealistic 3D Human Animation with Self-guided Stochastic Sampling, *Qi Sun, Can Wang, Jiaxiang Shang, Yingchun Liu, Jing Liao*

510 Variational Graph-based Normal Integration, *Lixiong Chen, Bohan Yu, Victor Adrian Prisacariu, Imari Sato*

511 Vinedresser3D: Towards Agentic Text-guided 3D Editing, *Yankuan Chi, Xiang Li, Zixuan Huang, James Matthew Rehg*

512 MV2UV: Generating High-quality UV Texture Maps with Multiview Prompts, *Zheng Zhang, Qinchuan Zhang, Yuteng Ye, Zhi Chen, Penglei Ji, Mengfei Li, Wenxiao Zhang, Yuan Liu*

513 ❊ Learning Hierarchical Hyperbolic Mixture Model for Part-aware 3D Generation, *Qitong Yang, Mingtao Feng, Zijie Wu, Huixin Zhu, Weisheng Dong, Yaonan Wang, Ajmal Mian*

514 ❊ MeshRipple: Structured Autoregressive Generation of Artist-Meshes, *Junkai Lin, Hang Long, Huipeng Guo, Jielei Zhang, Jiayi Yang, Tianle Guo, Yang Yang, Jianwen Li, Wenxiao ZHANG, Matthias Nießner, Wei Yang*

515 ❊ FACE: A Face-based Autoregressive Representation for High-Fidelity and Efficient Mesh Generation, *Hanxiao Wang, Yuan-Chen Guo, Ying-Tian Liu, Zi-Xin Zou, Biao Zhang, Weize Quan, Ding Liang, Yan-Pei Cao, Dong-Ming Yan*

516 Easy3E: Feed-Forward 3D Asset Editing via Rectified Voxel Flow, *Shimin Hu, Yuanyi Wei, Fei Zha, Yudong Guo, Juyong Zhang*

517 ❊ CUPID: Generative 3D Reconstruction via Joint Object and Pose Modeling, Binbin Huang, Haobin Duan, Yiqun Zhao, Zibo Zhao, Yi Ma, Shenghua Gao

518 3D-Fixer: Coarse-to-Fine In-place Completion for 3D Scenes from a Single Image, *Ze-Xin Yin, Liu Liu, Xinjie Wang, Wei Sui, Zhizhong Su, Jian Yang, Jin Xie*

519 DRM: Diffusion-based Reward Model With Step-wise Guidance, *Jaxon Zhang, Binxin Yang, Hubery Yin, Chen Li, Jing LYU*

520 Taming Preference Mode Collapse via Directional Decoupling Alignment in Diffusion Reinforcement Learning, *Chubin Chen, Sujie Hu, Jiashu Zhu, Meiqi Wu, Jintao Chen, Yanxun Li, Nisha Huang, Chengyu Fang, Jiahong Wu, Xiangxiang Chu, Xiu Li*

521 VA-π: Variational Policy Alignment for Pixel-Aware Autoregressive Generation, *Xinyao Liao, Qiyuan He, Kai Xu, Xiaoye Qu, Yicong Li, Wei Wei, Angela Yao*

522 SoliReward: Mitigating Susceptibility to Reward Hacking and Annotation Noise in Video Generation Reward Models, *Jiesong Lian, Ruizhe Zhong, Zixiang Zhou, Xiaoyue Mi, Long Hu, Yuan Zhou, Qinglin Lu, Yixue Hao, Junchi Yan*

523 AnyID: Ultra-Fidelity Universal Identity-Preserving Video Generation from Any Visual References, *Jiahao Wang, Hualian Sheng, Sijia Cai, Yuxiao Yang, Weizhan Zhang, Caixia Yan, Bing Deng, Jieping Ye*

524 Style-GRPO: Semantic-Aware Preference Optimization for Image Style Transfer Guided by Reward Modeling, *Jianbin Zhao, Chaoran Feng, Miao Yu, Yingtao Li, Zhenyu Tang, Wangbo Yu, Yian Zhao, Xiaomin Li, Li Yuan, Yonghong Tian*

525 LAMP: Language-Assisted Motion Planning for Controllable Video Generation, *Muhammed Burak Kizil, Enes Sanli, Niloy J. Mitra, Erkut Erdem, Aykut Erdem, Duygu Ceylan*

526 Diverse Video Generation with Determinantal Point Process-Guided Policy Optimization, *Tahira Kazimi, Connor Dunlop, Pinar Yanardag*

527 Spectral Scalpel: Amplifying Adjacent Action Discrepancy via Frequency-Selective Filtering for Skeleton-Based Action Segmentation, *Haoyu Ji, Bowen Chen, Zhihao Yang, Wenze Huang, Yu Gao, Xueting Liu, Weihong Ren, Zhiyong Wang, Honghai Liu*

528 DETACH : Decomposed Spatio-Temporal Alignment for Exocentric Video and Ambient Sensors with Staged Learning, *Junho Yoon, Jaemo Jeong, Hyunju Kim, Dongman Lee*

529 Learning a Unified Latent Action Space from Videos with Action-centric Cycle Consistency, *Guangyan Chen, Qi Shao, Te Cui, Zichen Zhou, Weixin Mao, Luojie Yang, Meiling Wang, Yi Yang, Hua Chen, Yufeng Yue*

530 ❊ VideoNet: A Large-Scale Dataset for Domain-Specific Action Recognition, *Tanush Yadav, Mohammadreza Salehi, Jae Sung Park, Vivek Ramanujan, Hannaneh Hajishirzi, Yejin Choi, Ali Farhadi, Rohun Tripathi, Ranjay Krishna*

531 BD-Merging: Bias-Aware Dynamic Model Merging with Evidence-Guided Contrastive Learning, *Yuhan Xie, Chen Lyu*

532 Dynamic Momentum Recalibration in Online Gradient Learning, *Zhipeng Yao, Rui Yu, Guisong Chang, Ying Li, Yu Zhang, Dazhou Li*

533 ❊ Spherical Leech Quantization for Visual Tokenization and Generation, *Yue Zhao, Hanwen Jiang, Zhenlin Xu, Chutong Yang, Ehsan Adeli, Philipp Kraehenbuehl*

534 ❊ MSPT: Efficient Large-Scale Physical Modeling via Parallelized Multi-Scale Attention, *Pedro M. P. Curvo, Jan-Willem van de Meent, Maksim Zhdanov*

535 GR-Gauge: Cost-efficient Training Configuration By Gauging the Gradient Redundancy, *Guanjie Wang, Chen Chen*

536 E^2-SCI: Elastic Edge–Cloud Speculative Decoding via Credit Inertia, *Senyao Li, Haozhao Wang, Zhaobai Jiang, Zhanbo Jin, Hao Fan, Ruixuan Li*

537 HyperNAS: Enhancing Architecture Representation for NAS Predictor via Hypernetwork, *Jindi Lv, Yuhao Zhou, Yuxin Tian, Qing Ye, Wentao Feng, Jiancheng Lv*
538 NeuroFlow: Toward Unified Visual Encoding and Decoding from Neural Activity, *Weijian Mai, Mu Nan, Yu Zhu, Jiahang Cao, Rui Zhang, Yuqin Dai, Chunfeng Song, Andrew Luo, Jiamin Wu*
539 ❊ Spectral Conformal Risk Control: Distribution-Free Tail Guarantees via Bayesian Quadrature, *Mohammad Mahdi Kazemi Esfeh, Qi Yan, Yongxing Zhang, Zahra Gholami, Renjie Liao, Purang Abolmaesumi*
540 Edge-RecViT: Efficient Vision Transformer via Semantic-Refined Dynamic Recursion, *YiZhou Li, Jinyi Xu, Mingyu Yin, Xianyi Zhao*
541 ERMoE: Eigen-Reparameterized Mixture-of-Experts for Stable Routing and Interpretable Specialization, *Anzhe Cheng, Shukai Duan, Shixuan Li, Chenzhong Yin, Mingxi Cheng, Heng Ping, Tamoghna Chattopadhyay, Sophia I. Thomopoulos, Shahin Nazarian, Paul Thompson, Paul Bogdan*
542 GUI-SAGE: Enhancing GUI Automation with Self-Explanatory Learning, *Fei Tang, Zhangxuan Gu, Zhengxi Lu, Shangzhan Zhang, Zhengwen Zeng, Shuheng Shen, Changhua Meng, Yuchen Yan, Wenqi Zhang, Yongliang Shen, Weiming Lu, Yueting Zhuang*
543 GUIDE: A Benchmark for Understanding and Assisting Users in Open-Ended GUI Tasks, *Saelyne Yang, Jaesang Yu, Yi-Hao Peng, Kevin Qinghong Lin, Jae Won Cho, Yale Song, Juho Kim*
544 HiconAgent: History Context-aware Policy Optimization for GUI Agents, *Xurui Zhou, Gongwei Chen, Yuquan Xie, Zaijing Li, Kaiwen Zhou, Shuai Wang, Shuo Yang, Zhuotao Tian, Rui Shao*
545 ❊ PET-DINO: Unifying Visual Cues into Grounding DINO with Prompt-Enriched Training, *Weifu Fu, Jinyang Li, Bin-Bin Gao, Jialin Li, Yuhuan Lin, Hanqiu Deng, Wenbing Tao, Yong Liu, Chengjie Wang*
546 SDDF: Specificity-Driven Dynamic Focusing for Open-Vocabulary Camouflaged Object Detection, *Jiaming Liang, Yifeng Zhan, Chunlin Liu, Weihua Zheng, Bingye Peng, Qiwei Liang, Boyang Cai, Xiaochun Mai, Qiang Nie*
547 Towards Open-Vocabulary Industrial Defect Understanding with a Large-Scale Multimodal Dataset, *Tsai-Ching Ni, Cheng-Chi Chen, Yuan-Fu Yang*
548 Common Inpainted Objects In-N-Out of Context, *Tianze Yang, Tyson Jordan, Ruitong Sun, Ninghao Liu, Jin Sun*
549 Prompt-Free Universal Region Proposal Network, *Qihong Tang, Changhan Liu, Shaofeng Zhang, Wenbin Li, Qi Fan, Yang Gao*
550 Rewis3d: Reconstruction Improves Weakly-Supervised Semantic Segmentation, *Jonas Ernst, Wolfgang Boettcher, Lukas Hoyer, Jan Eric Lenssen, Bernt Schiele*
551 PaNDaS: Learnable Shape Interpolation Modeling with Localized Control, *Thomas Besnier, Emery Pierson, Sylvain Arguillere, Maks Ovsjanikov, Mohamed Daoudi*
552 Hilbert Curve-Based Attention Enabling Topology-Preserving Image Tensor Representation for Semantic Segmentation Network, *Linkang Xu, Gang Li, Yue Song, Xiangxin Ji*
553 Towards High-Quality Image Segmentation: Improving Topology Accuracy by Penalizing Neighbor Pixels, *Juan Miguel Valverde, Dim P. Papadopoulos, Rasmus Larsen, Anders Bjorholm Dahl*
554 SAGE: Style-Adaptive Generalization for Privacy-Constrained Semantic Segmentation Across Domains, *Qingmei Li, Yang Zhang, Peifeng Zhang, Haohuan Fu, Juepeng Zheng*
555 ❊ Better than Average: Spatially-Aware Aggregation of Segmentation Uncertainty Improves Downstream Performance, *Vanessa Emanuela Guarino, Claudia Winklmayr, Jannik Franzen, Josef Lorenz Rumberger, Manuel Pfeuffer, Sonja Greven, Klaus Maier-Hein, Dagmar Kainmueller, Christoph Karg, Carsten T. Lüth*
556 Universal 3D Shape Matching via Coarse-to-Fine Language Guidance, *Qinfeng Xiao, Guofeng Mei, Bo Yang, Liying Zhang, Jian Zhang, Kit-lun Yick*
557 Direct Segmentation without Logits Optimization for Training-Free Open-Vocabulary Semantic Segmentation, *Jiahao Li, Yang Lu, Yachao Zhang, Fangyong Wang, Yuan Xie, Yanyun Qu*
558 CDICS: Delving Into Fine-Grained Attribute for In-Context Segmentation via Compositional Prompts and Phased Decoupling, *Zhiyu Li, Dianmo Sheng, Qi Chu, Shilong Chen, Tao Gong, Zhou Wei, Nenghai Yu*
559 Discriminative Perception via Anchored Description for Reasoning Segmentation, *Tao Yang, Qing Zhou, Yanliang Li, Qi Wang*
560 SegEarth-R2: Towards Comprehensive Language-guided Segmentation for Remote Sensing Images, *Zepeng Xin, Kaiyu Li, Luodi Chen, Wanchen Li, Xiao Yuchen, Hui Qiao, Weizhan Zhang, Deyu Meng, Xiangyong Cao*
561 Cross-Scale Pansharpening via ScaleFormer and the PanScale Benchmark, *Ke Cao, Xuanhua He, Xueheng Li, Lingting Zhu, Yingying Wang, Ao Ma, Zhanjie Zhang, Man Zhou, Chengjun Xie, Jie Zhang*
562 CrossEarth-Gate: Fisher-Guided Adaptive Tuning Engine for Efficient Adaptation of Cross-Domain Remote Sensing Semantic Segmentation, *Shilei Cao, Ziyang Gong, Hehai Lin, Yang Liu, Jiashun Cheng, Xiaoxing Hu, Haoyuan Liang, Guowen Li, Chengwei Qin, Hong Cheng, Xue Yang, Juepeng Zheng, Haohuan Fu*
563 Multigrain-aware Semantic Prototype Scanning and Tri-Token Prompt
❊ Learning Embraced High-Order RWKV for Pan-Sharpening, *Junfeng Li, Wenyang Zhou, Xueheng Li, Xuanhua He, Jianhou Gan, Wenqi Ren*
564 ACPV-Net: All-Class Polygonal Vectorization for Seamless Vector Map Generation from Aerial Imagery, *Weiqin Jiao, Hao Cheng, George Vosselman, Claudio Persello*
565 Beyond Endpoints: Path-Centric Reasoning for Vectorized Off-Road Network Extraction, *Wenfei Guan, Jilin Mei, Tong Shen, Xumin Wu, Shuo Wang, Chen Min, Yu Hu*
566 Rotation Invariant and Symmetry Aware Pixel Difference Network for Remote Sensing Object Detection, *Jialei Zhan, Li Liu, Jiehua Zhang, Yuhang Xie, Yongxiang Liu, Jiangming Chen, Ming-Ming Cheng*
567 F2Net: A Frequency-Fused Network for Ultra-High Resolution Remote Sensing Segmentation, *Hengzhi Chen, Liqian Feng, Wenhua Wu, Xiaogang Zhu, Qiuxia Wu, Lianlei Shan, Kun Hu*
568 RoadGIE: Towards A Global-Scale Aerial Benchmark for Generalizable Interactive Road Extraction, *Chenxu Peng, Chenxu Wang, Yimian Dai, Yongxiang Liu, Ming-Ming Cheng, Xiang Li*
569 ❊ PGA: Prior-free Generative Attack for Practical No-box Scenario, *Hongyu Peng, Xiang Yuan, Gong Cheng*
570 Lipschitz Optimization for Formal Verification of Homographies, *Jean-Guillaume Durand, Panagiotis Kouvaros, Maxime Gariel, Alessio Lomuscio*
571 ❊ Batman: Benign Knowledge Alignment Through Malicious Null Space in Federated Backdoor Attack, *Wenwen He, Wenke Huang, Yiyang Fang, Wenjie Qu, Jiaheng Zhang, Mang Ye*
572 Out of Sight, Out of Track: Adversarial Attacks on Propagation-based Multi-Object Trackers via Query State Manipulation, *Halima Bouzidi, Haoyu Liu, Yonatan Achamyeleh, Praneetsai Iddamsetty, Mohammad Al Faruque*
573 Eliminate Distance Differences Induced by Backdoor Attacks: Layer-Selective Training and Clipping to Mask Backdoor Models, *Xuzeng Li, Tao Zhang, Xiangyun Tang, Jiacheng Wang, Jian Wang, Jiawen Kang, Jiqiang Liu, Zhen Han, Dusit Niyato, Dong In Kim*
574 Mitigating Error Amplification in Fast Adversarial Training, *Mengnan Zhao, Lihe Zhang, Bo Wang, Tianhang Zheng, Hong Zhong, Geyong Min*
575 Physical Adversarial Clothing Evades Visible-Thermal Detectors via Non-Overlapping RGB-T Pattern, *Xiaopei Zhu, Guanning Zeng, Zhanhao Hu, Jun Zhu, Xiaolin Hu*
576 What Your Features Reveal: Data-Efficient Black-Box Feature Inversion Attack for Split DNNs, *Zhihan Ren, Lijun He, Jiaxi Liang, Xinzhu Fu, Haixia Bi, Fan Li*
577 Exposing Functional Fusion: A New Class of Strategic Backdoor in Dynamic Prompt Architectures, *Zeyao Liu, Zhendong Zhao, Xiaojun Chen, Xin Zhao, Yuexin Xuan, Xiaoshuang Ji*
578 Learning to See and Act: Task-Aware Virtual View Exploration for Robotic Manipulation, *Yongjie Bai, Zhouxia Wang, Yang Liu, Kaijun Luo, Yifan Wen, Mingtong Dai, Weixing Chen, Ziliang Chen, Lingbo Liu, Guanbin Li, Liang Lin*
579 Evo-1: Lightweight Vision-Language-Action Model with Preserved Semantic Alignment, *Tao Lin, Yilei Zhong, Yuxin Du, Jingjing Zhang, Jiting Liu, Yinxinyu Chen, Encheng Gu, Ziyan Liu, Hongyi Cai, Yanwen Zou, Lixing Zou, Zhaoye Zhou, Gen Li, Bo Zhao*
580 FM-Steer: Enhance Generalist Policies with Value-Guided Cascaded Denoising, *Haoming Song, Delin Qu, Yuanqi Yao, Qizhi Chen, Jiarui Li, Qi Lv, Yiwen Tang, Li Kang, Heng Zhou, Xianqiang Gao, Yuhang Tang, Xiaofan Li, Modi Shi, Guanghui Ren, Maoqing Yao, Bin Zhao, Dong Wang, Xuelong Li*

581 Bootstrap Dynamic-Aware 3D Visual Representation for Scalable Robot Learning, *Qiwei Liang, Boyang Cai, Minghao Lai, Sitong Zhuang, Tao Lin, Yan Qin, Yixuan Ye, Jiaming Liang, Renjing Xu*

582 Visual Sim-to-Real at Scale for Humanoid Loco-Manipulation, *Tairan He, Zi Wang, Haoru Xue, Qingwei Ben, Zhengyi Luo, Wenli Xiao, Ye Yuan, Xingye Da, Fernando Castañeda, Shankar Sastry, Changliu Liu, Guanya Shi, Linxi Fan, Yuke Zhu*

583 Contact-Aware Neural Dynamics, *Changwei Jing, Jai Krishna Bandi, Jianglong Ye, Yan Duan, Pieter Abbeel, Xiaolong Wang, Sha Yi*

584 ❊ AVA-VLA: Improving Vision-Language-Action models with Active Visual Attention, *Lei Xiao, Jifeng Li, Juntao Gao, Feiyang Ye, Yan Jin, Jingjing Qian, Jing Zhang, Yong Wu, Xiaoyuan Yu*

585 UAST: Unified Active Search and Tracking for Arbitrary Targets with UAVs, *Liang Qin, Min Wang, Xingyu Lu, Aowen Qiu, Wengang Zhou, Houqiang Li*

586 SwiftVLA: Unlocking Spatiotemporal Dynamics for Lightweight VLA Models at Minimal Overhead, *Chaojun Ni, Cheng Chen, Xiaofeng Wang, Zheng Zhu, Wenzhao Zheng, Boyuan Wang, Tianrun Chen, Guosheng Zhao, Haoyun Li, Zhehao Dong, Qiang Zhang, Yun Ye, Yang Wang, Guan Huang,*

587 ❊ Visual-RRT: Finding Paths toward Visual-Goals via Differentiable Rendering, *Sebin Lee, Jumin Lee, Taeyeon Kim, Youngju Na, Woobin Im, Sung-Eui Yoon*

588 ❊ Cross-Hand Latent Representation for Vision-Language-Action Models, *Guangqi Jiang, Yutong Liang, Jianglong Ye, Jia-Yang Huang, Changwei Jing, Rocky Duan, Pieter Abbeel, Xiaolong Wang, Xueyan Zou*

589 Beyond Success: Refining Elegant Robot Manipulation from Mixed-Quality Data via Just-in-Time Intervention, *Yanbo Mao, Jianlong Fu, Ruoxuan Zhang, Hongxia Xie, Meibao Yao*

590 Physically Ground Commonsense Knowledge for Articulated Object Manipulation with Analytic Concepts, *Jiude Wei, Yuxuan Li, Cewu Lu, Jianhua Sun*

591 ❊ GeoPredict: Leveraging Predictive Kinematics and 3D Gaussian Geometry for Precise VLA Manipulation, *Jingjing Qian, Boyao Han, Chen Shi, Lei Xiao, Long Yang, Shaoshuai Shi, Li Jiang*

592 From Manuals to Actions: A Unified VLA Model for Chain-of-Thought Manual Generation and Robotic Manipulation, *Chenyang Gu, Jiaming Liu, Hao Chen, Runzhong Huang, Qingpo Wuwu, Xiaoqi Li, Zhuoyang Liu, Ying Li, Renrui Zhang, Peng Jia, Pheng-Ann Heng, Shanghang Zhang*

593 ❊ Real-World Point Tracking with Verifier-Guided Pseudo-Labeling, *Görkay Aydemir, Fatma Güney, Weidi Xie*

594 Rethinking Occlusion Modeling for UAV Tracking, *Jian Zhang, Xincheng Yu, Yi Lin*

595 Adaptive Capacity Autoregressive Visual Tracking, *Tong Lin, Yifan Bai, Shiyi Liang, Ruigang Niu, Xing Wei*

596 Spatio-Temporal Conditional Denoising Transformer for Modality-Missing RGBT Tracking, *Andong Lu, Ziyi Zha, Jiandong Jin, Shihao Li, Chenglong Li, Jin Tang, Bin Luo*

597 Breaking Smooth-Motion Assumptions: A UAV Benchmark for Multi-Object Tracking in Complex and Adverse Conditions, *Jingtao Ye, Zhang Kexin, Xunchi Ma, Johann Li, Guangming Zhu, Peiyi Shen, Linhua Jiang, Xiangdong Zhang, Zhang Liang*

598 TrackMAE: Video Representation Learning via Track Mask and Predict, *Renaud Vandeghen, Fida Mohammad Thoker, Marc Van Droogenbroeck, Bernard Ghanem*

599 ❊ Dual-branch Distilled Transformer for Efficient Asymmetric UAV Tracking, *Hongtao Yang, Bineng Zhong, Qihua Liang, Yaozong Zheng, Xiantao Hu, Yuanliang Xue, Shuxiang Song*

600 Multi-view Crowd Tracking Transformer with View-Ground Interactions Under Large Real-World Scenes, *Qi Zhang, Jixuan Chen, Kaiyi Zhang, Xinquan Yu, Antoni B. Chan, Hui Huang*

601 Scaling Self-Supervised and Cross-Modal Pretraining for Volumetric CT Transformers, *Cris Claessens, Christiaan Viviers, Giacomo D'Amicantonio, Egor Bondarev, Fons van der Sommen*

602 ❊ MuViT: Multi-Resolution Vision Transformers for Learning Across Scales in Microscopy, *Albert Dominguez Mantes, Gioele La Manno, Martin Weigert*

603 ❊ SemVideo: Reconstructs What You Watch from Brain Activity via Hierarchical Semantic Guidance, *Minghan Yang, Lan Yang, Ke Li, Honggang Zhang, Kaiyue Pang, Yi-Zhe Song*

604 Multimodal Causality-Driven Representation Learning for Generalizable Medical Image Segmentation, *Xusheng Liang, Lihua Zhou, Nianxin Li, Miao Xu, Ziyang Song, Dong Yi, Jinlin Wu, Jiawei Ma, Hongbin Liu, Zhen Lei, Jiebo Luo*

605 Simple Agents Outperform Experts in Biomedical Imaging Workflow Optimization, *Xuefei Wang, Kai Horstmann, Ethan Lin, Jonathan Chen, Alexander Farhang, Sophia Stiles, Atharva Sehgal, Jonathan Light, David Van Valen, Yisong Yue, Jennifer J. Sun*

606 TopoSlide: Topologically-Informed Histopathology Whole Slide Image Representation Learning, *Shahira Abousamra, Asmita Sood, Sylvia Plevritis*

607 Beyond the Static-World: Lifelong Learning for All-in-One Medical Image Restoration, *Shihao Shan, Hongying Liu, Fanhua Shang, Liang Wan, Jingjing Deng*

608 ❊ Hyperbolic Relational Prompts for Intersectional Fairness in Medical VLMs, *Jiayu Qian, Zongxian Yang, Guanxing Chen, Pengwei Hu, KC Tan, Yan Wang, Yu-An Huang, Zhi-An Huang*

609 RNED: Rotary Number Encoding and Decoding for Quantitative Medical VLM Analysis, *Fengbei Liu, Sunwoo Kwak, Nusrat Nizam, Ilan Richter, Ashley Beecy, Jayant Raikhelkar, Deborah Estrin, Mert R. Sabuncu*

610 MLLM-HWSI: A Multimodal Large Language Model for Hierarchical Whole Slide Image Understanding, *Basit Alawode, Arif Mahmood, Muaz Khalifa Al Radi, Shahad Albastaki, Asim Khan, Muhammad Bilal, Moshira Ali Abdalla, Mohammed Bennamoun, Sajid Javed*

611 Learning Generalizable 3D Medical Image Representations from Mask-Guided Self-Supervision, *Yunhe Gao, Yabin Zhang, Chong Wang, Jiaming Liu, Maya Varma, Jean-Benoit Delbrouck, Akshay Chaudhari, Curtis Langlotz*

612 BiOTPrompt: Bidirectional Optimal Transport Guided Prompting for Disease Evolution-aware Radiology Report Generation, *Tengfei Liu, Yijian Fan, Boyue Wang, Yongli Hu, Mingjie Li, Jinghua Li, Junbin Gao, Xiaojun Chang, Zhihui Li, Baocai Yin*

613 ❊ Learning to See Through a Baby's Eyes: Early Visual Diets Enable Robust Visual Intelligence in Humans and Machines, *Yusen Cai, Qing Lin, Bhargava Satya Nunna, Mengmi Zhang*

614 UDAPose: Unsupervised Domain Adaptation for Low-Light Human Pose Estimation, *Haopeng Chen, Yihao Ai, Kabeen Kim, Robby T. Tan, Yixin Chen, Bo Wang*

615 Enhancing Accuracy of Uncertainty Estimation in Appearance-based Gaze Tracking with Probabilistic Evaluation and Calibration, *Qiaojie Zheng, Jiucai Zhang, Amy Zhang, Xiaoli Zhang*

616 SCAPO: Self-Supervised Category-Level Articulated Pose Estimation from a Single 3D Observation, *Can Zhang, Gim Hee Lee*

617 Composite-Attribute Person Re-Identification via Pose-Guided Disentanglement, *Kartik Patwari, Noranart Vesdapunt, Chien-Yi Wang, Dawei Li, Cong Phuoc Huynh, Ning Zhou, Chen-Nee Chuah, Kah Kuen Fu*

618 ❊ Representing 3D Faces with Learnable B-Spline Volumes, *Prashanth Chandran, Daoye Wang, Timo Bolkart*

619 RHINO: Reconstructing Human Interactions with Novel Objects from Monocular Videos, *Lixin Xue, Chengwei Zheng, Georgios Paschalidis, Chen Guo, Manuel Kaufmann, Juan Zarate, Dimitrios Tzionas*

620 HumanBA: Human-Aware Bundle Adjustment via Global Human-Camera Decoupling, *Fengyuan Yang, Tanuj Sur, Tze Ho Elden Tse, Angela Yao*

621 ❊ HamiPose: Hamiltonian Optimization for Unsupervised Domain Adaptive Pose Estimation, *Jiawen Li, Fei Jiang, Dandan Zhu, Aimin Zhou*

622 KASALv2: Fully Automatic 3D Rotational Symmetry Classification and Axis Localization, *Mengxin Zhang, Yulin Wang, Chen Luo, Yongzhe Li, Yijun Zhou*

623 AnyLift: Scaling Motion Reconstruction from Internet Videos via 2D Diffusion, *Hongjie Li, Heng Yu, Jiaman Li, Hong-Xing Yu, Ehsan Adeli, C. Karen Liu, Jiajun Wu*

624 Active Inference for Micro-Gesture Recognition: EFE-Guided Temporal Sampling and Adaptive Learning, *Weijia Feng, Jingyu Yang, Ruojia Zhang, Fengtao Sun, Qian Gao, Chenyang Wang, Tongtong Su, Jia Guo, Xiaobai Li, Minglai Shao*

625 ArtPro: Self-Supervised Articulated Object Reconstruction with Adaptive Integration of Mobility Proposals, *Xuelu Li, Zhaonan Wang, Xiaogang Wang, Lei Wu, Manyi Li, Changhe Tu*

626 Similarity-Consistent Likelihood Diffusion enables Hidden Person Detection from Wall Reflections, *Zhiwen Zheng, Hao Zhou, Huiyu Qi, Zhao Huang, Guangyuan Zhang, Shaowei Jiang, Wenwen Tang, Bin Yang, Jin Liu, Xiaoshuai Zhang, Xingru Huang*
627 VLM-Guided Group Preference Alignment for Diffusion-based Human Mesh Recovery, *Wenhao Shen, Hao Wang, Wanqi Yin, Fayao Liu, Xulei Yang, Chao Liang, Zhongang Cai, Guosheng Lin*
628 Occluded Human Body Capture with Frequency Domain Denoising Prior, *Buzhen Huang, Chongyang Xu, Wentao Tang, Yuan Shu, Jingyi Ju, Binghui Zuo, Yangang Wang*
629 ResiHMR: Residual-Limb Aware Single-Image 3D Human Mesh
❊ Recovery for Individuals with Limb Loss, *Jiaying Ying, Heming Du, Kaihao Zhang, Sean M. Tweedy, Xin Yu*
630 OnlineHMR: Video-based Online World-Grounded Human Mesh Recovery, *Yiwen Zhao, Ce Zheng, Yufu Wang, Hsueh-Han Daniel Yang, Liting Wen, László A. Jeni*
631 MimiCAT: Mimic with Correspondence-Aware Cascade-Transformer for Category-Free 3D Pose Transfer, *Zenghao Chai, Chen Tang, Yongkang Wong, Xulei Yang, Mohan Kankanhalli*
632 Exploring Adaptive Masked Reconstruction for Self-Supervised Skeleton-Based Action Recognition, *Shengkai Sun, Zhiyong Cheng, Zefan Zhang, Jianfeng Dong, Zhihui Li, Meng Wang*
633 DFD-HR: Generalizable Deepfake Detection via Hierarchical Routing Learning, *Jiamu Sun, Zhiyuan Yan, Ke-Yue Zhang, Taiping Yao, Shouhong Ding*
634 MGDHand: Multi-Granularity Prior-to-Inertial Distillation Framework for Sequential 3D Hand Pose Estimation from Sparse IMUs, *Xinyi Wang, Pengfei Ren, Haoyang Zhang, Hanling Zhan, Yingxi Li, Liang Xie, Yue Gao, Erwei Yin*
635 CARI4D: Category Agnostic 4D Reconstruction of Human-Object Interaction, *Xianghui Xie, Bowen Wen, Yan Chang, Hesam Rabeti, Jiefeng Li, Ye Yuan, Gerard Pons-Moll, Stan Birchfield*
636 E-3DPSM: A State Machine for Event-based Egocentric 3D Human Pose Estimation, *Mayur Deshmukh, Hiroyasu Akada, Helge Rhodin, Christian Theobalt, Vladislav Golyanik*
637 Bézier Degradation Modeling for LiDAR-based Human Motion Capture, *Xiaoqi An, Lin Zhao, Jun Li, Chen Gong, Jian Yang*
638 UniSH: Unifying Scene and Human Reconstruction in a Feed-Forward Pass, *Mengfei Li, Peng Li, Zheng Zhang, Jiahao Lu, Chengfeng Zhao, Wei Xue, Qifeng Liu, Sida Peng, Wenxiao Zhang, Wenhan Luo, Yuan Liu, Yike Guo*
639 Illumination-Consistent Human-Scene Reconstruction from Monocular Video, *Rongbin Zheng, Wensheng Li, Lingzhe Zeng, Dong Wang, Chengying Gao*
640 Attribution as Retrieval: Model-Agnostic AI-Generated Image Attribution, *Hongsong Wang, Renxi Cheng, Chaolei Han, Jie Gui*
641 Agent4FaceForgery: Multi-Agent LLM Framework for Realistic Face Forgery Detection, *Yingxin Lai, Zitong YU, Jun Wang, Linlin Shen, Yong Xu, Xiaochun Cao*
642 Enabling Supervised Learning of Generative Signatures for Generalized
❊ Synthetic Image Detection, *Jianwei Fei, Yunshu Dai, Xiaoyu Zhou, Zhihua Xia, Alessandro Piva*
643 DiffusionFF: A Diffusion-based Framework for Joint Face Forgery Detection and Fine-Grained Artifact Localization, *Siran Peng, Haoyuan Zhang, Li Gao, Tianshuo Zhang, Xiangyu Zhu, Bao Li, Weisong Zhao, Zhen Lei*
644 All in One: Unifying Deepfake Detection, Tampering Localization, and
❊ Source Tracing with a Robust Landmark-Identity Watermark, *Junjiang Wu, Liejun Wang, Zhiqing Guo*
645 Towards an Incremental Unified Multimodal Anomaly Detection: Augmenting Multimodal Denoising From an Information Bottleneck Perspective, *Kaifang Long, Lianbo Ma, Jiaqi Liu, Liming Liu, Guoyang Xie*
646 AG-VAS: Anchor-Guided Zero-Shot Visual Anomaly Segmentation with Large Multimodal Models, *Zhen Qu, Xian Tao, Xiaoyi Bao, Dingrong Wang, ShiChen Qu, Zhengtao Zhang, Xingang Wang*
647 Dual-Prototype-Guided Multi-task Learning for Unsupervised Anomaly Detection and Classification, *Qianhao Luo, Jiajia Mi, Mingtao Yan, JingSheng Liu, ShuYang Pang, Weiling Li*
648 The Road Less Seen: Segment Exploration for Weakly Supervised Video Anomaly Detection, *Anusha Acharya, Hitesh Sapkota, Qi Yu, Xumin Liu*
649 Omni-AD: A Large-scale and Versatile Benchmark for Industrial Anomaly Detection, *Dahu Shi, Chengshen He, Shaochen Zhang, Bo Qian, Xiaochen Quan, Wencong Zhang, Xing Wei*
650 Back to Point: Exploring Point-Language Models for Zero-Shot 3D Anomaly Detection, *Kaiqiang Li, Gang Li, Mingle Zhou, Min Li, Delong Han, Jin Wan*
651 Complementary Prototype Mapping for Efficient Multimodal Anomaly Detection, *Yuan Zhao, Xiaoqin Zhang, Huchuan Lu, Lihe Zhang*
652 LiDAS: Lighting-driven Dynamic Active Sensing for Nighttime Perception, *Simon de Moreau, Andrei Bursuc, Hafid El Idrissi, Fabien Moutarde*
653 Gau-Occ: Geometry-Completed Gaussians for Multi-Modal 3D Occupancy Prediction, *Chengxin Lv, Yihui Li, Hongyu Yang, YunHong Wang*
654 OpenVO: Open-World Visual Odometry with Temporal Dynamics Awareness, *Phuc Nguyen, Anh N. Nhu, Ming C. Lin*
655 An Instance-Centric Panoptic Occupancy Prediction Benchmark for Autonomous Driving, *Yi Feng, Junwu E, Zizhan Guo, Yu Ma, Hanli Wang, Rui Fan*
656 OneOcc: Semantic Occupancy Prediction for Legged Robots with a Single Panoramic Camera, *Hao Shi, Ze Wang, Shangwei Guo, Mengfei Duan, Song Wang, Teng Chen, Kailun Yang, Lin Wang, Kaiwei Wang*
657 ProOOD: Prototype-Guided Out-of-Distribution 3D Occupancy Prediction, *Yuheng Zhang, Mengfei Duan, Kunyu Peng, Yuhang Wang, Di Wen, Danda Pani Paudel, Luc Van Gool, Kailun Yang*

16:00 - 18:00 DEMOS (ExHall F)

1 FlowSLAM: Dense Visual SLAM on Edge Devices, *Aniket Gupta, Tianye Ding, Ajay Rajendra Kumar, Dennis Giaya, Pia Bideau, Charles Saunders, Qu Cao, Aruni RoyChowdhury, Hanumant Singh, Huaizu Jiang*
2 FoundYou: A Unified Model for Personalized Segmentation and Retrieval, *Gabriele Trivigno, Marcos Alfaro, Claudia Cuttano, Gabriele Berton, Luis Paya, Carlo Masone*
3 Code2Video: A Code-centric Paradigm for Educational Video Generation, *Yanzhe Chen, Kevin Qinghon Lin, Mike Zheng Shou*
4 MIBURI: Towards Expressive Interactive Gesture Synthesis, *M. Hamza Mughal, Rishabh Dabral, Vera Demberg, Christian Theobalt*
5 RnG, *Mochu Xiang*

Saturday, June 6

7:30 - 17:00 **Registration / Badge Pickup** (Lobby A)
7:00 - 17:00 **Press Room** (ExHall F)
7:00 - 17:00 **Mother's Room** (Adjacent to Room 102)
7:00 - 17:00 **Quiet Room** (Adjacent to Room 102)
7:00 - 17:00 **Prayer Room** (Room 206)
7:30 - 9:00 **Findings Posters** (ExHall A)
7:30 - 9:00 **Breakfast** (ExHall C)
7:30 - 17:00 **Poster Pickup / T-shirt Pickup** (ExHall A & F)
9:00 - 10:15 **Oral Session 3A: Generative Diffusion Modeling** (Bluebird Ballroom)
Oral Session 3B: Spatial Understanding (Four Seasons Ballroom)
Oral Session 3C: Generative Editing (Mile High Ballroom 1A - 2A)
Oral Session 3D: Multimodal Modeling (Mile High Ballroom 3A - 4A)
10:15 - 10:30 **Courtesy Break**
10:30 - 11:30 **KEYNOTE 2 - Jerry Chow, IBM, TJ Watson, CTO Quantum Centric Supercomputing; Transforming Computing with Quantum-Centric Supercomputing** (Bluebird Ballroom)
11:15 - 11:45 **Poster Setup** (ExHall A)
11:45 - 13:45 **Poster Session 3 & Exhibit Hall** (ExHall F)
11:45 - 18:00 **Art Program** (ExHall E)
11:45 - 13:45 **DEMOS** (ExHall F)
11:45 / 17:00 **Art Gallery Tour with Curator, Luba Elliott** (ExHall E) 30 mins each
11:45 - 13:30 **LUNCH** (ExHall C)
11:45 - 13:45 **Doctoral Consortium** (Room 207) By Invitation Only, Sponsored by NSF.
Supported by: U.S. National Science Foundation
14:00 - 15:15 **Oral Session 4A: Geometric Understanding** (Bluebird Ballroom)
Oral Session 4B: Embodied & Agentic Intelligence (Four Seasons Ballroom)
Oral Session 4C: Spatial Reasoning (Mile High Ballroom 1A - 2A)
Oral Session 4D: Visual Segmentation (Mile High Ballroom 3A - 4A)
13:45 - 14:45 **Art Panel** (Room 201)
15:15 - 15:30 **Courtesy Break**
15:30 - 16:30 **PAMI TC Meeting** (Four Seasons Ballroom)
16:15 - 16:45 **Poster Setup** (ExHall A)
16:45 - 18:45 **Poster Session 4 & Exhibit Hall w/ Coffee Break** (ExHall A)
16:45 - 18:45 **DEMOS**
19:00 - 21:00 **RECEPTION: Join us in the Bluebird Ballroom for drinks and live music. A quiet space for food and conversation will be in our catering hall, Exhibit Hall C** (Bluebird Ballroom & ExHall C)

7:30 - 9:00 **Findings Posters** (ExHall A)

1 Beyond Top-1: Forensic Analysis of Full Prediction Distributions Reveals Hidden Model Reasoning, *Minhyeok Lee*
2 Zero-Shot Textual Explanations via Translating Decision-Critical Features, *Toshinori Yamauchi, Hiroshi Kera, Kazuhiko Kawamoto*
3 DMin: Scalable Training Data Influence Estimation for Diffusion Models, *Huawei Lin, Yingjie Lao, Weijie Zhao*
4 A Framework for Evaluating Zero-Shot Image Generation in Concept-Based Explainability, *Giacomo Astolfi, Matteo Bianchi, Riccardo Campi, Antonio De Santis, Marco Brambilla*
5 Self-Guided Integrated Gradient Method for Attribution, *Sabrina Henry, Alice Ruget, Stirling Scholes, Jonathan Leach*
6 Circuit Tracing in Vision-Language Models: Understanding the Internal Mechanisms of Multimodal Thinking, *Jingcheng Yang, Tianhu Xiong, Shengyi Qian, Klara Nahrstedt, Mingyuan Wu*
7 Discovering Attention Head Interactions in Vision Transformers, *Zhenyu Lu, Yuheng Jia, Wei You, Hao Chen*
8 Value bounds and Convergence Analysis for Averages of LRP attributions, *Alexander Binder, Nastaran Takmil-Homayouni, Urun Dogan*
9 MReactor: Offline Multiple Appropriate Facial Reaction Generation with Hierarchical Cognitive Disentanglement, *Jiachen Luo, Jiajun He, Shuai Shen, Lin Wang, Huy Phan, Joshua Reiss, Lin Haijun, Bjoern Schuller, Zeyu Fu, Siyang Song*
10 B-MoE: A Body-Part-Aware Mixture-of-Experts "All Parts Matter" Approach to Micro-Action Recognition, *Nishit Poddar, Aglind Reka, Diana-Laura Borza, Snehashis Majhi, Michal Balazia, Abhijit Das, François Brémond*
11 Learning by Neighbor-Aware Semantics, Deciding by Open-Form Flows: Towards Robust Zero-Shot Skeleton Action Recognition, *Yang Chen, Miaoge Li, Zhijie Rao, Deze Zeng, Song Guo, Jingcai Guo*
12 Actionable Human Motion Generation via Latent Imitation and Fine-Grained Text Completion, *Feiyang Xie, Haoqi Yuan, Zongqing Lu*
13 GHOST: Fast Category-Agnostic Hand-Object Interaction Reconstruction from RGB Videos Using Gaussian Splatting, *Ahmed Tawfik Aboukhadra, Marcel Rogge, Nadia Robertini, Abdalla Arafa, Jameel Malik, Ahmed Elhayek, Didier Stricker*
14 Hoi3DGen: Generating High-Quality Human-Object-Interactions in 3D, *Agniv Sharma, Xianghui Xie, Tom Fischer, Eddy Ilg, Gerard Pons-Moll*
15 CoherentHand: Temporally Consistent 3D Hand Trajectory Synthesis with Semantic Motion Priors, *Bikram Boote, Junho Kim, Ozgur Kara, Sangmin Lee, James M Rehg*
16 Weakly Supervised Micro-Expression Spotting based on Boundary Refinement Mechanism and Cross-subject Learning Representation, *Zhihua Xie, Haolin Chang, Guohua Miao, Jianing Chen*
17 FUSION: Full-body Unified Motion Prior for Body and Hands Via Diffusion, *Enes Duran, Nikos Athanasiou, Muhammed Kocabas, Michael J. Black, Omid Taheri*
18 BridgeDiffusion: Latent Space Optimization for Independent Body-Part Generation with Motion Consistency Bridges in Interactive Dance, *Yufei Huo, Ao Li, Wenxun Dai, Songli Wu, Yansong Tang*
19 MARIO: Motion-Augmented Real-Time Multi-Sensor Inertial Odometry, *Yiquan Li, Taeyoung Yeon, Chenfeng Gao, Vasco Xu, Xuanyou Liu, Karan Ahuja*
20 BitTP: The Lightweight Trajectory Prediction Model with BitLLM for Edge-Devices, *Mincheol Kang, HyunJin Lim, Bomin Kang, Daehee Park*
21 WHOLE: World-Grounded Hand-Object Lifted from Egocentric Videos, *Yufei Ye, Jiaman Li, Ryan Rong, C. Karen Liu*
22 TalkVid: A Large-Scale Diversified Dataset for Audio-Driven Talking Head Synthesis, *Shunian Chen, Hejin Huang, Yexin Liu, Zihan Ye, Pengcheng Chen, Chenghao Zhu, Michael Guan, Rongsheng Wang, Junying Chen, Jianye Hou, Bo Li, Guanbin Li, Ser-Nam Lim, Harry Yang, Benyou Wang*
23 How2Sign-Synth3D: Markerless Holistic Sign Language Performance Capture and Synthetic Data for Dense Landmark Tracking, *Levente Tempfli, Stephan Huber, Oscar Koller, Amanda Duarte*
24 SocialMirror: Reconstructing 3D Human Interaction Behaviors from Monocular Videos with Semantic and Geometric Guidance, *Qi Xia, Peishan Cong, Ziyi Wang, Yujing Sun, Qin Sun, Xinge Zhu, Mao Ye, Ruigang Yang, Yuexin Ma*
25 EggHand: A Multimodal Foundation Model for Egocentric Hand Pose Forecasting, *Jaeyoung Choi, Hyeondong Kim, Yujin Kim, Daehee Park*
26 Towards Metric-Aware Multi-Person Mesh Recovery by Jointly Optimizing Human Crowd in Camera Space, *Kaiwen Wang, Kaili Zheng, Yiming Shi, Chenyi Guo, Ji Wu*
27 VoxFace: Streaming Audio-Visual Synthesis via Relay-Style Multi-Token Prediction for Interactive Conversation, *Junwen Xiong, Chuanyue Li, Peng Zhang*
28 OmniHead: A Unified Model for Dynamic Nonverbal Facial Behaviors, *Pierre Vuillecard, Jean-Marc Odobez*
29 Detecting Precise Hand Touch Moments in Egocentric Video, *Huy Anh Nguyen, Feras Dayoub, Minh Hoai*
30 Less is More: Multimodal Human Pose Estimation with Selective Fusion,

Yutong Xu, Qianyi Huang, Xu Chen

31 PHYLOMAN: Generative Behavior Control via Fusing LLM Planning and Physics-based Control, *Jusheng Zhang, Jinzhou Tang, Sidi Liu, Jian Wang, Keze Wang*

32 Contact Matrix: Enhancing Dance Motion Synthesis with Precise Interaction Modeling, *Xuhai Chen, Zhi Cen, Huaijin Pi, Sida Peng, Xiaowei Zhou, Yong Liu*

33 Learning Predictive Visuomotor Coordination, *Wenqi Jia, Bolin Lai, Xu Cao, Miao Liu, Danfei Xu, James M. Rehg*

34 FSMC-Pose: Frequency and Spatial Fusion with Multiscale Self-Calibration for Cattle Mounting Pose Estimation, *Fangjing Li, Zhihai Wang, Xinxin Ding, Haiyang Liu, Ronghua Gao, Rong Wang, Yao Zhu, Ming Jin*

35 Bootstrapping Sign Language Annotations with Sign Language Models, *Colin Lea, Vasileios Baltatzis, Connor Gillis, Raja Kushalnagar, Lorna Quandt, Leah Findlater*

36 OmniMotion-X: Versatile Multimodal Whole-Body Motion Generation, *Guowei Xu, Yuxuan Bian, Ailing Zeng, Zhuo Chen, Mingyi Shi, Shaoli Huang, Wen Li, Lixin Duan, Qiang Xu*

37 THOM: Generating Physically Plausible Hand-Object Meshes From Text, *Uyoung Jeong, Yihalem Yimolal Tiruneh, Hyung Jin Chang, Seungryul Baek, Kwang In Kim*

38 ExposeAnyone: Personalized Audio-to-Expression Diffusion Models Are Robust Zero-Shot Face Forgery Detectors, *Kaede Shiohara, Toshihiko Yamasaki, Vladislav Golyanik*

39 All-Age Human Mesh Recovery, *Laura Bravo-Sánchez, Matthieu Armando, Romain Brégier, Grégory Rogez, Serena Yeung-Levy, Fabien Baradel*

40 GeneFlow: Modeling Heredity and Variation via Flow Matching Transformers for Kinship Verification, *Yihang Wu, Xianxu Hou, Linlin Shen*

41 Dynamic Full-body Motion Agent with Object Interaction via Blending Pre-trained Modular Controllers, *Sanghyeok Nam, Byoungjun Kim, Daehyung Park, Tae-Kyun Kim*

42 MotionDuet: Dual-Conditioned 3D Human Motion Generation with Video-Regularized Text Learning, *Yi-Yang Zhang, Tengjiao Sun, Pengcheng Fang, Deng-Bao Wang, Xiaohao Cai, Min-Ling Zhang, Hansung Kim*

43 Fast-HOI: Fast Human-Object Interaction Synthesis via Distilled Interaction Prior and Physical Constrains, *Xiaokang Pan, Zhizhong Zhang, Yangyuan Liu, Zhuoran Chen, Zhiwei Zhang, Bin Ji, Mingang Chen, Yong Xie, Jingyu Gong, Xuhong Wang, Xin Tan, Yuan Xie*

44 HM-Talker: Hybrid Motion Modeling for High-Fidelity Talking Head Synthesis, *Shiyu Liu, Kui Jiang, Junjun Jiang, Xianming Liu, Xiaocheng Feng, Fei Ma, Hongxun Yao, Qi Tian*

45 GeoHOI: Geometry-Enhanced Human-Object Interaction Video Generation via Hierarchical Multi-Modal Injection, *Ziyi Xu, Zejing Rao, Juan Cao, Xiaoqiang Liu, Zhixue Fang, Haoxian Zhang, Songlin Tang, Fan Tang*

46 TAUE: Training-free Noise Transplant and Cultivation Diffusion Model, *Daichi Nagai, Ryugo Morita, Shunsuke Kitada, Hitoshi Iyatomi*

47 GR-Diffusion: Graph-Guided Relational-Aware Diffusion via Attention Alignment, *Xiaochen Liu, Xiaoting Xi, Chao Yin, Xiaoqiang Li, Daoguo Dong*

48 V-GRPO: Online Reinforcement Learning for Denoising Generative Models Is Easier than You Think, *Bingda Tang, Yuhui Zhang, Xiaohan Wang, Jiayuan Mao, Ludwig Schmidt, Serena Yeung-Levy*

49 FREESTYLE: An Anchor-Free Mechanism for Training-Free Style-Aligned Image Generation, *Minseok Oh, Jihun Park, Jongmin Gim, Minwoo Choi, Kyoungmin Lee, Ferdinando Fioretto, Sunghoon Im*

50 Is Your Text-to-Image Model Robust to Caption Noise?, *Weichen Yu, Ziyan Yang, Shanchuan Lin, Qi Zhao, Jianyi Wang, Liangke Gui, Matt Fredrikson, Lu Jiang*

51 Decomposing Subject-Driven Image Generation via Intermediate Structural Prediction, *Hanzhong Guo, Yizhou Yu*

52 RectifiedHR: Enable Efficient High-Resolution Synthesis via Energy Rectification, *Zhen Yang, Guibao Shen, Minyang Li, Liang Hou, Mushui Liu, Luozhou Wang, Xin Tao, Ying-Cong Chen*

53 ASTRA: Enhancing Multi-Subject Generation with Retrieval-Augmented Pose Guidance and Disentangled Position Embedding, *Tianze Xia, Zijian Ning, Zonglin Zhao, Mingjia Wang*

54 Objects in Generated Videos Are Slower Than They Appear: Models Suffer Sub-Earth Gravity and Don't Know Galileo's Principle...for now, *Varun Varma Thozhiyoor, Shivam Tripathi, Venkatesh Babu Radhakrishnan, Anand Bhattad*

55 Group Relative Attention Guidance for Image Editing, *Xuanpu Zhang, Xuesong Niu, Ruidong Chen, Dan Song, Jianhao Zeng, Penghui Du, Haoxiang Cao, Kai Wu, An-an Liu*

56 ControlPose: High-Fidelity Pose-Controlled Image Generation with Multi-Faceted Pose Disentanglement, *Zhongjing Du, Xiao Chen, Zhiwei Nie, Yuxuan Chen, Chang Liu, Xiangyang Ji, Jie Chen*

57 FlowC2S: Flowing from Current to Succeeding Frames for Fast and Memory-Efficient Video Continuation, *Hovhannes Margaryan, Quentin Bammey, Christian Sandor*

58 Latent-Compressed Variational Autoencoder for Video Diffusion Models, *Jiarui Guan, Wenshuai Zhao, Zhengtao Zou, Juho Kannala, Arno Solin*

59 Deep Parameter Interpolation for Scalar Conditioning, *Chicago Y. Park, Michael T. McCann, Cristina Garcia-Cardona, Brendt Wohlberg, Ulugbek S. Kamilov*

60 Mining Real-World Image Relations for Large-Scale Controllable Generation and Editing, *Hao Shao, Liyang Liu, Zhengxiong Luo, Zhuofan Zong, Hongsheng Li*

61 Disentangle Once, Control All: A Unified and Efficient Framework for Disentangling Multi-Condition Control in Human Video Generation, *Runqi Wang, Chuming Wang, Fangqiu Yi, Yuying Zhao, Jingyu Xu, Yuhang Dai, Zheng Wang, Chi Zhang*

62 HAM: A Training-Free Style Transfer Approach via Heterogeneous Attention Modulation for Diffusion Models, *Yeqi He, Liang Li, Zhiwen Yang, Xichun Sheng, Zhidong Zhao, Chenggang Yan*

63 Gaussian Shannon: High-Precision Diffusion Model Watermarking Based on Communication, *Yi Zhang, Hongbo Huang, Liang-Jie Zhang*

64 Video4Spatial: Towards Visuospatial Intelligence with Context-Guided Video Generation, *Zeqi Xiao, Yiwei Zhao, Lingxiao Li, Yushi Lan, Ning Yu, Rahul Garg, Mohammad H. Taghavi, Xingang Pan*

65 Text-Driven Reasoning Video Editing via Reinforcement Learning on Digital Twin Representations, *Yiqing Shen, Chenjia Li, Mathias Unberath*

66 Beyond Optimal Transport: Model-Aligned Coupling for Flow Matching, *Yexiong Lin, Yu Yao, Yang Zhou, Tongliang Liu*

67 Stochastic Perturbations Improve Distribution-to-Distribution Generative Models, *Shiye Su, Yuhui Zhang, Linqi Zhou, Rajesh Ranganath, Serena Yeung-Levy*

68 StereoSpace: Depth-Free Synthesis of Stereo Geometry via End-to-End Diffusion in a Canonical Space, *Tjark Behrens, Anton Obukhov, Bingxin Ke, Fabio Tosi, Matteo Poggi, Konrad Schindler*

69 FA-MoE: Improving Medical Image Generation Through Frequency-Aware Mixture of Experts, *Yifan Sun, Qingjie Meng, Tao Chen, Huiping Chen*

70 Generated Reality: Human-Centric World Simulation Using Interactive Video Generation with Hand and Camera Control, *Linxi Xie, Lisong C. Sun, Ashley Neall, Tong Wu, Shengqu Cai, Gordon Wetzstein*

71 VHOI: Controllable Video Generation of Human–Object Interactions from Sparse Trajectories via Motion Densification, *Wanyue Zhang, Lin Geng Foo, Thabo Beeler, Rishabh Dabral, Christian Theobalt*

72 LoViC: Efficient Long Video Generation with Context Compression, *Jiaxiu Jiang, Wenbo Li, Jingjing Ren, Yuping Qiu, Renjing Pei, Fenglong Song, Yong Guo, Xiaogang Xu, Han Wu, Wangmeng Zuo*

73 FedErase: Personalized Federated Unlearning for Text-to-Image Diffusion Models, *Tianyu Geng, Wenfei Liang, Sijie Wang, Rui She, Wee Peng Tay*

74 Zero4D: Training-Free 4D Video Generation From Single Video Using Off-the-Shelf Video Diffusion Models, *Jangho Park, Taesung Kwon, Jong Chul Ye*

75 Earthquake-Bench: Video Generation Benchmark for Earthquake Simulation, *Lei Bao, Hao Chen, Yuyan Chen, Kui Wu, Lijia Chen, Fangwei Zhong, Feiran Huang, Bo Song, Han Yang*

76 OmniInsert: Mask-Free Video Insertion of Any Reference via Diffusion Transformer Models, *Jinshu Chen, Xinghui Li, Xu Bai, Tianxiang Ma, Pengze Zhang, Mengtian Li, Zhuowei Chen, Gen Li, Lijie Liu, Songtao Zhao, Bingchuan Li, Qian He*

77 Block Cascading: Training Free Acceleration of Block-Causal Video Models, *Hmrishav Bandyopadhyay, Nikhil Pinnaparaju, Rahim Entezari, Jim Scott, Yi-Zhe Song, Varun Jampani*

78 Activation-Norm Maximization to Accelerate Training in Flow-Matching Transformers, *Yash Belhe, Wesley Chang, Tzu-Mao Li, Ravi Ramamoorthi, Michaël Gharbi*
79 FREE: Uncertainty-Aware Autoregression for Parallel Diffusion Transformers, *Xinwan Wen, Bowen Li, Jiajun Luo, Ye Li, Zhi Wang*
80 No Cache Left Idle: Accelerating diffusion model via Extreme-Slimming Caching, *Tingyan Wen, Haoyu Li, Yihuang Chen, Xing Zhou, Lifei Zhu, XueQian Wang*
81 Inference-Time Alignment of Diffusion Models with Evolutionary Algorithms, *Purvish Jajal, Nicholas John Eliopoulos, Benjamin Shiue-Hal Chou, George K Thiruvathukal, James C. Davis, Yung-Hsiang Lu*
82 TokenErase: Robust Concept Erasure via Visual-Injected Token Optimization, *Liangshun Zou, Zhangkai Ni, Hanli Wang*
83 VisionCreator: A Native Visual-Generation Agentic Model with Understanding, Thinking, Planning and Creation, *Jinxiang Lai, Zexin Lu, Jiajun He, Rongwei Quan, Wenzhe Zhao, Qinyu Yang, Qi Chen, Qin Lin, Chuyue Li, Tao Gao, Yuhao Shan, Song Guo, Qinglin Lu*
84 Animated-ART: Multi-Layer Transparent Video Generation, *Ziqiang Li, Yunnan Wang, Dong Chen, Yue Dong, Ji Li, Yuhui Yuan, Xin Jin*
85 Rethinking Conditioning in Diffusion Models: Dynamic Token Scheduling for Efficient and Aligned Text-to-Image Generation, *Jia Li, Xiaomeng Fu, Yizhao Gao, Jiaxu Wang, Xi Wang, Hayden Kwok-Hay So*
86 Attention-Guided Energy Optimization for Label-Aligned Anomaly Generation, *Zhibin Wan, Zhiqiang Gao, Mingjie Sun, Yupei Wu, Guohong Fu, Ran Yi*
87 USV: Unified Sparsification for Accelerating Video Diffusion Models, *Xinjian Wu, Hongmei Wang, Yuan Zhou, Qinglin Lu*
88 OminPSD: Layered PSD Generation with Diffusion Transformer, *Cheng Liu, Yiren Song, Haofan Wang, Mike Zheng Shou*
89 Video2LoRA: Unified Semantic-Controlled Video Generation via Per-Reference-Video LoRA, *Zexi Wu, Baolu Li, Jing Dai, Yiming Zhang, Yue Ma, Qinghe Wang, Xu Jia, Hongming Xu*
90 Depth Adaptive Efficient Visual Autoregressive Modeling, *Chunliang Li, Tianze Cao, Sanyuan Zhao*
91 Cross-Resolution Diffusion Models Via Network Pruning, *Jiaxuan Ren, Junhan Zhu, Huan Wang*
92 FrameDiT: Diffusion Transformer with Matrix Attention for Efficient Video Generation, *Minh Khoa Le, Kien Do, Duc Thanh Nguyen, Truyen Tran*
93 Understanding Reward Hacking in Text-to-Image Reinforcement Learning, *Yunqi Hong, Kuei-Chun Kao, Hengguang Zhou, Cho-Jui Hsieh*
94 OminiControl2: Efficient Conditioning for Diffusion Transformers, *Zhenxiong Tan, Qiaochu Xue, Xingyi Yang, Songhua Liu, Xinchao Wang*
95 Seen-to-Scene: Keep the Seen, Generate the Unseen for Video Outpainting, *Inseok Jeon, Minhyeok Lee, Seunghoon Lee, Minseok Kang, Suhwan Cho, Sangyoun Lee*
96 AdaGaR: Adaptive Gabor Representation for Dynamic Scene Reconstruction, *Jiewen Chan, Zhenjun Zhao, Yu-Lun Liu*
97 MoVieDrive: Urban Scene Synthesis with Multi-Modal Multi-View Video Diffusion Transformer, *Guile Wu, David Huang, Dongfeng Bai, Bingbing Liu*
98 InstaDA: Augmenting Instance Segmentation Data with Dual-Agent System, *Xianbao Hou, Yonghao He, Zeyd Boukhers, John See, Hu Su, Wei Sui, Cong Yang*
99 One Model for All: Unified Try-On and Try-Off in Any Pose via LLM-Inspired Bidirectional Tweedie Diffusion, *Jinxi Liu, Zijian He, Guangrun Wang, Guanbin Li, Liang Lin*
100 Adversarial Concept Distillation for One-Step Diffusion Personalization, *Yixiong Yang, Tao Wu, Senmao Li, Shiqi Yang, Yaxing Wang, Joost van de Weijer, Kai Wang*
101 DSA: Dynamic Step Allocation for Fast Autoregressive Video Generation, *Thanh-Tung Le, Yunhan Zhao, Menglei Chai, Zhengyang Shen, Zhe Cao, Danhang Tang, Xiaohui Xie, Deying Kong*
102 Anomaly Agent: Unified Anomaly Retrieval and Synthesis Before Manufacturing, *Xiangyue Li, Xiaoyang Wang, Siyue Yao, Mingjie Sun, Yupei Wu*
103 S^2DiT: Sandwich Diffusion Transformer for Mobile Streaming Video Generation, *Lin Zhao, Yushu Wu, Aleksei Lebedev, Dishani Lahiri, Meng Dong, Arpit Sahni, Michael Vasilkovsky, Hao Chen, Ju Hu, Aliaksandr Siarohin, Sergey Tulyakov, Yanzhi Wang, Anil Kag, Yanyu Li*
104 UniLat3D: Geometry-Appearance Unified Latents for Single-Stage 3D Generation, *Guanjun Wu, Jiemin Fang, Chen Yang, Sikuang Li, Taoran Yi, Jia Lu, Zanwei Zhou, Jiazhong Cen, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Xinggang Wang, Qi Tian*
105 ColorMam: Color-Aware State Space Model for Image Color Style Transfer, *Jian Li, Jiaxin Peng, Yuchen Li, Siwang Zhou*
106 NumeriKontrol: Adding Numeric Control to Diffusion Transformers for Instruction-based Image Editing, *Zhenyu Xu, Xiaoqi Shen, Haotian Nan, Xinyu Zhang*
107 Towards Source-Aware Object Swapping with Initial Noise Perturbation, *Jiahui Zhan, Xianbing Sun, Xiangnan Zhu, Yikun Ji, Ruitong Liu, Liqing Zhang, Jianfu Zhang*
108 SyntheticManga: Training-Free Manga Generation with Phased Diffusion, *Xuelei Peng, Chi-Keung Tang, Yu-Wing Tai*
109 Fast Autoregressive Video Generation with Diagonal Decoding, *Yang Ye, Junliang Guo, Haoyu Wu, Tianyu He, Tim Pearce, Tabish Rashid, Katja Hofmann, Jiang Bian*
110 E-GRPO: High Entropy Steps Drive Effective Reinforcement Learning for Flow Models, *Shengjun Zhang, Zhang Zhang, Chensheng Dai, Yueqi Duan*
111 Bind-Your-Avatar: Multi-Character-Talking Video Generation with Dynamic 3D-mask-based Embedding Router, *Yubo Huang, Weiqiang Wang, Sirui Zhao, Tong Xu, Lin Liu, Enhong Chen*
112 SCAIL: Towards Studio-Grade Character Animation via In-Context Learning of 3D-Consistent Pose Representations, *Wenhao Yan, Sheng Ye, Zhuoyi Yang, Jiayan Teng, ZhenHui Dong, Kairui Wen, Xiaotao Gu, Yong-Jin Liu, Jie Tang*
113 PEDRA: Evaluating the Realism of Pedestrian Dynamics in Video Generation, *Aaron Appelle, Jerome P. Lynch*
114 VideoCanvas: Unified Video Completion from Arbitrary Spatiotemporal Patches via In-Context Conditioning, *Minghong Cai, Qiulin Wang, Zongli Ye, Wenze Liu, Quande Liu, Weicai Ye, Xintao Wang, Pengfei Wan, Kun Gai, Xiangyu Yue*
115 Jano: Adaptive Diffusion Generation with Early-Stage Convergence Awareness, *Yuyang Chen, Linqian Zeng, Yijin Zhou, Hengjie Li, Jidong Zhai*
116 Low-Bitrate Video Compression through Semantic-Conditioned Diffusion, *Lingdong Wang, Guan-Ming Su, Divya Kothandaraman, Tsung-Wei Huang, Mohammad Hajiesmaili, Ramesh K. Sitaraman*
117 Decoupled Scale-wise Autoregressive Modeling for Visual Generation, *Sucheng Ren, Yaodong Yu, Nataniel Ruiz, Feng Wang, Cihang Xie*
118 TalkVerse: Democratizing Minute-Long Audio-Driven Video Generation, *Zhenzhi Wang, Jian Wang, Ke Ma, Dahua Lin, Bing Zhou*
119 Future Optical Flow Prediction Improves Robot Control and Video Generation, *Kanchana Ranasinghe, Honglu Zhou, Yu Fang, Luyu Yang, Le Xue, Ran Xu, Caiming Xiong, Silvio Savarese, Michael S Ryoo, Juan Carlos Niebles*
120 Stepper: Stepwise Immersive Scene Generation with Multiview Panoramas, *Felix Wimbauer, Fabian Manhardt, Michael Oechsle, Nikolai Kalischek, Christian Rupprecht, Daniel Cremers, Federico Tombari*
121 Drive-Cascade: Autoregressive Occupancy to LiDAR and Video Synthesis, *Shuangming Lei, Yuehao Huang, Yao Yi, Yijia Xie, Jingke Wang, Ruoyu Wang, Jiajun Lv, Guanglin Xu, AiXue Ye, Bingbing Liu, Siyuan Cheng, Hongbo Zhang, Yukai Ma, Yong Liu*
122 ADAPT: Attention Driven Adaptive Prompt Scheduling and InTerpolating Orthogonal Complements for Rare Concepts Generation, *Kwanyoung Lee, Hyunwoo Oh, SeungJu Cha, Sungho Koh, Dong-Jin Kim*
123 Concept Erasure via Attention Redirection, *Amit Schechter, Rinon Gal, Ofir Kedem, Gal Chechik, Daniel Cohen-Or*
124 Loom: Diffusion-Transformer for Interleaved Generation, *Mingcheng Ye, Jiaming Liu, Yiren Song*
125 Rethinking Training Dynamics in Scale-Wise Autoregressive Generation, *Gengze Zhou, Chongjian Ge, Hao Tan, Feng Liu, Yicong Hong*
126 HiStream: Efficient High-Resolution Video Generation via Redundancy Eliminated Streaming, *Haonan Qiu, Shikun Liu, Zijian Zhou, Zhaochong An, Weiming Ren, Zhiheng Liu, Jonas Schult, Sen He, Shoufa Chen, Yuren Cong, Tao Xiang, Ziwei Liu, Juan-Manuel Perez-Rua*
127 Eevee: Towards Close-up High-resolution Video-based Virtual

Try-on, *Jianhao Zeng, Yancheng Bai, Ruidong Chen, Xuanpu Zhang, Lei Sun, Dongyang Jin, Ryan Xu, Nannan Zhang, Dan Song, Xiangxiang Chu*

128 Consistent Video Editing as Flow-Driven Image-to-Video Generation, *Ge Wang, Songlin Fan, Hangxu Liu, Quanjian Song, Hewei Wang, Jinfeng Xu*

129 IM-Animation: An Implicit Motion Representation for Identity-Decoupled Character Animation, *Zhufeng Xu, Xuan Gao, Feng-Lin Liu, Haoxian Zhang, Zhixue Fang, Yu-Kun Lai, Xiaoqiang Liu, Pengfei Wan, Lin Gao*

130 UniTalking: A Unified Audio-Video Framework for Talking Portrait Generation, *Hebeizi Li, Benyuan Sun, Yi Yang, Zihao Liang, Zihao Yin, Xiao Sha, Chenliang Wang*

131 Generative Visual Chain-of-Thought for Image Editing, *Zijin Yin, Tiankai Hang, Yiji Cheng, Shiyi Zhang, Runze He, Yu Xu, Chunyu Wang, Bing Li, Zheng Chang, Kongming Liang, Qinglin Lu, Zhanyu Ma*

132 UniLayDiff: A Unified Diffusion Transformer for Content-Aware Layout Generation, *Zeyang Liu, Le Wang, Sanping Zhou, Yuxuan Wu, Xiaolong Sun, Gang Hua, Haoxiang Li*

133 Blend-Aware Latent Diffusion: Mitigating Stitched Seams in Image Inpainting, *Yunpeng Liu, Xingzhong Hou, Jie Wu, Boxiao Liu, Yi Zhang, Guanglu Song, Yu Liu, Changyao Tian, Gen Luo, Haihang You*

134 One Layer Is Enough: Adapting Pretrained Visual Encoders for Image Generation, *Yuan Gao, Chen Chen, Jiatao Gu*

135 RealDiffusion: Physics-informed Attention for Multi-character Storybook Generation, *Qi Zhao, Jun Chen, Ivor Tsang, Guang Dai*

136 SwiftPie: Lightning-fast Subject-driven Image Personalization via One step Diffusion, *Huy Duong, Trong-Tung Nguyen, Cuong Pham, Anh Tran, Khoi Nguyen, Minh Hoai*

137 Video Generation Models are Good Latent Reward Models, *Xiaoyue Mi, Wenqing Yu, Jiesong Lian, Shibo Jie, Ruizhe Zhong, Zijun Liu, Guozhen Zhang, Zixiang Zhou, Zhiyong Xu, Yuan Zhou, Qinglin Lu, Fan Tang*

138 Harnessing Layered Graphic Designs with Real Intentions for Text-to-Design Generation, *Xinya Song, Bo Yang, Ying Cao*

139 VeCoR — Velocity Contrastive Regularization for Flow Matching, *Zong-Wei Hong, Jing-Lun Li, Lin-Ze Li, Shen Zhang, Yao Tang*

140 CETCam: Camera-Controllable Video Generation via Consistent and Extensible Tokenization, *Zelin Zhao, Xinyu Gong, Bangya Liu, Ziyang Song, Jun Zhang, Suhui Wu, Yongxin Chen, Hao Zhang*

141 SafetyBPO: Bidirectional Preference Optimization for Safe Text-to-Image Generation, *You Wu, Beier Zhu, Chi Zhang*

142 Fashion130K: An E-commerce Fashion Dataset for Outfit Generation with Unified Multi-modal Condition, *Yu He, Ting Zhu, Yichun Liu, Lichen Ma, Xinyuan Shan, Jingling Fu, Yu Shi, Junshi Huang, Yan Li*

143 PoseGen: In-Context LoRA Finetuning for Pose-Controllable Long Human Video Generation, *Jingxuan He, Busheng Su, Finn Wong*

144 DebFilter: Eradicating Biases Stashed in Value, *Seung Hyuk Lee, Songkuk Kim*

145 PEdit: Pareto-Guided Image Editing via Dynamic Latent Trajectory Control, *Sooyeon Park, Jaeil Park, Sung-Bae Cho*

146 Prompt-Guided Image Editing with Masked Logit Nudging in Visual Autoregressive Models, *Amir El-Ghoussani, Marc Hölle, Gustavo Carneiro, Vasileios Belagiannis*

147 Pioneering Perceptual Video Fluency Assessment: A Novel Task with Benchmark Dataset and Baseline, *Qizhi Xie, Kun Yuan, Yunpeng Qu, Ming Sun, Chao Zhou, Jihong Zhu*

148 Adapting Large VLMs with Iterative and Manual Instructions for Generative Low-light Enhancement, *Xiaoran Sun, Liyan Wang, Yeying Jin, Kin-man Lam, Zhixun Su, Yang Yang, Jinshan Pan, Cong Wang*

149 Beyond Pixel Loss: Video-INRs Prefer Perceptual Optimization, *Junqi Shi, Wuyang Cong, Ming Lu, Bowei Xu, Zhan Ma*

150 MVSSM: Motion-aware Visual State Space Model for Efficient Video Deblurring, *Chen Zhou, Tao Wu, Wei Liu, Xi Wu, Ying Fu*

151 PrismNet: Semantic-Aware Image Enhancement via Vision Transformer and Zero-Cost Gating, *Ruichen Zhang*

152 FLAIR: Frequency- and Locality-Aware Implicit Neural Representations, *Sukhun Ko, Seokhyun Youn, Dahyeon Kye, Kyle Min, Chanho Eom, Jihyong Oh*

153 CtrlISP: Rescuing Low-Light RAW Images via Controllable Neural ISP, *Chi Zhang, Yachun Li, Hang Du, Shicai Yang, Di Xie, Jiang Zhu, Yang Yang*

154 Deepfake-Agent: Aggregating Semantic Forgery Clues for Generalizable Detection, *Xiao Guo, Yue Zhang, Mohit Bansal, Xiaoming Liu*

155 How far have we gone in Generative Image Restoration? A study on its capability, limitations and evaluation practices, *Xiang Yin, Jinfan Hu, Zhiyuan You, Kainan Yan, Yu Tang, Chao Dong, Jinjin Gu*

156 PCSTracker: Long-term Scene Flow Estimation for Point Cloud Sequences, *Min Lin, Gangwei Xu, Xianqi Wang, Yuyi Peng, Xin Yang*

157 POS-ISP: Pipeline Optimization at the Sequence Level for Task-aware ISP, *Jiyun Won, Heemin Yang, Woohyeok Kim, Jungseul Ok, Sunghyun Cho*

158 Semantic-Aware Spectral Reconstruction: A Spectral Library-Aided Unsupervised Method Based on the Diffusion Model, *Keli Deng, Yuntao Qian*

159 Linear Recurrent Unit with Semantic Modulation for Image Super-Resolution, *Mingyu Choi, Woo Kyoung Han, Sunghoon Im, Kyong Hwan Jin*

160 RodNet: Visual Pathway-Inspired Adaptive Sparse Network for Efficient Low-Light Image Enhancement, *Boheng Liu, Ziyu Li, Zhong Zhang, Mengrui Xu, Chenghua Duan, Dehao Liu, Qing Li, Xia Wu*

161 LWTformer: A Detail-Aware, Learnable Wavelet-Transformer for Ancient Chinese Character Image Restoration, *Wentao Ruan, Xinhui Li, Zhan Cheng, Cunhang Fan, Libao Tian, Zhao Lv*

162 SAT: Selective Aggregation Transformer for Image Super-Resolution, *Dinh Phu Tran, Thao Do, Saad Wazir, Seongah Kim, Seon Kwon Kim, Daeyoung Kim*

163 PhyFusion: Physics-Aware Infrared and Visible Image Fusion via Modality-Specific Physical Priors, *Haiyang Jiang, Huiqin Zhang, Yanduo Zhang, Jiayi Ma, Junjun Jiang, Huabing Zhou*

164 UnfoldIR: Rethinking Deep Unfolding Network in Illumination Degradation Image Restoration, *Chunming He, Rihan Zhang, Fengyang Xiao, Chengyu Fang, Longxiang Tang, Rui Zhang, Sina Farsiu*

165 Evaluating Low-Light Image Enhancement Across Multiple Intensity Levels, *Maria Pilligua, David Serrano-Lozano, Pai Peng, Ramon Baldrich, Michael S. Brown, Javier Vazquez-Corral*

166 FALCON: Fast Adaptive Lightweight Computation of Intensities and Events for Depth Estimation, *Sankarshana Venugopal, Mohammad Mostafavi, Jonghyun Choi*

167 Learning to Translate Noise for Robust Image Denoising, *Inju Ha, Donghun Ryou, Seonguk Seo, Bohyung Han*

168 QDM: Quadtree-Based Region-Adaptive Sparse Diffusion Models for Efficient Image Super-Resolution, *Donglin Yang, Paul Vicol, Xiaojuan Qi, Renjie Liao, Xiaofan Zhang*

169 AlignVAR: Towards Globally Consistent Visual Autoregression for Image Super-Resolution, *Cencen Liu, Dongyang Zhang, Wen Yin, Jielei Wang, Tianyu Li, Ji Guo, Wenbo Jiang, Guoqing Wang, Guoming Lu*

170 Optical Tolerance-Compensated Diffusion Model for Image Restoration, *Hongji Dong, Huihui Gong, Tanli Zuo, Yu Zhao, Jin Dai, Jingduo Tian, Kai Ni*

171 TinySR: Shallow Diffusion Transformers for Real-World Image Super-Resolution, *Linwei Dong, Qingnan Fan, Yuhang Yu, Qi Zhang, Jinwei Chen, Yawei Luo, Changqing Zou*

172 Inf-Dehaze: Beyond GPU Memory Constraints for Ultra-High-Resolution Image Dehazing, *Xinyu Yan, Jiuchen Chen, Qizhi Xu*

173 DenoiseGS: Gaussian Reconstruction Model for Burst Denoising, *Yongsen Cheng, Yuanhao Cai, Yulun Zhang*

174 FlowSteer: Conditioning Flow Field for Consistent Image Restoration, *Tharindu Wickremasinghe, Chenyang Qi, Harshana Weligampola, Zhengzhong Tu, Stanley H. Chan*

175 P^2CS: Parallel Point Cloud Pre-Training with Semantic Consistency, *Linshuang Diao, Sensen Song, Yuan Jia, Yurong Qian, Dayong Ren*

176 Towards Calibrated Gradient-based Multi-Task Learning, *Linxiao Cao, Mianzimei Yang, Zhipeng Zhou, Hong Xie, Defu Lian, Menglin Yang*

177 Brain-Inspired Multimodal Spike Neural Network for Image-Text Retrieval, *Xintao Zong, Wenxuan Liu, Jianhao Ding, Zhaofei Yu, Xian Zhong, Tiejun Huang*

178 Conformal Cross-Modal Active Learning, *Huy Hoang Nguyen, Cédric Jung, Shirin Salehi, Tobias Glück, Anke Schmeink, Andreas Kugi*

179 Deep-to-Shallow Knowledge Transfer:Multi-Scale Self-Distillation with Bidirectional Aware for 3D Brain Segmentation, *Ziwei Zhang, Dayu Tan, Xin Peng, Weimin Zhong*
180 MedSAD-CLIP: Supervised CLIP with Token-Patch Cross-Attention for Medical Anomaly Detection and Segmentation, *Thuy Truong Tran, Minh Kha Do, Phuc Nguyen Duy, Min Hun Lee*
181 Rethinking Whole-Body CT Image Interpretation: An Abnormality-Centric Approach, *Ziheng Zhao, Lisong Dai, Ya Zhang, Weidi Xie, Yanfeng Wang*
182 Generative Vision-Language Multiple Instance Learning for Weakly Supervised Neonatal Fundus Screening and Reporting, *Xiao Zhang, Guangshuang Tan, Jie Hu, Shichao Kan, Bing Jiang, Yixiong Liang*
183 Mitigating Batch Effects in Histopathology via Language-Mediated Robust Embedding Generation, *Yishu Zhang, Shushan Wu, Zhenzhong Zhang, Didong Li, Huaxiu Yao, Yun Li, Iain Carmichael, Katherine A Hoadley, Hongtu Zhu, Di Wu, Daiwei Zhang*
184 PTF-CT: Polar-Aware Temporal-Frequential Iterative Reconstruction for Sparse-View CT, *Borui Kang, Guanyi Qin, Chuanpu Li, Yueming Jin*
185 Learning from Noisy Prompts: Saliency-Guided Prompt Distillation for Robust Segmentation with SAM, *Jingxuan Kang, Ziqi Zhang, Shaoming Zheng, Shuang Li, Uday Bharat Patel, Alexander Harry Fitzhugh, Phillip Lung, Yusuf Kiberu, Nikesh Jathanna, Shahnaz Jamil-Copley, Bernhard Kainz, Chen Qin*
186 Towards Noise-Robust Medical Segmentation via Chebyshev-Attention-Based Asymmetric UNet, *Yue Xin, Ziyang Zheng, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong*
187 Two-Stage 3D Pulmonary Vessel Reconstruction via Trunk--Expansion Coupled Point Cloud Generation, *Jie Zhang, Yu Xin, Guoqing Li*
188 A Simple yet Effective Data Scaling Strategy for Semi-Supervised Medical Image Segmentation, *Yajun Liu*
189 DepthScopy: Decoupling Frequency for Endoscopic Depth Estimation in Sparsely-Textured Regions, *Minghai Shi, Xiaoxian Zhang, Xiaoyue Liu, Fan Yang, Lei Li*
190 ReCliFF: Adaptive Orthogonal Decoupling for Federated Fine-tuning of Medical MLLMs, *Yuncheng Jiang, Chun-Mei Feng, Rui Sun, Le Zhang*
191 Volumetrically Consistent Implicit Atlas Learning via Neural Diffeomorphic Flow for Placenta MRI, *Athena Taymourtash, S Mazdak Abulnaga, Esra Abaci-Turk, P Ellen Grant, Polina Golland*
192 Vision-Language Models for Automated 3D PET/CT Report Generation, *Wenpei Jiao, Ke Yan, Jiajin Zhang, Dakai Jin, Zhaoheng Xie*
193 PaM-MIL: Proliferation and Metastasis Enhanced Localization for Multiple Instance Learning on Pathology Images, *Pengyu Guo, Jiachuan Wang, Zhao CHEN, Caleb Chen Cao, Liping Wang, Tingyi Jiang, Lei Chen*
194 Surgical Procedural Planning as 3D World Modelling: Towards Automated Pulmonary Resection, *Zhen Zhang, Zhaorong Dong, Xiao Yang, Liqin Huang, Qiang Wu, Taidui Zeng, Hanyu Zheng, Mingjing Yang, Shaohua Zheng, Wangbin Ding, Lin Pan*
195 From Adaptation to Generalization: Adaptive Visual Prompting for Medical Image Segmentation, *Evren Çetinkaya, Sangmin Lee, Jung Uk Kim, Hong Joo Lee, Nassir Navab*
196 AceMIL: Ordinal-Aware Multiple Instance Learning for Pathological Progression Analysis, *Shijie Li, Yiming Chen, Yingyun Gong, Hongwen Zhou, Feng-Jung Chen, Xieping Gao, Zhineng Chen*
197 PhySe-RPO: Physics and Semantics Guided Relative Policy Optimization for Diffusion-Based Surgical Smoke Removal, *Zining Fang, Cheng Xue, Chunhui Liu, Bin Xu, Ming Chen, Xiaowei Hu*
198 Anatomy-CoT: Teaching MLLMs to Reason in Radiology, *Shengzhi Wang, Kai Wu, Lei Yang, Yiwen Ye, Zihan Wang, Yuhang Wu, Mingliang Xiong, Wen Fang, Mingqing Liu, Mengyuan Xu, Hao Deng, Gang Li, Haihua Yang, Qingwen Liu*
199 DELRER: Disease Evolution-Informed Longitudinal Radiology Report Generation, *Kaiyu Wang, Bing Wang, Changchun Li, You Lu, Yaning Wang, Huimao Zhang, Ximing Li*
200 M^4Fuse: Lightweight State-Space MoE with a Cross-Scale Gating Bridge for Brain Tumor Segmentation, *Meihua Zhou, Xinyu Tong, Li Yang*
201 DynaMind: Reconstructing Dynamic Visual Scenes from EEG by Aligning Temporal Dynamics and Multimodal Semantics to Guided Diffusion, *Junxiang Liu, Junming Lin, Jie Zhou, Wei Xiong, Jiangtong Li, Jie Li, Jie Zhuang, Hongfei Ji*
202 MAE-XNT: A Foundation Model for Segmenting Neuronal Tissue Volumes Generated with X-Ray Nanotomography, *Alfred Laugros, Sebastien Roig, Alexandra Pacureanu*
203 NAKUL-Med: Spectral-Graph State Space Models with Dynamics Kernels for Medical Signals, *Badri N Patro, Vijay S Agneeswaran*
204 Gaze into the Details: Locality-Sensitive Enhancement for OCTA Retinal Vessel Segmentation, *Tuopusen Huang, Ding Ma, Xiangqian Wu*
205 M^3D-BFS: a Multi-Stage Dynamic Fusion Strategy for Sample-Adaptive Multi-Modal Brain Network Analysis, *Rui Dong, Xiaotong Zhang, Jiaxing Li, Yueying Li, Jiayin Wei, Youyong Kong*
206 Multimodal Decoupled Dynamic Graph Learning for Brain Disease Diagnosis, *Aimei Dong, Yongxing Cai, Bin Liu, Jiale Sun, Guixin Zhao*
207 TICON: A Slide-Level Tile Contextualizer for Histopathology Representation Learning, *Varun Belagali, Saarthak Kapse, Pierre Marza, Srijan Das, Zilinghan Li, Sofiène Boutaj, Pushpak Pati, Srikar Yellapragada, Tarak Nath Nandi, Ravi K Madduri, Joel Saltz, Prateek Prasanna, Stergios Christodoulidis, Maria Vakalopoulou, Dimitris Samaras*
208 TP-Seg: Task-Prototype Framework for Unified Medical Lesion Segmentation, *Jiawei Xu, Qiangqiang Zhou, Dandan Zhu, Yong Chen, Yugen Yi, Xiaoqi Zhao*
209 C3-Diff: Super-resolving Spatial Transcriptomics via Cross-modal Cross-content Contrastive Diffusion Modelling, *Xiaofei Wang, Stephen J Price, Chao Li*
210 MeMix: Multi-Encoder Mixture Framework for Medical Report Generation, *Yiming Cao, Lizhen Cui, Zhiqi Shen*
211 Learning Spatial-Preserving Hierarchical Representations for Digital Pathology, *Weiyi Wu, Xingjian Diao, Chunhui Zhang, Chongyang Gao, Xinwen Xu, Siting Li, Jiang Gui*
212 Open-Set Spatial Gene Expression Prediction from Histological Images via Retrieval-Augmented Generation, *Chaochen Wu, Meiyun Zuo, Lei Xie*
213 Personalized Functional Brain Network Modeling with Adaptive Auto-Weighted Learning for Automatic Brain Disorder Diagnosis, *Yan Zhang, Kun Liu, Min Li*
214 Do Vision Models Perceive Illusory Motion in Static Images Like Humans?, *Isabella E. Rosario, Fan L. Cheng, Zitang Sun, Nikolaus Kriegeskorte*
215 Meta-CDMTransNet: Cross-Domain Multi-Scale Transformer Meta-Learning Framework for Few-Shot Breast Histopathological Image Classification, *Anindita Mohanta, Sourav Dey Roy, Priya Saha, Niharika Nath, Mrinal Kanti Bhowmik*
216 PLCReg: Correlation-Aware Polar-Linear Attention for Guiding Medical Image Registration, *Yedi Zhang, Wenhui Huang, Yuanjie Zheng*
217 A Denoising-Enhanced Multimodal Learning Framework for Robust Nasal Endoscopy Report Generation, *Xinpan Yuan, Mingzhu Huang, Liujie Hua, Jianuo Ju, Xiaowei Zhao, Lin Yuanbo Wu*
218 When Models Learn to Ask Why: Adaptive Causal Reasoning for Trustworthy Medical Vision–Language Models, *Jianxin Lin, Chunzheng Zhu, Peter J Kneuertz, Yunfei Bai, Yuan Xue*
219 PBSBench: A Multi-Level Vision-Language Framework and Benchmark for Hematopathology Whole Slide Image Interpretation, *Yuanlong Wang, Weichi Chen, Adrian Rajab, Wenfang Liu, Yulan Jin, Andrew Srisuwananukorn, Ping Zhang*
220 Gated Differential Linear Attention: A Linear-Time Decoder for High-Fidelity Medical Segmentation, *Hongbo Zheng, Afshin Bozorgpour, Dorit Merhof, Minjia Zhang*
221 Vision-Language Models Encode Clinical Guidelines for Concept-Based Medical Reasoning, *Mohamed Harmanani, Bining Long, Zhuoxin Guo, Paul F.R. Wilson, Amirhossein Sabour, Minh Nguyen Nhat To, Gabor Fichtinger, Purang Abolmaesumi, Parvin Mousavi*
222 PGDM: Physics-Guided Noise-Free Diffusion Model Based on Point Spread Function for Light-Scattering Removal in Unpaired Biomedical Images, *Jinze Zhao, Keyi Han, Qiushi Huang, Jie Tian, Zhenhua Hu*
223 Elicit and Enhance: Advancing Multimodal Reasoning in Medical Scenarios, *Zhongzhen Huang, Linjie Mu, Yannian Gu, Kangzhe Hu, Shengyi Hua, Xiaofan Zhang*
224 Anatomy-Aware Adaptive Feature Perturbation Framework for Semi-Supervised MRI Segmentation, *Ji Lin, Bo Peng, Suping Li,*

Qianni Zhang

225 EI: Early Intervention for Multimodal Imaging Based Disease Recognition, *Qijie Wei, HaiLan Lin, Xirong Li*

226 Rethinking Medical High-Modality Learning Under Missingness — A Long-Tailed Distribution Perspective, *Chenwei Wu, Zitao Shuai, Liyue Shen*

227 HazeMatching: Dehazing Light Microscopy Images with Guided Conditional Flow Matching, *Anirban Ray, Ashesh Ashesh, Florian Jug*

228 Learning Priors via Hybrid Visual Autoregressive Modeling for Medical Image to Image Translation, *Zhaohu Xing, Hongqiu Wang, Tian Ye, Sixiang Chen, Wenxue Li, Lihao Liu, Shuaibo Li, Lei Zhu*

229 BLEG: LLM Functions as Powerful fMRI Graph-Enhancer for Brain Network Analysis, *Rui Dong, Zitong Wang, Jiaxing Li, Weihuang Zheng, Youyong Kong*

230 RelativeFlow: Taming Medical Image Denoising Learning with Noisy Reference, *Yuxin Liu, Yiqing Dong, Wenxue Yu, Zhan Wu, Rongjun Ge, Yang Chen, Yuting He*

231 UGLMM: Towards Unified Vision Grounding with Large Multimodal Model, *Xiangheng Shan, Li Zhou, Zenghui Sun, Shichao Dong, Nong Sang, Jinsong Lan, Xiaoyong Zhu, Bo Zheng, Changxin Gao, Kaifu Zhang*

232 FIRE-CIR: Fine-grained Reasoning for Composed Fashion Image Retrieval, *François Gardères, Camille-Sovanneary Gauthier, Jean Ponce, Shizhe Chen*

233 Training-Free Cross-Modal Alignment via Anchor Profiles with Statistical Significance Testing, *Kuo Yang, Jianglin Lu, Yun Fu*

234 CREM: Compression-Driven Representation Enhancement for Multimodal Retrieval and Comprehension, *Lihao Liu, Biao Yang, Yan Wang, Da Li, Jiangxia Cao, Yuxiao Luo, Xiang Chen, Xiangyu Wu, Wei Yuan, Fan Yang, Guiguang Ding, Tingting Gao, Guorui Zhou*

235 LLM Guided Multi Style Typography and Layout Generation via Dynamic Direct Preference Optimization, *Chen Fu, Shengzhou Yi, Ling Xiao, Toshihiko Yamasaki*

236 FusionBridge: An Efficient Fusion Via Feature Disentanglement for Multi-Modal Object Re-Identification, *Yali Li, Qianru Han, Xinwei He, Zhi Liu, Jinhai Xiang*

237 LlamaRG: A Multi-View Large Language Model for Radiology Report Generation, *Tanuja Jayas, Aditya Rastogi, Pavithra Raghavan, Gianluca Brugnara, Kai Schlamp, Martha Foltyn-Dumitru, Philipp Vollmuth*

238 Mitigating Information Forgetting via Entropy-Driven Progressive Retrospection for Multimodal Long Reasoning, *Yifei Gao, Ning Xu, Guoqing Jin, Shenyuan Zhang, An-An Liu*

239 InternVL-X: Advancing and Accelerating InternVL Series with Efficient Visual Token Compression, *Dongchen Lu, Zilu Zhang, Leping Huang, Yuyao Sun, Jianliang Zeng, Mao Shu, Huo Cao*

240 R²MoE: Representation and Expert Selection Dual-Regularized Mixture-of-Experts for Multimodal Clinical Data, *Wajih Hassan Raza, Mya Schiess, Juan Martinez Lemus, Timothy Michael Ellmore, Charles Green, Claudio Soto, Xin Fu, Renjie Hu*

241 DUALVISION: RGB–Infrared Multimodal Large Language Models for Robust Visual Reasoning, *Abrar Majeedi, Zhiyuan Ruan, Ziyi Zhao, Hongcheng Wang, Jianglin Lu, Yin Li*

242 Parallel In-context Learning for Large Vision Language Models, *Shin'ya Yamaguchi, Daiki Chijiwa, Tamao Sakao, Taku Hasegawa*

243 Vision-R1: Evolving Human-Free Alignment in Large Vision-Language Models via Vision-Guided Reinforcement Learning, *Yufei Zhan, Yousong Zhu, Hongyin Zhao, Fan Yang, Shurong Zheng, Ming Tang, Jinqiao Wang*

244 Prototype and Sample Level Semantic Alignment for Incomplete Multi-View Clustering, *Zhengzhong Zhu, Pei Zhou, Lanxi Bai, Jia Nie, Li Cheng, Shiquan Min, Jiangping Zhu*

245 Rethinking VLMs for Image Forgery Detection and Localization, *Shaofeng Guo, Jiequan Cui, Richang Hong*

246 DiFlowDubber: Discrete Flow Matching for Automated Video Dubbing via Cross-Modal Alignment and Synchronization, *Ngoc-Son Nguyen, Thanh V. T. Tran, Jeongsoo Choi, Hieu-Nghia Huynh-Nguyen, Truong-Son Hy, Van Nguyen*

247 OTPrune: Distribution-Aligned Visual Token Pruning Via Optimal Transport, *Xiwen Chen, Wenhui Zhu, Gen Li, Xuanzhao Dong, Yujian Xiong, Hao Wang, Peijie Qiu, Qingquan Song, Zhipeng Wang, Shao Tang, Yalin Wang, Abolfazl Razi*

248 Causal Chain-Guided Reasoning for Modular and Explainable Causal-Why Video Question Answering, *Paritosh Parmar, Eric Peh, Basura Fernando*

249 Materialistic RIR: Material Conditioned Realistic RIR Generation, *Mahnoor Fatima Saad, Sagnik Majumder, Kristen Grauman, Ziad Al-Halah*

250 From Coarse to Precise: Rethinking and Bridging Localization in Multimodal Large Language Models, *Lysa Xiao, Veronica Liesaputra, Lech Szymanski, Stephen Cranefield*

251 Do Audio-Visual Large Language Models Really See and Hear?, *Ramaneswaran Selvakumar, Kaousheik Jayakumar, S Sakshi, Sreyan Ghosh, Ruohan Gao, Dinesh Manocha*

252 DraCo: Draft as CoT for Text-to-Image Preview and Rare Concept Generation, *Dongzhi Jiang, Renrui Zhang, Haodong Li, Zhuofan Zong, Ziyu Guo, Jun He, Claire Guo, Junyan Ye, Rongyao Fang, Weijia Li, Rui Liu, Hongsheng Li*

253 FastMMoE: Accelerating Multimodal Large Language Models through Dynamic Expert Activation and Routing-Aware Token Pruning, *Guoyang Xia, Yifeng Ding, Fengfa Li, Lei Ren, Wei Chen, Fangxiang Feng, Xiaojie Wang*

254 Anticipatory Planning for Multimodal AI Agents, *Yongyuan Liang, Shijie Zhou, Yu Gu, Hao Tan, Gang Wu, Franck Dernoncourt, Jihyung Kil, Ryan A. Rossi, Ruiyi Zhang*

255 Quantifying the Gap between Understanding and Generation within Unified Multimodal Models, *Chenlong Wang, Yuhang Chen, Zhihan Hu, Dongping Chen, Wenhu Chen, Sarah Wiegreffe, Tianyi Zhou*

256 VideoScaffold: Elastic-Scale Visual Hierarchies for Streaming Video Understanding in MLLMs, *Naishan Zheng, Qingpei Guo, Jie Huang, Feng Zhao*

257 Concise Geometric Description as a Bridge: Unleashing the Potential of LLM for Plane Geometric Problem Solving, *Jingyun Wang, Dian Li, Xiaohan Wang, Gang Liu, Jiahong Yan, Guoliang Kang*

258 HippoMM: Hippocampal-inspired Multimodal Memory for Long Audiovisual Event Understanding, *Yueqian Lin, Jingyang Zhang, Qinsi Wang, Hancheng Ye, Yuzhe Fu, Yudong Liu, Hai Helen Li, Yiran Chen*

259 A Diagnostic Study of Region-Based Representations in Multimodal LLMs, *Ji Li, Shengcao Cao, Yu-Xiong Wang*

260 HoliSafe: Holistic Safety Benchmarking and Modeling for Vision-Language Model, *Youngwan Lee, Kangsan Kim, Kwanyong Park, Ilchae Jung, Soojin Jang, Seanie Lee, Yong-Ju Lee, Sung Ju Hwang*

261 UMI-HOI: Unifying Multimodal Information with Semantic Multi-Head Attention for Human–Object Interaction Detection, *Yuankai Wu, Zhinan Li, Constantin Patsch, Marsil Zakour, Driton Salihu, Eckehard Steinbach*

262 AnatomiX, an Anatomy-Aware Grounded Multimodal Large Language Model for Chest X-Ray Interpretation, *Anees Ur Rehman Hashmi, Numan Saeed, Christoph Lippert*

263 Visual2Echo Compositional Contrastive Learning (V2E-CCL): Binaural Knowledge Distilled Network for Depth Prediction, *Nazrul Ismail, Owais Ahmed Malik, Ong Wee Hong*

264 TextBind: Your Vision-Language Models are Naturally Unified Multimodal Models, *Xu Ma, Yun Fu*

265 Learning to Walk the Right Paths: Task-Responsive Graph Reasoning for Multimodal Inference, *Xuecheng Li, Weikuan Jia, Yuanjie Zheng*

266 CLASH: A Benchmark for Cross-Modal Contradiction Detection, *Teodora Popordanoska, Jiameng Li, Matthew B. Blaschko*

267 DA-CLIP: Mitigating Granularity Mismatch in Zero-Shot Anomaly Detection via Decoupled Text-Visual Alignment, *Jianqin Liu, Peng Wang, Junming Huang, Xue Zhou, Li Yu*

268 HAIT: Hybrid Adversarial Iterative Training for Mitigating Object Hallucination in Large Vision–Language Models, *Liangjie Zhao, Liao Wenjie, Ming Feng, Xiaohui Song, Huafei Li, Haonan Lu*

269 Cross-Modal-Domain Generalization Through Semantically Aligned Discrete Representations, *Souptik Sen, Raneen Younis, Zahra Ahmadi*

270 CP-IMoE: Collaborative Prompt-Guided Interactive Mixture-of-Experts for Incomplete Multimodal Learning, *Jing Li, Dongbo Zhang, Yalin Zheng, Yanda Meng*

271 Vision Inference Former: Sustaining Visual Consistency in Multimodal Large Language Models, *Xinpeng Dong, Min Zhang, Kairong Han, Xu Tan, Fei Wu, Kun Kuang*
272 PaLMR: Towards Faithful Visual Reasoning via Multimodal Process Alignment, *Yantao Li, Chenyang Yan, Qiang Hui, Fang Zhao, Kanzhi Cheng, Chao Tan, Huanlin Gao, Jianbing Zhang, Kai Wang, Xinyu Dai, Shiguo Lian*
273 If you can describe it, they can see it: Cross-Modal Learning of Visual Concepts from Textual Descriptions, *Carlo Alberto Barbano, Luca Molinaro, Massimiliano Ciranni, Emanuele Aiello, Vito Paolo Pastore, Marco Grangetto*
274 LiteEmbed: Adapting CLIP to Rare Classes, *Aishwarya Agarwal, Srikrishna Karanam, Vineet Gandhi*
275 CADReasoner: Iterative Program Editing for CAD Reverse Engineering, *Soslan Kabisov, Vsevolod Kirichuk, Andrey Volkov, Marina Barannikov, Gennadiy Savrasov, Anton Konushin, Andrey Kuznetsov, Dmitrii Zhemchuzhnikov*
276 COSTA: Collaborative Open-Set Test-Time Adaptation Through Robust Prototype Learning, *Can Zhang, Ruirui Li*
277 Perturb and Recover: Fine-Tuning for Effective Backdoor Removal from CLIP, *Naman Deep Singh, Francesco Croce, Matthias Hein*
278 PrismPrune: Decoupling Saliency and Diversity in Attention for Efficient Visual Token Pruning in VLMs, *Ziniu Liu, Shuheng Zhou, Mingqing Liu, Hao Deng, Huijia Zhu*
279 Scaling Pre-training to One Hundred Billion Data for Vision Language Models, *Xiao Wang, Ibrahim Alabdulmohsin, Daniel Salz, Zhe Li, Keran Rong, Xiaohua Zhai*
280 HAFM: A Post-Fusion Gating Module for Haze-Aware RGB–Thermal Object Detection, *Juan M. Saeteros, Nick J. Arévalo, Boris X. Vintimilla*
281 CaptAin: Caption-driven Alignment for Bridging Modality Gaps in Partially Relevant Video Retrieval, *Chuanshen Chen, Kai Zhou, Feiqi Wang, Yutao Ning, Zhendong Xiong, Yirui Li, Zhiquan Wen, Mingkui Tan*
282 Dual Anchors, Do It Better: Hierarchical Group Merging for Zero-Shot Anomaly Detection, *Jimin Roh, Dongkyu Kim, Suk-Ju Kang*
283 HeartcareGPT: A Unified Multimodal ECG Suite for Dual Signal–Image Modeling and Understanding, *Yihan Xie, Sijing Li, Zhuonan Wang, Tianwei Lin, Chenglin Yang, Yu Zhong, Wenjie Yan, Wenqiao Zhang, Xiaogang Guo, Jun Xiao, Yueting Zhuang, Beng Chin Ooi*
284 Unbiased Dynamic Multimodal Fusion, *Shicai Wei, Kaijie Zhang, Luyi Chen, Tao He, Guiduo Duan*
285 Video Reasoning Without Training, *Deepak Sridhar, Kartikeya Bhardwaj, Jeya Pradha Jeyaraj, Nuno Vasconcelos, Ankita Nayak, Harris Teague*
286 Efficient Discrete Diffusion Model for Scalable Multi-Objective Traveling Salesman Problem, *Dawei Su, Zhanhong Fang, Junyi Luo, Debing Wang, Jinbiao Chen, Zizhen Zhang*
287 EpiMask: Leveraging Epipolar Distance Based Masks in Cross-Attention for Satellite Image Matching, *Rahul Deshmukh, Aditya Chauhan, Avinash Kak*
288 S³O: Selective Spatial-Spectral Operator for Cross-Scale Fusion, *Jieyuan Pei, Wei Li, Zhuoxuan Li, Junwei Zhu, Meiyi Lu, Jiawei Jiang, Chenyu Wang, Jianwei Zheng*
289 Fast Kernel-Space Diffusion for Remote Sensing Pansharpening, *Hancong Jin, Zihan Cao, Liang-Jian Deng, Jingjing Li*
290 Unified Urban Tuning: Co-Enhancing Satellite and Street View Reasoning with a Progressive Tuning Framework, *Yong Li, Weiyu Zhang, Ling Dai, Jian Yang, Dacheng Yin, Sirun Li, Jing Lyu, Fengyun Rao, Fan Zhang*
291 GReD-RSITR: A Generative Re-Examined Discriminative Framework for Remote Sensing Image-Text Retrieval, *Shuhuai Wang, Songwei Pei, Bingfeng Liu, Yuanzhou Huang, Qian Li, Shangguang Wang*
292 ZODS-RS — Zero-Training Oriented Detection & Segmentation for Remote Sensing, *Zuan Gu, Tianhan Gao, Langxu Zhao*
293 Turning Generators into Retrievers: Unlocking MLLMs for Natural Language-Guided Geo-Localization, *Yuqi Chen, Xiaohan Zhang, Ahmad Arrabi, Waqas Sultani, Chen Chen, Safwan Wshah*
294 Optimal-Transport-based Feature Alignment for Multimodal Change Detection, *Mengqi Huang, Jun Liu, Li Cui, Yuping Duan, Faqiang Wang*
295 HarmoniDiff-RS: Training-Free Diffusion Harmonization for Satellite Image Composition, *Xiaoqi Zhuang, Jefersson A Dos Santos, Jungong Han*
296 CrossWeaver: Towards Efficient Cross-Modal Interweaving and Decoupling for Weakly-Aligned Multispectral Object Detection, *Haitian Yang, Juan Fang, Yiren Zhu, Xudong Zhao, Yufei Guo, Xiaohan Zhang, Xiaoxing Hu, Xue Yang, Qi Ming*
297 ProSM: Progressive Soft Masking for Fine-Grained Remote Image Segmentation, *Bingkun Nian, Fenghe Tang, Zhiwei Ning, Dongsheng Jiang, Yin Li, JIE Yang, Rong Xiao, Shaohua Kevin Zhou, Wei Liu*
298 UniD-Shift: Towards Unified Semantic Segmentation via Interpretable Shared–Private Multimodal Decomposition, *Shuai Zhang, Zhecheng Shi, Zhuoxiao Li, Jing Ou, Tengxi Wang, Yuan Liu, Wufan Zhao*
299 OffNadirLoc: Benchmark and Framework for Challenging UAV-to-Satellite Geo-Localization under Large Off-Nadir Views, *Qian Qiao, Wenye Liu, Ting Liu, Jiuhe Shu, Peng Wang*
300 M-PhyGs: Multi-Material Object Dynamics from Video, *Norika Wada, Kohei Yamashita, Ryo Kawahara, Ko Nishino*
301 Diffusion^2: Turning 3D Environments into Radio Frequency Heatmaps, *Kyoungjun Park, Yifan Yang, Changhan Ge, Lili Qiu, Shiqi Jiang*
302 Controllable Radar Simulation with Waveform Parameter Embedding, *Weiqing Xiao, Hao Huang, Chonghao Zhong, Yujie Lin, Nan Wang, Xiaoxue Chen, Zhaoxi Chen, Saining Zhang, Shuocheng Yang, Pierre Merriaux, Lei Lei, Hao Zhao*
303 mmDiff: A Noise-Robust Differentiable Ray-Tracing Framework for mmWave Scene Calibration and Channel Prediction, *Haofan Lu, Yadi Cao, Wanghao Yi, Omid Abari*
304 GLOW: Global Illumination-Aware Inverse Rendering of Indoor Scenes Captured with Dynamic Co-Located Light & Camera, *Jiaye Wu, Saeed Hadadan, Geng Lin, Peihan Tu, Matthias Zwicker, David Jacobs, Roni Sengupta*
305 Scene-Level Heterogeneous Physics Simulation with 3D Gaussian Splats, *Xiaoyang Liu, Shangzhe Wu, Kai Han*
306 How to Achieve Prototypical Birth and Death for OOD Detection?, *Ningkang Peng, Qianfeng Yu, Xiaoqian Peng, Linjing Qian, Yafei Liu, Canran Xiao, Xinyu Lu, Tingyu Lu, Zhichao Zheng, Yanhui Gu*
307 Uncertainty-Aware Cross-Modal Opinion Interaction: A General Frameworkfor Visible-Infrared Vehicle and Person Re-Identification, *Shihao Shan, Hongying Liu, Fanhua Shang, Qian Wang, Yang Song*
308 EIRES:Training-free AI-Generated Image Detection via Edit-Induced Reconstruction Error Shift, *Wan Jiang, Jing Yan, Xiaojing Chen, Ling Shen, Chenhao Lin, Yunfeng Diao, Richang Hong*
309 Vote-in-Context: VLMs as Explainable Zero-Shot Rank Fusers, *Mohamed Eltahir, Ali Habibullah, Lama Ayash, Tanveer Hussain, Naeemullah Khan*
310 PRADA: Probability-Ratio-Based Attribution and Detection of Autoregressive-Generated Images, *Simon Damm, Jonas Ricker, Henning Petzka, Asja Fischer*
311 HypHOI: Exploring Hierarchical Hyperbolic Embeddings for Human-Object Interaction Detection, *Yixin Guo, Yu Liu, Weimin Wang, Yanming Guo, Qi Jia*
312 A Low-Rank Learning Framework Integrating Detection, Masking, and Recovery for Occluded Facial Expression Recognition, *Yanzhong Wang, Daming Shi*
313 DSAA: Dual-Stage Attribute Activation for Fine-Grained Open Vocabulary Detection, *Donghong Jiang, Endian Lin, Hanqing Liu, Mingjie Liu, Luoping Cui, Zhao Yang, Chuang Zhu*
314 ConSel: Concept-Aware Self-supervised Learning for Regression Beyond Ordinal Tasks, *Abdullah Tariq, Bisma Saleem, R Muhammad Atif Azad, Martin Masek, Syed Zulqarnain Gilani*
315 Rolling and Denoising: Rethinking Dynamic Modal Fusion for Multi-Modal Object Re-Identification, *Shihao Li, Huaibo Huang, Aihua Zheng, Jin Tang, Ran He*
316 Adapting with an Open Mind: Leveraging Open-Vocabulary Detectors for Closed Set Source-Free Domain Adaptive Object Detection, *Kaustubh R Borgavi, Sarvesh Shashikumar, Chetan Arora*
317 SFS-DETR: Spatial-Frequency Selection for UAV Object Detection, *Dingding Jia, Jiankang Wang, Longlong Zhang, Zhiheng Liu, Xuan Wang*
318 ForenDeX: Unlocking Forensic Insights for Explainable AI-Generated Image Detection, *Chuangchuang Tan, Jinglu Wang, Xiang Ming, Renshuai Tao, Yunchao Wei, Yao Zhao, Yan Lu*
319 Long-Tailed Out-of-Distribution Detection with Refined Separate Class Learning, *Shuai Feng, Yuxin Ge, Baoming Zhang, Yuntao Du, MingCai Chen, Chongjun Wang, Lei Feng*

320 Bridging Day and Night: Unsupervised Cross-Domain Re-Identification with Synergistic Prompt and Prototype Learning, *Jiyang Xu, Rui Liu, Hang Dai*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

10:15 - 11:30 Oral Session 3A: Generative Diffusion Modeling (Bluebird Ballroom)

🏆 - Award candidate paper

1 Breaking Semantic Boundaries: Distribution-Guided Semantic Exploration for Creative Generation, *Fu Feng, Yucheng Xie, Ruixiao Shi, Xu Yang, Jing Wang, Xin Geng*
2 Guiding a Diffusion Model by Swapping Its Tokens, *Weijia Zhang, Yuehao Liu, Shanyan Guan, Wu Ran, Yanhao Ge, Wei Li, Chao Ma*
3 🏆 PixelDiT: Pixel Diffusion Transformers for Image Generation, *Yongsheng Yu, Wei Xiong, Weili Nie, Yichen Sheng, Shiqiu Liu, Jiebo Luo*
4 🏆 SeaCache: Spectral-Evolution-Aware Cache for Accelerating Diffusion Models, *Jiwoo Chung, Sangeek Hyun, MinKyu Lee, Byeongju Han, Geonho Cha, Dongyoon Wee, Youngjun Hong, Jae-Pil Heo*
5 SenCache: Accelerating Diffusion Model Inference via Sensitivity-Aware Caching, *Yasaman Haghighi, Alexandre Alahi*
6 Streaming Diffusion Model for Fast Infrared and Visible Video Fusion, *Jinyuan Liu, Ludan Sun, Tengyu Ma, Chunyan Yang, Zhiying Jiang, Long Ma, Risheng Liu, Xin Fan*

10:15 - 11:30 Oral Session 3B: Spatial Understanding (Four Seasons Ballroom)

1 🏆 ComPose: A Unified Completion-Pose Framework for Robust Category-Level Object Pose Estimation, *Huan Ren, Yihan Chen, Chuxin Wang, Nailong Liu, Wenfei Yang, Tianzhu Zhang*
2 🏆 CoSMo3D: Open-World Promptable 3D Semantic Segmentation through LLM-Guided Canonical Spatial Modeling, *Li Jin, Weikai Chen, Yujie Wang, Yingda Yin, Zeyu Hu, Runze Zhang, Keyang Luo, Shengju Qian, Xin Wang, Xueying Qin*
3 🏆 GeoViS: Geospatially Rewarded Visual Search for Remote Sensing Visual Grounding, *Peirong Zhang, Yidan Zhang, Luxiao Xu, Jinliang Lin, Zonghao Guo, Fengxiang Wang, Xue Yang, Kaiwen Wei, Lei Wang*
4 RobotSeg: A Model and Dataset for Segmenting Robots in Image and Video, *Haiyang Mei, Qiming Huang, Hai Ci, Mike Zheng Shou*
5 S^2AM3D: Scale-controllable Part Segmentation of 3D Point Clouds, *Han Su, Tianyu Huang, Zichen Wan, Xiaohe Wu, Wangmeng Zuo*
6 Scalable Multi-View Subspace Clustering with Tensorized Anchor Guidance, *Miao Jia, Xingchen Hu, Jiyuan Liu, Siwei Wang, Min Wang, Zijian Chen*

10:15 - 11:30 Oral Session 3C: Generative Editing (Mile High Ballroom 1A - 2A)

1 3D-LATTE: Latent Space 3D Editing from Textual Instructions, *Maria Parelli, Michael Oechsle, Michael Niemeyer, Federico Tombari, Andreas Geiger*
2 AnchorFlow: Training-Free 3D Editing via Latent Anchor-Aligned Flows, *Zhenglin Zhou, Fan Ma, Chengzhuo Gui, Xiaobo Xia, Hehe Fan, Yi Yang, Tat-Seng Chua*
3 🏆 ChordEdit: One-Step Low-Energy Transport for Image Editing, *Liangsi Lu, Xuhang Chen, Minzhe Guo, Shichu Li, Jingchao Wang, Yang Shi*
4 Faithful Contouring: Near-Lossless 3D Voxel Representation Free from Iso-surface, *Yihao Luo, Xianglong He, Chuanyu Pan, Yiwen Chen, Jiaqi Wu, Yangguang Li, Wanli Ouyang, Yuanming Hu, Guang Yang, ChoonHwai Yap*
5 🏆 Native and Compact Structured Latents for 3D Generation, *Jianfeng Xiang, Xiaoxue Chen, Sicheng Xu, Ruicheng Wang, Zelong Lv, Yu Deng, Hongyuan Zhu, Yue Dong, Hao Zhao, Nicholas Jing Yuan, Jiaolong Yang*
6 SliderEdit: Continuous Image Editing with Fine-Grained Instruction Control, *Arman Zarei, Samyadeep Basu, Mobina Pournemat, Sayan Nag, Ryan A. Rossi, Soheil Feizi*

10:15 - 11:30 Oral Session 3D: Multimodal Modeling (Mile High Ballroom 3A - 4A)

1 🏆 Differentiable Vector Quantization for Rate-Distortion Optimization of Generative Image Compression, *Shiyin Jiang, Wei Long, Minghao Han, Zhenghao Chen, Ce Zhu, Shuhang Gu*
2 🏆 FINER: MLLMs Hallucinate under Fine-grained Negative Queries, *Rui Xiao, Sanghwan Kim, Yongqin Xian, Zeynep Akata, Stephan Alaniz*
3 MDCS-MoAME: Multi-directional Composite Scanning with Mixture of Attention and Mamba Experts for Cancer Survival Prediction, *Linjie Qu, Jin Xiao, Xiangrong Liu, Changming Sun, Hui Cui, Yuqi Fang, Ran Su, Qiangguo Jin, Leyi Wei*
4 PAS: A Training-Free Stabilizer for Temporal Encoding in Video LLMs, *Bowen Sun, Yujun Cai, Ming-Hsuan Yang, Hang Wu, Yiwei Wang*
5 PAVAS: Physics-Aware Video-to-Audio Synthesis, *Oh Hyun-Bin, Yuhta Takida, Toshimitsu Uesaka, Tae-Hyun Oh, Yuki Mitsufuji*
6 🏆 ProPhy: Progressive Physical Alignment for Dynamic World Simulation, *Zijun Wang, Panwen Hu, Jing Wang, Terry Jingchen Zhang, Yuhao Cheng, Long Chen, Yiqiang Yan, Zutao Jiang, Hanhui Li, Xiaodan Liang*

11:45 - 13:45 Poster Session 3 & Exhibit Hall (ExHall F)

❊ - Highlight paper 🏆 - Award candidate paper

1 Breaking Semantic Boundaries: Distribution-Guided Semantic Exploration for Creative Generation, *Fu Feng, Yucheng Xie, Ruixiao Shi, Xu Yang, Jing Wang, Xin Geng*
2 Guiding a Diffusion Model by Swapping Its Tokens, *Weijia Zhang, Yuehao Liu, Shanyan Guan, Wu Ran, Yanhao Ge, Wei Li, Chao Ma*
3 🏆 PixelDiT: Pixel Diffusion Transformers for Image Generation, *Yongsheng Yu, Wei Xiong, Weili Nie, Yichen Sheng, Shiqiu Liu, Jiebo Luo*
4 🏆 SeaCache: Spectral-Evolution-Aware Cache for Accelerating Diffusion Models, *Jiwoo Chung, Sangeek Hyun, MinKyu Lee, Byeongju Han, Geonho Cha, Dongyoon Wee, Youngjun Hong, Jae-Pil Heo*
5 SenCache: Accelerating Diffusion Model Inference via Sensitivity-Aware Caching, *Yasaman Haghighi, Alexandre Alahi*
6 Streaming Diffusion Model for Fast Infrared and Visible Video Fusion, *Jinyuan Liu, Ludan Sun, Tengyu Ma, Chunyan Yang, Zhiying Jiang, Long Ma, Risheng Liu, Xin Fan*
7 🏆 ComPose: A Unified Completion-Pose Framework for Robust Category-Level Object Pose Estimation, *Huan Ren, Yihan Chen, Chuxin Wang, Nailong Liu, Wenfei Yang, Tianzhu Zhang*
8 🏆 CoSMo3D: Open-World Promptable 3D Semantic Segmentation through LLM-Guided Canonical Spatial Modeling, *Li Jin, Weikai Chen, Yujie Wang, Yingda Yin, Zeyu Hu, Runze Zhang, Keyang Luo, Shengju Qian, Xin Wang, Xueying Qin*
9 🏆 GeoViS: Geospatially Rewarded Visual Search for Remote Sensing Visual Grounding, *Peirong Zhang, Yidan Zhang, Luxiao Xu, Jinliang Lin, Zonghao Guo, Fengxiang Wang, Xue Yang, Kaiwen Wei, Lei Wang*
10 RobotSeg: A Model and Dataset for Segmenting Robots in Image and Video, *Haiyang Mei, Qiming Huang, Hai Ci, Mike Zheng Shou*
11 S^2AM3D: Scale-controllable Part Segmentation of 3D Point Clouds *Han Su, Tianyu Huang, Zichen Wan, Xiaohe Wu, Wangmeng Zuo*
12 Scalable Multi-View Subspace Clustering with Tensorized Anchor Guidance, *Miao Jia, Xingchen Hu, Jiyuan Liu, Siwei Wang, Min Wang, Zijian Chen*
13 3D-LATTE: Latent Space 3D Editing from Textual Instructions, *Maria Parelli, Michael Oechsle, Michael Niemeyer, Federico Tombari, Andreas Geiger*
14 AnchorFlow: Training-Free 3D Editing via Latent Anchor-Aligned Flows, *Zhenglin Zhou, Fan Ma, Chengzhuo Gui, Xiaobo Xia, Hehe Fan, Yi Yang, Tat-Seng Chua*
15 🏆 ChordEdit: One-Step Low-Energy Transport for Image Editing, *Liangsi Lu, Xuhang Chen, Minzhe Guo, Shichu Li, Jingchao Wang, Yang Shi*
16 Faithful Contouring: Near-Lossless 3D Voxel Representation Free from Iso-surface, *Yihao Luo, Xianglong He, Chuanyu Pan, Yiwen Chen, Jiaqi Wu, Yangguang Li, Wanli Ouyang, Yuanming Hu, Guang Yang, ChoonHwai Yap*
17 🏆 Native and Compact Structured Latents for 3D Generation, *Jianfeng Xiang, Xiaoxue Chen, Sicheng Xu, Ruicheng Wang, Zelong Lv, Yu Deng, Hongyuan Zhu, Yue Dong, Hao Zhao, Nicholas Jing Yuan, Jiaolong Yang*
18 SliderEdit: Continuous Image Editing with Fine-Grained Instruction Control, *Arman Zarei, Samyadeep Basu, Mobina Pournemat, Sayan Nag, Ryan A. Rossi, Soheil Feizi*

19 Differentiable Vector Quantization for Rate-Distortion Optimization of Generative Image Compression, *Shiyin Jiang, Wei Long, Minghao Han, Zhenghao Chen, Ce Zhu, Shuhang Gu*

20 FINER: MLLMs Hallucinate under Fine-grained Negative Queries, *Rui Xiao, Sanghwan Kim, Yongqin Xian, Zeynep Akata, Stephan Alaniz*

21 MDCS-MoAME: Multi-directional Composite Scanning with Mixture of Attention and Mamba Experts for Cancer Survival Prediction, *Linjie Qu, Jin Xiao, Xiangrong Liu, Changming Sun, Hui Cui, Yuqi Fang, Ran Su, Qiangguo Jin, Leyi Wei*

22 PAS: A Training-Free Stabilizer for Temporal Encoding in Video LLMs, *Bowen Sun, Yujun Cai, Ming-Hsuan Yang, Hang Wu, Yiwei Wang*

23 PAVAS: Physics-Aware Video-to-Audio Synthesis, *Oh Hyun-Bin, Yuhta Takida, Toshimitsu Uesaka, Tae-Hyun Oh, Yuki Mitsufuji*

24 ProPhy: Progressive Physical Alignment for Dynamic World Simulation, *Zijun Wang, Panwen Hu, Jing Wang, Terry Jingchen Zhang, Yuhao Cheng, Long Chen, Yiqiang Yan, Zutao Jiang, Hanhui Li, Xiaodan Liang*

25 V-DPM: 4D Video Reconstruction with Dynamic Point Maps, *Edgar Sucar, Eldar Insafutdinov, Zihang Lai, Andrea Vedaldi*

26 Registration-Free Learnable Multi-View Capture of Faces in Dense Semantic Correspondence, *Panagiotis P. Filntisis, George Retsinas, Radek Danecek, Vanessa Sklyarova, Petros Maragos, Timo Bolkart*

27 Mesh4D: 4D Mesh Reconstruction and Tracking from Monocular Video, *Zeren Jiang, Chuanxia Zheng, Iro Laina, Diane Larlus, Andrea Vedaldi*

28 SPE-MVS: Spatial Position Encoding Enhanced Multi-View Stereo with Monocular Depth Priors, *Shaoqian Wang, Jiadai Sun, Bosen Hou, Qiang Wang, Bin Fan, Bo Li, Bin Lu, Yuchao Dai*

29 Block-Sparse Global Attention for Efficient Multi-View Geometry Transformers, *Chung-Shien Brian Wang, Christian Schmidt, Jens Piekenbrinck, Bastian Leibe*

30 SMVRT: Implicit Human 3D Modeling Using Sparse Multi-View Volumetric Reconstruction with Transformer Fusion, *Chuanmao Fan, Chenxi Zhao, Ye Duan*

31 LiDAR Prompted Spatio-Temporal Multi-View Stereo for Autonomous Driving, *Qihao Sun, Jiarun Liu, Ziqian Ni, Jianyun Xu, Sheng Yang, Tao Xie, Lijun Zhao, Ruifeng Li*

32 Any4D: Unified Feed-Forward Metric 4D Reconstruction, *Jay Karhade, Nikhil Keetha, Yuchen Zhang, Tanisha Gupta, Akash Sharma, Sebastian Scherer, Deva Ramanan*

33 Co-Me: Confidence Guided Token Merging for Visual Geometric Transformers, *Yutian Chen, Yuheng Qiu, Ruogu Li, Jay Patrikar, Sebastian Scherer*

34 Point4Cast: Streaming Dynamic Scene Reconstruction and Forecasting, *Xinhang Liu, Pedro Miraldo, Suhas Lohit, Huaizu Jiang, Naoko Sawada, Yu-Wing Tai, Chi-Keung Tang, Moitreya Chatterjee*

35 AMB3R: Accurate Feed-forward Metric-scale 3D Reconstruction with Backend, *Hengyi Wang, Lourdes Agapito*

36 AlignPose: Generalizable 6D Pose Estimation via Multi-view Feature-metric Alignment, *Anna Šárová Mikeštíková, Médéric Fourmy, Martin Cifka, Josef Sivic, Vladimir Petrik*

37 Parallelised Differentiable Straightest Geodesics for 3D Meshes, *Hippolyte Verninas, Caner Korkmaz, Stefanos Zafeiriou, Tolga Birdal, Simone Foti*

38 Geometry-Aligned and Anomaly-Aware Reconstruction for 3D Anomaly Detection, *Linchun Wu, Qin Zou, Yuanhao Yue, Zhongyuan Wang*

39 DVGT: Driving Visual Geometry Transformer, *Sicheng Zuo, Zixun Xie, Wenzhao Zheng, Shaoqing Xu, Fang Li, Shengyin Jiang, Long Chen, Zhi-Xin Yang, Jiwen Lu*

40 FMPose3D: monocular 3D pose estimation via flow matching, *Ti Wang, Xiaohang Yu, Mackenzie Weygandt Mathis*

41 MoRE: 3D Visual Geometry Reconstruction Meets Mixture-of-Experts, *Jingnan Gao, Zhe Wang, Xianze Fang, Xingyu Ren, Zhuo Chen, Shengqi Liu, Yuhao Cheng, Jiangjing Lyu, Xiaokang Yang, Yichao Yan*

42 Foundation Encoders Are All You Need for Preference-Aware Personalization, *Hyungjin Kim, Seokho Ahn, Young-Duk Seo*

43 Where Culture Fades: Revealing the Cultural Gap in Text-to-Image Generation, *Chuancheng Shi, Shangze Li, Shiming Guo, Simiao Xie, Wenhua Wu, Jingtong Dou, Chao Wu, Canran Xiao, Cong Wang, Zifeng Cheng, Fei Shen, Tat-Seng Chua*

44 ThinkGen: Generalized Thinking for Visual Generation, *Siyu Jiao, Yiheng Lin, Yujie Zhong, Qi She, Wei Zhou, Xiaohan Lan, Zilong Huang, Fei Yu, Yingchen Yu, Yunqing Zhao, Yao Zhao, Yunchao Wei*

45 CoLoGen: Progressive Learning of Concept–Localization Duality for Unified Image Generation, *Yuxin Song, Yu Lu, Haoyuan Sun, Huanjin Yao, Fanglong Liu, Yifan Sun, Haocheng Feng, Hang Zhou, Jingdong Wang*

46 Talk2Move: Reinforcement Learning for Text-Instructed Object-Level Geometric Transformation in Scenes, *Jing Tan, Zhaoyang Zhang, Yantao Shen, Jiarui Cai, Shuo Yang, Jiajun Wu, Wei Xia, Zhuowen Tu, Stefano Soatto*

47 When Safety Collides: Resolving Multi-Category Harmful Conflicts in Text-to-Image Diffusion via Adaptive Safety Guidance, *Yongli Xiang, Ziming Hong, Zhaoqing Wang, Xiangyu Zhao, Bo Han, Tongliang Liu*

48 PSR: Scaling Multi-Subject Personalized Image Generation with Pairwise Subject-Consistency Rewards, *Shulei Wang, Longhui Wei, Xin He, Jianbo Ouyang, Hui Lu, Zhou Zhao, Qi Tian*

49 HBridge: H-Shape Bridging of Heterogeneous Experts for Unified Multimodal Understanding and Generation, *Xiang Wang, Zhifei Zhang, He Zhang, Zhe Lin, Yuqian Zhou, Qing Liu, Shiwei Zhang, Yijun Li, Shaoteng Liu, Haitian Zheng, Jason Kuen, Yuehuan Wang, Changxin Gao, Nong Sang*

50 Multimodal Semantic Bias Mitigation for Diverse Text-To-3D Generation, *Yukuan Min, Muli Yang, Jinhao Zhang, Yuxuan Wang, Yihang Zhu, Jiexi Yan, Cheng Deng*

51 Visual Personalization Turing Test, *Rameen Abdal, James Burgess, Sergey Tulyakov, Kuan-Chieh Jackson Wang*

52 Composing Concepts from Images and Videos via Concept-prompt Binding, *Xianghao Kong, Zeyu Zhang, Yuwei Guo, Zhuoran Zhao, Songchun Zhang, Anyi Rao*

53 Less is More: Data-Efficient Adaptation for Controllable Text-to-Video Generation, *Shihan Cheng, Nilesh Kulkarni, David Hyde, Dmitriy Smirnov*

54 Semantic Derivative Flow: Graph-Guided Diffusion for Controllable Instance Interactions, *Shibin Mei, Hang Wang, Bingbing Ni*

55 Improving Text-to-Image Generation with Intrinsic Self-Confidence Rewards, *Seungwook Kim, Minsu Cho*

56 Hierarchical Enhancement of Semantic Priors for Disentangled Text-Driven Motion Generation, *Wenhan Lv, Shaopan Wang, Xiangyu Wu, Tianchu Hang, Zhongquan Jian, Qingqiang Wu*

57 Simpleposter: A Simple Baseline for Product Poster Generation, *Benlei Cui, Fangao Zeng, Weitao Jiang, Yuwen Zhai, Haiwen Hong, Longtao Huang, Hui Xue, Wenxiang Shang, Pipei Huang*

58 Prompt Yourself: Awakening Textual Semantics in 1D Visual Tokenizers, *Hualiang Wang, Siming Fu, Weinan Jia, Yuning Lu, Mu Liu, Jidong Jiang, Xiaomeng Li*

59 SkyReels-Text: Fine-Grained Font-Controllable Text Editing for Poster Design, *Yunjie Yu, Jingchen Wu, Junchen Zhu, Chunze Lin, Guibin Chen*

60 Image Generation from Contextually-Contradictory Prompts, *Saar Huberman, Or Patashnik, Omer Dahary, Ron Mokady, Daniel Cohen-Or*

61 PromptEnhancer: Taming Your Rewriter for Text-to-Image Generation via Fine-Grained Reward, *Linqing Wang, Zhiyong Xu, Ximing Xing, Yiji Cheng, Zhiyuan Zhao, Donghao Li, Tiankai Hang, Zhenxi Li, Jiale Tao, Qixun Wang, Ruihuang Li, Comi Chen, Xin Li, Mingrui Wu, Xinchi Deng, Shuyang Gu, Chunyu Wang, Qinglin Lu*

62 Aligning Text, Images and 3D Structure Token-by-Token, *Aadarsh Sahoo, Vansh Tibrewal, Georgia Gkioxari*

63 RefTon: Reference person shot assist virtual Try-on, *Liuzhuozheng Li, Yue Gong, Shanyuan Liu, Zanyi Wang, Dengyang Jiang, Leibucha Wu, Bo Cheng, Yuhang Ma, Dawei Leng, Yuhui Yin*

64 GaussianVision: Vision-Language Alignment from Compressed Image Representations using 2D Gaussian Splatting, *Yasmine Omri, Connor Ding, Tsachy Weissman, Thierry Tambe*

65 Copy-Transform-Paste: Zero-Shot Object-Object Alignment Guided by Vision-Language and Geometric Constraints, *Rotem Gatenyo, Ohad Fried*

66 Gravitation-Driven Semantic Alignment for Text Video Retrieval, *Yi Yang, Zheng Wang, Xing Xu, Jingkuan Song, Heng Tao Shen*

67 MoE-GRPO: Optimizing Mixture-of-Experts via Reinforcement Learning in Vision-Language Models, *Dohwan Ko, Jinyoung Park, Seoung Choi, Sanghyeok Lee, Seohyun Lee, Hyunwoo J. Kim*

68 M^3KG-RAG: Multi-hop Multimodal Knowledge Graph-enhanced Retrieval-Augmented Generation, *Hyeongcheol Park, Jiyoung Seo, Jaewon Mun, Hogun Park, Wonmin Byeon, Sung June Kim, Hyeonsoo Im, JeungSub Lee, Sangpil Kim*

69 Evolutionary Multimodal Reasoning via Hierarchical Semantic Representation for Intent Recognition, *Qianrui Zhou, Hua Xu, Yunjin Gu, Yifan Wang, Songze Li, Hanlei Zhang*

70 ReFAct: Empowering Multimodal Web Agents with Visual and Context Focusing, *Rui Wu, Shuo Zhang, Xiaoxuan Tang, Ruirui Zhang, Yi Liu, Tao Jiang, Wenhao Xu, Yong Li*

71 PersonaVLM: Long-Term Personalized Multimodal LLMs, *Chang Nie,* ❊ *Chaoyou Fu, Yifan Zhang, Haihua Yang, Caifeng Shan*

72 MR-RAG: Multimodal Relevance-Aware Retrieval-Augmented Generation for Medical Visual Question Answering, *Xuze Li, Haozhao Wang, Zhenyu Huang, Zhongxu Wang, Jinghua Zhang, Ruixuan Li*

73 Decoupling Stability and Plasticity for Multi-Modal Test-Time Adaptation, *Yongbo He, Zirun Guo, Tao Jin*

74 CUE: Concept-Aware Multi-Label Expansion to Mitigate Concept Confusion in Long-Tailed Learning, *Ruichi Zhang, Chikai Shang, Jiacheng Yang, Mengke Li, Yang Zhou, Junlong Gao, Yang Lu*

75 Energy Waveify and Redistribution for Test-Time Adaptation: A Control System Perspective, *Zhenbin Wang, Lei Zhang, Lituan Wang, Zhenwei Zhang, Guangwu Qian, Yan Wang, Wei Huang*

76 CD-Buffer: Complementary Dual-Buffer Framework for Test-Time ❊ Adaptation in Adverse Weather Object Detection, *Youngjun Song, Hyeongyu Kim, Dosik Hwang*

77 CoFiDA-M: Concept-Aware Feature Modulation for Cross-Domain Adaptation with Image-Only Inference, *Nurjahan Sultana, Moi Hoon Yap, Xinqi Fan, Wenqi Lu*

78 Towards Multimodal Domain Generalization with Few Labels, ❊ *Hongzhao Li, Hao Dong, Hualei Wan, Shupan Li, Mingliang Xu, Muhammad Haris Khan*

79 Reclaiming Lost Text Layers for Source-Free Cross-Domain Few-Shot Learning, *Zhenyu Zhang, Guangyao Chen, Yixiong Zou, Yuhua Li, Ruixuan Li*

80 Event6D: Event-based Novel Object 6D Pose Tracking, *Jae-Young Kang, Hoonhee Cho, Taeyeop Lee, Minjun Kang, Bowen Wen, Youngho Kim, Kuk-Jin Yoon*

81 EV-CGNet: Co-visible Focused 3D-guided 2D Event Keypoint Detection Network, *Yuan Gao, Tianle Ding, Yuqing Zhu, Tianzhu Zhang*

82 AE2VID: Event-based Video Reconstruction via Aperture Modulation, *Chenxu Bai, Boyu Li, Peiqi Duan, Xinyu Zhou, Hanyue Lou, Boxin Shi*

83 From Contrast to Consistency: Rethinking Event-based Continuous- ❊ Time Optical Flow Estimation, *Rui Hu, Song Wu, Wen Yang, Jinjian Wu*

84 Spike-driven Discrete Aggregation for Event-based Object Detection, *Huaning Li, Ziming Wang, Runhao Jiang, Yan Rui, Huajin Tang*

85 x^2-Fusion: Cross-Modality and Cross-Dimension Flow Estimation in ❊ Event Edge Space, *Ruishan Guo, Ciyu Ruan, Haoyang Wang, Zihang Gong, Jingao Xu, Xinlei Chen*

86 FloVerse: Floor Plan-Guided Multi-Modal Navigation, *Weiqi Huang, Shuangyi Dong, Jiaxin Li, Yifei Guo, Zan Wang, Wei Liang*

87 TrajRAG: Retrieving Geometric-Semantic Experience for Zero-Shot Object Navigation, *Yiyao Wang, Sixian Zhang, Keming Zhang, Xinhang Song, Songjie Du, Shuqiang Jiang*

88 History to Future: Evolving Agent with Experience and Thought for Zero-shot Vision-and-Language Navigation, *Guangzhao Dai, Shuo Wang, Zihan Wang, Guo-Sen Xie, Yang Yang, Jinshan Pan, Qianru Sun, Xiangbo Shu*

89 DreamSAC: Learning Hamiltonian World Models via Symmetry Exploration, *Jinzhou Tang, Fan Feng, Minghao Fu, Wenjun Lin, Jing Yang, Biwei Huang, Keze Wang*

90 Beyond Scanpaths: Graph-Based Gaze Simulation in Dynamic Scenes, *Luke Palmer, Petar Palasek, Hazem Abdelkawy*

91 CGL: Advancing Continual GUI Learning via Reinforcement Fine-Tuning, *Zhenquan Yao, Zitong Huang, Yihan Zeng, Jianhua Han, Hang Xu, Chun-Mei Feng, Jianwei Ma, Wangmeng Zuo*

92 Rethinking Visual Rearrangement from A Diffusion Perspective, *Tianliang Qi, Xinhang Song, Yuyi Liu, Shuqiang Jiang*

93 APEX: A Decoupled Memory-based Explorer for Asynchronous Aerial Object Goal Navigation, *Daoxuan Zhang, Ping Chen, Xiaobo Xia, Xiu Su, Ruichen Zhen, Jianqiang Xiao, Shuo Yang*

94 Bridging the 2D-3D Gap: A Hierarchical Semantic-Geometric Map for Vision Language Navigation, *Kailing Li, Tianwen Qian, Lijin Yang, Yuqian Fu, Jingyu Gong, Xiaoling Wang, Liang He*

95 InterAgent: Physics-based Multi-agent Command Execution via Diffusion on Interaction Graphs, *Bin Li, Ruichi Zhang, Han Liang, Jingyan Zhang, Juze Zhang, Xin Chen, Lan Xu, Jingyi Yu, Jingya Wang*

96 When Robots Should Say "I Don't Know": Benchmarking Abstention ❊ in Embodied Question Answering, *Tao Wu, Chuhao Zhou, Guangyu Zhao, Haozhi Cao, Yewen Pu, Jianfei Yang*

97 RoboAgent: Chaining Basic Capabilities for Embodied Task Planning, *Peiran Xu, Jiaqi Zheng, Yadong Mu*

98 Towards Training-free Scene Text Editing, *Yubo Li, Xugong Qin, Peng Zhang, Hailun Lin, Gangyan Zeng, Kexin Zhang*

99 VINS-120K: Ultra High-Resolution Image Editing with A Large-Scale Dataset, *Zhizhou Chen, Shanyan Guan, Zhanxin Gao, En Ci, Yanhao Ge, Wei Li, Zhenyu Zhang, Jian Yang, Ying Tai*

100 ArtiMuse: Fine-Grained Image Aesthetics Assessment with Joint Scoring and Expert-Level Understanding, *Shuo Cao, Nan Ma, Jiayang Li, Xiaohui Li, Lihao Shao, Kaiwen Zhu, Yu Zhou, Yuandong Pu, Jiarui Wu, Jiaquan Wang, Bo Qu, Wenhai Wang, Yu Qiao, Dajuin Yao, Yihao Liu*

101 Charge: A Comprehensive Novel View Synthesis Benchmark and Dataset to Bind Them All, *Michal Nazarczuk, Thomas Tanay, Arthur Moreau, Zhensong Zhang, Eduardo Pérez-Pellitero*

102 Region-Wise Correspondence Prediction between Manga Line Art Images, *Yingxuan Li, Jiafeng Mao, Qianru Qiu, Yusuke Matsui*

103 WEAVE: Unleashing and Benchmarking the In-context Interleaved Comprehension and Generation, *Wei Chow, Jiachun Pan, Yongyuan Liang, Mingze Zhou, Xue Song, Liyu Jia, Saining Zhang, Siliang Tang, Juncheng Li, Fengda Zhang, Weijia Wu, Hanwang Zhang, Tat-Seng Chua*

104 I2I-Bench: A Comprehensive Benchmark Suite for Image-to-Image Editing Models, *Juntong Wang, Jiarui Wang, Huiyu Duan, Jiaxiang Kang, Guangtao Zhai, Xiongkuo Min*

105 TokenGS: Decoupling 3D Gaussian Prediction from Pixels with Learnable ❊ Tokens, *Jiawei Ren, Michal Jan Tyszkiewicz, Jiahui Huang, Zan Gojcic*

106 Hermite Radial Basis Function for Surface Reconstruction via ❊ Differentiable Rendering, *Hugo Blanc, Jean-Emmanuel Deschaud, Alexis Paljic*

107 RF4D:Neural Radar Fields for Novel View Synthesis in Outdoor ❊ Dynamic Scenes, *Jiarui Zhang, Zhihao Li, Chong Wang, Bihan Wen*

108 Voxify3D: Pixel Art Meets Volumetric Rendering, *Yi-Chuan Huang, Jiewen Chan, Hao-Jen Chien, Yu-Lun Liu*

109 Node-RF: Learning Generalized Continuous Space-Time Scene Dynamics with Neural ODE-based NeRFs, *Hiran Sarkar, Liming Kuang, Yordanka Velikova, Benjamin Busam*

110 FluidGaussian: Propagating Simulation-Based Uncertainty Toward Functionally-Intelligent 3D Reconstruction, *Yuqiu Liu, Jialin Song, Marissa Ramirez de Chanlatte, Rochishnu Chowdhury, Rushil Paresh Desai, Wuyang Chen, Daniel Martin, Michael W. Mahoney*

111 GaussFusion: Improving 3D Reconstruction in the Wild with A Geometry-Informed Video Generator, *Liyuan Zhu, Manjunath Narayana, Michal Stary, Will Hutchcroft, Gordon Wetzstein, Iro Armeni*

112 LagerNVS: Latent Geometry for Fully Neural Real-time Novel View ❊ Synthesis, *Stanislaw Szymanowicz, Minghao Chen, Jianyuan Wang, Christian Rupprecht, Andrea Vedaldi*

113 Turbo-GS: Accelerating 3D Gaussian Fitting for High-Resolution ❊ Radiance Fields, *Ankit Dhiman, Tao Lu, R Srinath, Emre Arslan, Angela Xing, Yuanbo Xiangli, Venkatesh Babu Radhakrishnan, Srinath Sridhar*

114 BiProLoRA: Bilevel Prompt LoRA for Real Scene Recovery, *Nan An, Long Ma, Tengyu Ma, Zhu Liu, Yingchi Liu, Risheng Liu*

115 Degradation-Consistent Test-Time Adaptation for All-in-One Image Restoration, *Ni Tang, Shenghao Nie, Xiaotong Luo, Yuan Xie, Yanyun Qu*

116 CanonCGT: Reference-Based Color Grading via Canonical Pivot Representation, *Jinwon Ko, Keunsoo Ko, Chang-Su Kim*

117 2-Shots in the Dark: Low-Light Denoising with Minimal Data Acquisition, *Liying Lu, Raphael Achddou, Sabine Süsstrunk*

118 Restore, Assess, Repeat: A Unified Framework for Iterative Image Restoration, *I-Hsiang Chen, Isma Hadji, Enrique Sanchez, Adrian Bulat, Sy-Yen Kuo, Radu Timofte, Georgios Tzimiropoulos, Brais Martinez*
119 It Takes Two: A Duet of Periodicity and Directionality for Burst Flicker Removal, *Lishen Qu, Shihao Zhou, Jie Liang, Hui Zeng, Lei Zhang, Jufeng Yang*
120 Scan Clusters, Not Pixels: A Cluster-Centric Paradigm for Efficient Ultra-high-definition Image Restoration, *Chen Wu, Ling Wang, Zhuoran Zheng, Yuning Cui, Zhixiong Yang, Xiangyu Chen, Yue Zhang, Weidong Jiang, Jingyuan Xia*
121 Seeing Beyond 8bits: Subjective and Objective Quality Assessment of HDR-UGC Videos, *Shreshth Saini, Bowen Chen, Yilin Wang, Neil Birkbeck, Balu Adsumilli, Alan C. Bovik*
122 Dynamic Exposure Burst Image Restoration, *Woohyeok Kim, Jaesung Rim, Daeyeon Kim, Sunghyun Cho*
123 FAPE-IR: Frequency-Aware Planning and Execution Framework for All-in-One Image Restoration, *Jingren Liu, Shuning Xu, Qirui Yang, Yun Wang, Xiangyu Chen, Zhong Ji*
124 ColorFLUX: A Structure-Color Decoupling Framework for Old Photo
❊ Colorization, *Bingchen Li, Zhixin Wang, Fan Li, Jiaqi Xu, Jiaming Guo, Renjing Pei, Xin Li, Zhibo Chen*
125 VEMamba: Efficient Isotropic Reconstruction of Volume Electron Microscopy with Axial-Lateral Consistent Mamba, *Longmi Gao, Pan Gao*
126 Anatomica: Localized Control over Geometric and Topological
❊ Properties for Anatomical Diffusion Models, *Karim Kadry, Abdalla Abdelwahed, Ajay Manicka, Naravich Chutisilp, Farhad R. Nezami, Elazer R. Edelman*
127 EMGauss: Continuous Slice-to-3D Reconstruction via Dynamic Gaussian Modeling in Volume Electron Microscopy, *Yumeng He, Zanwei Zhou, Yekun Zheng, Chen Liang, Yunbo Wang, Xiaokang Yang*
128 Underground Plant Exploration: Non-Destructive 3D Root Assessment with GPR Based on Point Graph Neural Network, *Yuwei Zhou, Guoyu Lu*
129 Uni-Encoder Meets Multi-Encoders: Representation Before Fusion for Brain Tumor Segmentation with Missing Modalities, *Peibo Song, Xiaotian Xue, Jinshuo Zhang, Zihao Wang, Jinhua Liu, Shujun Fu, Fangxun Bao, Si Yong Yeo*
130 MicroFM: Physics-guided Flow Matching for Isotropic Microscopy Reconstruction, *Xingzu Zhan, Runmin Jiang, Vatsal Gupta, Tanush Swaminathan, Yanwen Wang, Genpei Zhang, Haili Wang, Min Xu*
131 Dynamic Stream Network for Combinatorial Explosion Problem in Deformable Medical Image Registration, *Shaochen Bi, Yuting He, Weiming Wang, Hao Chen*
132 PMRNet: Physics-informed Multi-scale Refinement Network for Medical Image Segmentation, *Boce Kang*
133 Towards Robust Vision Transformers: Path Dependency Analysis and a Simple Two-Stage Adversarial Training, *Seongmin Kim, Byung Cheol Song*
134 PA-Attack: Guiding Gray-Box Attacks on LVLM Vision Encoders with Prototypes and Attention, *Hefei Mei, Zirui Wang, Chang Xu, Jianyuan Guo, Minjing Dong*
135 When CLIP Sees More, It Fights Back Harder: Multi-View Guided Adaptive Counterattacks for Test-Time Adversarial Robustness, *Sunoh Kim, Daeho Um*
136 Hidden Dangers of Compositional Generation: Diagnosing Semantic Safety Failures in Text-to-Image Models, *Haoming Yang, Ke Ma, Ligong Zhang, Xiaojun Jia, Yingfei Sun, Qianqian Xu, Qingming Huang*
137 VisiLock: Authorizing Instruction-based Image editing with Dual Score Distillation, *Van Thanh Le, Yun Fu*
138 JANUS: A Lightweight Framework for Jailbreaking Text-to-Image Models via Distribution Optimization, *Haolun Zheng, Yu He, Tailun Chen, Shuo Shao, Zhixuan Chu, Hongbin Zhou, Lan Tao, Zhan Qin, Kui Ren*
139 GenBreak: Red Teaming Text-to-Image Generation Using Large Language Models, *Zilong Wang, Xiang Zheng, Xiaosen Wang, Bo Wang, Xingjun Ma*
140 TUNA: Taming Unified Visual Representations for Native Unified
❊ Multimodal Models, *Zhiheng Liu, Weiming Ren, Haozhe Liu, Zijian Zhou, Shoufa Chen, Haonan Qiu, Xiaoke Huang, Zhaochong An, Fanny Yang, Aditya Patel, Viktar Atliha, Tony Ng, Xiao Han, Chuyan Zhu, Chenyang Zhang, Ding Liu, Juan-Manuel Perez-Rua, Sen He, Jürgen Schmidhuber, Wenhu Chen, Ping Luo, Wei Liu, Tao Xiang, Jonas Schult, Yuren Cong*
141 Generate, Analyze, and Refine: Training-Free Sound Source Localization via MLLM Meta-Reasoning, *Subin Park, Jung Uk Kim*
142 MMCP-GEN: A Modality-Extensible Diffusion Language Model for Conditional Protein Sequence Generation, *Zeyu An, Wanyu Lin, Feng Tan, Shujun Wang*
143 Few-shot Acoustic Synthesis with Multimodal Flow Matching, *Amandine Brunetto*
144 CLIP-like Model as a Foundational Density Ratio Estimator, *Fumiya Uchiyama, Rintaro Yanagi, Shohei Taniguchi, Shota Takashiro, Masahiro Suzuki, Hirokatsu Kataoka, Yusuke Iwasawa, Yutaka Matsuo*
145 Learning What Matters: Prioritized Concept Learning via Relative Error-driven Sample Selection, *Qian Yang, Shivam Chandhok, Oscar Mañas, Kanishk Jain, Aishwarya Agrawal, Leonid Sigal*
146 EgoAVU: Egocentric Audio-Visual Understanding, *Ashish Seth,*
❊ *Xinhao Mei, Changsheng Zhao, Varun Nagaraja, Ernie Chang, Gregory P. Meyer, Gael Le Lan, Yunyang Xiong, Vikas Chandra, Yangyang Shi, Dinesh Manocha, Zhipeng Cai*
147 Dictionary-Aligned Concept Control for Safeguarding Multimodal LLMs, *Jinqi Luo, Jinyu Yang, Tal Neiman, Lei Fan, Bing Yin, Son Tran, Mubarak Shah, René Vidal*
148 Multimodal Protein Language Models for Enzyme Kinetic Parameters: From Substrate Recognition to Conformational Adaptation, *Fei Wang, Xinye Zheng, Kun Li, Yanyan Wei, Yuxin Liu, Ganpeng Hu, Tong Bao, Jingwen Yang*
149 Echoes Over Time: Unlocking Length Generalization in Video-to-Audio Generation Models, *Christian Simon, Masato Ishii, Wei-Yao Wang, Koichi Saito, Akio Hayakawa, Dongseok Shim, Zhi Zhong, Shuyang Cui, Takashi Shibuya, Shusuke Takahashi, Yuki Mitsufuji*
150 Adaptive Confidence Regularization for Multimodal Failure Detection, *Moru Liu, Hao Dong, Olga Fink, Mario Trapp*
151 Factorize, Reconstruct, Enhance: A Unified Framework for Multimodal Sentiment Analysis, *Zhilu Yang, Mingcheng Li*
152 PhenoYieldNet: Learning Crop-Aware Phenological Responses for Multi-Crop Yield Prediction, *Yu Luo, Xiaogang Zhu, Shan Zeng, Wei Xiang, Thomas Francis Bishop, Zhiyong Wang, Kun Hu*
153 Conflict-Aware Adaptive Cross-Reconstruction for Multimodal Sentiment Analysis, *Yan Wang, Fuyuan Cao, Xingwang Zhao*
154 EduDiag: A Benchmark for Educational Diagnostic Reasoning with Error Tracing and Correction on Large Multimodal Models, *Jiali Chen, Yuqi Xue, Xusen Hei, DingBa Fu, Yuancheng Wei, Jiayuan Xie, Yi Cai*
155 UniM: A Unified Any-to-Any Interleaved Multimodal Benchmark, *Yanlin Li, Minghui Guo, Kaiwen Zhang, Shize Zhang, Yiran Zhao, Haodong Li, Congyue Zhou, Weijie Zheng, Yushen Yan, Shengqiong Wu, Wei Ji, Lei Cui, Furu Wei, Hao Fei, Mong-Li Lee, Wynne Hsu*
156 Disentangle-then-Align: Non-Iterative Hybrid Multimodal Image Registration via Cross-Scale Feature Disentanglement, *Chunlei Zhang, Jiahao Xia, Yun Xiao, Bo Jiang, Jian Zhang*
157 ChartNet: A Million-Scale, High-Quality Multimodal Dataset for Robust Chart Understanding, *Jovana Kondic, Pengyuan Li, Dhiraj Joshi, Isaac Sanchez, Ben Wiesel, Shafiq Abedin, Amit Alfassy, Eli Schwartz, Daniel Caraballo, Yagmur Gizem Cinar, Florian Scheidegger, Steven I. Ross, Daniel Karl I. Weidele, Hang Hua, Ekaterina Arutyunova, Roei Herzig, Zihan Wang, Xinyue Yu, Yunfei Zhao, Sicong Jiang, Minghao Liu, Qunshu Lin, Aude Oliva, Rogerio Feris*
158 Cross-Modal Guided Visual Synthesis for Data-Efficient Multimodal Depression Recognition, *Shanliang Yang, Xiaoxiao Wang*
159 AffordGrasp: Cross-Modal Diffusion for Affordance-Aware Grasp Synthesis, *Xiaofei Wu, Yi Zhang, Yumeng Liu, Yuexin Ma, Yujiao Shi, Xuming He*
160 PAM: A Pose–Appearance–Motion Engine for Sim-to-Real HOI Video Generation, *Mingju Gao, Kaisen Yang, Huan-ang Gao, Bohan Li, Ao Ding, Wenyi Li, Yangcheng Yu, Jinkun Liu, Shaocong Xu, Yike Niu, Haohan Chi, Hao Chen, Hao Tang, Yu Zhang, Li Yi, Hao Zhao*
161 AffordGen: Generating Diverse Demonstrations for Generalizable Object Manipulation with Affordance Correspondence, *Jiawei Zhang,*

Kaizhe Hu, Yingqian Huang, Yuanchen Ju, Zhengrong Xue, Huazhe Xu

162 HandWorld: Hand-Centric Unified Video Action Generation, *Zhihao Sun, Zhiying Du, Xitong Yang, Zuxuan Wu*

163 HVG-3D: Bridging Real and Simulation Domains for 3D-Conditional Hand-Object Interaction Video Synthesis, *Mingjin Chen, Junhao Chen, Zhaoxin Fan, Yujian Lee, Zichen Dang, Lili Wang, Yawen Cui, Lap-Pui Chau, Yi Wang*

164 ❊ ArtHOI: Taming Foundation Models for Monocular 4D Reconstruction of Hand-Articulated-Object Interactions, *Zikai Wang, Zhilu Zhang, Yiqing Wang, Hui Li, Wangmeng Zuo*

165 LAM: Language Articulated Object Modelers, *Yipeng Gao, Yunhao Ge, Peilin Cai, Daniel Seita, Laurent Itti*

166 Haptic Neural Fields: Bringing Tactile Interactions to 3D Rendered Scenes, *Antonio Luigi Stefani, Niccolò Bisagno, Nicola Conci, Eckehard Steinbach, Francesco De Natale*

167 Open-world Hand-Object Interaction Video Generation Based on Structure and Contact-aware Representation, *Haodong Yan, Hang Yu, Zhide Zhong, Weilin Yuan, Xin Gong, Zehang Luo, Chengxi Heyu, Junfeng Li, Wenxuan Song, Shunbo Zhou, Haoang Li*

168 EgoEdit: Dataset, Real-Time Streaming Model, and Benchmark for Egocentric Video Editing, *Runjia Li, Moayed Haji-Ali, Ashkan Mirzaei, Chaoyang Wang, Arpit Sahni, Ivan Skorokhodov, Aliaksandr Siarohin, Tomas Jakab, Junlin Han, Sergey Tulyakov, Philip Torr, Willi Menapace*

169 From Inpainting to Layer Decomposition: Repurposing Generative Inpainting Models for Image Layer Decomposition, *Jingxi Chen, Yixiao Zhang, Xiaoye Qian, Zongxia Li, Cornelia Fermuller, Caren Chen, Yiannis Aloimonos*

170 ❊ Temporal Equilibrium MeanFlow: Bridging the Scale Gap for One-Step Generation, *Yuanpeng Tu, Yunpeng Chen, Xinyu Zhang, Chao Liao, Hengshuang Zhao*

171 PROMO: Promptable Outfitting for Efficient High-Fidelity Virtual Try-On, *Haohua Chen, Tianze Zhou, Wei Zhu, Runqi Wang, Yandong Guan, Dejia Song, Yibo Chen, Xu Tang, Yao Hu, Lu Sheng, Zhiyong Wu*

172 Harmony: Harmonizing Audio and Video Generation through Cross-Task Synergy, *Teng Hu, Zhentao Yu, Guozhen Zhang, Zihan Su, Zhengguang Zhou, Youliang Zhang, Yuan Zhou, Qinglin Lu, Ran Yi*

173 UniSER: A Foundation Model for Unified Soft Effects Removal, *Jingdong Zhang, Lingzhi Zhang, Qing Liu, Mang Tik Chiu, Connelly Barnes, Yizhou Wang, Haoran You, Xiaoyang Liu, Yuqian Zhou, Zhe Lin, Eli Shechtman, Sohrab Amirghodsi, Xin Li, Wenping Wang, Xiaohang Zhan*

174 EffectMaker: Unifying Reasoning and Generation for Customized Visual Effect Creation, *Shiyuan Yang, Ruihuang Li, Jiale Tao, Shuai Shao, Qinglin Lu, Jing Liao*

175 ❊ Inference-time Physics Alignment of Video Generative Models with Latent World Models, *Jianhao Yuan, Xiaofeng Zhang, Felix Friedrich, Nicolas Beltran-Velez, Melissa Hall, Reyhane Askari-Hemmat, Xiaochuang Han, Nicolas Ballas, Michal Drozdzal, Adriana Romero-Soriano*

176 SMRABooth: Subject and Motion Representation Alignment for Customized Video Generation, *Xuancheng Xu, Yaning Li, Sisi You, Bing-Kun Bao*

177 Plenoptic Video Generation, *Xiao Fu, Shitao Tang, Min Shi, Xian Liu, Jinwei Gu, Ming-Yu Liu, Dahua Lin, Chen-Hsuan Lin*

178 PyramidalWan: On Making Pretrained Video Model Pyramidal for Efficient Inference, *Denis Korzhenkov, Adil Karjauv, Animesh Karnewar, Mohsen Ghafoorian, Amirhossein Habibian*

179 AdapTok: Learning Adaptive and Temporally Causal Video Tokenization in a 1D Latent Space, *Yan Li, Changyao Tian, Renqiu Xia, Ning Liao, Weiwei Guo, Hongsheng Li, Jifeng Dai, Hao Li, Xue Yang*

180 OneStory: Coherent Multi-Shot Video Generation with Adaptive Memory, *Zhaochong An, Menglin Jia, Haonan Qiu, Zijian Zhou, Xiaoke Huang, Zhiheng Liu, Weiming Ren, Kumara Kahatapitiya, Ding Liu, Sen He, Chenyang Zhang, Tao Xiang, Fanny Yang, Serge Belongie, Tian Xie*

181 Flowception: Temporally Expansive Flow Matching for Video Generation, *Tariq Berrada Ifriqi, John Nguyen, Karteek Alahari, Jakob Verbeek, Ricky T. Q. Chen*

182 Qwen-Image-Layered: Towards Inherent Editability via Layer Decomposition, *Shengming Yin, Zekai Zhang, Zecheng Tang, Kaiyuan Gao, Xiao Xu, Kun Yan, Jiahao Li, Yilei Chen, Yuxiang Chen, Heung-Yeung Shum, Lionel M. Ni, Junyang Lin, Chenfei Wu*

183 Linear Image Generation by Synthesizing Exposure Brackets, *Yuekun Dai, Zhoutong Zhang, Shangchen Zhou, Nanxuan Zhao*

184 Low-Resolution Editing is All You Need for High-Resolution Editing, *Junsung Lee, Hyunsoo Lee, Yong Jae Lee, Bohyung Han*

185 UniGenDet: A Unified Generative-Discriminative Framework for Co-Evolutionary Image Generation and Generated Image Detection, *Yanran Zhang, Wenzhao Zheng, Yifei Li, Bingyao Yu, Yu Zheng, Lei Chen, Jiwen Lu, Jie Zhou*

186 iMontage: Unified, Versatile, Highly Dynamic Many-to-many Image Generation, *Zhoujie Fu, Xianfang Zeng, Jinghong Lan, Xinyao Liao, Cheng Chen, Junyi Chen, Jiacheng Wei, Wei Cheng, Shiyu Liu, Yunuo Chen, Gang Yu, Guosheng Lin*

187 ❊ VENI: Variational Encoder for Natural Illumination, *Paul Walker, James A. D. Gardner, Andreea Ardelean, William A. P. Smith, Bernhard Egger*

188 SketchAssist: A Practical Assistant for Semantic Edits and Precise Local Redrawing, *Han Zou, Yan Zhang, Ruiqi Yu, Cong Xie, Jie Huang, Zhenpeng Zhan*

189 MultiShotMaster: A Controllable Multi-Shot Video Generation Framework, *Qinghe Wang, Xiaoyu Shi, Baolu Li, Weikang Bian, Quande Liu, Huchuan Lu, Xintao Wang, Pengfei Wan, Kun Gai, Xu Jia*

190 MoCha: End-to-End Video Character Replacement without Structural Guidance, *Zhengbo Xu, Jie Ma, Ziheng Wang, Zhan Peng, Jun Liang, Jing Li*

191 Negative Binomial Variational Autoencoders for Overdispersed Latent Modeling, *Yixuan Zhang, Jinhao Sheng, Wenxin Zhang, Quyu Kong, Feng Zhou*

192 Training-free Detection of Generated Videos via Spatial-Temporal Likelihoods, *Omer Ben Hayun, Roy Betser, Meir Yossef Levi, Levi Kassel, Guy Gilboa*

193 VOSR: A Vision-Only Generative Model for Image Super-Resolution, *Rongyuan Wu, Lingchen Sun, Zhengqiang Zhang, Xiangtao Kong, Jixin Zhao, Shihao Wang, Lei Zhang*

194 Dual Graph Regularized Deep Unfolding Network for Guided Depth Map Super-resolution, *Zhiwei Zhong, Peilin Chen, Qiangqiang Shen, Bo Li, Shiqi Wang*

195 DUO-VSR: Dual-Stream Distillation for One-Step Video Super-Resolution, *Zhengyao Lv, Menghan Xia, Xintao Wang, Kwan-Yee K. Wong*

196 VSRELL: A Simple Baseline for Video Super-Resolution and Enhancement in Low-Light Environment, *Yanming Hui, Fanhua Shang, Hongying Liu, Ben Wang, Zhenwei Zhang, Liang Wan, Wei Feng, Tong Xue, Bingqin Lv*

197 Gradient Knows Best: Mixed-Precision Quantization via Gradient-Guided Bit Allocation for Super-Resolution, *Jun Young Kim, Joo Hyeon Jeon, Sangyeon Ahn, Yoonseo Park, Yong Seok Oh, Bogyeong Kim, Sung In Cho*

198 ❊ Toward Real-world Infrared Image Super-Resolution: A Unified Autoregressive Framework and Benchmark Dataset, *Yang Zou, Jun Ma, Zhidong Jiao, Xingyuan Li, Zhiying Jiang, Jinyuan Liu*

199 Next-Scale Autoregressive Models for Text-to-Motion Generation, *Zhiwei Zheng, Shibo Jin, Lingjie Liu, Mingmin Zhao*

200 Push-and-Step: From RL-Based Balance Recovery to Physical Simulation of Dense Crowds, *Alexis Jensen, Pei Xu, Ioannis Karamouzas, Charles Pontonnier, Julien Pettré*

201 Iterative Closed-Loop Motion Synthesis for Scaling the Capabilities of Humanoid Control, *Weisheng Xu, Qiwei Wu, Jiaxi Zhang, Jing Tan, Yangfan Li, Yuetong Fang, Jiaqi Xiong, Kai Wu, Rong Ou, Renjing Xu*

202 RoMo: A Large-Scale, Richly Organized Dataset and Semantic Taxonomy for Human Motion Generation, *Jiahao Zhang, Joseph Liu, Young-Yoon Lee, Seonghyeon Moon, Victor Zordan, Guy Tevet, C. Karen Liu, Stephen Gould, Oren Jacob, Haomiao Jiang, Mubbasir Kapadia, Yizhak Ben-Shabat*

203 FrankenMotion: Part-level Human Motion Generation and Composition, *Chuqiao Li, Xianghui Xie, Yong Cao, Andreas Geiger, Gerard Pons-Moll*

204 HSI-GPT2: A Dual-Granularity Large Motion Reasoning Model with Diffusion Refinement for Human–Scene Interaction, *Yuan Wang, Xiang Li, Yali Li, Xuege Hou, Shengjin Wang*

205 SceMoS: Scene-Aware 3D Human Motion Synthesis by Planning with Geometry-Grounded Tokens, *Anindita Ghosh, Vladislav Golyanik, Taku Komura, Philipp Slusallek, Christian Theobalt, Rishabh Dabral*
206 Progressive Guessing to Fixed Point: Rethinking Human Motion Prediction with Deep Equilibrium Models, *Dong Wei, Huaijiang Sun, Fan Liu, Yuhui Zheng*
207 Archon: A Unified Multimodal Model for Holistic Digital Human Generation, *Chong Bao, Shichen Liu, Lijun Yu, David Futschik, Stylianos Moschoglou, Shefali Srivastava, Ziqian Bai, Feitong Tan, Guofeng Zhang, Zhaopeng Cui, Sean Fanello, Yinda Zhang*
208 ReMoGen: Real-time Human Interaction-to-Reaction Generation via Modular Learning from Diverse Data, *Yaoqin Ye, Yiteng Xu, Qin Sun, Xinge Zhu, Yujing Sun, Yuexin Ma*
209 Towards Motion Turing Test: Evaluating Human-Likeness in Humanoid Robots, *Mingzhe Li, Mengyin Liu, Zekai Wu, Xincheng Lin, Junsheng Zhang, Ming Yan, Zengye Xie, Changwang Zhang, Chenglu Wen, Lan Xu, Siqi Shen, Cheng Wang*
210 PatchScene: Patch-based Voxel Diffusion Model for Large-Scale Scene Completion, *Qingdong Xu, Jiajun Zhu, Shilin Zhu, Xinjing He, Chao Lu, Huanran Wang, Jiyao Zhang*
211 Prototype-Guided Concept Erasure in Diffusion Models, *Yuze Cai, Jiahao Lu, Hongxiang Shi, Yichao Zhou, Hong Lu*
212 Any2Any 3D Diffusion Models with Knowledge Transfer: A Radiotherapy Planning Study, *Yuhan Wang, Zihan Li, Han Liu, Simon Arberet, Martin Kraus, Yuyin Zhou, Florin-Cristian Ghesu, Dorin Comaniciu, Ali Kamen, Riqiang Gao*
213 CARD: Correlation Aware Restoration with Diffusion, *Niki Nezakati, Arnab Ghosh, Amit Roy-Chowdhury, Vishwanath Saragadam*
214 DMAligner: Enhancing Image Alignment via Diffusion Model Based
❊ View Synthesis, *Xinglong Luo, Ao Luo, Zhengning Wang, Yueqi Yang, Chaoyu Feng, Lei Lei, Bing Zeng, Shuaicheng Liu*
215 DRiffusion: Draft-and-Refine Process Parallelizes Diffusion Models with Ease, *Runsheng Bai, Chengyu Zhang, Yangdong Deng*
216 Do Less, Achieve More: Do We Need Every-Step Optimization for RL Fine-tuning of Diffusion Models?, *Renye Yan, Jikang Cheng, Shikun Sun, Yi Sun, You Wu, Wei Peng, Zongwei Wang, Ling Liang, Junliang Xing, Yimao Cai*
217 CSF: Black-box Fingerprinting via Compositional Semantics for Text-to-Image Models, *Junhoo Lee, Mijin Koo, Nojun Kwak*
218 InstantViR: Real-Time Video Inverse Problem Solver with Distilled
❊ Diffusion Prior, *Weimin Bai, Suzhe Xu, Yiwei Ren, Jinhua Hao, Ming Sun, Wenzheng Chen, He Sun*
219 MMTIT-Bench: A Multilingual and Multi-Scenario Benchmark with Cognition–Perception–Reasoning Guided Text-Image Machine Translation, *Gengluo Li, Chengquan Zhang, Yupu Liang, Huawen Shen, Yaping Zhang, Pengyuan Lyu, Weinong Wang, Xingyu Wan, Gangyan Zeng, Han Hu, Can Ma, Yu Zhou*
220 M3DocDep: Multi-modal, Multi-page, Multi-document Dependency Chunking with Large Vision-Language Models, *Joongmin Shin, Jeongbae Park, Jaehyung Seo, Heuiseok Lim*
221 Towards Policy-Adaptive Image Guardrail: Benchmark and Method, *Caiyong Piao, Zhiyuan Yan, Haoming Xu, Yunzhen Zhao, Kaiqing Lin, Feiyang Xu, Shuigeng Zhou*
222 Flat-Pack Bench: Evaluating Spatio-Temporal Understanding in Large Vision-Language Models through Furniture Assembly, *Aditya Chetan, Eric Cai, Peeyush Kushwaha, Bharath Raj Nagoor Kani, Utkarsh Mall, Qianqian Wang, Noah Snavely, Bharath Hariharan*
223 TextFM: Robust Semi-dense Feature Matching with Language Guidance, *Zhihao Zheng, Jinglun Feng, Nirav Savaliya, Zheng-Hang Yeh, Bo Lang, Mooi Choo Chuah*
224 What's Wrong with Synthetic Data for Scene Text Recognition? A Strong Synthetic Engine with Diverse Simulations and Self-Evolution, *Xingsong Ye, Yongkun Du, JiaXin Zhang, Chen Li, Jing LYU, Zhineng Chen*
225 Boosting Document Parsing Efficiency and Performance with Coarse-
❊ to-Fine Visual Processing, *Cheng Cui, Ting Sun, Suyin Liang, Tingquan Gao, Zelun Zhang, Jiaxuan Liu, Xueqing Wang, Changda Zhou, Hongen Liu, Manhui Lin, Yue Zhang, Yubo Zhang, Jing Zhang, Jun Zhang, Xing Wei, Yi Liu, Dianhai Yu, Yanjun Ma*
226 SJD-PAC: Accelerating Speculative Jacobi Decoding via Proactive Drafting and Adaptive Continuation, *Jialiang Kang, Han Shu, Wenshuo Li, Yingjie Zhai, Xinghao Chen*
227 Point Cloud as a Foreign Language for Multi-modal Large Language Model, *Sneha Paul, Zachary Patterson, Nizar Bouguila*
228 Grounded 3D-Aware Spatial Vision-Language Modeling, *An-Chieh Cheng, Yang Fu, Yatai Ji, Ligeng Zhu, Guanqi Zhan, Zhuoyang Zhang, Zhaojing Yang, Song Han, Yao Lu, Pavlo Molchanov, Vidya Nariyambut Murali, Jan Kautz, Xiaolong Wang, Hongxu Yin, Sifei Liu*
229 SpatialTree: How Spatial Intelligence Branches Out in MLLMs,
❊ *Yuxi Xiao, Longfei Li, Shen Yan, Xinhang Liu, Sida Peng, Yunchao Wei, Xiaowei Zhou, Bingyi Kang*
230 TerraScope: Pixel-Grounded Visual Reasoning for Earth Observation, *Yan Shu, Bin Ren, Zhitong Xiong, Xiao Xiang Zhu, Begüm Demir, Nicu Sebe, Paolo Rota*
231 Beyond 3D VQAs: Injecting 3D Spatial Priors into Vision-Language Models for Enhanced Geometric Reasoning, *Chun-Hsiao Yeh, Shengyi Qian, Manchen Wang, Yi Ma, Joseph Tighe, Fanyi Xiao*
232 OpenVoxel: Training-Free Grouping and Captioning Voxels for Open-Vocabulary 3D Scene Understanding, *Sheng-Yu Huang, Jaesung Choe, Yu-Chiang Frank Wang, Cheng Sun*
233 BOP-ASK: Object-Interaction Reasoning for Vision-Language Models, *Vineet Bhat, Sungsu Kim, Valts Blukis, Greg Heinrich, Prashanth Krishnamurthy, Ramesh Karri, Stan Birchfield, Farshad Khorrami, Jonathan Tremblay*
234 Scalable Object Relation Encoding for Better 3D Spatial Reasoning in Large Language Models, *Shengli Zhou, Minghang Zheng, Feng Zheng, Yang Liu*
235 Eliciting Complex Spatial Reasoning in MLLMs through Wide-Baseline Matching, *Hao Zhong, Muzhi Zhu, Shenyan Zeng, Anzhou Li, Cong Chen, Hua Geng, Duochao Shi, Wentao Ye, Tao Lin, Hao Chen, Chunhua Shen*
236 REALM: An MLLM-Agent Framework for Open World 3D Reasoning Segmentation and Editing on Gaussian Splatting, *Changyue Shi, Minghao Chen, Yiping Mao, Chuxiao Yang, Xinyuan Hu, Jiajun Ding, Zhou Yu*
237 From Indoor to Open World: Revealing the Spatial Reasoning Gap in MLLMs, *Mingrui Wu, Zhaozhi Wang, Fangjinhua Wang, Jiaolong Yang, Marc Pollefeys, Tong Zhang*
238 MVGGT: Multimodal Visual Geometry Grounded Transformer for Multiview 3D Referring Expression Segmentation, *Changli Wu, Haodong Wang, Jiayi Ji, Yutian Yao, Chunsai Du, Jihua Kang, Yanwei Fu, Liujuan Cao*
239 SpaceMind: Camera-Guided Modality Fusion for Spatial Reasoning in Vision-Language Models, *Ruosen Zhao, Zhikang Zhang, Jialei Xu, Jiahao Chang, Dong Chen, Lingyun Li, Weijian Sun, Zizhuang Wei*
240 ReMatch: Boosting Representation through Matching for Multimodal Retrieval, *Qianying Liu, Xiao Liang, Zhiqiang Zhang, Yibo Chen, Xu Tang, Zhongfei Qing, Fengfan Zhou, Yao Hu, Paul Henderson*
241 RI-Mamba: Rotation-Invariant Mamba for Robust Text-to-Shape Retrieval, *Khanh Nguyen, Dasith de Silva Edirimuni, Ghulam Mubashar Hassan, Ajmal Mian*
242 Revisiting F-measure Optimization in Multi-Label Classification: A Sampling-based Approach, *Zixun Wang*
243 Thinking Beyond Labels: Vocabulary-Free Fine-Grained Recognition using Reasoning-Augmented LMMs, *Dmitry Demidov, Muhammad Zaigham Zaheer, Zongyan Han, Omkar Thawakar, Rao Anwer*
244 WISER: Wider Search, Deeper Thinking, and Adaptive Fusion for Training-Free Zero-Shot Composed Image Retrieval, *Tianyue Wang, Leigang Qu, Tianyu Yang, Xiangzhao Hao, Yifan Xu, Haiyun Guo, Jinqiao Wang*
245 Modeling the Visual Ambiguity of Human Sketches, *Yang Zhou, Ping Ni, Jin Wang, Senyun Jia, Jingdan Yan, Kaixiang Huang, Guodong Lu, Jingru Yang, Shengfeng He*
246 SATTC: Structure-Aware Label-Free Test-Time Calibration for Cross-Subject EEG-to-Image Retrieval, *Qunjie Huang, Weina Zhu*
247 ConeSep: Cone-based Robust Noise-Unlearning Compositional Network for Composed Image Retrieval, *Zixu Li, Yupeng Hu, Zhiwei Chen, Mingyu Zhang, Zhiheng Fu, Liqiang Nie*
248 V^2-SAM: Marrying SAM2 with Multi-Prompt Experts for Cross-View
❊ Object Correspondence, *Jiancheng Pan, Runze Wang, Tianwen Qian,*

Mohammad Mahdi, Yanwei Fu, Xiangyang Xue, Xiaomeng Huang, Luc Van Gool, Danda Pani Paudel, Yuqian Fu

249 WeaveTime: Streaming from Earlier Frames into Emergent Memory in VideoLLMs, *Yulin Zhang, Cheng Shi, Sibei Yang*

250 Streaming Video Crime Anticipation with Spatio-Temporal Causal Reasoning, *Yusong Wang, Zheyuan Gu, Keyu Mao, Minghao Shao, Mingkun Xu, Prayag Tiwari, Jiawei Shao, Qingsong Zhao*

251 Efficient Frame Selection for Long Video Understanding via Reinforcement Learning, *Yaxuan Qin, Hefei Li, Wenqi Mu, Yancheng He*

252 HieraMamba: Video Temporal Grounding via Hierarchical Anchor-❊ Mamba Pooling, *Joungbin An, Kristen Grauman*

253 InternVideo-Next: Towards World-Understanding Video Models, *Chenting Wang, Yuhan Zhu, Yicheng Xu, Jiange Yang, Ziang Yan, Yali Wang, Yi Wang, Limin Wang*

254 Condensed Test-Time Adaptation of VLMs for Action Recognition, *Wenxuan Ge, Hongyu Qu, Rui Yan, Guo-Sen Xie, Yazhou Yao, Xiangbo Shu, Jinhui Tang*

255 Test-time Ego-Exo-centric Adaptation for Action Anticipation via Multi-Label Prototype Growing and Dual-Clue Consistency, *Zhaofeng Shi, Heqian Qiu, Lanxiao Wang, Qingbo Wu, Fanman Meng, Lili Pan, Hongliang Li*

256 A Stitch in Time: Learning Procedural Workflow via Self-Supervised Plackett–Luce Ranking, *Chengan Che, Chao Wang, Xinyue Chen, Sophia Tsoka, Luis C. Garcia-Peraza-Herrera*

257 SurgCoT: Advancing Spatiotemporal Reasoning in Surgical Videos through a Chain-of-Thought Benchmark, *Gui Wang, YongSong Zhou, Kaijun Deng, Wooi Ping Cheah, Rong Qu, Jianfeng Ren, Linlin Shen*

258 Attend Before Attention: Efficient and Scalable Video Understanding ❊ via Autoregressive Gazing, *Baifeng Shi, Stephanie Fu, Long Lian, Hanrong Ye, David Eigen, Aaron Reite, Jan Kautz, Boyi Li, David M. Chan, Trevor Darrell, Pavlo Molchanov, Hongxu Yin*

259 Concept-Guided Fine-Tuning: Steering ViTs away from Spurious ❊ Correlations to Improve Robustness, *Yehonatan Elisha, Oren Barkan, Noam Koenigstein*

260 Explaining Object Detectors via Collective Contribution of ❊ Pixels, *Toshinori Yamauchi, Hiroshi Kera, Kazuhiko Kawamoto*

261 Where MLLMs Attend and What They Rely On: Explaining Autoregressive Token Generation, *Ruoyu Chen, Xiaoqing Guo, Kangwei Liu, Siyuan Liang, Shiming Liu, Qunli Zhang, Laiyuan Wang, Hua Zhang, Xiaochun Cao*

262 H-Sets: Hessian-Guided Discovery of Set-Level Feature Interactions in Image Classifiers, *Ayushi Mehrotra, Dipkamal Bhusal, Michael Clifford, Nidhi Rastogi*

263 Evaluating Generative Models via One-Dimensional Code ❊ Distributions, *Zexi Jia, Pengcheng Luo, Yijia Zhong, Jinchao Zhang, Jie Zhou*

264 TriDF: Evaluating Perception, Detection, and Hallucination for Interpretable DeepFake Detection, *Jian-Yu Jiang-Lin, Kang-Yang Huang, Ling Zou, Ling Lo, Sheng-Ping Yang, Yu-Wen Tseng, Kun-Hsiang Lin, Chia-Ling Chen, Yu-Ting Ta, Yan-Tsung Wang, Po-Ching Chen, Hongxia Xie, Hong-Han Shuai, Wen-Huang Cheng*

265 BuildAnyPoint: 3D Building Structured Abstraction from Diverse Point Clouds, *Tongyan Hua, Haoran Gong, Yuan Liu, Di Wang, Ying-Cong Chen, Wufan Zhao*

266 LiDAR-to-4DRadar Diffusion Bridge via Cross-Modal Alignment and Translation in Latent Space, *Dazhong Shen, Jingjing Gu, Qiang Zhou, Meng Zhao, Ying Sun*

267 Edges Compete for Trust: Group Relative Edge Optimization for Building Reconstruction from Point Clouds, *Yujun Liu, Ruisheng Wang, Xiang Ao, Haoyuan Shen, Kuihao Wang, Kun Zhou, Qingquan Li*

268 Unsupervised Monocular 3D Keypoint Discovery from Multi-View Diffusion Priors, *Subin Jeon, In Cho, Junyoung Hong, Woong Oh Cho, Seon Joo Kim*

269 QD-PCQA: Quality-Aware Domain Adaptation for Point Cloud Quality Assessment, *Guohua Zhang, Jian Jin, Meiqin Liu, Chao Yao, Weisi Lin*

270 L3DR: 3D-aware LiDAR Diffusion and Rectification, *Quan Liu, Xiaoqin Zhang, Ling Shao, Shijian Lu*

271 Ghost-FWL: A Large-Scale Full-Waveform LiDAR Dataset for Ghost Detection and Removal, *Kazuma Ikeda, Ryosei Hara, Rokuto Nagata, Ozora Sako, Zihao Ding, Takahiro Kado, Ibuki Fujioka, Taro Beppu, Mariko Isogawa, Kentaro Yoshioka*

272 Ghosts in the Point Clouds: De-glaring LiDAR in the Transient Domain, *Avery Gump, Connor Henley, Sungjin Cheong, Akarsh Prabhakara, Mohit Gupta*

273 MS^2Gait: A Multi-Scale Spatio-Temporal Fusion Network for LiDAR-based Gait Recognition, *Shenyin Xu, Yishan Wang, Xinyu Li, Rui Liu, Zhongyuan Wang, Xin Tian*

274 Foundry: Distilling 3D Foundation Models for the Edge, *Guillaume Letellier, Siddharth Srivastava, Frederic Jurie, Gaurav Sharma*

275 Learning to Identify Out-of-Distribution Objects for 3D LiDAR Anomaly Segmentation, *Simone Mosco, Daniel Fusaro, Alberto Pretto*

276 Dual-Level Confidence based Implicit Self-Refinement for Medical Visual Question Answering, *Meihong Pan, Yefeng Zheng*

277 FedMPT: Federated Multi-Label Prompt Tuning of Vision-Language Models, *Xucong Wang, Pengkun Wang, Zhe Zhao, Liheng Yu, Shuang Wang, Yang Wang*

278 Rethinking Model Selection in VLM Through the Lens of Gromov-❊ Wasserstein Distance, *Muyang Li, Yucheng Liu, Jianbo Ma, Elliot Osborne, Bo Han, Tongliang Liu*

279 NTK-Guided Implicit Neural Teaching, *Chen Zhang, Wei Zuo, Bingyang Cheng, Yikun Wang, Wei-Bin Kou, Yik-Chung Wu, Ngai Wong*

280 SynthRGB-T: Language-Vision Guided Image Translation for Diversity Synthesis, *Jiangang Ding, Yiquan Du, Pengxiang Li, Lili Pei, Yuanlin Zhao, Wei Li*

281 Text-Printed Image: Bridging the Image-Text Modality Gap for Text-centric Training of Large Vision-Language Models, *Shojiro Yamabe, Futa Waseda, Daiki Shiono, Tsubasa Takahashi*

282 Harmonious Parameter Adaptation in Continual Visual Instruction Tuning for Safety-Aligned MLLMs, *Ziqi Wang, Chang Che, Qi Wang, Hui Ma, Zenglin Shi, Cees G. M. Snoek, Meng Wang*

283 StructXLIP: Enhancing Vision-language Models with Multimodal Structural Cues, *Zanxi Ruan, Songqun Gao, Qiuyu Kong, Yiming Wang, Marco Cristani*

284 Same or Not? Enhancing Visual Perception in Vision-Language ❊ Models, *Damiano Marsili, Aditya Mehta, Ryan Y. Lin, Georgia Gkioxari*

285 Vector Prism: Animating Vector Graphics by Stratifying Semantic ❊ Structure, *Jooyeol Yun, Jaegul Choo*

286 AssemblyBench: Physics-Aware Assembly of Complex Industrial Objects, *Danrui Li, Jiahao Zhang, Bernhard Egger, Moitreya Chatterjee, Suhas Lohit, Tim K. Marks, Anoop Cherian*

287 Animator-Centric Skeleton Generation on Objects with Fine-Grained Details, *Mingze Sun, Cheng Zeng, Jiansong Pei, Junhao Chen, Chaoyue Song, Shaohui Wang, Tianyuan Chang, Bin Huang, Zijiao Zeng, Ruqi Huang*

288 Synthesizing Visual Concepts as Vision-Language Programs, *Antonia Wüst, Wolfgang Stammer, Hikaru Shindo, Lukas Helff, Devendra Singh Dhami, Kristian Kersting*

289 Self-Consistency for LLM-Based Motion Trajectory Generation and Verification, *Jiaju Ma, R. Kenny Jones, Jiajun Wu, Maneesh Agrawala*

290 Semantic Scale Space: A Framework for Controllable Image Abstraction, *Kazu Mishiba*

291 Pointer-CAD: Unifying B-Rep and Command Sequences via Pointer-based Edges & Faces Selection, *Dacheng Qi, Chenyu Wang, Jingwei Xu, Tianzhe Chu, Zibo Zhao, Wen Liu, Wenrui Ding, Yi Ma, Shenghua Gao*

292 DSFlash: Comprehensive Panoptic Scene Graph Generation in Realtime, *Julian Lorenz, Vladyslav Kovganko, Elias Kohout, Mrunmai Phatak, Daniel Kienzle, Rainer Lienhart*

293 SIF: Semantically In-Distribution Fingerprints for Large Vision-Language Models, *Yifei Zhao, Qian Lou, Mengxin Zheng*

294 Designing to Forget: Deep Semi-parametric Models for ❊ Unlearning, *Amber Yijia Zheng, Yu-Shan Tai, Raymond A. Yeh*

295 Meta-FC: Meta-Learning with Feature Consistency for Robust and Generalizable Watermarking, *Yuheng Li, Weitong Chen, Chengcheng Zhu, Jiale Zhang, Chunpeng Ge, Di Wu, Guodong Long*

296 PrivSynth: Alternating and Control-Based Optimization for Privacy and Utility in Synthetic Data, *Xinyuan Zhao, Hanlin Gu, Guibao Song, Gongxi Zhu, Yifei Zou, Lixin Fan, Yuxing Han*

297 Neighbor-Aware Localized Concept Erasure in Text-to-Image Diffusion Models, *Zhuan Shi, Alireza Dehghanpour Farashah, Rik de Vries, Golnoosh Farnadi*
298 EcoAlign: An Economically Rational Framework for Efficient LVLM Alignment, *Ruoxi Cheng, Hao-Xuan Ma, Teng Ma, Hongyi Zhang*
299 Activation Matters: Test-time Activated Negative Labels for OOD Detection with Vision-Language Models, *Yabin Zhang, Maya Varma, Yunhe Gao, Jean-Benoit Delbrouck, Jiaming Liu, Chong Wang, Curtis Langlotz*
300 A Polynomial Chaos Framework for Causal Discovery in Nonlinear Uncertain Systems, *Liang Cao*
301 Domain-Skewed Federated Learning with Feature Decoupling and Calibration, *Huan Wang, Jun Shen, Jun Yan, Guansong Pang*
302 From Selection to Scheduling: Federated Geometry-Aware
❊ Correction Makes Exemplar Replay Work Better under Continual Dynamic Heterogeneity, *Zhuang Qi, Ying-Peng Tang, Lei Meng, Guoqing Chao, Lei Wu, Han Yu, Xiangxu Meng*
303 Fine-Tuning Impairs the Balancedness of Foundation Models in Long-tailed Personalized Federated Learning, *Shihao Hou, Chikai Shang, Zhiheng Yang, Jiacheng Yang, Xinyi Shang, Junlong Gao, Yiqun Zhang, Yang Lu*
304 Few-for-Many Personalized Federated Learning, *Ping Guo, Tiantian Zhang, Xi Lin, Xiang Li, Zhi-Ri Tang, Qingfu Zhang*
305 ProxyFL: A Proxy-Guided Framework for Federated Semi-Supervised Learning, *Duowen Chen, Yan Wang*
306 Domain Sensitive Federated Learning with Fisher-Informed Pruning, *Chenchen Lin, Wenhao Yuan, Zhengji Xu, Xuehe Wang*
307 SPARROW: Learning Spatial Precision and Temporal Referential Consistency in Pixel-Grounded Video MLLMs, *Mohamad Alansari, Naufal Suryanto, Divya Velayudhan, Sajid Javed, Naoufel Werghi, Muzammal Naseer*
308 Bridging Facial Understanding and Animation via Language Models, *Luchuan Song, Pinxin Liu, Haiyang Liu, Zhenchao Jin, Yolo Yunlong Tang, Zichong Xu, Susan Liang, Jing Bi, Jason J Corso, Chenliang Xu*
309 AR²-4FV: Anchored Referring and Re-identification for Long-Term Grounding in Fixed-View Videos, *Teng Yan, Yihan Liu, Jiongxu Chen, Teng Wang, Jiaqi Li, Bingzhuo Zhong*
310 CVA: Context-aware Video-text Alignment for Video Temporal Grounding, *Sungho Moon, Seunghun Lee, Jiwan Seo, Sunghoon Im*
311 OmniGround: A Comprehensive Spatio-Temporal Grounding Benchmark for Real-World Complex Scenarios, *Hong Gao, Jingyu Wu, Xiangkai Xu, Kangni Xie, Yunchen Zhang, Bin Zhong, Xurui Gao, Min-Ling Zhang*
312 ST4R-Splat: Spatio-Temporal Referring Segmentation in 4D Gaussian Splatting, *Yuming Meng, Dong Wu, Hongbin Zha*
313 WeMMU: Enhanced Bridging of Vision-Language Models and Diffusion Models via Noisy Query Tokens, *Jian Yang, Dacheng Yin, Xiaoxuan He, Yong Li, Fengyun Rao, Jing Lyu, Wei Zhai, Yang Cao, Zheng-Jun Zha*
314 Rejection Mixing: Fast Semantic Propagation of Mask Tokens for Efficient DLLM Inference, *Yushi Ye, Feng Hong, Huangjie Zheng, Xu Chen, Zhiyong Chen, Yanfeng Wang, Jiangchao Yao*
315 Towards Unified Human Perception and Machine Understanding: Token Flow Guided Compression Framework, *Li Xu, Yingfu Zhang, Kepeng Xu, Gang He, Yunsong Li*
316 A More Word-like Image Tokenization for MLLMs, *Hyun Lee, Hyemin Jeong, Yejin Kim, Hyungwook Choi, Hyunsoo Cho, Soo Kyung Kim, Joonseok Lee*
317 DUET-VLM: Dual stage Unified Efficient Token reduction for VLM Training and Inference, *Aditya Kumar Singh, Hitesh Kandala, Pratik Prabhanjan Brahma, Zicheng Liu, Emad Barsoum*
318 Unified Spatiotemporal Token Compression for Video-LLMs at Ultra-Low Retention, *Junhao Du, Jialong Xue, Anqi Li, Jincheng Dai, Guo Lu*
319 One Layer's Trash is Another Layer's Treasure: Adaptive Layer-wise
❊ Visual Token Selection in LVLMs, *Yongru Chen, Kai Zhang, Zeliang Zong, Yuchen Lu, Wenming Tan, Ye Ren, Jilin Hu*
320 OmniZip: Audio-Guided Dynamic Token Compression for Fast Omnimodal Large Language Models, *Keda Tao, Kele Shao, Bohan Yu, Weiqiang Wang, Jian Liu, Huan Wang*
321 Tunable Soft Equivariance with Guarantees, *Md Ashiqur Rahman, Lim Jun Hao, Jeremiah Jiang, Teck-Yian Lim, Raymond A. Yeh*
322 Semi-Supervised Conformal Prediction With Unlabeled Nonconformity Score, *Xuanning Zhou, Zihao Shi, Hao Zeng, Xiaobo Xia, Bingyi Jing, Hongxin Wei*
323 Cluster-aware Anchor Learning for Multi-View Clustering, *Zhe Chen, Fanhui Meng, Tianyang Xu, Xiao-Jun Wu*
324 Revisiting Sparsity Constraint Under High-Rank Property in Partial Multi-Label Learning, *Chongjie Si, Yidan Cui, Fuchao Yang, Wei Shen*
325 Weight Space Representation Learning via Neural Field Adaptation, *Zhuoqian Yang, Mathieu Salzmann, Sabine Süsstrunk*
326 Recurrent Video Masked Autoencoders, *Daniel Zoran, Nikhil Parthasarathy, Yi Yang, Drew A Hudson, João Carreira, Andrew Zisserman*
327 Revisiting Unknowns: Towards Effective and Efficient Open-Set Active Learning, *Chen-Chen Zong, Yu-Qi Chi, Xie-Yang Wang, Yan Cui, Sheng-Jun Huang*
328 Seeing Through the Shift: Causality-Inspired Robust Generalized Category Discovery, *Wei Feng, Yiwen Jiang, Sijin Zhou, Zhuang Qi, Zhongxing Xu, Zhonghua Wang, Feilong Tang, Zongyuan Ge*
329 From Exploration to Exploitation: A Two-Stage Entropy RLVR Approach for Noise-Tolerant MLLM Training, *Donglai Xu, Hongzheng Yang, Yuzhi Zhao, Pingping Zhang, Jinpeng Chen, Wenao Ma, Zhijian Hou, Mengyang Wu, Xiaolei Li, Senkang Hu, Ziyi Guan, Jason Chun Lok Li, Lai-Man Po*
330 Spatial Retrieval Augmented Autonomous Driving, *Xiaosong Jia, Chenhe Zhang, Yule Jiang, Songbur Wong, Zhiyuan Zhang, Chen Chen, Shaofeng Zhang, Xuanhe Zhou, Xue Yang, Junchi Yan, Yu-Gang Jiang*
331 Scaling-Aware Data Selection for End-to-End Autonomous Driving
❊ Systems, *Tolga Dimlioglu, Nadine Chang, Maying Shen, Rafid Mahmood, Jose M. Alvarez*
332 ColaVLA: Leveraging Cognitive Latent Reasoning for Hierarchical Parallel Trajectory Planning in Autonomous Driving, *Qihang Peng, Xuesong Chen, Chenye Yang, Shaoshuai Shi, Hongsheng Li*
333 CARD: A Multi-Modal Automotive Dataset for Dense 3D Reconstruction
❊ in Challenging Road Topography, *Gasser Elazab, Frank Neuhaus, Tilman Koß, Malte Splietker, Aditya Date, Michael Unterreiner, Maximilian Jansen, Olaf Hellwich*
334 MindDriver: Introducing Progressive Multimodal Reasoning for Autonomous Driving, *Lingjun Zhang, Yujian Yuan, Changjie Wu, Xinyuan Chang, Xin Cai, Shuang Zeng, Linzhe Shi, Sijin Wang, Hang Zhang, Mu Xu*
335 WPT: World-to-Policy Transfer via Online World Model Distillation, *Guangfeng Jiang, Yueru Luo, Jun Liu, Yi Huang, Yiyao Zhu, Zhan Qu, Dave Zhenyu Chen, Bingbing Liu, Xu Yan*
336 ClimaOoD: Improving Anomaly Segmentation via Physically Realistic Synthetic Data, *Yuxing Liu, Zheng Li, Huanhuan Liang, Ji Zhang, Zeyu Sun, Yong Liu*
337 Recover to Predict: Progressive Retrospective Learning for Variable-
❊ Length Trajectory Prediction, *Hao Zhou, Lu Qi, Xiangtai Li, Jie Zhang, Yi Liu, Xu Yang, Mingyu Fan, Fei Luo*
338 URScenes: A Multi-scenario Dataset for Unstructured Road Environments, *Runsen Liu, Aizemaitijiang Baoerhan, Zhangyu Wang, Jie Wang, Jinghao Cui, Guizhen Yu, Songyue Yang, WanCheng Sun, Mingjun Tang, Zhanbo Hua, Wenwen Luo*
339 MeanFuser: Fast One-Step Multi-Modal Trajectory Generation and Adaptive Reconstruction via MeanFlow for End-to-End Driving, *Junli Wang, Yinan Zheng, Xueyi Liu, Zebin Xing, Pengfei Li, Kun Ma, Hangjun Ye, Guang Chen, Guang Li, Long Chen, Zhongpu Xia, Qichao Zhang*
340 SAMosaic3D: Modular Scene Assembly for Real-Time 3D Segment Anything, *Peng Wang, Yongcai Wang, Wang Chen, Hualong Cao, Kang Yang, Chunxu Li, Jie Wen, Deying Li*
341 Mitigating Objectness Bias and Region-to-Text Misalignment for
❊ Open-Vocabulary Panoptic Segmentation, *Nikolay Kormushev, Josip Šarić, Matej Kristan*
342 MV3DIS: Multi-View Mask Matching via 3D Guides for Zero-Shot 3D Instance Segmentation, *Yibo Zhao, Yigong Zhang, Jin Xie*
343 PEARL: Geometry Aligns Semantics for Training-Free Open-Vocabulary Semantic Segmentation, *Gensheng Pei, Xiruo Jiang, Xinhao Cai, Tao Chen, Yazhou Yao, Byeungwoo Jeon*

344 RAVEN: Radar Adaptive Vision Encoders for Efficient Chirp-wise ❊ Object Detection and Segmentation, *Anuvab Sen, Mir Sayeed Mohammad, Saibal Mukhopadhyay*
345 SAMIX: Reinforcing SAM2 with Semantic Adapter and Reference Selecting Policy for Mix-Supervised Segmentation, *Qiang Hu, Jiajie Wei, Zhenyu Yi, Zhifen Yan, Yingjie Guo, Hongkuan Shi, Ge-Peng Ji, Qiang Li, Zhiwei Wang*
346 MARSS: Radar Semantic Segmentation via Modular Attention and State Space Models, *Fengyu Chen, Tiao Tan, Teng Li, Yuantian Quan, Qingmin Liao*
347 MixerCSeg: An Efficient Mixer Architecture for Crack Segmentation via Decoupled Mamba Attention, *Zilong Zhao, Zhengming Ding, Pei Niu, Wenhao Sun, Feng Guo*
348 Exemplar-Free Class Incremental Learning via Preserving Class- ❊ Discriminative Structure, *Xin Zhang, Liang Bai, Guanchao Wang, Xian Yang*
349 Critical Patch-Aware Sparse Prompting with Decoupled Training for Continual Learning on the Edge, *Wonseon Lim, Jaesung Lee, Dae-Won Kim*
350 PACT: Phase-Like Transition Constraints in Adapter-Based Continual Learning of Vision-Language Models, *Xuan Wang, Guiguang Ding, Jungong Han*
351 Representation-Steered Incremental Adapter-Tuning for Class-Incremental Learning with Pre-Trained Models, *Jiarui Zhao, Libo Huang, Xiangqi Li, Zhulin An, Chuanguang Yang, Yu Wang, Boyu Diao, Yongjun Xu*
352 Re-evaluating Continual VQA: Toward Fair and Robust Evaluation for Multimodal Continual Learning, *Zijian Gao, Zicheng Sun, Xingxing Zhang, Kele Xu, Huaimin Wang*
353 Distilling Balanced Knowledge from a Biased Teacher, *Seonghak Kim*
354 Enhancing Continual Learning of Vision-Language Models via Dynamic Prefix Weighting, *Hyeonseo Jang, Hyuk Kwon, Kibok Lee*
355 Beyond Myopic Alignment: Lookahead Optimization for Online Class-Incremental Learning, *Song Lai, Zhe Zhao, Fei Zhu, Ji Cheng, Xi Lin, Qingfu Zhang, Gaofeng Meng*
356 EmoDiffTalk: Emotion-aware Diffusion for Editable 3D Gaussian Talking Head, *Chang Liu, Tianjiao Jing, Chengcheng Ma, Xuanqi Zhou, Zhengxuan Lian, Qin Jin, Hongliang Yuan, Shi-Sheng Huang*
357 Avatar Forcing: Real-Time Interactive Head Avatar Generation for Natural Conversation, *Taekyung Ki, Sangwon Jang, Jaehyeong Jo, Jaehong Yoon, Sung Ju Hwang*
358 D^3FER: Dual Channel and Dual Branch Network for Robust Facial Expression Recognition under Dual Challenges, *Hui Tang, Yifan He, Zhong Jin*
359 HumanNOVA: Photorealistic, Universal and Rapid 3D Human Avatar ❊ Modeling from a Single Image, *Hezhen Hu, Wangbo Zhao, Lanqing Guo, Hanwen Jiang, Jonathan C. Liu, Zhiwen Fan, Kai Wang, Zhangyang Wang, Georgios Pavlakos*
360 ExpPortrait: Expressive Portrait Generation via Personalized Representation, *Junyi Wang, Yudong Guo, Boyang Guo, Shengming Yang, Juyong Zhang*
361 PersonaLive! Expressive Portrait Image Animation for Live Streaming, *Zhiyuan Li, Chi-Man Pun, Chen Fang, Jue Wang, Xiaodong Cun*
362 ProFocus: Proactive Perception and Focused Reasoning in Vision-and-Language Navigation, *Wei Xue, Mingcheng Li, Xuecheng Wu, Jingqun Tang, Dingkang Yang, Lihua Zhang*
363 OptiMVMap: Offline Vectorized Map Construction via Optimal Multi-vehicle Perspectives, *Zedong Dan, Zijie Wang, Wei Zhang, Xiangru Lin, Weiming Zhang, Xiao Tan, Jingdong Wang, Liang Lin, Guanbin Li*
364 CogDriver: Integrating Cognitive Inertia for Temporally Coherent Planning in Autonomous Driving, *Pei Liu, Qingtian Ning, Xinyan Lu, Haipeng Liu, Weiliang Ma, Dangen She, Xianpeng Lang, Jun Ma*
365 TopoHR: Hierarchical Centerline Representation for Cyclic Topology Reasoning in Driving Scenes with Point-to-Instance Relations, *Yifeng Bai, Zhirong Chen, Bo Song, Erkang Cheng, Haibin Ling*
366 AURA: Multi-modal Shared Autonomy for Urban Navigation, *Yukai Ma, Honglin He, Selina Song, Wayne Wu, Bolei Zhou*
367 Zero-Shot Reconstruction of Animatable 3D Avatars with Cloth Dynamics from a Single Image, *Joohyun Kwon, Geonhee Sim, Gyeongsik Moon*
368 FlexAvatar: Learning Complete 3D Head Avatars with Partial Supervision, *Tobias Kirschstein, Simon Giebenhain, Matthias Nießner*
369 Large-scale Codec Avatars: The Unreasonable Effectiveness of Large-scale Avatar Pretraining, *Junxuan Li, Rawal Khirodkar, Egor Zakharov, Jihyun Lee, Zhaoen Su, Yuan Dong, Julieta Martinez, Kai Li, Qingyang Tan, Takaaki Shiratori, Matthew Hu, Peihong Guo, Xuhua Huang, Zhongshi Jiang, Lingchen Yang, Ariyan Zarei, Marco Pesavento, Yichen Xu, Chengan He, He Wen, Giljoo Nam, Teng Deng, Wyatt Borsos, Anjali Thakrar, Jean-Charles Bazin, Rinat Abdrashitov, Carsten Stoll, Ginés Hidalgo, James Booth, Lucy Wang, Xiaowen Ma, Yu Rong, Sairanjith Thalanki, Chen Cao, Christian Häne, Abhishek Kar, Sofien Bouaziz, Jason Saragih, Yaser Sheikh, Shunsuke Saito*
370 UIKA: Fast Universal Head Avatar from Pose-Free Images, *Zijian Wu,* ❊ *Boyao Zhou, Liangxiao Hu, Hongyu Liu, Yuan Sun, Xuan Wang, Xun Cao, Yujun Shen, Hao Zhu*
371 FlexAvatar: Flexible Large Reconstruction Model for Animatable Gaussian Head Avatars with Detailed Deformation, *Cheng Peng, Zhuo Su, Liao Wang, Chen Guo, Zhaohu Li, Chengjiang Long, Zheng Lv, Jingxiang Sun, Chenyangguang Zhang, Yebin Liu*
372 First Logit Boosting: Visual Grounding Method to Mitigate Object Hallucination in Large Vision-Language Models, *Jiwoo Ha, Jongwoo Baek, Jinhyun So*
373 Locate-then-Sparsify: Attribution Guided Sparse Strategy for Visual Hallucination Mitigation, *Tiantian Dang, Chao Bi, Shufan Shen, Jinzhe Liu, Qingming Huang, Shuhui Wang*
374 Envision, Attend, Then Respond: Counterfactual Hallucination Mitigation in Large Vision-Language Models, *Yuxuan Liang, Fan Shi, Rui Zhu, Xu Li, Xiaolei Chen, Zhe Liu, Bin Li, Xiangyang Xue*
375 PAS: Prelim Attention Score for Detecting Object Hallucinations in Large Vision-Language Models, *Nhat Hoang, Minh Vu, My T. Thai, Manish Bhattarai*
376 MoD-DPO: Towards Mitigating Cross-modal Hallucinations in Omni LLMs using Modality Decoupled Preference Optimization, *Ashutosh Chaubey, Jiacheng Pang, Mohammad Soleymani*
377 Fine-Grained Multi Image Object Hallucination Benchmark, *Joonki Min, Chaeyun Kim, Hyungwook Choi, Yejin Kim, Kihyun Kim, Yohan Jo, Joonseok Lee*
378 Generative Video Motion Editing with 3D Point Tracks, *Yao-Chih Lee, Zhoutong Zhang, Jiahui Huang, Jui-Hsien Wang, Joon-Young Lee, Jia-Bin Huang, Eli Shechtman, Zhengqi Li*
379 BulletTime: Decoupled Control of Time and Camera Pose for Video Generation, *Yiming Wang, Qihang Zhang, Shengqu Cai, Tong Wu, Jan Ackermann, Zhengfei Kuang, Yang Zheng, Frano Rajič, Siyu Tang, Gordon Wetzstein*
380 Learning to Generate Highly Dynamic Videos using Synthetic Motion Data, *Wonjoon Jin, Jiyun Won, Janghyeok Han, Qi Dai, Chong Luo, Seung-Hwan Baek, Sunghyun Cho*
381 Stereo World Model: Camera-Guided Stereo Video Generation, *Yang-Tian Sun, Zehuan Huang, Yifan Niu, Lin Ma, Yan-Pei Cao, Yuewen Ma, Xiaojuan Qi*
382 CG-Floor: Centroid-Guided Diffusion for Large-Scale Floorplan Generation, *Hongjin Lian, Jian Ma, Hongjie Chen, Jia Li, Ruizhen Hu, Yu-Kun Lai, Kun Li*
383 MAD: Motion Appearance Decoupling for efficient Driving World Models, *Ahmad Rahimi, Valentin Gerard, Eloi Zablocki, Matthieu Cord, Alexandre Alahi*
384 VDFE: Difference-Aware 3D Scene Editing with Non-Intrusive Video Diffusion Priors for Multi-View Consistency and Efficiency, *Chao Zhang, Fang Liu, Shuo Li, Yang Liu, Jiahao Wang, Xinyan Huang, Lingling Li, Puhua Chen, Xu Liu, Wenping Ma, Siqi Yu*
385 Endless World: Real-Time 3D-Aware Long Video Generation, *Ke Zhang, Jiacong Xu, Yiqun Mei, Vishal M. Patel*
386 SpatialDiff: 3D-Aware Object Movement via Implicit Spatial Modeling, *Zheng Liu, Zijian He, Huiguo He, Weizhi Zhong, Yejun Tang, Huan Yang, Kun Gai, Guanbin Li*
387 Towards Realistic and Consistent Orbital Video Generation via 3D Foundation Priors, *Rong Wang, Ruyi Zha, Ziang Cheng, Jiayu Yang, Pulak Purkait, Hongdong Li*

388 YOLO-ULM: Ultra-Lightweight Models for Real-Time Object Detection, *Shasha Han, Chong Li, Xinning Wang, Xuebo Li*
389 CHIRP dataset: towards long-term, individual-level, behavioral monitoring
❊ of bird populations in the wild, *Alex Hoi Hang Chan, Neha Singhal, Onur Kocahan, Andrea Meltzer, Saverio Lubrano, Miyako H. Warrington, Michael Griesser, Fumihiro Kano, Hemal Naik*
390 YOLO-Master: MOE-Accelerated with Specialized Transformers for Enhanced Real-time Detection, *Xu Lin, Jinlong Peng, Zhenye Gan, Jiawen Zhu, Jun Liu*
391 VLM4RSDet: Collaborative Optimization with Vision-Language Model
❊ for Enhancing Remote Sensing Object Detection, *Shuohao Shi, Qiang Fang, Xin Xu*
392 WiTTA-Bench: Benchmarking Test-Time Adaptation for WiFi Sensing, *Bing Li, Qiang Wang, Junda Lu, Le Zhang, Yun Liu, Ce Zhu, Wei Cui*
393 MFEN: Multi-Frequency Expert Network for Visible-Infrared Person
❊ Re-ID, *Xulin Li, Yan Lu, Bin Liu, Qinhong Yang, Qi Chu, Tao Gong, Nenghai Yu*
394 Object-Generalized Re-Identification: A Step Towards Universal Instance Perception, *Shuoyi Chen, Yurui Wu, Mang Ye*
395 When Transformers Meet Mamba: A Hybrid Transformer-Mamba Network for Video Object Detection, *Qiang Qi, Xiao Wang, Zongyuan Du, Yu Zhang*
396 Prompt-Anchored Vision–Text Distillation for Lifelong Person Re-identification, *Wen Wen, Hao Chen, Shiliang Zhang*
397 HyperGait: Unleashing the Power of Parsing for Gait Recognition in
❊ the Wild via Hypergraph, *Jinkai Zheng, Jiaqing Wei, Xinxiang Jin, Yaoqi Sun, Xichun Sheng, Ming Li, Liangqiong Qu, Xinchen Liu, Wu Liu*
398 Accelerating Streaming Video Large Language Models via Hierarchical Token Compression, *Yiyu Wang, Xuyang Liu, Xiyan Gui, Xinying Lin, Boxue Yang, Chenfei Liao, Tailai Chen, Linfeng Zhang*
399 Do You See What I Am Pointing At? Gesture-Based Egocentric Video Question Answering, *Yura Choi, Roy Miles, Rolandos Alexandros Potamias, Ismail Elezi, Jiankang Deng, Stefanos Zafeiriou*
400 Beyond Caption-Based Queries in Video Moment Retrieval, *David Pujol-Perich, Albert Clapés, Dima Damen, Sergio Escalera, Michael Wray*
401 Neural-Centric Video Processing Pipeline for Unified Multi-Task Inference, *Seyeon Lee, Juncheol Ye, Jaehong Kim, Dongsu Han*
402 VideoRealBench: A Chain-of-Thought Realism Evaluation Benchmark for Generated Human-Centric Videos, *Min Yang, Xinwen Zhang, Jialei Tang, Xin Zhou, Kehan Li, Zeyi Huang, Limin Wang*
403 VAST: Video Ability-Stratified Taxonomy for Data-Efficient Video Reasoning, *Zhongan Wang, Xiaoyu Wen, Lingxiao Du, Kun Li, Zhiliang Wu, Xingcheng Xu, Qiaosheng Zhang, Chaochao Lu, Hehe Fan*
404 An Empirical Study on How Video-LLMs Answer Video Questions, *Chenhui Gou, Ziyu Ma, Zicheng Duan, Haoyu He, Feng Chen, Akide Liu, Bohan Zhuang, Jianfei Cai, Hamid Rezatofighi*
405 FPSBench: A Benchmark for Video Understanding at High Frame Rates, *Rohan Choudhury, Jean-Sebastien Dandurand, Kai Qiu, Kshitij Madhav Bhat, Kartik Sharma, Liza Dahiya, Yizhou Zhao, Souraja Kundu, Chun-Hsien Lin, Kris M. Kitani, Laszlo A. Jeni*
406 UniComp: Rethinking Video Compression Through Informational Uniqueness, *Chao Yuan, Shimin Chen, Minliang Lin, Limeng Qiao, Guanglu Wan, Lin Ma*
407 NaTex: Seamless Texture Generation as Latent Color Diffusion,
❊ *Zeqiang Lai, Yunfei Zhao, Zibo Zhao, Xin Yang, Xin Huang, Jingwei Huang, Xiangyu Yue, Chunchao Guo*
408 Your Latent Mask is Wrong: Pixel-Equivalent Latent Compositing for Diffusion Models, *Rowan Bradbury, Dazhi Zhong*
409 Pluggable Pruning with Contiguous Layer Distillation for Diffusion
❊ Transformers, *Jian Ma, Qirong Peng, Xujie Zhu, Peixing Xie, Chen Chen, Haonan Lu*
410 Attribute-Preserving Pseudo-Labeling for Diffusion-Based Face Swapping, *Jiwon Kang, Yeji Choi, JoungBin Lee, Wooseok Jang, Jinhyeok Choi, Taekeun Kang, Yongjae Park, Myungin Kim, Seungryong Kim*
411 Delta Rectified Flow Sampling for Text-to-Image Editing, *Gaspard Beaudouin, Minghan Li, Jaeyeon Kim, Sung-Hoon Yoon, Mengyu Wang*
412 Training-free Mixed-Resolution Latent Upsampling for Spatially
❊ Accelerated Diffusion Transformers, *Wongi Jeong, Kyungryeol Lee, Hoigi Seo, Se Young Chun*
413 SpotEdit: Selective Region Editing in Diffusion Transformers, *Zhibin Qin, Zhenxiong Tan, Zeqing Wang, Songhua Liu, Xinchao Wang*
414 All-in-One Slider for Attribute Manipulation in Diffusion Models, *Weixin Ye, Hongguang Zhu, Wei Wang, Yahui Liu, Mengyu Wang, Xuecheng Nie*
415 DA-VAE: Plug-in Latent Compression for Diffusion via Detail Alignment, *Xin Cai, Zhiyuan You, Zhoutong Zhang, Tianfan Xue*
416 From Sketch to Fresco: Efficient Diffusion Transformer with Progressive Resolution, *Shikang Zheng, Guantao Chen, Landis He, Jiacheng Liu, Yuqi Lin, Chang Zou, Linfeng Zhang*
417 CATNet: Collaborative Alignment and Transformation Network for Cooperative Perception, *Gong Chen, Chaokun Zhang, Tao Tang, Pengcheng Lv, Feng Li, Xin Xie*
418 Scene Reconstruction as Mapping Priors for 3D Detection, *Yang Fu,*
❊ *Yuliang Zou, Hao Xiang, Xin Huang, Yijing Bai, Chen Song, Weijing Shi, Govind Thattai, Dragomir Anguelov, Mingxing Tan, Yingwei Li*
419 CCF: Complementary Collaborative Fusion for Domain Generalized Multi-Modal 3D Object Detection, *Yuchen Wu, Kun Wang, Yining Pan, Na Zhao*
420 Unleashing the Power of Chain-of-Prediction for Monocular 3D
❊ Object Detection, *Zhihao Zhang, Abhinav Kumar, Girish Chandar Ganesan, Xiaoming Liu*
421 R4Det: 4D Radar-Camera Fusion for High-Performance 3D Object Detection, *Zhongyu Xia, Yousen Tang, Yongtao Wang, Zhifeng Wang, Weijun Qin*
422 Revisiting Token Compression for Accelerating ViT-based Sparse Multi-View 3D Object Detectors, *Mingqian Ji, Shanshan Zhang, Jian Yang*
423 Few-Shot Incremental 3D Object Detection in Dynamic Indoor Environments, *Yun Zhu, Jianjun Qian, Jian Yang, Jin Xie, Na Zhao*
424 Learning from Synthetic Data via Provenance-Based Input Gradient Guidance, *Koshiro Nagano, Ryo Fujii, Ryo Hachiuma, Fumiaki Sato, Taiki Sekii, Hideo Saito*
425 Seeing Clearly, Reasoning Confidently: Plug-and-Play Remedies for Vision Language Model Blindness, *Xin Hu, Haomiao Ni, Yunbei Zhang, Jihun Hamm, Zechen Li, Zhengming Ding*
426 Draft and Refine with Visual Experts, *Sungheon Jeong, Ryozo Masukawa,*
❊ *Jihong Park, Sanggeon Yun, Wenjun Huang, Hanning Chen, Mahdi Imani, Mohsen Imani*
427 R2G: A Multi-View Circuit Graph Benchmark Suite from RTL to GDSII, *Zewei Zhou, Jiajun Zou, Jiajia Zhang, Ao Yang, Ruichao He, Haozheng Zhou, Ao Liu, Jiawei Liu, Leilei Jin, Shan Shen, Daying Sun*
428 VQ-VA World: Towards High-Quality Visual Question-Visual Answering, *Chenhui Gou, Zilong Chen, Zeyu Wang, Feng Li, Deyao Zhu, Zicheng Duan, Kunchang Li, Chaorui Deng, Hongyi Yuan, Haoqi Fan, Cihang Xie, Jianfei Cai, Hamid Rezatofighi*
429 Cross-Domain Demo-to-Code via Neurosymbolic Counterfactual Reasoning, *Jooyoung Kim, Wonje Choi, Younguk Song, Honguk Woo*
430 Beyond Multiple Choice: Verifiable OpenQA for Robust Vision-Language RFT, *Yesheng Liu, Hao Li, Haiyu Xu, Baoqi Pei, Jiahao Wang, Mingxuan Zhao, Jing-Shu Zheng, Zheqi He, JG Yao, Xi Yang, Bowen Qin, Jiajun Zhang*
431 See Further, Think Deeper: Advancing VLM's Reasoning Ability with Low-level Visual Cues and Reflection, *Zhiheng Wu, Tong Wang, Shuning Wang, Naiming Liu, Yumeng Zhang*
432 PDCR: Perception-Decomposed Confidence Reward for Vision-Language Reasoning, *Hee Suk Yoon, Eunseop Yoon, Ji Woo Hong, SooHwan Eom, Gwanhyeong Koo, Mark Hasegawa-Johnson, Qi Dai, Chong Luo, Chang D. Yoo*
433 µVLM: A Vision Language Model for µNPUs, *Zijie Chen, Guiyun Fan, Zhaoxing Yang, Rong Ding, Haiming Jin*
434 Gaussian Mapping for Evolving Scenes, *Vladimir Yugay, Thies Kersten, Luca Carlone, Theo Gevers, Martin R. Oswald, Lukas Schmid*
435 Part-aware Modeling of Articulated Objects using 3D Gaussian Splatting, *Tianjiao Yu, Vedant Shah, Muntasir Wahed, Ying Shen, Kiet A. Nguyen, Ismini Lourentzou*

436 AnchorSplat: Feed-Forward 3D Gaussian Splatting With 3D Geometric Priors, *Xiaoxue Zhang, Xiaoxu Zheng, Yixuan Yin, Tiao Zhao, Kaihua Tang, Michael Bi Mi, Zhan Xu, Dave Zhenyu Chen*
437 SGAD-SLAM: Splatting Gaussians at Adjusted Depth for Better Radiance Fields in RGBD SLAM, *Pengchong Hu, Zhizhong Han*
438 Faster-GS: Analyzing and Improving Gaussian Splatting Optimization, *Florian Hahlbohm, Linus Franke, Martin Eisemann, Marcus Magnor*
439 Layered 4D-Rotor Gaussian Splatting: A Compressed Representation for Long Dynamic Scenes, *Hanjie Xu, Yuanxing Duan, Qiyu Dai, Ge Li, Baoquan Chen, He Wang*
440 GaussianGrow: Geometry-aware Gaussian Growing from 3D Point Clouds with Text Guidance, *Weiqi Zhang, Junsheng Zhou, Haotian Geng, Kanle Shi, Shenkun Xu, Yi Fang, Yu-Shen Liu*
441 PhysGS: Bayesian-Inferred Gaussian Splatting for Physical Property Estimation, *Samarth Chopra, Jing Liang, Gershom Seneviratne, Dinesh Manocha*
442 3D Gaussian Splatting at Arbitrary Resolutions with Compact Proxy Anchors, *Mingyun Jeong, Seongro Yoon, Francois Bremond, Donghyeon Cho*
443 Stochastic Ray Tracing for the Reconstruction of 3D Gaussian Splatting, *Peiyu Xu, Shuang Zhao, Xin Sun, Krishna Mullia, Raymond Fei, Iliyan Georgiev*
444 AeroDGS: Physically Consistent Dynamic Gaussian Splatting for Single-Sequence Aerial 4D Reconstruction, *Hanyang Liu, Rongjun Qin*
445 GaussianPile: A Unified Sparse Gaussian Splatting Framework for Slice-based Volumetric Reconstruction, *Di Kong, Yikai Wang, Wenjie Guo, Yifan Bu, Boya Zhang, Yuexin Duan, Xiawei Yue, Wenbiao Du, Yiman Zhong, Yuwen Chen, Cheng Ma*
446 More Natural, More Real: Object-aware Gaussian Splatting for 3D Visual Decoding from Human Brain, *Haodong Jing, Dongyao Jiang, Jixin Wang, Junhao Jia, Yanshu Li, Yongqiang Ma, Nanning Zheng*
447 Eulerian Gaussian Splatting using Hashed Probability Pyramids, *Mia Gaia Polansky, George Kopanas, Stephan Garbin, Todd Zickler, Dor Verbin*
448 ❊ Confidence-Guided Multi-Scale Aggregation for Sparse-View High-Resolution 3D Gaussian Splatting, *Qinzheng Zhou, Zaychik Liu, Lijing Lu, Zhihang Li*
449 ❊ ULF-Loc: Unbiased Landmark Feature for Robust Visual Localization with 3D Gaussian Splatting, *Yingdong Gu, Shaocheng Yan, Zhenjun Zhao, Yuan Kou, Jianxin Luo, Pengcheng Shi, Jiayuan Li*
450 Robust3DGSW: Toward Robust Watermarking for Quantization-Aware 3D Gaussian Splatting, *Boyu Wang, Jun Xia, Mingsong Chen*
451 ParkGaussian: Surround-view 3D Gaussian Splatting for Autonomous Parking, *Xiaobao Wei, Zhangjie Ye, Yuxiang Gu, Zunjie Zhu, Yunfei Guo, Yingying Shen, Shan Zhao, Ming Lu, Haiyang Sun, Bing Wang, Guang Chen, Rongfeng Lu, Hangjun Ye*
452 Part-aware Modeling of Articulated Objects using 3D Gaussian Splatting, *Ashish Kumar, Rajagopalan N Ambasamduram*
453 Probabilistic Concept Graph Reasoning for Multimodal Misinformation Detection, *Ruichao Yang, Wei Gao, Xiaobin Zhu, Jing Ma, Hongzhan Lin, Ziyang Luo, Bo-Wen Zhang, Xu-Cheng Yin*
454 POINTS-Long: Adaptive Dual-Mode Visual Reasoning in MLLMs, *Haicheng Wang, Yuan Liu, Yikun Liu, Zhemeng Yu, Zhongyin Zhao, Yangxiu You, Zilin Yu, Le Tian, Zhou Xiao, Jie Zhou, Weidi Xie, Yanfeng Wang*
455 SegCompass: Exploring Interpretable Alignment with Sparse Autoencoders for Enhanced Reasoning Segmentation, *Zhenyu Lu, Liupeng Li, Jinpeng Wang, Haoqian Kang, Yan Feng, Ke Chen, Yaowei Wang*
456 CRIT: Graph-Based Automatic Data Synthesis to Enhance Cross-Modal Multi-Hop Reasoning, *Junyoung Sung, Seungwoo Lyu, Minjun Kim, Sumin An, Arsha Nagrani, Paul Hongsuck Seo*
457 DeepScan: A Training-Free Framework for Visually Grounded Reasoning in Large Vision-Language Models, *Yangfu Li, Hongjian Zhan, Jiawei Chen, Yuning Gong, Qi Liu, Yue Lu*
458 Locate-Then-Examine: Grounded Region Reasoning Improves Detection of AI-Generated Images, *Yikun Ji, Yan Hong, Bowen Deng, Jun Lan, Huijia Zhu, Weiqiang Wang, Liqing Zhang, Jianfu Zhang*
459 HUMORCHAIN: Theory-Guided Multi-Stage Reasoning for Interpretable Multimodal Humor Generation, *Jiajun Zhang, Shijia Luo, Ruikang Zhang, Qi Su*
460 CodeDance: A Dynamic Tool-integrated MLLM for Executable Visual Reasoning, *Qi Song, Honglin Li, Yingchen Yu, Haoyi Zhou, Lin Yang, Song Bai, Qi She, Zilong Huang, Yunqing Zhao*
461 Rethinking MLLM Itself as a Segmenter with a Single Segmentation Token, *Anqi Zhang, Xiaokang Ji, Guangyu Gao, Jianbo Jiao, Chi Harold Liu, Yunchao Wei*
462 Video-Only ToM: Enhancing Theory of Mind in Multimodal Large Language Models, *Siqi Liu, Xinyang Li, Bochao Zou, Junbao Zhuo, Huimin Ma, Jiansheng Chen*
463 Mario: Multimodal Graph Reasoning with Large Language Models, *Yuanfu Sun, Kang Li, Pengkang Guo, Jiajin Liu, Qiaoyu Tan*
464 Boosting Reasoning in Large Multimodal Models via Activation Replay, *Yun Xing, Xiaobin Hu, Qingdong He, Jiangning Zhang, Shuicheng Yan, Shijian Lu, Yu-Gang Jiang*
465 Rationale-Enhanced Decoding for Multi-modal Chain-of-Thought, *Shin'ya Yamaguchi, Kosuke Nishida, Daiki Chijiwa*
466 Mimic Human Cognition, Master Multi-Image Reasoning: A Meta-Action Framework for Enhanced Visual Understanding, *Jianghao Yin, Qingbin Li, Kun Sun, Cheng Ding, Jie Wang, Qin Chen, Jie Zhou, Nan Wang, Changqing Li, Pei Wu, Jian Xu, Zheming Yang, Liang He*
467 ❊ ROSE: Rotate Your Large Language Model to See, *Tongtian Yue, Xuange Gao, Longteng Guo, Zijia Zhao, Zikang Liu, Jie Jiang, Hua Huang, Jing Liu*
468 OpenMMReasoner: Pushing the Frontiers in Multimodal Reasoning with an Open and General Recipe, *Kaichen Zhang, Keming Wu, Zuhao Yang, Bo Li, Kairui Hu, Bin Wang, Xingxuan Li, Lidong Bing*
469 SelecTKD: Selective Token-Weighted Knowledge Distillation for LLMs, *Haiduo Huang, Jiangcheng Song, Yadong Zhang, Pengju Ren*
470 Sparsity as a Key: Unlocking New Insights from Latent Structures for Out-of-Distribution Detection, *Ahyoung Oh, Wonseok Shin, Songkuk Kim*
471 SparVAR: Exploring Sparsity in Visual AutoRegressive Modeling for Training-Free Acceleration, *Zekun Li, Ning Wang, Tongxin Bai, Changwang Mei, Peisong Wang, Shuang Qiu, Jian Cheng*
472 Suppressing Non-Semantic Noise in Masked Image Modeling Representations, *Martine Hjelkrem-Tan, Marius Aasan, Rwiddhi Chakraborty, Gabriel Y. Arteaga, Changkyu Choi, Adín Ramírez Rivera*
473 Block-based Learned Image Compression without Blocking Artifacts, *Jong Wook Kim, Suyong Bahk, TaeHwa Lee, HyunDong Cho, Donghyun Kim, Sung-Chang Lim, Jin Soo Choi, Hui Yong Kim*
474 DeDelayed: Deleting Remote Inference Delay via On-Device Correction, *Dan Jacobellis, Mateen Ulhaq, Fabien Racapé, Hyomin Choi, Neeraja J. Yadwadkar*
475 AdaRadar: Rate Adaptive Spectral Compression for Radar-based Perception, *Jinho Park, Se Young Chun, Mingoo Seok*
476 ❊ Gaussian Splatting-based Low-Rank Tensor Representation for Multi-Dimensional Image Recovery, *Yiming Zeng, Xi-Le Zhao, Wei-Hao Wu, Teng-Yu Ji, Chao Wang*
477 Precise Object and Effect Removal with Adaptive Target-Aware Attention, *Jixin Zhao, Zhouxia Wang, Peiqing Yang, Shangchen Zhou*
478 Decompose, Mix, Adapt: A Unified Framework for Parameter-Efficient Neural Network Recombination and Compression, *Nazia Tasnim, Shrimai Prabhumoye, Bryan A. Plummer*
479 ❊ FreqSIC: Frequency-aware Stereo Image Compression with Bi-directional Checkerboard Context Model, *Shiyu Qin, Yongkang Lu, Yimin Zhou, Jiawei Li, Yifan Ren, Yuerong Xue, Shu-Tao Xia, Bin Chen*
480 ❊ SinGeo: Unlock Single Model's Potential for Robust Cross-View Geo-Localization, *Yang Chen, Xieyuanli Chen, Junxiang Li, Jie Tang, Tao Wu*
481 Fusion of Depth and Semantics for Probabilistic Floorplan Localization, *Kecheng Ye, Mao Chen, Xiangkai Zhang, Xu Yang*
482 A2GC: Asymmetric Aggregation with Geometric Constraints for Locally Aggregated Descriptors, *Zhenyu Li, Tianyi Shang*
483 Geo2: Geometry-Guided Cross-view Geo-Localization and Image Synthesis, *Yancheng Zhang, Xiaohan Zhang, Guangyu Sun, Zonglin Lyu, Safwan Wshah, Chen Chen*
484 Coverage Optimization for Camera View Selection, *Timothy Chen, Adam Dai, Maximilian Adang, Grace Gao, Mac Schwager*

485 Resolving Evidence Sparsity: Agentic Context Engineering for Long-Document Understanding, *Keliang Liu, Zizhi Chen, Mingcheng Li, Jingqun Tang, Dingkang Yang, Lihua Zhang*
486 Reasoning Palette: Modulating Reasoning via Latent Contextualization for Controllable Exploration for (V)LMs, *Rujiao Long, Yang Li, Xingyao Zhang, Weixun Wang, Tianqianjin Lin, Xi Zhao, Yuchi Xu, Wenbo Su, Junchi Yan, Bo Zheng*
487 ORCA: Orchestrated Reasoning with Collaborative Agents for Document Visual Question Answering, *Aymen Lassoued, Mohamed Ali Souibgui, Yousri Kessentini*
488 MSJoE: Jointly Evolving MLLM and Sampler for Efficient Long-Form Video Understanding, *Wenhui Tan, Xiaoyi Yu, Jiaze Li, Yijing Chen, Jianzhong Ju, Zhenbo Luo, Ruihua Song, Jian Luan*
489 A Multi-Agent Perception-Action Alliance for Efficient Long Video Reasoning, *Yichang Xu, Gaowen Liu, Ramana Rao Kompella, Tiansheng Huang, Sihao Hu, Fatih Ilhan, Selim Furkan Tekin, Zachary Yahn, Ling Liu*
490 Saliency-Guided Representation with Consistency Policy Learning for Visual Unsupervised Reinforcement Learning, *Jingbo Sun, Qichao Zhang, Songjun Tu, Xing Fang, Yupeng Zheng, Haoran Li, Ke Chen, Dongbin Zhao*
491 LensWalk: Agentic Video Understanding by Planning How You See ❊ in Videos, *Keliang Li, Yansong Li, Hongze Shen, Mengdi Liu, Hong Chang, Shiguang Shan*
492 DPGF-Net: Dual-Prior Guided Fusion Network for Joint Assessment of Perceptual Quality and Semantic Consistency in AI-Generated Images, *Tao Li, Xingran Liao, Mingliang Zhou*
493 RegionFuse: Region-Adaptive Pixel Distribution Learning for Infrared and Visible Image Fusion, *Jianghan Xia, Hong Song, Jinfu Li, Yucong Lin, Shihan Ma, Jingfan Fan, Danni Ai, Tianyu Fu, Deqiang Xiao, Jian Yang*
494 Missing No More: Dictionary-Guided Cross-Modal Image Fusion ❊ under Missing Infrared, *Yafei Zhang, Meng Ma, Huafeng Li, Yu Liu*
495 VideoFusion: A Spatio-Temporal Collaborative Network for Multi-modal Video Fusion, *Linfeng Tang, Yeda Wang, Meiqi Gong, Zizhuo Li, Yuxin Deng, Xunpeng Yi, Chunyu Li, Han Xu, Hao Zhang, Jiayi Ma*
496 TAPE: Task-Adaptive Prototype Evolution in Audio-Language Models for Fully Few-shot Class-incremental Audio Classification, *Yunlong Gao, Wenxin Liang, Guanglu Wang, Senqi Guan, Linlin Zong, Dongyu Zhang, Xinyue Liu*
497 Remedying Target-Domain Astigmatism for Cross-Domain Few-Shot Object Detection, *Yongwei Jiang, Yixiong Zou, Yuhua Li, Ruixuan Li*
498 DDSF: Robust Few-Shot Learning via Disentangled Subspaces with Determinantal Point Process, *Xulun Ye, Yifan Mei, Kun Zhou, Zelei Wu, Jieyu Zhao*
499 Hyperbolic Defect Feature Synthesis for Few-Shot Defect Classification, *Huimin Li, Boxuan Hu, Yulin Zhang, Xiuzhuang Zhou, Junlin Hu*
500 Training-Only Heterogeneous Image-Patch-Text Graph Supervision for Advancing Few-Shot Learning Adapters, *Mohammed Rahman Sherif Khan Mohammad, Ardhendu Behera, Sandip Pradhan, Swagat Kumar, Amr Ahmed*
501 Learning to Learn Weight Generation via Local Consistency Diffusion, *Yunchuan Guan, Yu Liu, Ke Zhou, Zhiqi Shen, Jenq-Neng Hwang, Lei Li*
502 Balanced Dataset Distillation via Modeling Multiple Visual Pattern Distribution, *Guanghui Shi, Xuefeng Liang, Qixiang Wen*
503 Grid Distillation: Compositional Image Distillation via Structured Generative Grids, *Biplab Ch Das, Shouvik Das, Viswanath Gopalakrishnan*
504 Dataset Distillation by Influence Matching, *Haoru Tan, Wang Wang, Sitong Wu, Xiuzhe Wu, Yang-Tian Sun, Chirui Chang, Shaofeng Zhang, Xiaojuan Qi*
505 StableMaterials: Enhancing Diversity in Material Generation via Semi-Supervised Learning, *Giuseppe Vecchio*
506 Seeing Through Blur: Tackling Defocus in Spike-Based Imaging, *Xiantao Ma, Siwei Dong, Lin Zhu, Lizhi Wang, Hua Huang*
507 Distilling Quasi-Conformal Mapping: A Generalizable and Efficient Solution for Wide-Angle Correction, *Chengyang Liu, Zixuan Lin, Miaolin Han, Michael K. Ng, Huibin Li*
508 Lighting in Motion: Spatiotemporal HDR Lighting Estimation, *Christophe Bolduc, Julien Philip, Li Ma, Mingming He, Paul Debevec, Jean-François Lalonde*
509 LightRR: A Lightweight Network for Single Image Reflection Removal, *Wenbin Yin, Junkang Zhang, Sunzhe Yang, Faming Fang, Guixu Zhang*
510 HFR and HDR Video from Multi-Attenuated Spikes Using a Rapidly Rotating SpokeND Filter, *Yakun Chang, Zhaojun Huang, Siqi Yang, Yeliduosi Xiaokaiti, Shikui Wei, Yao Zhao, Tiejun Huang, Boxin Shi*
511 Coded-E2LF: Coded Aperture Light Field Imaging from Events, *Tomoya Tsuchida, Keita Takahashi, Chihiro Tsutake, Toshiaki Fujii, Hajime Nagahara*
512 TokenLight: Precise Lighting Control in Images using Attribute ❊ Tokens, *Sumit Chaturvedi, Yannick Hold-Geoffroy, Mengwei Ren, Jingyuan Liu, He Zhang, Yiqun Mei, Julie Dorsey, Zhixin Shu*
513 Kaleidoscopic Scintillation Event Imaging, *Alex Bocchieri,* ❊ *John Mamish, David Appleyard, Andreas Velten*
514 gQIR: Generative Quanta Image Reconstruction, *Aryan Garg,* ❊ *Sizhuo Ma, Mohit Gupta*
515 Solving Minimal Problems Without Matrix Inversion Using FFT-Based ❊ Interpolation, *Haidong Wu, Snehal Bhayani, Janne Heikkila*
516 Predicting Spatial Transcriptomics from Histology Images via High-Order Multi-Cell Interaction Modeling, *Youhan Sun, Jiahua Rao, Kangrui Du, Jiancong Xie, Yuedong Yang*
517 From Spots to Pixels: Dense Spatial Gene Expression Prediction from ❊ Histology Images, *Ruikun Zhang, Yan Yang, Liyuan Pan*
518 Cell-Type Prototype-Informed Neural Network for Gene Expression Estimation from Pathology Images, *Kazuya Nishimura, Ryoma Bise, Shinnosuke Matsuo, Haruka Hirose, Yasuhiro Kojima*
519 LightSplat: Fast and Memory-Efficient Open-Vocabulary 3D Scene Understanding in Five Seconds, *Jaehun Bang, Jinhyeok Kim, Minji Kim, Seungheon Jeong, Kyungdon Joo*
520 Guardians of the Hair: Rescuing Soft Boundaries in Depth, Stereo, and Novel Views, *Xiang Zhang, Studios blank, Yang Zhang, Studios blank, Lukas Mehl, Studios blank, Markus Gross, Studios blank, Christopher Schroers, Studios blank*
521 Zero-Shot Depth Completion with Vision-Language Model, *Zhiqiang Yan, Yuan Wu, Gim Hee Lee*
522 FE2E: From Editor to Dense Geometry Estimator, *Jiyuan Wang, Chunyu Lin, Lei Sun, Rongying Liu, Lang Nie, Mingxing Li, Kang Liao, Xiangxiang Chu*
523 Ego-1K – A Large-Scale Multiview Video Dataset for Egocentric Vision, *Jae Yong Lee, Daniel Scharstein, Akash Bapat, Hao Hu, Andrew Fu, Haoru Zhao, Paul Sammut, Xiang Li, Stephen Jeapes, Anik Gupta, Lior David, Saketh Madhuvarasu, Jay Girish Joshi, Jason Wither*
524 Edit-As-Act: Goal-Regressive Planning for Open-Vocabulary 3D Indoor Scene Editing, *Seongrae Noh, SeungWon Seo, Gyeong-Moon Park, HyeongYeop Kang*
525 VGGT-360: Geometry-Consistent Zero-Shot Panoramic Depth Estimation, *Jiayi Yuan, Haobo Jiang, De Wen Soh, Na Zhao*
526 NI-Tex: Non-isometric Image-based Garment Texture Generation, ❊ *Hui Shan, Ming Li, Haitao Yang, Kai Zheng, Sizhe Zheng, Yanwei Fu, Xiangru Huang*
527 Velox: Learning Representations of 4D Geometry and Appearance, *Anagh Malik, Dorian Chan, Xiaoming Zhao, David B. Lindell, Oncel Tuzel, Jen-Hao Rick Chang*
528 UniPixie: Unified and Probabilistic 3D Physics Learning via Flow ❊ Matching, *Qilin Huang, Quynh Anh Huynh, Long Le, Chen Wang, Chuhao Chen, Ryan Lucas, Eric Eaton, Lingjie Liu*
529 UniTEX: Universal High Fidelity Generative Texturing for 3D Shapes, *Yixun Liang, Kunming Luo, Xiao Chen, Rui Chen, Hongyu Yan, Weiyu Li, Jiarui Liu, Fei-Peng Tian, Ping Tan*
530 Points-to-3D: Structure-Aware 3D Generation with Point Cloud Priors, *Jiatong Xia, Zicheng Duan, Anton van den Hengel, Lingqiao Liu*
531 PartDiffuser: Part-wise 3D Mesh Generation via Discrete Diffusion, *Yichen Yang, Hong Li, Haodong Zhu, Linin Yang, Guojun Lei, Sheng Xu, Baochang Zhang*
532 LoST: Level of Semantics Tokenization for 3D Shapes, *Niladri Shekhar Dutt, Zifan Shi, Paul Guerrero, Chun-Hao Paul Huang, Duygu Ceylan, Niloy J. Mitra, Xuelin Chen*
533 Lafite: A Generative Latent Field for 3D Native Texturing, ❊ *Chia-Hao Chen, Yuan-Chen Guo, Zi-Xin Zou, Ze Yuan, Guan Luo,*

Xiaojuan Qi, Ding Liang, Yan-Pei Cao, Song-Hai Zhang

534 Image-Guided Geometric Stylization of 3D Meshes, *Changwoon Choi, Hyunsoo Lee, Clément Jambon, Yael Vinker, Young Min Kim*

535 LATTICE: Democratize High-Fidelity 3D Generation at Scale, *Zeqiang Lai, Yunfei Zhao, Zibo Zhao, Haolin Liu, Qingxiang Lin, Jingwei Huang, Chunchao Guo, Xiangyu Yue*

536 Dehallu3D: Hallucination-Mitigated 3D Generation from a Single Image via Cyclic View Consistency Refinement, *Xiwen Wang, Shichao Zhang, Ruowei Wang, Mao Li, Chenyu Zhou, Ji-Zhe Zhou, Qijun Zhao, Hailun Zhang*

537 MeshMosaic: Scaling Artist Mesh Generation via Local-to-Global Assembly, *Rui Xu, Tianyang Xue, Qiujie Dong, Le Wan, Zhe Zhu, Peng Li, Zhiyang Dou, Cheng Lin, Shiqing Xin, Yuan Liu, Wenping Wang, Taku Komura*

538 TacSlm: A Dataset and Benchmark for Football Tactical Style Imitation, *Peng Wen, Yuting Wang, Qiurui Wang*

539 DynamicsBoost: Dynamic Plausible Video Generation via Annotation-Free Continuation Preference Optimization, *Jiaxing Li, Jiepeng Wang, Junyao Gao, Yang Liu, Eric Li, Bo An, Hao-Xiang Guo*

540 Reinforcement-Guided Synthetic Data Generation for Privacy-Sensitive Identity Recognition, *Xuemei Jia, Jiawei Du, Hui Wei, Jun Chen, Joey Tianyi Zhou, Zheng Wang*

541 Fine-Grained GRPO for Precise Preference Alignment in Flow Models, *Yujie Zhou, Pengyang Ling, Jiazi Bu, Yibin Wang, Yuhang Zang, Jiaqi Wang, Li Niu, Guangtao Zhai*

542 Lighting-grounded Video Generation with Renderer-based Agent Reasoning, *Ziqi Cai, Taoyu Yang, Zheng Chang, Si Li, Han Jiang, Shuchen Weng, Boxin Shi*

543 RewardFlow: Generate Images by Optimizing What You Reward, *Onkar Susladkar, Dong-Hwan Jang, Tushar Prakash, Adheesh Juvekar, Vedant Shah, Ayush Barik, Nabeel Bashir, Muntasir Wahed, Ritish Shrirao, Ismini Lourentzou*

544 Goal Force: Teaching Video Models To Accomplish Physics-Conditioned Goals, *Nate Gillman, Yinghua Zhou, Zitian Tang, Evan Luo, Arjan Chakravarthy, Daksh Aggarwal, Michael Freeman, Chen Sun*

545 Self-Corrected Image Generation with Explainable Latent Rewards, *Yinyi Luo, Hrishikesh Gokhale, Marios Savvides, Jindong Wang, Shengfeng He*

546 Polyphony: Diffusion-based Dual-Hand Action Segmentation with Alternating Vision Transformer and Semantic Conditioning, *Hao Zheng, Hu Wang, Tiantian Zheng, Prajjwal Bhattarai, Tuka Alhanai*

547 ❊ Reading Your Actions: Learning Generalizable Action Representations via Pre-training AEMG, *Zhenghao Huang, Huilin Yao, Kaikai Wang, Lin Shu*

548 MA-Bench: Towards Fine-grained Micro-Action Understanding, *Kun Li, Jihao Gu, Fei Wang, Zhiliang Wu, Hehe Fan, Dan Guo*

549 OpenMarcie: Dataset for Multimodal Action Recognition in Industrial Environments, *Hymalai Bello, Lala Ray, Joanna Sorysz, Sungho Suh, Paul Lukowicz*

550 ❊ Action Motifs: Self-Supervised Hierarchical Representation of Human Body Movements, *Genki Kinoshita, Shu Nakamura, Ryo Kawahara, Shohei Nobuhara, Yasutomo Kawanishi, Ko Nishino*

551 DarkShake-DVS: Event-based Human Action Recognition under Low-light and Shaking Camera Conditions, *Jiaqi Chen, Qinfu Xu, Liyuan Pan*

552 Protect to Adapt: Subspace-Constrained Adaptation with Ranked Negative Prompt Feedback for Few-Shot Action Recognition, *Hantao Qi, Yan Yan, Junlong Gao, Hanzi Wang*

553 SkeletonContext: Skeleton-side Context Prompt Learning for Zero-Shot Skeleton-based Action Recognition, *Ning Wang, Tieyue Wu, Naeha Sharif, Farid Boussaid, Guangming Zhu, Lin Mei, Mohammed Bennamoun, Liang Zhang*

554 InTrain: Intrinsic Trainability for Zero-Cost Neural Architecture Search, *Qinqin Zhou, Fuhai Chen, Jipeng Wu, Zhiwei Chen, Zhikai Hu, Weiwei Cai*

555 S^2FT: Parameter-Efficient Fine-Tuning in Sparse Spectrum Domain, *Baoquan Zhang, Zhehao Yu, Lisai Zhang, Kenghong Lin, Tianran Chen, Yuxi Sun, Yunming Ye, Yao He*

556 Rethinking SNN Online Training and Deployment: Gradient-Coherent Learning via Hybrid-Driven LIF Model, *Zecheng Hao, Yifan Huang, Zijie Xu, Wenxuan Liu, Yuanhong Tang, Zhaofei Yu, Tiejun Huang*

557 ❊ Gated KalmaNet: A Fading Memory Layer through Test-time Ridge Regression, *Liangzu Peng, Aditya Chattopadhyay, Luca Zancato, Elvis Nunez, Wei Xia, Stefano Soatto*

558 Towards Efficient Medical Reasoning with Minimal Fine-Tuning Data, *Xinlin Zhuang, Feilong Tang, Haolin Yang, Xiwei Liu, Ming Hu, Huifa Li, Haochen Xue, Junjun He, Zongyuan Ge, Yichen Li, Ying Qian, Imran Razzak*

559 AdaBet: Gradient-free Layer Selection for Efficient Training of Deep Neural Networks, *Irene Tenison, Soumyajit Chatterjee, Fahim Kawsar, Mohammad Malekzadeh*

560 TAS-LoRA: Transformer Architecture Search with Mixture-of-LoRA Experts, *Jeimin Jeon, Hyunju Lee, Bumsub Ham*

561 QuCNet: Quantum Deep Learning Driven Multi-Circuit Network for Remote Sensing Image Classification, *Komal Komal, Mukul Gupta, Saumya Singh, Santosh Kumar Vipparthi, C.C. Reddy, Subrahmanyam Murala*

562 Learning to Solve PDEs on Neural Shape Representations, *Lilian Welschinger, Yilin Liu, Zican Wang, Niloy J. Mitra*

563 Frequency Switching Mechanism for Parameter-Efficient Multi-Task Learning, *Shih-Wen Liu, Yen-Chang Chen, Wei-Ta Chu, Fu-En Yang, Yu-Chiang Frank Wang*

564 Reconstructing Spiking Neural Networks Using a Single Neuron with Autapses, *Wuque Cai, Hongze Sun, Quan Tang, Shifeng Mao, Zhenxing Wang, Jiayi He, Duo Chen, Dezhong Yao, Daqing Guo*

565 ❊ Widget2Code: From Visual Widgets to UI Code via Multimodal LLMs, Houston H. Zhang, Tao Zhang, Baoze Lin, Yuanqi Xue, Yincheng Zhu, Huan Liu, Li Gu, Linfeng Ye, Ziqiang Wang, Xinxin Zuo, Yang Wang, Yuanhao Yu, Zhixiang Chi

566 GUI-CEval: A Hierarchical and Comprehensive Chinese Benchmark for Mobile GUI Agents, *Yang Li, Yuchen Liu, Haoyu Lu, Zhiqiang Xia, Hongzhen Wang, Kaiyang Han, Changpeng Yang, Jinyang Wu, Jiaming Xu, Runyu Shi, Ying Huang*

567 FocusUI: Efficient UI Grounding via Position-Preserving Visual Token Selection, *Mingyu Ouyang, Kevin Qinghong Lin, Mike Zheng Shou, Hwee Tou Ng*

568 Streamlined Open-Vocabulary Human-Object Interaction Detection, *Chang Sun, Dongliang Liao, Changxing Ding*

569 Decompose and Transfer: CoT-Prompting Enhanced Alignment for Open-Vocabulary Temporal Action Detection, *Sa Zhu, Wanqian Zhang, Lin Wang, Xiaohua Chen, Chenxu Cui, Jinchao Zhang, Bo Li*

570 Mitigating Simplicity Bias in OOD Detection through Object Co-occurrence Analysis, *Boyang Dai, Chaoqi Chen, Yizhou Yu*

571 Boosting Quantitive and Spatial Awareness for Zero-Shot Object Counting, *Da Zhang, Bingyu Li, Feiyu Wang, Zhiyuan Zhao, Junyu Gao*

572 Parameter-Efficient Semantic Augmentation for Enhancing Open-Vocabulary Object Detection, *Weihao Cao, Runqi Wang, Xiaoyue Duan, Jinchao Zhang, Ang Yang, Liping Jing*

573 WeDetect: Fast Open-Vocabulary Object Detection as Retrieval, *Shenghao Fu, Yukun Su, Fengyun Rao, Jing LYU, Xiaohua Xie, Wei-Shi Zheng*

574 ❊ Open-Vocabulary Domain Generalization in Urban-Scene Segmentation, Dong Zhao, Qi Zang, Nan Pu, Wenjing Li, Nicu Sebe, Zhun Zhong

575 OpenDPR: Open-Vocabulary Change Detection via Vision-Centric Diffusion-Guided Prototype Retrieval for Remote Sensing Imagery, *Qi Guo, Jue Wang, Yinhe Liu, Yanfei Zhong*

576 Annotation-Efficient Coreset Selection for Context-dependent Segmentation, *Jin Zhang, Zhe Cao, Biwen Yang, Ruiheng Zhang*

577 ALLNet: Multi-task Dense Prediction for Degraded Images, *Weiran Wang, Jialing Wu, Yaqi Chang, Gang He, Li Xu, Chang Wu, Yunsong Li*

578 Geometry-Aware Cross-Modal Graph Alignment for Referring Segmentation in 3D Gaussian Splatting, *Yuwen Tao, Kanglei Zhou, Chang Li, Liyuan Wang*

579 Volumetric Functional Maps, *Filippo Maggioli, Simone Melzi, Marco Livesu*

580 GenMask: Adapting DiT for Segmentation via Direct Mask Generation, *Yuhuan Yang, Xianwei Zhuang, Yuxuan Cai, Chaofan Ma, Shuai Bai, Jiangchao Yao, Ya Zhang, Junyang Lin, Yanfeng Wang*

581 Frequency-Aware Affinity for Weakly Supervised Semantic Segmentation, *Ziqian Yang, Xianglin Qiu, Xinqiao Zhao, Xiaolei Wang, Quan Zhang, Jimin Xiao*
582 Learning and Aligning Click-Aware Shape Prior for Interactive Amodal Instance Segmentation, *Junjie Chen, Junwei Lin, Ren Hong, Shengjie Liu, Yuming Fang, Feng Qian, Yifan Zuo*
583 Beyond Reassembly: Fractured Object Recovery with Missing Parts, *Qun-Ce Xu, Jiahui Li, Yan-Pei Cao, Weihao Cheng, Tai-Jiang Mu, Ying Shan, Chuan Li, Da Chen, Yong-Liang Yang, Shi-min Hu*
584 Best Segmentation Buddies for Image-Shape Correspondence, *Itai Lang, Dongwei Lyu, Dale Decatur, Rana Hanocka*
585 RMAE-ProGRess: Advancing Semantic Segmentation in Unstructured Environments, *Manish Bhurtel, Danda B. Rawat*
586 Local Precise Refinement: A Dual-Gated Mixture-of-Experts for Enhancing Foundation Model Generalization against Spectral Shifts, *Xi Chen, Maojun Zhang, Yu Liu, Shen Yan*
587 Orthogonal Spatial-Aware Multi-View Anchor Graph Clustering for Incomplete Remote Sensing Data, *Yongshan Zhang, Xiaohuan Lin, Lefei Zhang, Zhihua Cai*
588 SIGMA: A Physics-Based Benchmark for Gas Chimney Understanding in Seismic Images, *Bao Truong, Quang Nguyen, Baoru Huang, Jinpei Han, Van Nguyen, Ngan Le, Minh-Tan Pham, Doan Huy Hien, Anh Nguyen*
589 SkySense-VITA: Towards Universal In-context Segmentation of Multi-modal Remote Sensing Imagery, *Kang Wu, Lei Yu, Junwei Luo, Bo Dang, Junjian Zhang, Xiangyuan Cai, Hongwei Hu, Jingdong Chen, Yansheng Li*
590 ProM3E: Probabilistic Masked MultiModal Embedding Model for Ecology, *Srikumar Sastry, Subash Khanal, Aayush Dhakal, Jiayu Lin, Dan Cher, Phoenix Jarosz, Nathan Jacobs*
591 GeoCoT: Towards Reliable Remote Sensing Reasoning with Manifold
❊ Perspective, *Daixun Li, Zirui Li, Sibo He, Jiayun Tian, Mingxiang Cao, Weiying Xie, Yunke Wang, Xin Zhang, Yusi Zhang, Yunsong Li, Chang Xu, Leyuan Fang*
592 STCast: Adaptive Boundary Alignment for Global and Regional Weather
❊ Forecasting, *Hao Chen, Tao Han, Jie Zhang, Song Guo, Lei Bai*
593 NeighborMAE: Exploiting Spatial Dependencies between Neighboring Earth Observation Images in Masked Autoencoders Pretraining, *Liang Zeng, Valerio Marsocci, Wufan Zhao, Andrea Nascetti, Maarten Vergauwen*
594 GeoDiT: A Diffusion-based Vision-Language Model for Geospatial Understanding, *Jiaqi Liu, Ronghao Fu, Haoran Liu, Lang Sun, Qipeng Wang, Bo Yang*
595 Balanced Hierarchical Contrastive Learning with Decoupled Queries for Fine-grained Object Detection in Remote Sensing Images, *Jingzhou Chen, Dexin Chen, Fengchao Xiong, Yuntao Qian, Liang Xiao*
596 Generative Adversarial Perturbations with Cross-paradigm Transferability on Localized Crowd Counting, *Alabi Mehzabin Anisha, Guangjing Wang, Sriram Chellappan*
597 Improving Adversarial Transferability with Local Perturbation Augmentation, *Jian-Xun Mi, Xuanhui Zhong, Weisheng Li*
598 Echoes of Ownership: Adversarial-Guided Dual Injection for Copyright Protection in MLLMs, *Chengwei Xia, Fan Ma, Ruijie Quan, Yunqiu Xu, Kun Zhan, Yi Yang*
599 Stealing Split Learning Bottom Models by Recovering Embedding Geometry, *Qinbo Zhang, Yanhang Shi, Ziyi Zhang, Hao Wang, Sai Qian Zhang, Jian Li*
600 PoInit-of-View: Poisoning Initialization of Views Transfers Across Multiple 3D Reconstruction Systems, *Weijie Wang, Songlong Xing, Zhengyu Zhao, Nicu Sebe, Bruno Lepri*
601 No Way To Steal My Face: Proactive Defense Against Identity-Preserving Personalized Generation, *Lizhi Xiong, Jun Li, Ziqiang Li, Weiwei Jiang, Zhangjie Fu*
602 Towards Reliable Evaluation of Adversarial Robustness for Spiking Neural Networks, *Jihang Wang, Dongcheng Zhao, Ruolin Chen, Qian Zhang, Yi Zeng*
603 Where, What, Why: Toward Explainable 3D-GS Watermarking, *Mingshu Cai, Jiajun Li, Osamu Yoshie, Yuya Ieiri, Yixuan Li*
604 Robust Spiking Neural Networks by Temporal Mutual Information, *Mengting Xu, Shi Gu, Peng Lin, De Ma, Huajin Tang, Qian Zheng, Gang Pan*
605 TraceGen: World Modeling in 3D Trace Space Enables Learning from Cross-Embodiment Videos, *Seungjae Lee, Yoonkyo Jung, Inkook Chun, Yao-Chih Lee, Zikui Cai, Hongjia Huang, Aayush Talreja, Tan Dao, Yongyuan Liang, Jia-Bin Huang, Furong Huang*
606 HiF-VLA: Hindsight, Insight and Foresight through Motion Representation for Vision-Language-Action Models, *Minghui Lin, Pengxiang Ding, Shu Wang, Zifeng Zhuang, Yang Liu, Xinyang Tong, Wenxuan Song, Shangke Lyu, Siteng Huang, Donglin Wang*
607 AtomicVLA: Unlocking the Potential of Atomic Skill Learning in Robots, *Likui Zhang, Tao Tang, Zhihao Zhan, Xiuwei Chen, Zisheng Chen, Jianhua Han, Jiangtong Zhu, Pei Xu, Hang Xu, Hefeng Wu, Liang Lin, Xiaodan Liang*
608 Obstruction Reasoning for Robotic Grasping, *Runyu Jiao, Matteo Bortolon, Francesco Giuliari, Alice Fasoli, Sergio Povoli, Guofeng Mei, Yiming Wang, Fabio Poiesi*
609 PointWorld: Scaling 3D World Models for In-The-Wild Robotic
❊ Manipulation, *Wenlong Huang, Yu-Wei Chao, Arsalan Mousavian, Ming-Yu Liu, Dieter Fox, Kaichun Mo, Li Fei-Fei*
610 CycleManip: Enabling Cycle-based Manipulation via Effective
❊ History Perception and Understanding, *Yi-Lin Wei, Haoran Liao, Yuhao Lin, Pengyue Wang, Zhizhao Liang, Guiliang Liu, Wei-Shi Zheng*
611 SIMPACT: Simulation-Enabled Action Planning using Vision-Language Models, *Haowen Liu, Shaoxiong Yao, Haonan Chen, Jiawei Gao, Jiayuan Mao, Jia-Bin Huang, Yilun Du*
612 Adaptive Action Chunking at Inference-time for Vision-Language-Action Models, *Yuanchang Liang, Xiaobo Wang, Kai Wang, Shuo Wang, Xiaojiang Peng, Haoyu Chen, David Kim Huat Chua, Prahlad Vadakkepat*
613 Localizing, Structuring, and Rendering: Bridging 3D and 2D Vision-Language-Action Models for Robotic Manipulation, *Yunlong Zhao, Xiaoheng Deng, Yichao Cao, Yi Chen, Xiangjian He, Shan You, Shuo Yang, Lei Fan, Fei Wang, Xiu Su*
614 NIL: No-data Imitation Learning, *Mert Albaba, Chenhao Li, Markos Diomataris, Omid Taheri, Andreas Krause, Michael J. Black*
615 Humanoid Generative Pre-Training for Zero-Shot Motion Tracking, *Zekun Qi, Xuchuan Chen, Jilong Wang, Chenghuai Lin, Yunrui Lian, Wenyao Zhang, Xinqiang Yu, He Wang, Li Yi*
616 EnergyAction: Unimanual to Bimanual Composition with Energy-
❊ Based Models, *Mingchen Song, Xiang Deng, Jie Wei, Dongmei Jiang, Liqiang Nie, Weili Guan*
617 CUBic: Coordinated Unified Bimanual Perception and Control Framework, *Xingyu Wang, Pengxiang Ding, Jingkai Xu, Donglin Wang, Zhaoxin Fan*
618 RehearseVLA: Simulated Post-Training for VLAs with Physically-Consistent World Model, *Junjin Xiao, Yandan Yang, Xinyuan Chang, Ronghan Chen, Feng Xiong, Mu Xu, Wei-Shi Zheng, Qing Zhang*
619 GraspGen-X: Cross-Embodiment 6-DOF Diffusion-based Grasping, *Beining Han, Yu-Wei Chao, Erwin Coumans, Clemens Eppner, Jia Deng, Stan Birchfield, Adithyavairavan Murali*
620 UETrack: A Unified and Efficient Framework for Single Object Tracking, *Ben Kang, Jie Zhao, Xin Chen, Wanting Geng, Bin Zhang, Lu Zhang, Dong Wang, Huchuan Lu*
621 ProgTrack: A Multi-Object Tracking Algorithm with Progressive Matching Strategy, *Chenhui Zhang, Guoqing Dong, Weijie Peng*
622 Efficient Video Object Segmentation and Tracking with Recurrent Dynamic Submodel, *Weidong Tang, Zhiyuan Liang, Xinyan Wan, Chen Zhu, Zhaopan Xu, Pengfei Zhou, Yan Song, Yang You, Wangbo Zhao*
623 Learning to Track Instance from Single Nature Language Description,
❊ *Yaozong Zheng, Bineng Zhong, Qihua Liang, Shuimu Zeng, Haiying Xia, Shuxiang Song*
624 MV-TAP: Tracking Any Point in Multi-View Videos, *Jahyeok Koo, Inès Hyeonsu Kim, Mungyeom Kim, Junghyun Park, Seohyeon Park, Jaeyeong Kim, Jung Yi, Seokju Cho, Seungryong Kim*
625 Adaptive Depth Lightweight RGB-T Tracking with Holistic Token
❊ Routing, *Tian Ding, Hongtao Yang, Liangtao Shi, Jun Li, Xiantao Hu, Jian Yang, Ying Tai*
626 Content-Adaptive Hierarchical Hyperprior for Neural Video

Coding, *Junqi Liao, Yaojun Wu, Chaoyi Lin, Zhipin Deng, Li Li, Dong Liu, Xiaoyan Sun*

627 UTPTrack: Towards Simple and Unified Token Pruning for Visual Tracking, *Hao Wu, Xudong Wang, Jialiang Zhang, Junlong Tong, Xinghao Chen, Junyan Lin, Yunpu Ma, Xiaoyu Shen*

628 ❊ Similarity-as-Evidence: Calibrating Overconfident VLMs for Interpretable and Label-Efficient Medical Active Learning, *Zhuofan Xie, Zishan Lin, Jinliang Lin, Jie Qi, Shaohua Hong, Shuo Li*

629 From Infusion to Assimilation Distillation for Medical Image Segmentation, *Jiankang Hong, Ye Luo, Yinan Liu, Junsong Yuan*

630 IBISAgent: Reinforcing Pixel-Level Visual Reasoning in MLLMs for Universal Biomedical Object Referring and Segmentation, *Yankai Jiang, Qiaoru Li, Binlu Xu, Haoran Sun, Chao Ding, Junting Dong, Yuxiang Cai, Xuhong Zhang, Jianwei Yin*

631 Unlocking Positive Transfer in Incrementally Learning Surgical Instruments: A Self-reflection Hierarchical Prompt Framework, *Yu Zhu, Kang Li, Zheng Li, Pheng-Ann Heng*

632 ❊ Keep It Frozen: Domain-Routed Conditional Residual Modulation for Multi-Domain Vision Transformers, *Ufaq Khan, Umair Nawaz, Massimo Caputo, Muhammad Bilal, Junaid Qadir, Muhammad Haris Khan*

633 Virtual Full-stack Scanning of Brain MRI via Imputing Any Quantised Code, *Yicheng Wu, Tao Song, Zhonghua Wu, Jin Ye, Zongyuan Ge, Wenjia Bai, Zhaolin Chen, Jianfei Cai*

634 MedLoc-R1: Performance-Aware Curriculum Reward Scheduling for GRPO-Based Medical Visual Grounding, *Guangjing Yang, Ziyuan Qin, Chaoran Zhang, Chenlin Du, Jinglin Wang, Wanran Sun, Zhenyu Zhang, Bing Ji, Qicheng Lao*

635 Turning Pre-Trained Vision Transformers into End-to-End Histopathology Whole Slide Image Models for Survival Prediction, *Jiawen Li, Jiali Hu, Xitong Ling, Renao Yan, Yuxuan Chen, Tian Guan, Yonghong He*

636 A Supervised Multi-task Framework for Joint cryo-ET Restoration Enabled by Generative Physical Simulation, *Xinsheng Wang, Zhidong Yang, Xiaohua Wan, Renmin Han, Shuai Tang, Hao Dong, Fa Zhang, Bin Hu*

637 KAMP: Knowledge-Anchored Multimodal Pretraining Framework for Medical Image Representation, *Feiyu Huang, Jia Li, Zhao Chen, Yang Wu, Caleb Chen Cao, Lei Chen*

638 ❊ CARE: A Molecular-Guided Foundation Model with Adaptive Region Modeling for Whole Slide Image Analysis, *Di Zhang, Zhangpeng Gong, Xiaobo Pang, Jiashuai Liu, Junbo Lu, Hao Cui, Jiusong Ge, Zhi Zeng, Kai Yi, Yinghua Li, Si Liu, Tingsong Yu, Haoran Wang, Mireia Crispin-Ortuzar, Weimiao Yu, Chen Li, Zeyu Gao*

639 Contrastive Cross-Bag Augmentation for Multiple Instance Learning-based Whole Slide Image Classification, *Bo Zhang, Xinan Xu, Shuo Yan, Yu Bai, Zheng Zhang, Wufan Wang, Hui Gao, Wendong Wang*

640 OmniFM: Toward Modality-Robust and Task-Agnostic Federated Learning for Heterogeneous Medical Imaging, *Meilin Liu, Jiaying Wang, Jing Shan*

641 Learning complete and explainable visual representations from itemized text supervision, *Yiwei Lyu, Chenhui Zhao, Soumyanil Banerjee, Shixuan Liu, Akshay Rao, Akhil Kondepudi, Honglak Lee, Todd C. Hollon*

642 ❊ EgoPoseFormer v2: Accurate Egocentric Human Motion Estimation for AR/VR, *Zhenyu Li, Sai Kumar Dwivedi, Filip Maric, Carlos Chacón, Nadine Bertsch, Filippo Arcadu, Tomas Hodan, Michael Ramamonjisoa, Peter Wonka, Amy Zhao, Robin Kips, Cem Keskin, Anastasia Tkach, Chenhongyi Yang*

643 MetricHMSR: Metric Human Mesh and Scene Recovery from Monocular Images, *Chentao Song, He Zhang, Haolei Yuan, Haozhe Lin, Jianhua Tao, Hongwen Zhang, Tao Yu*

644 Differentially Private 2D Human Pose Estimation, *Kaushik Bhargav Sivangi, Paul Henderson, Fani Deligianni*

645 TROPHIES: Temporal Reconstruction of Places, Humans, and Cameras from Multi-view Videos, *Jinpeng Liu, Yukang Xu, Yutong Li, Xingyu Liu*

646 ❊ PoseD-Flow: Versatile and Guided Flow Matching Model of Human Pose, *Jebastin Nadar, Simone Foti, Tolga Birdal*

647 SIMSPINE: A Biomechanics-Aware Simulation Framework for 3D Spine Motion Annotation and Benchmarking, *Muhammad Saif Ullah Khan, Didier Stricker*

648 HUMAPS-4D: A Multimodal Dataset for HUman Motion Analysis with Physiological and Semantic informations, *Matthieu Dabrowski, Ouala Ben Jemaa, Benjamin Allaert*

649 ❊ PHASE-Net: Physics-Grounded Harmonic Attention System for Efficient Remote Photoplethysmography Measurement, *Bo Zhao, Dan Guo, Junzhe Cao, Yong Xu, Bochao Zou, Tao Tan, Yue Sun, Zitong Yu*

650 LAMP: Localization Aware Multi-camera People Tracking in Metric 3D World, *Nan Yang, Julian Straub, Fan Zhang, Richard Newcombe, Jakob Engel, Lingni Ma*

651 Expanding mmWave Datasets for Human Pose Estimation with Unlabeled Data and LiDAR Datasets, *Zhuoxuan Peng, Boan Zhu, Xingjian Zhang, Wenying Li, S.-H. Gary Chan*

652 Towards Balanced Multi-Modal Learning in 3D Human Pose Estimation, *Mengshi Qi, Jiaxuan Peng, Xianlin Zhang, Huadong Ma*

653 OMGTex: One-stage Multi-style Facial Texture Reconstruction without Geometry Guidance, *Xiao Zitong, Yuda Qiu, Zisheng Ye, Xiaoguang Han*

654 ❊ Human Interaction-Aware 3D Reconstruction from a Single Image, *Gwanghyun Kim, Junghun James Kim, Suh Yoon Jeon, Jason Park, Se Young Chun*

655 Towards Generalizable AI-Generated Image Detection via Image-Adaptive Prompt Learning, *Yiheng Li, Zichang Tan, Guoqing Xu, Zhen Lei, Xu Zhou, Yang Yang*

656 SAGA: Source Attribution of Generative AI Videos, *Rohit Kundu, Vishal Mohanty, Hao Xiong, Shan Jia, Athula Balachandran, Amit K. Roy-Chowdhury*

657 VMD-FACT: A New Video Dataset and MLLM-based method for Detecting Realistic AI-Generated Video Misinformation, *Yongkang Zhang, Dongyu She, Baiyu Ji, Qichuan Geng, Zhong Zhou, Yan Wang*

658 ReAlign: Generalizable Image Forgery Detection via Reasoning-Aligned Representation, *Qing Huang, Zhipei Xu, Xuanyu Zhang, Xiangyu Yu, Jian Zhang*

659 A Sanity Check for Multi-In-Domain Face Forgery Detection in the Real World, *Jikang Cheng, Renye Yan, Zhiyuan Yan, Yaozhong Gan, Xueyi Zhang, Zhongyuan Wang, Wei Peng, Ling Liang*

660 PPM-CLIP: Probabilistic Prompt Modeling for Generalizable AI-Generated Image Detection, *Xinyuan Wang, Yingxin Lai, Zhiming Luo, Zhihui Liu*

661 Learning from Noisy Supervision: A Denoising–Debiasing Framework for Weakly Supervised Video Anomaly Detection, *Yaxin Zhao, Yang Wang, Wenya Guo, Sihan Xu, Xiangrui Cai, Xi Lin, Ying Zhang, Xiaojie Yuan*

662 Anomaly as Non-Conformity via Training-Free Graph Laplacian Energy Minimization, *Jungwook Seo, Minjeong Kim, Younkwan Lee, Seungho Shin, Sungyong Baik*

663 VisualAD: Language-Free Zero-Shot Anomaly Detection via Vision Transformer, *Yanning Hou, Peiyuan Li, Zirui Liu, Yitong Wang, Yanran Ruan, Jianfeng Qiu, Ke Xu*

664 CHAL: Causal-guided Hierarchical Anomaly-aware Learning for Moving Infrared Small Target Detection, *Weiwei Duan, Luping Ji, Shipeng Lei, Sicheng Zhu, Jianghong Huang, Mao Ye*

665 RAID: Retrieval-Augmented Anomaly Detection, *Mingxiu Cai, Zhe Zhang, Gaochang Wu, Tianyou Chai, Xiatian Zhu*

666 ADSeeker: A Knowledge-Grounded Reasoning Framework for Industry Anomaly Detection and Reasoning, *Kai Zhang, Zekai Zhang, Xihe Sun, Anpeng Wang, Jingmeng Nie, Qinghui Chen, Han Hao, Jianyuan Guo, Jinglin Zhang*

667 InvAD: Inversion-based Reconstruction-Free Anomaly Detection with Diffusion Models, *Shunsuke Sakai, Xiangteng He, Chunzhi Gu, Leonid Sigal, Tatsuhito Hasegawa*

668 QueryOcc: Query-based Self-Supervision for 3D Semantic Occupancy, *Adam Lilja, Ji Lan, Junsheng Fu, Lars Hammarstrand*

669 ❊ GSV2X: Geometry-Aware Uncertainty Modeling and Orthogonal Fusion for Robust Roadside Perception, *Jianqiang Xu, Gensheng Pei, Huafeng Liu, Yazhou Yao*

670 Grounded Latents for Entity-Centric 4D Scene Generation, *Jinhyung Park, Navyata Sanghvi, Erica Weng, Shawn Hunt, Shinya Tanaka, Hironobu Fujiyoshi, Kris Kitani*

11:45 - 13:45 DEMOS (ExHall F)

1 Mobile-O: Real-Time Unified Multimodal Understanding and Generation on Mobile Devices, *Abdelrahman Shaker, Salman Khan*
2 Stereo Foveated Camera for High-quality Long-range Depth Sensing, *Yuxuan Zhang, Jacob Carter, Noah Ralph, Michael Tomadakis, Sanjeev Koppal*
3 Color Beyond Capture: A Two-Ink Printing Process, *Christopher Swift*
4 KV-Tracker: Real-Time Pose Tracking with Transformers, *Marwan Taher, Ignacio Alzugaray, Kirill Mazur, Xin Kong, Andrew J. Davison*
5 ARVRag: Real-Time Retrieval-Augmented Object Detection and Guidance in Augmented Reality, *Alireza Taheritajar, Jieqiong Zhao, Jason Orlosky*
6 Pulse of Motion: Measuring Physical Frame Rate from Visual Dynamics, *Xiangbo Gao, Mingyang Wu, Siyuan Yang, Jiongze Yu, Pardis Taghavi, Fangzhou Lin, Zhengzhong Tu*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

14:00 - 15:15 Oral Session 4A: Geometric Understanding (Bluebird Ballroom)

🏆 - Award candidate paper

1 🏆 Chorus: Multi-Teacher Pretraining for Holistic 3D Gaussian Scene Encoding, *Yue Li, Qi Ma, Runyi Yang, Mengjiao Ma, Bin Ren, Nikola Popovic, Nicu Sebe, Theo Gevers, Luc Van Gool, Danda Pani Paudel, Martin R. Oswald*
2 Featurising Pixels from Dynamic 3D Scenes with Linear In-Context Learners, *Nikita Araslanov, Martin Sundermeyer, Hidenobu Matsuki, David Joseph Tan, Federico Tombari*
3 From Pairs to Sequences: Track-Aware Policy Gradients for Keypoint Detection, *Yepeng Liu, Hao Li, Liwen Yang, Fangzhen Li, Xudi Ge, Yuliang Gu, Kuang Gao, Bing Wang, Guang Chen, Hangjun Ye, Yongchao Xu*
4 🏆 Linear Fundamental Matrix Estimation from 7 or 5 Points, *Taci Ata Kucukpinar, Juan Mogollon, Joshua Fraser, Timothy Duff, Kannappan Palaniappan*
5 OccuFly: A 3D Vision Benchmark for Semantic Scene Completion from the Aerial Perspective, *Markus Gross, Sai B. Matha, Aya Fahmy, Rui Song, Daniel Cremers, Henri Meeß*
6 🏆 VGGT-Ω, *Jianyuan Wang, Minghao Chen, Shangzhan Zhang, Nikita Karaev, Johannes Schönberger, Patrick Labatut, Piotr Bojanowski, David Novotny, Andrea Vedaldi, Christian Rupprecht*

14:00 - 15:15 Oral Session 4B: Embodied & Agentic Intelligence (Four Seasons Ballroom)

1 CodeV: Code with Images for Faithful Visual Reasoning via Tool-Aware Policy Optimization, *Xinhai Hou, Shaoyuan Xu, Manan Biyani, Moyan Li, Jia Liu, Todd C Hollon, Bryan Wang*
2 🏆 NitroGen: An Open Foundation Model for Generalist Gaming Agents, *Loïc Magne, Anas Awadalla, Guanzhi Wang, Yinzhen Xu, Joshua Belofsky, Fengyuan Hu, Joohwan Kim, Ludwig Schmidt, Georgia Gkioxari, Jan Kautz, Yisong Yue, Yejin Choi, Yuke Zhu, Linxi Fan*
3 🏆 PAI-Bench: A Comprehensive Benchmark For Physical AI, *Fengzhe Zhou, Jiannan Huang, Jialuo Li, Deva Ramanan, Humphrey Shi*
4 🏆 RefAV: Towards Planning-Centric Scenario Mining, *Cainan Davidson, Deva Ramanan, Neehar Peri*
5 🏆 SoccerMaster: A Vision Foundation Model for Soccer Understanding, *Haolin Yang, Jiayuan Rao, Haoning Wu, Weidi Xie*
6 VS-Bench: Evaluating VLMs for Strategic Abilities in Multi-Agent Environments, *Zelai Xu, Zhexuan Xu, Xiangmin Yi, Huining Yuan, Mo Guang, Kaiwen Long, Xinlei Chen, Yi Wu, Chao Yu, Yu Wang*

14:00 - 15:15 Oral Session 4C: Spatial Reasoning (Mile High Ballroom 1A - 2A)

1 Breaking the Scalability Limit of Multi-Projector Calibration with Embedded Cameras, *Takumi Kawano, Kohei Miura, Daisuke Iwai*
2 🏆 GaussianFluent: Gaussian Simulation for Dynamic Scenes with Mixed Materials, *Bei Huang, Yixin Chen, Ruijie Lu, Gang Zeng, Hongbin Zha, Yuru Pei, Siyuan Huang*
3 InfiniBench: Infinite Benchmarking for Visual Spatial Reasoning with Customizable Scene Complexity, *Haoming Wang, Qiyao Xue, Wei Gao*
4 MAGICIAN: Efficient Long-Term Planning with Imagined Gaussians for Active Mapping, *Shiyao Li, Antoine Guédon, Shizhe Chen, Vincent Lepetit*
5 🏆 Memory-Augmented Scene Understanding and Exploration for Open-World Aerial Object-Goal Navigation, *Jiacong Zhou, Jiaxu Miao, Yourun Lin, Xianyun Wang, Jun Xiao, Jun Yu*
6 🏆 Monocular Open Vocabulary Occupancy Prediction for Indoor Scenes, *Changqing Zhou, Yueru Luo, Han Zhang, Zeyu Jiang, Changhao Chen*

14:00 - 15:15 Oral Session 4D: Visual Segmentation (Mile High Ballroom 3A - 4A)

1 INSID3: Training-Free In-Context Segmentation with DINOv3, *Claudia Cuttano, Gabriele Trivigno, Christoph Reich, Daniel Cremers, Carlo Masone, Stefan Roth*
2 MARCO: Navigating the Unseen Space of Semantic Correspondence, *Claudia Cuttano, Gabriele Trivigno, Carlo Masone, Stefan Roth*
3 PR-MaGIC: Prompt Refinement Via Mask Decoder Gradient Flow For In-Context Segmentation, *Minjae Lee, Sungwoo Hur, Soojin Hwang, Won Hwa Kim*
4 🏆 R^2-Seg: Training-Free OOD Medical Tumor Segmentation via Anatomical Reasoning and Statistical Rejection, *Shuaike Shen, Ke Liu, Jiaqing Xie, Shangde Gao, Chunhua Shen, Ge Liu, Mireia Crispin-Ortuzar, Shangqi Gao*
5 🏆 The SA-FARI Dataset: Segment Anything in Footage of Animals for Recognition and Identification, *Dante Wasmuht, Otto Brookes, Maximilian Schall, Pablo Palencia, Christopher Beirne, Tilo Burghardt, Majid Mirmehdi, Hjalmar Kühl, Mimi Arandjelovic, Sam Pottie, Peter Bermant, Brandon Asheim, Yi Jin Toh, Adam Elzinga, Jason Allan Holmberg, Andrew Whitworth, Eleanor Flatt, Laura Gustafson, Chaitanya Ryali, Yuan-Ting Hu, Baishan Guo, Andrew Westbury, Kate Saenko, Didac Suris*
6 🏆 VGGT-Segmentor: Geometry-Enhanced Cross-View Segmentation, *Yulu Gao, Bohao Zhang, Zongheng Tang, Jitong Liao, Wenjun Wu, Si Liu*

16:45 - 18:45 Poster Session 4 & Exhibit Hall w/ Coffee Break (ExHall A)

❇ - Highlight paper 🏆 - Award candidate paper

1 🏆 Chorus: Multi-Teacher Pretraining for Holistic 3D Gaussian Scene Encoding, *Yue Li, Qi Ma, Runyi Yang, Mengjiao Ma, Bin Ren, Nikola Popovic, Nicu Sebe, Theo Gevers, Luc Van Gool, Danda Pani Paudel, Martin R. Oswald*
2 Featurising Pixels from Dynamic 3D Scenes with Linear In-Context Learners, *Nikita Araslanov, Martin Sundermeyer, Hidenobu Matsuki, David Joseph Tan, Federico Tombari*
3 From Pairs to Sequences: Track-Aware Policy Gradients for Keypoint Detection, *Yepeng Liu, Hao Li, Liwen Yang, Fangzhen Li, Xudi Ge, Yuliang Gu, Kuang Gao, Bing Wang, Guang Chen, Hangjun Ye, Yongchao Xu*
4 🏆 Linear Fundamental Matrix Estimation from 7 or 5 Points, *Taci Ata Kucukpinar, Juan Mogollon, Joshua Fraser, Timothy Duff, Kannappan Palaniappan*
5 OccuFly: A 3D Vision Benchmark for Semantic Scene Completion from the Aerial Perspective, *Markus Gross, Sai B. Matha, Aya Fahmy, Rui Song, Daniel Cremers, Henri Meeß*
6 🏆 VGGT-Ω, *Jianyuan Wang, Minghao Chen, Shangzhan Zhang, Nikita Karaev, Johannes Schönberger, Patrick Labatut, Piotr Bojanowski, David Novotny, Andrea Vedaldi, Christian Rupprecht*
7 CodeV: Code with Images for Faithful Visual Reasoning via Tool-Aware Policy Optimization, *Xinhai Hou, Shaoyuan Xu, Manan Biyani, Moyan Li, Jia Liu, Todd C Hollon, Bryan Wang*
8 🏆 NitroGen: An Open Foundation Model for Generalist Gaming Agents, *Loïc Magne, Anas Awadalla, Guanzhi Wang, Yinzhen Xu, Joshua Belofsky, Fengyuan Hu, Joohwan Kim, Ludwig Schmidt, Georgia Gkioxari, Jan Kautz, Yisong Yue, Yejin Choi, Yuke Zhu, Linxi Fan*
9 🏆 PAI-Bench: A Comprehensive Benchmark For Physical AI, *Fengzhe Zhou, Jiannan Huang, Jialuo Li, Deva Ramanan, Humphrey Shi*
10 🏆 RefAV: Towards Planning-Centric Scenario Mining, *Cainan Davidson, Deva Ramanan, Neehar Peri*
11 🏆 SoccerMaster: A Vision Foundation Model for Soccer Understanding, *Haolin Yang, Jiayuan Rao, Haoning Wu, Weidi Xie*

12 VS-Bench: Evaluating VLMs for Strategic Abilities in Multi-Agent Environments, *Zelai Xu, Zhexuan Xu, Xiangmin Yi, Huining Yuan, Mo Guang, Kaiwen Long, Xinlei Chen, Yi Wu, Chao Yu, Yu Wang*
13 Breaking the Scalability Limit of Multi-Projector Calibration with Embedded Cameras, *Takumi Kawano, Kohei Miura, Daisuke Iwai*
14 GaussianFluent: Gaussian Simulation for Dynamic Scenes with Mixed
🏆 Materials, *Bei Huang, Yixin Chen, Ruijie Lu, Gang Zeng, Hongbin Zha, Yuru Pei, Siyuan Huang*
15 InfiniBench: Infinite Benchmarking for Visual Spatial Reasoning with Customizable Scene Complexity, *Haoming Wang, Qiyao Xue, Wei Gao*
16 MAGICIAN: Efficient Long-Term Planning with Imagined Gaussians for Active Mapping, *Shiyao Li, Antoine Guédon, Shizhe Chen, Vincent Lepetit*
17 Memory-Augmented Scene Understanding and Exploration for Open-
🏆 *World Aerial Object-Goal Navigation, Jiacong Zhou, Jiaxu Miao, Yourun Lin, Xianyun Wang, Jun Xiao, Jun Yu*
18 Monocular Open Vocabulary Occupancy Prediction for Indoor Scenes,
🏆 *Changqing Zhou, Yueru Luo, Han Zhang, Zeyu Jiang, Changhao Chen*
19 INSID3: Training-Free In-Context Segmentation with DINOv3, *Claudia Cuttano, Gabriele Trivigno, Christoph Reich, Daniel Cremers, Carlo Masone, Stefan Roth*
20 MARCO: Navigating the Unseen Space of Semantic Correspondence, *Claudia Cuttano, Gabriele Trivigno, Carlo Masone, Stefan Roth*
21 PR-MaGIC: Prompt Refinement Via Mask Decoder Gradient Flow For In-Context Segmentation, *Minjae Lee, Sungwoo Hur, Soojin Hwang, Won Hwa Kim*
22 R^2-Seg: Training-Free OOD Medical Tumor Segmentation via
🏆 *Anatomical Reasoning and Statistical Rejection, Shuaike Shen, Ke Liu, Jiaqing Xie, Shangde Gao, Chunhua Shen, Ge Liu, Mireia Crispin-Ortuzar, Shangqi Gao*
23 The SA-FARI Dataset: Segment Anything in Footage of Animals for
🏆 Recognition and Identification, *Dante Wasmuht, Otto Brookes, Maximilian Schall, Pablo Palencia, Christopher Beirne, Tilo Burghardt, Majid Mirmehdi, Hjalmar Kühl, Mimi Arandjelovic, Sam Pottie, Peter Bermant, Brandon Asheim, Yi Jin Toh, Adam Elzinga, Jason Allan Holmberg, Andrew Whitworth, Eleanor Flatt, Laura Gustafson, Chaitanya Ryali, Yuan-Ting Hu, Baishan Guo, Andrew Westbury, Kate Saenko, Didac Suris*
24 VGGT-Segmentor: Geometry-Enhanced Cross-View Segmentation,
🏆 *Yulu Gao, Bohao Zhang, Zongheng Tang, Jitong Liao, Wenjun Wu, Si Liu*
25 DAGE: Dual-Stream Architecture for Efficient and Fine-Grained Geometry Estimation, *Tuan Duc Ngo, Jiahui Huang, Seoung Wug Oh, Kevin Blackburn-Matzen, Evangelos Kalogerakis, Chuang Gan, Joon-Young Lee*
26 Wave-Former: Through-Occlusion 3D Reconstruction via Wireless Shape Completion, *Laura Dodds, Maisy Lam, Waleed Akbar, Yibo Cheng, Fadel Adib*
27 Lite Any Stereo: Efficient Zero-Shot Stereo Matching, *Junpeng Jing, Weixun Luo, Ye Mao, Krystian Mikolajczyk*
28 MuM: Multi-View Masked Image Modeling for 3D Vision, *David Nordström, Johan Edstedt, Fredrik Kahl, Georg Bökman*
29 ZipMap: Linear-Time Stateful 3D Reconstruction via Test-Time Training, *Haian Jin, Rundi Wu, Tianyuan Zhang, Ruiqi Gao, Jonathan T. Barron, Noah Snavely, Aleksander Hołyński*
30 Scal3R: Scalable Test-Time Training for Large-Scale 3D Reconstruction,
❇ *Tao Xie, Peishan Yang, Yudong Jin, Yingfeng Cai, Wei Yin, Weiqiang Ren, Qian Zhang, Wei Hua, Sida Peng, Xiaoyang Guo, Xiaowei Zhou*
31 LaRP: Efficient Multi-View Inpainting with Latent Reprojection Priors, *Gaoyang Zhang, Xinguo Liu*
32 TopoMA: Topology-Guided Multi-Agent Dense RGB 3D Reconstruction via Distributed Inference, *Xuanxuan Zhang, ShuHui Shi, Tianxiang Zhang, Zhetao Guo, Huang Zixuan, You Li*
33 Sparse–View Localization via Online Neural 3D Regression, *Ludvig Dillén, Magnus Oskarsson, Viktor Larsson*
34 Dynamic Visual SLAM using a General 3D Prior, *Xingguang Zhong, Liren Jin, Marija Popovic, Jens Behley, Cyrill Stachniss*
35 Learning Scene Coordinate Reconstruction from Unposed Images via
❇ Pose Graph Optimization, *Tze Ho Elden Tse, Jizong Peng, Angela Yao*
36 FlashVGGT: Efficient and Scalable Visual Geometry Transformers with Compressed Descriptor Attention, *Zipeng Wang, Dan Xu*
37 No Calibration, No Depth, No Problem: Cross-Sensor View Synthesis
❇ with 3D Consistency, *Cho-Ying Wu, Zixun Huang, Xinyu Huang, Liu Ren*
38 UFO: Unifying Feed-Forward and Optimization-based Methods for Large Driving Scene Modeling, *Kaiyuan Tan, Yingying Shen, Ziyue Zhu, Mingfei Tu, Haohui Zhu, Haiyang Sun, Bing Wang, Guang Chen, Hangjun Ye*
39 Reliev3R: Relieving Feed-forward 3D Reconstruction from Multi-View Geometric Annotations, *Youyu Chen, Junjun Jiang, Yueru Luo, Kui Jiang, Xianming Liu, Xu Yan, Dave Zhenyu Chen*
40 TALO: Pushing 3D Vision Foundation Models Towards Globally Consistent Online Reconstruction, *Fengyi Zhang, Tianjun Zhang, Kasra Khosoussi, Zheng Zhang, Zi Huang, Yadan Luo*
41 Global Structure-from-Motion Meets Feedforward Reconstruction,
❇ *Linfei Pan, Johannes Schönberger, Marc Pollefeys*
42 POCA: Pareto-Optimal Curriculum Alignment for Visual Text Generation, *Yaohou Fan, Qingzhong Wang, Yongsong Huang, Junyi Liu, Tomo Miyazaki, Shinichiro Omachi*
43 DuoGen: Towards Autonomous Interleaved Multimodal Generation,
❇ *Min Shi, Xiaohui Zeng, Jiannan Huang, Yin Cui, Francesco Ferroni, Jialuo Li, Zhaoshuo Li, Yogesh Balaji, Haoxiang Wang, Tsung-Yi Lin, Xiao Fu, Yue Zhao, Chieh-Yun Chen, Ming-Yu Liu, Humphrey Shi*
44 Vibe Spaces for Creatively Connecting and Expressing Visual Concepts, *Huzheng Yang, Katherine Xu, Andrew Lu, Michael D. Grossberg, Yutong Bai, Jianbo Shi*
45 StoryTailor:A Zero-Shot Pipeline for Action-Rich Multi-Subject Visual Narratives, *Jinghao Hu, Yuhe Zhang, Guohua Geng, Kang Li, Han Zhang*
46 CREward: A Type-Specific Creativity Reward Model, *Jiyeon Han, Ali Mahdavi-Amiri, Hao Zhang, Haedong Jeong*
47 LumiX: Structured and Coherent Text-to-Intrinsic Generation, *Xu Han, Biao Zhang, Xiangjun Tang, Xianzhi Li, Peter Wonka*
48 Synthetic Curriculum Reinforces Compositional Text-to-Image Generation, *Shijian Wang, Runhao Fu, Siyi Zhao, Qingqin Zhan, Xingjian Wang, Jiarui Jin, Yuan Lu, Hanqian Wu, Cunjian Chen*
49 OmniGen2: Towards Instruction-Aligned Multimodal Generation, *Chenyuan Wu, Jiahao Wang, Pengfei Zheng, Ruiran Yan, Shitao Xiao, Xin Luo, Yueze Wang, Wanli Li, Xiyan Jiang, Yexin Liu, Junjie Zhou, Ziyi Xia, Ze Liu, Chaofan Li, Haoge Deng, Kun Luo, Bo Zhang, Jiajun Zhang, Dong Liu, Defu Lian, Xinlong Wang, Zhongyuan Wang, Tiejun Huang, Zheng Liu*
50 Selectively Extracting and Injecting Visual Attributes into Text-to-Image Models, *Seunghwan Choi, Jooyeol Yun, Youngdo Lee, Jaegul Choo*
51 LoFA: Learning to Predict Personalized Prior for Fast Adaptation of Visual Generative Models, *Yiming Hao, Mutian Xu, Chongjie Ye, Jie Qin, Shunlin Lu, Yipeng Qin, Xiaoguang Han*
52 UniVerse: Empower Unified Generation with Reasoning and Knowledge, *Kaiyue Sun, Weiyang Jin, Chengqi Duan, Rongyao Fang, Xian Liu, Yuwei Niu, Chunwei Wang, Aoxue Li, Xihui Liu*
53 UniVerse: A Unified Modulation Framework for Segmentation-Free, Disentangled Multi-Concept Personalization, *Quynh Phung, Sandesh Ghimire, Minsi Hu, Chung-Chi Tsai, Jia-Bin Huang*
54 Residual Decoder Adapter: ID-Preserving Tokenizer Adaption for Autoregressive Text Rendering, *Dongxing Mao, Alex Jinpeng Wang, Jiahao Tang, Kevin Qinghong Lin, Linjie Li, Zhengyuan Yang, Lijuan Wang, Min Li, Jingru Tan*
55 TGT: Text-Grounded Trajectories for Locally Controlled Video Generation, *Guofeng Zhang, Angtian Wang, Jacob Zhiyuan Fang, Liming Jiang, Haotian Yang, Bo Liu, Yiding Yang, Guang Chen, Longyin Wen, Alan Yuille, Chongyang Ma*
56 RAISE: Requirement-Adaptive Evolutionary Refinement for Training-Free Text-to-Image Alignment, *Liyao Jiang, Ruichen Chen, Chao Gao, Di Niu*
57 FlowFixer: Towards Detail-Preserving Subject-Driven Generation, *Jinyoung Jun, Won-Dong Jang, Wenbin Ouyang, Raghudeep Gadde, Jungbeom Lee*
58 TextPecker: Rewarding Structural Anomaly Quantification for Enhancing Visual Text Rendering, *Hanshen Zhu, Yuliang Liu, Xuecheng Wu, An-Lan Wang, Hao Feng, Dingkang Yang, Chao Feng, Can Huang, Jingqun Tang, Xiang Bai*
59 UltraFlux: Data-Model Co-Design for High-quality Native 4K Text-to-
❇ Image Generation across Diverse Aspect Ratios, *Tian Ye, Song Fei, Lei Zhu*
60 FEAT: Fashion Editing and Try-On from Any Design, *Soye Kwon,*
❇ *Keonyoung Lee, Dahuin Jung, Jaekoo Lee*

61 Rethinking Prompt Design for Inference-time Scaling in Text-to-Visual Generation, *Subin Kim, Sangwoo Mo, Mamshad Nayeem Rizve, Yiran Xu, Difan Liu, Jinwoo Shin, Tobias Hinz*
62 PointAlign: Feature-Level Alignment Regularization for 3D Vision-Language Models, *Yuanhao Su, Shaofeng Zhang, Xiaosong Jia, Qi Fan*
63 PowerCLIP: Powerset Alignment for Contrastive Pre-Training, *Masaki Kawamura, Nakamasa Inoue, Rintaro Yanagi, Hirokatsu Kataoka, Rio Yokota*
64 MoBind: Motion Binding for Fine-Grained IMU–Video Pose Alignment, *Duc Duy Nguyen, Tat-Jun Chin, Minh Hoai*
65 ❄ The Geometry of Robustness: Optimizing Loss Landscape Curvature and Feature Manifold Alignment for Robust Finetuning of Vision-Language Models, *Shivang Chopra, Shaunak Halbe, Chengyue Huang, Brisa Maneechotesuwan, Zsolt Kira*
66 Tackling Model Bias via Game-theoretic Multi-agent Collaboration Framework for Hateful Meme Classification, *Yiwei Wei, Zhengliang Guo, Shaozu Yuan, Chengyin Hu, Zhiyang Jia, Jiujiang Guo, Meng Chen, Peiying Wang, Longbiao Wang*
67 CCCaption: Dual-Reward Reinforcement Learning for Complete and Correct Image Captioning, *Zhijiang Tang, Linhua Wang, Jiaxin Qi, Weihao Jiang, Peng Hou, Anxiang Zeng, Jianqiang Huang*
68 MM-ReCoder: Advancing Chart-to-Code Generation with Reinforcement Learning and Self-Correction, *Zitian Tang, Xu Zhang, Jianbo Yuan, Yang Zou, Varad Gunjal, Songyao Jiang, Davide Modolo*
69 Learning to Generate via Understanding: Understanding-Driven Intrinsic Rewarding for Unified Multimodal Models, *Jiadong Pan, Liang Li, Yuxin Peng, Yu-Ming Tang, Shuohuan Wang, Yu Sun, Hua Wu, Qingming Huang, Haifeng Wang*
70 Hierarchical Process Reward Models are Symbolic Vision Learners, *Shan Zhang, Aotian Chen, Kai Zou, Jindong Gu, Yuan Xue, Anton van den Hengel*
71 ARM-Thinker: Reinforcing Multimodal Generative Reward Models with Agentic Tool Use and Visual Reasoning, *Shengyuan Ding, Xinyu Fang, Ziyu Liu, Yuhang Zang, Yuhang Cao, Xiangyu Zhao, Haodong Duan, Xiaoyi Dong, Jianze Liang, Bin Wang, Conghui He, Dahua Lin, Jiaqi Wang*
72 SG-LoRA: Semantic-guided LoRA Parameters Generation, *Miaoge Li, Yang Chen, Zhijie Rao, Can Jiang, Kang Wei, Jingcai Guo*
73 AcTTA: Rethinking Test-Time Adaptation via Dynamic Activation, *Hyeongyu Kim, Geonhui Han, Dosik Hwang*
74 Reframing Long-Tailed Learning via Loss Landscape Geometry, *Shenghan Chen, Yiming Liu, Yanzhen Wang, Yujia Wang, Xiankai Lu*
75 ❄ Cleaning the Pool: Progressive Filtering of Unlabeled Pools in Deep Active Learning, *Denis Huseljic, Marek Herde, Lukas Rauch, Paul Hahn, Bernhard Sick*
76 DC-Merge: Improving Model Merging with Directional Consistency, *Han-Chen Zhang, Zi-Hao Zhou, Mao-Lin Luo, Shimin Di, Min-Ling Zhang, Tong Wei*
77 TALON: Test-time Adaptive Learning for On-the-Fly Category Discovery, *Yanan Wu, Yuhan Yan, Tailai Chen, Zhixiang Chi, ZiZhang Wu, Yi Jin, Yang Wang, Zhenbo Li*
78 Event-Illumination Collaborative Low-light Image Enhancement with a High-resolution Real-world Dataset, *Senyan Xu, Zhijing Sun, Kean Liu, Xin Lu, Ruixuan Jiang, Xueyang Fu, Zheng-Jun Zha*
79 NEC-Diff: Noise-Robust Event-RAW Complementary Diffusion for Seeing Motion in Extreme Darkness, *Haoyue Liu, Jinghan Xu, Luxin Feng, Hanyu Zhou, Haozhi Zhao, Yi Chang, Luxin Yan*
80 Towards Persistence: Learning Topological Constraints for Event-based Small Object Detection, *Shiman He, Nuo Chen, Xinyi Ying, Yihang Luo, Yangsi Shi, Zaiping Lin, Miao Li*
81 ❄ Geometric-Photometric Event-based 3D Gaussian Ray Tracing, *Kai Kohyama, Yoshimitsu Aoki, Guillermo Gallego, Shintaro Shiba*
82 EventDrive: Event Cameras for Vision-Language Driving Intelligence, *Dongyue Lu, Rong Li, Ao Liang, Lingdong Kong, Wei Yin, Lai Xing Ng, Benoit R. Cottereau, Camille Simon Chane, Wei Tsang Ooi*
83 EventGait: Towards Robust Gait Recognition with Event Streams, *Senyan Xu, Shuai Chen, Chuanfu Shen, Kean Liu, Zhijing Sun, Chengzhi Cao, Xueyang Fu*
84 MergeVLA: Cross-Skill Model Merging Toward a Generalist Vision-Language-Action Agent, *Yuxia Fu, Zhizhen Zhang, Yuqi Zhang, Zijian Wang, Zi Huang, Yadan Luo*
85 Resolving the Stability-Plasticity Dilemma in Reinforcement Learning via Complementary Continual Critics, *Bo Sun, Peixi Peng, Guang Tan, Haoran Xu, Yaokun Li, Yiqian Chang, Shuaixian Wang, Luntong Li*
86 SAGE: Scalable Agentic 3D Scene Generation for Embodied AI, *Hongchi Xia, Xuan Li, Zhaoshuo Li, Qianli Ma, Jiashu Xu, Ming-Yu Liu, Yin Cui, Tsung-Yi Lin, Wei-Chiu Ma, Shenlong Wang, Shuran Song, Fangyin Wei*
87 Semantic Audio-Visual Navigation in Continuous Environments, *Yichen Zeng, Hebaixu Wang, Meng Liu, Yu Zhou, Chen Gao, Kehan Chen, Gongping Huang*
88 Unifying Perception and Action: A Hybrid-Modality Pipeline with Implicit Visual Chain-of-Thought for Robotic Action Generation, *Xiangkai Ma, Lekai Xing, Han Zhang, Wenzhong Li, Sanglu Lu*
89 FLARE: A Failure-Aware Framework for Autonomous Correction and Recovery in Visual-Language Robotic Manipulation, *Ganlong Zhao, Zijia Tang, Xingping Chen, Zhanghui Kuang, Ye Tian, Guanbin Li*
90 Learning to Adapt: Self-Improving Web Agent via Cognitive-Aware Exploration, *Weile Chen, Bingchen Miao, Qifan Yu, Wendong Bu, Guoming Wang, Wenqiao Zhang, Shengyu Zhang, Juncheng Li, Siliang Tang*
91 General Process Reward Modeling for Robotic Reinforcement Learning, *Huajie Tan, Sixiang Chen, Yijie Xu, Zixiao Wang, Cheng Chi, Yuheng Ji, Yaoxu Lyu, Zhongxia Zhao, Xiansheng Chen, Peterson Co, Shaoxuan Xie, Guocai Yao, Pengwei Wang, Zhongyuan Wang, Shanghang Zhang*
92 DynBridge: Bridging Imagination and Control through Interaction Dynamics for Robot Manipulation, *Alex Wang, Zhiwei Dong, Qicheng Bai, Chenshi Zhang, Yujie Yi, Guang Dai, Yong Liu, Mengmeng Wang*
93 Action-Sketcher: From Reasoning to Action via Visual Sketches for Robotic Manipulation, *Huajie Tan, Peterson Co, Yijie Xu, Shanyu Rong, Yuheng Ji, Cheng Chi, Xiansheng Chen, Zhongxia Zhao, Pengwei Wang, Zhongyuan Wang, Shanghang Zhang*
94 Thinking in 360°: Humanoid Visual Search in the Wild, *Heyang Yu, Yinan Han, Xiangyu Zhang, Baiqiao Yin, Bowen Chang, Xiangyu Han, Xinhao Liu, Jing Zhang, Marco Pavone, Chen Feng, Saining Xie, Yiming Li*
95 Learning from Semantic Dictionaries: Discriminative Codebook Contrastive Learning for Unified Visual Representation and Generation, *Imanol G. Estepa, Jesús M. Rodríguez-de-Vera, Bhalaji Nagarajan, Petia Radeva*
96 MagicQuill V2: Precise and Interactive Image Editing with Layered Visual Cues, *Zichen Liu, Yue Yu, Hao Ouyang, Qiuyu Wang, Shuailei Ma, Ka Leong Cheng, Wen Wang, Qingyan Bai, Yuxuan Zhang, Yanhong Zeng, Yixuan Li, Xing Zhu, Yujun Shen, Qifeng Chen*
97 Cycle-Consistent Tuning for Layered Image Decomposition, *Zheng Gu, Min Lu, Zhida Sun, Dani Lischinski, Daniel Cohen-Or, Hui Huang*
98 RealUnify: Do Unified Models Truly Benefit from Unification? A Comprehensive Benchmark, *Yang Shi, Yuhao Dong, Yue Ding, Yuran Wang, Xuanyu Zhu, Sheng Zhou, Wenting Liu, Haochen Tian, Rundong Wang, Huanqian Wang, Zuyan Liu, Bohan Zeng, Ruizhe Chen, Qixun Wang, Zhuoran Zhang, Xinlong Chen, Chengzhuo Tong, Bozhou Li, Qiang Liu, Haotian Wang, Wenjing Yang, Yuanxing Zhang, Pengfei Wan, Yi-Fan Zhang, Ziwei Liu*
99 Beyond Objects: Contextual Synthetic Data Generation for Fine-Grained Classification, *William Yang, Xindi Wu, Zhiwei Deng, Esin Tureci, Olga Russakovsky*
100 NEAF: Natural Image Editing with Attention Fusion for Generalizable Test-time Optimization in Text-Guided Image Editing, *Jisoo Kim, Heeseok Oh*
101 OntoAug: Rethinking Generative Data Augmentation via Ontology Guidance, *Shuo Wang, Zhichuan Wang, Jun Luo*
102 Spherical Voronoi: Directional Appearance as a Differentiable Partition of the Sphere, *Francesco Di Sario, Daniel Rebain, Dor Verbin, Marco Grangetto, Andrea Tagliasacchi*
103 4DSurf: High-Fidelity Dynamic Scene Surface Reconstruction, *Renjie Wu, Hongdong Li, Jose M. Alvarez, Miaomiao Liu*

104 Learning 3D Representations for Spatial Intelligence from Unposed Multi-View Images, *Bo Zhou, Qiuxia Lai, Zeren Sun, Xiangbo Shu, Yazhou Yao, Wenguan Wang*
105 Depth Peeling for High-Fidelity Gaussian-Enhanced Surfel Rendering, *Keyang Ye, Hongzhi Wu, Kun Zhou*
106 Intrinsic Image Fusion for Multi-View 3D Material Reconstruction, *Peter Kocsis, Lukas Höllein, Matthias Nießner*
107 PackUV: Packed Gaussian UV Maps for 4D Volumetric Video, *Aashish Rai, Angela Xing, Anushka Agarwal, Xiaoyan Cong, Zekun Li, Tao Lu, Aayush Prakash, Srinath Sridhar*
108 Opti-NeuS: Neural Reconstruction for Dual-Layered Transparent and Opaque Objects, *Yi Yang, Gaoyang Zhang, Jun Tan, Xinguo Liu*
109 PhysGaia: A Physics-aware Benchmark with Multi-Body Interactions for Dynamic Novel View Synthesis, *Mijeong Kim, Gunhee Kim, Jungyoon Choi, Wonjae Roh, Bohyung Han*
110 MatSpray: Fusing 2D Material World Knowledge on 3D Geometry, *Philipp Langsteiner, Jan-Niklas Dihlmann, Hendrik Lensch*
111 OMoBlur: An Object Motion Blur Dataset and Benchmark for Real-World Local Motion Deblurring, *Dingchuan Yu, Jiatong Li, Jingwen Zhou, Zhengyue Zhuge, Yueting Chen, Qi Li*
112 Hybrid Agents for Image Restoration, *Bingchen Li, Xin Li, Yiting Lu, Zhibo Chen*
113 Zero-Shot Image Denoising via Hybrid Prior-Guided Pseudo Sample Generation, *Xiaole Zhao, Qingsong Pang, Xiaobo Zhang, Xun Xu, Xun Gong, Yan Yang, Tianrui Li*
114 Self-supervised Dynamic Heterogeneous Degradation Modeling for Unified Zero-Shot Image Restoration, *XiaoWan Hu, Jing Yang, HeNan Liu, HuaQiu Li, Mai Xu*
115 Next-Scale Prediction: A Self-Supervised Approach for Real-World Image Denoising, *Yiwen Shan, Haiyu Zhao, Peng Hu, Xi Peng, Yuanbiao Gou*
116 PhaSR: Generalized Image Shadow Removal with Physically Aligned Priors, *Chia-Ming Lee, Yu-Fan Lin, Yu-Jou Hsiao, Jin-Hui Jiang, Yu-Lun Liu, Chih-Chung Hsu*
117 UARE: A Unified Vision-Language Model for Image Quality Assessment, Restoration, and Enhancement, *Weiqi Li, Xuanyu Zhang, Bin Chen, Jingfen Xie, Yan Wang, Kexin Zhang, Junlin Li, Li zhang, Jian Zhang, Shijie Zhao*
118 FastGaMer: Efficient GainMap Learning for Practical Inverse Tone Mapping, *Yuanshen Guan, Ruikang Xu, Chang Chen, Yinuo Liao, Dehua Song, Fenglong Song, Zhiwei Xiong*
119 MDS-VQA: Model-Informed Data Selection for Video Quality ❊ Assessment, *Jian Zou, Xiaoyu Xu, Zhihua Wang, Yilin Wang, Balu Adsumilli, Kede Ma*
120 Seeing through Light and Darkness: Sensor-Physics Grounded ❊ Deblurring HDR NeRF from Single-Exposure Images and Events, *Yunshan Qi, Lin Zhu, Nan Bao, Yifan Zhao, Jia Li*
121 Disentanglement-wise Image Dehazing through Cross-Domain Manifold Consensus, *Tianyi Lyu, Mingye Ju, Kai-Kuang Ma*
122 Unsupervised Multi-Scale Segmentation of 3D Subcellular World with ❊ Stable Diffusion Foundation Model, *Mostofa Rafid Uddin, HM Shadman Tabib, Thanh-Huy Nguyen, Kashish Gandhi, Min Xu*
123 EchoPOSE: 6D Pose Estimation of Sparse Echocardiograms for Left-Ventricular 3D Shape Reconstruction, *Lucas Iijima, Yihao Luo, Dario Sesia, Amit Kaura, Jamil Mayet, Choon Hwai Yap*
124 Spatial-SAM: Spatially Consistent 3D Electron Microscopy ❊ Segmentation with SDF Memory and Semi-Supervised Learning, *Yikai Huang, Renmin Han, Yuxuan Wang, Youcheng Cai, Ligang Liu*
125 LLaDA-MedV: Exploring Large Language Diffusion Models for Biomedical Image Understanding, *Xuanzhao Dong, Wenhui Zhu, Xiwen Chen, Zhipeng Wang, Peijie Qiu, Shao Tang, Xin Li, Yalin Wang*
126 TAlignDiff: Automatic Tooth Alignment assisted by Diffusion-based ❊ Transformation Learning, *Yunbi Liu, Enqi Tang, Shiyu Li, Hui Shuai, Lei Ma, Juncheng Li, Kuai Yu, Shu Lou, Yongchu Pan, Qingshan Liu*
127 Harmonized Feature Conditioning and Frequency-Prompt Personalization for Multi-Rater Medical Segmentation, *Sanaz Karimijafarbigloo, Armin Khosravi, Alireza Kheyrkhah, Reza Azad, Mauricio Reyes, Dorit Merhof*
128 Masked-Diffusion Autoencoders for 3D Medical Vision Representation Learning, *Jiachen Tu, Guanghui Qin, Theodore Zhengde Zhao, Jeya Maria Jose Valanarasu, Sheng Zhang, Tristan Naumann, Fan Lam, Sheng Wang, Hoifung Poon*
129 PGR-Net: Prior-Guided ROI Reasoning Network for Brain Tumor MRI Segmentation, *Jiacheng Lu, Hui Ding, Shiyu Zhang, Guoping Huo*
130 Test-Time Attention Purification for Backdoored Large Vision Language Models, *Zhifang Zhang, Bojun Yang, Shuo He, Weitong Chen, Wei Emma Zhang, Olaf Maennel, Lei Feng, Miao Xu*
131 AGFT: Alignment-Guided Fine-Tuning for Zero-Shot Adversarial Robustness of Vision-Language Models, *Yubo Cui, Xianchao Guan, Zijun Xiong, Zheng Zhang*
132 Towards Robust Multimodal Large Language Models Against Jailbreak Attacks, *Ziyi Yin, Yuanpu Cao, Han Liu, Ting Wang, Jinghui Chen, Fenglong Ma*
133 R^2TUA: Reconstruction-residual Based Targeted and Untargeted Attack Against Text-Image Person Re-Identification, *Yubo Wang, Yan Lu, Bin Liu, Xulin Li, Jixiang Niu*
134 When Robots Obey the Patch: Universal Transferable Patch Attacks ❊ on Vision-Language-Action Models, *Hui Lu, Yi Yu, Yiming Yang, Chenyu Yi, Qixin Zhang, Bingquan Shen, Alex C. Kot, Xudong Jiang*
135 FlowHijack: A Dynamics-Aware Backdoor Attack on Flow-Matching Vision-Language-Action Models, *Xinyuan An, Tao Luo, Gengyun Peng, Yaobing Wang, Kui Ren, Dongxia Wang*
136 Principled Steering via Null-space Projection for Jailbreak Defense in Vision-Language Models, *Xingyu Zhu, Beier Zhu, Shuo Wang, Junfeng Fang, Kesen Zhao, Hanwang Zhang, Xiangnan He*
137 Enhancing Part-Level Point Grounding for Any Open-Source MLLMs, *Jin-Cheng Jhang, Fu-En Wang, Xin Yang, Nan Qiao, Lu Xia, Min Sun, Cheng-Hao Kuo*
138 MeteorPred: A Meteorological Multimodal Large Model and Dataset ❊ for Severe Weather Event Prediction, *Shuo Tang, Jian Xu, Jiadong Zhang, Yi Chen, Qizhao Jin, Lingdong Shen, Chenglin Liu, Shiming Xiang*
139 YieldSAT: A Multimodal Benchmark Dataset for High-Resolution Crop Yield Prediction, *Miro Miranda, Deepak Pathak, Patrick Helber, Benjamin Bischke, Hiba Najjar, Francisco Mena, Cristhian Sanchez, Akshay Pai, Diego Arenas, Matias Valdenegro-Toro, Marcela Charfuelan, Marlon Nuske, Andreas Dengel*
140 How Far Can We Go With Synthetic Data for Audio-Visual Sound ❊ Source Localization?, *Arda Senocak, Sooyoung Park, Tae-Hyun Oh, Joon Son Chung*
141 Modeling Cross-vision Synergy for Unified Large Vision Model, *Shengqiong Wu, Lanhu Wu, Mingyang Bao, Wenhao Xu, Hanwang Zhang, Shuicheng Yan, Hao Fei, Tat-Seng Chua*
142 Beyond Missing Modalities: Hypergraph Conditioned Diffusion for Uncertainty-Aware Multimodal Emotion Recognition, *Xihang Qiu, Yuhao Fang, Qing Zhou, Bin Zhai, Jialong Hong, Wanpeng Zhang, Yao Lu, Ye Zhang, Chun Li*
143 Rosetta Stone For Unified MLLMs: A Unified Tokenizer to Decipher Understanding and Generation, *Wenyu Sun, Hufei Li, Ruijin Jin, Xiangheng Kong, Yuning Jiang*
144 MOON2.0: Dynamic Modality-balanced Multimodal Representation Learning for E-commerce Product Understanding, *Zhanheng Nie, Chenghan Fu, Daoze Zhang, Junxian Wu, Wanxian Guan, Pengjie Wang, Jian Xu, Bo Zheng*
145 Nano-EmoX: Unifying Multimodal Emotional Intelligence from Perception to Empathy, *Jiahao Huang, Fengyan Lin, Xuechao Yang, Chen Feng, Kexin Zhu, Xu Yang, Zhide Chen*
146 AMusE: Audio-Visual Benchmark and Alignment Framework for Agentic Multi-Speaker Understanding, *Sanjoy Chowdhury, Karren D Yang, Xudong Liu, Fartash Faghri, Pavan Kumar Anasosalu Vasu, Oncel Tuzel, Dinesh Manocha, Chun-Liang Li, Raviteja Vemulapalli*
147 Prototype-as-Prompt: Multimodal Sentiment Prototypes Endowing ❊ Large Language Models the Capability to Perform Multimodal Sentiment Analysis, *Xianbing Zhao, Lan Luo, Hengyang Lu, Buzhou Tang*
148 CF-IPT: Cross-Modal Fusion Interactive Prompt Tuning of Vision-Language Pre-Trained Model for Multisource Remote Sensing Data Classification, *Jinheng Ji, Jiahui Qu, Wenqian Dong, Yunsong Li*
149 EMAD: Evidence-Centric Grounded Multimodal Diagnosis for Alzheimer's Disease, *Qiuhui Chen, Xuancheng Yao, Zhenglei Zhou, Xinyue Hu, Yi Hong*

150 Multimodal Learning on Low-Quality Data with Conformal Predictive Self-Calibration, *Xun Jiang, Yufan Gu, Disen Hu, Yuqing Hou, Yazhou Yao, Fumin Shen, Heng Tao Shen, Xing Xu*
151 Cross-View Distillation and Adaptive Masking for Incomplete Multi-View Multi-Label Classification, *Yadong Liu, Qiaoqi Li, Yueying Wang, Lunke Fei, Jie Wen*
152 Bootstrap Your Own AV-Proxies: Adaptive Contrastive and Prototype Learning for Audio-Visual Segmentation, *Junbo Zhang, Hang Su, Zhaofan Li, Hang Dong, Chao Sun*
153 Multimodal Distribution Matching for Vision-Language Dataset Distillation, *Jongoh Jeong, Hoyong Kwon, Minseok Kim, Kuk-Jin Yoon*
154 M4-RAG: A Massive-Scale Multilingual Multi-Cultural Multimodal RAG, *David Anugraha, Patrick Amadeus Irawan, Anshul Singh, En-Shiun Annie Lee, Genta Indra Winata*
155 Text-Driven 3D Hand Motion Generation from Sign Language Data, *Léore Bensabath, Mathis Petrovich, Gül Varol*
156 Real2Edit2Real: Generating Robotic Demonstrations via a 3D Control Interface, *Yujie Zhao, Hongwei Fan, Di Chen, Shengcong Chen, Liliang Chen, Xiaoqi Li, Guanghui Ren, Hao Dong*
157 GenHOI: Towards Object-Consistent Hand–Object Interaction with Temporally Balanced and Spatially Selective Object Injection, *Xuan Huang, Mochu Xiang, Zhelun Shen, Jinbo Wu, Chenming Wu, Chen Zhao, Kaisiyuan Wang, Hang Zhou, Shanshan Liu, Haocheng Feng, Wei He, Jingdong Wang*
158 Clay-to-Stone: Phase-wise 3D Gaussian Splatting for Monocular Articulated Hand-Object Manipulation Modeling, *Xingyu Liu, Pengfei Ren, Qi Qi, Haifeng Sun, Zirui Zhuang, Jianxin Liao, Jingyu Wang*
159 Training-free Motion Factorization for Compositional Video Generation, *Zixuan Wang, Ziqin Zhou, Feng Chen, Duo Peng, Yixin Hu, Changsheng Li, Yinjie Lei*
160 Audio-sync Video Instance Editing with Granularity-Aware Mask Refiner, *Haojie Zheng, Shuchen Weng, Jingqi Liu, Siqi Yang, Boxin Shi, Xinlong Wang*
161 CaTok: Taming Mean Flows for One-Dimensional Causal Image Tokenization, *Yitong Chen, Zuxuan Wu, Xipeng Qiu, Yu-Gang Jiang*
162 FFP-300K: Scaling First-Frame Propagation for Generalizable Video Editing, *Xijie Huang, Chengming Xu, Donghao Luo, Xiaobin Hu, Peng Tang, Xu Peng, Jiangning Zhang, Chengjie Wang, Yanwei Fu*
163 V-RGBX: Video Editing with Accurate Controls over Intrinsic Properties, *Ye Fang, Tong Wu, Valentin Deschaintre, Duygu Ceylan, Iliyan Georgiev, Chun-Hao Paul Huang, Yiwei Hu, Xuelin Chen, Tuanfeng Yang Wang*
164 PoseAnything: General Pose-guided Video Generation with Part-aware Temporal Coherence, *Ruiyan Wang, Teng Hu, Kaihui Huang, Zihan Su, Ran Yi, Lizhuang Ma*
165 FastHybrid: Accelerating Hybrid Autoregressive Image Generation with Lookahead and Guided Decoding, *Zhengguo Jiang, Fang Zhang, Yongxiang Hua, Bocheng Li, Wentao Zhang, Linli Xu*
166 DPAR: Dynamic Patchification for Efficient Autoregressive Visual Generation, *Divyansh Srivastava, Akshay Mehra, Pranav Maneriker, Debopam Sanyal, Vishnu Raj, Vijay Kamarshi, Fan Du, Joshua Kimball*
167 AlcheMinT: Fine-grained Temporal Control for Multi-Reference Consistent Video Generation, *Sharath Girish, Viacheslav Ivanov, Tsai-Shien Chen, Hao Chen, Aliaksandr Siarohin, Sergey Tulyakov*
168 LeapAlign: Post-training Flow Matching Models at Any Generation Step by Building Two-Step Trajectories, *Zhanhao Liang, Tao Yang, Jie Wu, Chengjian Feng, Liang Zheng*
169 EVATok: Adaptive Length Video Tokenization for Efficient Visual Autoregressive Generation, *Tianwei Xiong, Jun Hao Liew, Zilong Huang, Zhijie Lin, Jiashi Feng, Xihui Liu*
170 Flow Matching for Multimodal Distributions, *Gaoxiang Luo, Frank Cole, Sihang Zhang, Yuxiang Wan, Yulong Lu, Ju Sun*
171 From Scale to Speed: Adaptive Test-Time Scaling for Image Editing, *Xiangyan Qu, Zhenlong Yuan, Jing Tang, Rui Chen, Datao Tang, Meng Yu, Lei Sun, Yancheng Bai, Xiangxiang Chu, Gaopeng Gou, Gang Xiong, Yujun Cai*
172 ReasonEdit: Towards Reasoning-Enhanced Image Editing Models, *Fukun Yin, Shiyu Liu, Yucheng Han, Zhibo Wang, Peng Xing, Rui Wang, Wei Cheng, Yingming Wang, Aojie Li, Zixin Yin, Pengtao Chen, Xianfang Zeng, Gang Yu, Daxin Jiang*
173 Cross-Subject EEG-to-Video Reconstruction and Beyond, *Runduo Han, Hongchen Tan*
174 Rethinking Position Embedding as a Context Controller for Multi-Reference and Multi-Shot Video Generation, *Binyuan Huang, Yuning Lu, Weinan Jia, Hualiang Wang, Mu Liu, Daiqing Yang*
175 Stand-In: A Lightweight and Plug-and-Play Identity Control for Video Generation, *Bowen Xue, Zheng-Peng Duan, Qixin Yan, Wenjing Wang, Hao Liu, Chun-Le Guo, Chongyi Li, Chen Li, Jing Lyu*
176 BiFM: Bidirectional Flow Matching for Few-Step Image Editing and Generation, *Yasong Dai, Zeeshan Hayder, David Ahmedt-Aristizabal, Hongdong Li*
177 DTG-Restore: Training-Free Diffusion Refinement for Generative Video Super-Resolution, *Hidir Yesiltepe, Koutilya PNVR, Gaurav Pathak, Navaneeth Bodla, Bharat Singh, Pinar Yanardag, Jinrong Xie*
178 VABench: A Comprehensive Benchmark for Audio-Video Generation, *Daili Hua, Xizhi Wang, Bohan Zeng, Xinyi Huang, Hao Liang, Junbo Niu, Xinlong Chen, Quanqing Xu, Wentao Zhang*
179 Relightful Video Portrait Harmonization, *Jun Myeong Choi, Jae Shin Yoon, Luchao Qi, Roni Sengupta, Joon-Young Lee*
180 DiT360: High-Fidelity Panoramic Image Generation via Hybrid Training, *Haoran Feng, Dizhe Zhang, Xiangtai Li, Bo Du, Lu Qi*
181 DVAR: Dynamic Visual Autoregressive Modeling for Image Super-Resolution, *Yu Zheng, Kai Zhang, Wei Zhu, Qingguo Liu, Xiantao Hu, Jun Li, Jian Yang*
182 Gated Condition Injection without Multimodal Attention: Towards Controllable Linear-Attention Transformers, *Yuhe Liu, Zhenxiong Tan, Yujia Hu, Songhua Liu, Xinchao Wang*
183 LinVideo: A Post-Training Framework towards O(n) Attention in Efficient Video Generation, *Yushi Huang, Xingtong Ge, Ruihao Gong, Chengtao Lv, Jun Zhang*
184 UCAN: Unified Convolutional Attention Network for Expansive Receptive Fields in Lightweight Super-Resolution, *Cao Thien Tan, Phan Thi Thu Trang, Do Nghiem Duc, Ho Ngoc Anh, Hanyang Zhuang, Nguyen Duc Dung*
185 EMR-Diff: Edge-aware Multimodal Residual Diffusion Model for Hyperspectral Image Super-resolution, *Tao Zhang, Shengtao Yao, Rong Zeng, Zunjie Zhu, Bolun Zheng, Yaoqi Sun, Ying Fu, Chenggang Yan*
186 RAW-Domain Degradation Models for Realistic Smartphone Super-Resolution, *Ali Mosleh, Faraz Ali, Fengjia Zhang, Stavros Tsogkas, Junyong Lee, Michael S. Brown, Alex Levinshtein*
187 One-Step Diffusion Transformer for Controllable Real-World Image Super-Resolution, *Yushun Fang, Yuxiang Chen, Shibo Yin, Qiang Hu, Jiangchao Yao, Ya Zhang, Xiaoyun Zhang, Yanfeng Wang*
188 FRAMER: Frequency-Aligned Self-Distillation with Adaptive Modulation Leveraging Diffusion Priors for Real-World Image Super-Resolution, *Seungho Choi, Jeahun Sung, Jihyong Oh*
189 HDW-SR: High-Frequency Guided Diffusion Model based on Wavelet Decomposition for Image Super-Resolution, *Chao Yang, Boqian Zhang, Jinghao Xu, Guang Jiang*
190 Unifying Precise Keyframes and Semantic Control via Multi-level Diffusion, *Linjun Wu, Jiejia Yu, Leyang Jin, He Wang, Bowen Zheng, Xu Yang, Hao Jiang, Fei Xia, Fei Ling, Jun Deng, Xiaogang Jin*
191 CIGPose: Causal Intervention Graph Neural Network for Whole-Body Pose Estimation, *Bohao Li, Zhicheng Cao, Huixian Li, Yangming Guo*
192 Pressure2Motion: Hierarchical Human Motion Reconstruction from Ground Pressure with Text Guidance, *Zhengxuan Li, Qinhui Yang, Yiyu Zhuang, Chuan Guo, Xinxin Zuo, Xiaoxiao Long, Yao Yao, Xun Cao, Qiu Shen, Hao Zhu*
193 From 3D Pose to Prose: Biomechanics-Grounded Vision–Language Coaching, *Yuyang Ji, Yixuan Shen, Shengjie Zhu, Yu Kong, Feng Liu*
194 InterPrior: Scaling Generative Control for Physics-Based Human-Object Interactions, *Sirui Xu, Samuel Schulter, Morteza Ziyadi, Xialin He, Xiaohan Fei, Yu-Xiong Wang, Liang-Yan Gui*
195 MoCoDiff: A Controllable Autoregressive Diffusion Model for Expressive Motion Generation, *Wenfeng Song, Xuehan Wang, Shuai Li, Yi Chen, Yuting Guo, Zhenyu Wu, Xingliang Jin, Chenglizhao Chen, Fei Hou, Hongyu Wu, Aimin Hao*

196 W2W: Language-Model-Based Trajectory Prediction with Reinforcement Learning, *Zirui Xu, Biao Yang, Rongrong Ni, Zhongkai Zhou, Shaobo Shen*
197 ParTY: Part-Guidance for Expressive Text-to-Motion Synthesis, *KunHo Heo, SuYeon Kim, Yonghyun Gwon, Youngbin Kim, MyeongAh Cho*
198 Interact2Ar: Full-Body Human-Human Interaction Generation via Autoregressive Diffusion Models, *Pablo Ruiz-Ponce, Sergio Escalera, José García-Rodríguez, Jiankang Deng, Rolandos Alexandros Potamias*
199 Unified Number-Free Text-to-Motion Generation Via Flow Matching, ❊ Guanhe Huang, Oya Celiktutan
200 Generative Diffusion Priors for 3D Mapping of the Dark Universe, ❊ *Brandon Zhao, Diana Scognamiglio, Olivier Doré, Katherine L. Bouman*
201 FlowPalm: Optical Flow Driven Non-Rigid Deformation for Geometrically ❊ Diverse Palmprint Generation, *Yuchen Zou, Huikai Shao, Lihuang Fang, Zhipeng Xiong, Dexing Zhong*
202 DiffuView: Multi-View Diffusion Pretraining for 3D Aware Robotic Manipulation, *Kaizhao Zhang, Tian Niu, Tianyu Liu, Chenen Guo, Zijun Xu, Qingda Hu, Wenchao Ding*
203 Circuit Mechanisms for Spatial Relation Generation in Diffusion ❊ Transformers, *Binxu Wang, Jingxuan Fan, Xu Pan*
204 Dual Ascent Diffusion for Inverse Problems, *Minseo Kim, Axel Levy, Gordon Wetzstein*
205 Forecast the Principal, Stabilize the Residual: Subspace-Aware Feature Caching for Diffusion Transformers, *Guantao Chen, Shikang Zheng, Yuqi Lin, Linfeng Zhang*
206 Spatial-Spectral Residuals Informed Diffusion Neural Operator for Pan-sharpening, *Jiahan Huang, Ran Ran, Junming Hou, Zihao Chen, Xiaofeng Cong, Junling Li, Liang-Jian Deng*
207 PhyOceanCast: Global Ocean Forecasting with Physics-Informed Diffusion, *Qixiu Li, Xiang Zhu, Xiaoyong Li, Xiaolong Xu*
208 Pixel Motion Diffusion is What We Need for Robot Control, *E-Ro Nguyen, Yichi Zhang, Kanchana Ranasinghe, Xiang Li, Michael S. Ryoo*
209 ORIC: Benchmarking Object Recognition under Contextual Incongruity in Large Vision-Language Models, *Zhaoyang Li, Zhan Ling, Yuchen Zhou, Litian Gong, Erdem Biyik, Hao Su*
210 M3Grounder: Mask-Based Multi-Span and Multi-Granular Grounding for Document QA, *Venkata Kesav Venna, Sai Madhusudan Gunda, Jyothi Swaroopa Jinka, Hrithik Sagar Rachakonda, Anirudh Srinivasan, Ravi Kiran Sarvadevabhatla*
211 BabyVLM-V2: Toward Developmentally Grounded Pretraining and Benchmarking of Vision Foundation Models, *Shengao Wang, Wenqi Wang, Zecheng Wang, Max Whitton, Michael Wakeham, Arjun Chandra, Joey Huang, Pengyue Zhu, Helen Chen, David Li, Jeffrey Li, Shawn Li, Andrew Zagula, Amy Zhao, Andrew Zhu, Sayaka Nakamura, Yuki Yamamoto, Jerry Jun Yokono, Aaron Mueller, Bryan A. Plummer, Kate Saenko, Venkatesh Saligrama, Boqing Gong*
212 Towards Real-World Document Parsing via Realistic Scene Synthesis and Document-Aware Training, *Gengluo Li, Pengyuan Lyu, Chengquan Zhang, Huawen Shen, Liang Wu, Xingyu Wan, Gangyan Zeng, Han Hu, Can Ma, Yu Zhou*
213 RoadSceneBench: A Lightweight Benchmark for Mid-Level Road Scene Understanding, *Xiyan Liu, Han Wang, Yuhu Wang, Junjie Cai, Zhe Cao, Jianzhong Yang, Zhen Lu*
214 UNICBench: UNIfied Counting Benchmark for MLLM, *Chenggang Rong, Tao Han, Zhiyuan Zhao, Yaowu Fan, Jia Wan, Song Guo, Yuan Yuan, Junyu Gao*
215 CaptionQA: Is Your Caption as Useful as the Image Itself?, *Shijia Yang, Yunong Liu, Bohan Zhai, Ximeng Sun, Zicheng Liu, Emad Barsoum, Manling Li, Chenfeng Xu*
216 EgoProx: Evaluating MLLMs on Egocentric 3D Proximity Reasoning Across a Cognitive Hierarchy, *Jinzhao Li, Yinuo Chen, Dongxu Piao, Panwang Pan, Yifan Yu, Dong Wang, Honglei Yan, Liang Yue, Shaofei Wang, Yixin Chen, Siyuan Huang, Miao Liu*
217 VULCAN: Tool-Augmented Multi Agents for Iterative 3D Object Arrangement, *Zhengfei Kuang, Rui Lin, Long Zhao, Gordon Wetzstein, Saining Xie, Sanghyun Woo*
218 EmbodiedSplat: Online Feed-Forward Semantic 3DGS for Open-Vocabulary 3D Scene Understanding, *Seungjun Lee, Zihan Wang, Yunsong Wang, Gim Hee Lee*
219 Efficient Encoder-Free Fourier-based 3D Large Multimodal Model, *Guofeng Mei, Wei Lin, Luigi Riz, Yujiao Wu, Yiming Wang, Fabio Poiesi*
220 Socratic-Geo: Synthetic Data Generation and Cross-Modal Geometric Reasoning via Multi-Agent Interaction, *Zhengbo Jiao, Zifan Zhang, Shaobo Wang, Wei Wang, Bing Zhao, Hu Wei, Linfeng Zhang*
221 HAMMER: Harnessing MLLMs via Cross-Modal Integration for Intention-Driven 3D Affordance Grounding, *Lei Yao, Yong Chen, Yuejiao Su, Yi Wang, Moyun Liu, Lap-Pui Chau*
222 Proxy3D: Efficient 3D Representations for Vision-Language Models via Semantic Clustering and Alignment, *Jerry Jiang, Haowen Sun, Denis Gudovskiy, Yohei Nakata, Tomoyuki Okuno, Kurt Keutzer, Wenzhao Zheng*
223 ReLaGS: Relational Language Gaussian Splatting, *Yaxu Xie, Abdalla Arafa, Alireza Javanmardi, Christen Millerdurai, Jia Cheng Hu, Shaoxiang Wang, Alain Pagani, Didier Stricker*
224 3D-IDE: 3D Implicit Depth Emergent, *Chushan Zhang, Ruihan Lu, Jinguang Tong, Yikai Wang, Hongdong Li*
225 FunFact: Building Probabilistic Functional 3D Scene Graphs via Factor-Graph Reasoning, *Zhengyu Fu, René Zurbrügg, Kaixian Qu, Marc Pollefeys, Marco Hutter, Hermann Blum, Zuria Bauer*
226 Parse, Search, and Confirmation: Training-Free Aerial Vision-and-Dialog Navigation with Chain-of-Thought Reasoning and Structured Spatial Memory, *Yu Qi, Hongyu Li, Shaofei Huang, Tianrui Hui, Yaxiong Wang, Lechao Cheng, Zhun Zhong, Si Liu, Meng Wang*
227 4DP-QA: Scalable QA for 4D Perception in Vision Language Models, *Seokju Cho, Abhishek Badki, Hang Su, Jindong Jiang, Ziyao Zeng, Seungryong Kim, Sifei Liu, Orazio Gallo*
228 LASAR: Towards Spatio-temporal Reasoning with Latent Cognitive Map, *Jinzhou Tang, Sidi Liu, Waikit Xiu, Weixing Chen, Keze Wang*
229 Text-Phase Synergy Network with Dual Priors for Unsupervised Cross-Domain Image Retrieval, *Jing Yang, Hui Xue, Shipeng Zhu, Pengfei Fang*
230 EagleNet: Energy-Aware Fine-Grained Relationship Learning Network for Text-Video Retrieval, *Yuhan Chen, Pengwen Dai, Chuan Wang, Dayan Wu, Xiaochun Cao*
231 PIX-TAB: Efficient PIXel-Precise TABle Structure Recognition Approach with Speculative Decoding and Region-Based Image Segmentation, *Viktor Zaytsev, Olena Vynokurova, Pavlo Tytarchuk, Dmytro Kozii, Vitalii Pohribnyi, Olga Radyvonenko, Artem Shcherbina*
232 CARLoS: Retrieval via Concise Assessment Representation of LoRAs at Scale, *Shahar Sarfaty, Adi Haviv, Uri Hacohen, Niva Elkin-Koren, Roi Livni, Amit H. Bermano*
233 Camouflage-aware Image-Text Retrieval via Expert Collaboration, *Yao Jiang, Zhongkuan Mao, Xuan Wu, Keren Fu, Qijun Zhao*
234 TriSim: Tri-Dimensional Similarity Modeling with Extreme Value Theory for False-Negative Mitigation in Remote Sensing Image-Text Retrieval, *Chengyu Zheng, Hanzhang Lu, Jie Nie, Shan Du*
235 TIGER: A Unified Framework for Time, Images and Geo-location Retrieval, *David G. Shatwell, Sirnam Swetha, Mubarak Shah*
236 Mistake Attribution: Fine-Grained Mistake Understanding in Egocentric Videos, *Yayuan Li, Aadit Jain, Filippos Bellos, Jason J. Corso*
237 VidTAG: Temporally Aligned Video to GPS Geolocalization with Denoising Sequence Prediction at a Global Scale, *Parth Parag Kulkarni, Rohit Gupta, Prakash Chandra Chhipa, Mubarak Shah*
238 Stitch-a-Demo: Creating Video Demonstrations from Multistep Descriptions, *Chi Hsuan Wu, Kumar Ashutosh, Kristen Grauman*
239 Prototypical Action Reasoning Facilitated by Vision-Language Alignment for Egocentric Action Anticipation, *Jiang Shao, Xinbo Zhao, Wenyin Tuo, Xiaochun Zou*
240 AdaSpot: Spend Resolution Where It Matters for Precise Event Spotting, *Artur Xarles, Sergio Escalera, Thomas B. Moeslund, Albert Clapés*
241 Unique Lives, Shared World: Learning from Single-Life Videos, *Tengda Han, Sayna Ebrahimi, Dilara Gokay, Li Yang Ku, Maks Ovsjanikov, Iva Babukova, Daniel Zoran, Viorica Patraucean, Joao Carreira, Andrew Zisserman, Dima Damen*
242 Symphony: A Cognitively-Inspired Multi-Agent System for Long-Video Understanding, *Haiyang Yan, Hongyun Zhou, Peng Xu, Xiaoxue Feng, Mengyi Liu*

243 VideoARM: Agentic Reasoning over Hierarchical Memory for Long-Form Video Understanding, *Yufei Yin, Qianke Meng, Minghao Chen, Jiajun Ding, Zhenwei Shao, Zhou Yu*
244 Wavelet-based Frame Selection by Detecting Semantic Boundary for Long Video Understanding, *Wang Chen, Yuhui Zeng, Yongdong Luo, Tianyu Xie, Luojun Lin, Jiayi Ji, Yan Zhang, Xiawu Zheng*
245 SVAgent: Storyline-guided Long Video Understanding via Cross-Modal Multi-Agent Collaboration, *Zhongyu Yang, Zuhao Yang, Shuo Zhan, Tan Yue, Wei Pang, Yingfang Yuan*
246 Frame2Freq: Spectral Adapters for Fine-Grained Video Understanding, *Thinesh Thiyakesan Ponbagavathi, Constantin Seibold, Alina Roitberg*
247 Structural Graph Probing of Vision–Language Models, *Haoyu He, Yue Zhuo, Yu Zheng, Qi R. Wang*
248 Saliency-R1: Enforcing Interpretable and Faithful Vision-language Reasoning via Saliency-map Alignment Reward, *Shizhan Gong, Minda Hu, Qiyuan Zhang, Chen Ma, Qi Dou*
249 Hidden Monotonicity: Explaining Deep Neural Networks via their DC
❊ Decomposition, *Jakob Paul Zimmermann, Georg Loho*
250 MaskDiME: Adaptive Masked Diffusion for Precise and Efficient Visual Counterfactual Explanations, *Changlu Guo, Anders Nymark Christensen, Anders Bjorholm Dahl, Morten Rieger Hannemose*
251 TRANSPORTER: Transferring Visual Semantics from VLM Manifolds, *Alexandros Stergiou*
252 Relational Visual Similarity, *Thao Nguyen, Sicheng Mo, Krishna Kumar Singh, Yilin Wang, Jing Shi, Nicholas Kolkin, Eli Shechtman, Yong Jae Lee, Yuheng Li*
253 PointCNN++: Performant Convolution on Native Points, *Lihan Li, Haofeng Zhong, Rui Bu, Mingchao Sun, Wenzheng Chen, Baoquan Chen, Yangyan Li*
254 Fast Markov Random Field Optimisation for Topologically Noisy 3D Shape Matching, *Paul Roetzer, Johan Thunberg, Zorah Lähner, Florian Bernard*
255 LitePT: Lighter Yet Stronger Point Transformer, *Yuanwen Yue,*
❊ *Damien Robert, Jianyuan Wang, Sunghwan Hong, Jan Dirk Wegner, Christian Rupprecht, Konrad Schindler*
256 SuP: Sub-cloud Driven Point Cloud Registration, *Sheldon Fung,*
❊ *Wei Pan, Ling Cao, Fei Hou, Ling Chen, Shasha Mao, Hongdong Li, Xuequan Lu*
257 PQDT: Pseudo-Query Dual Transformer for Robust Point Cloud Restoration, *Haoqing Wu, Alexa Nawotki, Jochen Garcke*
258 Test-Time Training for LiDAR Semantic Segmentation under Corruption via Geometric Inlier Discrimination, *Hyeonseong Kim, Hyun-Kurl Jang, Kuk-Jin Yoon*
259 MHopReg: Efficient Hierarchical Multi-Hop Graph Search for Point Cloud Registration, *Yue Wu, Feng Xiao, Yongzhe Yuan, Hao Li, Kaiyuan Feng, Maoguo Gong, Qiguang Miao, Wenping Ma*
260 GEM: Generating LiDAR World Model via Deformable Mamba, *Yang Wu, Zhaojiang Liu, Qiang Meng, Youquan Liu, Renliang Weng, Jianjun Qian, Jian Yang, Jin Xie*
261 Hybrid Robust Collaborative Perception with LiDAR-4D Radar Fusion under Adverse Weather Conditions, *Yuquan Yang, Hui Zhang, Wenyu Lu, Ziyin Zhang, Chuanming Zhang, Xiaohua Xu*
262 Task-Driven Implicit Representations for Automated Design of
❊ LiDAR Systems, *Nikhil Behari, Aaron Young, Tzofi Klinghoffer, Akshat Dave, Ramesh Raskar*
263 Hierarchical Point-Patch Fusion with Adaptive Patch Codebook for 3D Shape Anomaly Detection, *Xueyang Kang, Zizhao Li, Tian Lan, Dong Gong, Kourosh Khoshelham, Liangliang Nan*
264 When Numbers Speak: Aligning Textual Numerals and Visual Instances in Text-to-Video Diffusion Models, *Zhengyang Sun, Yu Chen, Xin Zhou, Xiaofan Li, Xiwu Chen, Dingkang Liang, Xiang Bai*
265 Beyond Layer-Wise Merging: Chain-of-Merging for Vision-Language Models, *Xinyu Zhang, Yuxuan Dong, Lingling Zhang, Chengyou Jia, ZhuoHang Dang, Yixing Yao, Yaqiang Wu, Basura Fernando, Jun Liu*
266 GazeShift: Unsupervised Gaze Estimation and Dataset for VR, *Gil Shapira, Ishay Goldin, Evgeny Artyomov, Donghoon Kim, Yosi Keller, Niv Zehngut*
267 Improving Calibration in Test-Time Prompt Tuning for Vision-Language Models via Data-Free Flatness-Aware Prompt Pretraining, *Hyeonseo Jang, Jaebyeong Jeon, Joong-Won Hwang, Kibok Lee*
268 Reevaluating the Intra-Modal Misalignment Hypothesis in CLIP, *Jonas Herzog, Yue Wang*
269 Dr. Seg: Revisiting GRPO Training for Visual Large Language Models through Perception-Oriented Design, *Haoxiang Sun, Tao Wang, Chenwei Tang, Li Yuan, Jiancheng Lv*
270 Soft Modality-Guided Expert Specialization in MoE-VLMs, *Zi-Hao Bo, Yaqian Li, Anzhou Hou, Rinyoichi Takezoe, Ertao Zhao, Tianxiang Pan, Jiale Yan, Mo Guang, Kaiwen Long*
271 CoVFT: Context-aware Visual Fine-tuning for Multimodal Large Language Models, *Nan Zhou, Huiqun Wang, Yaoyan Zheng, Di Huang*
272 Retrieving Counterfactuals Improves Visual In-Context Learning, *Guangzhi Xiong, Sanchit Sinha, Zhenghao He, Aidong Zhang*
273 AutoRegressive Generation with B-rep Holistic Token Sequence Representation, *Jiahao Li, Yunpeng Bai, Yongkang Dai, Hao Guo, Hongping Gan, Yilei Shi*
274 VecGlypher: Unified Vector Glyph Generation with Language Models, *Xiaoke Huang, Bhavul Gauri, Kam Woh Ng, Tony Ng, Mengmeng Xu, Zhiheng Liu, Weiming Ren, Zhaochong An, Zijian Zhou, Haonan Qiu, Yuyin Zhou, Sen He, Ziheng Wang, Tao Xiang, Xiao Han*
275 NERFIFY: A Multi-Agent Framework for Turning NeRF Papers
❊ into Code, *Seemandhar Jain, Keshav Gupta, Kunal Gupta, Manmohan Chandraker*
276 Diagram2Structure: Unlocking LLMs' Diagram Comprehension through DiagramDiff, an Offline Diagram Structuring Framework, *Haoxiang Hu, Yaokun Li, Zeyuan Huang, Cangjun Gao, Qiang He, Qingkun Li, Xiaoming Deng, Cuixia Ma, Yu-Kun Lai, Yong-Jin Liu, Hongan Wang*
277 ShowTable: Unlocking Creative Table Visualization with Collaborative Reflection and Refinement, *Zhihang Liu, Xiaoyi Bao, Pandeng Li, Junjie Zhou, Zhaohe Liao, Yefei He, Kaixun Jiang, Chen-Wei Xie, Yun Zheng, Hongtao Xie*
278 GardenDesigner: Encoding Aesthetic Principles into Jiangnan Garden
❊ Construction via a Chain of Agents, *Mengtian Li, Fan Yang, Ruixue Xiong, Yiyan Fan, Zhifeng Xie, Zeyu Wang*
279 ShadowDraw: From Any Object to Shadow-Drawing Compositional Art, *Rundong Luo, Noah Snavely, Wei-Chiu Ma*
280 End-to-End Hyper-Relational Information Extraction for Engineering Diagrams via Dynamically Tokenized Relation Transformer, *Tianyou Bai, Yan-Ming Zhang, Zixiang Zhang, Jibin Zhou, Fei Yin, Cheng-Lin Liu*
281 When Anonymity Breaks: Identifying Models Behind Text-to-Image Leaderboards, *Ali Naseh, Anshuman Suri, Yuefeng Peng, Harsh Chaudhari, Alina Oprea, Amir Houmansadr*
282 Bias at the End of the Score, *Salma Abdel Magid, Grace Guo, Esin Tureci, Amaya Dharmasiri, Vikram V. Ramaswamy, Hanspeter Pfister, Olga Russakovsky*
283 PECCVAI: Overcoming the Brittleness of AI Image Watermarking Under Visual Paraphrasing Attacks, *Shreyas Dixit, Ashhar Aziz, Shashwat Bajpai, Vasu Sharma, Aman Chadha, Vinija Jain, Amitava Das*
284 Dynamic Token Reweighting for Robust Vision-Language Models, *Tanqiu Jiang, Jiacheng Liang, Rongyi Zhu, Jiawei Zhou, Fenglong Ma, Ting Wang*
285 COPYLENS: Towards Copyrighted Characters Infringement Detection via Copyright-Aware Prompt Learning, *Yaoyu Jin, Xiaochun Yang, Hong Liu, Leixia Wang, Jian Li, Rui Ding, Bin Wang*
286 Closed-Form Concept Erasure via Double Projections, *Chi Zhang, Jingpu Cheng, Zhixian Wang, Ping Liu*
287 Adaptive Bayesian Early-Exit Networks for Efficient Non-Transferable Learning, *Siyu Luan, Yan Li, Zhong Chen, Zhenyi Wang*
288 Stake the Points: Structure-Faithful Instance Unlearning, *Kiseong Hong, JungKyoo Shin, Eunwoo Kim*
289 Federated Active Learning Under Extreme Non-IID and Global Class Imbalance, *Chen-Chen Zong, Sheng-Jun Huang*
290 FedRG: Unleashing the Representation Geometry for Federated Learning with Noisy Clients, *Tian Wen, Zhiqin Yang, Yonggang Zhang, Xuefeng Jiang, Hao Peng, Yuwei Wang, Bo Han*
291 FedCART: Tackling Long-Tailed Distributions in Federated Adversarial Training via Classifier Refinement, *Yuchen Qin, Yizhi Zhou, Junxiao Wang, Xin Xie, Heng Qi*

292 Generalized and Personalized Federated Learning with Black-Box Foundation Models via Orthogonal Transformations, *Eun Gyung Kong, Jewon Yeom, Yonghoon Jeon, Taesup Kim*
293 Fully Decentralized Certified Unlearning, *Hithem Lamri, Michail Maniatakos*
294 Fed-ADE: Adaptive Learning Rate for Federated Post-adaptation under Distribution Shift, *Heewon Park, Mugon Joe, Miru Kim, Kyungjin Im, Minhae Kwon*
295 Towards Streaming Referring Video Segmentation via Large Language Model, *Wenkang Zhang, Kaicheng Yang, Xiang An, Qiang Li, Ziyong Feng, Wankou Yang, Jiankang Deng*
296 Multi-speaker Attention Alignment for Multimodal Social Interaction, *Liangyang Ouyang, Yifei Huang, Mingfang Zhang, Caixin Kang, Ryosuke Furuta, Yoichi Sato*
297 OmniVTG: A Large-Scale Dataset and Training Paradigm for Open-World Video Temporal Grounding, *Minghang Zheng, Zihao Yin, Yi Yang, Yuxin Peng, Yang Liu*
298 SARL-STG: A Spatially Aware Reinforcement Learning Framework for Refining MLLMs in Spatio-Temporal Video Grounding, *Hong Gao, Xiangkai Xu, Bin Zhong, Junjie Yin, Fangyu Kang, Yutong Xu, Xiugang Dong, Xurui Gao, Min-Ling Zhang*
299 VideoITG: Multimodal Video Understanding with Instructed Temporal Grounding, *Shihao Wang, Guo Chen, De-An Huang, Zhiqi Li, Minghan Li, Guilin Liu, Jan Kautz, Jose M. Alvarez, Lei Zhang, Zhiding Yu*
300 DeRVOS: Decoupling Consistent Trajectory Generation and Multimodal Understanding for Referring Video Object Segmentation, *Wenxuan Cheng, Ming Dai, Huimin Lu, Wankou Yang*
301 UniCompress: Token Compression for Unified Vision–Language Understanding and Generation, *Ziyao Wang, Chen Chen, Jingtao Li, Weiming Zhuang, Jiabo Huang, Ang Li, Lingjuan Lyu*
302 StreamingTOM: Streaming Token Compression for Efficient Video Understanding, *Xueyi Chen, Keda Tao, Kele Shao, Huan Wang*
303 SCoRe: Salience-Coverage Reduction for Vision Token Pruning in Vision-Language Models, *Tong Xu, Hailong Shi, Xingyu Gao*
304 VLM-PTQ: Efficient Post-Training Quantization for Large Vision-Language Models, *Juncan Deng, Kejie Huang*
305 Aligning What Vision-Language Models See and Perceive with Adaptive Information Flow, *Chengxin Liu, Wonseok Choi, Chenshuang Zhang, Tae-Hyun Oh*
306 Quant Experts: Token-aware Adaptive Error Reconstruction with Mixture of Experts for Large Vision-Language Models Quantization, *Chenwei Jia, Baoting Li, Xuchong Zhang, Mingzhuo Wei, Bochen Lin, Hongbin Sun*
307 Rethinking Token Reduction for Large Vision-Language Models, *Yi Wang, Haofei Zhang, Qihan Huang, Anda Cao, Gongfan Fang, Wei Wang, Xuan Jin, Jie Song, Mingli Song, Xinchao Wang*
308 Prototype-based Causal Intervention for Multi-Label Image Classification, *Yanmin Li, Zhilong Mao, Mao Wang, Lihua Liu, Jibing Wu, Weidong Bao*
309 FAST: Topology-Aware Frequency-Domain Distribution Matching for Coreset Selection, *Jin Cui, Boran Zhao, Jiajun Xu, Jiaqi Guo, Shuo Guan, Pengju Ren*
310 Face-Guided Sentiment Boundary Enhancement for Weakly-Supervised Temporal Sentiment Localization, *Cailing Han, Zhangbin Li, Jinxing Zhou, Wei Qian, Jingjing Hu, Yanghao Zhou, Zhangling Duan, Dan Guo*
311 Evidential Deep Partial Label Learning to Quantify Disambiguation Uncertainty, *Jinfu Fan, Jiangnan Li, Xiaohui Zhong, Kangrui Ren, Zhencun Jiang, Min Gan, Tianhao Gu, Linqing Huang*
312 Unlocking Strong Supervision: A Data-Centric Study of General-Purpose Audio Pre-Training Methods, *Xuanru Zhou, Yiwen Shao, Wei-Cheng Tseng, Dong Yu*
313 Revisiting Learning with Noisy Labels: Active Forgetting and Noise Suppression, *Mengmeng Sheng, Zeren Sun, Tao Chen, Jinshan Pan, Yazhou Yao, Fumin Shen*
314 PAF: Perturbation-Aware Filtering for Open-Set Semi-Supervised Learning, *Yinan Han, Qing-Yuan Jiang*
315 Global-Graph Guided and Local-Graph Weighted Contrastive Learning for Unified Clustering on Incomplete and Noise Multi-View Data, *Hongqing He, Jie Xu, Wenyuan Yang, Yonghua Zhu, Guoqiu Wen, Xiaofeng Zhu*
316 Enhancing Out-of-Distribution Detection with Extended Logit Normalization, *Yifan Ding, Xixi Liu, Jonas Unger, Gabriel Eilertsen*
317 Unleashing VLA Potentials in Autonomous Driving via Explicit Learning from Failures, *Yuechen Luo, Fang Li, Qimao Chen, Shaoqing Xu, Jiaxin Liu, Ziying Song, Zhi-xin Yang, Fuxi Wen*
318 Unposed-to-3D: Learning Simulation-Ready Vehicles from Real-World Images, *Hongyuan Liu, Bochao Zou, Qiankun Liu, Haochen Yu, Qi Mei, Jianfei Jiang, Chen Liu, Cheng Bi, Zhao Wang, Xueyang Zhang, Yifei Zhan, Jiansheng Chen, Huimin Ma*
319 SafeDrive: Fine-Grained Safety Reasoning for End-to-End Driving in a Sparse World, *Jungho Kim, Jiyong Oh, Seunghoon Yu, Hongjae Shin, Donghyuk Kwak, Jun Won Choi*
320 RAG-TP: A General Framework for Vehicle Trajectory Prediction via Retrieval-Augmented Generation, *Ziyi Wang, Yang Zhang, Guijian Tang, Chao Zhang, Shibo Zhang, Xueqiong Li, Shaowu Yang*
321 Perceiving the Near, Reasoning the Distant: Coherent Long-Horizon Trajectory Prediction for Autonomous Driving, *Hua Hu, Zikang Zhou, Qian Zhou, Zihao Wen, Junjie Hu, Xinhong Chen, Zhengmin Jiang, Yung-Hui Li, Jianping Wang*
322 Dual-Agent Reinforcement Learning for Adaptive and Cost-Aware Visual–Inertial Odometry, *Feiyang Pan, Shenghe Zheng, Chunyan Yin, Guangbin Dou*
323 HorizonForge: Driving Scene Editing with Any Trajectories and Any Vehicles, *Yifan Wang, Francesco Pittaluga, Zaid Tasneem, Chenyu You, Manmohan Chandraker, Ziyu Jiang*
324 AMap: Distilling Future Priors for Ahead-Aware Online HD Map Construction, *Ruikai Li, Xinrun Li, Mengwei Xie, Hao Shan, Shoumeng Qiu, Xinyuan Chang, Yizhe Fan, Feng Xiong, Han Jiang, Yilong Ren, Haiyang Yu, Mu Xu, Yang Long, Varun Ojha, Zhiyong Cui*
325 WAM-Flow: Parallel Coarse-to-Fine Motion Planning via Discrete Flow Matching for Autonomous Driving, *Yifang Xu, Jiahao Cui, Zhihao Zhu, Hanlin Shang, Shan Luan, Mingwang Xu, Feipeng Cai, Neng Zhang, Yaoyi Li, Jia Cai, Siyu Zhu*
326 PlannerRFT: Reinforcing Diffusion Planners through Closed-Loop and Sample-Efficient Fine-Tuning, *Hongchen Li, Tianyu Li, Jiazhi Yang, Mingyang Shang, Gaoqiang Wu, Caojun Wang, Haochen Tian, Zengrong Lin, Zhihui Hao, XianPeng Lang, Jia Hu, Hongyang Li*
327 MARIS: Marine Open-Vocabulary Instance Segmentation, *Bingyu Li, Feiyu Wang, Da Zhang, Zhiyuan Zhao, Junyu Gao, Xuelong Li*
328 XSeg: A Large-scale X-ray Contraband Segmentation Benchmark For Real-World Security Screening, *Hongxia Gao, Yixin Chen, Jiali Wen, Litao Li, Qianyun Liu, Kaijie Zhang*
329 Training-Free Open-Vocabulary Camouflaged Object Segmentation via Fine-Grained Object Binding and Adaptive Hybrid Prompt, *Peng Ren, Cheng Jiang, Chuande Yang, Fuming Sun, Tian Bai*
330 M⁴-SAM: Multi-Modal Mixture-of-Experts with Memory-Augmented SAM for RGB-D Video Salient Object Detection, *Jiyuan Liu, Jia Lin, Xiaofei Zhou, Runmin Cong, Deyang Liu, Zhi Liu*
331 ReAttnCLIP: Training-Free Open-Vocabulary Remote Sensing Image Segmentation via Re-defined Attention in CLIP, *Xin Niu, Manqi Zhao, Dongsheng Jiang, Yingying Wu, Bing Su*
332 Mixture of Prototypes for Test-time Adaptive Segmentation, *Guangrui Li, Zhengyu Zhu, Yongxin Ge*
333 Reconstruction-Guided Slot Curriculum: Addressing Object Over-Fragmentation in Video Object-Centric Learning, *WonJun Moon, Hyun Seok Seong, Jae-Pil Heo*
334 ELVIS: Enhance Low-Light for Video Instance Segmentation in the Dark, *Joanne Lin, Ruirui Lin, Yini Li, David Bull, Nantheera Anantrasirichai*
335 Decouple Your Discovery and Memory in Continual Generalized Category Discovery, *Jiawei Yu, Zijian Gao, Xingxing Zhang, Xuan Liu, Huaimin Wang, Kele Xu*
336 Beyond the Static World: Continual Category Discovery under Visual Drift, *Wei Feng, Yiwen Jiang, Sijin Zhou, Zongyuan Ge*
337 Memory-Efficient Transfer Learning with Fading Side Networks via Masked Dual Path Distillation, *Yutong Zhang, Jiaxin Chen, Honglin Chen, Kaiqi Zheng, Shengcai Liao, Hanwen Zhong, Weixin Li, Yunhong Wang*

338 SAME: Sparse and Anchored Model Editing for Heterogeneous
❊ Incremental Learning under Limited Data, *Zixuan Duan, Zeyu Zhang, Fengyuan Lu, Shaofeng Zhang, Wenbin Li, Qi Fan, Yang Gao*
339 CHEEM: Continual Learning by Reuse, New, Adapt and Skip - A Hierarchical Exploration-Exploitation Approach, *Chinmay Savadikar, Michelle Dai, Tianfu Wu*
340 Exemplar-Free Continual Learning for State Space Models,
❊ *Isaac Ning Lee, Leila Mahmoodi, Trung Le, Mehrtash Harandi*
341 A Faster Path to Continual Learning, *Wei Li, Hangjie Yuan, Zixiang Zhao, Borui Kang, Ziwei Liu, Tao Feng*
342 Continual Learning for fMRI-Based Brain Disorder Diagnosis via Functional Connectivity Matrices Generative Replay, *Qianyu Chen, Shujian Yu*
343 BeautyGRPO: Aesthetic Alignment for Face Retouching via Dynamic
❊ Path Guidance and Fine-Grained Preference Modeling, *Jiachen Yang, Xianhui Lin, Yi Dong, Zebiao Zheng, Xing Liu, Hong Gu, Yanmei Fang*
344 SyncDreamer: Controllable and Expressive Avatar Generation Beyond
❊ the Talking Head, *Fatemeh Nazarieh, Zhenhua Feng, Diptesh Kanojia, Josef Kittler, Muhammad Awais*
345 PerformRecast: Expression and Head Pose Disentanglement for Portrait Video Editing, *Jiadong Liang, Bojun Xiong, Jie Tian, Hua Li, Xiao Long, Yong Zheng, Huan Fu*
346 UniLS: End-to-End Audio-Driven Avatars for Unified Listening and Speaking, *Xuangeng Chu, Ruicong Liu, Yifei Huang, Yun Liu, Yichen Peng, Bo Zheng*
347 PC-Talk: Precise Facial Animation Control for Audio-Driven Talking Face Generation, *Baiqin Wang, Xiangyu Zhu, Fan Shen, Hao Xu, Zhen Lei*
348 FlashPortrait: 6x Faster Infinite Portrait Animation with Adaptive Latent Prediction, *Shuyuan Tu, Yueming Pan, Yinming Huang, Xintong Han, Zhen Xing, Qi Dai, Kai Qiu, Chong Luo, Zuxuan Wu*
349 DriveVLN: Towards Mapless Vision-and-Language Navigation in Autonomous Driving, *Dongqian Guo, Haoran Wei, Wencheng Han, Runzhou Tao, Zhongying Qiu, Jianfei Yang, Jianbing Shen*
350 Towards Open Environments and Instructions: General Vision-Language Navigation via Fast-Slow Interactive Reasoning, *Yang Li, Aming Wu, Zihao Zhang, Yahong Han*
351 Unifying Language-Action Understanding and Generation for
❊ Autonomous Driving, *Xinyang Wang, Qian Liu, Wenjie Ding, Zhao Yang, Wei Li, Chang Liu, Bailin Li, Kun Zhan, Xianpeng Lang, Wei Chen*
352 Drive My Way: Preference Alignment of Vision-Language-Action Model for Personalized Driving, *Zehao Wang, Huaide Jiang, Shuaiwu Dong, Yuping Wang, Hang Qiu, Jiachen Li*
353 Prune2Drive: A Plug-and-Play Framework for Accelerating Vision-Language Models in Autonomous Driving, *Minhao Xiong, Zichen Wen, Zhuangcheng Gu, Xuyang Liu, Rui Zhang, Hengrui Kang, Jiabing Yang, Junyuan Zhang, Weijia Li, Conghui He, Linfeng Zhang*
354 CGHair: Compact Gaussian Hair Reconstruction with Card Clustering, *Haimin Luo, Srinjay Sarkar, Albert Mosella-Montoro, Francisco Vicente Carrasco, Fernando De la Torre*
355 HyperGaussians: High-Dimensional Gaussian Splatting for High-Fidelity Animatable Face Avatars, *Gent Serifi, Marcel C. Buehler*
356 Skullptor: High Fidelity 3D Head Reconstruction in Seconds with Multi-View Normal Prediction, *Noé Artru, Rukhshanda Hussain, Emeline Got, Alexandre Messier, David B. Lindell, Abdallah Dib*
357 RelightAnyone: A Generalized Relightable 3D Gaussian Head Model, *Yingyan Xu, Studios 0000-0002-8076-1947, Pramod Rao, Sebastian Weiss, Studios blank, Gaspard Zoss, Studios blank, Markus Gross, Studios; ETH Zurich blank, Christian Theobalt, Marc Habermann, Derek Bradley, Studios blank*
358 Feed-forward Gaussian Registration for Head Avatar Creation and Editing, *Malte Prinzler, Paulo Gotardo, Siyu Tang, Timo Bolkart*
359 Residual Decoding: Mitigating Hallucinations in Large Vision-Language Models via History-Aware Residual Guidance, *Xinrong Chen, Xu Chu, Yingmin Qiu, Hengyuan Zhang, Jing Xiong, Shiyu Tang, Shuai Liu, Shaokang Yang, Cheng Yang, Hayden Kwok-Hay So, Ngai Wong*
360 Prefill-Time Intervention for Mitigating Hallucination in Large Vision-Language Models, *Chengsheng Zhang, Chenghao Sun, Xinyan Jiang, Wei Li, Xinmei Tian*
361 SVHalluc: Benchmarking Speech–Vision Hallucination in Audio-Visual Large Language Models, *Chenshuang Zhang, Kyeong Seon Kim, Chengxin Liu, Tae-Hyun Oh*
362 Same Attention, Different Truths: Put Logit-Lens over Visual
❊ Attention to Detect and Mitigate LVLM Object Hallucination, *Zichuan Wang, Songlin Yang, Bo Peng, Zhenchen Tang, Yang Li, Beibei Dong, Jing Dong*
363 Understanding the Role of Hallucination in Reinforcement Post-Training of Multimodal Reasoning Models, *Gengwei Zhang, Jie Peng, Zhen Tan, Mufan Qiu, Hossein Nourkhiz Mahjoub, Vaishnav Tadiparthi, Kwonjoon Lee, Yanyong Zhang, Tianlong Chen*
364 Lyapunov Probes for Hallucination Detection in Large Foundation Models, *Bozhi Luan, Gen Li, Yalan Qin, Jifeng Guo, Yun Zhou, Faguo Wu, Hongwei Zheng, Wenjun Wu, Zhaoxin Fan*
365 Captain Safari: A World Engine with Pose-Aligned 3D Memory, *Yu-Cheng Chou, Xingrui Wang, Yitong Li, Jiahao Wang, Hanting Liu, Cihang Xie, Alan Yuille, Junfei Xiao*
366 Gen3R: 3D Scene Generation Meets Feed-Forward Reconstruction, *Jiaxin Huang, Yuanbo Yang, Bangbang Yang, Lin Ma, Yuewen Ma, Yiyi Liao*
367 PerpetualWonder: Long-horizon Action-conditioned 4D Scene
❊ Generation, *Jiahao Zhan, Zizhang Li, Hong-Xing Yu, Jiajun Wu*
368 CineScene: Implicit 3D as Effective Scene Representation for
❊ Cinematic Video Generation, *Kaiyi Huang, Yukun Huang, Yu Li, Jianhong Bai, Xintao Wang, Zinan Lin, Xuefei Ning, Jiwen Yu, Yu Wang, Xihui Liu*
369 DreamStereo: Towards Real-Time Stereo Inpainting for HD Videos, *Yuan Huang, Sijie Zhao, Jing Cheng, Hao Xu, Shaohui Jiao*
370 SeeThrough3D: Occlusion Aware 3D Control in Text-to-Image Generation, *Vaibhav Agrawal, Rishubh Parihar, Pradhaan S Bhat, Ravi Kiran Sarvadevabhatla, Venkatesh Babu Radhakrishnan*
371 RecEdit-Drive: 3D Reconstruction-Guided Spatiotemporal Video Editing for Autonomous Driving Scenes, *Yipeng Wu, Xin Wang, Chenghan Yang, Chong Wang, Dongdong Wu, Wanchao Su, Hengshuang Zhao, Wei Feng, Kairui Yang, Di Lin*
372 RAYNOVA: Scale-Temporal Autoregressive World Modeling in Ray Space, *Yichen Xie, Chensheng Peng, Mazen Abdelfattah, Yihan Hu, Jiezhi Yang, Eric Higgins, Ryan Brigden, Masayoshi Tomizuka, Wei Zhan*
373 RigMo: Unifying Rig and Motion Learning for Generative Animation, *Hao Zhang, Jiahao Luo, Bohui Wan, Yizhou Zhao, Zongrui Li, Michael Vasilkovsky, Chaoyang Wang, Jian Wang, Narendra Ahuja, Bing Zhou*
374 LaVR: Scene Latent Conditioned Generative Video Trajectory Re-Rendering using Large 4D Reconstruction Models, *Mingyang Xie, Numair Khan, Tianfu Wang, Naina Dhingra, Seonghyeon Nam, Haitao Yang, Zhuo Hui, Christopher Metzler, Andrea Vedaldi, Hamed Pirsiavash, Lei Luo*
375 WHU-MARS: A Multispectral Aerial-Ground Benchmark Towards Any-
❊ Scenario Person Re-Identification, *Yuxuan Zhao, Zhongao Zhou, Bin Yang, He Li, Jian Liang, Jun Chen, Bo Du, Mang Ye*
376 Detect Anything via Next Point Prediction, *Qing Jiang, Junan Huo, Xingyu Chen, Yuda Xiong, Zhaoyang Zeng, Yihao Chen, Tianhe Ren, Junzhi Yu, Lei Zhang*
377 Text-guided Feature Disentanglement for Cross-modal Gait
❊ Recognition, *Zhiyang Lu, Ming Cheng*
378 Distribution-Aligned Multimodal Fusion for Robust Object Detection, *Xiaohui Hao, Yanglin Pu, Yongjun Wang, Rui She*
379 PaQ-DETR: Learning Pattern and Quality-Aware Dynamic Queries for Object Detection, *Zhengjian Kang, Jun Zhuang, Kangtong Mo, Qi Chen, Rui Liu, Ye Zhang*
380 Portable Active Learning for Object Detection, *Rashi Sharma,*
❊ *Justin Timothy C. Bersamin, Karthikk Subramanian*
381 Efficiency Follows Global-Local Decoupling, *Zhenyu Yang, Gensheng Pei, Tao Chen, Yichao Zhou, Tianfei Zhou, Yazhou Yao, Fumin Shen*
382 VRCLIP: Multimodal Canonical Correlation Alignment for CLIP-Driven Vision-Radio Person Re-Identification, *Rui Zhang, Yaqi Wang, Yadong Li, Ruixu Geng, Jianyang Wang, Qijun Ying, Dongheng Zhang, Yang Hu, Yan Chen*
383 EReCu: Pseudo-label Evolution Fusion and Refinement with Multi-Cue Learning for Unsupervised Camouflage Detection, *Shuo Jiang, Gaojia Zhang, Min Tan, Yufei Yin, Gang Pan*

384 Expert-Teacher-Student Collaborative Learning for Domain Adaptive Object Detection, *Yiming Cui, Liang Li, Haibing Yin, Yuhan Gao, Xichun Sheng, Chenggang Yan*
385 CI-VID: A Coherent Interleaved Text-Video Dataset, *Yiming Ju, Jijin Hu, Zhengxiong Luo, Haoge Deng, hanyu Zhao, Li Du, Wenbo Xiao, Chengwei Wu, Donglin Hao, Xinlong Wang, Tengfei Pan*
386 Generalizable Video Quality Assessment via Weak-to-Strong Learning, *Linhan Cao, Wei Sun, Xiangyang Zhu, Kaiwei Zhang, Jun Jia, Yicong Peng, Dandan Zhu, Guangtao Zhai, Xiongkuo Min*
387 ❊ EgoSound: Benchmarking Sound Understanding in Egocentric Videos, *Bingwen Zhu, Yuqian Fu, Qiaole Dong, Guolei Sun, Tianwen Qian, Yuzheng Wu, Danda Pani Paudel, Yanwei Fu, Xiangyang Xue*
388 ❊ WorldMM: Dynamic Multimodal Memory Agent for Long Video Reasoning, *Woongyeong Yeo, Kangsan Kim, Jaehong Yoon, Sung Ju Hwang*
389 GIFT: Global Irreplaceability Frame Targeting for Efficient Video Understanding, *Junpeng Ma, Sashuai Zhou, Guanghao Li, Xin Gao, Yue Cao, Hengyu Zeng, Yuxiang Yan, Zhibin Wang, Jun Song, Bo Zheng, Shanghang Zhang, Jian Pu*
390 Select Less, Reason More: Prioritizing Evidence Purity for Video Reasoning, *Xuchen Li, Xuzhao Li, Shiyu Hu, Kaiqi Huang*
391 Ego2Web: A Web Agent Benchmark Grounded in Egocentric Videos, *Shoubin Yu, Lei Shu, Antoine Yang, Yao Fu, Srinivas Sunkara, Maria Wang, Jindong Chen, Mohit Bansal, Boqing Gong*
392 Compositional Transformation Reasoning for Composed Video Retrieval, *Sihong Huang, Jiaxin Wu, Dongmei Jiang, Yi Cai, Yaowei Wang, Xiaoyong Wei*
393 UniVBench: Towards Unified Evaluation for Video Foundation Models, *Jianhui Wei, Xiaotian Zhang, Yichen Li, Yuan Wang, Yan Zhang, Ziyi Chen, Zhihang Tang, Wei Xu, Zuozhu Liu*
394 NAMI: Efficient Image Generation via Bridged Progressive Rectified Flow Transformers, *Yuhang Ma, Bo Cheng, Shanyuan Liu, Hongyi Zhou, Liebucha Wu, Dawei Leng, Yuhui Yin*
395 InverFill: One-Step Inversion for Enhanced Few-Step Diffusion Inpainting, *Duc Vu, Kien Nguyen, Trong-Tung Nguyen, Ngan Nguyen, Phong Nguyen, Khoi Nguyen, Cuong Pham, Anh Tran*
396 ❊ TimeRipples: Accelerating vDiTs by Understanding the Spatio-Temporal Correlations in Latent Space, *Wenxuan Mao, Yulin Sun, Aiyue Chen, Jing Lin, Yiwu Yao, Yiming Gan, Jieru Zhao, Jingwen Leng, Minyi Guo, Yu Feng*
397 ProcessMaker: A Generalized Process Visualization Framework with Adaptive Sequence Steps on Diffusion Transformers, *Mengling Xu, Sisi You, Yaning Li, Bing-Kun Bao*
398 MeanFlow Transformers with Representation Autoencoders, *Zheyuan Hu, Chieh-Hsin Lai, Ge Wu, Yuki Mitsufuji, Stefano Ermon*
399 DiT-IC: Aligned Diffusion Transformer for Efficient Image Compression, *Junqi Shi, Ming Lu, Xingchen Li, Anle Ke, Ruiqi Zhang, Zhan Ma*
400 FARMER: Flow AutoRegressive Transformer over Pixels, *Guangting Zheng, Qinyu Zhao, Tao Yang, Fei Xiao, Zhijie Lin, Jie Wu, Jiajun Deng, Yanyong Zhang, Rui Zhu*
401 Probabilistic Precipitation Nowcasting with Rectified Flow Transformers, *Johannes Schusterbauer, Jannik Wiese, Nick Stracke, Timy Phan, Björn Ommer*
402 ❊ FlowDC: Flow-Based Decoupling-Decay for Complex Image Editing, *Yilei Jiang, Zhen Wang, Yanghao Wang, Jun Yu, Yueting Zhuang, Jun Xiao, Long Chen*
403 High-Fidelity Diffusion Face Swapping with ID-Constrained Facial Conditioning, *Dailan He, Xiahong Wang, Shulun Wang, Hao Shao, Bingqi Ma, Guanglu Song, Yu Liu, Hongsheng Li*
404 ❊ 3D-Object Perception Transformer (3PT), *Agastya Kalra, Tim Salzmann, Guy Stoppi, Dmitrii Marin, Rishav Agarwal, Vage Taamazyan, Martin Bokeloh, Stefan Hinterstoisser, Anton Boykov, Alberto Dall'Olio, Pravin Dangol, Kartik Venkataraman, Huaijin Chen*
405 SemLT3D: Semantic-Guided Expert Distillation for Camera-only Long-Tailed 3D Object Detection, *Hao Vo, Khoa Vo, Thinh Phan, Ngo Xuan Cuong, Gianfranco Doretto, Hien Nguyen, Anh Nguyen, Ngan Le*
406 Spe-BEVHead: Rethinking the Detection Head Design for Bird's-Eye-View Object Detection, *Junshu Zhang, Sicheng Zhao, Xin Zhao, Fan Yang, Ruike Chen, Jungong Han, Guiguang Ding*
407 Unsupervised Multi-agent and Single-agent Perception from Cooperative Views, *Haochen Yang, Baolu Li, Lei Li, Delin Ren, Jiacheng Guo, Minghai Qin, Tianyun Zhang, Hongkai Yu*
408 Zoo3D: Zero-Shot 3D Object Detection at Scene Level, *Andrey Lemeshko, Bulat Gabdullin, Nikita Drozdov, Anton Konushin, Danila Rukhovich, Maksim Kolodiazhnyi*
409 Beyond Appearance: Camouflaged Object Detection via Geometric Structure, *Jinyu Han, Changguang Wu, Fuming Sun, Jinhui Tang*
410 ❊ SABER: Spatially Consistent 3D Universal Adversarial Objects for BEV Detectors, *Aixuan Li, Mochu Xiang, Bosen Hou, Zhexiong Wan, Jing Zhang, Yuchao Dai*
411 AceTone: Bridging Words and Colors for Conditional Image Grading, *Tianren Ma, Mingxiang Liao, Xijin Zhang, Qixiang Ye*
412 Do VLMs Perceive or Recall? Probing Visual Perception vs. Memory with Classic Visual Illusions, *Xiaoxiao Sun, Mingyang Li, Kun Yuan, Min Woo Sun, Mark Endo, Shengguang Wu, Changlin Li, Yuhui Zhang, Zeyu Wang, Serena Yeung-Levy*
413 Pixels Don't Lie (But Your Detector Might): Bootstrapping MLLM-as-a-Judge for Trustworthy Deepfake Detection and Reasoning Supervision, *Kartik Kuckreja, Parul Gupta, Muhammad Haris Khan, Abhinav Dhall*
414 UI-Lens: Assessing General MLLMs' Potential to Automate UI Display Quality Assurance, *Wei Xiang, Yexinrui Wu, Xinli Chen, Xinran Li, Shi Chen*
415 Seeing is Improving: Visual Feedback for Iterative Text Layout Refinement, *Junrong Guo, Shancheng Fang, Yadong Qu, Hongtao Xie*
416 Is your VLM Sky-Ready? A Comprehensive Spatial Intelligence Benchmark for UAV Navigation, *Lingfeng Zhang, Yuchen Zhang, Hongsheng Li, Haoxiang Fu, Yingbo Tang, Hangjun Ye, Long Chen, Xiaojun Liang, Xiaoshuai Hao, Wenbo Ding*
417 Linking Perception, Confidence and Accuracy in MLLMs, *Yuetian Du, Yucheng Wang, Rongyu Zhang, Zhijie Xu, Boyu Yang, Ming Kong, Jie Liu, Qiang Zhu*
418 AVA-Bench: Atomic Visual Ability Benchmark for Vision Foundation Models, *Zheda Mai, Arpita Chowdhury, Zihe Wang, Sooyoung Jeon, Lemeng Wang, Jiacheng Hou, Jihyung Kil, Wei-Lun Chao*
419 ❊ Learning to Focus and Precise Cropping: A Reinforcement Learning Framework with Information Gaps and Grounding Loss for MLLMs, *Xuanpu Zhao, Zhentao Tan, Dianmo Sheng, Tianxiang Chen, Yao Liu, Yue Wu, Tao Gong, Qi Chu, Nenghai Yu*
420 From Pixel to Precision: Enhancing Handwritten Mathematical Expression Recognition with Image-Level Reward, *Ze Liu, Kai Zhang, Xianquan Wang, Shuochen Liu, Jiaxian Yan, Yupeng Han, Qi Liu*
421 Rethinking Pose Refinement in 3D Gaussian Splatting under Pose Prior and Geometric Uncertainty, *Mangyu Kong, Jaewon Lee, Seongwon Lee, Euntai Kim*
422 Revisiting Pose Sensitivity in Splat-based Computed Tomography under Sparse-view Reconstruction, *Kiseok Choi, Hyeongjun Cho, Inchul Kim, Min H. Kim*
423 ❊ Seele: A Unified Acceleration Framework for Real-Time Gaussian Splatting on Mobile Devices, *He Zhu, Xiaotong Huang, Zihan Liu, Weikai Lin, Xiaohong Liu, Zhezhi He, Jingwen Leng, Minyi Guo, Yu Feng*
424 GHPT: Real-Time Relightable Gaussian Splatting using Hybrid Path Tracing, *Jinyang Bo, Fan Dou, Wenrui Quan, Shangxun Liu, Yang Xu, Yuhe Zhang, Kang Li, Guohua Geng*
425 PolarGuide-GSDR: 3D Gaussian Splatting Driven by Polarization Priors and Deferred Reflection for Real-World Reflective Scenes, *Derui Shan, Qian Qiao, Hao Lu, Tao Du, Peng Lu*
426 ❊ EcoSplat: Efficiency-controllable Feed-forward 3D Gaussian Splatting from Multi-view Images, *Minh-Quan Viet Bui, Jongmin Park, Juan Luis Gonzalez, Jaeho Moon, Jihyong Oh, Munchurl Kim*
427 SGS-Intrinsic: Semantic-Invariant Gaussian Splatting for Sparse-View Indoor Inverse Rendering, *Jiahao Niu, Rongjia Zheng, Wenju Xu, Wei-Shi Zheng, Qing Zhang*
428 GIFSplat: Generative Prior-Guided Iterative Feed-Forward 3D Gaussian Splatting from Sparse Views, *Tianyu Chen, Wei Xiang, Kang Han, Yu Lu, Di Wu, Gaowen Liu, Ramana Rao Kompella*
429 3D Gaussian Splatting with Self-Constrained Priors for High Fidelity Surface Reconstruction, *Takeshi Noda, Yu-Shen Liu, Zhizhong Han*
430 FilterGS: Traversal-Free Parallel Filtering and Adaptive Shrinking for Large-Scale LoD 3D Gaussian Splatting, *Yixian Wang, Haolin Yu, Jiadong Tang, Yu Gao, Xihan Wang, Yufeng Yue, Yi Yang*

431 TWINGS: Thin Plate Splines Warp-aligned Initialization for Sparse-View Gaussian Splatting, *Hyeseong Kim, Geonhui Son, Deukhee Lee, Dosik Hwang*
432 VarSplat: Uncertainty-aware 3D Gaussian Splatting for Robust RGB-D SLAM, *Anh Thuan Tran, Jana Kosecka*
433 SpeeDe3DGS: Speedy Deformable 3D Gaussian Splatting with Temporal Pruning and Motion Grouping, *Allen Tu, Haiyang Ying, Alex Hanson, Yonghan Lee, Tom Goldstein, Matthias Zwicker*
434 FastGS: Training 3D Gaussian Splatting in 100 Seconds, *Shiwei Ren,* ❊ *Tianci Wen, Yongchun Fang, Biao Lu*
435 BrepGaussian: CAD reconstruction from Multi-View Images with Gaussian Splatting, *Jiaxing Yu, Dongyang Ren, Hangyu Xu, Zhouyuxiao Yang, Yuanqi Li, Jie Guo, Zhengkang Zhou, Yanwen Guo*
436 ODGS-SLAM: Omnidirectional Gaussian Splatting SLAM, *Stefan Spiss, Joey Hieronimy, Marcel Ritter, Matthias Harders*
437 BA-GS: Bayesian Adaptive Gaussian Splatting for SFM-Free 3D Reconstruction, *Zhongjie Ma, Di Lin, Xin Wang, Haotian Dong, Chong Wang, Dongdong Wu, Changqing Zhang*
438 FSFSplatter: Geometrically Accurate Reconstruction with Free Sparse-view Images within 2 minutes, *Yibin Zhao, Yihan Pan, Jun Nan, Liwei Chen, Jianjun Yi*
439 ViRC: Enhancing Visual Interleaved Mathematical CoT with Reason Chunking, *Lihong Wang, Liangqi Li, Weiwei Feng, Jiamin Wu, Changtao Miao, Tieru Wu, Rui Ma, Bo Zhang, Zhe Li*
440 When Visualizing is the First Step to Reasoning: MIRA, a Benchmark for Visual Chain-of-Thought, *Yiyang Zhou, Haoqin Tu, Zijun Wang, Zeyu Wang, Niklas Muennighoff, Fan Nie, Chaorui Deng, Shen Yan, Haoqi Fan, Yejin Choi, James Zou, Cihang Xie, Huaxiu Yao, Qinghao Ye*
441 PixDLM: A Dual-Path Multimodal Language Model for UAV Reasoning ❊ Segmentation, *Shuyan Ke, Yifan Mei, Changli Wu, Yonghan Zheng, Jiayi Ji, Liujuan Cao, Rongrong Ji*
442 Can a Second-View Image Be a Language? Geometric and Semantic Cross-Modal Reasoning for X-ray Prohibited Item Detection, *Chuang Peng, Renshuai Tao, Zhongwei Ren, Xianglong Liu, Yunchao Wei*
443 VCU-Bridge: Hierarchical Visual Connotation Understanding via Semantic Bridging, *Ming Zhong, Yuanlei Wang, Liuzhou Zhang, Ruichuan An, Renrui Zhang, Hao Liang, Ming Lu, Ying Shen, Wentao Zhang*
444 Learning to See through Illumination Extremes with Event Streaming in Multimodal Large Language Models, *Baoheng Zhang, Jiahui Liu, Gui Zhao, Weizhou Zhang, Yixuan Ma, Jun Jiang, Yingxian Chen, Wilton W.T. Fok, Xiaojuan Qi, Hayden Kwok-Hay So*
445 VOLD: Reasoning Transfer from LLMs to Vision-Language Models via On-Policy Distillation, *Walid Bousselham, Hilde Kuehne, Cordelia Schmid*
446 Cut to the Chase: Training-free Multimodal Summarization via Chain-of-Events, *Xiaoxing You, Qiang Huang, Lingyu Li, Xiaojun Chang, Jun Yu*
447 UVU: Improving Multimodal Understanding via Vision-Language Unified Autoregressive Paradigm, *Zhehan Kan, Xinghua Jiang, Yanlin Liu, Xiaochen Yang, Zhixiang Wei, Shifeng Liu, Yubo Zhu, Qingmin Liao, Wenming Yang, Xin Li, Yinsong Liu, Deqiang Jiang, Xing Sun*
448 PointThinker: Point-Incentivized Parallel Thinking for Multimodal Large Language Model, *Zhengdong Hu, Chao Wang, Fengyun Rao, Jing LYU, Hehe Fan, Yi Yang*
449 OctoMed: Data Recipes for State-of-the-Art Multimodal Medical Reasoning, *Timothy Ossowski, Sheng Zhang, Qianchu Liu, Guanghui Qin, Reuben Tan, Tristan Naumann, Junjie Hu, Hoifung Poon*
450 HoneyBee: Data Recipes for Vision-Language Reasoners, *Hritik Bansal, Devendra Singh Sachan, Kai-Wei Chang, Aditya Grover, Gargi Ghosh, Wen-tau Yih, Ramakanth Pasunuru*
451 VisPlay: Self-Evolving Vision-Language Models, *Yicheng He, Chengsong Huang, Zongxia Li, Jiaxin Huang, Yonghui Yang*
452 Chart-FR1: Visual Focus-Driven Fine-Grained Reasoning on Dense Charts, *Hongkun Pan, Yuwei Wu, Wanyi Hong, Shenghui Hu, Qitong Yan, Yi Yang, Rufei Han, Changju Zhou, Minfeng Zhu, Dongming Han, Wei Chen*
453 Thinking-while-Generating: Interleaving Textual Reasoning throughout Visual Generation, *Ziyu Guo, Renrui Zhang, Hongyu Li, Manyuan Zhang, Xinyan Chen, Sifan Wang, Yan Feng, Peng Pei, Pheng-Ann Heng*
454 ApET: Approximation-Error Guided Token Compression for Efficient ❊ VLMs, *Qiankun Ma, Ziyao Zhang, Haofei Wang, Zhen Song, Jie Chen, Hairong Zheng*
455 Granulon: Awakening Pixel-Level Visual Encoders with Adaptive Multi-Granularity Semantics for MLLM, *Junyuan Mao, Qiankun Li, Linghao Meng, Zhicheng He, Xinliang Zhou, Kun Wang, Yang Liu, Yueming Jin*
456 Vision Transformers Need More Than Registers, *Cheng Shi, Yizhou Yu, Sibei Yang*
457 Head-wise Adaptive Rotary Positional Encoding for Fine-Grained Image Generation, *Jiaye Li, Baoyou Chen, Hui Li, Zilong Dong, Jingdong Wang, Siyu Zhu*
458 PRISM: Video Dataset Condensation with Progressive Refinement and Insertion for Sparse Motion, *Jaehyun Choi, Jiwan Hur, Gyojin Han, Jaemyung Yu, Junmo Kim*
459 AdaSVD: Singular Value Decomposition with Adaptive Mechanisms for Large Multimodal Models, *Zhiteng Li, Mingyuan Xia, Jingyuan Zhang, Zheng Hui, Haotong Qin, Linghe Kong, Yulun Zhang, Xiaokang Yang*
460 ReFTA: Breaking the Weight Reconstruction Bottleneck in Tensorized Parameter-Efficient Fine-Tuning, *Jingjing Zheng, Anda Tang, Qiangqiang Mao, Zhouchen Lin, Yankai Cao*
461 HTTM: Head-wise Temporal Token Merging for Faster VGGT, *Weitian Wang, Lukas Meiner, Rai Shubham, Cecilia De La Parra, Akash Kumar*
462 Reparameterized Tensor Ring Functional Decomposition for Multi-Dimensional Data Recovery, *Yangyang Xu, Junbo Ke, You-Wei Wen, Chao Wang*
463 Self-Attention Driven Tensor Representation for High-Order Data ❊ Recovery, *Zhi-Wei Shi, Yu-Bang Zheng, Heng-Chao Li*
464 PlanaReLoc: Camera Relocalization in 3D Planar Primitives via Region-Based Structure Matching, *Hanqiao Ye, Yuzhou Liu, Yangdong Liu, Shuhan Shen*
465 MOGeo: Beyond One-to-One Cross-View Object Geo-localization, *Bo Lv, Qingwang Zhang, Le Wu, Yuanyuan Li, Yingying Zhu*
466 Homaloidal parametrization for detecting critical two-view ❊ configurations, *Rakshith Madhavan, Matteo Forlivesi, Marina Bertolini, Cristina Turrini, Federica Arrigoni, Luca Magri*
467 AsymLoc: Towards Asymmetric Feature Matching for Efficient Visual ❊ Localization, *Mohammad Omama, Gabriele Berton, Eric Foxlin, Yelin Kim*
468 MMLandmarks: a Cross-View Instance-Level Benchmark for Geo-Spatial Understanding, *Oskar Kristoffersen, Alba Reinders Sánchez, Morten Rieger Hannemose, Anders Bjorholm Dahl, Dim P. Papadopoulos*
469 Asking like Socrates: Socrates helps VLMs understand remote sensing images, *Run Shao, Ziyu Li, Zhaoyang Zhang, Linrui Xu, Xinran He, Hongyuan Yuan, Bolei He, Yongxing Dai, Yiming Yan, Yijun Chen, Wang Guo, Haifeng Li*
470 GTR-Turbo: Merged Checkpoint is Secretly a Free Teacher for Agentic VLM Training, *Tong Wei, Yijun Yang, Changhao Zhang, Junliang Xing, Yuanchun Shi, Zongqing Lu, Deheng Ye*
471 Let VLMs Grade Their Own Thoughts: A Self-Quantification Approach to Reasoning-Aware Reward Modeling, *Xing Xi, Yu Qiu, Ronghua Luo, Peixian Chen, peilin tong*
472 SciEducator: Scientific Video Understanding and Educating via Deming-Cycle Multi-Agent System, *Zhiyu Xu, Weilong Yan, Yufei Shi, Xin Meng, Tao He, Huiping Zhuang, Ming Li, Hehe Fan*
473 SenseSearch: Empowering Vision-Language Models with High-❊ Resolution Agentic Search-Reasoning via Reinforcement Learning, *Yong Xien Chng, Tao Hu, Wenwen Tong, Xueheng Li, Jiandong Chen, Haojia Yu, Jiefan Lu, Hewei Guo, Hanming Deng, Chengjun Xie, Gao Huang, Lewei Lu*
474 Scaling Agentic Reinforcement Learning for Tool-Integrated Reasoning in VLMs, *Meng Lu, Ran Xu, Yi Fang, Wenxuan Zhang, Yue Yu, Gaurav Srivastava, Yuchen Zhuang, Mohamed Elhoseiny, Charles Fleming, Carl Yang, Zhengzhong Tu, Yang Xie, Guanghua Xiao, Di Jin, Wenqi Shi, Xuan Wang*
475 VideoSSR: Video Self-Supervised Reinforcement Learning, *Zefeng He, Xiaoye Qu, Yafu Li, Siyuan Huang, Daizong Liu, Yu Cheng*
476 Neurodynamics-Driven Coupled Neural P Systems for Multi-Focus Image Fusion, *Bo Li, Yunkuo Lei, Tingting Bao, Hang Yan, Yaxian Wang, Weiping Fu, Lingling Zhang, Jun Liu*

477 MagicFuse: Single Image Fusion for Visual and Semantic Reinforcement, *Hao Zhang, Yanping Zha, Zizhuo Li, Meiqi Gong, Jiayi Ma*
478 Bridging Pixels and Words: Mask-Aware Local Semantic Fusion for Multimodal Media Verification, *Zizhao Chen, Ping Wei, Ziyang Ren, Huan Li, Xiangru Yin*
479 ❊ Human-Centric Multi-Exposure Fusion: Benchmark and Bi-level Cognition Distillation Framework, *Jingjie Shang, Tengyu Ma, Heng Zhang, Jinyuan Liu, Risheng Liu, Yuan Wang, Xiaochen Bo*
480 ConceptPose: Training-Free Zero-Shot Object Pose Estimation using Concept Vectors, *Liming Kuang, Yordanka Velikova, Mahdi Saleh, Jan-Nico Zaech, Danda Pani Paudel, Benjamin Busam*
481 A Closer Look at Cross-Domain Few-Shot Object Detection: Fine-Tuning Matters and Parallel Decoder Helps, *Xuanlong Yu, Youyang Sha, Longfei Liu, Xi Shen, Di Yang*
482 ❊ NAF: Zero-Shot Feature Upsampling via Neighborhood Attention Filtering, *Loïck Chambon, Paul Couairon, Éloi Zablocki, Alexandre Boulch, Nicolas Thome, Matthieu Cord*
483 Universal-to-Specific: Dynamic Knowledge-Guided Multiple Instance Learning for Few-Shot Whole Slide Image Classification, *Junjian Li, Hulin Kuang, Jin Liu, Hailin Yue, Mengshen He, Jianxin Wang*
484 SOTA: Self-adaptive Optimal Transport for Zero-Shot Classification with Multiple Foundation Models, *Zhanxuan Hu, Qiyu Xu, Yu Duan, Yonghang Tai, Huafeng Li*
485 Uni-DAD: Unified Distillation and Adaptation of Diffusion Models for Few-step Few-shot Image Generation, *Yara Bahram, Mélodie Desbos, Mohammadhadi Shateri, Eric Granger*
486 Streamlined Knowledge Distillation, *Hyeon-Jin Jeong, Han-Jin Lee, Seok-Hwan Choi*
487 Generalizable Knowledge Distillation from Vision Foundation Models for Semantic Segmentation, *Chonghua Lv, Dong Zhao, Shuang Wang, Dou Quan, Ning Huyan, Nicu Sebe, Zhun Zhong*
488 IMS3: Breaking Distributional Aggregation in Diffusion-Based Dataset Distillation, *Chenru Wang, Yunyi Chen, Zijun Yang, Joey Tianyi Zhou, Chi Zhang*
489 Continuous Exposure-Time Modeling for Realistic Atmospheric Turbulence Synthesis, *Junwei Zeng, Dong Liang, Sheng-Jun Huang, Kun Zhan, Songcan Chen*
490 ❊ 240FPS Stereo Vision from Monocular Mixed Spikes, *Yeliduosi Xiaokaiti, Yakun Chang, Yang Bai, Zhaojun Huang, Peiqi Duan, Boxin Shi*
491 D^2-FOSA: Dual-Diffusion Guided EEG-to-Image Reconstruction with Frequency-Oriented Semantic Alignment, *Chenglong Yu, Shuai Shen, Xiangsheng Li, Yang Li*
492 Self-Diffusion Driven Blind Imaging, *Yanlong Yang, Guanxiong Luo*
493 Differentiable Stroke Planning with Dual Parameterization for Efficient and High-Fidelity Painting Creation, *Jinfan Liu, Wuze Zhang, Zhangli Hu, Zhehan Zhao, Ye Chen, Bingbing Ni*
494 Solvability of the Viewing Graph Under the Affine Camera Model, *Gabriele Pedroni, Rakshith Madhavan, Federica Arrigoni*
495 DiffBMP: Differentiable Rendering with Bitmap Primitives, *Seongmin Hong, Junghun James Kim, Daehyeop Kim, Insoo Chung, Se Young Chun*
496 Splat-Based Metal Artifact Reduction in Cone-Beam CT via Compact Attenuation Modeling, *Kiseok Choi, Jaemin Cho, Inchul Kim, Min H. Kim*
497 Lumosaic: Hyperspectral Video via Active Illumination and Coded-Exposure Pixels, *Dhruv Verma, Andrew Qiu, Roberto Rangel, Ayandev Barman, Hao Yang, Chenjia Hu, Fengqi Zhang, Roman Genov, David B. Lindell, Kiriakos N. Kutulakos, Alex Mariakakis*
498 Towards Universal Computational Aberration Correction in Photographic Cameras: A Comprehensive Benchmark Analysis, *Xiaolong Qian, Qi Jiang, Yao Gao, Lei Sun, Zhonghua Yi, Kailun Yang, Luc Van Gool, Kaiwei Wang*
499 Multi-View Hierarchical Alignment Learning for Spatial Transcriptomics, *Zhengzhong Zhu, Liangjin Liu, Pei Zhou, Shiquan Min, Jiangping Zhu*
500 FEAST: Fully Connected Expressive Attention for Spatial Transcriptomics, *Taejin Jeong, Joohyeok Kim, Jinyeong Kim, Chanyoung Kim, Seong Jae Hwang*
501 ❊ TRIDENT: A Trimodal Cascade Generative Framework for Drug and RNA-Conditioned Cellular Morphology Synthesis, *Rui Peng, Ziru Liu, Lingyuan Ye, Yuxing Lu, Boxin Shi, Jinzhuo Wang*
502 OrienPose: Orientation-Guided Novel View Synthesis for Single-Image Unseen Object Pose Estimation, *Yating Liu, Zhaoshuai Qi, Yang Zou, Yongnan Yang, Shizhou Zhang, Yanning Zhang*
503 Illustrator's Depth: Monocular Layer Index Prediction for Image Decomposition, *Nissim Maruani, Peiying Zhang, Siddhartha Chaudhuri, Matthew Fisher, Nanxuan Zhao, Vladimir G. Kim, Pierre Alliez, Mathieu Desbrun, Wang Yifan*
504 Depth Any Panoramas: A Foundation Model for Panoramic Depth Estimation, *Xin Lin, Meixi Song, Dizhe Zhang, Wenxuan Lu, Haodong Li, Bo Du, Ming-Hsuan Yang, Truong Nguyen, Lu Qi*
505 Seeing Depth Through Frequency and Motion: A Progressive Training Paradigm for Monocular Depth Estimation, *Ke Li, Bolin Song, Hongbo Liu*
506 GeoGuide: Hierarchical Geometric Guidance for Open-Vocabulary 3D Semantic Segmentation, *Xujing Tao, Chuxin Wang, Yubo Ai, Zhixin Cheng, Zhuoyuan Li, Liangsheng Liu, Yujia Chen, Xinjun Li, Qiao Li, Wenfei Yang, Tianzhu Zhang*
507 ❊ B^3-Seg: Camera-Free, Training-Free 3DGS Segmentation via Analytic EIG and Beta-Bernoulli Bayesian Updates, *Hiromichi Kamata, Samuel Arthur Munro, Fuminori Homma*
508 PE3R: Perception-Efficient 3D Reconstruction, *Jie Hu, Shizun Wang, Xinchao Wang*
509 GS-ASM: 2DGS-Supervised Active Stereo Matching, *Zhengling Wu, Rongfeng Lu, Quan Chen, Longjian Zeng, Ming Lu, Yaoqi Sun, Yahong Chen, Baofeng Ji, Chenggang Yan*
510 Real2Sim2Real: RetinalDepth-64K for Depth Estimation in Posterior Segment Ophthalmic Surgery, *Bingwen Dong, Gan Liu, Xiaoxi Lu, Guangcheng Chen, Jialu Zhang, Yan Hu, Xiaoqing Zhang, Jiang Liu*
511 Iris: Bringing Real-World Priors into Diffusion Model for Monocular Depth Estimation, *Xinhao Cai, Gensheng Pei, Zeren Sun, Yazhou Yao, Fumin Shen, Wenguan Wang*
512 InfiniDepth: Arbitrary-Resolution and Fine-Grained Depth Estimation with Neural Implicit Fields, *Hao Yu, Haotong Lin, Jiawei Wang, Jiaxin Li, Yida Wang, Xueyang Zhang, Yue Wang, Xiaowei Zhou, Ruizhen Hu, Sida Peng*
513 AirSim360: A Panoramic Simulation Platform within Drone View, *Xian Ge, Yuling Pan, Yuhang Zhang, Xiang Li, Weijun Zhang, Dizhe Zhang, Zhaoliang Wan, Xin Lin, Xiangkai Zhang, Juntao Liang, Xiangtai Li, WenJie Jiang, Bo Du, Ming-Hsuan Yang, Lu Qi*
514 Radar-Guided Polynomial Fitting for Metric Depth Estimation, *Patrick Rim, Hyoungseob Park, Vadim Ezhov, Jeffrey Moon, Alex Wong*
515 UniDAC: Universal Metric Depth Estimation for Any Camera, *Girish Chandar Ganesan, Yuliang Guo, Liu Ren, Xiaoming Liu*
516 SCE-Depth: A Spherical Compound Eye Framework for Wide FOV Depth Estimation, *Yi Zhu, Hao Xiong, Lin Xiao, Ranfeng Shi, Qinying Gu, Leilei Gu*
517 I-Scene: 3D Instance Models are Implicit Generalizable Spatial Learners, *Lu Ling, Yunhao Ge, Yichen Sheng, Aniket Bera*
518 REVIVE 3D: Refinement via Encoded Voluminous Inflated prior for Volume Enhancement, *Hankyeol Lee, Wooyeol Baek, Seongdo Kim, Jongyoo Kim*
519 Muses: Designing, Composing, Generating Nonexistent Fantasy 3D Creatures without Training, *Hexiao Lu, Xiaokun Sun, Zeyu Cai, Hao Guo, Ying Tai, Jian Yang, Zhenyu Zhang*
520 EI-Part: Explode for Completion and Implode for Refinement, *Wanhu Sun, Zhongjin Luo, Heliang Zheng, Jiahao Chang, Chongjie Ye, Huiang He, Shengchu Zhao, Rongfei Jia, Xiaoguang Han*
521 MorphAny3D: Unleashing the Power of Structured Latent in 3D Morphing, *Xiaokun Sun, Zeyu Cai, Hao Tang, Ying Tai, Jian Yang, Zhenyu Zhang*
522 Fast3Dcache: Training-free 3D Geometry Synthesis Acceleration, *Mengyu Yang, Yanming Yang, Chenyi Xu, Chenxi Song, Yufan Zuo, Tong Zhao, Ruibo Li, Chi Zhang*
523 ViLearn: Accelerating Training Convergence of Image-to-3D Generation via Visibility Learning, *Rui Chen, Jianfeng Zhang, Jing Lin, Xuanyu Yi, Yixun Liang, Guan Luo, Xiu Li, Zeming Li, Ping Tan*
524 FlashMesh: Faster and Better Autoregressive Mesh Synthesis via Structured Speculation, *Tingrui Shen, Yiheng Zhang, Chen Tang, Chuan Ping, Zixing Zhao, Le Wan, Yuwang Wang, Ronggang Wang, Shengfeng He*

525 X-Part: High Fidelity And Structure Coherent Shape Decomposition And Completion, *Xinhao Yan, Jiachen Xu, Yang Li, Changfeng Ma, Yunhan Yang, Chunshi Wang, Zibo Zhao, Zeqiang Lai, Yunfei Zhao, Zhuo Chen, Chunchao Guo*
526 Realiz3D: 3D Generation Made Photorealistic via Domain-Aware Learning, *Ido Sobol, Kihyuk Sohn, Yoav Blum, Egor Zakharov, Max Bluvstein, Andrea Vedaldi, Or Litany*
527 TopoMesh: High-Fidelity Mesh Autoencoding via Topological Unification, *Guan Luo, Xiu Li, Rui Chen, Xuanyu Yi, Jing Lin, Chia Hao Chen, Jiahang Liu, Song-Hai Zhang, Jianfeng Zhang*
528 Nestwork: Conditional 3D Furnished House Layout Generation through Latent Heterogeneous Graph Diffusion, *Shuhan Miao, Biru Cao, Junling Zhuang*
529 TEXTRIX: Latent Attribute Grid for Native Texture Generation and Beyond, *Yifei Zeng, Yajie Bao, Jiachen Qian, Shuang Wu, Youtian Lin, Hao Zhu, Buyu Li, Feihu Zhang, Xun Cao, Yao Yao*
530 Beyond Geometry: Artistic Disparity Synthesis for Immersive 2D-to-3D, *Ping Chen, Zezhou Chen, Xingpeng Zhang, Yanlin Qian, Huan Hu, Xiang Liu, Zipeng Wang, Xin Wang, Zhaoxiang Liu, Kai Wang, Shiguo Lian*
531 WorldGen: From Text to Traversable and Interactive 3D Worlds, *Dilin Wang, Hyunyoung Jung, Tom Monnier, Kihyuk Sohn, Chuhang Zou, Xiaoyu Xiang, Yu-Ying Yeh, Di Liu, Zixuan Huang, Thu Nguyen-Phuoc, Yuchen Fan, Sergiu Oprea, Ziyan Wang, Roman Shapovalov, Nikolaos Sarafianos, Thibault Groueix, Antoine Toisoul, Prithviraj Dhar, Xiao Chu, Minghao Chen, Geon Yeong Park, Rakesh Ranjan, Andrea Vedaldi*
532 ExMesh: EXplicit Mesh Reconstruction with Topology Adaptation, *Chuanjin Fan, Lifan Wu, Wenjie Chang, Hanzhi Chang, Wenfei Yang, Tianzhu Zhang*
533 SceneMaker: Open-set 3D Scene Generation with Decoupled De-occlusion and Pose Estimation Model, *Yukai Shi, Weiyu Li, Zihao Wang, Hongyang Li, Xingyu Chen, Ping Tan, Lei Zhang*
534 ShapeR: Robust Conditional 3D Shape Generation from Casual Captures, *Yawar Siddiqui, Duncan Frost, Samir Aroudj, Armen Avetisyan, Henry Howard-Jenkins, Daniel DeTone, Pierre Moulon, Qirui Wu, Zhengqin Li, Julian Straub, Richard Newcombe, Jakob Engel*
535 SwiftTailor: Efficient 3D Garment Generation with Geometry Image
❊ Representation, *Phuc Pham, Uy Dieu Tran, Binh-Son Hua, Phong Nguyen*
536 3DrawAgent: Teaching LLM to Draw in 3D with Early Contrastive
❊ Experience, *Hongcan Xiao, Xinyue Xiao, Yilin Wang, Yue Zhang, Yonggang Qi*
537 Sculpt4D: Generating 4D Shapes via Sparse-Attention Diffusion Transformers, *Minghao Yin, Wenbo Hu, Jiale Xu, Ying Shan, Kai Han*
538 HiFi-BRep: High-Fidelity Latent Representation for Robust B-Rep Generation, *Junhao Hou, Chenqi Luo, Pufan Wang, Jiaying Lu, Yusheng Liu, Feiwei Qin, Meie Fang, Kun Zhou*
539 PhysGen: Physically Grounded 3D Shape Generation for Industrial Design, *Yingxuan You, Chen Zhao, Hantao Zhang, Ming Xu, Pascal Fua*
540 Perceptual 3D Simulation With Physical World Modeling, *Wanhee Lee, Klemen Kotar, Rahul Mysore Venkatesh, Jared Watrous, Daniel LK Yamins*
541 EchoFoley: Event-Centric Hierarchical Control for Video Grounded Creative Sound Generation, *Bingxuan Li, Yiming Cui, Yicheng He, Yiwei Wang, Shu Zhang, Longyin Wen, Yulei Niu*
542 Active Intelligence in Video Avatars via Closed-loop World Modeling, *Xuanhua He, Tianyu Yang, Ke Cao, Ruiqi Wu, Cheng Meng, Yong Zhang, Zhuoliang Kang, Xiaoming Wei, Qifeng Chen*
543 Enhancing Spatial Understanding in Image Generation via Reward Modeling, *Zhenyu Tang, Chaoran Feng, Yufan Deng, Jie Wu, Xiaojie Li, Rui Wang, Yunpeng Chen, Daquan Zhou*
544 Seeing What Matters: Visual Preference Policy Optimization for Visual Generation, *Ziqi Ni, Yuanzhi Liang, Rui Li, Yi Zhou, Haibin Huang, Chi Zhang, Xuelong Li*
545 TAG-MoE: Task-Aware Gating for Unified Generative Mixture-of-Experts, *Yu Xu, Hongbin Yan, Juan Cao, Yiji Cheng, Tiankai Hang, Runze He, Zijin Yin, Shiyi Zhang, Yuxin Zhang, Jintao Li, Chunyu Wang, Qinglin Lu, Tong-Yee Lee, Fan Tang*
546 Identity-Preserving Image-to-Video Generation via Reward-Guided Optimization, *Liao Shen, Wentao Jiang, Yiran Zhu, Jiahe Li, Tiezheng Ge, Zhiguo Cao, Bo Zheng*
547 JarvisEvo: Towards a Self-Evolving Photo Editing Agent with Synergistic Editor-Evaluator Optimization, *Yunlong Lin, Linqing Wang, Kunjie Lin, Zixu Lin, Kaixiong Gong, Wenbo Li, Bin Lin, Zhenxi Li, Shiyi Zhang, Yuyang Peng, Wenxun Dai, Xinghao Ding, Chunyu Wang, Qinglin Lu*
548 Learning Latent Proxies for Controllable Single-Image Relighting, *Haoze Zheng, Zihao Wang, Xianfeng Wu, Yajing Bai, Yexin Liu, Yun Li, Xiaogang Xu, Harry Yang*
549 MoVie: Broaden Your Views with Human Motion for Action Detection, *Di Yang, Mahmoud Ali, Xuanlong Yu, Xi Shen, Quan Kong, Gianpiero Francesca, François Brémond*
550 MooCap: A Multi-View Benchmark for Cow-Object-Human Interaction and Behavior Dynamics, *Ian Noronha, Heather Neave, Upinder Kaur*
551 LAOF: Robust Latent Action Learning with Optical Flow Constraints, *Xizhou Bu, Jiexi Lyu, Fulei Sun, Ruichen Yang, Zhiqiang Ma, Wei Li*
552 DarkAct: A RGB-Thermal Dataset and Fusion Framework for Multimodal Low-Light Action Recognition, *Yuanjun Tan, Aoran Xiao, Liqian Deng, Zhigang Tu*
553 Random Wins All: Rethinking Grouping Strategies for Vision Tokens, *Qihang Fan, Yuang Ai, Huaibo Huang, Ran He*
554 Steering Where to Diffuse: Generative Modeling of Phenotypic Response Simulation with Steered Diffusion Bridge, *Rongchao Zhang, Chengxin Li, Yiwei Lou, Yuling Shi, Hanpin Wang, Yu Huang*
555 Deep Feature Deformation Weights, *Richard Liu, Itai Lang,*
❊ *Rana Hanocka*
556 Resolving Endpoint Underfitting in Diffusion Bridges via Noise Alignment, *Yurong Gao, Zicheng Zhang, Congying Han, Tiande Guo, Xinmin Qiu*
557 RNN as Linear Transformer: A Closer Investigation into Representational Potentials of Visual Mamba Models, *Timing Yang, Feng Wang, Guoyizhe Wei*
558 Coupling Liquid Time-Constant Encoders with Modern Hopfield Memory, *Bishal Ranjan Swain, Kyung Joo Cheoi, Jaepil Ko*
559 Stronger Normalization-Free Transformers, *Mingzhi Chen, Taiming Lu, Jiachen Zhu, Mingjie Sun, Zhuang Liu*
560 HCL-FF: Hierarchical and Contrastive Learning for Forward-Forward Algorithm, *Jie-En Yao, Hong-En Chen, C.-C. Jay Kuo*
561 Can You Learn to See Without Images? Procedural Warm-Up for Vision Transformers, *Zachary Shinnick, Liangze Jiang, Hemanth Saratchandran, Damien Teney, Anton van den Hengel*
562 Convolutional Neural Networks Driven by Content Similarity, *Ligeng Zou, Guihu Zhao*
563 MorphSeek: Fine-grained Latent Representation-Level Policy
❊ Optimization for Deformable Image Registration, *Runxun Zhang, Yizhou Liu, Dongrui Li, Bo Xu, Jingwei Wei*
564 HATS: Hardness-Aware Trajectory Synthesis for GUI Agents, *Rui Shao, Ruize Gao, Bin Xie, Yixing Li, Kaiwen Zhou, Shuai Wang, Weili Guan, Gongwei Chen*
565 MVP: Multiple View Prediction Improves GUI Grounding, *Yunzhu Zhang, Zeyu Pan, Zhengwen Zeng, Shuheng Shen, Changhua Meng, Linchao Zhu*
566 Towards GUI Agents: Vision-Language Diffusion Models for GUI Grounding, *Shrinidhi Kumbhar, Haofu Liao, Srikar Appalaraju, Kunwar Yashraj Singh*
567 ProactiveMobile: A Comprehensive Benchmark for Boosting Proactive Intelligence On Mobile Devices, *Dezhi Kong, Zhengzhao Feng, Qiliang Liang, Hao Wang, Haofei Sun, Changpeng Yang, Yang Li, Peng Zhou, Shuai Nie, Hongzhen Wang, Linfeng Zhou, Hao Jia, Jiaming Xu, Runyu Shi, Ying Huang*
568 OS-Oracle: A Comprehensive Framework for Cross-Platform GUI Critic Models, *Zhenyu Wu, Jingjing Xie, Zehao Li, Bowen Yang, Qiushi Sun, Zhaoyang Liu, Zhoumianze Liu, Yu Qiao, Xiangyu Yue, Zun Wang, Zichen Ding*
569 Training High-Level Schedulers with Execution-Feedback Reinforcement Learning for Long-Horizon GUI Automation, *Zehao Deng, Tianjie Ju, Zheng Wu, Zhuosheng Zhang, Gongshen Liu*
570 See, Think, Act: Teaching Multimodal Agents to Effectively Interact with GUI by Identifying Toggles, *Zongru Wu, Rui Mao,*

Zhiyuan Tian, Pengzhou Cheng, Tianjie Ju, Zheng Wu, Lingzhong Dong, Haiyue Sheng, Zhuosheng Zhang, Gongshen Liu

571 Beyond Weak Supervision: MLLMs-Guided Graded Knowledge Distillation for Unsupervised Camouflaged Object Detection, *Huafeng Chen, Chenguang Zhu, Yueming Lyu, Caifeng Shan*

572 Detecting Unknown Objects via Energy-based Separation for Open World Object Detection, *Jun-Woo Heo, Keonhee Park, Gyeong-Moon Park*

573 Beyond Prompt Degradation: Prototype-guided Dual-pool Prompting for Incremental Object Detection, *Yaoteng Zhang, Qing Zhou, Junyu Gao, Qi Wang*

574 SPAR: Single-Pass Any-Resolution ViT for Open-vocabulary Segmentation, *Naomi Kombol, Ivan Martinović, Siniša Šegvić, Giorgos Tolias*

575 TTL: Test-time Textual Learning for OOD Detection with Pretrained Vision-Language Models, *Jinlun Ye, Jiang Liao, Runhe Lai, Xinhua Lu, Jiaxin Zhuang, Zhiyong Gan, Ruixuan Wang*

576 Parameterized Prompt for Incremental Object Detection, *Zijia An, Boyu Diao, Ruiqi Liu, Libo Huang, Chuanguang Yang, Fei Wang, Zhulin An, Yongjun Xu*

577 SRA-Det: Learning Omni-Grained Open-Vocabulary Detection Beyond Category Names, *Li Yang, Boyu Cai, Wei Liu, Yan Wang, Chunfeng Yuan, Bing Li, Weiming Hu*

578 ❊ Retrieve and Segment: Are a Few Examples Enough to Bridge the Supervision Gap in Open-Vocabulary Segmentation?, *Tilemachos Aravanis, Vladan Stojnić, Bill Psomas, Nikos Komodakis, Giorgos Tolias*

579 PCA-Seg: Revisiting Cost Aggregation for Open-Vocabulary Semantic and Part Segmentation, *Jianjian Yin, Tao Chen, Yi Chen, Gensheng Pei, Xiangbo Shu, Yazhou Yao, Fumin Shen*

580 Partial Weakly-Supervised Oriented Object Detection, *Mingxin Liu, Peiyuan Zhang, Yuan Liu, Wei Zhang, Yue Zhou, Ning Liao, Ziyang Gong, Junwei Luo, Zhirui Wang, Yi Yu, Xue Yang*

581 Seeing Both Sides: Towards Bidirectional Semantic Alignment for Open-Vocabulary Camouflaged Object Segmentation, *Guohui Zhang, Fuming Sun, Yu Zhao, Yuqiu Kong, Jing Sun, Fasheng Wang*

582 Towards Robust Multi-Modal Semantic Segmentation with Teacher-Student Framework and Hybrid Prototype Distillation, *Jiaqi Tan, Xu Zheng, Yang Liu*

583 REL-SF4PASS: Panoramic Semantic Segmentation with REL Depth Representation and Spherical Fusion, *Xuewei Li, Xinghan Bao, Zhimin Chen, Xi Li*

584 Looking Beyond the Window: Global-Local Aligned CLIP for Training-free Open-Vocabulary Semantic Segmentation, *ByeongCheol Lee, Hyun Seok Seong, Sangeek Hyun, Gilhan Park, WonJun Moon, Jae-Pil Heo*

585 From Softmax to Dirichlet: Evidential Learning for Semi-supervised Semantic Segmentation, *Huayu Mai, Rui Sun, Yujia Chen, Wangkai Li, Bingzhou Wang, Aibing Li, Zhangyu He, Yuan Wang*

586 Particulate: Feed-Forward 3D Object Articulation, *Ruining Li, Yuxin Yao, Chuanxia Zheng, Christian Rupprecht, Joan Lasenby, Shangzhe Wu, Andrea Vedaldi*

587 HOPS: Hierarchical Open-vocabulary Part Segmentation with Attention-Aware Filtering and Affinity-Guided Enhancement, *Xinlong Li, Di Lin, Shaoyiyi Gao, Yaxuan Liu, Jixian He, Jiaxin Li, Ruonan Liu, Qing Guo, Kairui Yang, Wei Feng*

588 Shape-of-You: Fused Gromov-Wasserstein Optimal Transport for Semantic Correspondence in-the-Wild, *Jiin Im, Sisung Liu, Je Hyeong Hong*

589 MEMO: Human-like Crisp Edge Detection Using Masked Edge Prediction, *Jiaxin Cheng, Yue Wu, Yicong Zhou*

590 MUFASA: A Multi-Layer Framework for Slot Attention, *Sebastian Bock, Leonie Schüßler, Krishnakant Singh, Simone Schaub-Meyer, Stefan Roth*

591 ChangeBridge: Spatiotemporal Image Generation with Multimodal Controls for Remote Senisng, *Zhenghui Zhao, Chen Wu, Xiangyong Cao, Di Wang, Hongruixuan Chen, Datao Tang, Liangpei Zhang, Zhuo Zheng*

592 MOMO: Mars Orbital MOdel Foundation Model for Mars Orbital Applications, *Mirali Purohit, Bimal Gajera, Irish Mehta, Bhanu Tokas, Jacob Adler, Steven Lu, Scott Dickenshied, Serina Diniega, Brian Bue, Umaa Rebbapragada, Hannah Kerner*

593 Seeing Through the Noise: Improving Infrared Small Target Detection and Segmentation from Noise Suppression Perspective, *Maoxun Yuan, Duanni Meng, Ziteng Xi, Tianyi Zhao, Shiji Zhao, Yimian Dai, Xingxing Wei*

594 GeoBridge: A Semantic-Anchored Multi-View Foundation Model Bridging Images and Text for Geo-Localization, *Zixuan Song, Jing Zhang, Di Wang, Zidie Zhou, Wenbin Liu, Haonan Guo, En Wang, Bo Du*

595 GeoSANE: Learning Geospatial Representations from Models, Not Data, *Joëlle Hanna, Damian Falk, Stella X. Yu, Damian Borth*

596 ❊ Brewing Stronger Features: Dual-Teacher Distillation for Multispectral Earth Observation, *Filip Wolf, Blaž Rolih, Luka Čehovin Zajc*

597 Spectral Super-Resolution via Adversarial Unfolding and Data-Driven Spectrum Regularization: From Multispectral Satellite Data to NASA Hyperspectral Image, *Si-Sheng Young, Chia-Hsiang Lin*

598 RAMEN: Resolution-Adjustable Multimodal Encoder for Earth Observation, *Nicolas Houdré, Diego Marcos, Hugo Riffaud de Turckheim, Dino Ienco, Laurent Wendling, Camille Kurtz, Sylvain Lobry*

599 ORSATR-X: A Foundation Model based on Differential-and-Excitation Networks for Optical Remote Sensing Object Recognition, *Canyu Mo, Yongxiang Liu, Jiehua Zhang, Zilong Yu, Zhen Liu, Tianpeng Liu, Li Liu*

600 SEBA: Sample-Efficient Black-Box Attacks on Visual Reinforcement Learning, *Tairan Huang, Yulin Jin, Junxu Liu, Qingqing Ye, Haibo Hu*

601 IAG: Input-aware Backdoor Attack on VLM-based Visual Grounding, *Junxian Li, Beining Xu, Simin Chen, Jiatong Li, Jingdi Lei, Haodong Zhao, Di Zhang*

602 DASH: A Meta-Attack Framework for Synthesizing Effective and Stealthy Adversarial Examples, *Abdullah Al Nomaan Nafi, Habibur Rahaman, Zafaryab Haider, Tanzim Mahfuz, Fnu Suya, Swarup Bhunia, Prabuddha Chakraborty*

603 AdapAction: Adaptive Target Action Backdoor Attack against GUI Agents, *Baicheng Chen, Mingda Zhang, Min Zhang, Haizhou Li, Baoyuan Wu*

604 Phantom: Physical Object Interactions as Dynamic Triggers for NMS-Exploited Backdoors, *Tianlin Huo, Dongchuan Ran, Ranjie Duan, Yao Zhu, Peilun Du, Ningbo Yao, Huanqian Yan, Xu Han, Qiang Yun, Yuzheng Tan, Yang Bao, Yuan He*

605 ❊ Verifying Neural Network Robustness with Dual Perturbations, *Hai Duong, Lam Nguyen, Thanh Le, ThanhVu Nguyen*

606 Defending Unauthorized Model Merging via Dual-Stage Weight Protection, *Wei-Jia Chen, Min-Yan Tsai, Cheng-Yi Lee, Chia-Mu Yu*

607 AntiStyler: Defending Object Detection Models Against Adversarial Patch Attacks Using Style Removal, *Idan Yankelev, Edita Grolman, Yarin Yerushalmi Levi, Amit Giloni, Omer Hofman, Toshiya Shimizu, Yuval Elovici, Asaf Shabtai*

608 On the Role of Temporal Granularity in the Robustness of Spiking Neural Networks, *Mengting Xu, Shi Gu, Peng Lin, De Ma, Huajin Tang, Qian Zheng, Gang Pan*

609 ❊ Boosting Vision-Language-Action Finetuning with Feasible Action Neighborhood Prior, *Haochen Niu, Kanyu Zhang, Shuyu Yin, Qinghai Guo, Peilin Liu, Fei Wen*

610 Exploring Conditions for Diffusion Models in Robotic Control, *Heeseong Shin, Byeongho Heo, Dongyoon Han, Seungryong Kim, Taekyung Kim*

611 ❊ A Frame is Worth One Token: Efficient Generative World Modeling with Delta Tokens, *Tommie Kerssies, Gabriele Berton, Ju He, Qihang Yu, Wufei Ma, Daan de Geus, Gijs Dubbelman, Liang-Chieh Chen*

612 ❊ Efficient Hybrid SE(3)-Equivariant Visuomotor Flow Policy via Spherical Harmonics for Robot Manipulation, *Qinglun Zhang, Shen Cheng, Tian Dan, Haoqiang Fan, Guanghui Liu, Shuaicheng Liu*

613 TSTM: Temporal Segmentation for Task-relevant Mask in Visual Reinforcement Learning Generalization, *Weicheng Du, Wenjia Meng, Zhengzhe Zhang, Yilong Yin, Xiankai Lu*

614 Scaling Spatial and Temporal Context for Robotic Imitation Learning Policies With Scene Graphs, *Jianing Qian, Qinhe Peng, Emmanuel Panov, Leonor Fermoselle, Dinesh Jayaraman, Bernadette Bucher, Tarik Kelestemur*

615 AdaDexTrack: Dynamic Modulation for Adaptive and Generalizable Dexterous Manipulation Tracking, *Jianibieke Adalibieke, Qianwei Han, Xueyi Liu, Yuzhe Qin, Li Yi*

616 GraspLDP: Towards Generalizable Grasping Policy via Latent Diffusion, *Enda Xiang, Haoxiang Ma, Xinzhu Ma, Zicheng Liu, Di Huang*
617 MoEActok: A MoE-based Action Tokenizer for Vision-Language-Action Models, *Chunpu Xu, Zhixuan Liang, Tianshuo Yang, Chi-Min Chan, Yang Xiao, Jessie Wang, Xiaokang Yang, Yao Mu*
618 A Cross-view Fusion Framework for Robust 6-DoF Grasp Pose Estimation, *Kangjian Zhu, Haobo Jiang, Jianjun Qian, Jin Xie*
619 SAVA-X: Ego-to-Exo Imitation Error Detection via Scene-Adaptive View Alignment and Bidirectional Cross View Fusion, *Xiang Li, Heqian Qiu, Lanxiao Wang, Benliu Qiu, Fanman Meng, Linfeng Xu, Hongliang Li*
620 PromptDepth: Efficient and Promptable Geometric 3D Vision Model for Embodied Intelligence, *Xianyun Wang, Jiaxu Miao, Tian Xu, Siyuan Wang, Yuehao Li, Haoyang Hu, Jun Xiao, Yonghong Tian, Jun Yu*
621 Gallant: Voxel Grid-based Humanoid Locomotion and Local-
❊ navigation across 3-D Constrained Terrains, *Qingwei Ben, Botian Xu, Kailin Li, Feiyu Jia, Wentao Zhang, Jingping Wang, Jingbo Wang, Dahua Lin, Jiangmiao Pang*
622 PALM: Progress-Aware Policy Learning via Affordance Reasoning for Long-Horizon Robotic Manipulation, *Yuanzhe Liu, Jingyuan Zhu, Yuchen Mo, Gen Li, Xu Cao, Jin Jin, Yifan Shen, Zhengyuan Li, Tianjiao Yu, Wenzhen Yuan, Fangqiang Ding, Ismini Lourentzou*
623 IGen: Scalable Data Generation for Robot Learning from Open-World Images, *Chenghao Gu, Haolan Kang, Junchao Lin, Jinghe Wang, Duo Wu, Shuzhao Xie, Fanding Huang, Junchen Ge, Ziyang Gong, Letian Li, Hongying Zheng, Changwei Lv, Zhi Wang*
624 Hypergraph-State Collaborative Reasoning for Multi-Object Tracking, *Zikai Song, Junqing Yu, Yi-Ping Phoebe Chen, Wei Yang, Xinchao Wang*
625 TGTrack: Temporal Generative Learning for Unified Single Object Tracking, *Wanting Geng, Xin Chen, Chuanyu Sun, Jie Zhao, Ben Kang, Dong Wang, Huchuan Lu*
626 GeoMotion: Rethinking Motion Segmentation via Latent 4D Geometry, *Xiankang He, Peile Lin, Ying Cui, Dongyan Guo, Chunhua Shen, Xiaoqin Zhang*
627 Generalizable Structure-Aware Keypoint Correspondence for Category-Unified 3D Single Object Tracking, *Jie Xiao, Yinchao Ma, Yuyang Tang, Dengqing Yang, Jianpeng Yang, Xu Zhou, Qiao Li, Wenfei Yang, Tianzhu Zhang*
628 Generative Point Tracking and Forecasting, *Xuanchen Lu, Ang Cao, Chao Feng, Andrew Owens*
629 RAGTrack: Language-aware RGBT Tracking with Retrieval-Augmented Generation, *Hao Li, Yuhao Wang, Wenning Hao, Pingping Zhang, Dong Wang, Huchuan Lu*
630 Dual-level Adaptation for Multi-Object Tracking: Building Test-Time Calibration from Experience and Intuition, *Wen Guo, Pengfei Zhao, Zongmeng Wang, Yufan Hu, Junyu Gao*
631 GMT: Effective Global Framework for Multi-Target Multi-Camera Tracking, *Yihao Zhen, Mingyue Xu, Qiang Wang, Baojie Fan, Jiahua Dong, Tinghui Zhao, Huijie Fan*
632 Bridging Brain and Semantics: A Hierarchical Framework for Semantically Enhanced fMRI-to-Video Reconstruction, *Yujie Wei, Chenglong Ma, Jianxiong Gao, Chenhui Wang, Shiwei Zhang, Biao Gong, Shuai Tan, Hangjie Yuan, Hongming Shan*
633 GraPHFormer: A Multimodal Graph Persistent Homology Transformer for the Analysis of Neuroscience Morphologies, *Uzair Shah, Marco Agus, Mahmoud Gamal, Mahmood Alzubaidi, Corrado Cali, Pierre J. Magistretti, Abdesselam Bouzerdoum, Mowafa Househ*
634 DARC: Dual Adjustment Reasoning with Counterfactuals for Trustworthy Chest X-ray Classification, *Zhifang Liao, Junhao Li, HaoKang Ding, Yucheng Song*
635 Every Error has Its Magnitude: Asymmetric Mistake Severity Training for Multiclass Multiple Instance Learning, *Sungrae Hong, Jiwon Jeong, Jisu Shin, Donghee Han, Sol Lee, Kyungeun Kim, Mun Yong Yi*
636 Phrase-grounded APO for Improving Chest X-ray Report Generation, *Raziuddin Mahmood, Tanveer Syeda-Mahmood*
637 Focus-to-Perceive Representation Learning: A Cognition-Inspired
❊ Hierarchical Framework for Endoscopic Video Analysis, *Yuan Zhang, Sihao Dou, Kai Hu, Shuhua Deng, Chunhong Cao, Fen Xiao, Xieping Gao*
638 OraPO: Oracle-educated Reinforcement Learning for Data-efficient and Factual Radiology Report Generation, *Zhuoxiao Chen, Hongyang Yu, Ying Xu, Yadan Luo, Long Duong, Yuan-Fang Li*
639 FluoCLIP: Stain-Aware Focus Quality Assessment in Fluorescence
❊ Microscopy, *Hyejin Park, Jiwon Yoon, Sumin Park, Suree Kim, Sinae Jang, Eunsoo Lee, Dongmin Kang, Dongbo Min*
640 CryoKRAQEN: Kernel-Regularized Annealing for Quantized Embedding Networks in Cryo-EM Heterogeneous Reconstruction, *Wenyuan Gao, Yutan Wu, Xuming He*
641 Building Robust Vision Encoders for Cross-Dataset Evaluation in Immunofluorescent Microscopy, *Umar Marikkar, Syed Sameed Husain, Muhammad Awais, Sara Atito*
642 H2-Surv: Hierarchical Hyperbolic Multimodal Representation Learning for Survival Prediction, *Jiaqi Yang, Wenting Chen, Xiangjian He, Yuanbai Li, Sen Yang, Linlin Shen, Xiaohan Xing*
643 Dual-Level Hypergraph Generation for Addressing Feature Scarcity in Whole-Slide Image Classification, *Shuilian Yao, Qi Jia, Yu Liu, Pengshuo Zhang, Lili Sun, Weimin Wang, Yanmei Zhu, Bo Zhang, Xin Fan*
644 Temporal Inversion for Learning Interval Change in Chest X-Rays, *Hanbin Ko, Kyeongmin Jeon, Doowoong Choi, Chang Min Park*
645 JUMP-Hand: Learning Joint-wise Uncertainty to Gate Mixture of View Experts for Multi-View 3D Hand Reconstruction, *Haohong Kuang, Yang Xiao, Changlong Jiang, Jinghong Zheng, Hang Xu, Ran Wang, Zhiguo Cao, Joey Tianyi Zhou*
646 PAD-Hand: Physics-Aware Diffusion for Hand Motion
❊ Recovery, *Elkhan Ismayilzada, Yufei Zhang, Zijun Cui*
647 Anatomical Domain Shifts: Test-time Heterogeneous Adaptation for 3D Human Pose Prediction, *Qiongjie Cui, Pan Zhou, Jingjing Chen, Na Zhao*
648 Unlocking Motion from Large Vision Models with a Semantic and
❊ Kinematic Duality for Gait Recognition, *Zhanbo Huang, Dingqiang Ye, Xiaoming Liu, Yu Kong*
649 Learning 3D Shape Fidelity Metric from Real-world Distortions, *Xuelu Feng, Tianyu Luan, Zixin Zhu, Akshobhya Sharma, Phani Nuney, Junsong Yuan, Chunming Qiao*
650 BarbieGait: An Identity-Consistent Synthetic Human Dataset with Versatile Cloth-Changing for Gait Recognition, *Qingyuan Cai, Saihui Hou, Xuecai Hu, Yongzhen Huang*
651 FisherPoser: Human Motion Estimation from Sparse Observations with Hierarchical Region-Wise Fisher-Matrix Uncertainty Modeling, *Songpengcheng Xia, Qingyu Zhang, Zhuo Su, Jiarui Yang, Zengyuan Lai, Qi Wu, Ling Pei*
652 EmbodMocap: In-the-Wild 4D Human-Scene Reconstruction for Embodied Agents, *Wenjia Wang, Liang Pan, Huaijin Pi, Yuke Lou, Xuqian Ren, Yifan Wu, Zhouyingcheng Liao, Lei Yang, Rishabh Dabral, Christian Theobalt, Taku Komura*
653 Ground Reaction Inertial Poser: Physics-based Human Motion Capture from Sparse IMUs and Insole Pressure Sensors, *Ryosuke Hori, Jyun-Ting Song, Zhengyi Luo, Jinkun Cao, Soyong Shin, Hideo Saito, Kris Kitani*
654 FUN REC Reconstructing Functional 3D Scenes from Egocentric Interaction Videos, *Alexandros Delitzas, Chenyangguang Zhang, Alexey Gavryushin, Tommaso Di Mario, Boyang Sun, Rishabh Dabral, Leonidas Guibas, Christian Theobalt, Marc Pollefeys, Francis Engelmann, Daniel Barath*
655 VIMCAN: Visual-Inertial 3D Human Pose Estimation with Hybrid Mamba-Cross-Attention Network, *Zepeng Yang, Junxuan Bai, Hao Li, Ju Dai, Junjun Pan, Yongfeng Yin, Bin Li*
656 Bringing Your Portrait to 3D Presence, *Jiawei Zhang, Lei Chu, Jiahao Li, Zhenyu Zang, Chong Li, Xiao Li, Xun Cao, Hao Zhu, Yan Lu*
657 FLOW: Feature-Level Optimal Warping for Generalized Remote Physiological Measurement, *Bo Zhao, Junzhe Cao, Dan Guo, Dongmin Huang, Wenjin Wang, Tao Tan, Yue Sun, Zitong Yu*
658 One-to-More: High-Fidelity Training-Free Anomaly Generation with
❊ Attention Control, *Haoxiang Rao, Zhao Wang, Chenyang Si, Yan Lyu, Yuanyi Duan, Fang Zhao, Caifeng Shan*
659 UniMMAD: Unified Multi-Modal and Multi-Class Anomaly Detection via MoE-Driven Feature Decompression, *Yuan Zhao, Youwei Pang, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Huchuan Lu, Xiaoqi Zhao*

660 BUSSARD: Normalizing Flows for Bijective Universal Scene-Specific Anomalous Relationship Detection, *Melissa Schween, Mathis Kruse, Bodo Rosenhahn*
661 Multi-Prototype Compactness and Boundary-Aware Synthesis for Unsupervised Anomaly Detection, *Kailun Liao, Jianfeng Yang, Tao Tao, Wenfei Wu, Jiaming Jiang, Jinsheng Xiao*
662 PDD: Manifold-Prior Diverse Distillation for Medical Anomaly Detection, *Xijun Lu, Hongying Liu, Fanhua Shang, Yanming Hui, Liang Wan*
663 Weakly Supervised Video Anomaly Detection with Anomaly-Connected Components and Intention Reasoning, *Yu Wang, Shengjie Zhao*
664 SubspaceAD: Training-Free Few-Shot Anomaly Detection via Subspace Modeling, *Camile Lendering, Erkut Akdag, Egor Bondarau*
665 Learning Spatial-Temporal Consistency for 3D Semantic Scene Completion, *Yujie Xue, Meng Wang, Ruihui Li, Fan Wu, Zhizhong Liu, Zhuo Tang, Kenli Li*
666 Generalizing Visual Geometry Priors to Sparse Gaussian Occupancy Prediction, *Changqing Zhou, Yueru Luo, Changhao Chen*
667 Deformable Gaussian Occupancy: Decoupling Rigid and Nonrigid Motion with Factorized Distillation, *Yang Gao, Wuyang Li, Po-Chien Luan, Alexandre Alahi*
668 OccAny: Generalized Unconstrained Urban 3D Occupancy, *Anh-Quan Cao, Tuan-Hung Vu*
669 Dr.Occ: Depth- and Region-Guided 3D Occupancy from Surround-View Cameras for Autonomous Driving, *Xubo Zhu, Haoyang Zhang, Fei He, Rui Wu, Yanhu Shan, Wen Yang, Huai Yu*
670 ShelfOcc: Native 3D Supervision beyond LiDAR for Vision-Based Occupancy Estimation, *Simon Boeder, Fabian Gigengack, Simon Roesler, Holger Caesar, Benjamin Risse*

16:45 - 18:45 DEMOS

1 MonoDepth.zip: Zero-Shot Real-Time Monocular Depth on Mobile Devices, *Fabio Tosi, Luca Bartolomei, Matteo Poggi, Stefano Mattoccia*
2 DiffBMP: Differentiable Rendering with Bitmap Primitives, *Seongmin Hong, Junghun James Kim, Se Young Chun*
3 Occlusion Aware 3D Scene Control in Text-to-Image Generation, *Vaibhav Agrawal, Rishubh Parihar, Pradhaan S Bhat, Ravi Kiran S, Venkatesh Babu R*
4 RAVEN: Real-Time Adaptive Radar Perception on Edge Devices from Raw ADC Signals on a TI IWR1443BOOST Sensor, *Mir Sayeed Mohammad, Anuvab Sen*
5 AutoLabel Pro: A Real-Time Human-in-the-Loop System for Data-Centric Object Detection, *Lu Gan, Xi Li*
6 RADx: Hand X-Ray Rheumatoid Arthritis Severity Assessment Tool, *Ganlin Feng, Yuxi Long, Anna Liu, Liam O'Neil, Carol Hitchon, Pingzhao Hu*

Notes:

Sunday, June 7

7:30 - 13:00 **Registration / Badge Pickup** (Lobby A)
7:30 - 18:00 **Luggage Check** (Lobby A)
7:00 - 17:00 **Press Room** (ExHall F)
7:00 - 17:00 **Mother's Room** (Adjacent to Room 102)
7:00 - 17:00 **Quiet Room** (Adjacent to Room 102)
7:00 - 17:00 **Prayer Room** (Room 206)
7:30 - 9:00 **Findings Posters** (ExHall A)
7:30 - 9:00 **Breakfast** (ExHall C)
7:30 - 17:00 **Poster Pickup / T-shirt Pickup** (ExHall A & F)
9:00 - 10:15 **Oral Session 5A: Dynamic Perception** (Bluebird Ballroom)
Oral Session 5B: Generalization and Adaptation (Four Seasons Ballroom)
Oral Session 5C: Geometry and Robotics (Mile High Ballroom 1A - 2A)
Oral Session 5D: Human-Centric Modeling & Lighting (Mile High Ballroom 3A - 4A)
10:15 - 10:30 **Courtesy Break**
10:30 - 11:30 **KEYNOTE 3 - Thomas Serre, Professor of Science and Professor of Cognitive & Psychological Sciences; Scaling Laws vs. Neural Laws: Toward More Natural Artificial Vision** (Bluebird Ballroom)
11:15 - 11:45 **Poster Setup** (ExHall A)
11:45 - 13:45 **Poster Session 5 & Exhibit Hall** (ExHall F)
11:45 - 13:45 **DEMOS**
11:45 - 15:00 **Art Program & Exhibits** (ExHall EF)
11:45 **Art Gallery Tour with Curator, Luba Elliott** (ExHall E) 30 mins
11:45 - 13:30 **LUNCH** (ExHall C)
14:00 - 15:15 **Oral Session 6A: Geometric Learning** (Bluebird Ballroom)
Oral Session 6B: Multimodal Reasoning (Four Seasons Ballroom)
Oral Session 6C: Medical Vision (Mile High Ballroom 1A - 2A)
Oral Session 6D: Large-Scale Neural Modeling (Mile High Ballroom 3A - 4A)
15:15 - 15:30 **Courtesy Break**
15:00 - 15:30 **Poster Setup** (ExHall A)
15:30 - 17:30 **Poster Session 6** (ExHall A)

7:30 - 9:00 **Findings Posters** (ExHall A)

1 Advancing Open-Set Detection and Segmentation via Disentangled Representations, *Haokang Zhang, Yuchen Guan, Runxi Cheng, Yujiu Yang*
2 Disrupting Positional Encoding for Effective Open Set Recognition, *Yu Wang, Jiabo Xie, Yucan Zhou, Junxian Mu, Qinghua Hu, Pengfei Zhu*
3 ODOV: Benchmark the Open-Domain Open-Vocabulary Object Detection, *Yupeng Zhang, Ruize Han, Fangnan Zhou, Wei Feng, Liang Wan*
4 Leave No Stone Unturned: Uncovering Holistic Audio-Visual Intrinsic Coherence for Deepfake Detection, *Jielun Peng, Yabin Wang, Yaqi Li, Long Kong, Xiaopeng Hong*
5 Region-Aware Hierarchical Sub-Feature Alignment for Robust EEG-Based Visual Decoding, *Yanan Zhu, Ziwei Xiang, Jiamin Wu, Jinyang Guo, Hongyuan Zhang, Chunfeng Song, Hongjian Fang, Yufei Guo, Xianglong Liu*
6 Super Sparse DETR: YOLO-Competitive Convergence and Acceleration, *Hebao Zhu*
7 Bi-Level Optimization for Single Domain Generalization, *Marzi Heidari, Hanping Zhang, Hao Yan, Yuhong Guo*
8 SA-Matching DETR: A Lightweight Transformer Detector with Enhanced Scale Adaptive Matching, *Chengshan Yang, Pengnian Zhang, Jinjing Zhao*
9 Asymmetric Collaborative Distillation for Asymmetric Image Retrieval, *Yi Xie, Huaidong Zhang, Xuandi Luo, Yan Zhou, Shengfeng He*
10 OKGraph: Online Knowledge Graph Probing for Open-vocabulary Recognition, *Junhui Yin, Zhizhen Cai, Puze Wang, Guanzhou Ke, Jianhua Yang, Man Zhang, Qiang Zhang, Shengfeng He*
11 Large Multimodal Models as General In-Context Classifiers, *Marco Garosi, Matteo Farina, Alessandro Conti, Massimiliano Mancini, Elisa Ricci*
12 Indexing Multimodal Language Models for Large-scale Image Retrieval, *Bahey Tharwat, Giorgos Kordopatis-Zilos, Pavel Suma, Ian Reid, Giorgos Tolias*
13 EvoPrompt-ReID: A Bilevel Optimization Framework for Prompt-Encoder Co-evolution in Image Re-Identification, *Yuanlin He, Zhenchuan Wang, Jun Chen, Yingying He, Jiabao Wang, Weiwen Wang, Kun Xu, zijin zhou, Xiaoxiao Wang, Mingju chen, Tingting Liu, Zhisong Pan*
14 Leveraging Arbitrary Data Sources for AI-Generated Image Detection Without Sacrificing Generalization, *Qinghui He, Haifeng Zhang, Xiuli Bi, Bo Liu, Chi-Man Pun, Bin Xiao*
15 OmniGCD: Abstracting Generalized Category Discovery for Modality Agnosticism, *Jordan Shipard, Arnold Wiliem, Kien Nguyen Thanh, Wei Xiang, Clinton Fookes*
16 PTAD: Pose and Texture Agnostic Anomaly Detection, *Wei Zhuo, Jianen Xiang, Miaomiao Liu, Huajun Lu*
17 Mitigating the ID–OOD Tradeoff in Open-Set Test-Time Adaptation, *Wenjie Zhao, Jia Li, Xin Dong, Yapeng Tian, Yu Xiang, Yunhui Guo*
18 Towards Universal Open-Set Visual Font Recognition Via Augmented Synthetic Similarity, *Peicheng Zhou, Shancheng Fang, Chenhui Jin, Bowei Pu, Hongtao Xie*
19 VR-CLIP: Visual Refinement of CLIP for Zero-Shot Semantic Segmentation, *Haitao Jiang, Xu Li, Yuanyang Cao, Ying Zhang, Jianji Wang*
20 DynProto: Dynamic Prototype Evolution for Out-of-Distribution Detection, *Yanqi Wu, Xinhua Lu, Runhe Lai, Qichao Chen, Jia-Xin Zhuang, Wei-Shi Zheng, Ruixuan Wang*
21 Exploring Hierarchical Consistency and Unbiased Objectness for Open-Vocabulary Object Detection, *Sanghoon Lee, Geon Lee, Hyekang Park, Bumsub Ham*
22 Once for All: An End-to-End Paradigm for VLM-Based Domain-Generalized Object Detection, *Peng Zhang, Xiang Yuan, Cong Li, Junwei Han, Gong Cheng*
23 SoREL: Soft-Label Refurbishment with Ensemble Learning for Noisy Long-Tailed Classification, *Jun Wei Hsieh, Ying-Hsuan Wu, Yi-Kuan Hsieh, Xin Li, Kuan-Chuan Peng, Ming-Ching Chang*
24 Unsupervised Graph Partitioning Framework for Background Suppression in Multi-Query Vehicle Re-Identification, *Yichun Hu, Zixuan Hu, Ling-Yu Duan*
25 Revisiting Real-Time Detection Transformer with Efficient Encoder Design, *Jiannan Huang, Aditya Kane, Fengzhe Zhou, Yunchao Wei, Humphrey Shi*
26 PASR: Pose-Aware 3D Shape Retrieval from Occluded Single Views, *Jiaxin Shi, Guofeng Zhang, Wufei Ma, Naifu Liang, Adam Kortylewski, Alan Yuille*
27 Ninja Codes: Neurally Generated Fiducial Markers for Stealthy 6-DoF Tracking, *Yuichiro Takeuchi, Yusuke Imoto, Shunya Kato*
28 DetRefiner: Model-Agnostic Detection Refinement with Feature Fusion Transformer, *Soichiro Okazaki, Tatsuya Sasaki, Hiroki Ohashi*
29 SpHOR: A Representation Learning Perspective on Open-set Recognition for Identifying Unknown Classes in Deep Neural Networks, *Thiru Thillai Nadarasar Bahavan, Sachith Seneviratne, Saman Halgamuge*
30 Complexity of Linear Regions in Self-supervised Deep ReLU Networks, *Mufhumudzi Muthivhi, Terence L. van Zyl*
31 Decoupled Sub-Feature Uncertainty Modeling for Robust Multimodal Representation Learning, *Aoqiang Zhu, Min Hu, Yan Xing, Yiming Tang*
32 Pre-trained Models Can Count (Almost): Exploring Quantitative Structure in Visual Representations, *Toshimichi Aota, Akinori Hashimoto, Naoto Sekizuka, Takayuki Okatani*
33 A-SelecT: Automatic Timestep Selection for Diffusion Transformer Representation Learning, *Changyu Liu, James Chenhao Liang, Wenhao Yang, Yiming Cui, Jinghao Yang, Tianyang Wang, Qifan Wang, Dongfang Liu, Cheng Han*
34 HyperFM: A Efficient Hyperspectral Foundation Model with Spectral Grouping, *Zahid Hassan Tushar, Sanjay Purushotham*
35 Seeing Through Fog: Towards Fog-Invariant Action Recognition, *Enqi Liu, Liyuan Pan, Zhi Gao, Lingzhi Li, Qing Li*
36 Reversing the Flow: Generation-to-Understanding Synergy in Large Multimodal Models, *Yujun Tong, Dongliang Chang, Zijin Yin, Xintong Liu, Yuanchen Fang, Zhanyu Ma*

37 FedAR: Attribute-Guided Representation Learning for Heterogeneous Federated Learning, *Mengjie Li, Liu Yang, Qi Shen*
38 ZeroDiff++: Balancing Semantic Diffusion Dynamics for Robust Zero-Shot Learning, *Qin Li, Qi Li, Limei Liu, Junfeng Yang, Han Peng*
39 Equivariant Unsupervised Object Detection with Learnable Riesz Transform and Composite Spatial Transformers, *Sayan Kumar Chaki, Thierry Fournel, Rémi Emonet*
40 MART: Mechanism-disentanglement Anchor-Routed Training for Learning with Open-World Noisy Data, *Changhui Hu, Bhalaji Nagarajan, Ricardo Marques, Petia Radeva*
41 Online Interpretable Matrix Decomposition for Large-Scale Streaming Data, *Muhammad A. A. Abdelgawad, Abdelrahman B. M. Eldaly, Meng Xinmin, Peng Jing, Abdurrashid Ibrahim Sanka, Ray C.C. Cheung, Hong Yan*
42 Object-Centric Vision Token Pruning for Vision Language Models, *Guangyuan Li, Rongzhen Zhao, Jinhong Deng, Yanbo Wang, Joni Pajarinen*
43 BrainStack: Neuro-MoE with Functionally Guided Expert Routing for EEG-Based Language Decoding, *Ziyi Zhao, Jinzhao Zhou, Xiaowei Jiang, Beining Cao, Wenhao Ma, Yang Shen, Ren Li, Yu-Kai Wang, Chin-teng Lin*
44 BiomedHELIX : HiErarchical-Local Interaction eXploration for Biomedical Vision-Language Models, *Ziheng Zhu, Yuncheng Guo, Jie Xu, Xiaodong Gu*
45 From Fewer Samples to Fewer Bits: Reframing Dataset Distillation as Joint Optimization of Precision and Compactness, *My H. Dinh, Aditya Sant, Akshay Malhotra, Keya Patani, Shahab Hamidi-Rad*
46 Seeing Helps Reasoning in Language Models, *Yulu Gan, Kaiya Ivy Zhao, Tomaso Poggio, Phillip Isola*
47 Layer Embedding Deep Fusion Graph Neural Network, *Taihua Xu, Genhao Tian, Jicong Fan, Xibei Yang, Qinghua Zhang, Yun Cui*
48 From Horizontal to Rotated: Cross-View Object Geo-Localization with Orientation Awareness, *Chenlin Fu, Ao Gong, Xingtao Ling, Yingying Zhu*
49 LinkedOut: Linking World Knowledge Representation Out of Video LLM for Next-Generation Video Recommendation, *Haichao Zhang, Yao Lu, Lichen Wang, Yunzhe Li, Daiwei Chen, Yunpeng Xu, Yun Fu*
50 Learning to Reason: Targeted Knowledge Discovery and Fuzzy Logic Update for Robust Image Recognition, *Gurucharan Srinivas, Joshua Niemeijer, Frank Köster*
51 GaussFiller: Unleashing VLM-Expert Guidance for 3D Scene Completion with 3D Gaussian Splatting, *Yuhan Ping, Cheng Lin, Yuan Liu, Zhiyang Dou, Jia Pan, Wenping Wang*
52 GEODE: Geometry-Guided Discrete Diffusion for Open-Vocabulary 3D Scene Graph Generation, *Changqun Feng, Wangxiandi Yin, Xin Hu, Lei Zhao, Dongyang Zhang, Tao He*
53 Map2Thought: Explicit 3D Spatial Reasoning via Metric Cognitive Maps, *Xiangjun Gao, Zhensong Zhang, Dave Zhenyu Chen, Songcen Xu, Long Quan, Eduardo Pérez-Pellitero, Youngkyoon Jang*
54 SCP: Spatial Causal Prediction in Video, *Yanguang Zhao, Jie Yang, Shengqiong Wu, Shutong Hu, Hongbo Qiu, Yu Wang, Guijia Zhang, Tan Kai Ze, Hao Fei, Chia-Wen Lin, Mong-Li Lee, Wynne Hsu*
55 SpatialDreamer: Incentivizing Spatial Reasoning via Active Mental Imagery, *Meng Cao, Xingyu Li, Xue Liu, Ian Reid, Xiaodan Liang*
56 Entropy-Constrained Information Optimal Transport for Multi-View Geo-Localization, *Xiaoxi Yang, Bo Sun, Yisheng An, Ganchao Liu*
57 Revisiting Image Manipulation Localization under Realistic Manipulation Scenarios, *Xuekang Zhu, Ji-Zhe Zhou, Kaiwen Feng, Chenfan Qu, Xiwen Wang, Yunfei Wang, Liting Zhou, Jian Liu*
58 Learning to Wander: Improving the Global Image Geolocation Ability of LMMs via Actionable Reasoning, *Yushuo Zheng, Huiyu Duan, Zicheng Zhang, Xiaohong Liu, Xiongkuo Min*
59 CADRNet: Cognitively-Inspired Active Vision for 3D Reasoning Segmentation via Differentiable Rendering, *Zai Yang Yu, Changshuo Wang, Yuan Shi, Linjun Sun, Shu Wei, Tingran Wang, Wangyu Wu, Yanjie Li, Weijun Li*
60 Direct Language Embedding Enables Gaussian Splatting for Large Scenes, *Zhida Li, Jianqiao Zhu, Hejin Huang, Yipeng Qin, Sibei Yang, Guanbin Li*
61 CogNet: Multi-Agent Collaborative Reasoning and Verification for Salient Object Ranking, *Zhenyu Wu, Tengfei Shi, Xuehao Wang, Ming Li, Chenglizhao Chen, Wenfeng Song, Aimin Hao*
62 MPerS: Dynamic MLLM MixExperts Perception-Guided Remote Sensing Scene Segmentation, *Ziyi Wang, Xianping Ma, Ziyao Wang, Hongyang Zhang, Man On Pun*
63 Towards Generalization of Scene Text Tampering Localization via Causal Invariance, *Huiru Shao, Bin Dong, Kaizhu Huang, Xiaowei Huang, Qiufeng Wang*
64 Background-Compensated Audio-Visual Semantic Modulation Framework for Audio-Visual Event Localization, *Chao Sun, Junbo Zhang, Chuanbo Zhu, Mingjun Huang, Bo Du*
65 POMA-3D: The Point Map Way to 3D Scene Understanding, *Ye Mao, Weixun Luo, Ranran Huang, Junpeng Jing, Krystian Mikolajczyk*
66 Gazemo: Mimicking Human Saccades via Foveal-Peripheral Feature Modeling for Lightweight Semantic Segmentation, *Mian Muhammad Naeem Abid, Radu Timofte*
67 MoonSeg3R: Monocular Online Zero-Shot Segment Anything in 3D with Reconstructive Foundation Priors, *Zhipeng Du, Duolikun Danier, Jan Eric Lenssen, Hakan Bilen*
68 AuralSAM2: Enabling SAM2 Hear Through Pyramid Audio-Visual Feature Prompting, *Yuyuan Liu, Yuanhong Chen, Chong Wang, Junlin Han, Junde Wu, Can Peng, Jingkun Chen, Yu Tian, Gustavo Carneiro*
69 PrAda: Few-Shot Visual Adaptation for Text-Prompted Segmentation, *Gabriele Rosi, Fabio Cermelli, Carlo Masone, Barbara Caputo*
70 SAGE: Shape-Adapting Gated Experts for Adaptive Histopathology Image Segmentation, *Gia Huy Thai, Hoang-Nguyen Vu, Anh-Minh Phan, Quang-Thinh Ly, Thi-Ngoc-Truc Nguyen, Nhat Ho*
71 Prompt-driven Small Object Instance Segmentation in Earth Observation, *Chenhao Wang, Yingrui Ji, Yu Meng, Yunjian Zhang, Yao Zhu*
72 OV-Stitcher: A Global Context-Aware Framework for Training-Free Open Vocabulary Semantic Segmentation, *Seungjae Moon, Seunghyun Oh, Youngmin Ro*
73 SCOPE: Scene-Contextualized Incremental Few-Shot 3D Segmentation, *Vishal Thengane, Zhaochong An, Tianjin Huang, Son Lam Phung, Abdesselam Bouzerdoum, Lu Yin, Na Zhao, Xiatian Zhu*
74 Towards Complete Activation: Foreground-Background Multi-Perspective Guided Cross-Support for Few-Shot Segmentation, *Yi Yang, Qiang Jiao, Mengrui Shi, Qiang Zhang*
75 MHMamba: Multi-Head Mamba for 3D Brain Tumor Segmentation, *Hanjun Tao, Hua Wang, Fan Zhang*
76 ROSE: Retrieval-Oriented Segmentation Enhancement, *Song Tang, Guangquan Jie, Henghui Ding, Yu-Gang Jiang*
77 ConInfer: Context-Aware Inference for Training-Free Open-Vocabulary Remote Sensing Segmentation, *Wenyang Chen, Zhanxuan Hu, Yaping Zhang, Hailong Ning, Yonghang Tai*
78 Unify the Views: View-Consistent Prototype Learning for Few-Shot Segmentation, *Hongli Liu, Yu Wang, Shengjie Zhao*
79 Autoregressive Universal Video Segmentation Model, *Miran Heo, Sukjun Hwang, Min-Hung Chen, Yu-Chiang Frank Wang, Albert Gu, Seon Joo Kim, Ryo Hachiuma*
80 FCL-COD: Weakly Supervised Camouflaged Object Detection with Frequency-aware and Contrastive Learning, *Jingchen Ni, Quan Zhang, Dan Jiang, Keyu Lv, Ke Zhang, Chun Yuan*
81 Counterfactual Segmentation Reasoning: Diagnosing and Mitigating Pixel-Grounding Hallucination, *Xinzhuo Li, Adheesh Juvekar, Jiaxun Zhang, Xingyou Liu, Muntasir Wahed, Kiet A. Nguyen, Yifan Shen, Tianjiao Yu, Ismini Lourentzou*
82 Weakly-Supervised Referring Video Object Segmentation Through Text Supervision, *Miaojing Shi, Jun Huang, Zijie Yue, Hanli Wang*
83 TALENT: Target-Aware Efficient Tuning for Referring Image Segmentation, *Shuo Jin, Siyue Yu, Bingfeng Zhang, Chao Yao, Meiqin Liu, Jimin Xiao*
84 DeepDP-TGMM: Amortized Non-Parametric Clustering for Hyperspherical Self-Supervised Representations, *Cyril Kana Tepakbong, Kévin Bouchard, Julien Maitre*
85 Proto-SaGa: Prototype-based 3D Scene Segmentation with Semantic-aware Gaussian Grouping, *Youngmin Oh, Changjae Oh, Bumsub Ham*
86 RecycleLoRA: Rank-Revealing QR-Based Dual-LoRA Subspace Adaptation for Domain Generalized Semantic Segmentation, *Chanseul Cho, Seokju Yun, Jaesung Jun, Seungjae Moon, Youngmin Ro*

87 Instruction-Focus-Prompt: Semantics--Driven Structural Prompts for Universal SAM Segmentation, *Shuqi Xia, Guangze Shi, Jiarui Cao, Aoyuan Shi, Meilin Liu, Xiaoyi Zhang, Yujie Wang, Xueyu Liu, Cai Zhao, Ziyuan He, Yongfei Wu, Mingqiang Wei*
88 Continual Alignment for SAM: Rethinking Foundation Models for Medical Image Segmentation in Continual Learning, *Jiayi Wang, Wei Dai, Haoyu Wang, Sihan Yang, Haixia Bi, Jian Sun*
89 VirPro: Visual-Referred Probabilistic Prompt Learning for Weakly-Supervised Monocular 3D Detection, *Chupeng Liu, Jiyong Rao, Shangquan Sun, Runkai Zhao, Weidong Cai*
90 A Single Pixel is All You Need: Weakly Supervised Medical Image Segmentation using Discrete Denoising Diffusion Models, *Mehmet Demirel, Christos Kyrkou*
91 AdaMeta: Adaptive Meta-Learning with Dynamic Task Relational Inference for Few-shot learning, *Xingyu Yang, Yidan Ma, Hanzhang Qu, Jianfu Cao*
92 NRFP: A Noise-Robust Feature Plugin for Source-Free Domain Adaptation, *Huanxin Zou, Zhize Wu, Yue Jiang, Jijian Zhou, Zhiwei Xu, Teng Li, Jianhua Shu, Fan Cheng*
93 Label-Agnostic Category Discovery, *Yuwei Bian, Shidong Wang, Chunming Li, Haofeng Zhang*
94 Learning from Label Proportion with Dual-Proportion Constraints, *Tianhao Ma, Ximing Li, Changchun Li, Renchu Guan*
95 Test-Time Distillation for Continual Model Adaptation, *Xiao Chen, Jiazhen Huang, Zhiming Liu, Qinting Jiang, Fanding Huang, Jingyan Jiang, Zhi Wang*
96 Another BRIXEL in the Wall: Towards Cheaper Dense Features, *Alexander Lappe, Martin A. Giese*
97 Task-Specific Knowledge Improves Generalization: A Logits-Based Framework for Continual Learning of Vision-Language Models, *Sijie Wang, Yingying Zhu*
98 DARN: Dynamic Adaptive Regularization Networks for Efficient and Robust Foundation Model Adaptation, *Dhenenjay Yadav, Rohan Sawai*
99 Training-Free Uncertainty-guided Logit Adjustment for Few-Shot Class-Incremental Learning, *Sungwon Woo, Dongjun Hwang, Shiwon Kim, Junsuk Choe, Jongho Nang*
100 Model Merging on Loss Landscapes: A Geometric Perspective, *Juanwu Lu, Anand Bhaskar, Brian Axelrod, Ekaterina Tolstaya, Tristan Emrich*
101 DGD: Density Gradient-guided Diffusion for Long-Tailed Clustering, *Xulun Ye, Yuanyuan Deng, Kun Zhou*
102 DGP: Dynamic Gradient Projection for Task-Adaptive Continual Learning, *Qier Meng, Cheng Deng*
103 Bootstrap Your Own Classifier: Your Pretrained Vision Models are Secretly Strong Continual Learners, *Yizheng Gong, Xiaoyang Wang, Siyue Yu, Waleed Al-Nuaimy, Jimin Xiao*
104 Memory-efficient Continual Learning with Prototypical Exemplar Condensation, *M.-Duong Nguyen, Thien-Thanh Dao, Le-Tuan Nguyen, Dung D. Le, Kok-Seng Wong*
105 Continual Adaptation of Vision Foundational Models for Semantic Segmentation in Adverse Weather, *Nikhil Kumar Jangamreddy, Mahsa Baktashmotlagh, Chetan Arora*
106 ReMem: A Dynamic Memory Evolution Detector for Zero-Shot Anomaly Detection, *Ling Yi, Zhe Chen, Gaochang Wu, Jinliang Ding, Xiaojie Wang, Zhaolong Ning*
107 CurrMix: Curriculum-Enhanced MixUp for Long-Tailed Visual Recognition, *Zhongquan Jian, Yanhao Chen, Bingbing Hu, Wenhan Lv, Shaopan Wang, Jipeng Wu, Junfeng Yao, Yang Lu, Qingqiang Wu*
108 Class-Aware Drift Compensation for Non-Uniform Semantic Shift in Continual Learning, *Fankang Xu, Lu Jin, Yanpeng Sun, Shiyu Xuan, Zechao Li*
109 Onboarding Without Forgetting: Hypernetwork Personalization with Data-Free Replay for Personalized Federated Learning, *Thinh Nguyen, Le Huy Khiem, Van-Tuan Tran, Khoa D Doan, Nitesh V. Chawla, Kok-Seng Wong*
110 FedNPC: Stochastic Noise-driven Post-hoc Classifier Calibration Method for Federated Long-tailed Learning, *Jintong Gao, He Zhao, Yibo Yang, Dandan Guo*
111 Learning Multi-Modal Prototypes for Cross-Domain Few-Shot Object Detection, *Wanqi Wang, Jingcai Guo, Yuxiang Cai, Zhi Chen*
112 MuSCM: Mutual Spatial Correlation Mapping for Class Incremental Detection Transformer, *Jian Zhong, Yifan Jiao, Xi Shao, Bing-Kun Bao*
113 AFCL: Achieving Spatio-Temporal Invariance to Data Heterogeneity in Federated Continual Learning, *Jianheng Tang, Jingyu He, Kejia Fan, Run He, Jingchao Wang, Anfeng Liu, Houbing Herbert Song, Leye Wang, Zhanxing Zhu, Huiping Zhuang, Yunhuai Liu*
114 SAGA: Semantic Anchor-Guided Alignment for Multi-Source Domain Adaptive Object Detection, *Yongchao Feng, Ziyue Huang, Jinqing Zhang, Wenrui Cai, Qingjie Liu*
115 DEED: Dual-Channel Enhanced Ensemble Distillation for Uncertainty-Aware Recognition, *Yang Yang, Kai Xu, Junyao Hou, Miao Zhang, Xiang Li, Zhenghua Chen, Yingxue Gao, Min Wu*
116 Wake the Sleeping Weights: Sparsely-Activated Continual Test-Time Adaptation for Medical Image Segmentation, *Jianhang Ji, Zhiming Cheng, Jianxiang Zhao, Bingtao Ma, Hao Chen, Yuhan Gao, Lian Zhang, Zuobin Ying, Shuai Wang*
117 Dynamic Pseudo-Label Assignment and Consistent Prototypical Learning for Few-Shot Class-Incremental Learning, *Zhilong Mao, Hang Zhang, Yanmin Li, Lihua Liu, Jibing Wu, Mao Wang*
118 Hold-One-Shot-Out (HOSO) for Validation-Free Few-Shot CLIP Adapters, *Chris Vorster, Mayug Maniparambil, Noel O'Connor, Noel Murphy, Derek Molloy*
119 Learning through Creation: A Hash-Free Framework for On-the-Fly Category Discovery, *Bohan Zhang, Weidong Tang, Zhixiang Chi, Yi Jin, Zhenbo Li, Yang Wang, Yanan Wu*
120 Frequency-Guided Iterative Bi-directional Exchange Network for Cross-Domain Few-Shot Segmentation, *Yadang Chen, Qi Liu, Guoqing Zhang, Le Sun, Yuhui Zheng*
121 Revisiting Prototype Rehearsal for Exemplar-Free Continual Learning: Manifold-Aware Boundary Sampling with Adaptive Class-Balanced Loss, *Hongye Xu, Bartosz Krawczyk*
122 SCOPE: Spatially Ordered Continual Learning for 3D Segmentation, *Wenhao Xu, Huaidong Zhang, Weipeng Zhang, Qianle Zhang, Shengfeng He*
123 Learning to Propose Pose for Category-Agnostic Objects via Joint Refinement with Co-Matching Supervision, *Junjie Chen, Zezheng Liu, Runxiang Liu, Yuming Fang, Yifan Zuo, Jiebin Yan*
124 Is Prompt Selection Necessary for Task-Free Online Continual Learning?, *Seoyoung Park, Haemin Lee, Hankook Lee*
125 ReConText3D: Replay-based Continual Text-to-3D Generation, *Muhammad Ahmed Ullah Khan, Muhammad Haris Bin Amir, Didier Stricker, Muhammad Zeshan Afzal*
126 Now You See It, Now You Don't: Instant Concept Erasure for Safe Text-to-Image and Video Generation, *Shristi Das Biswas, Arani Roy, Kaushik Roy*
127 ECOC-IL: Robust and Efficient Label LDP for Imbalanced Learning, *Mengyang Li, Ou Wu*
128 Safe Codebook: Token-Level Moderation for Safer Visual Autoregressive Generation, *Jiaxuan Zhang, Qianqian Xu, Peisong Wen, Siran Dai, Yang Liu, Qingming Huang*
129 Towards Universal and Lightweight Coverless Image Steganography with Multimodal Large Language Models Assistance, *Jia Li, Zhankai Li, Yongqiang Yu, Xuehu Yan, Yuliang Lu*
130 A Visual Semantic Adaptive Watermark Grounded by Prefix-Tuning for Large Vision-Language Model, *Qi Zheng, Shuliang Liu, Yu Huang, Sihang Jia, Jungang Li, Lyuhao Chen, Junhao Chen, Hanqian Li, Aiwei Liu, Yibo Yan, Xuming Hu*
131 TriGuard-FL: A User-Centric Trust Triad in Federated Learning via Auditable Data, Verifiable Contributions, and Antidote-Driven Mitigation, *K Naveen Kumar, Mohsen Guizani*
132 Assessing the Reliability of Image Quality Metrics and Mitigating Quality Bias in Generative Models, *Hoin Jung, Shenyu Lu, De Wang, Xiaoqian Wang*
133 Efficient Unlearning through Maximizing Relearning Convergence Delay, *Khoa Tran, Simon S. Woo*
134 Robust Continual Unlearning against Knowledge Erosion and Forgetting Reversal, *Eun-Ju Park, Youjin Shin, Simon S. Woo*
135 Memorization In Stable Diffusion Is Unexpectedly Driven by CLIP Embeddings, *Bumjun Kim, Albert No*
136 RAZOR: Ratio-Aware Layer Editing for Targeted Unlearning in Vision

Transformers and Diffusion Models, *Ravi Ranjan, Utkarsh Grover, Xiaomin Lin, Agoritsa Polyzou*
137 FedOrtho: Efficient Federated Unlearning Via Orthogonal Convolution and Adaptive Soft Pruning, *Qinghui Gong, Xue Yang, Xunlei Chen, Jinshan Lai, Hua Meng, Xiaohu Tang*
138 Improving Synthesized Image Detection by Disentangling Generator-Shared and Generator-Specific Image Artifacts, *Yongqi Yang, Yuke Li, Heng Huang, Zhihui Li, Bo Du, Yu Wu*
139 PLR-Gate: Real-Time Gradient Privacy Assessment and Gated Transmission for Secure Federated Learning, *Tao Huang, Jiayang Meng, Hong Chen, Chen Hou, Guolong Zheng, Xu Yang*
140 A Unified Privacy-Utility Framework for Collaborative Inference via Randomized Smoothing, *Shiwei Ding, Lan Zhang, Zhenlin Wang, Xiaoyong Yuan*
141 Verify Claimed Text-to-Image Models Via Boundary-Aware Prompt Optimization, *Zidong Zhao, Yihao Huang, Qing Guo, Tianlin Li, Anran Li, Kailong Wang, Jin Song Dong, Geguang Pu*
142 Towards Robust Content Watermarking Against Removal and Forgery Attacks, *Yifan Zhu, Yihan Wang, Xiao-Shan Gao*
143 Revisiting Model Inversion Evaluation: From Misleading Standards to Reliable Privacy Assessment, *Sy-Tuyen Ho, Koh Jun Hao, Ngoc-Bao Nguyen, Alexander Binder, Ngai-Man Cheung*
144 CBDC: Clean Bias Direction Construction for Unsupervised Debiasing in Vision-Language Models, *DoYoung Kim, SungJoon Hwang, Byung-Joon Lee, Joohyeon Lee, Jee-Hyong Lee*
145 Erased, But Not Forgotten: Erased Rectified Flow Transformers Still Remain Unsafe Under Concept Attack, *Nanxiang Jiang, Zhaoxin Fan, Enhan Kang, Daiheng Gao, Yun Zhou, Yanxia Chang, Zheng Zhu, Yeying Jin, Wenjun Wu*
146 Leveraging Unlabeled Data from Unknown Sources via Dual-Path Guidance for Deepfake Face Detection, *Zhiqiang Yang, Renshuai Tao, Chunjie Zhang, Guodong Yang, Xiaolong Zheng, Yao Zhao*
147 SEM: Sparse Embedding Modulation for Post-Hoc Debiasing of Vision-Language Models, *Quentin Guimard, Federico Bartsch, Simone Caldarella, Rahaf Aljundi, Elisa Ricci, Massimiliano Mancini*
148 When Agents Steer Human Perception: How AI-Selected Images Can Convertly Alter Disagreements, *Chi Zhang, Yulang Gao, Jiachen Zou, Chen Wei, Quanying Liu*
149 UniShield: An Adaptive Multi-Agent Framework for Unified Forgery Image Detection and Localization, *Qing Huang, Zhipei Xu, Xuanyu Zhang, Xiangyu Yu, Jian Zhang*
150 On the Group Disparities Arising from Machine Unlearning, *Zijie Pan, Zuobin Ying, Yajie Wang, Liehuang Zhu, Wanlei Zhou*
151 Count What Repeats: Period-Adaptive Multi-Scale Consistency for Self-Supervised Repetitive Action Counting, *Shizhao Gao, Jun Li, Qiming Li*
152 Taming Hallucinations: Boosting MLLMs' Video Understanding via Counterfactual Video Generation, *Zhe Huang, Hao Wen, Aiming Hao, Bingze Song, Meiqi Wu, Jiahong Wu, Xiangxiang Chu, Sheng Lu, Haoqian Wang*
153 ConfDiff: Confidence-Guided Representation Diffusion for Video Moment Retrieval, *Haiming Zhao, Tai Wang*
154 Evolutionary Multi-Agent Collaboration for Real-World Video Face Restoration, *Bowen Tang, Tao Wang, Miao Zhang, Xin Yu, Jinwei Chen, Bo Li, Kaihao Zhang*
155 STS-Mixer: Spatio-Temporal-Spectral Mixer for 4D Point Cloud Video Understanding, *Wenhao Li, Xueying Jiang, Gongjie Zhang, Xiaoqin Zhang, Ling Shao, Shijian Lu*
156 HiVid-Narrator: Hierarchical Video Narrative Generation with Scene-Primed ASR-anchored Compression, *Haoxuan Li, Mengyan Li, Junjun Zheng*
157 D^2-STX: Decoupling Spatial-Temporal Cross-Attention for Dual-branch Repetitive Action Counting, *Xiaoai Wang, Hang Wang, Yan Liu, Huan Hu, Bruce X.B. Yu*
158 Group-DINOmics: Incorporating People Dynamics into DINO for Self-supervised Group Activity Feature Learning, *Ryuki Tezuka, Chihiro Nakatani, Norimichi Ukita*
159 Mamba-VMR: Multimodal Query Augmentation Via Generated Videos for Precise Temporal Grounding, *Yunzhuo Sun, Xinyue Liu, Yanyang Li, Nanding Wu, Linlin Zong, Xianchao Zhang, Wenxin Liang*
160 VideoThinker: Building Agentic VideoLLMs with LLM-Guided Tool Reasoning, *Chenglin Li, Qianglong Chen, Feng Han, Yikun Wang, Xingxi Yin, Yan Gong, Ruilin Li, Yin Zhang, Jiaqi Wang*
161 TP^2-DETR: Unlocking Deformable DETR for Zero-Shot Temporal Action Proposal Generation with Temporal Feature Pyramids, *Ya-Yun Cheng, Kan Tippayamontri, Chih-Yuan Yang, Jane Yung-jen Hsu*
162 QENN: A Quantum Entanglement-Inspired Neural Network for Interaction and Relationship Prediction in Story Videos, *Zijun Xu, Zhengqian Wu, Chunjie Zhang, Zhongyuan Wang, Chunxia Xiao, Chao Liang*
163 FineGrade: A Rule-Consistent Scoring Framework for Fine-Grained Action Quality Assessment, *Yicong Li, Howard Leung*
164 One Identity, Many Roles: Multimodal Entity Coreference for Enhanced Video Situation Recognition, *Balaji Darur, Amanmeet Garg, Makarand Tapaswi*
165 REBA: Residual Mixture-of-Experts and Bidirectional Video–Text Alignment for Better Fine-grained Weakly Supervised Video Anomaly Detection, *Chengxi Chu, Nurul Japar, Chee Kau Lim*
166 ReFoCUS: Reinforcement-guided Frame Optimization for Contextual Understanding, *Hosu Lee, Junho Kim, Hyunjun Kim, Yong Man Ro*
167 VIDEOP2R: Video Understanding from Perception to Reasoning, *Yifan Jiang, Yueying Wang, Rui Zhao, Toufiq Parag, Zhimin Chen, Zhenyu Liao, Jayakrishnan Unnikrishnan*
168 Video-R4: Reinforcing Text-Rich Video Reasoning with Visual Rumination, *Yolo Yunlong Tang, Daiki Shimada, Hang Hua, Chao Huang, Jing Bi, Rogerio Feris, Chenliang Xu*
169 ForestPrune: High-ratio Visual Token Compression for Video Multimodal Large Language Models Via Spatial-Temporal Forest Modeling, *Shaobo Ju, Baiyang Song, Tao Chen, Jiapeng Zhang, Qiong Wu, Chao Chang, Huaixi Wang, Yiyi Zhou, Rongrong Ji*
170 HARP: Hierarchical Adaptive Ranking with Probabilistic Modeling for Skill Determination, *Hui Yu, Xiao Ke, Zhihong Zeng, Huangbiao Xu, Huanqi Wu*
171 STORM: End-to-End Referring Multi-Object Tracking in Videos, *Zijia Lu, Jingru Yi, Jue Wang, Yuxiao Chen, Junwen Chen, Xinyu Li, Davide Modolo*
172 Extending Segment Anything Model 2 to Multi-Object Tracking by Optimizing Hierarchical Trajectory Memory, *Cheng-Yen Yang, Hsiang-Wei Huang, Kuang-Ming Chen, Kunjun Li, Jenq-Neng Hwang*
173 NCSTR: Node-Centric Decoupled Spatio-Temporal Reasoning for Video-based Human Pose Estimation, *Quang Dang Huynh, Xuefei Yin, Andrew Busch, Hugo G. Espinosa, Alan Wee-Chung Liew, Matthew T.O. Worsey, Yanming Zhu*
174 MOSSTrack : Modality-Specific Spatio-Temporal Context Learning for RGB-T Tracking, *Yisong Liu, He Yao, Junlong Cheng, Yujie Lu, Junqi Bai, Min Zhu*
175 Temporally Consistent Long-Term Memory for 3D Single Object Tracking, *Jaejoon Yoo, SuBeen Lee, Yerim Jeon, Miso Lee, Jae-Pil Heo*
176 DM^3T: Harmonizing Modalities via Diffusion for Multi-Object Tracking, *Weiran Li, Yeqiang Liu, Yijie Wei, Mina Han, Qiannan Guo, Zhenbo Li*
177 IRDINO: Adapting DINOv3 with Second-Order Motion Awareness for Moving Infrared Small Target Detection, *Qian Xu, Shuaipeng Fan, Fei Gao, Mingjin Zhang*
178 SemanticMoments: Training-Free Motion Similarity via Third Moment Features, *Saar Huberman, Kfir Goldberg, Or Patashnik, Sagie Benaim, Ron Mokady*
179 TAPNext++: What's Next for Tracking Any Point (TAP)?, *Sebastian Jung, Artem Zholus, Martin Sundermeyer, Carl Doersch, Ross Goroshin, David Joseph Tan, Sarath Chandar, Rudolph Triebel, Federico Tombari*
180 ARGS: Auto-Regressive Gaussian Splatting via Parallel Progressive Next-Scale Prediction, *Quanyuan Ruan, Kewei Shi, Jiabao Lei, Xifeng Gao, Xiaoguang Han*
181 100Editor: 100+ Views per Batch and Minute-Scale View-Consistent 3D Editing, *Cunqi Wu, Peng Zhou, Jie Qin, Qi Tian*
182 DIAMOND-SSS: Diffusion-Augmented Multi-View Optimization for Data-efficient SubSurface Scattering, *Guillermo Figueroa Araneda, Iris Dania Jimenez, Florian Hofherr, Manny Ko, Hector Andrade-Loarca, Daniel Cremers*

183 Reason-SVG: Enhancing Structured Reasoning for Vector Graphics Generation with Reinforcement Learning, *Ximing Xing, Ziteng Xue, Yandong Guan, Jing Zhang, Dong Xu, Qian Yu*
184 Harmonized Multi-Layer Text-to-Image Generation with Generative Priors, *Yusuf Dalva, Yijun Li, Qing Liu, Nanxuan Zhao, Jianming Zhang, Zhe Lin, Pinar Yanardag*
185 StabiGS: Video Stabilization through Rendering-Aware Trajectory Optimization in 3DGS-Reconstructed Scenes, *Souheib Ben Mabrouk, Jean-Emmanuel Deschaud, Eva Coupeté, Thomas Derbanne, Nicolas Rahmouni*
186 More Traces Better: Unified Artifact Modeling for Generalizable and Robust AI-generated Image Detection, *Ruiqi Liu, Xiaolei Lv, Zhiyuan Yan, Yi Han, Boyi Sun, Bo Li, Jun Gao, Lubin Weng, Yan Wang, Shu Wu*
187 Predicting Gene Expression in Spatially Resolved Transcriptomics Across Samples Through Probabilistic Fusion of Hierarchical Histology and Spatial Information, *Yinbo Liu, Qi Wu, Keyang Ye, Xiao He, Tian Tian*
188 Don't Let the Information Slip Away, *Taozhe Li, Guansu Wang, Bo Yu, Yiming Liu, Wei Sun*
189 FraQAT: Quantization Aware Training with Fractional Bits, *Luca Morreale, Alberto Gil C P Ramos, Malcolm Chadwick, Mehdi Noroozi, Ruchika Chavhan, Abhinav Mehrotra*
190 MAGIC: Few-Shot Mask-Guided Anomaly Inpainting with Prompt Perturbation, Spatially Adaptive Guidance, and Context Awareness, *JaeHyuck Choi, Minjun Kim, Je Hyeong Hong*
191 Video Inspector: An Agentic-RL Framework and Benchmark for Human-Aligned Generative Video Evaluation, *Jacey Somers, Harrison Zale, Janine Mason, Tina Walker, Eddie Quinn, Felix Lewis, Gavin Wright, Yvonne Young, Charles Sullivan, Wayne Carter, Julian Foster*
192 From Pixels to Nucleotides: End-to-End Token-Based Video Compression for DNA Storage, *Cihan Ruan, Lebin Zhou, Bingqing Zhao, Rongduo Han, Qiming Yuan, Chenchen Zhu, Linyi Han, Liang Yang, Wei Wang, Wei Jiang, Nam Ling*
193 CoPS: Conditional Prompt Synthesis for Zero-Shot Anomaly Detection, *Qiyu Chen, Zhen Qu, Wei Luo, Haiming Yao, Yunkang Cao, Yuxin Jiang, Yinan Duan, Huiyuan Luo, Chengkan Lv, Zhengtao Zhang*
194 PSIM: Perceptual Similarity Index Measure, *Md Eimran Hossain Eimon, Hari Kalva*
195 UI-AGILE: Advancing GUI Agents with Effective Reinforcement Learning and Precise Inference-Time Grounding, *Shuquan Lian, Yuhang Wu, Jia Ma, Yifan Ding, Zihan Song, Bingqi Chen, Xiawu Zheng, Hui Li, Rongrong Ji*
196 Beyond Static Artifacts: A Forensic Benchmark for Video Deepfake Reasoning in Vision Language Models, *Zheyuan Gu, Qingsong Zhao, Yusong Wang, Zhaohong Huang, Xinqi Li, Chen Yuan, Jiawei Shao, Chi Zhang, Xuelong Li*
197 GreenPlanner: Practical Floorplan Layout Generation via an Energy-Aware and Function-Feasible Generative Framework, *Pengyu Zeng, Yuqin Dai, Jun Yin, Jing Zhong, Ziyang Han, Chaoyang Shi, ZhanXiang Jin, Maowei Jiang, Yuxing Han, Shuai Lu*
198 Dual-Stage Parameter-Efficient Fine-Tuning for Consistent Spatial and Temporal Representation, *Junhao Xia, Chaoyang Zhang, Yecheng Zhang, Chengyang Zhou, Zhichang Wang, Bochun Liu, Dongshuo Yin*
199 WideEye: Achieving Wide Field-of-view Traffic Video Analytics With Dynamic Orientation Adaptation, *Z. Jonny Kong, Sibendu Paul, Y. Charlie Hu*
200 Restore-R1: Efficient Image Restoration Agents via Reinforcement Learning with Multimodal LLM Perceptual Feedback, *Jianglin Lu, Yuanwei Wu, Ziyi Zhao, Hongcheng Wang, Felix Jimenez, Abrar Majeedi, Yun Fu*
201 Pose-dIVE: Pose-Diversified Augmentation for Person Re-Identification, *Inès Hyeonsu Kim, Woojeong Jin, Soowon Son, Junyoung Seo, Seokju Cho, JeongYeol Baek, Byeongwon Lee, JoungBin Lee, Seungryong Kim*
202 Adaptive Reinforcement for Open-ended Medical Reasoning via Semantic-Guided Reward Collapse Mitigation, *Yizhou Liu, Dingkang Yang, Zizhi Chen, Minghao Han, Xukun Zhang, Keliang Liu, Jingwei Wei, Lihua Zhang*
203 BrandFusion: A Multi-Agent Framework for Seamless Brand Integration in Text-to-Video Generation, *Zihao Zhu, Ruotong Wang, Siwei Lyu, Min Zhang, Baoyuan Wu*
204 IEA: Amateur-Friendly Conversational Image Editing Agent via Three Stages of Multitask Alignment, *Zichen Zhu, Yuheng Sun, Mingxuan Zhu, Wenjie Ma, Situo Zhang, Zhexiang Wang, Ziyue Yang, Danyang Zhang, Kunyao Lan, Zihan Zhao, Dingye Liu, Siqi Xiang, Lu Chen, Kai Yu*
205 QuPAINT: Physics-Aware Instruction Tuning Approach to Quantum Material Discovery, *Xuan Bac Nguyen, Hoang-Quan Nguyen, Sankalp Pandey, Tim Faltermeier, Nicholas Borys, Hugh Churchill, Khoa Luu*
206 CLIPtone-GO: Geometry-Aware, Gradient-Orthogonalized Text-Guided Color Tone Adjustment, *Satyam Merothiya, Chanda Grover Kamra, Indra Deep Mastan*
207 Generative Digital Twins: Vision-Language Simulation Models for Executable Industrial Systems, *YuChe Hsu, AnJui Wang, TsaiChing Ni, YuanFu Yang*
208 Exploiting the Source-Asymmetry Confidence Gap for Generalizable AI-Generated Image Detection, *Ziyang Zheng, Weiyan Chen, Yao Xiao, Zijie Cao, Dongyu Zhang, Pengxu Wei*
209 CineMatte: Background Matting for Virtual Production and Beyond, *Yuanjian He, Chen Zhang, Fasheng Chen, Jiangbo Cao*
210 GATE: Gaussian-Attentive Transformer for Uncertainty-Aware Age Estimation, *Chaewon Lee, JunHyeok Heo, Chang-Su Kim*
211 GRAFT: Graph-Based Affordance Transfer via Part Correspondence, *Mengying Lin, Utkarsh Mishra, Ajay Mandlekar, Danfei Xu*
212 Face Time Traveller : Travel Through Ages Without Losing Identity, *Purbayan Kar, Ayush Ghadiya, Vishal Chudasama, Pankaj Wasnik, C.V. Jawahar*
213 KGGAT: Knowledge-Guided Graph Attention Network for Multi-Label Image Classification, *Christine Dewi, Dhananjay R Thiruvady, Nayyar Zaidi*
214 IntentEdit: Multi-Agent Reasoning for Intent-Driven Complex Image Editing, *Yuxuan Zhang, Shijia Huang, Liwei Wang*
215 Gen-n-Val: Agentic Image Data Generation and Validation, *Jing-En Huang, I-Sheng Fang, Tzuhsuan Huang, Yu-Lun Liu, Chih-Yu Wang, Jun-Cheng Chen*
216 SignReasoner: Compositional Reasoning for Complex Traffic Sign Understanding Via Functional Structure Units, *Ruibin Wang, Zhenyu Lin, Xinhai Zhao*
217 DARTS: Distance-Aware Robust Training for Selective Classification, *A. Q. M. Sazzad Sayyed, Nathaniel D. Bastian, Francesco Restuccia*
218 Modulate-and-Map: Crossmodal Feature Mapping with Cross-View Modulation for 3D Anomaly Detection, *Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, Luigi Di Stefano*
219 PestVL-Net: Enabling Multimodal Pest Learning Via Fine-grained Vision-Language Interaction, *Xueheng Li, Tao Hu, Ke Cao, Runsheng Qi, Huixin Zhang, Rui Li, Jie Zhang, Chengjun Xie*
220 Plug-and-Play Dynamic In-context Learning with Stochastic Regularization for Screen Content Image Super-Resolution, *Yuexin Wang, Xiaolei Wang, Guangliang Cheng, Huihui Bai, Tammam Tillo, Jimin Xiao*
221 EscherNet++: A Scalable Multi-View Framework for Amodal Completion, Novel View Synthesis and Feed-Forward 3D Reconstruction, *Xinan Zhang, Muhammad Zubair Irshad, Anthony Yezzi, Yi-Chang Tsai, Zsolt Kira*
222 Human-Intervention Segmentation via Federated Intent Embedding and Multi-Mask Recommendation, *Yeongsu Kim, Seo-Yeon Choi, Kyungsu Lee*
223 Di3PO - Diptych Diffusion DPO for Targeted Improvements in Image Generation, *Sanjana Reddy, Ishaan Malhi, Sally Ma, Praneet Dutta*
224 Robust Image Self-Recovery against Tampering using Watermark Generation with Pixel Shuffling, *Minyoung Kim, Paul Hongsuck Seo*
225 Learning to Select, Learning to Judge: Active Preference Alignment for Mars Terrain Segmentation, *JunJie Li, Miyu Li, Jiawei Wang, Yu Liu, Yumei Wang*
226 Attention Never Lie: Visual Attention Defocus Reveals and Rectifies Hallucinations in MLLMs, *Chenxi Zhao, Yan Zhou, Jufeng Yang*

227 Organizing Unstructured Image Collections using Natural Language, *Mingxuan Liu, Zhun Zhong, Jun Li, Gianni Franchi, Subhankar Roy, Elisa Ricci*
228 Thinking with Blueprints: Assisting Vision–Language Models in Spatial Reasoning via Structured Object Representation, *Weijian Ma, Shizhao Sun, Tianyu Yu, Ruiyu Wang, Tat-Seng Chua, Jiang Bian*
229 Mitigating Object Hallucinations in LVLMs via Attention Imbalance Rectification, *Han Sun, Qin Li, Peixin Wang, Min Zhang*
230 Efficient3D : A Unified Framework for Adaptive and Debiased Token Reduction in 3D MLLMs, *Yuhui Lin, Siyue Yu, Yuxing Yang, Guangliang Cheng, Jimin Xiao*
231 HiViS: Hiding Visual Tokens from the Drafter for Speculative Decoding in Vision-Language Models, *Zhinan Xie, Peisong Wang, Shuang Qiu, Jian Cheng*
232 Visual Funnel: Resolving Contextual Blindness in Multimodal Large Language Models, *Woojun Jung, Jaehoon Go, Mingyu Jeon, Sunjae Yoon, Junyeong Kim*
233 Video Parallel Scaling: Aggregating Diverse Frame Subsets for VideoLLMs, *Hyungjin Chung, Hyelin Nam, Jiyeon Kim, Hyojun Go, Byeongjun Park, Junho Kim, Joonseok Lee, Seongsu Ha, Byung-Hoon Kim*
234 Reasoning for Mobile User Experience with Multimodal LLMs: Task, Benchmark, and Approach, *Ruichao Mao, Zhou Fang, Teng Guo, Hao Yang, Yaping Li, Shaohua Peng, Maji Huang, Xiaoyu Lin, Shuoyang Liu, Xuepeng Li, Yuyu Zhang, Hai Rao*
235 Visual Reasoning Through Tool-Supervised Reinforcement Learning, *Qihua Dong, Gozde Sahin, Pei Wang, Zhaowei Cai, Robik Shrestha, Hao Yang, Davide Modolo*
236 VSI: Visual–Subtitle Integration for Keyframe Selection to Enhance Long Video Understanding, *Jianxiang He, Meisheng Hong, Jungang Li, Weiyu Guo, Xuming Hu, Hui Xiong*
237 ProGAL-VLA: Grounded Alignment through Prospective Reasoning in Vision-Language-Action Models, *Nastaran Darabi, Amit Ranjan Trivedi*
238 Myopia Rectification: KV Cache Pruning for MLLMs Via Dynamic Attention Subsidy and Token Reclamation, *Jiedong Zhuang, Lu Lu, Ming Dai, Jian Chen, Qiang Liu, Haoji Hu*
239 NaiLIA: Multimodal Nail Design Retrieval Based on Dense Intent Descriptions and Palette Queries, *Kanon Amemiya, Daichi Yashima, Kei Katsumata, Takumi Komatsu, Ryosuke Korekata, Seitaro Otsuki, Komei Sugiura*
240 Logical Consistency Optimization for Few-Shot Weakly Supervised Video Anomaly Detection, *Hantao Zheng, Ning Han, Yawen Zeng, Hegui Zhu, Hao Chen*
241 VEGAS: Mitigating Hallucinations in Large Vision-Language Models via Vision-Encoder Attention Guided Adaptive Steering, *Zihu Wang, Boxun Xu, Yuxuan Xia, Peng Li*
242 COOPER: A Unified Model for Cooperative Perception and Reasoning in Spatial Intelligence, *Zefeng Zhang, Xiangzhao Hao, Hengzhu Tang, Zhenyu Zhang, Jiawei Sheng, Xiaodong Li, Zhenyang Li, Li Gao, Daiting Shi, Dawei Yin, Tingwen Liu*
243 CoVCR: Bridging Visual Narrative Gaps via Context Generation for Robust Commonsense Reasoning, *Xinyu Li, Shiliang Sun*
244 MMR1: Enhancing Multimodal Reasoning with Variance-Aware Sampling, *Sicong Leng, Jing Wang, Jiaxi Li, Hao Zhang, Zhiqiang Hu, Boqiang Zhang, Yuming Jiang, Hang Zhang, Xin Li, Deli Zhao, Wei Lu, Yu Rong, Aixin Sun, Shijian Lu*
245 Active Video Perception: Iterative Evidence Seeking for Agentic Long Video Understanding, *Ziyang Wang, Honglu Zhou, Shijie Wang, Junnan Li, Caiming Xiong, Silvio Savarese, Mohit Bansal, Michael S. Ryoo, Juan Carlos Niebles*
246 VoQA: Visual-only Question Answering, *Jianing An, Luyang Jiang, Jie Luo, Wenjun Wu, Lei Huang*
247 Benchmarking Vision-Language Models under Contradictory Virtual Content Attacks in Augmented Reality, *Yanming Xiu, Zhengyuan Jiang, Neil Zhenqiang Gong, Maria Gorlatova*
248 Language-Augmented Semantic Priors for B-Spline Surface Fitting, *Yunzhong Lou, Yusheng Luo, Jiahao Li, Yu Song, Xiangdong Zhou*
249 Flash-Unified: A Training-Free and Task-Aware Acceleration Framework for Native Unified Models, *Junlong Ke, Zichen Wen, Boxue Yang, Yantai Yang, Xuyang Liu, Chenfei Liao, Zhaorun Chen, Shaobo Wang, Linfeng Zhang*
250 Trajectory-Diversity-Driven Robust Vision-and-Language Navigation, *Jiangyang Li, Cong Wan, SongLin Dong, Chenhao Ding, Qiang Wang, Zhiheng Ma, Yihong Gong*
251 V-STaR: Benchmarking Video-LLMs on Video Spatio-Temporal Reasoning, *Zixu Cheng, Jian Hu, Ziquan Liu, Chenyang Si, Wei Li, Shaogang Gong*
252 Distilling Counterfactual Reasoning from Language to Vision: Causal Graph-Guided Post-Training for Video Understanding, *Yuefei Chen, Jiang Liu, Xiaodong Lin, Ruixiang Tang*
253 Exploring Physics-aware Video Generation through Reinforcement Learning with Autoregressive Tokens, *Wang Lin, Liyu Jia, Wentao Hu, Kaihang Pan, Zhongqi Yue, Fengda Zhang, Wei Zhao, Jingyuan Chen, Fei Wu, Hanwang Zhang*
254 Are Video Models Ready as Zero-Shot Reasoners? An Empirical Study with the MME-CoF Benchmark, *Ziyu Guo, Xinyan Chen, Renrui Zhang, Ruichuan An, Yu Qi, Dongzhi Jiang, Xiangtai Li, Manyuan Zhang, Hongsheng Li, Pheng-Ann Heng*
255 Overthinking Causes Hallucination: Tracing Confounder Propagation in Vision Language Models, *Abin Shoby, Ta Duc Huy, Tuan Dung Nguyen, Minh Khoi Ho, Qi Chen, Anton van den Hengel, Phi Le Nguyen, Johan W. Verjans, Vu Minh Hieu Phan*
256 GDP: Graph-Based Dynamic Personalization for Multimodal Large Language Models, *Cong Ray, Xiangwen Deng, Feice Huang, ZhengXian Wu, Shen'ao Jiang, Peng Jiao, Zhifang Liu, Haoqian Wang*
257 AnyExperts: On-Demand Expert Allocation for Multimodal Language Models with Mixture of Experts, *Yuting Gao, Lan Wang, Hengyuan Zhao, Linjiang Huang, Si Liu, Qingpei Guo*
258 Scaling Spatial Reasoning in MLLMs through Programmatic Data Synthesis, *Helu Zhi, Jingjing Huang, Wang Xu, Yangbin Xu, Yibin Huang, Wanyue Zhang, Baoyang Jiang, Shirui Deng, Liang Zhu, FangFang Li, Tiejun Zhao, Yankai Lin, Yuan Yao*
259 Reasoning Within the Mind: Dynamic Multimodal Interleaving in Latent Space, *Chengzhi Liu, Yuzhe Yang, Yue Fan, Qingyue Wei, Sheng Liu, Xin Eric Wang*
260 RVLF: A Reinforcing Vision–Language Framework for Gloss-Free Sign Language Translation, *Zhi Rao, Yucheng Zhou, Benjia Zhou, Yiqing Huang, Sergio Escalera, Jun Wan*
261 Weaver: End-to-End Agentic System Training for Video Interleaved Reasoning, *Yudi Shi, Shangzhe Di, Qirui Chen, Qinian Wang, Jiayin Cai, Xiaolong Jiang, Yao Hu, Weidi Xie*
262 AITP: Traffic Accident Responsibility Allocation via Multimodal Large Language Models, *Zijin Zhou, Songan Zhang*
263 Grounding Hierarchical Vision-Language-Action Models Through Explicit Language-Action Alignment, *Theodor Wulff, Federico Tavella, Rahul Singh Maharjan, Manith Adikari, Angelo Cangelosi*
264 Fine-Grained Visual Prompt and Region Self-Distillation for Retrieval-Augmented VQA, *Yujie Wang, Hu Zhang, Jiye Liang, Zhiqiang Wang, Hongye Tan, Ru Li*
265 RADSeg: Unleashing Parameter and Compute Efficient Zero-Shot Open-Vocabulary Segmentation Using Agglomerative Models, *Omar Alama, Darshil Jariwala, Avigyan Bhattacharya, Seungchan Kim, Wenshan Wang, Sebastian Scherer*
266 Modality-Aware Bit Allocation for Mixed-Precision Quantization of Vision-Language Models, *Xi Zhang, Hanwei Zhu, Jiamang Wang, Xiaolin Wu, Weisi Lin*
267 Switch-KD: Visual-Switch Knowledge Distillation for Vision-Language Models, *Haoyi Sun, Xiaoxiao Wang, Ning Mao, Qian Wang, Lifu Mu, Wen Zheng, Tao Wei, Wei Chen*
268 Analyzing and Enhancing Visual Learning in LLM-based Radiology Report Generation, *Zailong Chen, Peng Gao, Johan Barthelemy, Luping Zhou, Lei Wang*
269 DocSLM: A Small Vision-Language Model for Long Multimodal Document Understanding, *Tanveer Hannan, Dimitrios Mallios, Parth Pathak, Faegheh Sardari, Thomas Seidl, Gedas Bertasius, Mohsen Fayyaz, Sunando Sengupta*
270 EchoTrail-GUI: Building Actionable Memory for GUI Agents via Critic-Guided Self-Exploration, *Runze Li, Yuwen Zhai, Bo Xu, Liwu Xu, Nian Shi, Wei Zhang, Ran Lin, Liang Wang*

271 Semantic Guided Feature Disentanglement and Reconstruction for Domain Adaptive Object Detection, *Xiaowei Zhao, Zhide Liu, Yuqing Ma, Xianglong Liu*
272 Dual-Modality Anchor-Guided Filtering for Test-Time Prompt Tuning, *Jungwon Choi, Eunwoo Kim*
273 Towards Efficient Multimodal Unified Reasoning Model via Model Merging, *Qixiang Yin, Huanjin Yao, Jianghao Chen, Jiaxing Huang, Zhicheng Zhao, Fei Su*
274 DeepSketcher: Internalizing Visual Manipulation for Multimodal Reasoning, *Chi Zhang, Haibo Qiu, Qiming Zhang, Zhixiong Zeng, Lin Ma, Jing Zhang*
275 Can Textual Reasoning Improve the Performance of MLLMs on Fine-Grained Visual Classification?, *Jie Zhu, Yiyang Su, Xiaoming Liu*
276 VRSA: Jailbreaking Multimodal Large Language Models through Visual Reasoning Sequential Attack, *Shiji Zhao, Shukun Xiong, Yao Huang, Jin Yan, Zhenyu Wu, Jiyang Guan, Ranjie Duan, Jialing Tao, Hui Xue, Xingxing Wei*
277 StreamEQA: Towards Streaming Video Understanding for Embodied Scenarios, *Yifei Wang, Zhenkai Li, Tianwen Qian, Huanran Zheng, Zheng Wang, Yuqian Fu, Xiaoling Wang*
278 MASS: Motion-Aware Spatial–temporal Grounding for Physics Reasoning and Comprehension in Vision-Language Models, *Xiyang Wu, Zongxia Li, Jihui Jin, Gouthaman KV, Vishnu Raj, Nilotpal Sinha, Jingxi Chen, Fan Du, Dinesh Manocha*
279 Beyond Syntax: Action Semantics Learning for App Agents, *Bohan Tang, Dezhao Luo, Jianheng Liu, Jingxuan Chen, Shaogang Gong, Jianye Hao, Jun Wang, Kun Shao*
280 Learning to Select Visual In-Context Demonstrations, *Eugene Lee, Yu-Chi Lin, Jiajie Diao*
281 CheXmix: Unified Generative Pretraining for Vision Language Models in Medical Imaging, *Ashwin Kumar, Robbie Holland, Corey Barrett, Jangwon Kim, Maya Varma, Zhihong Chen, Yunhe Gao, Greg Zaharchuk, Tara Taghavi, Krishnaram Kenthapadi, Akshay Chaudhari*
282 Mull-Tokens: Modality-Agnostic Latent Thinking, *Arijit Ray, Ahmed Abdelkader, Chengzhi Mao, Bryan A. Plummer, Kate Saenko, Ranjay Krishna, Leonidas Guibas, Wen-Sheng Chu*
283 SPHINX: A Synthetic Environment for Visual Perception and Reasoning, *Md Tanvirul Alam, Saksham Aggarwal, Justin Yang Chae, Nidhi Rastogi*
284 It's Time to Get It Right: Improving Analog Clock Reading and Clock-Hand Spatial Reasoning in Vision-Language Models, *Jaeha Choi, Jin Won Lee, Siwoo You, Jangho Lee*
285 Uncertainty-Guided Graph Formulation via MWIS for Token Pruning in LVLMs, *Jouwon Song, Sohyeon Kim, Kyeongbo Kong*
286 From Alignment to Reason: Multi-Agent Debate for Tactical Badminton Video Retrieval, *Yi-Xiang Zhang, Yu-Shuen Wang*
287 Distilling Out-of-Distribution Knowledge from Large Language Models for CLIP Generalization, *Qiji Ma, Chuanguang Yang, Zhulin An, Libo Huang, Erhu Zhao, Yuqi Li, Yongjun Xu*
288 Multimodal Reasoning with Explicit Reasoning Patterns and Rewards, *Han Qiu, Sheng Jin, Zhongrong Zuo, Ziyue Wang, Qi She, Ling Shao, Shijian Lu*
289 VRAG-DFD: Verifiable Retrieval-Augmentation for MLLM-based Deepfake Detection, *Hui Han, Shunli Wang, Yandan Zhao, Taiping Yao, Shouhong Ding*
290 MIRA: Multimodal Iterative Reasoning Agent for Image Editing, *Ziyun Zeng, Hang Hua, Jiebo Luo*
291 Mitigating Visual Context Degradation in Large Multimodal Models: A Training-Free Decoupled Agentic Framework, *Hongrui Jia, Chaoya Jiang, Shikun Zhang, Wei Ye*
292 CodePlot-CoT: Mathematical Visual Reasoning by Thinking with Code-Driven Images, *Chengqi Duan, Kaiyue Sun, Rongyao Fang, Manyuan Zhang, Yan Feng, Ying Luo, Yufang Liu, Ke Wang, Peng Pei, Xunliang Cai, Hongsheng Li, Yi Ma, Xihui Liu*
293 Do All Individual Layers Help? An Empirical Study of Task-Interfering Layers in Vision-Language Model, *Zhiming Liu, Yujie Wei, Lei Feng, Xiu Su, Xiaobo Xia, Weili Guan, Zeke Xie, Shuo Yang*
294 Recursive Think-Answer Process for LLMs and VLMs, *Byung-Kwan Lee, Youngchae Chee, Yong Man Ro*
295 GenSRL: Generative Spatiotemporal Representation Learning for Ophthalmic Prognosis Prediction, *Wanyu Zhang, Yanzhao Shi, Chengxin Zheng, Hua Wang, Jianing Wang, Yue Zhang, Xiaobing Yu, Xiaodan Zhang*
296 Ramen: Robust Test-Time Adaptation of Vision-Language Models with Active Sample Selection, *Wenxuan Bao, Yanjun Zhao, Xiyuan Yang, Jingrui He*
297 LED: LLM Enhanced Open-Vocabulary Object Detection without Human Curated Data Generation, *Yang Zhou, Shiyu Zhao, Yuxiao Chen, Zhenting Wang, Can Jin, Mingyu Zhao, Dimitris N. Metaxas*
298 VSAS-Bench: Real-Time Evaluation of Visual Streaming Assistant Models, *Pavan Kumar Anasosalu Vasu, Cem Koc, Fartash Faghri, Chun-Liang Li, Bo Feng, Zhengfeng Lai, Meng Cao, Oncel Tuzel, Hadi Pouransari*
299 Mitigating Vision-Text Order Bias in Vision-Language Model, *Weilin Gan, Yifan Song, Zhuocheng Yu, Sujian Li*
300 MARS-RL: Enhancing Multi-Agent RAG Systems for Multi-Modal Documents via Strategic Reasoning with Reinforcement Learning, *Zhongyu Wang, Pengbo Liu*
301 Beyond Single Object: Learning 3D Relations with Large Language Models, *Kohsuke Ide, Ryousuke Yamada, Yue Qiu, Xianzheng Ma, Yoshihiro Fukuhara, Hirokatsu Kataoka, Yutaka Satoh*
302 CarePilot: A Multi-Agent Framework for Long-Horizon Computer Task Automation in Healthcare, *Akash Ghosh, Tajamul Ashraf, Rishu Kumar Singh, Numan Saeed, Sriparna Saha, Xiuying Chen, Salman Khan*
303 Attention-Space Contrastive Guidance for Efficient Hallucination Mitigation in LVLMs, *Yujin Jo, Sangyoon Bae, Taesup Kim*
304 UnrealSpace: Analyzing Spatial Understanding and Reasoning in Controllable Simulation, *Wufei Ma, Sky Cen, Jianzhi Shen, Rex Lee, León Begiristain, Yan Zhuang, Jiawei Peng, Zhifei Yu, Tianao Song, Xinyuan Qi, Tianmin Shu, Adam Kortylewski, Alan Yuille*
305 Frequency-Modulated Visual Restoration for Matryoshka Large Multimodal Models, *Qingtao Pan, Zhihao Dou, Shuo Li*
306 Learning When to Look: A Disentangled Curriculum for Strategic Perception in Multimodal Reasoning, *Siqi Yang, Zilve Gao, Haibo Qiu, Fanfan Liu, Peng Shi, Zhixiong Zeng, Qingmin Liao, Lin Ma*
307 Hierarchical Textual Knowledge for Enhanced Image Clustering, *Yijie Zhong, Yunfan Gao, Weipeng Jiang, Haofen Wang*
308 Are Multimodal Large Language Models Ready for Omnidirectional Spatial Reasoning?, *Zihao Dongfang, Xu Zheng, Ziqiao Weng, Yuanhuiyi Lyu, Danda Pani Paudel, Luc Van Gool, Kailun Yang, Xuming Hu*
309 Entropy-Based Visual Re-perception Inference for Multimodal Models, *Jia Liufu, Qiangyu Yan, Zhehan Kan, Wenming Yang, Hailin Hu, Xinghao Chen, Borui Jiang*
310 VACoT: Rethinking Visual Data Augmentation with VLMs, *Zhengzhuo Xu, Chong Sun, SiNan Du, Chen Li, Jing Lyu, Chun Yuan*
311 Open World Image Aesthetic Assessment, *Mingxiang Liao, Tianren Ma, Xijin Zhang*
312 coDrawAgents: A Multi-Agent Dialogue Framework for Compositional Image Generation, *Chunhan Li, Qifeng Wu, Jia-Hui Pan, Ka-Hei Hui, Jingyu Hu, Yuming Jiang, Bin Sheng, Xihui Liu, Wenjuan Gong, Zhengzhe Liu*
313 PosterGen: Aesthetic-Aware Multi-Modal Paper-to-Poster Generation Via Multi-Agent LLMs, *Zhilin Zhang, Xiang Zhang, Jiaqi Wei, Yiwei Xu, Chenyu You*
314 Euclid's Gift: Enhancing Spatial Perception and Reasoning in Vision-Language Models via Geometric Surrogate Tasks, *Shijie Lian, Changti Wu, Laurence Tianruo Yang, Hang Yuan, Bin Yu, Lei Zhang, Kai Chen*
315 Why MLLMs Struggle to Determine Object Orientations, *Anju Gopinath, Nikhil Krishnaswamy, Bruce Draper*
316 VADE: Variance-Aware Dynamic Sampling via Online Sample-Level Difficulty Estimation for Multimodal Reinforcement Learning, *Zengjie Hu, Jiantao Qiu, Tianyi Bai, Haojin Yang, Binhang Yuan, Qi Jing, Conghui He, Wentao Zhang*
317 Less is More: Token-Efficient Video-QA via Adaptive Frame-Pruning and Semantic Graph Integration, *Shaoguang Wang, Weiyu Guo, Ziyang Chen, Yijie Xu, Xuming Hu, Hui Xiong*
318 Alleviating Hallucinations in Large Vision-Language Models via Decoding-Time Perturbation Adaptation, *Jiaqi Bai, Hongcheng Guo, Jiaheng Liu, Zhibo Zhou, Jian Yang, Feiran Huang*

319 RISE: Enhancing VLM Image Annotation with Self-Supervised Reasoning, *Suhang Hu, Wei Hu, Yuhang Su, Fan Zhang*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

10:15 - 11:30 Oral Session 5A: Dynamic Perception
(Bluebird Ballroom)
🏆 - Award candidate paper

1 🏆 Evidential Neural Radiance Fields, *Ruxiao Duan, Alex Wong*
2 🏆 Global-Aware Edge Prioritization for Pose Graph Initialization, *Tong Wei, Giorgos Tolias, Jiri Matas, Daniel Barath*
3 🏆 Molmo2: Open Weights and Data for Vision-Language Models with Video Understanding and Grounding, *Christopher Clark, Jieyu Zhang, Zixian Ma, Jae Sung Park, Rohun Tripathi, Sangho Lee, Mohammadreza Salehi, Jason Ren, Chris Dongjoo Kim, Yinuo Yang, Vincent Shao, Yue Yang, Weikai Huang, Ziqi Gao, Taira Anderson, Jianrui Zhang, Jitesh Jain, George Stoica, Ali Farhadi, Ranjay Krishna*
4 Optical Flow Matching: Reframing Optical Flow as Continuous Transport Dynamics, *Ao Luo, Xin Li, Fan Yang, Yuezun Li, Zhaoquan Yuan, Shan Zhao, Bing Su, Xiao Wu*
5 SEATrack: Simple, Efficient, and Adaptive Multimodal Tracker, *Junbin Su, Ziteng Xue, Shihui Zhang, Kun Chen, Weiming Hu, Zhipeng Zhang*
6 🏆 U^2Flow: Uncertainty-Aware Unsupervised Optical Flow Estimation, *Xunpei Sun, Wenwei Lin, Yi Chang, Gang Chen*

10:15 - 11:30 Oral Session 5B: Generalization and Adaptation
(Four Seasons Ballroom)

1 AToken: A Unified Tokenizer for Vision, *Jiasen Lu, Liangchen Song, Mingze Xu, Byeongjoo Ahn, Yanjun Wang, Chen Chen, Afshin Dehghan, Yinfei Yang*
2 Confusion-Aware Spectral Regularizer for Long-Tailed Recognition, *Ziquan Zhu, Gaojie Jin, Hanruo Zhu, Si-Yuan Lu, Yunxiao Zhang, Zeyu Fu, Ronghui Mu, Guoqiang Zhang, Zhao Sun, Yuhang Xia, Jiaxing Shang, Xiang Li, Lu Liu, Tianjin Huang*
3 🏆 Learning Latent Concepts for Detecting Out-of-Distribution Objects, *Ting Peng, Junhao Dong, Yew-Soon Ong*
4 Learning Like Humans: Analogical Concept Learning for Generalized Category Discovery, *Jizhou Han, Chenhao Ding, Yuhang He, Qiang Wang, Shaokun Wang, SongLin Dong, Yihong Gong*
5 Understanding and Enforcing Weight Disentanglement in Task Arithmetic, *Shangge Liu, Yuehan Yin, Lei Wang, Qi Fan, Yinghuan Shi, Wenbin Li, Yang Gao, Dacheng Tao*
6 Understanding Task Transfer in Vision-Language Models, *Bhuvan Sachdeva, Karan Uppal, Abhinav Java, Vineeth N. Balasubramanian*

10:15 - 11:30 Oral Session 5C: Geometry and Robotics
(Mile High Ballroom 1A - 2A)

1 AT-VLA: Adaptive Tactile Injection for Enhanced Feedback Reaction in Vision-Language-Action Models, *Xiaoqi Li, Muhe Cai, Jiadong Xu, Juan Zhu, Hongwei Fan, Yan Shen, Guangrui Ren, Hao Dong*
2 🏆 Learning Diffeomorphism for Medical Image Registration with Time-Embedded Architectures Using Semigroup Regularization, *Mohammadjavad Matinkia, Nilanjan Ray*
3 🏆 QuadSync: Quadrifocal Tensor Synchronization via Tucker Decomposition, *Daniel Miao, Gilad Lerman, Joe Kileel*
4 🏆 SocialNav: Training Human-Inspired Foundation Model for Socially-Aware Embodied Navigation, *Ziyi Chen, Yingnan Guo, Zedong Chu, Minghua Luo, Yanfen Shen, Mingchao Sun, Junjun Hu, Shichao Xie, Yang Kuan, Pei Shi, Zhining Gu, Lu Liu, Honglin Han, Xiaolong Wu, Mu Xu, Yu Zhang*
5 Structural Action Transformer for 3D Dexterous Manipulation, *Xiaohan Lei, Min Wang, Bohong Weng, Wengang Zhou, Houqiang Li*
6 TESO: Online Tracking of Essential Matrix by Stochastic Optimization, *Jaroslav Moravec, Radim Sara, Akihiro Sugimoto*

10:15 - 11:30 Oral Session 5D: Human-Centric Modeling & Lighting
(Mile High Ballroom 3A - 4A)

1 BoostSLT: Boosting Sign Language Translation via a Plug-and-Play Diffusion-Based Semantic Enhancer, *Changzhou Han, Wanlun Ma, Xi Tang, Kun Hu, Sheng Wen, Yang Xiang*
2 🏆 ImmerIris: A Large-Scale Dataset and Benchmark for Off-Axis and Unconstrained Iris Recognition in Immersive Applications, *Yuxi Mi, Qiuyang Yuan, Zhizhou Zhong, Xuan Zhao, Jiaogen Zhou, Fubao Zhu, Jihong Guan, Shuigeng Zhou*
3 🏆 OLATverse: A Large-scale Real-world Object Dataset with Precise Lighting Control, *Xilong Zhou, Jianchun Chen, Pramod Rao, Timo Teufel, Linjie Lyu, Tigran Minasian, Oleksandr Sotnychenko, Xiao-Xiao Long, Marc Habermann, Christian Theobalt*
4 🏆 OpenDance: Multimodal Controllable 3D Dance Generation with Large-scale Internet Data, *Jinlu Zhang, Zixi Kang, Libin Liu, Jianlong Chang, Qi Tian, Feng Gao, Yizhou Wang*
5 POLAR: A Portrait OLAT Dataset and Generative Framework for Illumination-Aware Face Modeling, *Zhuo Chen, Chengqun Yang, Zhuo Su, Zheng Lv, Jingnan Gao, Xiaoyuan Zhang, Xiaokang Yang, Yichao Yan*
6 🏆 Relightable Holoported Characters: Capturing and Relighting Dynamic Human Performance from Sparse Views, *Kunwar Maheep Singh, Jianchun Chen, Vladislav Golyanik, Stephan J. Garbin, Thabo Beeler, Rishabh Dabral, Marc Habermann, Christian Theobalt*

11:45 - 13:45 Poster Session 5 & Exhibit Hall (ExHall F)
❊ - Highlight paper 🏆 - Award candidate paper

1 🏆 Evidential Neural Radiance Fields, *Ruxiao Duan, Alex Wong*
2 🏆 Global-Aware Edge Prioritization for Pose Graph Initialization, *Tong Wei, Giorgos Tolias, Jiri Matas, Daniel Barath*
3 🏆 Molmo2: Open Weights and Data for Vision-Language Models with Video Understanding and Grounding, *Christopher Clark, Jieyu Zhang, Zixian Ma, Jae Sung Park, Rohun Tripathi, Sangho Lee, Mohammadreza Salehi, Jason Ren, Chris Dongjoo Kim, Yinuo Yang, Vincent Shao, Yue Yang, Weikai Huang, Ziqi Gao, Taira Anderson, Jianrui Zhang, Jitesh Jain, George Stoica, Ali Farhadi, Ranjay Krishna*
4 Optical Flow Matching: Reframing Optical Flow as Continuous Transport Dynamics, *Ao Luo, Xin Li, Fan Yang, Yuezun Li, Zhaoquan Yuan, Shan Zhao, Bing Su, Xiao Wu*
5 SEATrack: Simple, Efficient, and Adaptive Multimodal Tracker, *Junbin Su, Ziteng Xue, Shihui Zhang, Kun Chen, Weiming Hu, Zhipeng Zhang*
6 🏆 U^2Flow: Uncertainty-Aware Unsupervised Optical Flow Estimation, *Xunpei Sun, Wenwei Lin, Yi Chang, Gang Chen*
7 AToken: A Unified Tokenizer for Vision, *Jiasen Lu, Liangchen Song, Mingze Xu, Byeongjoo Ahn, Yanjun Wang, Chen Chen, Afshin Dehghan, Yinfei Yang*
8 Confusion-Aware Spectral Regularizer for Long-Tailed Recognition, *Ziquan Zhu, Gaojie Jin, Hanruo Zhu, Si-Yuan Lu, Yunxiao Zhang, Zeyu Fu, Ronghui Mu, Guoqiang Zhang, Zhao Sun, Yuhang Xia, Jiaxing Shang, Xiang Li, Lu Liu, Tianjin Huang*
9 🏆 Learning Latent Concepts for Detecting Out-of-Distribution Objects, *Ting Peng, Junhao Dong, Yew-Soon Ong*
10 Learning Like Humans: Analogical Concept Learning for Generalized Category Discovery, *Jizhou Han, Chenhao Ding, Yuhang He, Qiang Wang, Shaokun Wang, SongLin Dong, Yihong Gong*
11 Understanding and Enforcing Weight Disentanglement in Task Arithmetic, *Shangge Liu, Yuehan Yin, Lei Wang, Qi Fan, Yinghuan Shi, Wenbin Li, Yang Gao, Dacheng Tao*
12 Understanding Task Transfer in Vision-Language Models, *Bhuvan Sachdeva, Karan Uppal, Abhinav Java, Vineeth N. Balasubramanian*
13 AT-VLA: Adaptive Tactile Injection for Enhanced Feedback Reaction in Vision-Language-Action Models, *Xiaoqi Li, Muhe Cai, Jiadong Xu, Juan Zhu, Hongwei Fan, Yan Shen, Guangrui Ren, Hao Dong*
14 🏆 Learning Diffeomorphism for Medical Image Registration with Time-Embedded Architectures Using Semigroup Regularization, *Mohammadjavad Matinkia, Nilanjan Ray*
15 🏆 QuadSync: Quadrifocal Tensor Synchronization via Tucker Decomposition, *Daniel Miao, Gilad Lerman, Joe Kileel*

16 SocialNav: Training Human-Inspired Foundation Model for Socially-Aware Embodied Navigation, *Ziyi Chen, Yingnan Guo, Zedong Chu, Minghua Luo, Yanfen Shen, Mingchao Sun, Junjun Hu, Shichao Xie, Yang Kuan, Pei Shi, Zhining Gu, Lu Liu, Honglin Han, Xiaolong Wu, Mu Xu, Yu Zhang*

17 Structural Action Transformer for 3D Dexterous Manipulation, *Xiaohan Lei, Min Wang, Bohong Weng, Wengang Zhou, Houqiang Li*

18 TESO: Online Tracking of Essential Matrix by Stochastic Optimization, *Jaroslav Moravec, Radim Sara, Akihiro Sugimoto*

19 BoostSLT: Boosting Sign Language Translation via a Plug-and-Play Diffusion-Based Semantic Enhancer, *Changzhou Han, Wanlun Ma, Xi Tang, Kun Hu, Sheng Wen, Yang Xiang*

20 ImmerIris: A Large-Scale Dataset and Benchmark for Off-Axis and Unconstrained Iris Recognition in Immersive Applications, *Yuxi Mi, Qiuyang Yuan, Zhizhou Zhong, Xuan Zhao, Jiaogen Zhou, Fubao Zhu, Jihong Guan, Shuigeng Zhou*

21 OLATverse: A Large-scale Real-world Object Dataset with Precise Lighting Control, *Xilong Zhou, Jianchun Chen, Pramod Rao, Timo Teufel, Linjie Lyu, Tigran Minasian, Oleksandr Sotnychenko, Xiao-Xiao Long, Marc Habermann, Christian Theobalt*

22 OpenDance: Multimodal Controllable 3D Dance Generation with Large-scale Internet Data, *Jinlu Zhang, Zixi Kang, Libin Liu, Jianlong Chang, Qi Tian, Feng Gao, Yizhou Wang*

23 POLAR: A Portrait OLAT Dataset and Generative Framework for Illumination-Aware Face Modeling, *Zhuo Chen, Chengqun Yang, Zhuo Su, Zheng Lv, Jingnan Gao, Xiaoyuan Zhang, Xiaokang Yang, Yichao Yan*

24 Relightable Holoported Characters: Capturing and Relighting Dynamic Human Performance from Sparse Views, *Kunwar Maheep Singh, Jianchun Chen, Vladislav Golyanik, Stephan J. Garbin, Thabo Beeler, Rishabh Dabral, Marc Habermann, Christian Theobalt*

25 Scaling View Synthesis Transformers, *Evan Kim, Hyunwoo Ryu, Thomas W. Mitchel, Vincent Sitzmann*

26 WildPose: A Unified Framework for Robust Pose Estimation in the Wild, *Jianhao Zheng, Liyuan Zhu, Zihan Zhu, Iro Armeni*

27 MoRe: Motion-aware Feed-forward 4D Reconstruction Transformer, *Juntong Fang, Zequn Chen, Weiqi Zhang, Donglin Di, Xuancheng Zhang, Chengmin Yang, Yu-Shen Liu*

28 Revisiting Monocular SLAM with Spatio-Temporal Scene Modeling, *Valter Piedade, Lalit Manam, Masashi Yamazaki, Pedro Miraldo*

29 Minimal Constraint Relaxation for Multiview Autocalibration, *Norio Kosaka, Timothy Duff, Tomas Pajdla*

30 Motion 3-to-4: 3D Motion Reconstruction for 4D Synthesis, *Hongyuan Chen, Xingyu Chen, Zexiang Xu, Anpei Chen*

31 GGPT: Geometry-Grounded Point Transformer, *Yutong Chen, Yiming Wang, Xucong Zhang, Sergey Prokudin, Siyu Tang*

32 MERG3R: A Divide-and-Conquer Approach to Large-Scale Neural Visual Geometry, *Leo Kaixuan Cheng, Abdus Shaikh, Ruofan Liang, Zhijie Wu, Yushi Guan, Nandita Vijaykumar*

33 Unlocking the Power of Critical Factors for 3D Visual Geometry Estimation, *Guangkai Xu, Hua Geng, Huanyi Zheng, Songyi Yin, Yanlong Sun, Hao Chen, Chunhua Shen*

34 KV-Tracker: Real-Time Pose Tracking with Transformers, *Marwan Taher, Ignacio Alzugaray, Kirill Mazur, Xin Kong, Andrew Davison*

35 InstructMix2Mix: Consistent Sparse-View Editing Through Multi-View Model Personalization, *Daniel Gilo, Or Litany*

36 From Rays to Projections: Better Inputs for Feed-Forward View Synthesis, *Zirui Wu, Zeren Jiang, Martin R. Oswald, Jie Song*

37 SLARM: Streaming and Language-Aligned Reconstruction Model for Dynamic Scenes, *Zhicheng Qiu, Jiarui Meng, Tong-an Luo, Yican Huang, Xuan Feng, Xuanfu Li, Zhan Xu*

38 Parallel Rigidity Matters for Bundle Adjustment, *Lalit Manam, Venu Madhav Govindu*

39 Simple but Effective Triplet-Based Compression Strategies for Compact Visual Localization, *Torsten Sattler, Zuzana Kukelova*

40 VIAFormer: Voxel-Image Alignment Transformer for High-Fidelity Voxel Refinement, *Tiancheng Fang, Bowen Pan, Lingxi Chen, Jiangjing Lyu, Chengfei Lv, Chaoyue Niu, Fan Wu*

41 Mining Attribute Subspaces for Efficient Fine-tuning of 3D Foundation Models, *Yu Jiang, Hanwen Jiang, Ahmed Abdelkader, Wen-Sheng Chu, Brandon Feng, Zhangyang Wang, Qixing Huang*

42 DualPrim: Compact 3D Reconstruction with Positive and Negative Primitives, *Xiaoxu Meng, Zhongmin Chen, Bo Yang, Weikai Chen, Weixiao Liu, Lin Gao*

43 StyleGallery: Training-free and Semantic-aware Personalized Style Transfer from Arbitrary Image References, *Boyu He, Yunfan Ye, Chang Liu, Weishang Wu, Fang Liu, Zhiping Cai*

44 DynFusion: Rethinking Condition Fusion for Adaptive Multi-Conditional Text-to-Image Generation, *Zheng Fang, Lichuan Xiang, Xu Cai, Bing Wang, Bo Yang, Hongkai Wen*

45 Agentic Retoucher for Text-To-Image Generation, *Shaocheng Shen, Jianfeng Liang, Chunlei Cai, Cong Geng, Huiyu Duan, Xiaoyun Zhang, Qiang Hu, Guangtao Zhai*

46 StyleDoctor: Towards Specialist Reward Model for Style-centric Generation Tasks, *Xilin He, Xiaole Xian, Xiangyu Yue, Muhammad Haris Khan*

47 SwitchCraft: Training-Free Multi-Event Video Generation with Attention Controls, *Qianxun Xu, Chenxi Song, Yujun Cai, Chi Zhang*

48 Premier: Personalized Preference Modulation with Learnable User Embedding in Text-to-Image Generation, *Zihao Wang, Yuxiang Wei, Xinpeng Zhou, Tianyu Zhang, Tao Liang, Yalong Bai, Hongzhi Zhang, Wangmeng Zuo*

49 Paper2Figure: A Multi-Agent Collaborative System for Figure Generation Towards Academic Research Paper, *Siwei Han, Haonian Ji, Siyang Xin, Juanquan Shi, Shi Qiu, Xinyu Ye, Peng Xia, Jiaqi Liu, Zhaorun Chen, Yiyang Zhou, Linjie Li, Lijuan Wang, Huaxiu Yao*

50 Adapting In-context Generation for Enhanced Composed Image Retrieval, *Haiwen Li, Zining Chen, Delong Liu, Zhaohui Hou, Zhicheng Zhao, Fei Su*

51 Transition Models: Rethinking the Generative Learning Objective, *Zidong Wang, Yiyuan Zhang, Xiaoyu Yue, Xiangyu Yue, Yangguang Li, Wanli Ouyang, Lei Bai*

52 Rethinking Glyph Spatial Information in Font Generation, *Peng Su, Xi Yang*

53 StreamDiT: Real-Time Streaming Text-to-Video Generation, *Akio Kodaira, Tingbo Hou, Ji Hou, Markos Georgopoulos, Felix Juefei-Xu, Masayoshi Tomizuka, Yue Zhao*

54 ChArtist: Generating Pictorial Charts with Unified Spatial and Subject Control, *Shishi Xiao, Tongyu Zhou, David H. Laidlaw, Gromit Yeuk-Yin Chan*

55 Camera Control for Text-to-Image Generation via Learning Viewpoint Tokens, *Xinxuan Lu, Charless Fowlkes, Alexander C. Berg*

56 3D Space as a Scratchpad for Editable Text-to-Image Generation, *Oindrila Saha, Vojtech Krs, Radomir Mech, Subhransu Maji, Matheus Gadelha, Kevin Blackburn-Matzen*

57 Aligning Multi-Character Narrative Image Generation with Multi-Aspect Human Preferences, *Ziyi Gao, Zhipeng Wei, Jingjing Chen, Zhiyu Tan, Hao Li, Yi-Ping Phoebe Chen*

58 FoleyDirector: Directing Temporal Controllable Video-to-Audio Generation via Fine-Grained Temporal Scripts, *You Li, Dewei Zhou, Fan Ma, Fu Li, Dongliang He, Yi Yang*

59 DCoAR: Deep Concept Injection into Unified Autoregressive Models for Personalized Text-to-Image Generation, *Fangtai Wu, Mushui Liu, Weijie He, Zhao Wang, Yunlong Yu*

60 DreamOmni2: Multimodal Instruction-based Generation and Editing, *Bin Xia, bohao peng, Yuechen Zhang, Junjia Huang, Jiyang Liu, Jingyao Li, Haoru Tan, Sitong Wu, Chengyao Wang, Yitong Wang, Bei Yu, Jiaya Jia*

61 AutoDebias: An Automated Framework for Detecting and Mitigating Backdoor Biases in Text-to-Image Models, *Hongyi Cai, Mohammad Mahdinur Rahman, MingKang Dong, Muxin Pu, Moayad Aloqaily, Jie Li, Xinfeng Li, Jialie Shen, Meikang Qiu, Qingsong Wen*

62 PosterIQ: A Design Perspective Benchmark for Poster Understanding and Generation, *Yuheng Feng, Wen Zhang, Haodong Duan, Xingxing Zou*

63 IVAAN: Instance-level Vision-Language Alignment via Attribute-Guided Text Prompts Generation for Nuclei Analysis, *Jaehoon Jeong, Yi Hu, Soopil Kim, Jongseong Jang, Soonyoung Lee, Sang Hyun Park*

64 IsoCLIP: Decomposing CLIP Projectors for Efficient Intra-modal Alignment, *Simone Magistri, Dipam Goswami, Marco Mistretta, Bartłomiej Twardowski, Joost van de Weijer, Andrew D. Bagdanov*

65 TIPSv2: Advancing Vision-Language Pretraining with Enhanced Patch-Text Alignment, *Bingyi Cao, Koert Chen, Kevis-Kokitsi Maninis, Kaifeng Chen, Arjun Karpur, Ye Xia, Sahil Dua, Tanmaya Dabral, Guangxing Han, Bohyung Han, Joshua Ainslie, Alex Bewley, Mithun Jacob, René Wagner, Washington Ramos, Krzysztof Choromanski, Mojtaba Seyedhosseini, Howard Zhou, Andre Araujo*

66 ❊ BioVITA: Biological Dataset, Model, and Benchmark for Visual-Textual-Acoustic Alignment, *Risa Shinoda, Kaede Shiohara, Nakamasa Inoue, Kuniaki Saito, Hiroaki Santo, Fumio Okura*

67 ❊ Boosting Visual Reprogramming for CLIP with Dual Granularity Alignment, *Jiayang Wu, Xinyang Chen, Ke Lv, Weili Guan*

68 Decouple to Generalize: Context-First Self-Evolving Learning for Data-Scarce Vision-Language Reasoning, *Tingyu Li, Zheng Sun, Jingxuan Wei, Conghui He, Lijun Wu, Cheng Tan*

69 UniGen-1.5: Enhancing Image Generation and Editing through Reward Unification in RL, *Rui Tian, Mingfei Gao, Haiming Gang, Jiasen Lu, Zhe Gan, Yinfei Yang, Zuxuan Wu, Afshin Dehghan*

70 PolySLGen: Online Multimodal Speaking-Listening Reaction Generation in Polyadic Interaction, *Zhi-Yi Lin, Thomas Markhorst, Jouh Yeong Chew, Xucong Zhang*

71 Label What Matters: Modality-Balanced and Difficulty-Aware Multimodal Active Learning, *Yuqiao Zeng, Xu Wang, Tengfei Liang, Yiqing Hao, Yi Jin, Hui Yu*

72 Unified Personalized Understanding, Generating and Editing, *Yu Zhong, Tianwei Lin, Ruike Zhu, Yuqian Yuan, Haoyu Zheng, Liang Liang, Wenqiao Zhang, Feifei Shao, Haoyuan Li, Wanggui He, Hao Jiang, Yueting Zhuang*

73 MSRL: Scaling Generative Multimodal Reward Modeling via Multi-Stage Reinforcement Learning, *Chenglong Wang, Yifu Huo, Yang Gan, Qiaozhi He, Qi Meng, Bei Li, Yan Wang, Junfu Liu, Tianhua Zhou, Jingbo Zhu, Tong Xiao*

74 Towards Uncertainty-aware Unsupervised Domain Adaptation for Videos and Time-Series with Causal Optimal Transport, *Khushboo Mishra, Varun Trivedi, Tanima Dutta*

75 Foundation Model Priors Enhance Object Focus in Feature Space for Source-Free Object Detection, *Sairam VCR, Rishabh Lalla, Aveen Dayal, Tejal Kulkarni, Anuj Lalla, Vineeth N. Balasubramanian, Muhammad Haris Khan*

76 Decision Boundary-aware Generation for Long-tailed Learning, *Jiacheng Yang, Ruichi Zhang, Chikai Shang, Mengke Li, Xinyi Shang, Junlong Gao, Yonggang Zhang, Yang Lu*

77 Towards Stable Federated Continual Test-Time Adaptation in Wild World, *Liwen Wang, Xingbo Dong, Iman Yi Liao, Zhe Jin*

78 HyCal: A Training-Free Prototype Calibration Method for Cross-Discipline Few-Shot Class-Incremental Learning, *Eunju Lee, MiHyeon Kim, JuneHyoung Kwon, Yoonji Lee, JiHyun Kim, Soojin Jang, YoungBin Kim*

79 ACE-Merging: Data-Free Model Merging with Adaptive Covariance Estimation, *Bo Xu, Haotian Wu, Hehai Lin, Weiquan Huang, Beier Zhu, Yao Shu, Chengwei Qin*

80 CHIPS: Efficient CLIP Adaptation via Curvature-aware Hybrid Influence-based Data Selection, *Xinlin Zhuang, Yichen Li, Xiwei Liu, Haolin Yang, Yifan Lu, Ziyun Zou, Yulong Li, Huifa Li, Dongliang Chen, Qinglei Wang, Weiyang Liu, Ying Qian, Jiangming Shi, Imran Razzak*

81 Addressing Exacerbated Attention Sink for Source-Free Cross-Domain Few-Shot Learning, *Shuai Yi, Yixiong Zou, Yuhua Li, Ruixuan Li*

82 ❊ Depth Hypothesis Guided Iterative Refinement for Event–Image Monocular Depth Estimation, *Daikun Liu, Teng Wang, Changyin Sun*

83 ❊ High-Quality and Efficient Turbulence Mitigation with Events, *Xiaoran Zhang, Jian Ding, Yuxing Duan, Haoyue Liu, Gang Chen, Yi Chang, Luxin Yan*

84 Tracking through Severe Occlusion via Event-Derived Transient Cues, *Hao Dong, Yujin Liu, Haoyue Liu, Zhenyu Wang, Shihan Peng, Zhiwei Shi, Yi Chang, Luxin Yan*

85 FastEventDGS: Deformable Gaussian Splatting for Fast Dynamic Scenes from a Single Event Camera, *Zijia Dai, Nico Messikommer, Rong Zou, Nikola Zubic, Davide Scaramuzza, Laurent Kneip*

86 Event-Based Motion Deblurring Using Task-Oriented 3D Gaussian Event Representations, *Shengdong Xue, Haoxiang Ma, Hao Chen, Zhen Yang, Yongjian Deng*

87 ❊ From Corners to Fiducial Tags: Revisiting Checkerboard Calibration for Event Cameras, *Taehun Ryu, Changwoo Kang, Kyungdon Joo*

88 Extending Embodied Question Answering from Perception to Decision, *Xicheng Gong, Qiwei Li, Peiran Xu, Yadong Mu*

89 Dejavu: Towards Experience Feedback Learning for Embodied Intelligence, *Shaokai Wu, Yanbiao Ji, Qiuchang Li, Zhiyi Zhang, Qichen He, Wenyuan Xie, Guodong Zhang, Bayram Bayramli, Yue Ding, Hongtao Lu*

90 Demo2Tutorial: From Human Experience to Multimodal Software Tutorials, *Zechen Bai, Zhiheng Chen, Yiqi Lin, Kevin Qinghong Lin, Difei Gao, Xiangwu Guo, Xin Wang, Mike Zheng Shou*

91 MaskDexGrasp: Generative Masked Modeling for Part-Aware Dexterous Grasp Synthesis, *Binghui Zuo, Lin Zhou, Haoxuan Xu, Jianan Yan, Zhipeng Yu, Zekai Liu, Yangang Wang*

92 Predict Before You Explore: Predictive Planning with Specialized Memory for Embodied Question Answering, *Bowen Yuan, Sisi You, Bing-Kun Bao*

93 VideoWeaver: Multimodal Multi-View Video-to-Video Transfer for Embodied Agents, *George Eskandar, Fengyi Shen, Mohammad Altillawi, Dong Chen, Yang Bai, Liudi Yang, Ziyuan Liu*

94 MindPower: Enabling Theory-of-Mind Reasoning in VLM-based Embodied Agents, *Ruoxuan Zhang, Qiyun Zheng, Zhiyu Zhou, Ziqi Liao, Siyu Wu, Jian-Yu Jiang-Lin, Bin Wen, Hongxia Xie, Jianlong Fu, Wen-Huang Cheng*

95 Align While Search: Belief-Guided Exploratory Inference for World-Grounded Embodied Agents, *Seohui Bae, Jeonghye Kim, Youngchul Sung, Woohyung Lim*

96 Rethinking Intermediate Representation for VLM-based Robot Manipulation, *Weiliang Tang, Jialin Gao, Jia-Hui Pan, Gang Wang, Li Erran Li, Yun-Hui Liu, Mingyu Ding, Pheng-Ann Heng, Chi-Wing Fu*

97 Dexterous World Models, *Byungjun Kim, Taeksoo Kim, Junyoung Lee, Hanbyul Joo*

98 FantasyVLN: Unified Multimodal Chain-of-Thought Reasoning for Vision-and-Language Navigation, *Jing Zuo, Lingzhou Mu, Fan Jiang, Chengcheng Ma, Mu Xu, Yonggang Qi*

99 ❊ UniLight: A Unified Representation for Lighting, *Zitian Zhang, Iliyan Georgiev, Michael Fischer, Yannick Hold-Geoffroy, Jean-Francois Lalonde, Valentin Deschaintre*

100 MICo-150K: A Comprehensive Dataset Advancing Multi-Image Composition, *Xinyu Wei, Kangrui Cen, Hongyang Wei, Zhen Guo, Bairui Li, Zeqing Wang, Jinrui Zhang, Lei Zhang*

101 Upsample Anything: A Simple and Hard to Beat Baseline for Feature Upsampling, *Minseok Seo, Mark Hamilton, Changick Kim*

102 Hist2Style: Histogram-Guided Stylization with Bilateral Grids, *Dekel Galor, Adam Pikielny, Zhoutong Zhang, Ke Wang, Laura Waller, Jiawen Chen, Ilya Chugunov*

103 ❊ Harmonic Canvas: Inversion-Free Editing for Visually-Guided Music Style Transfer, *Yue Lei, Siqi Yang, Ting Zhong, Fan Zhou*

104 ❊ How to Take a Memorable Picture? Empowering Users with Actionable Feedback, *Francesco Laiti, Davide Talon, Jacopo Staiano, Elisa Ricci*

105 UniEdit-I: Training-free Image Editing for Unified VLM via Iterative Understanding, Editing and Verifying, *Chengyu Bai, Jintao Chen, Xiang Bai, Yilong Chen, Qi She, Ming Lu, Shanghang Zhang*

106 SCIEval: Evaluating and Benchmarking the Faithfulness of Scientific Image Generation and Interpretation with Large Multimodal Models, *Guanghui Ye, Huan Zhao, Zhixue Zhao, Tengfei Ma, Kehan Wang, Steffen Eger, Zhihua Jiang*

107 ❊ GeoRelight: Learning Joint Geometrical Reconstruction and Relighting with Flexible Multi-Modal Diffusion Transformers, *Yuxuan Xue, Ruofan Liang, Egor Zakharov, Timur Bagautdinov, Chen Cao, Giljoo Nam, Shunsuke Saito, Gerard Pons-Moll, Javier Romero*

108 HAD: Hallucination-Aware Diffusion Priors for 3D Reconstruction, *Xi Liu, Weiwei Sun, Zhou Ren, Chris Broaddus, Siyu Huang, Laurent Guigues*

109 Catalyst4D: High-Fidelity 3D-to-4D Scene Editing via Dynamic Propagation, *Shifeng Chen, Yihui Li, Jun Liao, Hongyu Yang, Di Huang*

110 ReFlow: Self-correction Motion Learning for Dynamic Scene Reconstruction, *Yanzhe Liang, Ruijie Zhu, Hanzhi Chang, Zhuoyuan Li, Jiahao Lu, Tianzhu Zhang*

111 Semantic Foam: Unifying Spatial and Semantic Scene Decomposition, ❊ *Amr Sharafeldin, Aryan Mikaeili, Thomas Walker, Shrisudhan Govindarajan, Daniel Rebain, Kwang Moo Yi, Andrea Tagliasacchi*
112 NVGS: Neural Visibility for Occlusion Culling in 3D Gaussian Splatting, *Brent Zoomers, Florian Hahlbohm, Joni Vanherck, Lode Jorissen, Marcus Magnor, Nick Michiels*
113 NeAR: Coupled Neural Asset–Renderer Stack, ❊ *Hong Li, Chongjie Ye, Houyuan Chen, Weiqing Xiao, Ziyang Yan, Lixing Xiao, Zhaoxi Chen, Jianfeng Xiang, Shaocong Xu, Xuhui Liu, Yikai Wang, Baochang Zhang, Xiaoguang Han, Jiaolong Yang, Hao Zhao*
114 Thermal is Always Wild: Characterizing and Addressing Challenges ❊ in Thermal-Only Novel View Synthesis, *M. Kerem Aydin, Vishwanath Saragadam, Emma Alexander*
115 PhysGM: Large Physical Gaussian Model for Feed-Forward 4D ❊ Synthesis, *Chunji Lv, Zequn Chen, Donglin Di, Weinan Zhang, Hao Li, Chen Wei, Yinjie Lei, Changsheng Li*
116 Life-IQA: Boosting Blind Image Quality Assessment through GCN- ❊ enhanced Layer Interaction and MoE-based Feature Decoupling, *Long Tang, Huiyu Duan, Guoquan Zheng, Jianbo Zhang, Jie Hao, Liang Yuan*
117 TM-BSN: Triangular-Masked Blind-Spot Network for Real-World Self-Supervised Image Denoising, *Junyoung Park, Youngjin Oh, Nam Ik Cho*
118 Multinex: Lightweight Low-light Image Enhancement via Multi-prior Retinex, *Alexandru Brateanu, Tingting Mu, Codruta O. Ancuti, Cosmin Ancuti*
119 Beyond Ground-Truth: Leveraging Image Quality Priors for Real-World Image Restoration, *Fengyang Xiao, Peng Hu, Lei Xu, XingE Guo, Guanyi Qin, Yuqi Shen, Chengyu Fang, Rihan Zhang, Chunming He, Sina Farsiu*
120 ExpoCM: Exposure-Aware One-Step Generative Single-Image HDR Reconstruction, *Aoyu Liu, Zhen Liu, Ziyi Wang, Dian Chen, Bing Zeng, Shuaicheng Liu*
121 Physically-Grounded Turbulence Mitigation with Frame-Shared ❊ Degradation Parameters, *Dongxin Xie, Yan Huang, Yong Xu, Hui Ji*
122 Convexity-Aware Noise Calibration: A Self-Supervised Framework for Noise-Level-Unknown Image Denoising, *Zhan Wang, Leiquan Wang, Chunlei Wu, Yu Meng*
123 UCMNet: Uncertainty-Aware Context Memory Network for Under-Display Camera Image Restoration, *Daehyun Kim, Youngmin Kim, Yoon Ju Oh, Tae Hyun Kim*
124 Beyond the Ground Truth: Enhanced Supervision for Image Restoration, *Donghun Ryou, Inju Ha, Sanghyeok Chu, Bohyung Han*
125 ShiftLUT: Spatial Shift Enhanced Look-Up Tables for Efficient Image Restoration, *Xiaolong Zeng, Yitong Yu, Shiyao Xiong, Jinhua Hao, Ming Sun, Chao Zhou, Bin Wang*
126 Bilevel Layer-Positioning LoRA for Real Image Dehazing, *Yan Zhang, Long Ma, Yuxin Feng, Zhe Huang, Fan Zhou, Zhuo Su*
127 SD-FSMIS: Adapting Stable Diffusion for Few-Shot Medical Image Segmentation, *Meihua Li, Yang Zhang, Weizhao He, Hu Qu, Yisong Li*
128 GeoSemba: Reconstructing State Space Model for Cross Paradigm Representation in Medical Image Segmentation, *Xutao Sun, Jiarui Li, Junwen Liu, Yonggong Ren*
129 SHAPE: Structure-aware Hierarchical Unsupervised Domain Adaptation with Plausibility Evaluation for Medical Image Segmentation, *Linkuan Zhou, Yinghao Xia, Yufei Shen, Xiangyu Li, Wenjie Du, Cong Cong, Leyi Wei, Ran Su, Qiangguo Jin*
130 Delving Aleatoric Uncertainty in Medical Image Segmentation via Vision Foundation Models, *Ruiyang Li, Fang Liu, Licheng Jiao, Xinglin Xie, Jiayao Hao, Shuo Li, Xu Liu, Jingyi Yang, Lingling Li, Puhua Chen, Wenping Ma*
131 Revisiting 2D Foundation Models for Scalable 3D Medical Image Classification, *Han Liu, Bogdan Georgescu, Yanbo Zhang, Youngjin Yoo, Michael Baumgartner, Riqiang Gao, Jianing Wang, Gengyan Zhao, Eli Gibson, Dorin Comaniciu, Sasa Grbic*
132 Focus on Background: Exploring SAM's Potential in Few-shot Medical Image Segmentation with Background-centric Prompting, *Yuntian Bo, Yazhou Zhu, Piotr Koniusz, Haofeng Zhang*
133 Simple-ViLMedSAM: Simple Text Prompts Meet Vision-Language Models for Medical Image Segmentation, *Chengcan Qian, Dong Nie, Geng Chen, Daoqiang Zhang, Xuyun Wen*
134 NeuroSeg Meets DINOv3: Transferring 2D Self-Supervised Visual Priors to 3D Neuron Segmentation via DINOv3 Initialization, *Yik San Cheng, Runkai Zhao, Weidong Cai*
135 Multi-Paradigm Collaborative Adversarial Attack Against Multi-Modal Large Language Models, *Yuanbo Li, Tianyang Xu, Cong Hu, Tao Zhou, Xiao-Jun Wu, Josef Kittler*
136 TINA: Text-Free Inversion Attack for Unlearned Text-to-Image Diffusion Models, *Qianlong Xiang, Miao Zhang, Haoyu Zhang, Kun Wang, Junhui Hou, Liqiang Nie*
137 Jailbreaking Vision-Language Models via Dissonance-Guided Suffix Optimization and Image–Phrase Injection, *Jiacheng Pi, Zhiguo Yang, Xingxing Huang, Dongsheng Xu, Ruizhi Zhong, Wenjie Ruan*
138 BlackMirror: Black-Box Backdoor Detection for Text-to-Image Models via Instruction-Response Deviation, *Feiran Li, Qianqian Xu, Shilong Bao, Zhiyong Yang, Xilin Zhao, Xiaochun Cao, Qingming Huang*
139 VCP-Attack: Visual-Contrastive Projection for Transferable Black-Box Targeted Attacks on Large Vision-Language Models, *Jiawei Zhao, Minjie Du, Zihan Qin, Zhuoran Wang, Lizhe Xie, Yining Hu*
140 Adapter Shield: A Unified Framework with Built-in Authentication for Preventing Unauthorized Zero-Shot Image-to-Image Generation, *Jun Jia, Hongyi Miao, Yingjie Zhou, Wangqiu Zhou, Jianbo Zhang, Linhan Cao, Dandan Zhu, Hua Yang, Xiongkuo Min, Wei Sun, Guangtao Zhai*
141 LLaVAShield: Safeguarding Multimodal Multi-Turn Dialogues in Vision-Language Models, *Guolei Huang, Qinzhi Peng, Gan Xu, Yao Huang, Yuxuan Lu, Yongjun Shen*
142 Transform to Transfer: Boosting Adversarial Attack Transferability on Vision-Language Pre-training Models, *Yang Li, Jia-Li Yin, Luojun Lin, Wei Lin*
143 Mask to Align, Weight to Disambiguate: Reliable Unsupervised ❊ Cross-Modal Hashing with Masked-Weight Contrast, *Fan Yang, Yuanzhi Zhao, Haimei Zhao, Yudong Zhao, Haikun Xu*
144 Reliable Clustering Number Estimation for Contrastive Multi-View ❊ Clustering, *Zhengzhong Zhu, Pei Zhou, Lanxi Bai, Li Cheng, Jia Nie, Shiquan Min, Jiangping Zhu*
145 Pushing the Frontier of Audiovisual Perception with Large-Scale Multimodal Correspondence Learning, *Apoorv Vyas, Heng-Jui Chang, Cheng-Fu Yang, Po-Yao Huang, Luya Gao, Julius Richter, Sanyuan Chen, Matthew Le, Piotr Dollár, Christoph Feichtenhofer, Ann Lee, Wei-Ning Hsu*
146 Enhance-then-Balance Modality Collaboration for Robust Multimodal Sentiment Analysis, *Kang He, Yuzhe Ding, Xinrong Wang, Fei Li, Chong Teng, Donghong Ji*
147 SonoWorld: From One Image to a 3D Audio-Visual Scene, *Derong Jin, Xiyi Chen, Ming C. Lin, Ruohan Gao*
148 MoDES: Accelerating Mixture-of-Experts Multimodal Large Language Models via Dynamic Expert Skipping, *Yushi Huang, Zining Wang, Zhihang Yuan, Yifu Ding, Ruihao Gong, Jinyang Guo, Xianglong Liu, Jun Zhang*
149 EXOTIC: External Vision-driven Incomplete Multi-view Classification, ❊ *Shilin Xu, Dezhong Peng, Zhenwen Ren, Yuan Sun*
150 Easy2Hard: From Partially to Fully Unmatched Modalities as Negative Samples in Contrastive Learning, *Zhicheng Yang, Yichen Liu, Chang Ge, Xiaopeng Jiang*
151 OneCAT: Decoder-Only Auto-Regressive Model for Unified Understanding and Generation, *Han Li, Xinyu Peng, Yaoming Wang, Zelin Peng, Xin Chen, Rongxiang Weng, Jingang Wang, Xunliang Cai, Wenrui Dai, Hongkai Xiong*
152 BALM: A Model-Agnostic Framework for Balanced Multimodal Learning under Imbalanced Missing Rates, *Phuong-Anh Nguyen, Tien Anh Pham, Duc-Trong Le, Cam-Van Thi Nguyen*
153 UniT: Unified Multimodal Chain-of-Thought Test-time Scaling, *Leon Liangyu Chen, Haoyu Ma, Zhipeng Fan, Ziqi Huang, Animesh Sinha, Xiaoliang Dai, Jialiang Wang, Zecheng He, Jianwei Yang, Chunyuan Li, Junzhe Sun, Chu Wang, Serena Yeung-Levy, Felix Juefei-Xu*
154 Multi-modal Test-time Adaptation via Adaptive Probabilistic Gaussian Calibration, *Jinglin Xu, Yi Li, Chuxiong Sun, Xiao Xu, Jiangmeng Li, Fanjiang Xu*
155 Information-Theoretic Decomposition for Multimodal Interaction Learning, *Zequn Yang, Yake Wei, Haotian Ni, Zhihao Xu, Di Hu*
156 Is the Modality Gap a Bug or a Feature? A Robustness Perspective, ❊ *Rhea Chowers, Oshri Naparstek, Udi Barzelay, Yair Weiss*

157 Omni-Fake: Benchmarking Unified Multimodal Social Media Deepfake Detection, *Tianxiao Li, Zhenglin Huang, Haiquan Wen, Yiwei He, Xinze Li, Bingyu Zhu, Wuhui Duan, Congang Chen, Zeyu Fu, Yi Dong, Baoyuan Wu, Xiangtai Li, Guangliang Cheng*
158 MUST: Modality-Specific Representation-Aware Transformer for Diffusion-Enhanced Survival Prediction with Missing Modality, *Kyungwon Kim, Dosik Hwang*
159 VQRAE: Representation Quantization Autoencoders for Multimodal Understanding, Generation and Reconstruction, *Sinan Du, Jiahao Guo, Bo Li, Shuhao Cui, Zhengzhuo Xu, Yifu Luo, Yongxian Wei, Kun Gai, Xinggang Wang, Kai Wu, Chun Yuan*
160 MOS: Mitigating Optical-SAR Modality Gap for Cross-Modal Ship Re-Identification, *Yujian Zhao, Hankun Liu, Guanglin Niu*
161 SeD-UD: An Influence-Driven and Hierarchically-Decoupled Information Bottleneck for Multimodal Intent Recognition, *Qin Li, Wenbo Zhang, Limei Liu, Han Peng, Junfeng Yang, Guanying Xu*
162 MultiModalPFN: Extending Prior-Data Fitted Networks for Multimodal Tabular Learning, *Wall Kim, Chaeyoung Song, Hanul Kim*
163 LacTokGen: Latent Consistency Tokenizer for 1024-pixel Image Generation by 256 Tokens, *Qingsong Xie, Luyuan Zhang, Zhao Zhang, Siyuan Li, Zhe Huang, Zhenyu Yang, Haonan Lu*
164 FlowSteer: Guiding Few-Step Image Synthesis with Authentic Trajectories, *Lei Ke, Hubery Yin, Gongye Liu, Zhengyao Lv, Jingcai Guo, Chen Li, Wenhan Luo, Yujiu Yang, Jing Lyu*
165 Visual Autoregressive Modeling via Next Focus Prediction, *Xiaofan Li, Chenming Wu, Yanpeng Sun, Jiaming Zhou, Delin Qu, Yansong Qu, Weihao Bo, Haibao Yu, Dingkang Liang*
166 Semantic Context Matters: Improving Conditioning for Autoregressive Models, *Dongyang Jin, Ryan Xu, Jianhao Zeng, Rui Lan, Yancheng Bai, Lei Sun, Xiangxiang Chu*
167 TempoMaster: Efficient Long Video Generation via Next-Frame-Rate Prediction, *Yukuo Ma, Cong Liu, Junke Wang, Junqi Liu, Haibin Huang, Zuxuan Wu, Chi Zhang, Xuelong Li*
168 FlashIn: Fast and Accurate Image Inversion for Real-time Image Editing, *Guangzhi Wang*
169 EasyV2V: A High-quality Instruction-based Video Editing Framework, *Jinjie Mai, Chaoyang Wang, Gordon Guocheng Qian, Willi Menapace, Sergey Tulyakov, Bernard Ghanem, Peter Wonka, Ashkan Mirzaei*
170 One Algorithm to Align Them All, *Boyi Pang, Savva Ignatyev, Vladimir Ippolitov, Ramil Khafizov, Yurii Melnik, Oleg Voynov, Maksim Nakhodnov, Aibek Alanov, Xiaopeng Fan, Peter Wonka, Evgeny Burnaev*
171 VGA-Bench: A Unified Benchmark and Multi-Model Framework for Video Aesthetics and Generation Quality Evaluation, *Longteng Jiang, DanDan Zheng, Qianqian Qiao, Heng Huang, Huaye Wang, Yihang Bo, Bao Peng, Jingdong Chen, Jun Zhou, Xin Jin*
172 ❊ Improved Mean Flows: On the Challenges of Fastforward Generative Models, *Zhengyang Geng, Yiyang Lu, Zongze Wu, Eli Shechtman, J. Zico Kolter, Kaiming He*
173 SynMotion: Semantic-Visual Adaptation for Motion Customized Video Generation, *Shuai Tan, Biao Gong, Yujie Wei, Shiwei Zhang, Zhuoxin Liu, Ke Ma, Yan Wang, Kecheng Zheng, Xing Zhu, Yujun Shen, Hengshuang Zhao*
174 Match-and-Fuse: Consistent Generation from Unstructured Image Sets, *Kate Feingold, Omri Kaduri, Tali Dekel*
175 Mixture of Style Experts for Diverse Image Stylization, *Shihao Zhu, Ziheng Ouyang, Yijia Kang, Qilong Wang, Mi Zhou, Bo Li, Ming-Ming Cheng, Qibin Hou*
176 Mirai: Autoregressive Visual Generation Needs Foresight, *Yonghao Yu, Lang Huang, Zerun Wang, Runyi Li, Toshihiko Yamasaki*
177 Align Images Before You Generate, *Shihua Zhang, Qiuhong Shen, Xinchao Wang*
178 Bridging the Perception Gap in Image Super-Resolution Evaluation, *Shaolin Su, Josep M. Rocafort, Danna Xue, David Serrano-Lozano, Lei Sun, Javier Vazquez-Corral*
179 Time-Aware One Step Diffusion Network for Real-World Image Super-Resolution, *Tianyi Zhang, Zheng-Peng Duan, Chun-Le Guo, Peng-Tao Jiang, Bo Li, Ming-Ming Cheng, Chongyi Li*
180 Restore Text First, Enhance Image Later: Two-Stage Scene Text Image Super-Resolution with Glyph Structure Guidance, *Minxing Luo, Linlong Fan, Qiushi Wang, Ge Wu, Yiyan Luo, Yuhang Yu, Jinwei Chen, Yaxing Wang, Qingnan Fan, Jian Yang*
181 IAFMNet: Information-Aware Feature Modulation for Efficient Super-Resolution, *Junwei Xu, Mengzu Liu, Zhenyu Wang, Fangfang Wu, Sijia Wu, Tao Huang, Weisheng Dong*
182 Physics-Consistent Diffusion for Efficient Fluid Super-Resolution via Multiscale Residual Correction, *Zhihao Li, Shengwei Dong, Chuang Yi, Junxuan Gao, Zhilu Lai, Zhiqiang Liu, Wei Wang, Guangtao Zhang*
183 Bridging Fidelity-Reality with Controllable One-Step Diffusion for Image Super-Resolution, *Hao Chen, Junyang Chen, Jinshan Pan, Jiangxin Dong*
184 Omni-Supervised Motion Editing: Balancing Change and Invariance through Positive-Negative Learning, *Zhenwu Shi, Jingyu Gong, Peiwei Wang, Xingzan Wang, Tianwen Qian, Wenxi Li, Yuan Fang, Jiao Xie, Lizhuang Ma, Shaohui Lin*
185 FaceCam: Portrait Video Camera Control via Scale-Aware Conditioning, *Weijie Lyu, Ming-Hsuan Yang, Zhixin Shu*
186 Cross-Axis Feature Fusion with Joint-Wise Motion Difference Prediction for Text-Based 3D Human Motion Editing, *Gyojin Han, Junmo Kim*
187 MotionMaster: Generalizable Text-Driven Motion Generation and Editing, *Nan Jiang, Yunhao Li, Lexi Pang, Zimo He, Siyuan Huang, Yixin Zhu*
188 OpenT2M: No-frill Motion Generation with Open-source, Large-scale, High-quality Data, *Bin Cao, Sipeng Zheng, Hao Luo, Boyuan Li, Jing Liu, Zongqing Lu*
189 Towards Decompositional Human Motion Generation with Energy-Based Diffusion Models, *Jianrong Zhang, Hehe Fan, Yi Yang*
190 PAMotion: Physics-Aware Motion Generation for Full-Body Interaction with Multiple Objects, *Yan Di, Yuheng Li, Yaoxing Wang, Mengge Liu, Shan Gao, Xiangyang Ji*
191 Sketch2Colab: Sketch-Conditioned Multi-Human Animation via Controllable Flow Distillation, *Divyanshu Daiya, Aniket Bera*
192 ViHOI: Human-Object Interaction Synthesis with Visual Priors, *Songjin Cai, Linjie Zhong, Ling Guo, Changxing Ding*
193 CLEP: Contrastive Language-Pose Pretraining, *Sen Jia, Huayu Wang, Hsiang-Wei Huang, Zhaochong An, Jenq-Neng Hwang, Huaping Zhang, Lei Li*
194 OpenFS: Multi-Hand-Capable Fingerspelling Recognition with Implicit Signing-Hand Detection and Frame-Wise Letter-Conditioned Synthesis, *Junuk Cha, Jihyeon Kim, Han-Mu Park*
195 ARMFlow: AutoRegressive MeanFlow for Online 3D Human Reaction Generation, *Zichen Geng, Zeeshan Hayder, Wei Liu, Hesheng Wang, Ajmal Saeed Mian*
196 InterPhys: Physics-aware Human Motion Synthesis in a Dynamic Scene, *Chaoyue Xing, Wei Mao, Miaomiao Liu*
197 Beyond Mimicry: Learning Whole-Body Human-Humanoid Interaction from Human-Human Demonstrations, *Wei-Jin Huang, Yue-Yi Zhang, Yi-Lin Wei, Zhi-Wei Xia, Juantao Tan, Yuan-Ming Li, Zhilin Zhao, Wei-Shi Zheng*
198 PHAC: Promptable Human Amodal Completion, *Seung Young Noh, Ju Yong Chang*
199 ❊ CoordSpeaker: Exploiting Gesture Captioning for Coordinated Caption-Empowered Co-Speech Gesture Generation, *Fengyi Fang, Sicheng Yang, Wenming Yang*
200 ❊ IntrinsicWeather: Controllable Weather Editing in Intrinsic Space, *Yixin Zhu, Zuo-Liang Zhu, Jian Yang, Miloš Hašan, Jin Xie, Beibei Wang*
201 Outlier-Robust Diffusion Solvers for Inverse Problems, *Yang Zheng, Jiahua Liu, Tongyao Pang, Wen Li, Zhaoqiang Liu*
202 Beyond Fixed Formulas: Data-Driven Linear Predictor for Efficient Diffusion Models, *Zhirong Shen, Rui Huang, Jiacheng Liu, Chang Zou, Peiliang Cai, Shikang Zheng, Zhengyi Shi, Liang Feng, Linfeng Zhang*
203 ReasonX: MLLM-Guided Intrinsic Image Decomposition, *Alara Dirik, Tuanfeng Yang Wang, Duygu Ceylan, Stefanos Zafeiriou, Anna Frühstück*
204 Diff-SemiER: Transparency-Aware Adaptive Fusion Diffusion Model with Generative Prior for Semi-Transparent Eyeglasses Removal, *Jiahao Li, Shiqi Yin, Zhenxiang Lian, Jingtao Guo*
205 KLIP: Localized Distribution Shift Detection via KL-Divergence with Diffusion Priors in Inverse Problems, *Alireza Kheirandish, Jihoon Hong, Sara Fridovich-Keil*

206 Elucidating the Design Space of Arbitrary-Noise-Based Diffusion Models, *Xingyu Qiu, Mengying Yang, Xinghua Ma, Dong Liang, Fanding Li, Gongning Luo, Wei Wang, Kuanquan Wang, Shuo Li*
207 Taming Generative Diffusion Model for Task-Oriented Infrared Imaging, *Tengyu Ma, Zhilong Dai, Yubo Diao, Guanming An, Long Ma, Jinyuan Liu, Risheng Liu*
208 Attention, May I Have Your Decision? Localizing Generative Choices in Diffusion Models, *Katarzyna Zaleska, Łukasz Popek, Monika Wysoczańska, Kamil Deja*
209 RxnCaption: Reformulating Reaction Diagram Parsing as Visual Prompt Guided Captioning, *Jiahe Song, Chuang Wang, Bowen Jiang, Yinfan Wang, Hao Zheng, Xingjian Wei, Chengjin Liu, Rui Nie, Junyuan Gao, Jiaxing Sun, Yubin Wang, Lijun Wu, Zhenhua Huang, Jiang Wu, Qian Yu, Conghui He*
210 More than the Sum: Panorama-Language Models for Adverse Omni-Scenes, *Weijia Fan, Ruiping Liu, Jiale Wei, Yufan Chen, Junwei Zheng, Zichao Zeng, Jiaming Zhang, Qiufu Li, Linlin Shen, Rainer Stiefelhagen*
211 DiGraphHal-Bench: Evaluating Multimodal Large Language Models on Complex Directed Graphs, *Yixin Fan, Zhao He, Yuxin Hou, Changhua Zhou, Zihao Liu, Peng Wang, Chenglong Lu, Xu Zhang, Wei Wang*
212 SEA-Vision: A Multilingual Benchmark for Comprehensive Document and Scene Text Understanding in Southeast Asia, *Pengfei Yue, Xingran Zhao, Juntao Chen, Peng Hou, Wang Longchao, Jianghang Lin, Shengchuan Zhang, Anxiang Zeng, Liujuan Cao*
213 Time Blindness: Why Video-Language Models Can't See What Humans Can?, *Ujjwal Upadhyay, Mukul Ranjan, Zhiqiang Shen, Mohamed Elhoseiny*
214 Spot The Ball: A Benchmark for Visual Social Inference, *Neha Balamurugan, Sarah Wu, Cristobal Eyzaguirre, Tobias Gerstenberg*
215 MM-SeR: Multimodal Self-Refinement for Lightweight Image Captioning, *Junha Song, Yongsik Jo, So Yeon Min, Quanting Xie, Taehwan Kim, Yonatan Bisk, Jaegul Choo*
216 E-comIQ-ZH: A Human-Aligned Dataset and Benchmark for Fine-Grained Evaluation of E-commerce Posters with Chain-of-Thought, *Meiqi Sun, Mingyu Li, Junxiong Zhu*
217 GeoWorld: Geometric World Models, *Zeyu Zhang, Danning Li, Ian Reid, Richard Hartley*
218 ORD: Object-Relation Decoupling for Generalized 3D Visual Grounding, *Ronggang Huang, Fansen Meng, Huaidong Zhang, Xuemiao Xu*
219 Benchmarking PhD-Level Coding in 3D Geometric Computer Vision, *Wenyi Li, Renkai Luo, Yue Yu, Huan-ang Gao, Mingju Gao, Li Yuan, Chaoyou Fu, Hao Zhao*
220 MonoVLM: Monocular 3D Visual Grounding with Vision Language Models, *Huaizhi Qu, Hossein Nourkhiz Mahjoub, Vaishnav Tadiparthi, Kwonjoon Lee, Tianlong Chen*
221 Curvature-Aware Captioning: Leveraging Geodesic Attention for 3D
❊ Scene Understanding, *Ziyao He, Yingjie Liu, Zhang Yangrui, Mingsong Chen, Xuan Tang, Xian Wei*
222 SPREAD: Spatial-Physical REasoning via geometry Aware Diffusion, *Minzhang Li, Kuixiang Shao, Xuebing Li, Yuyang Jiao, Yinuo Bai, Hengan Zhou, Sixian Shen, Jiayuan Gu, Jingyi Yu*
223 ExtrinSplat: Decoupling Geometry and Semantics for Open-Vocabulary Understanding in 3D Gaussian Splatting, *Jiayu Ding, Xinpeng Liu, Zhiyi Pan, Shiqiang Long, Ge Li*
224 SpatialScore: Towards Comprehensive Evaluation for Spatial
❊ Intelligence, *Haoning Wu, Xiao Huang, Yaohui Chen, Ya Zhang, Yanfeng Wang, Weidi Xie*
225 4D-RGPT: Toward Region-level 4D Understanding via Perceptual
❊ Distillation, *Chiao-An Yang, Ryo Hachiuma, Sifei Liu, Subhashree Radhakrishnan, Raymond A. Yeh, Yu-Chiang Frank Wang, Min-Hung Chen*
226 VLM-3R: Vision-Language Models Augmented with Instruction-Aligned 3D Reconstruction, *Zhiwen Fan, Jian Zhang, Renjie Li, Junge Zhang, Runjin Chen, Hezhen Hu, Kevin Wang, Peihao Wang, Huaizhi Qu, Shijie Zhou, Dilin Wang, Zhicheng Yan, Hongyu Xu, Justin Theiss, Tianlong Chen, Jiachen Li, Zhengzhong Tu, Zhangyang Wang, Rakesh Ranjan*
227 Merge3D: Efficient 3D Multimodal LLMs via Joint 2D-3D Token Merging, *Tianbo Pan, Xingyi Yang, Xinchao Wang*
228 Multi-SpatialMLLM: Multi-Frame Spatial Understanding with Multi-Modal Large Language Models, *Runsen Xu, Weiyao Wang, Hao Tang, Xingyu Chen, Xiaodong Wang, Fu-Jen Chu, Matt Feiszli, Kevin J. Liang*
229 LocateAnything3D: Vision-Language 3D Detection with Chain-of-Sight, *Yunze Man, Shihao Wang, Guowen Zhang, Johan Bjorck, Liang-Yan Gui, Jim Fan, Jan Kautz, Yu-Xiong Wang, Zhiding Yu*
230 Quota-Calibrated Fine-Grained Alignment with Context-Aware Marginals for Text-based Person Retrieval, *Dongsheng Li, Xinyuan Guo, Huijie Zhang, Pingting Hao, Qiushi Xia*
231 Evo-Retriever: LLM-Guided Curriculum Evolution with Viewpoint-Pathway Collaboration for Multimodal Document Retrieval, *Weiqing Li, Jinyue Guo, Yaqi Wang, Haiyang Xiao, Yuewei Zhang, Guohua Liu, Hao Henry Wang*
232 Taxonomy-Aware Representation Alignment for Hierarchical Visual Recognition with Large Multimodal Models, *Hulingxiao He, Zhi Tan, Yuxin Peng*
233 FAAR: Efficient Frequency-Aware Multi-Task Fine-Tuning via Automatic Rank Selection, *Maxime Fontana, Michael Spratling, Miaojing Shi*
234 Model Merging in the Essential Subspace, *Longhua Li, Lei Qi, Qi Tian, Xin Geng*
235 Beyond Semantic Search: Towards Referential Anchoring in Composed
❊ Image Retrieval, *Yuxin Yang, Yinan Zhou, Yuxin Chen, Ziqi Zhang, Zongyang Ma, Chunfeng Yuan, Bing Li, Jun Gao, Weiming Hu*
236 SAVE: Speech-Aware Video Representation Learning for Video-Text Retrieval, *Ruixiang Zhao, Zhihao Xu, Bangxiang Lan, Zijie Xin, Jingyu Liu, Xirong Li*
237 MarkushGrapher-2: End-to-end Multimodal Recognition of Chemical Structures, *Tim Strohmeyer, Lucas Morin, Gerhard Ingmar Meijer, Valery Weber, Ahmed Nassar, Peter Staar*
238 Progressive Cross-Modal Causal Intervention for Long-Term Action Recognition, *Shaowu Xu, Xibin Jia, Chao Fan, Junyu Gao, Jing Chang, Qianmei Sun*
239 EthoCLIP: Ontology-Enhanced Video-Language Pretraining for
❊ Animal Behavior Understanding, *Yinuo Jing, Jinyan Wu, Zixi Yang, Kongming Liang, Xiatian Zhu, Zhanyu Ma*
240 TrajTok: Learning Trajectory Tokens Enhances Video Understanding, *Chenhao Zheng, Jieyu Zhang, Jianing Zhang, Weikai Huang, Ashutosh Kumar, Quan Kong, Oncel Tuzel, Chun-Liang Li, Ranjay Krishna*
241 Streaming Video Instruction Tuning, *Jiaer Xia, Peixian Chen, Mengdan Zhang, Xing Sun, Kaiyang Zhou*
242 VidPrism: Heterogeneous Mixture of Experts for Image-to-Video Transfer, *Rui Lin, Chuanming Wang, Huadong Ma*
243 ViterbiPlanNet: Injecting Procedural Knowledge via Differentiable
❊ Viterbi for Planning in Instructional Videos, *Luigi Seminara, Davide Moltisanti, Antonino Furnari*
244 From Static to Dynamic: Exploring Self-supervised Image-to-Video Representation Transfer Learning, *Yang Liu, Qianqian Xu, Peisong Wen, Siran Dai, Xilin Zhao, Qingming Huang*
245 Learnable Motion-Focused Tokenization for Effective and Efficient Video Unsupervised Domain Adaptation, *Tzu Ling Liu, Ian Stavness, Mrigank Rochan*
246 FluxMem: Adaptive Hierarchical Memory for Streaming Video Understanding, *Yiweng Xie, Bo He, Junke Wang, Xiangyu Zheng, Ziyi Ye, Zuxuan Wu*
247 Learning Transferable Temporal Primitives for Video Reasoning via Synthetic Videos, *Songtao Jiang, Sibo Song, Chenyi Zhou, Yuan Wang, Ruizhe Chen, Tongkun Guan, Ruilin Luo, Yan Zhang, Zhihang Tang, Yuchong Sun, Hang Zhang, Zhibo Yang, Shuai Bai, Junyang Lin, Zuozhu Liu*
248 Video Panels for Long Video Understanding, *Lars Doorenbos, Federico Spurio, Juergen Gall*
249 Gaze Target Estimation Anywhere with Concepts, *Xu Cao, Houze Yang, Vipin Gunda, Zhongyi Zhou, Tianyu Xu, Adarsh Kowdle, Inki Kim, James M. Rehg*
250 Select, Hypothesize and Verify: Towards Verified Neuron Concept Interpretation, *ZeBin Ji, Yang Hu, Xiuli Bi, Bo Liu, Bin Xiao*
251 Finding Distributed Object-Centric Properties in Self-Supervised

❊ Transformers, *Samyak Rawlekar, Amitabh Swain, Yujun Cai, Yiwei Wang, Ming-Hsuan Yang, Narendra Ahuja*
252 Explaining CLIP Zero-shot Predictions Through Concepts, *Onat Ozdemir, Anders Christensen, Stephan Alaniz, Zeynep Akata, Emre Akbas*
253 See Through the Noise: Improving Domain Generalization in Gaze Estimation, *Yanming Peng, Shijing Wang, Yaping Huang, Yi Tian*
254 Mechanisms of Object Localization in Vision–Language Models,
❊ *Timothy Schaumlöffel, Martina G. Vilas, Gemma Roig*
255 mmWaveFlow: Unified Enhancement and Generation of mmWave Human Point Clouds, *Chang Su, Beihong Jin, Qiwen Shi, Zhi Wang*
256 From Feature Learning to Spectral Basis Learning: A Unifying and Flexible Framework for Efficient and Robust Shape Matching, *Feifan Luo, Hongyang Chen*
257 Topology-aware Feature Propagation for Unsupervised Non-rigid Point Cloud Correspondence, *Haozhe Chen, Rui Li, Zhengbao Wang, Xinhao Zhu, Linjie Li, Tianyu Xiong, Xuan Ouyang, Jiaqi Yang*
258 BEV-SLD: Self-Supervised Scene Landmark Detection for Global Localization with LiDAR Bird's-Eye View Images, *David Skuddis, Vincent Ress, Wei Zhang, Vincent Ofosu Nyako, Norbert Haala*
259 SAG-GNN: Semantic-Aware Guided GNN for Descriptor-Free 2D-3D Matching, *Shihua Zhang, Tianhao Xu, Zizhuo Li, Qing Ma, Jiayi Ma*
260 LiREC-Net: A Target-Free and Learning-Based Network for LiDAR, RGB, and Event Calibration, *Aditya Ranjan Dash, Ramy Battrawy, René Schuster, Didier Stricker*
261 GM-R^2: Generative Matching Learning for Unsupervised Geometric
❊ Representation and Registration, *Haobo Jiang, Liang Yu, Jianmin Zheng*
262 4D Local Modeling Toward Dynamic Global Perception for Ambiguity-free Rotation-Invariant Point Cloud Analysis, *Jiaxun Guo, Wentao Fan, Manar Amayri, Nizar Bouguila*
263 PointNSP: Autoregressive 3D Point Cloud Generation with Next-Scale Level-of-Detail Prediction, *Ziqiao Meng, Qichao Wang, Zhiyang Dou, Zixing Song, Zhipeng Zhou, Irwin King, Peilin Zhao*
264 MORE-STEM: Long-Short MemOry REcall and Spatio-TEmporal Consistency Model for Query-Driven 3D/4D Point Cloud Segmentation, *Chade Li, Haida Feng, Pengju Zhang, Yihong Wu*
265 Low-Rank Test-Time Training for Pre-Trained Point Cloud Models, *Ouyangzi Ye, Feifei Shao, Kexin Li, Yawei Luo, Zikai Song, Ping Liu, Fengda Zhang, Hongwei Wang, Jun Xiao*
266 STAR: Test-Time Adaptation Can Enhance Universal Prompt Learning for Vision-Language Models, *Yiwei Fu, Hui Wan, Xiao Luo, Minghua Deng*
267 Exploring Visual Pretraining for Learning Language Intelligence, *Zhonghan Zhao, Yiming Zhang, Wenwei Zhang, Haiteng Zhao, Xingguang Wei, Zhangwei Gao, Kuikun Liu, Yuzhe Gu, Size Wu, Haian Huang, Jianfei Gao, Haijun Lv, Demin Song, Yunhua Zhou, Qipeng Guo, Gaoang Wang, Kai Chen*
268 VL-Eraser: Vacuum Distillation for Machine Unlearning in Vision-Language Models, *Yili Wang, Lu Dai, Tairan Huang, Yijie Xu, Hui Xiong*
269 DeAR: Fine-Grained VLM Adaptation by Decomposing Attention Head Roles, *Yiming Ma, Hongkun Yang, Lionel Z. Wang, Bin Chen, Weizhi Xian, Jianzhi Teng*
270 SynCLIP: Synonym-Coherent Language-Image Pretraining for Robust Open-Vocabulary Dense Perception, *Mingjie Xie, Guangjun He, Dongli Xu, Youtian Lin, Hongjue Li, Pengming Feng, Jian Guan, Yue Deng*
271 MODIX: A Training-Free Multimodal Information-Driven Positional
❊ Index Scaling for Vision-Language Models, *Ruoxiang Huang, Zhen Yuan*
272 VisMem: Latent Vision Memory Unlocks Potential of Vision-Language Models, *Xinlei Yu, Chengming Xu, Guibin Zhang, Zhangquan Chen, Yudong Zhang, Yongbo He, Peng-Tao Jiang, Jiangning Zhang, Xiaobin Hu, Shuicheng Yan*
273 ORION: ORthonormal Text Encoding for Universal VLM AdaptatION, *Omprakash Chakraborty, Jose Dolz, Ismail Ben Ayed*
274 CASPA: Graph-Structured Concept Anchors for Modality-Agnostic Adaptation in Vision–Language Models, *Abhiroop Chatterjee, Susmita Ghosh, Ashish Ghosh, Emmett Ientilucci*
275 Mirror Illusion Art, *Xiaopei Zhu, Zeyuan Li,*
❊ *Jun Zhu, Xiaolin Hu*
276 HOG-Layout: Hierarchical 3D Scene Generation, Optimization and Editing via Vision-Language Models, *Haiyan Jiang, Deyu Zhang, Dongdong Weng, Weitao Song, Henry Been-Lirn Duh*
277 Towards Human-Like Robot Handwriting via Contour-Aware
❊ Generation, *Yutao Qin, Gang Dai, Yifan Zhang, Youwei Han, Qisheng He, Shuangping Huang*
278 MajutsuCity: Language-driven Aesthetic-adaptive City Generation with Controllable 3D Assets and Layouts, *Zilong Huang, Jun He, Xiaobin Huang, Ziyi Xiong, Yang Luo, Junyan Ye, Weijia Li, Yiping Chen, Ting Han*
279 VectorArk: Learning Practical Image Vectorization with Rounded Polygon Representation, *Tarun Gehlaut, Difan Liu, Charu Bansal, Krutik Malani, Souymodip Chakraborty, Ankit Phogat, Matthew Fisher, Vineet Batra*
280 OctoT2I: A Self-Evolving Agentic Text-to-Image Router, *Xu Jiang, Bin Chen, Gehui Li, Yule Duan, Ronggang Wang, Jian Zhang*
281 LottieGPT: Tokenizing Vector Animation for Autoregressive Generation, *Junhao Chen, Kejun Gao, Yuehan Cui, Mingze Sun, Mingjin Chen, Shaohui Wang, Xiaoxiao Long, Fei Ma, Qi Tian, Hao Zhao, Ruqi Huang*
282 SEA: Evaluating Sketch Abstraction Efficiency via Element-level Commonsense Visual Question Answering, *Jiho Park, Sieun Choi, Jaeyoon Seo, Minho Sohn, Yeana Kim, Jihie Kim*
283 Selective Amnesia using Contrastive Subnet Erasure for Class Level
❊ Unlearning in Vision Models, *Vishal Pramanik, Maisha Maliha, Susmit Jha, Alvaro Velasquez, Olivera Kotevska, Sumit Kumar Jha*
284 A Closed-Form Solution for Debiasing Vision-Language Models with Utility Guarantees Across Modalities and Tasks, *Tangzheng Lian, Guanyu Hu, Yijing Ren, Dimitrios Kollias, Oya Celiktutan*
285 Rank-Guided Pseudo-Bias Learning for Robust Black-Box Adaptation, *Rajeev Ranjan Dwivedi, Anshuman Dangwal, Vinod K Kurmi*
286 Diagnosing and Repairing Unsafe Channels in Vision-Language Models via Causal Discovery and Dual-Modal Safety Subspace Projection, *Jinhu Fu, Yihang Lou, Qingyi Si, Shudong Zhang, Sen Su*
287 WaTeRFlow: Watermark Temporal Robustness via Flow Consistency, *Utae Jeong, Sumin In, Hyunju Ryu, Jaewan Choi, Feng Yang, Jongheon Jeong, Seungryong Kim, Sangpil Kim*
288 DSO: Direct Steering Optimization for Bias Mitigation, *Lucas Monteiro Paes, Nivedha Sivakumar, Yinong Oliver Wang, Masha Fedzechkina, Barry-John Theobald, Luca Zappella, Nicholas Apostoloff*
289 SWIFT: Sliding Window Reconstruction for Few-Shot Training-Free Generated Video Attribution, *Chao Wang, Zijin Yang, Yaofei Wang, Yuang Qi, Weiming Zhang, Nenghai Yu, Kejiang Chen*
290 SineProject: Machine Unlearning for Stable Vision-Language Alignment, *Arpit Garg, Hemanth Saratchandran, Simon Lucey*
291 HiLoRA: Hierarchical Low-Rank Adaptation for Personalized
❊ Federated Learning, *Zihao Peng, Nan Zou, Jiandian Zeng, Guo Li, Ke Chen, Boyuan Li, Tian Wang*
292 OS-Fed: One Snapshot Is All You Need, *Xuwei Qian, Jinghui Zhang, Yuchuan Tan, Wenbo Huang, Zhen Wu, Shen Zhou, LiSha Gao, Ding Ding, Fang Dong*
293 FedAlign: Differentially Private Distribution Alignment for Non-IID Federated Learning, *Peng Wu, Jiapeng Zhang, Yingjie Song, Xiong Xiao, Zhuo Tang*
294 Guiding Diffusion Models with Fine-Grained Conditions and Semantics-Preserving Sampling for One-Shot Federated Learning, *Xiaojun Deng, Tianchi Liao, Zhiyuan Liu, Chuan Chen, Zibin Zheng*
295 Personalized Federated Training of Diffusion Models with Privacy Guarantees, *Kumar Kshitij Patel, Bingqing Jiang, A F M Mahfuzul Kabir, Weitong Zhang, Difan Zou, Lingxiao Wang*
296 FedRAC: Rolling Submodel Allocation for Collaborative Fairness in Federated Learning, *Zihui Wang, Yuhang Fu, Mengmeng Du, Zhimin Yuan, Yachen Liu, Weisheng Liao, Kaiyu Wang, Zheng Wang*
297 Understanding Temporal Logic Consistency in Video-Language Models through Cross-Modal Attention Discriminability, *Chengzhi Li, Heyan Huang, Ping Jian, Zhen Yang, Yaning Tian, Zhongbin Guo*

298 Small Object, Great Challenge: A Benchmark for Small Object Visual Grounding, *Wenqi Jia, Ruifan Li, Pengyue Lin, Fangxiang Feng, Zhanyu Ma, Xiaojie Wang*
299 UFVideo: Towards Unified Fine-Grained Video Cooperative Understanding with Large Language Models, *Hewen Pan, Cong Wei, Dashuang Liang, Zepeng Huang, Pengfei Gao, Ziqi Zhou, Lulu Xue, Pengfei Yan, Xiaoming Wei, Minghui Li, Shengshan Hu*
300 ReMoRa: Multimodal Large Language Model based on Refined Motion Representation for Long-Video Understanding, *Daichi Yashima, Shuhei Kurita, Yusuke Oda, Komei Sugiura*
301 CaST-Bench: Benchmarking Causal Chain-Grounded Spatio-Temporal Reasoning for Video Question Answering, *Mingfang Zhang, Jingjing Pan, Ashutosh Kumar, Rajat Saini, Mustafa Erdogan, Hsuan-Kung Yang, Caixin Kang, Yifei Huang, Yoichi Sato, Quan Kong*
302 HERO: Hierarchical Embedding-Refinement for Open-Vocabulary Temporal Sentence Grounding in Videos, *Tingting Han, Xinsong Tao, Yufei Yin, Min Tan, Sicheng Zhao, Zhou Yu*
303 ❊ Scaling the Long Video Understanding of Multimodal Large Language Models via Visual Memory Mechanism, *Tao Chen, Kun Zhang, Qiong Wu, Xiao Chen, Chao Chang, Xiaoshuai Sun, Yiyi Zhou, Rongrong Ji*
304 Hybrid Token Compression for Vision-Language Models, *Jusheng Zhang, Xiaoyang Guo, Kaitong Cai, Qinhan Lv, Yijia Fan, Wenhao Chai, Jian Wang, Keze Wang*
305 Focus, Don't Prune: Identifying Instruction-Relevant Regions for Information-Rich Image Understanding, *Mincheol Kwon, Minseung Lee, Seonga Choi, Miso Choi, Kyeongjin Oh, Hyunyoung Lee, Cheonyoung Park, Yongho Song, Seunghyun Park, Jinkyu Kim*
306 When Token Pruning is Worse than Random: Understanding Visual Token Information in VLLMs, *Yahong Wang, Juncheng Wu, Zhangkai Ni, Longzhen Yang, Yihang Liu, Chengmei Yang, Ying Wen, Lianghua He, Xianfeng Tang, Hui Liu, Yuyin Zhou*
307 VISion On Request: Enhanced VLLM efficiency with sparse, dynamically selected, vision-language interactions, *Adrian Bulat, Alberto Baldrati, Ioannis Maniadis Metaxas, Yassine Ouali, Georgios Tzimiropoulos*
308 BiGain: Unified Token Compression for Joint Generation and Classification, *Jiacheng Liu, Shengkun Tang, Jiacheng Cui, Dongkuan Xu, Zhiqiang Shen*
309 Hi-Lo Prune: Look at What You'll Lose before Pruning with Hierarchical Token Selection, *Zixun Sun, Yubo Dong, Hehe Fan, Yi Yang*
310 VLM-Pruner: Buffering for Spatial Sparsity in an Efficient VLM Centrifugal Token Pruning Paradigm, *Zhenkai Wu, Xiaowen Ma, Zhenliang Ni, Dengming Zhang, Han Shu, Xin Jiang, Xinghao Chen*
311 Bridge: Basis-Driven Causal Inference Marries VFMs for Domain Generalization, *Mingbo Hong, Feng Liu, Caroline Gevaert, George Vosselman, Hao Cheng*
312 In Pursuit of Pixel Supervision for Visual Pre-training, *Lihe Yang, Shang-Wen Li, Yang Li, Xinjie Lei, Dong Wang, Abdelrahman Mohamed, Saining Xie, Hengshuang Zhao, Kaiming He, Hu Xu*
313 GaussianMatch: Semi-Supervised Regression with Pseudo-Label Filtering via Multi-View Gaussian Consistency, *Yin Wang, Hao Lu, Zixuan Wang, Zhen Qin, Li Kuang, Mengchu Zhou, Shuiguang Deng*
314 TAR: Token-Aware Refinement for Fine-grained Generalized Category Discovery, *Xingyu Yang, Yu Zhang, Siya Mi, Xiu-Shen Wei*
315 Semantic Noise Reduction via Teacher-Guided Dual-Path Audio-Visual Representation Learning, *Linge Wang, Yingying Chen, Bingke Zhu, Lu Zhou, Jinqiao Wang*
316 The Universal Normal Embedding, *Chen Tasker, Roy Betser, Eyal Gofer, Meir Yossef Levi, Guy Gilboa*
317 Bypassing the Transport Plan: Dynamic Reweighting for Out-of-Distribution Detection with Optimal Transport, *Yang Xiao, Weiming Liu, Jun Dan, Tengyue Xu, Fan Wang, Hua Yu, Junhao Dong, Jiao Liu, Shunjie Dong, Lianyong Qi*
318 Cross-domain Dual-stream Feature Disentanglement for Brain Disorder Prediction with Sparsely Labeled PET, *Huabin Wang, Xinyu Chen, Yuan Zhou, Fei Liu*
319 Debiased Sample Selection for Learning with Noisy Labels, *Weiran Pan, Wei Wei, Wenfeng Xie*
320 Driving on Registers, *Ellington Kirby, Alexandre Boulch, Yihong Xu, Yuan Yin, Gilles Puy, Éloi Zablocki, Andrei Bursuc, Spyros Gidaris, Renaud Marlet, Florent Bartoccioni, Anh-Quan Cao, Nermin Samet, Tuan-Hung VU, Matthieu Cord*
321 Open-Ended Instruction Realization with LLM-Enabled Multi-Planner Scheduling in Autonomous Vehicles, *Jiawei Liu, Xun Gong, Fen Fang, Muli Yang, Bohao Qu, Yunfeng Hu, Hong Chen, Xulei Yang, Qing Guo*
322 EE-RL: Vision Language Guided Reinforcement Learning with Explorer and Expert model for End-to-End Autonomous Driving, *Xiaolong Li, Lan Yang, Ruyang Li, Shan Fang, Yang Liu, Xiangmo Zhao*
323 Sensor2Sensor: Cross-Embodiment Sensor Conversion for Autonomous Driving, *Jiahao Wang, Bo Sun, Yijing Bai, Vincent Casser, Songyou Peng, Zehao Zhu, Meng-Li Shih, Xander Masotto, Shih-Yang Su, Kanaad Parvate, Tiancheng Ge, Linn Bieske, Dragomir Anguelov, Mingxing Tan, Chiyu Max Jiang*
324 SHARP: Short-Window Streaming for Accurate and Robust Prediction in Motion Forecasting, *Alexander Prutsch, Christian Fruhwirth-Reisinger, David Schinagl, Horst Possegger*
325 DriveCombo: Benchmarking Compositional Traffic Rule Reasoning in Autonomous Driving, *Enhui Ma, Jiahuan Zhang, Guantian Zheng, Tao Tang, Shengbo Eben Li, Yuhang Lu, Xia Zhou, Xueyang Zhang, Yifei Zhan, Kun Zhan, Zhihui Hao, Xianpeng Lang, Kaicheng Yu*
326 ❊ CausalVAD: De-confounding End-to-End Autonomous Driving via Causal Intervention, *Jiacheng Tang, Zhiyuan Zhou, Zhuolin He, Jia Zhang, Kai Zhang, Jian Pu*
327 Reliable Policy Transfer for Safety-Aware End-to-End Driving with Deep Reinforcement Learning, *Uddin Md. Borhan, Arif Raza, Zhiliang Lin, Lu Wang, Jianqiang Li, Jie Chen*
328 Learning to Drive is a Free Gift: Large-Scale Label-Free Autonomy Pretraining from Unposed In-The-Wild Videos, *Matthew Strong, Wei-Jer Chang, Quentin Herau, Jiezhi Yang, Yihan Hu, Chensheng Peng, Wei Zhan*
329 WhisperNet: A Scalable Solution for Bandwidth-Efficient Collaboration, *Gong Chen, Chaokun Zhang, Xinyan Zhao*
330 Efficient Equivariant Transformer for Self-Driving Agent Modeling, *Scott Xu, Dian Chen, Kelvin Wong, Chris Zhang, Kion Fallah, Raquel Urtasun*
331 Generalizable Co-Salient Object Detection via Mixed Content-Style Modulation, *Guanting Guo, Shenglong Hu, Kaihua Zhang, Guangcan Liu, Min Xia*
332 Saliency-Driven Token Merging for Vision Transformers, *Weiying Xie, Xiaoyu Chen, Xin Zhang, Chenhe Hao, Jitao Ma, Yunsong Li, Leyuan Fang*
333 RISE: Single Static Radar-based Indoor Scene Understanding, *Kaichen Zhou, Laura Dodds, Sayed Saad Afzal, Fadel Adib*
334 Mixture-of-Experts based Feature Decoupling for Open Vocabulary Scene Graph Generation, *Yiming Li, Sisi You, Bing-Kun Bao*
335 TF-SSD: A Strong Pipeline via Synergic Mask Filter for Training-free Co-salient Object Detection, *Zhijin He, Shuo Jin, Siyue Yu, Shuwei Wu, Bingfeng Zhang, Li Yu, Jimin Xiao*
336 ❊ Denoise and Align: Towards Source-Free UDA for Robust Panoramic Semantic Segmentation, *Yaowen Chang, Zhen Cao, Xu Zheng, Xiaoxin Mi, Zhen Dong*
337 SPOT: Spatiotemporal Prompt Optimization for Motion-Stabilized MLLM-Guided Video Segmentation, *Jiayi Fan, Zheyun Qin, Xiaoming Xi, Xiushan Nie, Yilong Yin*
338 ❊ Changes in Real Time: Online Scene Change Detection with Multi-View Fusion, *Chamuditha Jayanga Galappaththige, Jason Lai, Lloyd Windrim, Donald Dansereau, Niko Suenderhauf, Dimity Miller*
339 Subspace Alignment for CLIP-based Continual Learning via Canonical Correlation Analysis, *Huan Zhang, Shuyu Dong, Yujin Zheng, Dingwen Wang, Shenghua Fan, Fan Lyu*
340 DGS: Dual Gradient and Semantic-Shift Guided Low-Rank Adaptation for Class Incremental Learning, *Kai Li, Jiafeng Li, Lianghua He, Ying Wen*
341 Dynamic Magic: Unleashing Restricted Knowledge for Lifelong Person Re-Identification, *Jinjia Peng, Jican Tan, Jiazuo Yu, Zeze Tao, Huibing Wang*

342 Which Concepts to Forget and How to Refuse? Decomposing Concepts for Continual Unlearning in Large Vision-Language Models, *Hyundong Jin, Dongyoon Han, Eunwoo Kim*
343 Temporal Imbalance of Positive and Negative Supervision in Class-Incremental Learning, *Jinge Ma, Fengqing Zhu*
344 ❊ Forging a Dynamic Memory: Retrieval-Guided Continual Learning for Generalist Medical Foundation Models, *Zizhi Chen, Yizhen Gao, Minghao Han, Yizhou Liu, Zhaoyu Chen, Dingkang Yang, Lihua Zhang*
345 Dance Across Shifts: Forward-Facilitation Continual Test-Time Adaptation through Dynamic Style Bridging, *Zhilin Zhu, Yabin Wang, Zhiheng Ma, Yaguang Song, Yaowei Wang, Xiaopeng Hong*
346 Few-Shot Hybrid Incremental Learning: Continually Learning under Data Scarcity and Task Uncertainty, *Yan Li, Yuzhu Shi, Kan Zhou, Shu Zhang, Diqi He, Dingwen Zhang, Junwei Han*
347 High-Fidelity Mobile Avatars with Pruned Local Blendshapes, *Youyi Zhan, He Wang, Tianjia Shao, Kun Zhou*
348 ❊ PhysSkin: Real-Time and Generalizable Physics-Based Animation via Self-Supervised Neural Skinning, *Yuanhang Lei, Tao Cheng, Xingxuan Li, Boming Zhao, Siyuan Huang, Ruizhen Hu, Peter Yichen Chen, Hujun Bao, Zhaopeng Cui*
349 Bridging Privacy and Provenance: Traceable Virtual Identity Generation, *Xianhan Zeng, Xiaoxiao Hu, Sheng Li, Zhenxing Qian, Xinpeng Zhang*
350 PortraitDirector: A Hierarchical Disentanglement Framework for Controllable and Real-time Facial Reenactment, *Chaonan Ji, Jinwei Qi, Sheng Xu, Peng Zhang, Bang Zhang*
351 Dynamic Label Noise Suppression with Optimal Teacher Pool for Facial Expression Recognition, *Yuzhuang Yang, Xiaolin Tian, Qigong Sun*
352 MimicTalker: A Multimodal Interactive and Memory-Enhanced Framework for Real-Time Dyadic 3D Head Generation, *Yinuo Wang, Yanbo Fan, Xuan Wang, Boyao Zhou, Yu Guo, Yujun Shen, Fei Wang*
353 DecoVLN: Decoupling Observation, Reasoning, and Correction for Vision-and-Language Navigation, *Zihao Xin, Wentong Li, Yixuan Jiang, Bin Wang, Runmin Cong, Jie Qin, Shengjun Huang*
354 HybridDriveVLA: Vision-Language-Action Model with Visual CoT reasoning and ToT Evaluation for Autonomous Driving, *Yipene Cedric Francois Bassole, Sungwoo Kim, Jiwoo Jung, Yunsick Sung*
355 NavForesee: A Unified Vision-Language World Model for Hierarchical Planning and Dual-Horizon Navigation Prediction, *Fei Liu, Shichao Xie, Minghua Luo, Zedong Chu, Junjun Hu, Xiaolong Wu, Mu Xu*
356 LookasideVLN: Direction-Aware Aerial Vision-and-Language Navigation, *Yuwei Ning, Ganlong Zhao, Yipeng Qin, Si Liu, Yang Liu, Liang Lin, Guanbin Li*
357 MAPS: Preserving Vision-Language Representations via Module-Wise Proximity Scheduling for Better Vision-Language-Action Generalization, *Chengyue Huang, Mellon M. Zhang, Robert Azarcon, Glen Chou, Zsolt Kira*
358 D3D-VLP: Dynamic 3D Vision-Language-Planning Model for Embodied Grounding and Navigation, *Zihan Wang, Seungjun Lee, Guangzhao Dai, Gim Hee Lee*
359 FreeForm: Reduced-Order Deformable Simulation from Particle-Based Skinning Eigenmodes, *Donglai Xiang, Vismay Modi, Rishit Dagli, Ty Trusty, Gilles Daviet, Anka He Chen, Nicholas Sharp, David I.W. Levin*
360 GeoDiff4D: Geometry-Aware Diffusion for 4D Head Avatar Reconstruction, *Chao Xu, Xiaochen Zhao, Xiang Deng, Jingxiang Sun, Donglin Di, Zhuo Su, Yebin Liu*
361 4DEquine: Disentangling Motion and Appearance for 4D Equine Reconstruction from Monocular Video, *Jin Lyu, Liang An, Pujin Cheng, Yebin Liu, Xiaoying Tang*
362 PhysHO: Physics-Based Dynamic 3D Gaussian Human and Object from Monocular Video, *Suyi Jiang, Gim Hee Lee*
363 ProgressiveAvatars: Progressive Animatable 3D Gaussian Avatars, *Kaiwen Song, Jinkai Cui, Juyong Zhang*
364 ZINA: Multimodal Fine-grained Hallucination Detection and Editing, *Yuiga Wada, Kazuki Matsuda, Komei Sugiura, Graham Neubig*
365 Mitigating Multimodal Hallucinations via Gradient-based Self-Reflection, *Shan Wang, Maying Shen, Nadine Chang, Chuong Nguyen, Hongdong Li, Jose M. Alvarez*
366 HalluGen: Synthesizing Realistic and Controllable Hallucinations for Evaluating Image Restoration, *Seunghoi Kim, Henry F. J. Tregidgo, Chen Jin, Matteo Figini, Daniel C. Alexander*
367 KVSmooth: Mitigating Hallucination in Multi-modal Large Language Models through Key-Value Smoothing, *Siyu Jiang, Feiyang Chen, Xiaojin Zhang, Kun He*
368 ELV-Halluc: Benchmarking Semantic Aggregation Hallucinations in Video Understanding, *Hao Lu, Jiahao Wang, Yaolun Zhang, Ruohui Wang, Xuanyu Zheng, Yepeng Tang, Dahua Lin, Lewei Lu*
369 Tell Model Where to Look: Mitigating Hallucinations in MLLMs by Vision-Guided Attention, *Jianfei Zhao, Feng Zhang, Xin Sun, Chong Feng, Zhixing Tan*
370 Circular-DPO: Aligning Multi-Stage 3D Generative Models via Preference Feedback Loop, *Zejian Li, Jiarui Ma, Han Xu, Weiting Zheng, Yangrui Zhu, Chenye Meng, Pei Chen, Ling Yang, Zhiyuan Yang, Changyuan Yang, Guang Yang, Immanuel Koh, Lingyun Sun*
371 Cloning Deterministic Worlds: The Critical Role of Latent Geometry in Long-Horizon World Models, *Zaishuo Xia, Yukuan Lu, Xinyi Li, Yifan Xu, Yubei Chen*
372 PrITTI: Primitive-based Generation of Controllable and Editable 3D Semantic Urban Scenes, *Christina Ourania Tze, Daniel Dauner, Yiyi Liao, Dzmitry Tsishkou, Andreas Geiger*
373 CubeComposer: Spatio-Temporal Autoregressive 4K 360° Video Generation from Perspective Video, *Lingen Li, Guangzhi Wang, Xiaoyu Li, Zhaoyang Zhang, Qi Dou, Jinwei Gu, Tianfan Xue, Ying Shan*
374 ExPose: Reinforcing Video Generation Models for Extreme Pose Estimation, *Youngho Yoon, Wonjune Cho, Hyunho Ha, Sujung Kim, Kuk-Jin Yoon*
375 Choreographing a World of Dynamic Objects, *Yanzhe Lyu, Chen Geng, Karthik Dharmarajan, Yunzhi Zhang, Hadi Alzayer, Shangzhe Wu, Jiajun Wu*
376 SounDiT: Geo-Contextual Soundscape-to-Landscape Generation, *Junbo Wang, Haofeng Tan, Bowen Liao, Albert Jiang, Teng Fei, Qixing Huang, Bing Zhou, Zhengzhong Tu, Shan Ye, Yuhao Kang*
377 ❊ Vista4D: Video Reshooting with 4D Point Clouds, *Kuan Heng Lin, Zhizheng Liu, Pablo Salamanca, Yash Kant, Ryan Burgert, Yuancheng Xu, Koichi Namekata, Yiwei Zhao, Bolei Zhou, Micah Goldblum, Paul Debevec, Ning Yu*
378 CamDirector: Towards Long-Term Coherent Video Trajectory Editing, *Kejia Yin, Zhihao Shi, Weilin Wan, Yuhongze Zhou, Yuanhao Yu, Xinxin Zuo, Qiang Sun, Juwei Lu*
379 ❊ Elastic3D: Controllable Stereo Video Conversion with Guided Latent Decoding, *Nando Metzger, Prune Truong, Goutam Bhat, Konrad Schindler, Federico Tombari*
380 Decoupling Bias, Aligning Distributions: Synergistic Fairness Optimization for Deepfake Detection, *Feng Ding, Wenhui Yi, Yunpeng Zhou, Xinan He, Hong Rao, Shu Hu*
381 Target-Aware Invertible Encoder with Reconstruction Guidance for Infrared Small Target Detection, *Shule Yan, Zetian Zhang, Xiao Ma, Zexuan Ji*
382 BDNet:Bio-Inspired Dual-Backbone Small Object Detection Network, *Wenchao Guan, Chuan Lin, Sihan Huang, Xiongzhen Wang, Xintao Pang*
383 ElasticFormer: Detecting Objects in HRW Shots via Elastic Computing Vision Transformer, *Wenxi Li, Jingchen Huang, Chenyang Lyu, Moran Liu, Haozhe Lin, Guiguang Ding, Yuchen Guo*
384 RGB-Event based Pedestrian Attribute Recognition: A Benchmark Dataset and An Asymmetric RWKV Fusion Framework, *Xiao Wang, Haiyang Wang, Shiao Wang, Qiang Chen, Jiandong Jin, Haoyu Song, Bo Jiang, Chenglong Li*
385 FusionAgent: A Multimodal Agent with Dynamic Model Selection for Human Recognition, *Jie Zhu, Xiao Guo, Yiyang Su, Anil Jain, Xiaoming Liu*
386 Free-Grained Hierarchical Visual Recognition, *Seulki Park, Zilin Wang, Stella X. Yu*
387 ❊ URICA: A Uniformity Region Affine Identifier Capture Algorithm for Arbitrary Region Retrieval in Pathology Images, *Ri Su, Zhao Chen, Caleb Chen Cao, Lei Chen*

388 Online Data Curation for Object Detection via Marginal Contributions ❊ to Dataset-level Average Precision, *Zitang Sun, Masakazu Yoshimura, Junji Otsuka, Atsushi Irie, Takeshi Ohashi*
389 DetAny4D: Detect Anything 4D Temporally in a Streaming RGB Video, *Jiawei Hou, Shenghao Zhang, Can Wang, Zheng Gu, Yonggen Ling, Taiping Zeng, Xiangyang Xue, Jingbo Zhang*
390 Follow the Saliency: Supervised Saliency for Retrieval-augmented Dense Video Captioning, *Seung hee Choi, MinJu Jeon, Hyunwoo Oh, Jihwan Lee, Dong-Jin Kim*
391 Video-CoE: Reinforcing Video Event Prediction via Chain of Events, *Qile Su, Jing Tang, Rui Chen, Lei Sun, Xiangxiang Chu*
392 VideoAuto-R1: Video Auto Reasoning via Thinking Once, Answering Twice, *Shuming Liu, Mingchen Zhuge, Changsheng Zhao, Jun Chen, Lemeng Wu, Zechun Liu, Chenchen Zhu, Zhipeng Cai, Chong Zhou, Haozhe Liu, Ernie Chang, Saksham Suri, Hongyu Xu, Qi Qian, Wei Wen, Balakrishnan Varadarajan, Zhuang Liu, Hu Xu, Florian Bordes, Raghuraman Krishnamoorthi, Bernard Ghanem, Vikas Chandra, Yunyang Xiong*
393 VRR-QA: Visual Relational Reasoning in Videos Beyond Explicit ❊ Cues, *Sirnam Swetha, Rohit Gupta, Parth Parag Kulkarni, David G Shatwell, Jeffrey A Chan Santiago, Nyle Siddiqui, Joseph Fioresi, Mubarak Shah*
394 Question-guided Visual Compression with Memory Feedback for Long-Term Video Understanding, *Sosuke Yamao, Natsuki Miyahara, Yuankai Qi, Shun Takeuchi*
395 CURVE: A Benchmark for Cultural and Multilingual Long Video ❊ Reasoning, *Darshan Singh, Arsha Nagrani, Kawshik Manikantan, Harman Singh, Dinesh Tewari, Tobias Weyand, Cordelia Schmid, Anelia Angelova, Shachi Dave*
396 SVBench: Evaluation of Video Generation Models on Social Reasoning, *Wenshuo Peng, Gongxuan Wang, Tianmeng Yang, Chuanhao Li, Xiaojie Xu, Hui He, Kaipeng Zhang*
397 Hierarchical Long Video Understanding with Audiovisual Entity Cohesion and Agentic Search, *Xinlei Yin, Xiulian Peng, Xiao Li, Zhiwei Xiong, Yan Lu*
398 LifeEval: A Multimodal Benchmark for Assistive AI in Egocentric Daily Life Tasks, *Hengjian Gao, Kaiwei Zhang, Shibo Wang, Mingjie Chen, Qihang Cao, Xianfeng Wang, Yucheng Zhu, Xiongkuo Min, Wei Sun, Dandan Zhu, Guangtao Zhai*
399 Thinking With Videos: Multimodal Tool-Augmented Reinforcement Learning for Long Video Reasoning, *Haoji Zhang, Xin Gu, Jiawen Li, Chixiang Ma, Sule Bai, Chubin Zhang, Bowen Zhang, Zhichao Zhou, Dongliang He, Yansong Tang*
400 Attention Surgery: An Efficient Recipe to Linearize Your Video Diffusion Transformer, *Mohsen Ghafoorian, Denis Korzhenkov, Amirhossein Habibian*
401 YOSE: You Only Select Essential Tokens for Efficient DiT-based Video Object Removal, *Chenyang Wu, Lina Lei, Fan Li, Chunle Guo, Dehong Kong, Xinran Qin, Zhixin Wang, Mingming Cheng, Chongyi Li*
402 CADC: Content Adaptive Diffusion-Based Generative Image ❊ Compression, *Xihua Sheng, Lingyu Zhu, Tianyu Zhang, Dong Liu, Shiqi Wang, Jing Wang*
403 FG-Portrait: 3D Flow Guided Editable Portrait Animation, *Yating Xu, Yunqi Miao, Evangelos Ververas, Jiankang Deng, Jifei Song*
404 ResCa: Residual Caching for Diffusion Transformers Acceleration, *Haipeng Fang, Yu Li, Fan Tang, Yixing Lu, Juan Cao, Sheng Tang*
405 IP-Adapter Is All You Need: Towards Fine-Tuning-Free Diffusion-Based Talking Face Generation, *Hao Wu, Xiangyang Luo, Hao Wang, Jiawei Zhang, Yi Zhang, Jinwei Wang*
406 SRA 2: Variational Autoencoder Self-Representation Alignment for Efficient Diffusion Training, *Mengmeng Wang, Dengyang Jiang, Liuzhuozheng Li, Yucheng Lin, Guojiang Shen, Xiangjie Kong, Yong Liu, Guang Dai, Jingdong Wang*
407 InnoAds-Composer: Efficient Condition Composition for E-Commerce Poster Generation, *Yuxin Qin, Ke Cao, Haowei Liu, Ao Ma, Fengheng Li, Honghe Zhu, Zheng Zhang, Run Ling, Wei Feng, Xuanhua He, Zhanjie Zhang, Zhen Guo, Haoyi Bian, Jingjing Lv, Junjie Shen, Ching Law*
408 Multi-Patch Global-to-Local Transformer Architecture For Efficient Flow Matching and Diffusion Model, *Quan Dao, Dimitris Metaxas*
409 SODA: Sensitivity-Oriented Dynamic Acceleration for Diffusion Transformer, *Tong Shao, Yusen Fu, Guoying Sun, Jingde Kong, Zhuotao Tian, Jingyong Su*
410 DSERT-RoLL: Robust Multi-Modal Perception for Diverse Driving Conditions with Stereo Event-RGB-Thermal Cameras, 4D Radar, and Dual-LiDAR, *Hoonhee Cho, Jae-Young Kang, Yuhwan Jeong, Yunseo Yang, Wonyoung Lee, Youngho Kim, Kuk-Jin Yoon*
411 A Semantically Disentangled Unified Model for Multi-category 3D ❊ Anomaly Detection, *SuYeon Kim, Wongyu Lee, MyeongAh Cho*
412 ReManNet: A Riemannian Manifold Network for Monocular 3D Lane Detection, *Chengzhi Hong, Bijun Li*
413 PanDA: Unsupervised Domain Adaptation for Multimodal 3D Panoptic Segmentation in Autonomous Driving, *Yining Pan, Shijie Li, Yuchen Wu, Xulei Yang, Na Zhao*
414 STUR3D: Spatio-Temporal Unified Representation Learning for 3D Object Detection, *Huijie Fan, Pengrui Huang, Qiang Wang, Baojie Fan, Jiahua Dong, Liangqiong Qu*
415 Exploring 6D Object Pose Estimation with Deformation, *Zhiqiang Liu, Rui Song, Duanmu Chuangqi, Jiaojiao Li, David Ferstl, Yinlin Hu*
416 SearchAD: Large-Scale Rare Image Retrieval Dataset for Autonomous Driving, *Felix Embacher, Jonas Uhrig, Marius Cordts, Markus Enzweiler*
417 Improving Vision-language Models with Perception-centric Process Reward Models, *Yingqian Min, Kun Zhou, Yifan Li, Yuhuan Wu, Han Peng, Yifan Du, Wayne Xin Zhao, Min Yang, Ji-Rong Wen*
418 X-PCR: A Benchmark for Cross-modality Progressive Clinical Reasoning in Ophthalmic Diagnosis, *Gui Wang, Zehao Zhong, YongSong Zhou, Yudong Li, Ende Wu, Wooi Ping Cheah, Rong Qu, Jianfeng Ren, Linlin Shen*
419 Better, Stronger, Faster: Tackling the Trilemma in MLLM-based Segmentation with Simultaneous Textual Mask Prediction, *Jiazhen Liu, Mingkuan Feng, Long Chen*
420 PhysInOne: Visual Physics Learning and Reasoning in One Suite, *Siyuan Zhou, Hejun Wang, Hu Cheng, Jinxi Li, Dongsheng Wang, Junwei Jiang, Yixiao Jin, Jiayue Huang, Shiwei Mao, Shangjia Liu, Yafei Yang, Hongkang Song, Shenxing Wei, Zihui Zhang, DataTeam vLAR, Bing Wang, Zhihua Wang, Chuhang Zou, Bo Yang*
421 AviaSafe: A Physics-Informed Data-Driven Model for Aviation Safety–Critical Cloud Forecasts, *Zijian Zhu, Qiusheng Huang, Anboyu Guo, Xiaohui Zhong, Hao Li*
422 TTRV: Test-Time Reinforcement Learning for Vision Language Models, *Akshit Singh, Shyam Marjit, Wei Lin, Paul Gavrikov, Serena Yeung-Levy, Hilde Kuehne, Rogerio Feris, Sivan Doveh, James Glass, M. Jehanzeb Mirza*
423 Reading or Reasoning? Format Decoupled Reinforcement Learning for Document OCR, *Yufeng Zhong, Lei Chen, Zhixiong Zeng, Xuanle Zhao, Deyang Jiang, Liming Zheng, Jing Huang, Haibo Qiu, Peng Shi, Siqi Yang, Lin Ma*
424 QUANTIPHY: A Quantitative Benchmark Evaluating Physical Reasoning Abilities of Vision-Language Models, *Li Puyin, Tiange Xiang, Ella Mao, Shirley Wei, Xinye Chen, Adnan Masood, Li Fei-Fei, Ehsan Adeli*
425 VisRes Bench: On Evaluating the Visual Reasoning Capabilities of VLMs, *Brigitta Malagurski Törtei, Yasser Dahou, Ngoc Dung Huynh, Wamiq Reyaz Para, Phúc H. Lê Khac, Ankit Singh, Sofian Chaybouti, Sanath Narayan*
426 TRivia: Self-supervised Fine-tuning of Vision-Language Models for Table Recognition, *Junyuan Zhang, Bin Wang, Qintong Zhang, Fan Wu, Zichen Wen, Jialin Lu, Junjie Shan, Ziqi Zhao, Shuya Yang, Ziling Wang, Ziyang Miao, Huaping Zhong, Yuhang Zang, Xiaoyi Dong, Ka-Ho Chow, Conghui He*
427 Urban-GS: A Unified 3D Gaussian Splatting Framework for Compact ❊ and High-Fidelity Aerial-to-Street Reconstruction, *Meng Wang, Changqun Xia, Yuze Wang, Junyi Wang, Wantong Duan, Xinxiong Xie, Yue Qi*
428 Generalizable Sparse-View 3D Reconstruction from Unconstrained Images, *Vinayak Gupta, Chih-Hao Lin, Shenlong Wang, Anand Bhattad, Jia-Bin Huang*
429 RemedyGS: Defend 3D Gaussian Splatting Against Computation Cost Attacks, *Yanping Li, Zhening Liu, Zijian Li, Zehong Lin, Jun Zhang*
430 SparseCam4D: Spatio-Temporally Consistent 4D Reconstruction from Sparse Cameras, *Weihong Pan, Xiaoyu Zhang, Zhuang Zhang, Zhichao Ye, Nan Wang, Haomin Liu, Guofeng Zhang*

431 IDESplat: Iterative Depth Probability Estimation for Generalizable 3D Gaussian Splatting, *Wei Long, Haifeng Wu, Shiyin Jiang, Jinhua Zhang, Xinchun Ji, Shuhang Gu*
432 GS^2: Graph-based Spatial Distribution Optimization for Compact 3D Gaussian Splatting, *Xianben Yang, Tao Wang, Yuxuan Li, Yi Jin, Haibin Ling*
433 OnlinePG: Online Open-Vocabulary Panoptic Mapping with 3D Gaussian Splatting, *Hongjia Zhai, Qi Zhang, Xiaokun Pan, Xiyu Zhang, Yitong Dong, Huaqi Zhang, Dan Xu, Guofeng Zhang*
434 ❊ Uni3R: Unified 3D Reconstruction and Semantic Understanding via Generalizable Gaussian Splatting from Unposed Multi-View Images, *Xiangyu Sun, Haoyi Jiang, Liu Liu, Seungtae Nam, Gyeongjin Kang, Xinjie Wang, Wei Sui, Zhizhong Su, Wenyu Liu, Xinggang Wang, Eunbyung Park*
435 Learning Explicit Continuous Motion Representation for Dynamic Gaussian Splatting from Monocular Videos, *Xuankai Zhang, Junjin Xiao, Shangwei Huang, Wei-shi Zheng, Qing Zhang*
436 MLLMSplat: A 2D MLLM-Powered Framework for 3D Gaussian Splatting Understanding, Generation, and Editing, *Jingqiao Xiu, Can Wang, Dong Xu*
437 Dropping Anchor and Spherical Harmonics for Sparse-view Gaussian Splatting, *Shuangkang Fang, I-Chao Shen, Xuanyang Zhang, Zesheng Wang, Yufeng Wang, Wenrui Ding, Gang YU, Takeo Igarashi*
438 RAP: Fast Feedforward Rendering-Free Attribute-Guided Primitive Importance Score Prediction for Efficient 3D Gaussian Splatting Processing, *Kaifa Yang, Qi Yang, Yiling Xu, Zhu Li*
439 Plug-and-Play PDE Optimization for 3D Gaussian Splatting: Toward High-Quality Rendering and Reconstruction, *Yifan Mo, Youcheng Cai, Ligang Liu*
440 PointGS: Semantic-Consistent Unsupervised 3D Point Cloud Segmentation with 3D Gaussian Splatting, *Yixiao Song, Qingyong Li, Wen Wang, Zhicheng Yan*
441 Scene Grounding in the Wild, *Tamir Cohen, Leo Segre, Shay Shomer-Chai, Shai Avidan, Hadar Averbuch-Elor*
442 ❊ Flow4DGS-SLAM: Optical Flow-Guided 4D Gaussian Splatting SLAM, *Yunsong Wang, Gim Hee Lee*
443 Revisiting 3D Reconstruction Kernels as Low-Pass Filters, *Shengjun Zhang, Min Chen, Yibo Wei, Mingyu Dong, Yueqi Duan*
444 SR3R: Rethinking Super-Resolution 3D Reconstruction With Feed-Forward Gaussian Splatting, *Xiang Feng, Xiangbo Wang, Tieshi Zhong, Chengkai Wang, Yiting Zhao, Tianxiang Xu, Zhenzhong Kuang, Feiwei Qin, Xuefei Yin, Yanming Zhu*
445 GP-4DGS: Probabilistic 4D Gaussian Splatting from Monocular Video via Variational Gaussian Processes, *Mijeong Kim, Jungtaek Kim, Bohyung Han*
446 VisRef: Visual Refocusing while Thinking Improves Test-Time Scaling in Multi-Modal Large Reasoning Models, *Soumya Suvra Ghosal, Youngeun Kim, Zhuowei Li, Ritwick Chaudhry, Linghan Xu, Hongjing Zhang, Jakub Zablocki, Yifan Xing, Qin Zhang*
447 IPR-1: Interactive Physical Reasoner, *Mingyu Zhang, Lifeng Zhuo, Tianxi Tan, Guocan Xie, Xian Nie, Yan Li, Renjie Zhao, Zizhu He, Ziyu Wang, Jiting Cai, Yong-Lu Li*
448 VIRO: Robust and Efficient Neuro-Symbolic Reasoning with Verification for Referring Expression Comprehension, *Hyejin Park, Junhyuk Kwon, Suha Kwak, Jungseul Ok*
449 Fuel Gauge: Estimating Chain-of-Thought Length Ahead of Time in Large Multimodal Models, *Yuedong Yang, Xiwen Wei, Mustafa Munir, Radu Marculescu*
450 Thinking in Dynamics: How Multimodal Large Language Models Perceive, Track, and Reason Dynamics in Physical 4D World, *Yuzhi Huang, Kairun Wen, Rongxin Gao, Dongxuan Liu, Yibin Lou, Jie Wu, Jing Xu, Jian Zhang, Zheng Yang, Yunlong Lin, Chenxin Li, Panwang Pan, Junbin Lu, Jingyan Jiang, Xinghao Ding, Yue Huang, Zhi Wang*
451 Latent Implicit Visual Reasoning, *Kelvin Li, Chuyi Shang, Leonid Karlinsky, Rogerio Feris, Trevor Darrell, Roei Herzig*
452 Thinking with Programming Vision: Towards a Unified View for Thinking with Images, *Zirun Guo, Minjie Hong, Feng Zhang, Kai Jia, Tao Jin*
453 AV-Reasoner: Improving and Benchmarking Clue-Grounded Audio-Visual Counting for MLLMs, *Lidong Lu, Guo Chen, Zhu Wei, Zhiqi Li, Yicheng Liu, Tong Lu*
454 ❊ All Roads Lead to Rome: Incentivizing Divergent Thinking in Vision-Language Models, *Xinyu Tian, Shu Zou, Zhaoyuan Yang, Mengqi He, Peter Tu, Jing Zhang*
455 See Less, See Right: Bi-directional Perceptual Shaping For Multimodal Reasoning, *Shuoshuo Zhang, Yizhen Zhang, Jingjing Fu, Lei Song, Jiang Bian, Yujiu Yang, Rui Wang*
456 Machine Mental Imagery: Empower Multimodal Reasoning with Latent Visual Tokens, *Zeyuan Yang, Xueyang Yu, Delin Chen, Maohao Shen, Chuang Gan*
457 ReaGEN: Adaptive Generation of Structured Chains-of-Thought for Efficient Multimodal Reasoning, *Ruiqing Tian, Mohan Sai Singamsetti, Di Niu, Bahador Rashidi*
458 Breaking the Regional Perception Bottleneck of Multimodal Large Language Models via External Reasoning Framework, *Jinrong Zhang, Zhaoyang Xu, Xusheng He, Xinrui Li, Na Zheng, Jianlong Wu*
459 CodePercept: Code-Grounded Visual STEM Perception for MLLMs, *Tongkun Guan, Zhibo Yang, Jianqiang Wan, Mingkun Yang, Zhentao Guo, Zijian Hu, Ruilin Luo, Ruizhe Chen, Songtao Jiang, Peng Wang, Wei Shen, Junyang Lin, Xiaokang Yang*
460 TableMix: Enhancing Multimodal Table Reasoning in MLLMs from a Data-Centric Perspective, *Chaohu Liu, Shida Wang, Yubo Wang, Linli Xu*
461 Harnessing Chain-of-Thought Reasoning in Multimodal Large Language Models for Face Anti-Spoofing, *Honglu Zhang, Zhiqin Fang, Ningning Zhao, Saihui Hou, Long Ma, Renwang Pei, Zhaofeng He*
462 Grounded Chain-of-Thought for Multimodal Large Language Models, *Qiong Wu, Xiangcong Yang, Yiyi Zhou, Chenxin Fang, Baiyang Song, Xiaoshuai Sun, Rongrong Ji*
463 LS-ViT: Least-Squares Hessian Based Block Reconstruction for Low-Bit Post-Training Quantization of Vision Transformers, *Hyunha Hwang, Xuan Truong Nguyen, Hyuk-Jae Lee*
464 SegMo: Co-Designing Content-Aware Sparsity and Locally-Cohesive Segment Parallelism for Efficient VLM Inference, *Haojuan Li, Ruohan Tang, Dongzhou Cheng, Zongpu Zhang, Jian Li, Jiaqi Wang*
465 Rethinking Asymmetric Quantization: Hidden Symmetry in Vision Model Weights, *Masafumi Mori, Shinya Gongyo, Mitsuru Ambai*
466 ❊ Compressed-Domain-Aware Online Video Super-Resolution, *Yuhang Wang, Hai Li, Shujuan Hou, Zhetao Dong, Xiaoyao Yang*
467 CAR-SAM: Cross-Attention Reconstruction for Post-Training Quantization of the Segment Anything Model, *Houji Wen, Jiangyong Yu, Dawei Yang, Jun Li*
468 Is Bin Generation Indispensable? A Bin-Generation-Free Dataset Quantization via Semantic Perspective, *Maijie Deng, Yuhua Li, Yixiong Zou, Yao Wu, Chenru Ma*
469 ❊ High Resolution Neural Video Coding with Bi-directional Confidence-Guided Reference Information Modeling, *Feng Ye, Kai Zhang, Li Zhang, Chuanmin Jia*
470 Distributed Image Compression with Multimodal Side Information at Extremely Low Bitrates, *Guojun Xu, Mingyang Zhang, Jianwen Xiang, Cheng Tan, Yanchao Yang, Junwei Zhou*
471 Task-Aware Image Signal Processor for Advanced Visual Perception, *Kai Chen, Jin Xiao, Leheng Zhang, Kexuan Shi, Shuhang Gu*
472 ❊ Enhancing Video Vision Language Model with Hippocampal Sensing, *Xu Cao*
473 ❊ VIRD: View-Invariant Representation through Dual-Axis Transformation for Cross-View Pose Estimation, *Juhye Park, Wooju Lee, Dasol Hong, Changki Sung, Youngwoo Seo, Dongwan Kang, Hyun Myung*
474 WRIVINDER: Towards Spatial Intelligence for Geo-locating Ground Images onto Satellite Imagery, *Chandrakanth Gudavalli, Tajuddin Manhar Mohammed, Abhay Yadav, Ananth Vishnu Bhaskar, Hardik Prajapati, Cheng Peng, Rama Chellappa, Shivkumar Chandrasekaran, B.S. Manjunath*
475 SoPE: Spherical Coordinate-Based Positional Embedding for Enhancing Spatial Perception of 3D LVLMs, *Koonting Yip, Qiyan Zhao, Wenhao Yu, Liangyu Yuen, Mingkai Li, Xiaofeng Zhang, Jianmin Ji, Yanyong Zhang, Qing Jiang, Ka-Veng Yuen*
476 RHO: Robust Holistic OSM-Based Metric Cross-View Geo-Localization, *Junwei Zheng, Ruize Dai, Ruiping Liu, Zichao Zeng, Yufan Chen, Fangjinhua Wang, Kunyu Peng, Kailun Yang, Jiaming Zhang, Rainer Stiefelhagen*

477 EfficientVPR: Toward Efficient Visual Place Recognition via Scene-Aware Prompt Tuning and Adaptive Feature Enhancement, *Wenjing Tang, Chuanguang Yang, Zhulin An, Libo Huang, Boyu Diao, Yongjun Xu*
478 Universal Guideline-Driven Image Clustering via a Hybrid LLM Agent, *Wenliang Zhong, Rob Barton, Lucas Goncalves, Kushal Kumar, Feng Jiang, Hehuan Ma, Yuzhi Guo, Vidit Bansal, Karim Bouyarmane, Junzhou Huang*
479 ReLaX: Reasoning with Latent Exploration for Large Reasoning Models, *Shimin Zhang, Xianwei Chen, Yufan Shen, Ziyuan Ye, Jibin Wu*
480 VideoChat-M1: Collaborative Policy Planning for Video Understanding via Multi-Agent Reinforcement Learning, *Boyu Chen, Zikang Wang, Zhengrong Yue, Kainan Yan, Chenyun Yu, Yi Huang, Zijun Liu, Yafei Wen, Xiaoxin Chen, Yang Liu, Peng Li, Yali Wang*
481 Think, Then Verify: A Hypothesis–Verification Multi-Agent Framework for Long Video Understanding, *Zheng Wang, Haoran Chen, Haoxuan Qin, Zhipeng Wei, Tianwen Qian, Cong Bai*
482 Reinforce to Learn, Elect to Reason: A Dual Paradigm for Video ❊ Reasoning, *Songyuan Yang, Weijiang Yu, Jilin Ma, Ziyu Liu, Guijian Tang, Wenjing Yang, Huibin Tan, Nong Xiao*
483 Graph-to-Frame RAG: Visual-Space Knowledge Fusion for Training-Free and Auditable Video Reasoning, *Songyuan Yang, Weijiang Yu, Ziyu Liu, Guijian Tang, Wenjing Yang, Huibin Tan, Nong Xiao*
484 LongVT: Incentivizing "Thinking with Long Videos" via Native Tool Calling, *Zuhao Yang, Sudong Wang, Kaichen Zhang, Keming Wu, Sicong Leng, Yifan Zhang, Bo Li, Chengwei Qin, Shijian Lu, Xingxuan Li, Lidong Bing*
485 Multi-Modal Image Fusion via Intervention-Stable Feature Learning, *Xue Wang, Zheng Guan, Wenhua Qian, Chengchao Wang, Runzhuo Ma*
486 ReCoFuse: Ultra-Robust Image Fusion via Restorative Multi-Modal Diffusion Reciprocal Coupling, *Hao Zhang, Shuhan Yang, Linfeng Tang, Xunpeng Yi, Jiayi Ma*
487 Degradation-Robust Fusion: An Efficient Degradation-Aware Diffusion Framework for Multimodal Image Fusion in Arbitrary Degradation Scenarios, *Yu Shi, Yu Liu, Zhong-Cheng Wu, Juan Cheng, Huafeng Li, Xun Chen*
488 DF^2-VB: Dual-level Fuzzy Fusion with View-specific Boosting for Multi-view Multi-label Classification, *Yuena Lin, Haichun Cai, Yi Shan, Hao Wei, Yongjian Deng, Zhen Yang, Gengyu Lyu*
489 UniFusion: A Unified Image Fusion Framework with Robust ❊ Representation and Source-Aware Preservation, *Xingyuan Li, Songcheng Du, Yang Zou, Haoyuan Xu, Zhiying Jiang, Jinyuan Liu*
490 Self-guided Semantic Inspection for Zero-Shot Composed Image Retrieval, *Jingjing Zhang, Lei Zhang, Zheren Fu, Bo Hu, Zhendong Mao*
491 G-MIXER: Geodesic Mixup-based Implicit Semantic Expansion and Explicit Semantic Re-ranking for Zero-Shot Composed Image Retrieval, *Jiyoung Lim, Heejae Yang, Jee-Hyong Lee*
492 No Hard Negatives Required: Concept Centric Learning Leads to Compositionality without Degrading Zero-shot Capabilities of Contrastive Models, *Hai X. Pham, David T. Hoffmann, Ricardo Guerrero, Brais Martinez*
493 MUSE: Harnessing Precise and Diverse Semantics for Few-Shot Whole Slide Image Classification, *Jiahao Xu, Sheng Huang, Xin Zhang, Zhixiong Nan, Jiajun Dong, Nankun Mu*
494 Pointing at Parts: Training-Free Few-Shot Grounding in Multimodal LLMs, *Shiang-Feng Tsai, Yuan-Hong Liao, Jin-Cheng Jhang, Nan Qiao, Min Sun*
495 Graph Attention Prototypical Network for Robust Few-Shot Classification, *Tingyun Liu, Licheng Liu, Qibin Zhang, Qiying Feng, C. L. Philip Chen*
496 Mitigating The Distribution Shift of Diffusion-based Dataset Distillation, *Yue Xu, Chenyu Hu, Pengyu An, Yong-Lu Li*
497 EVLF: Early Vision-Language Fusion for Generative Dataset Distillation, *Wenqi Cai, Yawen Zou, Guang Li, Chunzhi Gu, Chao Zhang*
498 Fixed Anchors Are Not Enough: Dynamic Retrieval and Persistent Homology for Dataset Distillation, *Muquan Li, Hang Gou, Yingyi Ma, Rongzheng Wang, Ke Qin, Tao He*
499 Flow Map Distillation Without Data, *Shangyuan Tong, Nanye Ma, Saining Xie, Tommi Jaakkola*
500 F^2HDR: Two-Stage HDR Video Reconstruction via Flow Adapter and Physical Motion Modeling, *Huanjing Yue, Dawei Li, Shaoxiong Tu, Jingyu Yang*
501 Learning Latent Transmission and Glare Maps for Lens Veiling Glare ❊ Removal, *Xiaolong Qian, Qi Jiang, Lei Sun, Zongxi Yu, Kailun Yang, Peixuan Wu, Jiacheng Zhou, Yao Gao, Yaoguang Ma, Ming-Hsuan Yang, Kaiwei Wang*
502 Inter-Photon-Limited Videography, *Andrew Xie, Dongyu Du,* ❊ *Sotiris Nousias, David B. Lindell, Kiriakos N. Kutulakos*
503 A Bit is All You Need! Efficient Video Capture via Single Bit Imaging, *Kanchana Vaishnavi Gandikota, Michael Moeller, Andreas Kolb, Bhaskar Choubey, Paramanand Chandramouli*
504 From Events to Clarity: The Event-Guided Diffusion Framework for Dehazing, *Ling Wang, Yunfan Lu, Wenzong Ma, Huizai Yao, Pengteng Li, Hui Xiong*
505 Electromagnetic Inverse Scattering from a Single Transmitter, ❊ *Yizhe Cheng, Chunxun Tian, Haoru Wang, Wentao Zhu, Xiaoxuan Ma, Yizhou Wang*
506 Statistical Characteristic-Guided Denoising for Rapid High-Resolution Transmission Electron Microscopy Imaging, *Hesong Li, Ziqi Wu, Ruiwen Shao, Ying Fu*
507 Physics-Guided Multistep Deformation Reversal for Ancient Bamboo Slip Restoration, *Qianqian Tang, Jinchi Zhu, Xiaolu Zhou, Yongchao Xu*
508 cryoSENSE: Compressive Sensing Enables High-throughput Microscopy with Sparse and Generative Priors on the Protein Cryo-EM Image Manifold, *Zain Shabeeb, Daniel Saeedi, Darin Tsui, Vida Jamali, Amirali Aghazadeh*
509 SGDE: Self-supervised Geometry Degradation Estimation Framework for Coded Aperture Compressive Spectral Imaging, *Yuqiao He, Xiaoyan Liu, Jianxu Mao, Yaonan Wang, Hui Zhang, Lizhu Liu, Yurong Chen, Wenbin He*
510 Factorized Context Aggregation for Robust Cancer Risk Estimation via Soft Re-Ranked Retrieval and Hierarchical Anchors, *Puria Azadi Moghadam, Ali Khajegili Mirabadi, Behnam Maneshgar, Hossein Farahani, Ali Bashashati*
511 UniMERNet: A Universal Network for Real-World Mathematical Expression Recognition, *Zhuangcheng Gu, Guang Liang, Bin Wang, Zhiyuan Zhao, Qintong Zhang, Weijia Li, Chao Xu, Bo Zhang, Botian Shi, Jiang Wu, Wentao Zhang, Conghui He*
512 GeneVAR: Causal MeanFlow for Autoregressive Gene-to-WSI Tile Synthesis, *Jianwei Zhao, Fan Yang, Xin Li, Qiang Zhai, Ao Luo, Ziqi Ren, Zhicheng Jiao, Hong Cheng*
513 Depth Any Endoscopy: Towards Self-Supervised Generalizable Depth ❊ Estimation in Monocular Endoscopy, *Shuwei Shao, Kejin Zhu, Shixing Ma, Xinzhe Du, Baochang Zhang, Zhe Min*
514 RoSAMDepth: Robust Self-supervised Depth Estimation Leveraging Segment Anything Model, *Xuanang Gao, Zhiwei Ning, Gengming Zhang, Jiaxi Cao, Runze Yang, Zhonglong Zheng, Jie Yang, Rong Xiao, Wei Liu*
515 AdaSFormer: Adaptive Serialized Transformers for Monocular Semantic Scene Completion from Indoor Environments, *Xuzhi Wang, Xinran Wu, Song Wang, Lingdong Kong, Ziping Zhao*
516 Dark3R: Learning Structure from Motion in the Dark, *Andrew Y. Guo,* ❊ *Anagh Malik, SaiKiran Tedla, Yutong Dai, Yiqian Qin, Zach Salehe, Benjamin Attal, Sotiris Nousias, Kiriakos N. Kutulakos, David B. Lindell*
517 What Makes Good Synthetic Training Data for Zero-Shot Stereo ❊ Matching?, *David Yan, Alexander Raistrick, Jia Deng*
518 TR2M: Transferring Monocular Relative Depth to Metric Depth with Language Descriptions and Dual-Level Scale-Oriented Contrast, *Beilei Cui, Yiming Huang, Long Bai, Hongliang Ren*
519 Iris: Integrating Language into Diffusion-based Monocular Depth Estimation, *Ziyao Zeng, Jingcheng Ni, Daniel Wang, Patrick Rim, Younjoon Chung, Fengyu Yang, Byung-Woo Hong, Alex Wong*
520 Ov3R: Open-Vocabulary Semantic 3D Reconstruction from RGB ❊ Videos, *Ziren Gong, Xiaohan Li, Fabio Tosi, Jiawei Han, Stefano Mattoccia, Jianfei Cai, Matteo Poggi*
521 M3DLayout: A Multi-Source Dataset of 3D Indoor Layouts and ❊ Structured Descriptions for 3D Generation, *Yiheng Zhang, Zhuojiang Cai, Mingdao Wang, Meitong Guo, Tianxiao Li, Li Lin, Yuwang Wang*
522 UniPart: Part-Level 3D Generation with Unified 3D Geom–Seg ❊ Latents, *Xufan He, Yushuang Wu, Xiaoyang Guo, Chongjie Ye, Jiaqing Zhou, Tianlei Hu, Xiaoguang Han, Dong Du*
523 Photo3D: Advancing Photorealistic 3D Generation through Structure-Aligned Detail Enhancement, *Xinyue Liang, Zhiyuan Ma, Lingchen Sun, Yanjun Guo, Lei Zhang*

524 Mesh-Pro: Asynchronous Advantage-guided Ranking Preference Optimization for Artist-style Quadrilateral Mesh Generation, *Zhen Zhou, Jian Liu, Biwen Lei, Jing Xu, Haohan Weng, Yiling Zhu, Zhuo Chen, Junfeng Fan, Yunkai Ma, Dazhao Du, Song Guo, Fengshui Jing, Chunchao Guo*

525 Order Matters: 3D Shape Generation from Sequential VR Sketches, *Yizi Chen, Sidi Wu, Tianyi Xiao, Nina Wiedemann, Loic Landrieu*

526 Think-Then-Generate: Structural Chain-of-Thought Reasoning for ❊ Consistent 3D Generation, *Xinyue Liu, Jin Liu, Hongbo Wang, Ran He, Huaibo Huang*

527 ArtLLM: Generating Articulated Assets via 3D LLM, *Penghao Wang, Siyuan Xie, Hongyu Yan, Xianghui Yang, Jingwei Huang, Chunchao Guo, Jiayuan Gu*

528 PoseMaster: A Unified 3D Native Framework for Stylized Pose Generation, *Hongyu Yan, Kunming Luo, Weiyu Li, Kaiyi Zhang, Yixun Liang, Jingwei Huang, Chunchao Guo, Ping Tan*

529 2D-LFM: Lifting Foundation Model without 3D Supervision, ❊ *Mosam Dabhi, Irhas Gill, László A. Jeni, Simon Lucey*

530 ActionMesh: Animated 3D Mesh Generation with Temporal 3D Diffusion, *Remy Sabathier, David Novotny, Niloy J. Mitra, Tom Monnier*

531 4DWorldBench: A Comprehensive Evaluation Framework for 3D/4D World Generation Models, *Yiting Lu, Wei Luo, Peiyan Tu, Haoran Li, Hanxin Zhu, Zihao Yu, Xingrui Wang, Xinyi Chen, Xinge Peng, Xin Li, Zhibo Chen*

532 FabricGen: Microstructure-Aware Woven Fabric Generation, ❊ *Yingjie Tang, Di Luo, Zixiong Wang, Xiaoli Ling, Jian Yang, Beibei Wang*

533 Leveraging Verifier-Based Reinforcement Learning in Image Editing, *Hanzhong Guo, Jie Wu, Jie Liu, Yu Gao, Zilyu Ye, Linxiao Yuan, Xionghui Wang, Yizhou Yu, Weilin Huang*

534 PaCo-RL: Advancing Reinforcement Learning for Consistent Image Generation with Pairwise Reward Modeling, *Bowen Ping, Chengyou Jia, Minnan Luo, Changliang Xia, Xin Shen, Zhuohang Dang, Hangwei Qian*

535 VIVA: VLM-Guided Instruction-Based Video Editing with Reward Optimization, *Xiaoyan Cong, Haotian Yang, Angtian Wang, Yizhi Wang, Yiding Yang, Canyu Zhang, Chongyang Ma*

536 MapReduce LoRA: Advancing the Pareto Front in Multi-Preference ❊ Optimization for Generative Models, *Chieh-Yun Chen, Zhonghao Wang, Qi Chen, Zhifan Ye, Min Shi, Yue Zhao, Yinan Zhao, Hui Qu, Wei-An Lin, Yiru Shen, Ajinkya Kale, Irfan Essa, Humphrey Shi*

537 Reward Forcing: Efficient Streaming Video Generation with Rewarded ❊ Distribution Matching Distillation, *Yunhong Lu, Yanhong Zeng, Haobo Li, Hao Ouyang, Qiuyu Wang, Ka Leong Cheng, Jiapeng Zhu, Hengyuan Cao, Zhipeng Zhang, Xing Zhu, Yujun Shen, Min Zhang*

538 C^2FG: Control Classifier-Free Guidance via Score Discrepancy ❊ Analysis, *Jiayang Gao, Tianyi Zheng, Jiayang Zou, Fengxiang Yang, Shice Liu, Luyao Fan, Zheyu Zhang, Hao Zhang, Jinwei Chen, Peng-Tao Jiang, Bo Li, Jia Wang*

539 Learning What to Trust: Bayesian Prior-Guided Optimization for Visual Generation, *Ruiying Liu, Yuanzhi Liang, Haibin Huang, Tianshu Yu, Chi Zhang*

540 Unified Customized Generation by Disentangled Reward Modeling, *Shaojin Wu, Mengqi Huang, Yufeng Cheng, Wenxu Wu, Jiahe Tian, Yiming Luo, Fei Ding, Qian He*

541 Region-Aware Instance Consistency Learning for Micro-Expression Recognition, *Yaomin Cai, C. L. Philip Chen, Shiting Xu, Haiqi Liu, Tong Zhang*

542 MPL: Match-guided Prototype Learning for Few-shot Action Recognition, *Feng Yang, Jie Zhao, Fulin Luo, Anyong Qin, Tiecheng Song, Yue Zhao, Chenqiang Gao, Junwei Han*

543 LaDy: Lagrangian-Dynamic Informed Network for Skeleton-based Action Segmentation via Spatial-Temporal Modulation, *Haoyu Ji, Xueting Liu, Yu Gao, Wenze Huang, Zhihao Yang, Weihong Ren, Zhiyong Wang, Honghai Liu*

544 LA-Pose: Latent Action Pretraining Meets Pose Estimation, *Zhengqing Wang, Saurabh Nair, Prajwal Chidananda, Pujith Kachana, Samuel Li, Matthew Brown, Yasutaka Furukawa*

545 RAAS: LLM Agentic System Architecture Search with GRPO, *Jiayi Yang, Guancheng Wan, Man Zhang, Mang Ye*

546 Temporal Representation Enhancement (TRE): Learning to Forget ❊ Dominant Patterns for Enhanced Temporal Spiking Features, *Wei Liu, Li Yang, Yufei Wang, Han Xiao, Boyu Cai, Weiming Hu*

547 Chain-of-Models Pre-Training: Rethinking Training Acceleration of Vision Foundation Models, *Jiawei Fan, Shigeng Wang, Chao Li, Xiaolong Liu, Anbang Yao*

548 Unlocking Pre-trained Weights: Parameter Inheritance for Zero-Shot Initialization, *Jiaze Xu, Shiyu Xia, Jiaqi Lv, Xin Geng*

549 Deconstructing the Failure of Ideal Noise Correction: A Three-Pillar Diagnosis, *Chen Feng, Zhuo Zhi, Zhao Huang, Jiawei Ge, Ling Xiao, Nicu Sebe, Georgios Tzimiropoulos, Ioannis Patras*

550 Progressive Neural Architecture Generation, *Caiyang Yu, Chen Huang, Yun Liu, Chenwei Tang, Wei Ju, Jiancheng Lv*

551 A Unified Framework for Knowledge Transfer in Bidirectional Model Scaling, *Jianlu Shen, Fu Feng, Jiaze Xu, Yucheng Xie, Jiaqi Lv, Xin Geng*

552 When Do Models Actually Decide? Mapping the Layer-Wise Decision ❊ Timeline in Pretrained Neural Networks, *Minhyeok Lee*

553 Temporal Interaction in Spiking Transformers with Multi-Delay Mixer, *Kexin Shi, Hanwen Liu, Zeyang Song, Yang Liu, Jieyuan Zhang, Shuai Wang, Jibin Wu, Malu Zhang, Yang Yang*

554 Consensus vs. Controversy: Mapping the Decision Space Where ❊ Architectures Diverge, *Minhyeok Lee*

555 Sparsely Timing the Change: A Spiking Temporal Framework for Remote Sensing Interpretation, *Shilong Li, Xiurui Xie, Qiugang Zhan, Luochao Wang, Yong Deng, Guisong Liu*

556 ProSoftArena: Benchmarking Hierarchical Capabilities of Multi-modal Agents in Professional Software Environments, *Jiaxin Ai, Yukang Feng, Fanrui Zhang, Jianwen Sun, Zizhen Li, Chuanhao Li, Yifan Chang, Wenxiao Wu, Ruoxi Wang, Mingliang Zhai, Kaipeng Zhang*

557 BAMI: Training-Free Bias Mitigation in GUI Grounding, *Borui Zhang, Bo Zhang, Bo Wang, Wenzhao Zheng, Yuhao Cheng, Liang Tang, Yiqiang Yan, Jie Zhou, Jiwen Lu*

558 DRS-GUI: Dynamic Region Search for Training-Free GUI Grounding, *Yichao Liu, Huawen Shen, Liu Yu, Shiyu Liu, Zeyu Chen, Yu Zhou*

559 Consistency Beyond Contrast: Enhancing Open-Vocabulary Object Detection Robustness via Contextual Consistency Learning, *Bozhao Li, Shaocong Wu, Tong Shao, Senqiao Yang, Qiben Shan, Zhuotao Tian, Jingyong Su*

560 Thermal-Det: Language-Guided Cross-Modal Distillation for Open-Vocabulary Thermal Object Detection, *Yasiru Ranasinghe, Elim Schenck, Florence Yellin, Shuowen Hu, Christopher Funk, Vishal M. Patel*

561 Geometry-driven OOD Detectors Are Class-Incremental Learners, *Wangwang Jia, Zijian Gao, Tianjiao Wan, Yuan Cao, Yong Dou, Kele Xu*

562 Mind the Way You Select Negative Texts: Pursuing the Distance Consistency in OOD Detection with VLMs, *Zhikang Xu, Qianqian Xu, Zitai Wang, Cong Hua, Sicong Li, Zhiyong Yang, Qingming Huang*

563 Prompt-Free Unknown Label Generation for Open World Detection in Remote Sensing, *Abdullah Azeem, Ruisheng Wang, Qingquan Li, Abubakar Siddique*

564 Learning to Diversify and Focus: A Reinforcement Framework for ❊ Open-Vocabulary HOI Detection, *Yongchao Xu, Jiawei Liu, Junfeng Wang, Sen Tao, Na Jiang, Zheng-Jun Zha*

565 RINO: Rotation-Invariant Non-Rigid Correspondences, *Maolin Gao,* ❊ *Shao Jie Hu-Chen, Congyue Deng, Riccardo Marin, Leonidas Guibas, Daniel Cremers*

566 Hyperbolic Prototype Learning with Uncertainty-Aware Consistency for Continual Test-Time Segmentation, *Siddhant Gole, Akash Pal, Amit More, S Divakar Bhat, Subhasis Chaudhuri, Biplab Banerjee*

567 DINO Eats CLIP: Adapting Beyond Knowns for Open-set 3D Object Retrieval, *Xinwei He, Yansong Zheng, Qianru Han, Zhichuan Wang, Yuxuan Cai, Yang Zhou, Jingbo Xia, Yulong Wang, Jinhai Xiang, Xiang Bai*

568 Leveraging Class Distributions in CLIP for Weakly Supervised Semantic Segmentation, *Ziqian Yang, Xinqiao Zhao, Xiaolei Wang, Quan Zhang, Jimin Xiao*

569 CompetitorFormer: Mitigating Query Conflicts for 3D Instance Segmentation via Competitive Strategy, *Duanchu Wang, Junjie Yang, Haoran Gong, Jing Liu, Di Wang*

570 D2Dewarp: Dual Dimensions Geometric Representation Learning Based Document Image Dewarping, *Heng Li, Xiangping Wu, Qingcai Chen*
571 Discover, Segment, and Select: A Progressive Mechanism for Zero-shot Camouflaged Object Segmentation, *Yilong Yang, Jianxin Tian, Shengchuan Zhang, Liujuan Cao*
572 ❊ D-Convexity: A Unified Differentiable Convex Shape Prior via Quasi-Concavity for Data-driven Image Segmentation, *Shengzhe Chen, Hao Yan*
573 Fast Reasoning Segmentation for Images and Videos, *Yiqing Shen, Mathias Unberath*
574 Structure-Aware Representation Distillation for Tiny-Dense Object Segmentation, *Xuesong Liu, Anke Xu, Wenbo Cao, Emmett Ientilucci*
575 CRFT: Consistent–Recurrent Feature Flow Transformer for Cross-Modal Image Registration, *Xuecong Liu, Mengzhu Ding, Zixuan Sun, Zhang Li, Xichao Teng*
576 FireScope: Wildfire Risk Raster Prediction With a Chain-of-Thought Oracle, *Mario Markov, Stefan Ailuro, Luc Van Gool, Konrad Schindler, Danda Pani Paudel*
577 OlmoEarth: Stable Latent Image Modeling for Multimodal Earth Observation, *Henry Herzog, Favyen Bastani, Yawen Zhang, Gabriel Tseng, Joseph Redmon, Hadrien Sablon, Ryan Park, Jacob Morrison, Alexandra Buraczynski, Karen Farley, Josh Hansen, Andrew Howe, Patrick Alan Johnson, Mark Otterlee, Ted Schmitt, Hunter Pitelka, Stephen Daspit, Rachel Ratner, Christopher Wilhelm, Sebastian Wood, Mike Jacobi, Hannah Kerner, Evan Shelhamer, Ali Farhadi, Ranjay Krishna, Patrick Beukema*
578 TESSERA: Temporal Embeddings of Surface Spectra for Earth Representation and Analysis, *Zhengpeng Feng, Clement Atzberger, Sadiq Jaffer, Jovana Knezevic, Silja Sormunen, Robin Young, Madeline C. Lisaius, Markus Immitzer, Toby Jackson, James Ball, David A. Coomes, Anil Madhavapeddy, Andrew Blake, Srinivasan Keshav*
579 Regulating Rather than Constraining: Adaptive Guidance for Complex Spectral Reconstruction in Pansharpening, *Zhuwei Wen, Zimin Xia, He Chen, Linwei Yue, Xianwei Zheng*
580 ❊ GeoMMBench and GeoMMAgent: Toward Expert-Level Multimodal Intelligence in Geoscience and Remote Sensing, *Aoran Xiao, Shihao Cheng, Yonghao Xu, Yexian Ren, Hongruixuan Chen, Naoto Yokoya*
581 ❊ Revisiting the Necessity of Full Accuracy: Weakly Supervised Object-Level Offset Correction for Misaligned Building Labels, *Junda Xu, Yanmeng Liu, Xiangqiang Zeng, Jinrong Wu, Ying Qu, Libao Zhang*
582 UniGeoSeg: Towards Unified Open-World Segmentation for Geospatial Scenes, *Shuo Ni, Di Wang, He Chen, Haonan Guo, Ning Zhang, Jing Zhang*
583 ❊ ZoomEarth: Active Perception for Ultra-High-Resolution Geospatial Vision-Language Tasks, *Ruixun Liu, Bowen Fu, Jiayi Song, Kaiyu Li, Wanchen Li, Lanxuan Xue, Hui Qiao, Weizhan Zhang, Deyu Meng, Xiangyong Cao*
584 Unleashing Stealthy Backdoor Pandemic by Infecting a Single Diffusion Model, *Mohaiminul Al Nahian, Abeer Matar Almalky, Sabbir Ahmed, Abdullah Al Arafat, Mamshad Nayeem Rizve, Adnan Siraj Rakin*
585 Taming the Long Tail: Rebalancing Adversarial Training via Adaptive Perturbation, *Lilin Zhang, Yimo Guo, Yue Li, Jiancheng Shi, Xianggen Liu*
586 Robustness Under Data Scarcity: Few-Shot Continual Adversarial Training for Evolving Threats, *Wenxuan Wang, Chenglei Wang, Chengzhi Yan, Xuelin Qian, Yanning Zhang*
587 Logit-Margin Repulsion for Backdoor Defense, *Zhiguo Yang, Dongsheng Xu, Ruizhi Zhong, Jiacheng Pi, Xingxing Huang, Wenjie Ruan*
588 ❊ Thermally Activated Dual-Modal Adversarial Clothing against AI Surveillance Systems, *Jiahuan Long, Tingsong Jiang, Hanqing Liu, Chao Ma, Weien Zhou, Yang Yang, Wen Yao*
589 ❊ Immunizing Models Against Harmful Long-Horizon Fine-Tuning via Contractive Optimization Dynamics, *Najibul Haque Sarker, Zaber Ibn Abdul Hakim, Ali Asgarov, Chia-Wei Tang, Alvi Md Ishmam, Chris Thomas*
590 Towards Stealthy and Effective Backdoor Attacks on Lane Detection: A Naturalistic Data Poisoning Approach, *Yifan Liao, Yuxin Cao, Yedi Zhang, Wentao He, Yan Xiao, Xianglong Du, Zhiyong Huang, Jin Song Dong*
591 Red-teaming Retrieval-Augmented Diffusion Models via Poisoning Knowledge Bases, *Xinqi Lyu, Yihao Liu, Dong Wang, Bin Xiao*
592 Latent Diffusion Inversion Requires Understanding the Latent Space, *Mingxing Rao, Bowen Qu, Daniel Moyer*
593 ❊ Fractal Camouflage: A Bio-Inspired Approach for Multi-Scale Adversarial Attacks in the Infrared Domain, *Chengyin Hu, Xin Wang, Rui Qiu, Zhe Jia, Yingying Zhao, Kai Wang, Xu Kang, Yiwei Wei*
594 EgoRoC: Towards Egocentric Robotic Control via Task-Agnostic Visual Alignment, *Wei Feng, Chi Zhang, Nan Li, Qian Zhang, Qi Zhang, Mingyan Li*
595 Describe Anything Anywhere At Any Moment, *Nicolas Gorlo, Lukas Schmid, Luca Carlone*
596 ❊ StaMo: Unsupervised Learning of Generalizable Robot Motion from Compact State Representation, *Mingyu Liu, Jiuhe Shu, Hui Chen, Zeju Li, Canyu Zhao, Jiange Yang, Shenyuan Gao, Hao Chen, Chunhua Shen*
597 VLA Models Are More Generalizable Than You Think: Revisiting Physical and Spatial Modeling, *Weiqi Li, Quande Zhang, Ruifeng Zhai, Liang Lin, Guangrun Wang*
598 Action–Geometry Prediction with 3D Geometric Prior for Bimanual Manipulation, *Chongyang Xu, Haipeng Li, Shen Cheng, Haoqiang Fan, Ziliang Feng, Shuaicheng Liu*
599 Joint-Aligned Latent Action: Towards Scalable VLA Pretraining in the Wild, *Hao Luo, Ye Wang, Wanpeng Zhang, Haoqi Yuan, Yicheng Feng, Haiweng Xu, Sipeng Zheng, Zongqing Lu*
600 Rethinking Camera Choice: An Empirical Study on Fisheye Camera Properties in Robotic Manipulation, *Han Xue, Nan Min, Xiaotong Liu, Wendi Chen, Yuan Fang, Jun Lv, Cewu Lu, Chuan Wen*
601 INSIGHT Bench: Towards Grounded IN-SItu Guidance for Robotic ManipulaTion, *Seonho Kim, Junhyeong Hong, Kyungjae Lee, Yoonseon Oh*
602 MM-ACT: Learn from Multimodal Parallel Generation to Act, *Haotian Liang, Xinyi Chen, Bin Wang, Mingkang Chen, Yitian Liu, Yuhao Zhang, Zanxin Chen, Tianshuo Yang, Yilun Chen, Jiangmiao Pang, Dong Liu, Xiaokang Yang, Yao Mu, Wenqi Shao, Ping Luo*
603 HQC-NBV: A Hybrid Quantum-Classical View Planning Approach, *Xiaotong Yu, Chang Wen Chen*
604 Motus: A Unified Latent Action World Model, *Hongzhe Bi, Hengkai Tan, Shenghao Xie, Zeyuan Wang, Shuhe Huang, Haitian Liu, Ruowen Zhao, Yao Feng, Chendong Xiang, Yinze Rong, Hongyan Zhao, Hanyu Liu, Zhizhong Su, Lei Ma, Hang Su, Jun Zhu*
605 SE(3)-Equivariance with Geometric and Topological Guidance for Category-Level Object Pose Estimation, *Sheng Yu, Di-Hua Zhai, Yuanqing Xia*
606 SPEAR-1: Scaling Beyond Robot Demonstrations via 3D Understanding, *Nikolay Nikolov, Giuliano Albanese, Sombit Dey, Aleksandar Yanev, Luc Van Gool, Jan-Nico Zaech, Danda Pani Paudel*
607 Global Prior Meets Local Consistency: Dual-Memory Augmented Vision-Language-Action Model for Efficient Robotic Manipulation, *Zaijing Li, Bing Hu, Rui Shao, Gongwei Chen, Dongmei Jiang, Pengwei Xie, Jianye Hao, Liqiang Nie*
608 RoboTAG: End-to-end Robot Pose Estimation via Topological Alignment Graph, *Yifan Liu, Fangneng Zhan, Wanhua Li, Haowen Sun, Katerina Fragkiadaki, Hanspeter Pfister*
609 MVLM: Template-Free Tracking via Vision–Language Margin Confidence and Memory-Gated Tracking, *Dae-Hyeon Park, Mina Baek, Jeong-Hun Ha, Chan-Seop Park, Jamshidjon Ganiev, Seung-Hwan Bae*
610 Interactive Tracking: A Human-in-the-Loop Paradigm with Memory-Augmented Adaptation, *Yuqing Huang, Guotian Zeng, Zhenqiao Yuan, Zhenyu He, Xin Li, Yaowei Wang, Ming-Hsuan Yang*
611 VidEoMT: Your ViT is Secretly Also a Video Segmentation Model, *Narges Norouzi, Idil Esen Zulfikar, Niccolò Cavagnero, Tommie Kerssies, Bastian Leibe, Gijs Dubbelman, Daan de Geus*
612 Matching Every Pair to Track Every Point: PairFormer for All-Pairs Tracking and Video Trajectory Fields, *Guangyang Wu, Youran Ding, Xinyu Che, Benyuan Sun, Yi Yang, Xiaohong Liu*
613 Boosting Self-Supervised Tracking with Contextual Prompts and Noise Learning, *Yaozong Zheng, Qihua Liang, Bineng Zhong, Shuimu Zeng, Yuanliang Xue, Ning Li, Shuxiang Song*
614 Progressive Multi-cue Alignment for Unaligned RGBT Tracking,

Jiandong Jin, Chenglong Li, Hao Feng, Andong Lu, Lili Huang, Jin Tang

615 Real-Time Neural Video Compression with Unified Intra and Inter ❊ Coding, *Hui Xiang, Yifan Bian, Li Li, Jingran Wu, Xianguo Zhang, Dong Liu*

616 Adapting Lightweight Image-based Counting Models for Video ❊ Crowd Counting, *Weibo Shu, Antoni B. Chan*

617 Sparse Task Vector Mixup with Hypernetworks for Efficient Knowledge Transfer in Whole-Slide Image Prognosis, *Pei Liu, Xiangxiang Zeng, Tengfei Ma, Yucheng Xing, Xuanbai Ren, Yiping Liu*

618 MedTVT-R1: A Multimodal LLM Empowering Medical Reasoning and Diagnosis, *Yuting Zhang, Kaishen Yuan, Hao Lu, Yutao Yue, Jintai Chen, Kaishun Wu*

619 MedKCO: Medical Vision-Language Pretraining via Knowledge-Driven Cognitive Orchestration, *Chenran Zhang, Ruiqi Wu, Tao Zhou, Yi Zhou*

620 Toward Generalizable Whole Brain Representations with High-Resolution Light-Sheet Data, *Minyoung E. Kim, Dae Hee Yun, Aditi V. Patel, Madeline Hon, Webster Guan, Taegeon Lee, Brian Nguyen*

621 CryoHype: Reconstructing a thousand cryo-EM structures with transformer-based hypernetworks, *Jeffrey Gu, Minkyu Jeon, Ambri Ma, Serena Yeung-Levy, Ellen D. Zhong*

622 GenTract: Generative Global Tractography, *Alec Sargood, Lemuel* ❊ *Puglisi, Elinor Thompson, Mirco Musolesi, Daniel C. Alexander*

623 LUMINA: A Multi-Vendor Mammography Benchmark with Energy Harmonization Protocol, *Hongyi Pan, Gorkem Durak, Halil Ertugrul Aktas, Andrea M. Bejar, Baver Tutun, Emre Uysal, Ezgi Bulbul, Mehmet Fatih Dogan, Berrin Erok, Berna Akkus Yildirim, Sukru Mehmet Erturk, Ulas Bagci*

624 Virtual Immunohistochemistry Staining with Dual-Aligned Multi-Task Feature Guidance, *Shigeng Xie, Hongming Xu, Guiyang Jiang, Tuomo Rossi, Tommi Kärkkäinen, Fengyu Cong*

625 Can Natural Image Autoencoders Compactly Tokenize fMRI Volumes for Long-Range Dynamics Modeling?, *Peter Yongho Kim, Juhyeon Park, Jungwoo Park, Jubin Choi, Jungwoo Seo, Jiook Cha, Taesup Moon*

626 IEBGL:An Interpretability-Enhanced Brain Graph Learning Framework with LLM-Instructed Topology and Literature-Augmented Semantics, *Yihang Duan, Shuo Huang, Li Zhang, Meiling Wang, Li Zhang*

627 Few-Shot Hybrid Incremental Learning: Continually Learning under Data Scarcity and Task Uncertainty, *Bin Pu, Xusheng Liang, Xinpeng Ding, Jinlin Wu, Zhen Lei, Shengli Li, Kenli Li, Jiawei Ma*

628 Sparse Spectral LoRA: Routed Experts for Medical VLMs, *Omid Nejatimanzari, Hojat Asgariandehkordi, Taha Koleilat, Yiming Xiao, Hassan Rivaz*

629 SAT-RRG: LLM-Guided Self-Adaptive Training for Radiology Report Generation with Token-Level Push–Pull Optimization, *Yunyi Liu, Yingshu Li, Tong Chen, Lingqiao Liu, Lei Wang, Luping Zhou*

630 OralGPT-Plus: Learning to Use Visual Tools via Reinforcement ❊ Learning for Panoramic X-ray Analysis, *Yuxuan Fan, Jing Hao, Hong Chen, Jiahao Bao, Yihua Shao, Yuci Liang, Kuo Feng Hung, Hao Tang*

631 Structural–Semantic Perception for Diffusion-Guided Temporal Forgery Localization, *Ligong Cao, Yeting Guo, Haoang Chi*

632 Forensic-Friendly Image Manipulation via Controllable Latent Diffusion, *Hanyu Chen, Haiwei Wu, Jinyu Tian, Jianqing LI, Jiantao Zhou*

633 IncreFA: Breaking the Static Wall of Generative Model Attribution, *Haotian Qin, Dongliang Chang, Yueying Gao, Yuexuan Tan, Lei Chen, Zhanyu Ma*

634 AVFakeBench: A Comprehensive Audio-Video Forgery Detection Benchmark for AV-LMMs, *Shuhan Xia, Peipei Li, Xuannan Liu, Dongsen Zhang, Xinyu Guo, Zekun Li*

635 Detecting Compressed AI-Generated Images via Phase Spectrum Robustness, *Kai Li, Wenqi Ren, Wei Wang, Xiaochun Cao*

636 Detect Any AI-Counterfeited Text Image, *Chenfan Qu, Yiwu Zhong, Xuekang Zhu, Junchi Li, Changjiang Jiang, Jian liu, Lianwen Jin*

637 DeepfakeImpact: A Two-Stage Benchmark with Real-World Impact in Deepfake Detection, *Chaoyu Gong, Han Zhang, Siqiang Luo*

638 Enhancing the Security of Visual Speaker Authentication Based on Dynamic Lip-Print Analysis, *Yi He, Lei Yang, Bofan Chen, Shilin Wang*

639 SimLBR: Learning to Detect Fake Images by Learning to Detect Real Images, *Aayush Dhakal, Subash Khanal, Srikumar Sastry, Jacob Arndt, Philipe Dias, Dalton Lunga, Nathan Jacobs*

640 Editprint: General Digital Image Forensics via Editing Fingerprint with ❊ Self-Augmentation Training, *Haiwei Wu, Kemou Li, Yuanman Li, Jiantao Zhou*

641 Detecting AI-Generated Forgeries via Iterative Manifold Deviation Amplification, *Jiangling Zhang, Shuxuan Gao, Bofan Liu, Siqiang Feng, Jirui Huang, Yaxiong Chen, Ziyu Chen*

642 Goldilocks Test Sets for Face Verification, *Haiyu Wu, Sicong Tian, Aman Bhatta, Jacob Gutierrez, Grace Bezold, Genesis Argueta, Karl Ricanek, Michael C. King, Kevin Bowyer*

643 Fine-VAD: Towards Fine-Grained Video Anomaly Detection via ❊ Progressive Cross-Granularity Learning, *Menghao Zhang, Yiyan Zhu, Pengfei Ren, Haifeng Sun, Qi Qi, Zirui Zhuang, Huazheng Wang, Lei Zhang, Jianxin Liao, Jingyu Wang*

644 DLVP-CLIP: Enhancing Fine-Grained Zero-Shot Anomaly Detection via Dynamic Local Visual Prompting, *Gaowei Zhang, Lihe Zhang*

645 MoECLIP: Patch-Specialized Experts for Zero-shot Anomaly Detection, *Jun Yeong Park, JunYoung Seo, Minji Kang, Yu Rang Park*

646 Alert-CLIP: Abnormality-aware Latent-Enhanced Representation Tuning of CLIP for Video Anomaly Detection, *Yiyan Zhu, Menghao Zhang, Haifeng Sun, Pengfei Ren, Xianao Chu, Chenye Xu, Hong Tan, Jinghan Wang, Qi Qi, Jingyu Wang*

647 AnomalyVFM -- Transforming Vision Foundation Models into Zero-Shot Anomaly Detectors, *Matic Fučka, Vitjan Zavrtanik, Danijel Skočaj*

648 LayoutAD: Exploring Semantic-Geometric Misalignment Reasoning for Scene Layout Anomaly Detection, *Zhichao Zeng, Jiasheng Zhang, Jiyun Sun, Jiangtao Cui, Xiaotian Qiao*

649 Bidirectional Multimodal Prompt Learning with Scale-Aware Training for Few-Shot Multi-Class Anomaly Detection, *Yujin Lee, Sewon Kim, Daeun Moon, Seoyoon Jang, Hyunsoo Yoon*

650 GS-CLIP: Zero-shot 3D Anomaly Detection by Geometry-Aware Prompt and Synergistic View Representation Learning, *Zehao Deng, An Liu, Yan Wang*

651 TLMA: Mitigating the Impact of Weakly Labeled Information for Video Anomaly Detection, *Rong Xu, Runqi Wang, Yingjun Zhang, Tao Tao, Xiaomeng Li, Liping Jing*

652 Defect Cue-Preserved Structural Feature Refinement for Few-Shot Anomaly Detection, *Le Jiang, Yan Huang, Zhen Xu, Yong Xu, Hau-San Wong, Si Wu*

653 Anomaly-Related Residual Fields for Cross-domain Anomaly Detection, *Kewei Gao, Jiayi Xie, Zhengda Shen, Weijun Qin, Lingxiang Jia, Kejia Chen, Zunlei Feng, Yijun Bei*

654 From Attraction to Equilibrium: Physics-Inspired Semantic Gravitons for Zero-Shot Anomaly Detection, *Yuwen Pan, Yuan Wang, Shaohui Li, Zhi Li, Yu Liu, You He*

655 Joint Learning of General and Diverse Patterns with Mixture of Memory Experts for Weakly-Supervised Video Anomaly Detection, *Bo Sun, Junxi Chen, Zhe Wu, Feng Gao, Fan Yang, Li Su, Yaowei Wang*

656 No Need For Real Anomaly: MLLM Empowered Zero-Shot Video Anomaly Detection, *Zunkai Dai, Ke Li, Jiajia Liu, Jie Yang, Yuanyuan Qiao*

657 FB-CLIP: Fine-Grained Zero-Shot Anomaly Detection with Foreground-Background Disentanglement, *Ming Hu, Yongsheng Huo, Mingyu Dou, Jianfu Yin, Peng Zhao, Yao Wang, Cong Hu, Bingliang Hu, Quan Wang*

658 DynamicVGGT: Learning Dynamic Point Maps for 4D Scene Reconstruction in Autonomous Driving, *Zhuolin He, Jing Li, Guanghao Li, Xiaolei Chen, Jiacheng Tang, Siyang Zhang, Zhounan Jin, Feipeng Cai, Bin Li, Jian Pu, Jia Cai, Xiangyang Xue*

659 GenieDrive: Towards Physics-Aware Driving World Model with 4D Occupancy Guided Video Generation, *Zhenya Yang, Zhe Liu, Yuxiang Lu, Liping Hou, Chenxuan Miao, Siyi Peng, Bailan Feng, Xiang Bai, Hengshuang Zhao*

660 Test-Time 3D Occupancy Prediction, *Fengyi Zhang, Xiangyu Sun, Huitong Yang, Zheng Zhang, Zi Huang, Yadan Luo*

661 Group Diffusion: Enhancing Image Generation by Unlocking Cross-Sample Collaboration, *Sicheng Mo, Thao Nguyen, Richard Zhang, Nick Kolkin, Siddharth Srinivasan Iyer, Eli Shechtman, Krishna Kumar Singh, Yong Jae Lee, Bolei Zhou, Yuheng Li*

662 Diffusion Mental Averages, *Phonphrm Thawatdamrongkit, Sukit Seripanitkarn, Supasorn Suwajanakorn*

663 dMLLM-TTS: Self-Verified and Efficient Test-Time Scaling for Diffusion Multi-Modal Large Language Models, *Yi Xin, Siqi Luo, Tianxiang Xu, Qi Qin, Haoxing Chen, Kaiwen Zhu, Zhiwei Zhang, Yangfan He, Rongchao Zhang, Jinbin Bai, Shuo Cao, Bin Fu, Junjun He, Yihao Liu, Yuewen Cao, Xiaohong Liu*
664 RegionRoute: Regional Style Transfer with Diffusion Model, *Bowen Chen, Jake Zuena, Alan C. Bovik, Divya Kothandaraman*
665 Low-Rank Residual Diffusion Models, *Junfu Tan, Jiang Yuan*
666 RDF-MIG: A Robust Diffusion Framework for Masked Image Generation to Augment Semantic Segmentation and Change Detection, *Zian Cao, Wei Wei, Qingshan Gao, Yuanyuan Fu*
667 TC-Padé: Trajectory-Consistent Padé Approximation for Diffusion Acceleration, *Shaoxuan He, Benlei Cui, Bukun Huang, Zhizeng Ye, Yunyun Sun, Longtao Huang, Hui Xue, Yang Yang, Haiwen Hong, Jingqun Tang, Zhou Zhao*
668 Bi-directional Autoregressive Diffusion for Large Complex Motion Interpolation, *Yongrui Ma, Shijie Zhao, Mingde Yao, Junlin Li, Li Zhang, Xiaohong Liu, Qi Dou, Jinwei Gu, Tianfan Xue*
669 Guiding Token-Sparse Diffusion Models, *Felix Krause, Stefan Andreas Baumann, Johannes Schusterbauer, Olga Grebenkova, Ming Gui, Vincent Tao Hu, Björn Ommer*
670 Accelerating Diffusion-based Video Editing via Heterogeneous Caching: Beyond Full Computing at Sampled Denoising Timestep, *Tianyi Liu, Ye Lu, Linfeng Zhang, Chen Cai, Jianjun Gao, Yi Wang, Kim-Hui Yap, Lap-Pui Chau*
671 See and Fix the Flaws: Enabling VLMs and Diffusion Models to Comprehend Visual Artifacts via Agentic Data Synthesis, *Jaehyun Park, Minyoung Ahn, Minkyu Kim, Jonghyun Lee, Jae-Gil Lee, Dongmin Park*
672 High-Fidelity Virtual Try-On beyond Paired Data Scarcity via Diffusion-based Cycle-Consistent Learning, *Jia Wu, Yijing Dai, Tingfeng Cao, Meiling Wu, Tao Luo, Jian Dong Zhang, Guangming Lu, Xiaoyi Zeng*
673 Sampling-Aware Quantization for Diffusion Models, *Qian Zeng, Jie Song, Yuanyu Wan, Huiqiong Wang, Mingli Song*
674 CRAFT: Aligning Diffusion Models with Fine-Tuning Is Easier Than You Think, *Zening Sun, Zhengpeng Xie, Lichen Bai, Shitong Shao, Shuo Yang, Zeke Xie*
675 Scale Space Diffusion, *Soumik Mukhopadhyay, Prateksha Udhayanan, Abhinav Shrivastava*
676 Making Training-Free Diffusion Segmentors Scale with the Generative Power, *Benyuan Meng, Qianqian Xu, Zitai Wang, Xiaochun Cao, Longtao Huang, Qingming Huang*
677 Roots Beneath the Cut: Uncovering the Risk of Concept Recovery in Pruning-Based Unlearning for Diffusion Models, *Ci Zhang, Zhaojun Ding, Chence Yang, Jun Liu, Xiaoming Zhai, Shaoyi Huang, Beiwen Li, Xiaolong Ma, Jin Lu, Geng Yuan*
678 Few-Step Diffusion Sampling Through Instance-Aware Discretizations, *Liangyu Yuan, Ruoyu Wang, Tong Zhao, Dingwen Fu, Mingkun Lei, Beier Zhu, Chi Zhang*
679 SpeeDiff: Scalable Pixel-Anchored End-to-End Latent Diffusion Model, *Bingliang Zhang, Wenda Chu, Yizhuo Li, Linjie Yang, Yisong Yue, Katherine Bouman, Yang Song, Qiushan Guo*
680 Structure-to-Intensity Diffusion for Adverse-Weather LiDAR Generation, *Peiyang Ni, Longyu Yang, Lu Zhang, Kuniaki Saito, Yap-Peng Tan, Fumin Shen, Heng Tao Shen, Xiaofeng Zhu, Ping Hu*
681 Focal–General Diffusion Model with Semantic Consistent Guidance for Sign Language Production, *Yiheng Yu, Sheng Liu, Yuan Feng, Zhelun Jin, Yining Jiang, Min Xu*
682 Diffusion Probe: Generated Image Result Prediction Using CNN Probes, *Bukun Huang, Benlei Cui, Zhizeng Ye, Xuemei Dong, Tuo Chen, Hui Xue, Dingkang Yang, Longtao Huang, Haiwen Hong, Jingqun Tang*
683 Content-Aware Dynamic Patchification for Efficient Video Diffusion, *Sheng Li, Connelly Barnes, Mamshad Nayeem Rizve, Hongwu Peng, Zhengang Li, Ohi Dibua, Alireza Ganjdanesh, Xulong Tang, Yan Kang, Yifan Gong*
684 PixelRush: Ultra-Fast, Training-Free High-Resolution Image
❊ Generation via One-step Diffusion, *Hong-Phuc Lai, Phong Nguyen, Anh Tran*
685 Diffusion-Based sRGB Real Noise Generation via Prompt-Driven
❊ Noise Representation Learning, *Jaekyun Ko, Dongjin Kim, Soomin Lee, Guanghui Wang, Tae Hyun Kim*
686 Decoupled Residual Denoising Diffusion Models for Unified and Data Efficient Image-to-Image Translation, *Ziyue Lin, Jiahe Hou, Hongyu Xia, Xinrui Xie, Feifei Wang, Yuyin Zhou, Wei Wang, Jiawei Liu, Liangqiong Qu*
687 GROW: Watermark Generation with Progressive Guidance for Diffusion Models, *Pengcheng Luo, Zexi Jia, Yijia Zhong, Jinchao Zhang, Jie Zhou*
688 MotionV2V: Editing Motion in a Video, *Ryan Burgert, Charles Herrmann, Forrester Cole, Michael S Ryoo, Neal Wadhwa, Andrey Voynov, Nataniel Ruiz*
689 Mind the Generative Details: Direct Localized Detail Preference Optimization for Video Diffusion Models, *Zitong Huang, Kaidong Zhang, Yukang Ding, Chao Gao, Rui Ding, Ying Chen, Wangmeng Zuo*
690 OrthoFuse: Training-free Riemannian Fusion of Orthogonal Style-Concept Adapters for Diffusion Models, *Ali Aliev, Kamil Garifullin, Nikolay Yudin, Vera Soboleva, Alexander Molozhavenko, Ivan Oseledets, Aibek Alanov, Maxim Rakhuba*
691 DreamStyle: A Unified Framework for Video Stylization, *Mengtian Li, Jinshu Chen, Songtao Zhao, Wanquan Feng, Pengqi Tu, Qian He*
692 Diffusion Sampling Path Tells More: An Efficient Plug-and-Play Strategy for Sample Filtering, *Sixian Wang, Zhiwei Tang, Tsung-Hui Chang*
693 Designing Instance-Level Sampling Schedules via REINFORCE with James-Stein Shrinkage, *Peiyu Yu, Suraj Kothawade, Sirui Xie, Ying Nian Wu, Hongliang Fei*
694 Reward Sharpness-Aware Fine-Tuning for Diffusion Models, *Kwanyoung Kim, Byeongsu Sim*
695 DBMSolver: A Training-free Diffusion Bridge Sampler for High-Quality Image-to-Image Translation, *Sankarshana Venugopal, Mohammad Mostafavi, Jonghyun Choi*
696 Cubic Discrete Diffusion: Discrete Visual Generation on High-
❊ Dimensional Representation Tokens, *Yuqing Wang, Chuofan Ma, Zhijie Lin, Yao Teng, Lijun Yu, Shuai Wang, Jiaming Han, Jiashi Feng, Yi Jiang, Xihui Liu*
697 TAP: A Token-Adaptive Predictor Framework for Training-Free Diffusion Acceleration, *Haowei Zhu, Tingxuan Huang, Xing Wang, Tianyu Zhao, Jiexi Wang, Weifeng Chen, Xurui Peng, Fangmin Chen, Junhai Yong, Bin Wang*
698 Cross-modal Representation Learning for Diffusion-generated Image Detection, *Tao Gong, Dayong Wang, Qi Chu, Bin Liu, Nenghai Yu*
699 Sparse-LaViDa: Sparse Multimodal Discrete Diffusion Language
❊ Models, *Shufan Li, Jiuxiang Gu, Kangning Liu, Zhe Lin, Zijun Wei, Aditya Grover, Jason Kuen*
700 Back to Basics: Let Denoising Generative Models Denoise, *Tianhong Li, Kaiming He*
701 CaricHarmony: Contrastive Diffusion Paths for Identity-Preserving Caricature Synthesis, *Dongyu Wang, Dar-Yen Chen, Yi-Zhe Song*
702 DiP: Taming Diffusion Models in Pixel Space, *Zhennan Chen, Junwei Zhu, Xu Chen, Jiangning Zhang, Xiaobin Hu, Hanzhen Zhao, Chengjie Wang, Jian Yang, Ying Tai*
703 RAPID: Reusing Attention Sparsity with Inter-step Adaptation for Efficient Video Diffusion, *Shangran Lin, Lu Lu, Jian Chen, Qiang Liu*
704 Efficient and Training-Free Single-Image Diffusion Models,
❊ *HaojunQiu, Kiriakos N. Kutulakos, David B. Lindell*

11:45 - 13:45 DEMOS

1 FOVEA: Flexible Ontology Visual Event Analyzer, *Aaron Steven White*
2 Rapid 3D Object Annotation through In-Situ Geometric Grounding, *Narges Honarvar Nazari*
3 EMMA: Extracting Multiple physical parameters from Multimodal Data, *Farhat Shaikh, Ayan Banerjee, Sandeep Gupta*
4 SparkVSR: Interactive Video Super-Resolution via Sparse Keyframe Propagation, *Jiongze Yu, Xiangbo Gao, Pooja Verlani, Akshay Gadde, Yilin Wang, Balu Adsumilli, Zhengzhong Tu*
5 EgoMedAgent: Towards Evidence-based Egocentric Assistant for Clinical Perception and Action, *Chen Fang, Xu Cao, Houze Yang, Yifan Shen*
6 Authenticating Matryoshka Nesting Dolls via Zero-Shot 3D Completion, *Yulia Kumar, Srotriyo Sengupta*

The papers in each oral session will also be presented as a poster in the following poster session. **Presentation order is subject to change. Please see mobile app or website for final presentation order.**

14:00 - 15:15 Oral Session 6A: Geometric Learning
(Bluebird Ballroom)
🏆 - Award candidate paper

1 Differentiable Laplacian Matrix Guided Superpixel Segmentation, *Jeremy Juybari, Josh Hamilton, Shuvra Das, Chaofan Chen, Andre Khalil, Yifeng Zhu*
2 FILTR: Extracting Topological Features from Pretrained 3D Models, *Louis Martinez, Maks Ovsjanikov*
3 Learning Convex Decomposition via Feature Fields, *Yuezhi Yang, Qixing Huang, Mikaela Angelina Uy, Nicholas Sharp*
4 🏆 Learning Eigenstructures of Unstructured Data Manifolds, *Roy Velich, Arkadi Piven, David Bensaïd, Daniel Cremers, Thomas Dagès, Ron Kimmel*
5 🏆 Mapping Networks, *Lord Sen, Shyamapada Mukherjee*

14:00 - 15:15 Oral Session 6B: Multimodal Reasoning
(Four Seasons Ballroom)

1 🏆 CineBrain: A Large-Scale Multi-Modal Audiovisual Brain Dataset for Brain-Conditioned Video Generation, *Jianxiong Gao, Yichang Liu, Baofeng Yang, Jianfeng Feng, Yanwei Fu*
2 🏆 Hearing the Room Through the Shape of the Drum: Modal-Guided Sound Recovery from Multi-Point Surface Vibrations, *Shai Bagon, Matan Kichler, Mark Sheinin*
3 SDTrack: A Baseline for Event-based Tracking via Spiking Neural Networks, *Yimeng Shan, Zhenbang Ren, Haodi Wu, Wenjie Wei, Rui-Jie Zhu, Shuai Wang, Dehao Zhang, Yichen Xiao, Jieyuan Zhang, Kexin Shi, Jingzhinan Wang, Jason K. Eshraghian, Haicheng Qu, Malu Zhang*
4 🏆 Thinking with Drafts: Speculative Temporal Reasoning for Efficient Long Video Understanding, *Pengfei Hu, Meng Cao, Yingyao Wang, Yi Wang, Jiahua Dong, Jun Song, Yu Cheng, Bo Zheng, Xiaodan Liang*
5 Wan-Weaver: Interleaved Multi-modal Generation via Decoupled Training, *Jinbo Xing, Zeyinzi Jiang, Yuxiang Tuo, Chaojie Mao, Xiaotang Gai, Xi Chen, Jingfeng Zhang, Yulin Pan, Zhen Han, Jie Xiao, Keyu Yan, Chenwei Xie, Chongyang Zhong, Kai Zhu, Tong Shen, Lianghua Huang, Yu Liu, Yujiu Yang*

14:00 - 15:15 Oral Session 6C: Medical Vision
(Mile High Ballroom 1A - 2A)

1 CURE: Curriculum-guided Multi-task Training for Reliable Anatomy Grounded Report Generation, *Pablo Messina, Andrés Villa, Juan Leon Alcazar, Karen Sanchez, Carlos Hinojosa, Denis Parra, Alvaro Soto, Bernard Ghanem*
2 DK-DDIL: Adaptive Knowledge Retention for Dynamic Domain-Incremental Learning in Medical Imaging, *Yuxi Ma, Sujie Liu, Jing Yang, Jiacheng Wang, Yiping Chen, Baptiste Magnier, Liansheng Wang*
3 🏆 Dual-level Adapter Boosting Prompt-free Curvilinear Structure Segmentation, *Kai Zhu, Li Chen, Jun Cheng*
4 LATA: Laplacian-Assisted Transductive Adaptation for Conformal Uncertainty in Medical VLMs, *Behzad Bozorgtabar, Dwarikanath Mahapatra, Sudipta Roy, Muzammal Naseer, Imran Razzak, Zongyuan Ge*
5 🏆 Medic-AD: Towards Medical Vision-Language Model's Clinical Intelligence, *Woohyeon Park, Jaeik Kim, Sunghwan Steve Cho, Pa Hong, Wookyoung Jeong, Yoojin Nam, Namjoon Kim, Ginny Y. Wong, Ka Chun Cheung, Jaeyoung Do*
6 SegMoTE: Token-Level Mixture of Experts for Medical Image Segmentation, *Yujie Lu, Jingwen Li, Sibo Ju, Yanzhou Su, He Yao, Yisong Liu, Min Zhu, Junlong Cheng*

14:00 - 15:15 Oral Session 6D: Large-Scale Neural Modeling
(Mile High Ballroom 3A - 4A)

1 🏆 Efficient Unrolled Networks for Large-Scale 3D Inverse Problems, *Romain Vo, Julián Tachella*
2 Federated Optimization, *Wenjie Hou, Tianxiang Chen, Feng Wang, Tiantong Wu, Zhiming Zheng, Shaoting Tang, Wei Yang Bryan Lim*
3 SimScale: Learning to Drive via Real-World Simulation at Scale, *Haochen Tian, Tianyu Li, Haochen Liu, Jiazhi Yang, Yihang Qiu, Guang Li, Junli Wang, Yinfeng Gao, Zhang Zhang, Liang Wang, Hangjun Ye, Long Chen, Hongyang Li*
4 🏆 Texvent: Asynchronous Event Data Simulation via Text Prompt, *Ruofei Wang, Peiqi Duan, Ka Chun Cheung, Simon See, Boxin Shi, Renjie Wan*
5 WorldLens: Full-Spectrum Evaluations of Driving World Models in Real World, *Ao Liang, Lingdong Kong, Tianyi Yan, Hongsi Liu, Yu Yang, Ziqi Huang, Wei Yin, Jialong Zuo, Yixuan Hu, Dekai Zhu, Dongyue Lu, Youquan Liu, Guangfeng Jiang, Linfeng Li, Xiangtai Li, Long Zhuo, Lai Xing Ng, Benoit R. Cottereau, Changxin Gao, Liang Pan, Wei Tsang Ooi, Ziwei Liu*

15:30 - 17:30 Poster Session 6 (ExHall A)
❊ - Highlight paper 🏆 - Award candidate paper

1 Differentiable Laplacian Matrix Guided Superpixel Segmentation, *Jeremy Juybari, Josh Hamilton, Shuvra Das, Chaofan Chen, Andre Khalil, Yifeng Zhu*
2 FILTR: Extracting Topological Features from Pretrained 3D Models, *Louis Martinez, Maks Ovsjanikov*
3 Learning Convex Decomposition via Feature Fields, *Yuezhi Yang, Qixing Huang, Mikaela Angelina Uy, Nicholas Sharp*
4 🏆 Learning Eigenstructures of Unstructured Data Manifolds, *Roy Velich, Arkadi Piven, David Bensaïd, Daniel Cremers, Thomas Dagès, Ron Kimmel*
5 🏆 Mapping Networks, Lord Sen, Shyamapada Mukherjee
6 🏆 CineBrain: A Large-Scale Multi-Modal Audiovisual Brain Dataset for Brain-Conditioned Video Generation, *Jianxiong Gao, Yichang Liu, Baofeng Yang, Jianfeng Feng, Yanwei Fu*
7 🏆 Hearing the Room Through the Shape of the Drum: Modal-Guided Sound Recovery from Multi-Point Surface Vibrations, *Shai Bagon, Matan Kichler, Mark Sheinin*
8 SDTrack: A Baseline for Event-based Tracking via Spiking Neural Networks, *Yimeng Shan, Zhenbang Ren, Haodi Wu, Wenjie Wei, Rui-Jie Zhu, Shuai Wang, Dehao Zhang, Yichen Xiao, Jieyuan Zhang, Kexin Shi, Jingzhinan Wang, Jason K. Eshraghian, Haicheng Qu, Malu Zhang*
9 🏆 Thinking with Drafts: Speculative Temporal Reasoning for Efficient Long Video Understanding, *Pengfei Hu, Meng Cao, Yingyao Wang, Yi Wang, Jiahua Dong, Jun Song, Yu Cheng, Bo Zheng, Xiaodan Liang*
10 Wan-Weaver: Interleaved Multi-modal Generation via Decoupled Training, *Jinbo Xing, Zeyinzi Jiang, Yuxiang Tuo, Chaojie Mao, Xiaotang Gai, Xi Chen, Jingfeng Zhang, Yulin Pan, Zhen Han, Jie Xiao, Keyu Yan, Chenwei Xie, Chongyang Zhong, Kai Zhu, Tong Shen, Lianghua Huang, Yu Liu, Yujiu Yang*
11 CURE: Curriculum-guided Multi-task Training for Reliable Anatomy Grounded Report Generation, *Pablo Messina, Andrés Villa, Juan Leon Alcazar, Karen Sanchez, Carlos Hinojosa, Denis Parra, Alvaro Soto, Bernard Ghanem*
12 DK-DDIL: Adaptive Knowledge Retention for Dynamic Domain-Incremental Learning in Medical Imaging, *Yuxi Ma, Sujie Liu, Jing Yang, Jiacheng Wang, Yiping Chen, Baptiste Magnier, Liansheng Wang*
13 🏆 Dual-level Adapter Boosting Prompt-free Curvilinear Structure Segmentation, *Kai Zhu, Li Chen, Jun Cheng*
14 LATA: Laplacian-Assisted Transductive Adaptation for Conformal Uncertainty in Medical VLMs, *Behzad Bozorgtabar, Dwarikanath Mahapatra, Sudipta Roy, Muzammal Naseer, Imran Razzak, Zongyuan Ge*
15 🏆 Medic-AD: Towards Medical Vision-Language Model's Clinical Intelligence, *Woohyeon Park, Jaeik Kim, Sunghwan Steve Cho, Pa Hong, Wookyoung Jeong, Yoojin Nam, Namjoon Kim, Ginny Y. Wong, Ka Chun Cheung, Jaeyoung Do*
16 SegMoTE: Token-Level Mixture of Experts for Medical Image Segmentation, *Yujie Lu, Jingwen Li, Sibo Ju, Yanzhou Su, He Yao, Yisong Liu, Min Zhu, Junlong Cheng*
17 🏆 Efficient Unrolled Networks for Large-Scale 3D Inverse Problems, *Romain Vo, Julián Tachella*

18 FedAdamom: Adaptive Momentum for Improved Generalization in Federated Optimization, *Wenjie Hou, Tianxiang Chen, Feng Wang, Tiantong Wu, Zhiming Zheng, Shaoting Tang, Wei Yang Bryan Lim*
19 SimScale: Learning to Drive via Real-World Simulation at Scale, *Haochen Tian, Tianyu Li, Haochen Liu, Jiazhi Yang, Yihang Qiu, Guang Li, Junli Wang, Yinfeng Gao, Zhang Zhang, Liang Wang, Hangjun Ye, Long Chen, Hongyang Li*
20 Texvent: Asynchronous Event Data Simulation via Text Prompt, *Ruofei Wang, Peiqi Duan, Ka Chun Cheung, Simon See, Boxin Shi, Renjie Wan*
21 WorldLens: Full-Spectrum Evaluations of Driving World Models in Real World, *Ao Liang, Lingdong Kong, Tianyi Yan, Hongsi Liu, Yu Yang, Ziqi Huang, Wei Yin, Jialong Zuo, Yixuan Hu, Dekai Zhu, Dongyue Lu, Youquan Liu, Guangfeng Jiang, Linfeng Li, Xiangtai Li, Long Zhuo, Lai Xing Ng, Benoit R. Cottereau, Changxin Gao, Liang Pan, Wei Tsang Ooi, Ziwei Liu*
22 BuildingGPT: Auto-Regressive Building Wireframe Reconstruction Model with Reinforcement Learning, *Yuzhou Liu, Lingjie Zhu, Hanqiao Ye, Yujun Liu, Shangfeng Huang, Xiang Gao, Ruisheng Wang, Shuhan Shen*
23 Emergent Extreme-View Geometry in 3D Foundation Models, *Yiwen Zhang, Joseph Tung, Ruojin Cai, David Fouhey, Hadar Averbuch-Elor*
24 LiteVGGT: Boosting Vanilla VGGT via Geometry-aware Cached Token Merging, *Zhijian Shu, Cheng Lin, Tao Xie, Wei Yin, Ben Li, Zhiyuan Pu, Weize Li, Yao Yao, Xun Cao, Xiaoyang Guo, Xiao-Xiao Long*
25 LASER: Layer-wise Scale Alignment for Training-Free Streaming 4D Reconstruction, *Tianye Ding, Yiming Xie, Yiqing Liang, Moitreya Chatterjee, Pedro Miraldo, Huaizu Jiang*
26 PanoVGGT: Feed-Forward 3D Reconstruction from Panoramic Imagery, *Yijing Guo, Mengjun Chao, Luo Wang, Tianyang Zhao, Haizhao Dai, Yingliang Zhang, Jingyi Yu, Yujiao Shi*
27 Rascene: High-Fidelity 3D Scene Imaging with mmWave Communication Signals, *Kunzhe Song, Geo Jie Zhou, Xiaoming Liu, Huacheng Zeng*
28 VGG-T^3: Offline Feed-Forward 3D Reconstruction at Scale, *Sven Elflein, Ruilong Li, Sérgio Agostinho, Zan Gojcic, Laura Leal-Taixé, Qunjie Zhou, Aljosa Osep*
29 SEA-Flow3D: Simplified, Efficient, and Accurate Scene Flow via Spatial Vector Sampling and Multi-scale Refinement, *Han Ling, Quansen Sun, Yinghua Yao, Ivor Tsang, Yinghui Sun*
30 OmniVGGT: Omni-Modality Driven Visual Geometry Grounded Transformer, *Haosong Peng, Hao Li, Yalun Dai, Yushi Lan, Yihang Luo, Tianyu Qi, Zhengshen Zhang, Yufeng Zhan, Junfei Zhang, Wenchao Xu, Ziwei Liu*
31 DROID-SLAM in the Wild, *Moyang Li, Zihan Zhu, Marc Pollefeys, Daniel Barath*
32 HeSS: Head Sensitivity Score for Sparsity Redistribution in VGGT, *Yongsung Kim, Wooseok Song, Jaihyun Lew, Hun Hwangbo, Jaehoon Lee, Sungroh Yoon*
33 Dense Metric Depth Completion from Sparse Direct Time-of-Flight Sensors, *Hakyeong Kim, Ruicheng Wang, Chengtang Yao, Jiaolong Yang, Min H. Kim*
34 Online3R: Online Learning for Consistent Sequential Reconstruction Based on Geometry Foundation Model, *Shunkai Zhou, Zike Yan, Fei Xue, Dong Wu, Yuchen Deng, Hongbin Zha*
35 Neu-PiG: Neural Preconditioned Grids for Fast Dynamic Surface Reconstruction on Long Sequences, *Julian Kaltheuner, Hannah Dröge, Markus Plack, Patrick Stotko, Reinhard Klein*
36 Learning 3D Reconstruction with Priors in Test Time, *Lei Zhou, Haoyu Wu, Akshat Dave, Dimitris Samaras*
37 ArchSym: Detecting 3D-Grounded Architectural Symmetries in the Wild, *Hanyu Chen, Ruojin Cai, Steve Marschner, Noah Snavely*
38 PointTPA: Dynamic Network Parameter Adaptation for 3D Scene Understanding, *Siyuan Liu, Chaoqun Zheng, Xin Zhou, Tianrui Feng, Dingkang Liang, Xiang Bai*
39 tttLRM: Test-Time Training for Long Context and Autoregressive 3D Reconstruction, *Chen Wang, Hao Tan, Wang Yifan, Zhiqin Chen, Yuheng Liu, Kalyan Sunkavalli, Sai Bi, Lingjie Liu, Yiwei Hu*
40 Hint2Gen: Bridging Understanding and Generation via Code-structured Hints, *Yuanpeng Tu, Yunpeng Chen, Xi Chen, Liang Li, Hengshuang Zhao*
41 Compositional Text-to-Image Generation Via Region-aware Bimodal Direct Preference Optimization, *Zhuohan Liu, Wujian Peng, Yitong Chen, Zuxuan Wu*
42 Learning by Analogy: A Causal Framework for Compositional Generalization, *Lingjing Kong, Shaoan Xie, Yang Jiao, Yetian Chen, Yanhui Guo, Simone Shao, Yan Gao, Guangyi Chen, Kun Zhang*
43 ID-Crafter: VLM-Grounded Online RL for Compositional Multi-Subject Video Generation, *Panwang Pan, Jingjing Zhao, Yuchen Lin, Chenguo Lin, Chenxin Li, Hengyu Liu, Tingting Shen, Yadong Mu*
44 GenColorBench: A Color Evaluation Benchmark for Text-to-Image Generation, *Muhammad Atif Butt, Alexandra Gomez-Villa, Tao Wu, Javier Vazquez-Corral, Joost Van De Weijer, Kai Wang*
45 Extending One-Step Image Generation from Class Labels to Text via Discriminative Text Representation, *Chenxi Zhao, Chen Zhu, Xiaokun Feng, Aiming Hao, Jiashu Zhu, Jiachen Lei, Jiahong Wu, Xiangxiang Chu, Jufeng Yang*
46 When Pretty Isn't Useful: Investigating Why Modern Text-to-Image Models Fail as Reliable Training Data Generators, *Krzysztof Adamkiewicz, Brian B. Moser, Stanislav Frolov, Tobias Christian Nauen, Federico Raue, Andreas Dengel*
47 TempoControl: Temporal Attention Guidance for Text-to-Video Models, *Shira Schiber, Ofir Lindenbaum, Idan Schwartz*
48 Hear What Matters! Text-conditioned Selective Video-to-Audio Generation, *Junwon Lee, Juhan Nam, Jiyoung Lee*
49 MultiCrafter: High-Fidelity Multi-Subject Generation via Disentangled Attention and Identity-Aware Preference Alignment, *Tao Wu, Yibo Jiang, Yehao Lu, Zhizhong Wang, Zeyi Huang, Zequn Qin, Xi Li*
50 Resolving the Identity Crisis in Text-to-Image Generation, *Shubhankar Borse, Farzad Farhadzadeh, Munawar Hayat, Fatih Porikli*
51 DiffGraph: An Automated Agent-driven Model Merging Framework for In-the-Wild Text-to-Image Generation, *Zhuoling Li, Hossein Rahmani, Jiarui Zhang, Yu Xue, Majid Mirmehdi, Jason Kuen, Jiuxiang Gu, Jun Liu*
52 Gloria: Consistent Character Video Generation via Content Anchors, *Yuhang Yang, Fan Zhang, Huaijin Pi, Ailing Zeng, Shuai Guo, Guowei Xu, Wei Zhai, Yang Cao, Zheng-Jun Zha*
53 DreamShot: Personalized Storyboard Synthesis with Video Diffusion Prior, *Junjia Huang, Binbin Yang, Pengxiang Yan, Jiyang Liu, Bin Xia, Zhao Wang, Yitong Wang, Liang Lin, Guanbin Li*
54 M4V: Multimodal Mamba for Efficient Text-to-Video Generation, *Jiancheng Huang, Gengwei Zhang, Zequn Jie, Siyu Jiao, Yinlong Qian, Ling Chen, Yunchao Wei, Lin Ma*
55 Property-Informed Diffusion-Based Text-to-Microstructure Generation, *Bingxuan Dai, Hongsong Wang, Jie Gui*
56 DreamingComics: A Story Visualization Pipeline via Subject and Layout Customized Generation using Video Models, *Patrick Kwon, Chen Chen*
57 Mixture of States: Routing Token-Level Dynamics for Multimodal Generation, *Haozhe Liu, Ding Liu, Mingchen Zhuge, Zijian Zhou, Tian Xie, Sen He, Yukang Yang, Shuming Liu, Yuren Cong, Jiadong Guo, Hongyu Xu, Ke Xu, Kam-Woh Ng, Juan C. Perez, Juan-Manuel Perez-Rua, Tao Xiang, Wei Liu, Shikun Liu, Jürgen Schmidhuber*
58 HiCoGen: Hierarchical Compositional Text-to-Image Generation in Diffusion Models via Reinforcement Learning, *Hongji Yang, Yucheng Zhou, Wencheng Han, Runzhou Tao, Zhongying Qiu, Jianfei Yang, Jianbing Shen*
59 TherA: Thermal-Aware Visual-Language Prompting for Controllable RGB-to-Thermal Infrared Translation, *Dong-Guw Lee, Tai Hyoung Rhee, Hyunsoo Jang, Young-Sik Shin, Ukcheol Shin, Ayoung Kim*
60 See What I Mean: Aligning Vision and Language Representations for Video Fine-grained Object Understanding, *Boyuan Sun, Bo-Wen Yin, Yuan-Ming Li, Xihan Wei, Qibin Hou*
61 CoV-Align: Efficient Fine-grained Cross-Modal Alignment with Cohesive Visual Semantics Priority, *Hengqi Liu, Wanting Zhou, Longteng Kong, Fangxiang Feng, Lei Ren, Wei Chen, Xiaojie Wang*
62 TDATR: Improving End-to-End Table Recognition via Table Detail-Aware Learning and Cell-Level Visual Alignment, *Chunxia Qin, Chenyu Liu, Pengcheng Xia, Jun Du, Baocai Yin, Bing Yin, Cong Liu*
63 A Mixed Diet Makes DINO An Omnivorous Vision Encoder, *Rishabh Kabra, Maks Ovsjanikov, Drew A. Hudson, Ye Xia,*

Skanda Koppula, Andre Araujo, Joao Carreira, Niloy J. Mitra

64 Uncertainty-guided Compositional Alignment with Part-to-Whole
❊ Semantic Representativeness in Hyperbolic Vision-Language Models, *Hayeon Kim, Ji Ha Jang, Junghun James Kim, Se Young Chun*

65 TaskForce: Cooperative Multi-agent Reinforcement Learning for Multi-task Optimization, *Wonhyeok Choi, Kyumin Hwang, Jihun Park, Kyoungmin Lee, Seunghun Lee, Jaeyeul Kim, Minwoo Choi, Sunghoon Im*

66 PhyCritic: Multimodal Critic Models for Physical AI, *Tianyi Xiong, Shihao Wang, Guilin Liu, Yi Dong, Ming Li, Heng Huang, Jan Kautz, Zhiding Yu*

67 R-C2: Cycle-Consistent Reinforcement Learning Improves Multimodal Reasoning, *Zirui Zhang, Haoyu Dong, Kexin Pei, Chengzhi Mao*

68 Multimodal RewardBench 2: Evaluating Omni Reward Models for Interleaved Text and Image, *Yushi Hu, Reyhane Askari-Hemmat, Melissa Hall, Emily Dinan, Luke Zettlemoyer, Marjan Ghazvininejad*

69 Unified Generation and Self-Verification for Vision-Language Models
❊ via Advantage Decoupled Preference Optimization, *Xinyu Qiu, Heng Jia, Zhengwen Zeng, Shuheng Shen, Changhua Meng, Yi Yang, Linchao Zhu*

70 Anchoring the Mind of Multimodal Reasoners: Cognitive Bias as a Vector for Jailbreak Attacks, *Linhua Cong, Bingrui Sima, Kun He*

71 InsCal: Calibrated Multi-Source Fully Test-Time Prompt Tuning for Object Detection, *Xiaofan Que, Dingrong Wang, Xumin Liu, Qi Yu*

72 Why Not Hyperparameter-Friendly Optimisation? A Monotonic Adaptive Norm Rescaling Approach For Long-Tailed Recognition, *Shuo Zhang, Chenqi Li, Tingting Zhu*

73 Decoupling Vision and Language: Codebook Anchored Visual Adaptation, *Jason Wu, Tianchen Zhao, Chang Liu, Jiarui Cai, Zheng Zhang, Zhuowei Li, Aaditya Singh, Xiang Xu, Mani Srivastava, Jonathan Wu*

74 MemFlow: A Lightweight Forward Memorizing Framework for Quick Domain Adaptive Feature Mapping, *Jianming Lv, Chengjun Wang, Depin Liang, Qianli Ma, Wei Chen, Xueqi Cheng*

75 Mind the Discriminability Trap in Source-Free Cross-domain Few-shot Learning, *Zhenyu Zhang, Yixiong Zou, Yuhua Li, Ruixuan Li, Guangyao Chen*

76 Vision-Language Model Guided Source-Free Domain Adaptation via Optimal Transport, *Shuo Han, Xu Tang, Jingjing Ma, Xiangrong Zhang*

77 Masked Representation Modeling for Domain-Adaptive Segmentation, *Wenlve Zhou, Zhiheng Zhou, Tiantao Xian, Yikui Zhai, Weibin Wu, Biyun MA*

78 TaskIT: Memory-Efficient Fine-Tuning of Multi-LoRA LLMs via Cross-Task Importance Transfer, *Cheng Fang, Zimu Zhou, Ke Ma, Bin Guo*

79 ARES: Unifying Asymmetric RGB-Event Stereo for Probabilistic Scene Flow Estimation, *Jie Long Lee, Gim Hee Lee*

80 MER-Tracker: Towards High-Speed 3D Point Tracking via Multi-View Event-RGB Hybrid Cameras, *Yiqian Chang, Qinghong Ye, Haoran Xu, Jianing Li, Dongyang Ma, Xuan Wang, Wei Zhang, Yonghong Tian, Peixi Peng*

81 Moving Border Ownership for Event-based Motion Segmentation, *Zhiyuan Hua, Cornelia Fermüller, Yiannis Aloimonos*

82 TTAPFormer: Robust Arbitrary Point Tracking via Transient Asynchronous Fusion of Frames and Events, *Jiaxiong Liu, Zhen Tan, Jinpu Zhang, Yi Zhou, Hui Shen, Xieyuanli Chen, Dewen Hu*

83 EventHub: Data Factory for Generalizable Event-Based Stereo Networks without Active Sensors, *Luca Bartolomei, Fabio Tosi, Matteo Poggi, Stefano Mattoccia, Guillermo Gallego*

84 Seeing Motion Through Polarity for Event-based Action Recognition, *Meiqi Cao, Jiachao Zhang, Xin Jiang, Rui Yan, Yazhou Yao, Zechao Li, Xiangbo Shu*

85 Multi-Scale Gaussian-Language Map for Zero-shot Embodied Navigation
❊ and Reasoning, *Sixian Zhang, Yiyao Wang, Xinhang Song, Keming Zhang, Zijian Xu, Shuqiang Jiang*

86 Explore with Long-term Memory: A Benchmark and Multimodal LLM-based Reinforcement Learning Framework for Embodied Exploration, *Sen Wang, Bangwei Liu, Zhenkun Gao, Lizhuang Ma, Xuhong Wang, Yuan Xie, Xin Tan*

87 SpaceTools: Tool-Augmented Spatial Reasoning via Double Interactive RL, *Siyi Chen, Mikaela Angelina Uy, Chan Hee Song, Faisal Ladhak, Adithyavairavan Murali, Qing Qu, Stan Birchfield, Valts Blukis, Jonathan Tremblay*

88 TeamHOI: Learning a Unified Policy for Cooperative Human-Object Interactions with Any Team Size, *Stefan Lionar, Gim Hee Lee*

89 AREA3D: Active Reconstruction Agent with Unified Feed-Forward 3D Perception and Vision-Language Guidance, *Tianling Xu, Shengzhe Gan, Leslie Gu, Yuelei Li, Fangneng Zhan, Hanspeter Pfister*

90 Experience Transfer for Multimodal LLM Agents in Minecraft Game, *Chenghao Li, Jun Liu, Songbo Zhang, Huadong Jian, Hao Ni, Lik-Hang Lee, Sung-Ho Bae, Guoqing Wang, Yang Yang, Chaoning Zhang*

91 MSGNav: Unleashing the Power of Multi-modal 3D Scene Graph for Zero-Shot Embodied Navigation, *Xun Huang, Shijia Zhao, Yunxiang Wang, Xin Lu, Wanfa Zhang, Rongsheng Qu, Weixin Li, Yunhong Wang, Chenglu Wen*

92 SaPaVe: Towards Active Perception and Manipulation in Vision-
❊ Language Action Models for Robotics, *Mengzhen Liu, Enshen Zhou, Cheng Chi, Yi Han, Shanyu Rong, Liming Chen, Pengwei Wang, Zhongyuan Wang, Shanghang Zhang*

93 MANSION: Multi-floor lANguage-to-3D Scene generatIOn for loNg-horizon tasks, *Lirong Che, Shuo Wen, Shan Huang, Chuang Wang, Yuzhe Yang, Gregory Dudek, Xueqian Wang, Jian Su*

94 RealAppiance: Let High-fidelity Appliance Assets Controllable and
❊ Workable as Aligned Real Manauls, *Yuzheng Gao, Yuxing Long, Lei Kang, Yuchong Guo, Ziyan Yu, Shangqing Mao, Jiyao Zhang, Ruihai Wu, Dongjiang Li, Hui Shen, Hao Dong*

95 ForeAct: Steering Your VLA with Efficient Visual Foresight Planning,
❊ *Zhuoyang Zhang, Shang Yang, Qinghao Hu, Luke J. Huang, James Hou, Yufei Sun, Yao Lu, Song Han*

96 Affordance Field Intervention: Enabling VLAs to Escape Memory Traps in Robotic Manipulation, *Siyu Xu, Zijian Wang, Yunke Wang, Chenghao Xia, Tao Huang, Chang Xu*

97 MERIT: Multi-domain Efficient RAW Image Translation, *Wenjun Huang, Shenghao Fu, Yian Jin, Yang Ni, Ziteng Cui, Hanning Chen, Yirui He, Yezi Liu, Sanggeon Yun, SungHeon Jeong, Ryozo Masukawa, William Youngwoo Chung, Mohsen Imani*

98 Pico-Banana-400K: A Large-Scale Dataset for Text-Guided Image Editing, *Yusu Qian, Eli Bocek-Rivele, Liangchen Song, Jialing Tong, Yinfei Yang, Jiasen Lu, Wenze Hu, Zhe Gan*

99 Probabilistic Prompt Adaptation for Unified Image Aesthetics and Quality Assessment, *Takayuki Hara, Yuya Otsuka*

100 EMMA: Concept Erasure Benchmark with Comprehensive Semantic Metrics and Diverse Categories, *Lu Wei, Yuta Nakashima, Noa Garcia*

101 Too Vivid to Be Real? Benchmarking and Calibrating Generative Color
❊ Fidelity, *Zhengyao Fang, Zexi Jia, Yijia Zhong, Pengcheng Luo, Jinchao Zhang, Guangming Lu, Jun Yu, Wenjie Pei*

102 WiseEdit: Benchmarking Cognition- and Creativity-Informed Image Editing, *Kaihang Pan, Weile Chen, Haiyi Qiu, Qifan Yu, Wendong Bu, Zehan Wang, Yun Zhu, Juncheng Li, Siliang Tang*

103 UnicEdit-10M: A Dataset and Benchmark Breaking the Scale-Quality Barrier via Unified Verification for Reasoning-Enriched Edits, *Keming Ye, Zhipeng Huang, Canmiao Fu, Qingyang Liu, Jiani Cai, Zheqi Lv, Chen Li, Jing LYU, Zhou Zhao, Shengyu Zhang*

104 Inter-Edit: First Benchmark for Interactive Instruction-Based Image Editing, *Delong Liu, Haotian Hou, Zhaohui Hou, Zhiyuan Huang, Shihao Han, Mingjie Zhan, Zhicheng Zhao, Fei Su*

105 PR-IQA: Partial-Reference Image Quality Assessment for Diffusion-Based Novel View Synthesis, *Inseong Choi, Siwoo Lee, Seung-Hun Nam, Soohwan Song*

106 LumiMotion: Improving Gaussian Relighting with Scene Dynamics,
❊ *Joanna Kaleta, Piotr Wójcik, Kacper Marzol, Tomasz Trzcinski, Kacper Kania, Marek Kowalski*

107 Let it Snow! Animating 3D Gaussian Scenes with Dynamic Weather Effects via Physics-Guided Score Distillation, *Gal Fiebelman, Hadar Averbuch-Elor, Sagie Benaim*

108 iLRM: An Iterative Large 3D Reconstruction Model, *Gyeongjin Kang, Seungtae Nam, Seungkwon Yang, Xiangyu Sun, Sameh Khamis, Abdelrahman Mohamed, Eunbyung Park*

109 MVInverse: Feed-forward Multiview Inverse Rendering in Seconds, *Xiangzuo Wu, Chengwei Ren, Jun Zhou, Xiu Li, Yuan Liu*

110 From None to All: Self-Supervised 3D Reconstruction via Novel View Synthesis, *Ranran Huang, Weixun Luo, Ye Mao, Krystian Mikolajczyk*
111 MoRel: Long-Range Flicker-Free 4D Motion Modeling via Anchor
❊ Relay-based Bidirectioanl Blending with Hierarchical Densification, *Sangwoon Kwak, Weeyoung Kwon, Jun Young Jeong, Geonho Kim, Won-Sik Cheong, Jihyong Oh*
112 Multi-view Pyramid Transformer: Look Coarser to See Broader, *Gyeongjin Kang, Seungkwon Yang, Seungtae Nam, Younggeun Lee, Jungwoo Kim, Eunbyung Park*
113 CaT-GS: Efficient 3DGS Rendering for Large Scale Scenes via Inter-frame Caching and Tile Scheduling, *Tingjia Zhang, Bo Chen, Shengzhong Liu, Fan Wu, Guihai Chen*
114 RL-ScanIQA: Reinforcement-Learned Scanpaths for Blind 360° Image Quality Assessment, *Yujia Wang, Yuyan Li, Jiuming Liu, Fang-Lue Zhang, Xinhu Zheng, Neil.A Dodgson*
115 Benchmarking Endoscopic Surgical Image Restoration and Beyond, *Jialun Pei, Diandian Guo, Donghui Yang, Zhixi Li, Yuxin Feng, Long Ma, Bo Du, Pheng-Ann Heng*
116 SDUIE: Semi-Supervised Diffusion for Underwater Image Enhancement
❊ with Quant-Text Dual Control, *Xiaofeng Cong, Yu-Xin Zhang, Hao Shen, Yeying Jin, Junming Hou, Jie Gui*
117 HiDRA: Hierarchical Degradation Representation and Adaptation with Generative Priors for Enhancing Infrared Vision, *Zihang Chen, Zhu Liu, Changbo Yan, Jinyuan Liu, Risheng Liu*
118 BluRef: Unsupervised Image Deblurring with Dense-Matching References, *Bang-Dang Pham, Anh Tran, Cuong Pham, Minh Hoai*
119 Bi-Bridge: Bidirectional Diffusion Bridges for Low-Light Image Enhancement, *Zeyu Hua, Hui Li, Yu Wang, Song Wang, Congchao Zhu, Caixia Zheng*
120 UniLDiff: Unlocking the Power of Diffusion Priors for All-in-One Image Restoration, *Zihan Cheng, Liangtai Zhou, Dian Chen, Ni Tang, Xiaotong Luo, Yuan Xie, Yanyun Qu*
121 MatAnyone 2: Scaling Video Matting via a Learned Quality
❊ Evaluator, *Peiqing Yang, Shangchen Zhou, Kai Hao, Qingyi Tao*
122 SelfHVD: Self-Supervised Handheld Video Deblurring, *Honglei Xu, Zhilu Zhang, Junjie Fan, Xiaohe Wu, Wangmeng Zuo*
123 Spatio-Temporal Difference Guided Motion Deblurring with the Complementary Vision Sensor, *Yapeng Meng, Lin Yang, Yuguo Chen, Xiangru Chen, Taoyi Wang, Lijian Wang, Zheyu Yang, Yihan Lin, Rong Zhao*
124 Learning Where to Look and How to Judge: Resolution-agnostic Image Quality Assessment with Quality-aware Saliency, *Hakan Emre Gedik, Shashank Gupta, Alan Bovik*
125 Bridging RGB and Hematoxylin Components: An Interleaved Guidance and Fusion Framework for Point Supervised Nuclei Segmentation, *Zihan Huan, Xipeng Pan, Hualong Zhang, Siyang Feng, Rushi Lan, Huadeng Wang, Haoxiang Lu, Zhenbing Liu*
126 Virtual Nodes Guided Dynamic Graph Neural Network for Brain Tumor Segmentation with Missing Modalities, *Sha Tao, Jiao Pan, Yu Guo, Chao Yao*
127 VoxTell: Free-Text Promptable Universal 3D Medical Image Segmentation, *Maximilian Rokuss, Moritz Langenberg, Yannick Kirchhoff, Fabian Isensee, Benjamin Hamm, Constantin Ulrich, Sebastian Regnery, Lukas Bauer, Efthimios Katsigiannopulos, Tobias Norajitra, Klaus Maier-Hein*
128 Photo-Guided Tooth Segmentation on 3D Oral Scan Model, *Shaojie Zhuang, Guangshun Wei, Jiangxin He, Yuanfeng Zhou*
129 Breaking the Continuum: Discrete Distribution Learning for Structural MRI Reconstruction, *Tianle Lyu, Mengjingcheng Mo, Ting Wen, Zhen Song, Zinan Xiong, Yanjie Zhu*
130 Uni-Hema: Unified Model for Digital Hematopathology, *Abdul Rehman, Iqra Rasool, Ayisha Imran, Mohsen Ali, Waqas Sultani*
131 Post-training Feature Pruning for Fundus Images Classification, *Van-Nguyen Pham, Duc-Tai Le, Junghyun Bum, Hyunseung Choo*
132 Sketch2CT: Multimodal Diffusion for Structure-Aware 3D Medical Volume Generation, *Delin An, Chaoli Wang*
133 SafeLogo: Turning Your Logos into Jailbreak Shields via Micro-Regional Adversarial Training, *Zhiyi Duan, Xiaoyue Zhang, Tianxing Man*
134 Anti-I2V: Safeguarding your Photos from Malicious Image-to-video Generation, *Duc Vu, Anh Nguyen, Chi Tran, Anh Tran*
135 UniGame: Turning a Unified Multimodal Model Into Its Own Adversary, *Zhaolong Su, Wang Lu, Hao Chen, Sharon Li, Jindong Wang*
136 Hierarchically Robust Zero-shot Vision-language Models, *Junhao Dong, Yifei Zhang, Hao Zhu, Yew-Soon Ong, Piotr Koniusz*
137 Beyond Text Prompts: Precise Concept Erasure through Text–Image Collaboration, *Jun Li, Lizhi Xiong, Ziqiang Li, Weiwei Jiang, Zhangjie Fu, Yong Li, Guo-Sen Xie*
138 AGENTSAFE: Benchmarking the Safety of Embodied Agents on Hazardous Instructions, *Zonghao Ying, Le Wang, Yisong Xiao, Jiakai Wang, Yuqing Ma, Jinyang Guo, Zhenfei Yin, Mingchuan Zhang, Aishan Liu, Xianglong Liu*
139 ReMoE: Region-Mixture Experts for Adversarially-Robust Vision Transformers, *Qinghao Zhong, Bingzhi Chen, Yishu Liu, Minhua Lu, Guangming Lu*
140 TreeTeaming: Autonomous Red-Teaming of Vision-Language Models via Hierarchical Strategy Exploration, *Chunxiao Li, Lijun Li, Jing Shao*
141 SO-Bench: A Structural Output Evaluation of Multimodal LLM, *Di Feng, Kaixin Ma, Feng Nan, Haofeng Chen, Bohan Zhai, David Griffiths, Mingfei Gao, Zhe Gan, Eshan Verma, Yinfei Yang, Zhifeng Chen, Afshin Dehghan*
142 Chain-of-Thought Guided Multi-Modal Object Re-Identification, *Ya Gao, Shihao Li, Zhaojun Liu, Aihua Zheng, Chenglong Li, Jin Tang*
143 When Lines Meet Textures: Spatial-Frequency Aligned Diffusion Features for Cross-Sparsity Correspondence, *Mingrui Zhu, Fengzhi Wang, Xin Wei, Jun Wang, Nannan Wang, Xinbo Gao*
144 CountGD++: Generalized Prompting for Open-World Counting, *Niki Amini-Naieni, Andrew Zisserman*
145 AudioStory: Generating Long-Form Narrative Audio with Large Language Models, *Yuxin Guo, Teng Wang, Yuying Ge, Shijie Ma, Yixiao Ge, Wei Zou*
146 Parameter-Efficient Adaptation for MLLMs via Implicit Modality Decomposition, *Mingfang Zhang, Yunhong Wang, Lu Wang, Jiaxin Chen*
147 Hyperbolic Gramian Volumes for Multimodal Alignment, *Saiyang Na, Feng Jiang, Qifeng Zhou, Wenliang Zhong, Thao M. Dang, Yuzhi Guo, Hehuan Ma, Chunyuan Li, Weizhi An, Junzhou Huang*
148 Venus: Benchmarking and Empowering Multimodal Large Language Models for Aesthetic Guidance and Cropping, *Tianxiang Du, Hulingxiao He, Yuxin Peng*
149 AutoCut: End-to-end advertisement video editing based on multimodal discretization and controllable generation, *Milton Zhou, Sizhong Qin, Yongzhi Li, Quan Chen, Peng Jiang*
150 StableMTL: Repurposing Latent Diffusion Models for Multi-Task Learning from Partially Annotated Synthetic Datasets, *Anh-Quan Cao, Ivan Lopes, Raoul de Charette*
151 CaReFlow: Cyclic Adaptive Rectified Flow for Multimodal Fusion, *Sijie Mai, Shiqin Han*
152 Lenses: Toward Polysemous Vision–Language Understanding, *Hani Alomari, Ali Asgarov, Chris Thomas*
153 CoRiM: Conflict-driven Risk Minimization for Dynamic Multimodal Fusion, *Shihao Zou, Wei Wei*
154 Uncertainty-Aware Exploratory Direct Preference Optimization for Multimodal Large Language Models, *Huatian Zhang, Zhendong Mao, Lei Zhang, Yongdong Zhang*
155 CICA: Coupling Confidence-Aware Pretraining with Confidence-Informed Attention for Robust Multimodal Sentiment Analysis, *Haoyu Jiang, Xiaoliang Chen, Duoqian Miao, Xiaolin Qin, Xianyong Li, Yajun Du*
156 SAMTok: Representing Any Mask with Two Words, *Yikang Zhou,*
❊ *Tao Zhang, Dengxian Gong, Yuanzheng Wu, Ye Tian, Haochen Wang, Haobo Yuan, Jiacong Wang, Lu Qi, Hao Fei, Shunping Ji, Anran Wang, Zhuochen Wang, Yujing Wang, Cheng Chen, Xiangtai Li*
157 Multi-Metric Representation Learning Strategy Based on Clustering for Fine-Grained Multimodal Sentiment Analysis, *Yidan Wang, Zongheng Wang, Hongjie Xing, Chunguo Li, Xiaoxiao Liu*
158 Cinematic Audio Source Separation Using Visual Cues, *Kang Zhang, Suyeon Lee, Arda Senocak, Joon Son Chung*
159 MMSD3.0: A Multi-Image Benchmark for Real-World Multimodal Sarcasm Detection, *Haochen Zhao, Yuyao Kong, Yongxiu Xu, Gaopeng Gou, Hongbo Xu, Yubin Wang, Haoliang Zhang*
160 Anchor-Guided Gradient Alignment for Incomplete Multimodal Learning, *Zhi-Hao Guan, Longfei Huang, Yang Yang*
161 PyraTok: Language-Aligned Pyramidal Tokenizer for Video Understanding and Generation, *Onkar Susladkar, Tushar Prakash,*

Adheesh Juvekar, Kiet A. Nguyen, Dong-Hwan Jang, Inderjit S Dhillon, Ismini Lourentzou

162 VDE: Training-Free Accelerating Rectified Flow Model via Velocity Decomposition and Estimation, *Junwen Tan, Jinglin Liang, Hongyuan Chen, Shuangping Huang*

163 Kontinuous Kontext: Continuous Strength Control for Instruction-based Image Editing, *Rishubh Parihar, Or Patashnik, Daniil Ostashev, Venkatesh Babu Radhakrishnan, Daniel Cohen-Or, Kuan-Chieh Jackson Wang*

164 ❊ VideoCoF: Unified Video Editing with Temporal Reasoner, *Xiangpeng Yang, Ji Xie, Yiyuan Yang, Yue Ma, Yan Huang, Min Xu, Qiang Wu*

165 Progressive Supernet Training for Efficient Visual Autoregressive Modeling, *Xiaoyue Chen, Yuling Shi, Kaiyuan Li, Huandong Wang, Yong Li, Xiaodong Gu, Xinlei Chen, Mingbao Lin*

166 CoT-Edit: Let CoT Guide Instruction Video Editing, *Sen Liang, Fengbin Guan, Youliang Zhang, Xin Li, Zhibo Chen*

167 ❊ Scaling Instruction-Based Video Editing with a High-Quality Synthetic Dataset, *Qingyan Bai, Qiuyu Wang, Hao Ouyang, Yue Yu, Hanlin Wang, Wen Wang, Ka Leong Cheng, Shuailei Ma, Yanhong Zeng, Zichen Liu, Yinghao Xu, Yujun Shen, Qifeng Chen*

168 Test-Time Instance-Specific Parameter Composition: A New Paradigm for Adaptive Generative Modeling, *Minh-Tuan Tran, Xuan-May Le, Quan Hung Tran, Mehrtash Harandi, Dinh Phung, Trung Le*

169 Understanding, Accelerating, and Improving MeanFlow Training, *Jin-Young Kim, Hyojun Go, Lea Bogensperger, Julius Erbach, Nikolai Kalischek, Federico Tombari, Konrad Schindler, Dominik Narnhofer*

170 Meta-CoT: Enhancing Granularity and Generalization in Image Editing, *Shiyi Zhang, Yiji Cheng, Tiankai Hang, Zijin Yin, Runze He, Yu Xu, Wenxun Dai, Yunlong Lin, Chunyu Wang, Qinglin Lu, Yansong Tang*

171 Dual-Granularity Memory for Efficient Video Generation, *Hongjun Wang, Lin Liu, Jianguo Li, Tao Lin*

172 Unified Camera Positional Encoding for Controlled Video Generation, *Cheng Zhang, Boying Li, Meng Wei, Yan-Pei Cao, Camilo Gambardella, Dinh Phung, Jianfei Cai*

173 EditMGT: Unleashing Potentials of Masked Generative Transformers in Image Editing, *Wei Chow, Linfeng Li, Lingdong Kong, Zefeng Li, Qi Xu, Hang Song, Tian Ye, Xian Wang, Jinbin Bai, Shilin Xu, Xiangtai Li, Junting Pan, Shaoteng Liu, Ran Zhou, Tianshu Yang, Songhua Liu*

174 MU-GeNeRF: Multi-view Uncertainty-guided Generalizable Neural Radiance Fields for Distractor-aware Scene, *Wenjie Mu, Zhan Li, Chuanzhou Su, Xuanyi Shen, Ziniu Liu, Fan Lu, Yujian Mo, Junqiao Zhao, Tiantian Feng, Chen Ye, Guang Chen*

175 PLACID: Identity-Preserving Multi-Object Compositing via Video Diffusion with Synthetic Trajectories, *Gemma Canet Tarrés, Manel Baradad, Francesc Moreno-Noguer, Yumeng Li*

176 Object-WIPER: Training-Free Object and Associated Effect Removal in Videos, *Saksham Singh Kushwaha, Sayan Nag, Yapeng Tian, Kuldeep Kulkarni*

177 Mobile-VTON: High-Fidelity On-Device Virtual Try-On, *Zhenchen Wan, Ce Chen, Runqi Lin, Jiaxin Huang, Tianxi Chen, Yanwu Xu, Tongliang Liu, Mingming Gong*

178 Progress by Pieces: Test-Time Scaling for Autoregressive Image Generation, *Joonhyung Park, Hyeongwon Jang, Joowon Kim, Eunho Yang*

179 Towards Robust Sequential Decomposition for Complex Image Editing, *Zilai Zeng, Mingdeng Cao, Zijie Li, Xiaochen Lian, Yichun Shi, Peihao Zhu, Chen Sun, Peng Wang*

180 Layer Consistency Matters: Elegant Latent Transition Discrepancy for Generalizable Synthetic Image Detection, *Yawen Yang, Feng Li, Shuqi Kong, Yunfeng Diao, Xinjian Gao, Zenglin Shi, Meng Wang*

181 Chain of Event-Centric Causal Thought for Physically Plausible Video Generation, *Zixuan Wang, Yixin Hu, Haolan Wang, Feng Chen, Yan Liu, Wen Li, Yinjie Lei*

182 LoL: Longer than Longer, Scaling Video Generation to Hour, *Justin Cui, Jie Wu, Ming Li, Tao Yang, Xiaojie Li, Rui Wang, Andrew Bai, Yuanhao Ban, Cho-Jui Hsieh*

183 FlowMotion: Training-Free Flow Guidance for Video Motion Transfer, *Zhen Wang, Youcan Xu, Jun Xiao, Long Chen*

184 Learning Straight Flows: Variational Flow Matching for Efficient Generation, *Chenrui Ma, Xi Xiao, Tianyang Wang, Xiao Wang, Yanning Shen*

185 SIGMA: Selective-Interleaved Generation with Multi-Attribute Tokens, *Xiaoyan Zhang, Zechen Bai, Haofan Wang, Yiren Song*

186 DNF-SR: Dual-Input and Negative-Aware Feature Fine-Tuning for Real-World Image Super-Resolution, *Shuhao Han, Wenjie Liao, Hayden Vance, Hang Dong, Rui Zhang, Chun-Le Guo, Chongyi Li*

187 IFCSR: Inference-Free Fidelity-Realism Control for One-Step Diffusion-based Real-World Image Super-Resolution, *Jonghee Back, Jongju Kim, Jeong-Uk Kim, Eunjin Kim, Minyong Jeon*

188 Edge-Focused Super-Resolution for Omnidirectional Images with Spherical Geometric Augmentation, *Shaolin Wang, Yuying Li, Lei Zhong, Shigang Li, Jianfeng Li*

189 TUDSR: Twice Upsampling-Diffusion for Higher Super-Resolution, *Zhiqiang Wu, Yitong Dong, Xian Wei*

190 PS-SR: Pseudo-Single-Step Video Super-Resolution via Speculative Diffusion, *Aiqiu Wu, Zhaofan Qiu, Ting Yao, Tao Mei*

191 Disentangled Textual Priors for Diffusion-based Image Super-Resolution, *Lei Jiang, Xin Liu, Xinze Tong, Zhiliang Li, Jie Liu, Jie Tang, Gangshan Wu*

192 Remote Sensing Image Super-Resolution for Imbalanced Textures: A Texture-Aware Diffusion Framework, *Enzhuo Zhang, Sijie Zhao, Dilxat Muhtar, Zhenshi Li, Xueliang Zhang, Pengfeng Xiao*

193 Rethinking Diffusion Model-Based Video Super-Resolution: Leveraging Dense Guidance from Aligned Features, *Jingyi Xu, Meisong Zheng, Ying Chen, Minglang Qiao, Xin Deng, Mai Xu*

194 DreamSR: Towards Ultra-High-Resolution Image Super-Resolution via a Receptive-Field Enhanced Diffusion Transformer, *Qingji Dong, Hang Dong, Mingqin Chen, Rui Zhang, Yitong Wang*

195 FiDeSR: High-Fidelity and Detail-Preserving One-Step Diffusion Super-Resolution, *Aro Kim, Myeongjin Jang, Chaewon Moon, Youngjin Shin, Jinwoo Jeong, Sang-hyo Park*

196 STCDiT: Spatio-Temporally Consistent Diffusion Transformer for High-Quality Video Super-Resolution, *Junyang Chen, Jiangxin Dong, Long Sun, Yixin Yang, Jinshan Pan*

197 Towards Highly-Constrained Human Motion Generation with Retrieval-Guided Diffusion Noise Optimization, *Hanchao Liu, Fang-Lue Zhang, Shining Zhang, Tai-Jiang Mu, Shi-Min Hu*

198 Learning to Control Physically-simulated 3D Characters via Generating and Mimicking 2D Motions, *Jianan Li, Xiao Chen, Tao Huang, Tien-Tsin Wong*

199 Human Geometry Distribution for 3D Animation Generation, *Xiangjun Tang, Biao Zhang, Peter Wonka*

200 A Temporal and Content Co-Awareness Latent Diffusion for Controllable Hand Image Generation, *Shuang Hao, Pengfei Ren, Haifeng Sun, Ting Pan, Qi Qi, Lei Zhang, Cong Liu, Jianxin Liao, Jingyu Wang*

201 Superman: Unifying Skeleton and Vision for Human Motion Perception and Generation, *Xinshun Wang, Peiming Li, Ziyi Wang, Zhongbin Fang, Zhichao Deng, Songtao Wu, Jason Li, Mengyuan Liu*

202 Learning to Assist: Physics-Grounded Human-Human Control via Multi-Agent Reinforcement Learning, *Yuto Shibata, Kashu Yamazaki, Lalit Jayanti, Yoshimitsu Aoki, Mariko Isogawa, Katerina Fragkiadaki*

203 Stability-Driven Motion Generation for Object-Guided Human-Human Co-Manipulation, *Jiahao Xu, Xiaohan Yuan, Xingchen Wu, Chongyang Xu, Kun Li, Buzhen Huang*

204 Causal Motion Diffusion Models for Autoregressive Motion Generation, *Qing Yu, Akihisa Watanabe, Kent Fujiwara*

205 Towards Storytelling Animations: Joint Synthesis of Human and Camera Motions, *Boyuan Cheng, Yingjie Xi, Rui He, Jinhe Na, Ying Cao, Pengjie Wang, Jian J. Zhang, Xiaosong Yang*

206 MoLingo: Motion–Language Alignment for Text-to-Human Motion Generation, *Yannan He, Garvita Tiwari, Xiaohan Zhang, Pankaj Bora, Tolga Birdal, Jan Eric Lenssen, Gerard Pons-Moll*

207 ❊ End-to-End Language-Action Model for Humanoid Whole Body Control, *Yuxuan Wang, Haobin Jiang, Shiqing Yao, Ziluo Ding, Zongqing Lu*

208 Toward Early Quality Assessment of Text-to-Image Diffusion Models, *Huanlei Guo, Hongxin Wei, Bingyi Jing*

209 CoD: A Diffusion Foundation Model for Image Compression, *Zhaoyang Jia, Zihan Zheng, Naifu Xue, Jiahao Li, Bin Li, Zongyu Guo, Xiaoyi Zhang, Houqiang Li, Yan Lu*

210 Diffusion MRI Transformer with a Diffusion Space Rotary Positional Embedding (D-RoPE), *Gustavo Chau Loo Kung, Mohammad Abbasi, Camila Blank, Juze Zhang, Alan Q. Wang, Sophie Ostmeier, Akshay Chaudhari, Kilian Pohl, Ehsan Adeli*
211 Language-Guided One-Step Diffusion Model for Nighttime Flare Removal, *Aoxiang Ning, Kailong Yu, Minglong Xue, Liyuan Pan, Jinhong He, Wenchao Yan, Mingliang Zhou, Yirui Wu*
212 SpiralDiff: Spiral Diffusion with LoRA for RGB-to-RAW Conversion Across Cameras, *Huanjing Yue, Shangbin Xie, Cong Cao, Qian Wu, Lei Zhang, Lei Zhao, Jingyu Yang*
213 PnP-CM: Consistency Models as Plug-and-Play Priors for Inverse Problems, *Merve Gulle, Junno Yun, Yasar Utku Alcalar, Mehmet Akcakaya*
214 Landscape-Awareness for Geometric View Diffusion Model, *Yan-Ting Chen, Hao-Wei Chen, Tsu-Ching Hsiao, Chun-Yi Lee*
215 Otil: Accelerating Diffusion Model Inference via Communication-Efficient Multi-GPU Parallelism, *Xin Li, Shujun Tian, Tao Lu, Han Bao, Zonghui Wang, Wenzhi Chen*
216 REACH: Explicit Recovery Behavior for Diffusion Policies, *Zundong Ke, Junlin Chen, Jiayi Zhu, Kuanhao Xia, Boyi Zhao, Jiayuan Gu*
217 OralGPT-Omni: A Versatile Dental Multimodal Large Language Model, *Jing Hao, Yuci Liang, Lizhuo Lin, Yuxuan Fan, Wenkai Zhou, Kaixin Guo, Zanting Ye, Yanpeng Sun, Xinyu Zhang, Yanqi Yang, Qiankun Li, Hao Tang, James Kit-Hon Tsoi, Linlin Shen, Kuo Feng Hung*
218 CrossHOI-Bench: A Unified Benchmark for HOI Evaluation across Vision-Language Models and HOI-Specific Methods, *Qinqian Lei, Bo Wang, Robby T. Tan*
219 The LLM Bottleneck: Why Open-Source Vision LLMs Struggle with Hierarchical Visual Recognition, *Yuwen Tan, Yuan Qing, Boqing Gong*
220 Do Vision-Language Models Measure Up? Benchmarking Visual Measurement Reading with MeasureBench, *Fenfen Lin, Yesheng Liu, Haiyu Xu, Yue Chen, Zheqi He, Mingxuan Zhao, Miguel Hu Chen, Jin-Ge Yao, Xi Yang*
221 KαLOS finds Consensus: A Meta-Algorithm for Evaluating Inter-Annotator Agreement in Complex Vision Tasks, *David Tschirschwitz, Volker Rodehorst*
222 Beyond Single Images: A Comprehensive Benchmark for Album-Level ❊ Vision-Language Understanding, *Shawn Huang, Brian Price, Yifei Fan, Bryan Morse*
223 LIBERO-Plus: A Progressive Robustness Benchmark for Visual-Language-Action Models, *Senyu Fei, Siyin Wang, Junhao Shi, Zihao Dai, Jikun Cai, Pengfang Qian, Li Ji, Xinzhe He, Shiduo Zhang, Zhaoye Fei, Jinlan Fu, Jingjing Gong, Xipeng Qiu*
224 Scenes as Tokens: Multi-Scale Normal Distributions Transform Tokenizer for General 3D Vision–Language Understanding, *Yutao Tang, Cheng Zhao, Gaurav Mittal, Rohith Kukkala, Rama Chellappa, Cheng Peng, Mei Chen*
225 LangRef3DGS: Natural Language-Guided 3D Referential Segmentation from Partial Observations via 3D Gaussian Splatting, *Xulun Ye, Qin Zhang, Kun Zhou*
226 Hear you are: Teaching LLMs Spatial Reasoning with Vision and Spatial Sound, *Hyeonggon Ryu, Joon Son Chung, David Harwath*
227 EgoMind: Activating Spatial Cognition through Linguistic Reasoning in MLLMs, *Zhenghao Chen, Huiqun Wang, Di Huang*
228 SAQN: Semantic-based Adaptive Query Network for 3D Referring ❊ Expression Segmentation, *Jiale Huang, Shangfei Wang*
229 EagleVision: A Dual-Stage Framework with BEV-grounding-based Chain-of-Thought for Spatial Intelligence, *Jiaxu Wan, Xu Wang, Mengwei Xie, Hang Zhang, Mu Xu, Yang Han, Ding Yuan, Hong Zhang, Yifan Yang*
230 Abstract 3D Perception for Spatial Intelligence in Vision-Language Models, *Yifan Liu, Fangneng Zhan, Kaichen Zhou, Yilun Du, Paul Pu Liang, Hanspeter Pfister*
231 PV-Ground: Text-Guided Point-Voxel Interaction for 3D Visual Grounding, *Junpeng Shang, Feifei Shao, Jun Xiao, Lin Li, Hongwei Wang, Dongfang Ma*
232 Masking Matters: Unlocking the Spatial Reasoning Capabilities of LLMs for 3D Scene-Language Understanding, *Yerim Jeon, Miso Lee, WonJun Moon, Jae-Pil Heo*
233 SpatialStack: Layered Geometry-Language Fusion for 3D VLM Spatial Reasoning, *Jian Zhang, Shijie Zhou, Bangya Liu, Achuta Kadambi, Zhiwen Fan*
234 Geometrically-Constrained Agent for Spatial Reasoning, *Zeren Chen, Xiaoya Lu, Zhijie Zheng, Pengrui Li, Lehan He, Yijin Zhou, Jing Shao, Bohan Zhuang, Lu Sheng*
235 PARSE: Part-Aware Relational Spatial Modeling, *Yinuo Bai, Peijun Xu, Kuixiang Shao, Yuyang Jiao, Jingxuan Zhang, Kaixin Yao, Jiayuan Gu, Jingyi Yu*
236 R4: Retrieval-Augmented Reasoning for Vision-Language Models in 4D Spatio-Temporal Space, *Tin Stribor Sohn, Maximilian Dillitzer, Jason J. Corso, Eric Sax*
237 MCHDoc: A Comprehensive Benchmark for Reading Multi-Carrier Chinese Historical Documents, *Yijun Sheng, Shipeng Zhu, Ruijia Zuo, Na Nie, Hui Xue*
238 Cross-modal Fuzzy Alignment Network for Text-Aerial Person Retrieval and A Large-scale Benchmark, *Yifei Deng, Chenglong Li, Yuyang Zhang, Guyue Hu, Jin Tang*
239 CodeMMR: Bridging Natural Language, Code, and Image for Unified Retrieval, *Jiahui Geng, Qing Li, Fengyu Cai, Fakhri Karray*
240 DiT-Distill: Open-Set Fine-Grained Retrieval via Generative Curriculum Knowledge, *Xin Jiang, Hao Tang, Meiqi Cao, Junyao Gao, Fei Shen, Zechao Li*
241 ReCALL: Recalibrating Capability Degradation for MLLM-based Composed Image Retrieval, *Tianyu Yang, Chenwei He, Xiangzhao Hao, Tianyue Wang, Jiarui Guo, Haiyun Guo, Leigang Qu, Jinqiao Wang, Tat-Seng Chua*
242 Love Me, Love My Label: Rethinking the Role of Labels in Prompt Retrieval for Visual In-Context Learning, *Tianci Luo, Haohao Pan, Jinpeng Wang, Niu Lian, Xinrui Chen, Bin Chen, Shu-Tao Xia, Chun Yuan*
243 Rethinking BCE Loss for Multi-Label Image Recognition with Fine-Tuning, *Ao Zhou, Zhiwei Jiang, Zifeng Cheng, Cong Wang, Yafeng Yin, Shufan Yang, Qing Gu*
244 CAST: Context-Aware Dynamic Latent Space Transformation for Interactive Text-to-Image Retrieval, *Xuanzuo Lin, Min Zhang, Daizong Liu, Zhiwen Zuo, Xun Yang, Changting Lin, Xun Wang, Jianfeng Dong*
245 PriVi: Towards a General-Purpose Video Model for Primate Behavior in the Wild, *Felix B. Mueller, Jan F. Meier, Timo Lueddecke, Richard Vogg, Roger L. Freixanet, Valentin Hassler, Tiffany Bosshard, Elif Karakoc, William J. O'Hearn, Sofia M. Pereira, Sandro Sehner, Kaja Wierucka, Judith Burkart, Claudia Fichtel, Julia Fischer, Alexander Gail, Catherine Hobaiter, Julia Ostner, Liran Samuni, Oliver Schülke, Neda Shahidi, Erin G. Wessling, Alexander S. Ecker*
246 Seeing Conversations: Communication Context Identification in Egocentric Video, *Tobias Dorszewski, Jens Hjortkjær*
247 Interactive Episodic Memory with User Feedback, *Nikesh Subedi, Loris Bazzani, Ziad Al-Halah*
248 Seeing without Pixels: Perception from Camera Trajectories, *Zihui Xue, Kristen Grauman, Dima Damen, Andrew Zisserman, Tengda Han*
249 PFGNet: A Fully Convolutional Frequency-Guided Peripheral Gating Network for Efficient Spatiotemporal Predictive Learning, *Xinyong Cai, Changbin Sun, Yong Wang, Hongyu Yang, Yuankai Wu*
250 Minerva-Ego: Spatiotemporal Hints for Egocentric Video Understanding, *Arsha Nagrani, Jasper Uijlings, Shyamal Buch, Tobias Weyand, Sudheendra Vijayanarasimhan, Bo Hu, Ramin Mehran, David A Ross, Cordelia Schmid*
251 StreamRAG: Enhancing Real-Time Video Understanding with Retrieval Augmentation, *Junlin Xie, Quanlong Zheng, Ruifei Zhang, Kuo Wang, Yanhao Zhang, Jinguo Luo, Haonan Lu, Xiang Wan, Guanbin Li*
252 ViKey: Enhancing Temporal Understanding in Videos via Visual Prompting, *Yeonkyung Lee, Dayun Ju, Youngmin Kim, Seil Kang, Seong Jae Hwang*
253 SkillSight: Efficient First-Person Skill Assessment with Gaze, *Chi Hsuan Wu, Ashutosh Kumar, Kristen Grauman*
254 BriMA: Bridged Modality Adaptation for Multi-Modal Continual Action Quality Assessment, *Kanglei Zhou, Chang Li, Qingyi Pan, Liyuan Wang*

255 Video-as-Answer: Predict and Generate Next Video Event with Joint-GRPO, *Junhao Cheng, Liang Hou, Xin Tao, Jing Liao*
256 MedLIME: A Distribution-Aligned and Evidence-Supported Framework
❊ for Medical Saliency Explanations, *Raghav Magazine, Xingjian Li, Min Xu*
257 Inside-Out: Measuring Generalization in Vision Transformers Through
❊ Inner Workings, *Yunxiang Peng, Mengmeng Ma, Ziyu Yao, Xi Peng*
258 Language Models Can Explain Visual Features via Steering, *Javier Ferrando, Enrique Lopez-Cuena, Pablo Agustin Martin-Torres, Daniel Hinjos, Anna Arias-Duart, Dario Garcia-Gasulla*
259 Making the Classification Explanation Faithful to the Confidence Score, *Jian-Xun Mi, Lu Pan, Weisheng Li*
260 Intrinsic Concept Extraction Based on Compositional Interpretability, *Hanyu Shi, Hong Tao, Guoheng Huang, Jianbin Jiang, Xuhang Chen, Chi-Man Pun, Shanhu Wang, Pan Pan*
261 Attribution-Guided Model Rectification of Unreliable Neural Network Behaviors, *Peiyu Yang, Naveed Akhtar, Jiantong Jiang, Ajmal Mian*
262 Measuring the (Un)Faithfulness of Concept-Based Explanations, *Shubham Kumar, Narendra Ahuja*
263 Deformation-based In-Context Learning for Point Cloud Understanding, *Chengxing Lin, Jinhong Deng, Yinjie Lei, Wen Li*
264 ELiC: Efficient LiDAR Geometry Compression via Cross-Bit-depth Feature Propagation and Bag-of-Encoders, *Junsik Kim, Gun Bang, Soowoong Kim*
265 ESAM++: Efficient Online 3D Perception on the Edge, *Qin Liu, Lavisha Aggarwal, Saptarashmi Bandyopadhyay, Vikas Bahirwani, Marc Niethammer, Ehsan Adeli, Andrea Colaco*
266 DualReg: Dual-Space Filtering and Reinforcement for Rigid Registration, *Jiayi Li, Yuxin Yao, Qiuhang Lu, Juyong Zhang*
267 Hg-I2P: Bridging Modalities for Generalizable Image-to-Point-Cloud Registration via Heterogeneous Graphs, *Pei An, Junfeng Ding, Jiaqi Yang, Yulong Wang, Jie Ma, Liangliang Nan*
268 Rethinking 2D-3D Registration: A Novel Network for High-Value Zone Selection and Representation Consistency Alignment, *Zhixin Cheng, Bohao Liao, Jiacheng Deng, Xiaotian Yin, Xinjun Li, Yujia Chen, Baoqun Yin, Tianzhu Zhang*
269 Adaptive 3D Perception for Small Aerial Targets Under Sparse Sampling via Reinforcement Learning, *Shenghai Yuan, Wei Yihan, Jason Yee, Zhuoran Qiao, boyang lou, Enwen Hu*
270 3D sans 3D Scans: Scalable Pre-training from Video-Generated Point Clouds, *Ryousuke Yamada, Kohsuke Ide, Yoshihiro Fukuhara, Hirokatsu Kataoka, Gilles Puy, Andrei Bursuc, Yuki M. Asano*
271 StreamVLO: Streaming Visual–LiDAR Odometry with Cumulative Drift
❊ Compensation, *Mengmeng Liu, Jiuming Liu, Michael Ying Yang, Chaokang Jiang, Jiangtao Li, Yunpeng Zhang, Hesheng Wang, Francesco Nex, Hao Cheng*
272 Mamba Learns in Context: Structure-Aware Domain Generalization
❊ for Multi-Task Point Cloud Understanding, *Jincen Jiang, Qianyu Zhou, Yuhang Li, Kui Su, Meili Wang, Jian Chang, Jian Jun Zhang, Xuequan Lu*
273 Routing on Demand: DSNet for Efficient Progressive Point Cloud Denoising, *Xiaoqian Cheng, Dong Xiao, Husen Li, Zheng Liu, Renjie Chen*
274 Hyper-PCN: Hypergraph-Based Point Cloud Completion via High-Order Correlation Modeling, *Linfei Li, Pei Tan, Siqi Li, Changqing Zou, Yue Gao*
275 Towards Calibrating Prompt Tuning of Vision- Language Models, *Ashshak Sharifdeen, Fahad Shamshad, Muhammad Akhtar Munir, Abhishek Basu, Mohamed Ismithdeen, Jeyapriyan Jeyamohan, Chathurika Silva, Karthik Nandakumar, Muhammad Haris Khan*
276 DEVA: Fine-tuning Multimodal Large Language Models for Visual Perception Tasks, *Debasmit Das, Munawar Hayat, Fatih Porikli*
277 LOREAL: Mitigating Low-Resolution Challenges in Vision-Language
❊ Models with Attribute-driven Prompt Self-Distillation, *Xucong Wang, Pengkun Wang, Zhe Zhao, Liheng Yu, Rui Mao, Yang Wang*
278 OpenVision 2: A Family of Generative Pretrained Visual Encoders for Multimodal Learning, *Yanqing Liu, Xianhang Li, Letian Zhang, Zirui Wang, Zeyu Zheng, Yuyin Zhou, Cihang Xie*
279 Language-guided Frequency Modulation for Large Vision-Language Models, *Shuyi Ouyang, Gongfan Fang, Xinyin Ma, Yen-Wei Chen, Lanfen Lin, Xinchao Wang*
280 TANGO: Text-Anchored Guided Optimization for Robust Fine-tuning Vision-Language Models under Label Noise, *Tengfei Ma, Weiran Pan, Wei Wei*
281 Cluster-Wise Spatio-Temporal Masking for Efficient Video-Language Pretraining, *Weijun Zhuang, Yuqing Huang, Weikang Meng, Xin Li, Ming Liu, Xiaopeng Hong, Yaowei Wang, Wangmeng Zuo*
282 Reconstructing CLIP for Open-Vocabulary Dense Perception, *Yajie Liu, Jinjin Zhang, Qingjie Liu, Di Huang*
283 DPL: Decoupled Prototype Learning for Enhancing Robustness of Vision–Language Transformers to Missing Modalities, *Jueqing Lu, Yuanyuan Qi, Xiaohao Yang, Shuaicheng Niu, Fucai Ke, Shujie Zhou, Wei Tan, Jionghao Lin, Wray Buntine, Hamid Rezatofighi, Lan Du*
284 BrepVGAE: Variational Graph Autoencoder with Unified Latent Representation for B-rep, *Hao Guo, Liyuan Deng, Yongkang Dai, Ruohan Wang, Jiahao Li, Yunpeng Bai, Yilei Shi*
285 NeuROK: Generative 4D Neural Object Kinematics, *Chen Geng, Guangzhao He, Yue Gao, Yunzhi Zhang, Shangzhe Wu, Jiajun Wu*
286 BrickNet: Graph-Backed Generative Brick Assembly, *Peter Kulits, Cordelia Schmid*
287 Unified Vector Floorplan Generation via Markup Representation,
❊ *Kaede Shiohara, Toshihiko Yamasaki*
288 CME-CAD: Heterogeneous Collaborative Multi-Expert Reinforcement Learning for CAD Code Generation, *Ke Niu, Haiyang Yu, Zhuofan Chen, Zhengtao Yao, Weitao Jia, Xiaodong Ge, Jingqun Tang, Benlei Cui, Bin Li, Xiangyang Xue*
289 Robo-SGG: Exploiting Layout-Oriented Normalization and Restitution Can Improve Robust Scene Graph Generation, *Changsheng Lv, Zijian Fu, Mengshi Qi*
290 OmniLottie: Generating Vector Animations via Parameterized Lottie Tokens, *Yiying Yang, Wei Cheng, Sijin Chen, Honghao Fu, Xianfang Zeng, Yujun Cai, Gang Yu, Xingjun Ma*
291 EpiAgent: An Agent-Centric System for Ancient Inscription Restoration, *Shipeng Zhu, Ang Chen, Na Nie, Pengfei Fang, Min-Ling Zhang, Hui Xue*
292 The Invisible Gorilla Effect in Out-of-distribution Detection,
❊ *Harry Anthony, Ziyun Liang, Hermione Warr, Konstantinos Kamnitsas*
293 Interpretable Debiasing of Vision-Language Models for Social Fairness, *Na Min An, Yoonna Jang, Yusuke Hirota, Ryo Hachiuma, Isabelle Augenstein, Hyunjung Shim*
294 Image-based Outlier Synthesis With Training Data, *Sudarshan Regmi*
295 SALMUBench: A Benchmark for Sensitive Association-Level Multimodal Unlearning, *Cai Selvas-Sala, Lei Kang, Lluis Gomez*
296 Scaling Test-Time Robustness of Vision-Language Models via Self-Critical Inference Framework, *Kaihua Tang, Jiaxin Qi, Jinli Ou, Yuhua Zheng, Jianqiang Huang*
297 When Understanding Becomes a Risk: Authenticity and Safety Risks
❊ in the Emerging Image Generation Paradigm, *Ye Leng, Junjie Chu, Mingjie Li, Chenhao Lin, Chao Shen, Michael Backes, Yun Shen, Yang Zhang*
298 IrisFP: Adversarial-Example-based Model Fingerprinting with Enhanced Uniqueness and Robustness, *Ziye Geng, Guang Yang, Yihang Chen, Changqing Luo*
299 Mark4D: Temporally-Consistent Watermarking for 4D Gaussian Splatting, *Jaejin Lee, Minjae Jeong, Joonhyuk Park, Yechan Hwang, Seunghun Baek, Won Hwa Kim*
300 Machine Unlearning via Adaptive Gradient Reweighting and Multi-stage Objective Optimization, *Juxin Lu, Haoyu Shi, Mengyao Wang, Huaiwen Zhang*
301 Taming Noise-Induced Prototype Degradation for Privacy-Preserving
❊ Personalized Federated Fine-Tuning, *Yuhua Wang, Qinnan Zhang, Xiaodong Li, Huan Zhang, Yifan Sun, Wangjie Qiu, Hainan Zhang, Yongxin Tong, Zhiming Zheng*
302 FedMOP: Achieving Enhanced Privacy and Performance in Federated Learning via Momentum Orthogonal Projection, *Yunlong Zhao, Xiaoheng Deng, Hongyan Xu, Zhuohua Qiu, Xiaowen Hu, Shan You, Yi Chen, Chang Xu, Xiu Su*
303 HFedATM: Hierarchical Federated Domain Generalization via Optimal Transport and Regularized Mean Aggregation, *Thinh Nguyen, Trung Phan, Binh Nguyen, Khoa D Doan, Kok-Seng Wong*
304 Single-Round Scalable Analytic Federated Learning, *Alan T. L. Bacellar, Mustafa Munir, Felipe M. G. França, Priscila M. V. Lima, Radu Marculescu, Lizy K. John*

305 Controllable Federated Prompt Learning at Test Time, *Rui Zhu, Liang Bai, Yanming Guo, Yirun Ruan, Tianyuan Yu, Zhihe Lu*
306 FedRE: A Representation Entanglement Framework for Model-Heterogeneous Federated Learning, *Yuan Yao, Lixu Wang, Jiaqi Wu, Jin Song, Simin Chen, Zehua Wang, Zijian Tian, Wei Chen, Huixia Li, Xiaoxiao Li*
307 Conversational Image Segmentation: Grounding Abstract Concepts with Scalable Supervision, *Aadarsh Sahoo, Georgia Gkioxari*
308 Spatial Matters: Position-Guided 3D Referring Expression Segmentation, *Yabing Wang, Zhuotao Tian, Le Wang, Zheng Qin, Sanping Zhou*
309 Long-RVOS: A Comprehensive Benchmark for Long-term Referring Video Object Segmentation, *Tianming Liang, Haichao Jiang, Yuting Yang, Chaolei Tan, Shuai Li, Wei-Shi Zheng, Jian-Fang Hu*
310 Refer-Agent: A Collaborative Multi-Agent System with Reasoning and Reflection for Referring Video Object Segmentation, *Haichao Jiang, Tianming Liang, Wei-Shi Zheng, Jian-Fang Hu*
311 CaptionFormer: Unified Segmentation, Tracking, and Captioning for Spatio-Temporal Objects, *Gabriel Fiastre, Antoine Yang, Cordelia Schmid*
312 TransPrune: Token Transition Pruning for Efficient Large Vision-Language Model, *Ao Li, Yuxiang Duan, Jinghui Zhang, Congbo Ma, Yutong Xie, Gustavo Carneiro, Mohammad Yaqub, Hu Wang*
313 QuantVLA: Scale-Calibrated Post-Training Quantization for Vision-Language-Action Models, *Jingxuan Zhang, Yunta Hsieh, Zhongwei Wan, Haokun Lin, Xin Wang, Ziqi Wang, Yingtie Lei, Mi Zhang*
314 Revisiting Multimodal KV Cache Compression: A Frequency-Domain-Guided Outlier-KV-Aware Approach, *Yaoxin Yang, Peng Ye, Xudong Tan, Chongjun Tu, Maosen Zhao, Jia Hao, Tao Chen*
315 Collaborative Multi-Mode Pruning for Vision-Language Models, *Zimeng Wu, Yunhong Wang, Donghao Wang, Jiaxin Chen*
316 ZOO-Prune: Training-Free Token Pruning via Zeroth-Order Gradient Estimation in Vision-Language Models, *Youngeun Kim, Youjia Zhang, Huiling Liu, Aecheon Jung, Sunwoo Lee, Sungeun Hong*
317 HAWK: Head Importance-Aware Visual Token Pruning in Multimodal Models, *Qihui Zhu, Tao Zhang, Yuchen Wang, Shuangwu Chen, Xiaobin Tan, Jian Yang, Yang Liu, Yinfei Pan*
318 CORE: Compact Object-centric REpresentations as a New Paradigm for Token Merging in LVLMs, *Jingyu Lei, Gaoang Wang, Der-Horng Lee*
319 Imbalanced View Contribution Evaluation and Refinement for Deep Incomplete Multi-View Clustering, *Taichun Zhou, Zhibin Dong, Hao Tan, Siwei Wang, Xinwang Liu, En Zhu, Di Hu, Tianrui Liu, Chuankun Li, Kunlun He*
320 Multi-Hierarchical Contrastive Spectral Fusion for Multi-View Clustering, *Bing Cai, Xiaoli Wang, Gui-Fu Lu, Zechao Li*
321 SECOS: Semantic Capture for Rigorous Classification in Open-World Semi-Supervised Learning, *Hezhao Liu, Jiacheng Yang, Junlong Gao, Mengke Li, Yiqun Zhang, Shreyank N Gowda, Yang Lu*
322 Multi-Modal Representation Learning via Semi-Supervised Rate Reduction for Generalized Category Discovery, *Wei He, Xianghan Meng, Zhiyuan Huang, Xianbiao Qi, Rong Xiao, Chun-Guang Li*
323 TimeBridge: Self-Supervised Video Representation Learning via Start-End Joint Embedding and In-Between Frame Prediction, *Qin Wang, Abigail Morrison, Hanno Scharr, Kai Krajsek*
324 ❊ Mitigating Instance Entanglement in Instance-Dependent Partial Label Learning, *Rui Zhao, Bin Shi, Kai Sun, Bo Dong*
325 Residual Connections Harm Generative Representation Learning, *Xiao Zhang, Ruoxi Jiang, William Gao, Rebecca Willet, Michael Maire*
326 Neural Mixture Density Processes, *Yi Ding, Qi Tao, Xingxing Liang, Longfei Zhang, Yiqin Lv, Weitao Song, Fangjie Yang, Cheems Wang, Guangquan Cheng*
327 Large-scale Robust Enhanced Ensemble Clustering via Outlier Decoupling, *Jiaxuan Xu, Lei Duan, Xinye Wang, Liang Du*
328 DriveLaW: Unifying Planning and Video Generation in a Latent Driving World, *Tianze Xia, Yongkang Li, Lijun Zhou, Jingfeng Yao, Kaixin Xiong, Haiyang Sun, Bing Wang, Kun Ma, Guang Chen, Hangjun Ye, Wenyu Liu, Xinggang Wang*
329 DLWM: Dual Latent World Models enable Holistic Gaussian-centric Pre-training in Autonomous Driving, *Yiyao Zhu, Ying Xue, Haiming Zhang, Guangfeng Jiang, Wending Zhou, Xu Yan, Jiantao Gao, Yingjie Cai, Bingbing Liu, Zhen Li, Shaojie Shen*
330 Latent Chain-of-Thought World Modeling for End-to-End Driving, *Shuhan Tan, Kashyap Chitta, Yuxiao Chen, Ran Tian, Yurong You, Yan Wang, Wenjie Luo, Yulong Cao, Philipp Krähenbühl, Marco Pavone, Boris Ivanovic*
331 ❊ RLFTSim: Realistic and Controllable Multi-Agent Traffic Simulation via Reinforcement Learning Fine-Tuning, *Ehsan Ahmadi, Hunter Schofield, Behzad Khamidehi, Fazel Arasteh, Jinjun Shan, Lili Mou, Dongfeng Bai, Kasra Rezaee*
332 TrafficAlign: Aligning Large Language Models for Traffic Scenario Generation, *Zhi Tu, Liangkun Niu, Tianyi Zhang*
333 Failure Modes for Deep Learning–Based Online Mapping: How to Measure and Address Them, *Michael Hubbertz, Qi Han, Tobias Meisen*
334 Linking Modality Isolation in Heterogeneous Collaborative Perception, *Changxing Liu, Zichen Chao, Siheng Chen*
335 LEAD: Minimizing Learner-Expert Asymmetry in End-to-End Driving, *Long Nguyen, Micha Fauth, Bernhard Jaeger, Daniel Dauner, Maximilian Igl, Andreas Geiger, Kashyap Chitta*
336 DriverGaze360: OmniDirectional Driver Attention with Object-Level Guidance, *Shreedhar Govil, Didier Stricker, Jason Rambach*
337 Diffusion Forcing Planner: History-Annealed Planning with Time-Dependent Guidance for Autonomous Driving, *Zehan Zhang, Yaoyi Li, Neng Zhang, Jia Cai*
338 DIMOS: Disentangling Instance-level Moving Object Segmentation, *Hongxiang Huang, Hongwei Ren, Xiaopeng Lin, Yulong Huang, Zeke Xie, Bojun Cheng*
339 EvObj: Learning Evolving Object-centric Representations for 3D Instance Segmentation without Scene Supervision, *Jiahao Chen, Zihui Zhang, Yafei Yang, Jinxi Li, Shenxing Wei, Zhixuan Sun, Bo Yang*
340 Live Interactive Training for Video Segmentation, *Xinyu Yang, Haozheng Yu, Yihong Sun, Bharath Hariharan, Jennifer J. Sun*
341 Robust Promptable Video Object Segmentation, *Sohyun Lee, Yeho Gwon, Lukas Hoyer, Konrad Schindler, Christos Sakaridis, Suha Kwak*
342 Scene-VLM: Multimodal Video Scene Segmentation via Vision-Language Models, *Nimrod Berman, Adam Botach, Emanuel Ben-Baruch, Shunit Haviv Hakimi, Asaf Gendler, Ilan Naiman, Erez Yosef, Igor Kviatkovsky*
343 Concept-Aware LoRA for Domain-Aligned Segmentation Dataset Generation, *Minho Park, Sunghyun Park, Jungsoo Lee, Hyojin Park, Kyuwoong Hwang, Fatih Porikli, Jaegul Choo, Sungha Choi*
344 BEV-CAR: Enhancing Monocular Bird's Eye View Segmentation with Context-Aware Rasterization, *Yixin Xiong, Ke Wang, Tongtong Cheng, Chunhui Liu, Kai Liu*
345 Exploring the Underwater World Segmentation without Extra Training, *Bingyu Li, Tao Huo, Da Zhang, Zhiyuan Zhao, Junyu Gao, Xuelong Li*
346 ❊ Learning from Oblivion: Predicting Knowledge-Overflowed Weights via Retrodiction of Forgetting, *Jinhyeok Jang, Jaehong Kim, Jung Uk Kim*
347 Cross-Architecture Adaptation: Cloud-Edge Continual Test-Time Adaptation with Dynamic Sampling and Heterogeneous Distillation, *Zirui Xu, Xianhang Chu, Jiahao Li, Xu Yang, Cheng Deng*
348 Towards Dynamic Modality Alignment in Multimodal Continual Learning, *Jiayao Tan, Fan Lyu, Tianle Liu, Fuyuan Hu, Wei Feng*
349 φ-DPO: Fairness Direct Preference Optimization Approach to Continual Learning in Large Multimodal Models, *Thanh-Dat Truong, Huu-Thien Tran, Jackson Cothren, Bhiksha Raj, Khoa Luu*
350 Incremental Object Detection via Future-Aware Decoupled Cross-Head Distillation, *Chenfeng Yin, De Cheng, Wenlong Luo, Mingyue Zeng, Shizhou Zhang, Nannan Wang, Xinbo Gao*
351 Smart Replay: Adaptive Scheduling of Memory Rehearsal for Computational Resource-Aware Incremental Learning, *Jianting Chen, Dianzhi Yu, Irwin King*
352 ReBaPL: Repulsive Bayesian Prompt Learning, *Yassir Bendou, Omar Ezzahir, Eduardo Montesuma, Gabriel Mahuas, Victoria Shevchenko, Mike Gartrell*
353 Spectral Mixture-of-Experts for Continual Learning, *Chen Yin, Xingbo Dong, Xuelin Shen, Zhe Jin*
354 ActAvatar: Temporally-Aware Precise Action Control for Talking Avatars, *Ziqiao Peng, Yi Chen, Yifeng Ma, Guozhen Zhang, Zhiyao Sun,*

Zixiang Zhou, Youliang Zhang, Zhengguang Zhou, Zhaoxin Fan, Hongyan Liu, Yuan Zhou, Qinglin Lu, Jun He

355 ViBES: A Conversational Agent with Behaviorally-Intelligent 3D Virtual Body, *Juze Zhang, Changan Chen, Xin Chen, Heng Yu, Tiange Xiang, Ali Sartaz Khan, Shrinidhi K. Lakshmikanth, Ehsan Adeli*

356 ❊ DeX-Portrait: Disentangled and Expressive Portrait Animation via Explicit and Latent Motion Representations, *Yuxiang Shi, Zhe Li, Yanwen Wang, Hao Zhu, Xun Cao, Ligang Liu*

357 ❊ SketchFaceGS: Real-Time Sketch-Driven Face Editing and Generation with Gaussian Splatting, *Bo Li, Jiahao Kang, Yubo Ma, Feng-Lin Liu, Bin Liu, Fang-Lue Zhang, Lin Gao*

358 MIBURI: Towards Expressive Interactive Gesture Synthesis, *M. Hamza Mughal, Rishabh Dabral, Vera Demberg, Christian Theobalt*

359 Personalized Image Descriptions from Attention Sequences, *Ruoyu Xue, Hieu Le, Jingyi Xu, Sounak Mondal, Abe Leite, Gregory Zelinsky, Minh Hoai, Dimitris Samaras*

360 GA-VLN: Geometry-Aware BEV Representation for Efficient Vision-Language Navigation, *Jiahao Yang, Zihan Wang, Xiangyang Li, Xing Zhu, Yujun Shen, Yinghao Xu, Shuqiang Jiang*

361 IMAIA: Interactive Maps AI Assistant for Travel Planning and Geo-Spatial Intelligence, *Jieren Deng, Zhizhang Hu, Ziyan He, Aleksandar Cvetkovic, Pak Kiu Chung, Dragomir Yankov, Chiqun Zhang*

362 OctoNav: Towards Generalist Embodied Navigation, *Chen Gao, Liankai Jin, Xingyu Peng, Jiazhao Zhang, Yue Deng, Annan Li, He Wang, Si Liu*

363 WalkGPT: Grounded Vision–Language Conversation with Depth-Aware Segmentation for Pedestrian Navigation, *Rafi Ibn Sultan, Hui Zhu, Xiangyu Zhou, Chengyin Li, Prashant Khanduri, Marco Brocanelli, Dongxiao Zhu*

364 SpaceDrive: Infusing Spatial Awareness into VLM-based Autonomous Driving, *Peizheng Li, Zhenghao Zhang, David Holtz, Hang Yu, Yutong Yang, Yuzhi Lai, Rui Song, Andreas Geiger, Andreas Zell*

365 SMAP: Semantic Route Planning with Map-Grounded Multimodal Alignment, *Wenjie Zhang, Chen Yang, Xin Lu, Zhen Wang, Yue Liu, Bobo Xi, Pengbo Zhang*

366 IDperturb: Enhancing Variation in Synthetic Face Generation via Angular Perturbations, *Fadi Boutros, Eduarda Caldeira, Tahar Chettaoui, Naser Damer*

367 ❊ Fresco: Frequency–Spatial Consistent Optimization for Fine-Grained Head Avatar Modeling, *Shikun Zhang, Yong Li, Yiqun Wang, Qiuhong Ke, Cunjian Chen*

368 Motion-Aware Animatable Gaussian Avatars Deblurring, *Muyao Niu, Yifan Zhan, Qingtian Zhu, Zhuoxiao Li, Wei Wang, Zhihang Zhong, Xiao Sun, Yinqiang Zheng*

369 ELITE: Efficient Gaussian Head Avatar from a Monocular Video via Learned Initialization and Test-time Generative Adaptation, *Kim Youwang, Lee Hyoseok, Park Subin, Gerard Pons-Moll, Tae-Hyun Oh*

370 Multi-view Consistent 3D Gaussian Head Avatars 'without' Multi-view Generation, *Aviral Chharia, Fernando De la Torre*

371 MAD: Modality-Adaptive Decoding for Mitigating Cross-Modal Hallucinations in Multimodal Large Language Models, *Sangyun Chung, Se Yeon Kim, Youngchae Chee, Yong Man Ro*

372 Cross-Modal Attention Calibration for LVLM Hallucination Mitigation, *Jiaming Li, Jiacheng Zhang, Zequn Jie, Lin Ma, Ming Li, Xiaonan Luo, Guanbin Li*

373 3D-VCD: Hallucination Mitigation in 3D-LLM Embodied Agents through Visual Contrastive Decoding, *Makanjuola Adekunmi Ogunleye, Eman Abdelrahman, Ismini Lourentzou*

374 Exposing and Evaluating Hallucinations for GUI Grounding, *Zicheng Zhang, Hongyi Jing, Rui Lv, Shuo Fang, Shiai Zhu, Junying Wang, Chunyi Li, Xiaohong Liu, Chenguang Ma, Guangtao Zhai*

375 Understanding and Mitigating Hallucinations in Multimodal Chain-of-Thought Models, *Ji Ma, Wei Suo, Peng Wang, Yanning Zhang*

376 Beyond the Global Scores: Fine-Grained Token Grounding as a Robust Detector of LVLM Hallucinations, *Tuan Dung Nguyen, Minh Khoi Ho, Qi Chen, Yutong Xie, Cam-Tu Nguyen, Minh Khoi Nguyen, Dang Huy Pham Nguyen, Anton van den Hengel, Johan Verjans, Phi Le Nguyen, Vu Minh Hieu Phan*

377 StereoWorld: Geometry-Aware Monocular-to-Stereo Video Generation, *Ke Xing, Longfei Li, Yuyang Yin, Hanwen Liang, Guixun Luo, Chen Fang, Jue Wang, Konstantinos N. Plataniotis, Xiaojie Jin, Yao Zhao, Yunchao Wei*

378 Infinity-RoPE: Action-Controllable Infinite Video Generation Emerges From Autoregressive Self-Rollout, *Hidir Yesiltepe, Tuna Meral, Adil Kaan Akan, Kaan Oktay, Pinar Yanardag*

379 AniMimic: Imitating 3D Animation from Video Priors, *Tianyi Xie, Yunuo Chen, Yaowei Guo, Yin Yang, Bolei Zhou, Demetri Terzopoulos, Ying Jiang, Chenfanfu Jiang*

380 VerseCrafter: Dynamic Realistic Video World Model with 4D Geometric Control, *Sixiao Zheng, Minghao Yin, Wenbo Hu, Xiaoyu Li, Ying Shan, Yanwei Fu*

381 ScenDi: 3D-to-2D Scene Diffusion Cascades for Urban Generation, *Hanlei Guo, Jiahao Shao, Xinya Chen, Xiyang Tan, Sheng Miao, Yujun Shen, Yiyi Liao*

382 ❊ MotionCrafter: Dense Geometry and Motion Reconstruction with a 4D VAE, *Ruijie Zhu, Jiahao Lu, Wenbo Hu, Xiaoguang Han, Jianfei Cai, Ying Shan, Chuanxia Zheng*

383 GeodesicNVS: Probability Density Geodesic Flow Matching for Novel View Synthesis, *Xuqin Wang, Tao Wu, Yanfeng Zhang, Lu Liu, Mingwei Sun, Yongliang Wang, Niclas Zeller, Daniel Cremers*

384 WorldStereo: Bridging Controllable Video Generation and Scene Reconstruction via 3D Geometric Memories, *Yisu Zhang, Chenjie Cao, Tengfei Wang, Xuhui Zuo, Junta Wu, Jianke Zhu, Chunchao Guo*

385 ❊ NeoVerse: Enhancing 4D World Model with in-the-wild Monocular Videos, *Yuxue Yang, Lue Fan, Ziqi Shi, Junran Peng, Feng Wang, Zhaoxiang Zhang*

386 ❊ Taming Video Models for 3D and 4D Generation via Zero-Shot Camera Control, *Chenxi Song, Yanming Yang, Tong Zhao, Ruibo Li, Chi Zhang*

387 ❊ Improving Motion in Image-to-Video Models via Adaptive Low-Pass Guidance, *June Suk Choi, Kyungmin Lee, Sihyun Yu, Yisol Choi, Jinwoo Shin, Kimin Lee*

388 SANER: Switchable Adapter with Non-parametric Enhanced Routing for Person De-Reidentification, *Yimin Liu, Nan Pu, Fengxiang Yang, Wenjing Li, Zhihui Li, Zhun Zhong*

389 BIT: Matching-based Bi-directional Interaction Transformation Network for Visible-Infrared Person Re-Identification, *Haoxuan Xu, Guanglin Niu*

390 Vision-Language Attribute Disentanglement and Reinforcement for Lifelong Person Re-Identification, *Kunlun Xu, Haotong Cheng, Jiangmeng Li, Xu Zou, Jiahuan Zhou*

391 Diversity over Uniformity: Rethinking Representation in Generated Image Detection, *Qinghui He, Haifeng Zhang, Qiao Qin, Bo Liu, Xiuli Bi, Bin Xiao*

392 Mining Instance-Centric Vision–Language Contexts for Human–Object Interaction Detection, *Soo Won Seo, KyungChae Lee, Hyungchan Cho, Taein Son, Nam Ik Cho, Jun Won Choi*

393 FSLoRA: Harmonizing Detection and Re-Identification via Freq-Spatial Low-Rank Adapter for One-Stage Person Search, *Yanling Tian, Shanshan Zhang, Di Chen, Jian Yang*

394 EEGiT: Teaching Vision Transformers to Understand the EEG signal, *Jiahao Zhou, Chenghao Xu, Wei Wang, Erkun Yang, Cheng Deng*

395 FedBPrompt: Federated Domain Generalization Person Re-Identification via Body Distribution Aware Visual Prompts, *Xin Xu, Weilong Li, Wei Liu, Wenke Huang, Zhixi Yu, Bin Yang, Xiaoying Liao, Kui Jiang*

396 Pose-guided Enriched Feature Learning for Federated-by-camera Person Re-identification, *JooHyung Oh, Minyoung Oh, Sung Whan Yoon, Jae-Young Sim*

397 UAV-CB: A Complex-Background RGB–T Dataset and Local Frequency Bridge Network for UAV Detection, *Shenghui Huang, Menghao Hu, Longkun Zou, Hongyu Chi, Zekai Li, Feng Gao, Fan Yang, Qingyao Wu, Ke Chen*

398 TimeViper: A Hybrid Mamba-Transformer Vision-Language Model for Efficient Long Video Understanding, *Boshen Xu, Zihan Xiao, Jiaze Li, Jianzhong Ju, Zhenbo Luo, Jian Luan, Qin Jin*

399 StreamReady: Learning What to Answer and When in Long Streaming Videos, *Shehreen Azad, Vibhav Vineet, Yogesh S Rawat*

400 LongVideo-R1: Smart Navigation for Low-cost Long Video Understanding, *Jihao Qiu, Lingxi Xie, Xinyue Huo, Qi Tian, Qixiang Ye*

401 Agentic Video Summarization via Self-Reflecting Multimodal Understanding, *Miaotian Guo, Shuguang Dou, Yin Li, Aidong Men, Dongsheng Jiang*
402 Self-Critical Distillation Network for Video-based Commonsense Captioning, *Mengqi Yuan, Gengyun Jia, Bing-Kun Bao*
403 Ego-Grounding for Personalized Question-Answering in Egocentric Videos, *Junbin Xiao, Shenglang Zhang, Pengxiang Zhu, Angela Yao*
404 AdaSpark: Adaptive Sparsity for Efficient Long-Video Understanding,
❊ *Handong Li, Zikang Liu, Longteng Guo, Tongtian Yue, Yepeng Tang, Xinxin Zhu, Chuanyang Zheng, Ziming Wang, Zhibin Wang, Jun Song, Cheng Yu, Bo Zheng, Jing Liu*
405 EarlyTom: Early Token Compression Completes Fast Video Understanding, *Hesong Wang, Xin Jin, Lu Lu, Chenhaowen Li, Jian Chen, Qiang Liu, Huan Wang*
406 VideoWorld 2: Learning Transferable Knowledge from Real-world Videos, *Zhongwei Ren, Yunchao Wei, Xiao Yu, Guixun Luo, Yao Zhao, Bingyi Kang, Jiashi Feng, Xiaojie Jin*
407 VirtueBench: Evaluating Trustworthiness under Uncertainty in Long Video Understanding, *Xueqing Yu, Bohan Li, Yan Li, Zhenheng Yang*
408 DiverseDiT: Towards Diverse Representation Learning in Diffusion Transformers, *Mengping Yang, Zhiyu Tan, Binglei Li, Xiaomeng Yang, Hesen Chen, Hao Li*
409 RenderFlow: Single-Step Neural Rendering via Flow Matching, *Shenghao Zhang, Runtao Liu, Christopher Schroers, Yang Zhang*
410 ResDiT: Evoking the Intrinsic Resolution Scalability in Diffusion Transformers, *Yiyang Ma, Feng Zhou, Xuedan Yin, Pu Cao, Yonghao Dang, Jianqin Yin*
411 Masked Region Transformer for Layered Image Generation and Editing at Scale, *Zhicong Tang, Jingye Chen, Zhao Zhang, Mohan Zhou, Yuchi Liu, Yifan Pu, Yalong Bai, Ethan Smith, Yuhui Yuan*
412 DDT: Decoupled Diffusion Transformer, *Shuai Wang, Zhi Tian, Weilin Huang, Limin Wang*
413 Just-in-Time: Training-Free Spatial Acceleration for Diffusion Transformers, *Wenhao Sun, Ji Li, Zhaoqiang Liu*
414 Preserving Source Video Realism: High-Fidelity Face Swapping for Cinematic Quality, *Zekai Luo, Zongze Du, Zhouhang Zhu, Hao Zhong, Muzhi Zhu, Wen Wang, Yuling Xi, Chenchen Jing, Hao Chen, Chunhua Shen*
415 ShapeAR: Generating Editable Shape Layers via Autoregressive Diffusion, *Souymodip Chakraborty, Ankur Singh, Amit Vikram Singh, Vineet Batra, Ankit Phogat*
416 ReHyAt: Recurrent Hybrid Attention for Video Diffusion Transformers, *Mohsen Ghafoorian, Amirhossein Habibian*
417 RecTok: Reconstruction Distillation along Rectified Flow, *Qingyu Shi, Size Wu, Jinbin Bai, Kaidong Yu, Yujing Wang, Yunhai Tong, Xiangtai Li, Xuelong Li*
418 EgoXtreme: A Dataset for Robust Object Pose Estimation in
❊ Egocentric Views under Extreme Conditions, *Taegyoon Yoon, Yegyu Han, Seojin Ji, Jaewoo Park, Sojeong Kim, Taein Kwon, Hyung-Sin Kim*
419 CoIn3D: Revisiting Configuration-Invariant Multi-Camera 3D Object Detection, *Zhaonian Kuang, Rui Ding, Haotian Wang, Xinhu Zheng, Meng Yang, Gang Hua*
420 H^2A^2: Homogeneity-Aware and Heterogeneity-Aware Feature Perception for Unified Indoor 3D Object Detection, *Tao Xie, Tao An, Feng Liu, Jin Wensheng, Zhengyu Li, Lijun Zhao, Ruifeng Li*
421 Cov2Pose: Leveraging Spatial Covariance for Direct Manifold-aware
❊ 6-DoF Object Pose Estimation, *Nassim Ali Ousalah, Peyman Rostami, Vincent Gaudillière, Emmanuel Koumandakis, Anis Kacem, Enjie Ghorbel, Djamila Aouada*
422 Towards Intrinsic-Aware Monocular 3D Object Detection, *Zhihao Zhang, Abhinav Kumar, Xiaoming Liu*
423 SToRe3D: Sparse Token Relevance in ViTs for Efficient Multi-View 3D Object Detection, *Sandro Papais, Lezhou Feng, Charles Cossette, Lingting Ge*
424 SPAN: Spatial-Projection Alignment for Monocular 3D Object Detection, *Yifan Wang, Yian Zhao, Fanqi Pu, Xiaochen Yang, Yang Tang, Xi Chen, Wenming Yang*
425 DSCA: Dynamic Subspace Concept Alignment for Lifelong VLM Editing, *Gyanendra Das, Sai Jena*
426 FailureAtlas: Mapping the Failure Landscape of T2I Models via Active Exploration, *Muxi Chen, Zhaohua Zhang, Chenchen Zhao, Mingyang Chen, Wenyu Jiang, Tianwen Jiang, Jianhuan Zhuo, Yu Tang, Qiuyong Xiao, Jihong Zhang, Qiang Xu*
427 HDR-VLM: HDR-Domain Adaptation of VLMs and Preference-Aligned
❊ Quality Assessment for HDR Video Color Grading, *Hao Yuan, Jiabin Zhang, Yajing Wu, Ruixuan Pang, Jing Li*
428 RobustVisRAG: Causality-Aware Vision-Based Retrieval-Augmented Generation under Visual Degradations, *I-Hsiang Chen, Yu-Wei Liu, Tse-Yu Wu, Yu-Chien Chiang, Jen-Chieh Yang, Wei-Ting Chen*
429 BiomedCCPL: Causal Conditional Prompt Learning for Biomedical Vision-Language Models, *Xueliang Cui, Juncai Zhang, Jiacheng Hou, Dan Lu, Hao Zhang, Ruxin Wang*
430 DynamicGTR: Leveraging Graph Topology Representation Preferences to Boost VLM Capabilities on Graph QAs, *Yanbin Wei, Jiangyue Yan, Chun Kang, Yang Chen, Hua Liu, James Kwok, Yu Zhang*
431 VisualOverload: Probing Visual Understanding of VLMs in Really Dense Scenes, *Paul Gavrikov, Wei Lin, M. Jehanzeb Mirza, Soumya Jahagirdar, Muhammad Huzaifa, Sivan Doveh, James Glass, Serena Yeung-Levy, Hilde Kuehne*
432 Revisiting Visual Corruptions in LVLMs: A Shape–Texture Perspective on Model Failures, *Xinkuan Qiu, Meina Kan, Zhenliang He, Yongbin Zhou, Shiguang Shan*
433 From Intuition to Investigation: A Tool-Augmented Reasoning MLLM Framework for Generalizable Face Anti-Spoofing, *Haoyuan Zhang, Keyao Wang, Guosheng Zhang, Haixiao Yue, Zhiwen Tan, Siran Peng, Tianshuo Zhang, Xiao Tan, Kunbin Chen, Wei He, Jingdong Wang, Ajian Liu, Xiangyu Zhu, Zhen Lei*
434 Trust-calibrated Collaborative Learning for Long-Tailed Visual Recognition, *Hao Zhou, Tingjin Luo*
435 SunFaded: Illumination-Aware Gaussian Splatting for Dark Scenes with Camera-Mounted Active Lighting, *Wenjie Chang, Tianle Ding, Wenfei Yang, Tianzhu Zhang*
436 TokenSplat: Token-aligned 3D Gaussian Splatting for Feed-forward
❊ Pose-free Reconstruction, *Yihui Li, Chengxin Lv, Zichen Tang, Hongyu Yang, Di Huang*
437 GOR-IS: 3D Gaussian Object Removal In the Intrinsic Space,
❊ *Yonghao Zhao, Yupeng Gao, Jian Yang, Jin Xie, Beibei Wang*
438 AeroGS: Scale-Aware Gaussian Splatting for Pose-Free Dynamic UAV
❊ Scene Reconstruction, *Tingyun Li, Xinyi Liu, Yongjun Zhang, Yi Wan, Xiaoan Liu, Weiwei Fan, Jiahao Liu*
439 Intrinsic Geometry-Appearance Consistency Optimization for Sparse-View Gaussian Splatting, *Kaiqiang Xiong, Rui Peng, Jiahao Wu, Zhanke Wang, Jie Liang, Xiaoyun Zheng, Feng Gao, Ronggang Wang*
440 AERGS-SLAM: Auto-Exposure-Robust Stereo 3D Gaussian Splatting SLAM, *Zhiyu Zhou, Feng Hui, Yu Liu*
441 Learning Differentiable Hierarchies in 3D Gaussian Splatting,
❊ *Youqi Pan, Wugen Zhou, Hongbin Zha*
442 WeatherCity: Urban Scene Reconstruction with Controllable Multi-Weather Transformation, *Wenhua Wu, Huai Guan, Zhe Liu, Hesheng Wang*
443 Cross-View Splatter: Feed-Forward View Synthesis with Georeferenced Images, *Matias Turkulainen, Akshay Krishnan, Filippo Aleotti, Mohamed Sayed, Guillermo Garcia-Hernando, Juho Kannala, Arno Solin, Gabriel Brostow, Daniyar Turmukhambetov*
444 TagSplat: Topology-Aware Gaussian Splatting for Dynamic Mesh
❊ Modeling and Tracking, *Hanzhi Guo, Dongdong Weng, Mo Su, Yixiao Chen, Xiaonuo Dongye, Chenyu Xu*
445 Hierarchical Visual Relocalization with Nearest View Synthesis from Feature Gaussian Splatting, *Huaqi Tao, Bingxi Liu, Guangcheng Chen, Fulin Tang, Li He, Hong Zhang*
446 Tracking-Guided 4D Generation: Foundation-Tracker Motion Priors for 3D Model Animation, *Su Sun, Cheng Zhao, Himangi Mittal, Gaurav Mittal, Rohith Kukkala, Yingjie Victor Chen, Mei Chen*
447 3D Gaussian Splatting from Unposed Spike Stream, *Yijia Guo, Tong Hu, Liwen Hu, Lei Ma, Tiejun Huang*
448 SparseOIT: Improving Order-Independent Transparency 3DGS via Active
❊ Set Method, *Wentao Yang, Fanzhen Kong, Zejian Kang, Xiangru Huang*
449 ClipGStream: Clip-Stream Gaussian Splatting for Any Length and Any Motion Multi-View Dynamic Scene Reconstruction, *Jie Liang,*

Jiahao Wu, Chao Wang, Jiayu Yang, Xiaoyun Zheng, Kaiqiang Xiong, Zhanke Wang, Jinbo Yan, Feng Gao, Ronggang Wang

450 Space-Time Forecasting of Dynamic Scenes with Motion-aware Gaussian Grouping, *Junmyeong Lee, Hoseung Choi, Minsu Cho*

451 MoRGS: Efficient Per-Gaussian Motion Reasoning for Streamable Dynamic 3D Scenes, *Wonjoon Lee, Sungmin Woo, Donghyeong Kim, Jungho Lee, Sangheon Park, Sangyoun Lee*

452 ❊ BEA-GS: BEyond RAdiance Supervision in 3DGS for Precise Object Extraction, *Alessio Mazzucchelli, Maria Naranjo-Almeida, Jorge Bustos-Sanchez, Mariella Dimiccoli, Francesc Moreno-Noguer, Jordi Sanchez-Riera, Adrian Penate-Sanchez*

453 EDGS: Eliminating Densification for Efficient Convergence of 3DGS, *Dmytro Kotovenko, Olga Grebenkova, Björn Ommer*

454 ReasonMap: Towards Fine-Grained Visual Reasoning from Transit Maps, *Sicheng Feng, Song Wang, Shuyi Ouyang, Lingdong Kong, Zikai Song, Jianke Zhu, Huan Wang, Xinchao Wang*

455 Conan: Progressive Learning to Reason Like a Detective over Multi-Scale Visual Evidence, *Kun Ouyang, Yuanxin Liu, Linli Yao, Yishuo Cai, Hao Zhou, Fandong Meng, Jie Zhou, Xu Sun*

456 DialogueVPR: Towards Conversational Visual Place Recognition, *Yukun Song, Changwei Wang, Xingtian Pei, Shibiao Xu, Wenhao Xu, Shunpeng Chen, Yu Zhang, Ke Zhang, Rongtao Xu, Xuxiang Feng, Pengyang Wang*

457 Perceptual-Evidence Anchored Reinforced Learning for Multimodal Reasoning, *Chi Zhang, Haibo Qiu, Qiming Zhang, Yufei Xu, Zhixiong Zeng, Siqi Yang, Peng Shi, Lin Ma, Jing Zhang*

458 Thinking with Video: Video Generation as a Promising Multimodal Reasoning Paradigm, *Jingqi Tong, Yurong Mou, Hangcheng Li, Mingzhe Li, Yongzhuo Yang, Ming Zhang, Qiguang Chen, Tianyi Liang, Xiaomeng Hu, Yining Zheng, Xinchi Chen, Jun Zhao, Xuanjing Huang, Xipeng Qiu*

459 VinQA: Visual Elements Interleaved Long-form Answer Generation for Real-World Multimodal Document QA, *Young Rok Jang, Hyesoo Kong, Kyunghwan An, Jae Sub Huh, Gyeonghun KIM, Stanley Jungkyu Choi*

460 ❊ DocSeeker: Structured Visual Reasoning with Evidence Grounding for Long Document Understanding, *Hao Yan, Yuliang Liu, Xingchen Liu, Yuyi Zhang, Minghui Liao, Jihao Wu, Wei Chen, Xiang Bai*

461 Recurrent Reasoning with Vision-Language Models for Estimating Long-Horizon Embodied Task Progress, *Yuelin Zhang, Sijie Cheng, Chen Li, Zongzhao Li, Yuxin Huang, Yang Liu, Wenbing Huang*

462 VGent: Visual Grounding via Modular Design for Disentangling Reasoning and Prediction, *Weitai Kang, Jason Kuen, Mengwei Ren, Zijun Wei, Yan Yan, Kangning Liu*

463 Grounding Everything in Tokens for Multimodal Large Language Models, *Xiangxuan Ren, Zhongdao Wang, Liping Hou, Pin Tang, Guoqing Wang, Chao Ma*

464 Evolving Contextual Safety in Multi-Modal Large Language Models via Inference-Time Self-Reflective Memory, *Ce Zhang, Jinxi He, Junyi He, Katia Sycara, Yaqi Xie*

465 ChartR: Evaluating Reasoning Accuracy and Robustness in Chart Question Answering, *Xiaojun Chen, Sixiao Luo, Ziqi Liu, Min Yang, Qin Zhang, Liang-Jie Zhang*

466 Think Visually, Reason Textually: Vision-Language Synergy in Abstract Reasoning, *Beichen Zhang, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, Jiaqi Wang*

467 VKG-QA: Visual Knowledge Graph-based Question Answer for Large Multimodal Models, *Yuntao Du, Yiming Wang, Renshuo Yuan, Jincheng Yue, Yijing Chen, Yue Fan, Bo Zhang, Qian Li, Lizhen Cui*

468 Med-CMR: A Fine-Grained Benchmark Integrating Visual Evidence and Clinical Logic for Medical Complex Multimodal Reasoning, *Haozhen Gong, Xiaozhong Ji, Yuansen Liu, Wenbin Wu, Xiaoxiao Yan, Jingjing Liu, Kai Wu, Jiazhen Pan, Bailiang Jian, Jiangning Zhang, Xiaobin Hu, Hongwei Bran Li*

469 Human-like Abstract Visual Reasoning via Understanding and Solving Reasoning Loop, *Xinwang Chen, Xiuxing Li, Qing Li, Ziyue Zhuang, Yutong Wu, Ziyu Li, Zhuo Wang, Kai Li, Jianye Hao, Xia Wu*

470 VITAL: Vision-Encoder-centered Pre-training for LMMs in Visual Quality Assessment, *Ziheng Jia, Linhan Cao, Jinliang Han, Zicheng Zhang, Jiaying Qian, Jiarui Wang, Zijian Chen, Guangtao Zhai, Xiongkuo Min*

471 Generative Video Compression with One-Dimensional Latent Representation, *Zihan Zheng, Zhaoyang Jia, Naifu Xue, Jiahao Li, Bin Li, Zongyu Guo, Xiaoyi Zhang, Zhenghao Chen, Houqiang Li, Yan Lu*

472 Markovian Scale Prediction: A New Era of Visual Autoregressive Generation, *Yu Zhang, Jingyi Liu, Yiwei Shi, Qi Zhang, Duoqian Miao, Changwei Wang, Longbing Cao*

473 Learned Image Compression via Sparse Attention and Adaptive Frequency, *Huidong Ma, Xinyan Shi, Hui Sun, Xiaofei Yue, Xiaoguang Liu, Gang Wang, Wentong Cai*

474 UPLiFT: Efficient Pixel-Dense Feature Upsampling with Local Attenders, *Matthew Walmer, Saksham Suri, Anirud Aggarwal, Abhinav Shrivastava*

475 VecAttention: Vector-wise Sparse Attention for Accelerating Long Context Inference, *Anmin Liu, Ruixuan Yang, Huiqiang Jiang, Bin Lin, Minmin Sun, Yong Li, Chen Zhang, Tao Xie*

476 Ultra-Fast Neural Video Compression, *Jiahao Li, Wenxuan Xie, Zhaoyang Jia, Bin Li, Zongyu Guo, Xiaoyi Zhang, Yan Lu*

477 Parallax to Align Them All: An OmniParallax Attention Mechanism for Distributed Multi-View Image Compression, *Haotian Zhang, Feiyue Long, Yixin Yu, Jian Xue, Haocheng Tang, Tongda Xu, Zhenning Shi, Yan Wang, Siwei Ma, Jiaqi Zhang*

478 Scaling Parallel Sequence Models to Vision Foundation Models, *Yitong Jiang, Collin McCarthy, Hongjun Wang, Hanrong Ye, Qi Dou, Tianfan Xue, Jinwei Gu, Jan Kautz, Hongxu Yin, Pavlo Molchanov, Sifei Liu*

479 Revisiting Model Stitching In the Foundation Model Era, *Zheda Mai, Ke Zhang, Fu-En Wang, Zixiao Ken Wang, Albert Y. C. Chen, Lu Xia, Min Sun, Wei-Lun Chao, Cheng-Hao Kuo*

480 ❊ GeoAgent: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics, *Modi Jin, Yiming Zhang, Boyuan Sun, Dingwen Zhang, Ming-Ming Cheng, Qibin Hou*

481 ❊ VLM-Loc: Localization in Point Cloud Maps via Vision-Language Models, *Shuhao Kang, Youqi Liao, Peijie Wang, Wenlong Liao, Qilin Zhang, Benjamin Busam, Xieyuanli Chen, Yun Liu*

482 ❊ HOLO: Homography-Guided Pose Estimator Network for Fine-Grained Visual Localization on SD Maps, *Xuchang Zhong, Xu Cao, Jinke Feng, Hao Fang*

483 ❊ TriLite: Efficient Weakly Supervised Object Localization with Universal Visual Features and Tri-Region Disentanglement, *Arian Sabaghi, Jose Oramas*

484 GeoSURGE: Geo-localization using Semantic Fusion with Hierarchy of Geographic Embeddings, *Angel Daruna, Nicholas Meegan, Han-Pang Chiu, Supun Samarasekera, Rakesh Kumar*

485 Towards Visual Query Localization in the 3D World, *Liang Peng, Bohan Tan, Zhipeng Zhang, Haobo Li, Yifan Jiao, Xingping Dong, Libo Zhang*

486 OVOD-Agent: A Markov–Bandit Framework for Proactive Visual Reasoning and Self-Evolving Detection, *Chujie Wang, Jianyu Lu, Zhiyuan Luo, Xi Chen, Chu He*

487 Pixel2Phys: Distilling Governing Laws from Visual Dynamics, *Ruikun Li, Jun Yao, Yingfan Hua, Shixiang Tang, Biqing Qi, Bin Liu, Wanli Ouyang, Yan Lu*

488 Tutor-Student Reinforcement Learning: A Dynamic Curriculum for Robust Deepfake Detection, *Zhanhe Lei, Zhongyuan Wang, Jikang Cheng, Baojin Huang, Yuhong Yang, Zhen Han, Chao Liang, Dengpan Ye*

489 Seeing as Experts Do: A Knowledge-Augmented Agent for Open-Set Fine-Grained Visual Understanding, *Junhan Chen, Zilu Zhou, Yujun Tong, Dongliang Chang, Yitao Luo, Zhanyu Ma*

490 Dynamic Important Example Mining for Reinforcement Finetuning, *Haoru Tan, Sitong Wu, Yanfeng Chen, Shizhen Zhao, Yang-Tian Sun, Tianjia Liu, Chirui Chang, Shaofeng Zhang, Samm Sun, Xiuzhe Wu, Ruobing Xie, Xiaojuan Qi*

491 ❊ Specificity-aware reinforcement learning for fine-grained open-world classification, *Samuele Angheben, Davide Berasi, Alessandro Conti, Elisa Ricci, Yiming Wang*

492 SAGE: Training Smart Any-Horizon Agents for Long Video Reasoning with Reinforcement Learning, *Jitesh Jain, Allen AI blank, Jialuo Li, Zixian Ma, Jieyu Zhang, Chris Dongjoo Kim, Sangho Lee, Rohun Tripathi, Tanmay Gupta, Christopher Clark, Humphrey Shi*

493 Uncertainty-Aware Modality Fusion for Unaligned RGB-T Salient Object Detection, *Mianzhao Wang, Fan Shi, Xu Cheng, Chen Jia, Shengyong Chen*
494 Fusion in Your Way: Aligning Image Fusion with Heterogeneous Demands via Direct Preference Optimization, *Weijian Su, Songqian Zhang, Yuqi Han, Jian Zhuang, Yongdong Huang, Qiang Zhang*
495 More Than Meets the Eye: A Unified Image Fusion Framework via Semantic-Pixel Entropy Trade-off for Zero-Shot Generalization, *Xiaowen Liu, Jing Li, Hongtao Huo, Haozhe Cao, Renhua Wang, Xu Dong*
496 Beyond Sequential Tools: A Unified VLM Agent System for Photographic Post-Processing via Dynamic Multi-Expert Fusion, *Honglin Xiong, Chenjie Zhu, Jianbiao Ding, Zixuan Ni, Wei Li, Zhenpeng Mi, Qian Wang*
497 Multi-modal Frequency Decomposition Network for Semantic Scene Completion, *Die Zuo, Lubo Wang, Ruonan Liu, Qing Guo, Chong Wang, Dongdong Wu, Wei Feng, Kairui Yang, Di Lin*
498 BiEvLight: Bi-level Learning of Task-Aware Event Refinement for Low-Light Image Enhancement, *Zishu Yao, Xiang-Xiang Su, Shengning Zhou, Guang-Yong Chen, Guodong Fan, Xing Chen*
499 FusionRegister: Every Infrared and Visible Image Fusion Deserves Registration, *Congcong Bian, Haolong Ma, Hui Li, Zhongwei Shen, Xiaoqing Luo, Xiaoning Song, Xiao-jun Wu*
500 OmniFood8K: Single-Image Nutrition Estimation via Hierarchical
❊ Frequency-Aligned Fusion, *Dongjian Yu, Weiqing Min, Qian Jiang, Xing Lin, Xin Jin, Shuqiang Jiang*
501 Enhancing Unregistered Hyperspectral Image Super-Resolution via Unmixing-based Abundance Fusion Learning, *Yingkai Zhang, Tao Zhang, Jing Nie, Ying Fu*
502 LRHDR: Learning Representation-enhanced HDR Video Reconstruction,
❊ *Chenzhuo Liao, Xin Chen, Bingchen Li, Yu Meng, Tao Yue, Xuemei Hu*
503 Cross-Domain Few-Shot Segmentation via Multi-view Progressive Adaptation, *Jiahao Nie, Guanqiao Fu, Wenbin An, Yap-Peng Tan, Alex C. Kot, Shijian Lu*
504 Interpretable Cross-Domain Few-Shot Learning with Rectified Target-Domain Local Alignment, *Yaze Zhao, Yixiong Zou, Yuhua Li, Ruixuan Li*
505 PP-Brep: Few-Shot B-rep Classification with Hybrid Graph Representation, *Jiacheng Hao, Chunying Liu, Hao Guo, Ruohan Wang, Hongping Gan, Yilei Shi*
506 AgentDet: A Shared-Blackboard Multi-Agent Framework for Zero-/Few-Shot Object Detection, *Haolin Li, Yaohua Wang, Ze Yan, Lijie Wen, Biqing Huang*
507 SFR-Net: Steering-Fusion-Refining Network in Multi-label Zero-Shot Sewer Defect Detection, *Zhao-Min Chen, Xinjian Huang, Yisu Ge, Yu Li*
508 Noise-Aware Few-Shot Learning through Bi-directional Multi-View Prompt Alignment, *Lu Niu, Cheng Xue*
509 Learnability-Guided Diffusion for Dataset Distillation, *Jeffrey A. Chan-Santiago, Mubarak Shah*
510 Phased DMD: Few-step Distribution Matching Distillation via Score Matching within Subintervals, *Xiangyu Fan, Zesong Qiu, Zhuguanyu Wu, Fanzhou Wang, Zhiqian Lin, Tianxiang Ren, Dahua Lin, Ruihao Gong, Lei Yang*
511 Progressive Mask Distillation for Self-supervised Video Representation, *Kewei Wu, Chong Liang, Zhao Xie, Dan Guo*
512 HierAmp: Coarse-to-Fine Autoregressive Amplification for Generative Dataset Distillation, *Lin Zhao, Xinru Jiang, Xi Xiao, Qihui Fan, Lei Lu, Yanzhi Wang, Xue Lin, Octavia Camps, Pu Zhao, Jianyang Gu*
513 SpiderCam: Low-Power Snapshot Depth from Differential Defocus,
❊ *Marcos A. Ferreira, Tianao Li, John Mamish, Josiah Hester, Yaman Sangar, Qi Guo, Emma Alexander*
514 Computational Speckle Pattern Interferometry, *Shengxi Wu, Sophia Yang, Dorian Chan, Matthew O'Toole*
515 DetectSCI: Toward Object-Guided ROI Reconstruction for High-Resolution Video Snapshot Compressive Imaging, *Xingjian Jiang, Lishun Wang, Ping Wang, Xin Yuan*
516 Solving a Nonlinear Blind Inverse Problem for Tagged MRI with Physics and Deep Generative Priors, *Zhangxing Bian, Shuwen Wei, Samuel W. Remedios, Junyu Chen, Aaron Carass, Blake Dewey, Jerry L Prince*
517 Nonlinear Color Transfer via Learnable Bezier Flows, *Junhyoung Lee, Seongwoon Jo, JeongHun Park, Yeonji Ryou, Jeongha Yang, Jangho Kim*
518 VT-Intrinsic: Physics-Based Decomposition of Reflectance and Shading using a Single Visible-Thermal Image Pair, *Zeqing Yuan, Mani Ramanagopal, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan*
519 GH-NAF: Grid-Adaptive Hash-Level–Attended Neural Attenuation Fields for Discrepancy-Aware CBCT, *Seong Je Oh, Ju Hwan Lee, Chae Yeon Lim, Donghwan Lee, Myung Jin Chung, Kyungsu Kim*
520 Computer Vision with a Superpixelation Camera, *Sasidharan Mahalingam, Rachel Brown, Atul Ingle*
521 Color-Encoded Illumination for High-Speed Volumetric Scene
❊ Reconstruction, *David Novikov, Eilon Vaknin, Narek Tumanyan, Mark Sheinin*
522 Multi-Scale Gradient-Guided Unrolling Architecture with Adaptive Mamba for Compressive Sensing, *Le Yang, Hongping Gan*
523 Deciphering Genotype-Phenotype Mechanisms from High-Content Profiling via Knowledge-Guided Multi-modal Graph Learning, *Hanjing Lin, Jiahua Rao, Youhan Sun, Jiancong Xie, Yuedong Yang*
524 Bulk RNA-seq Guided Multi-modal Detection of Anomalous Regions in Human Cancer via Spatial Transcriptomics, *Hang Shi, Ruocheng Yang, Wenjie You, Zhilin Huang, Daoqiang Zhang, Wei Shao*
525 Intervention-Aware Multiscale Representation Learning from Imaging Phenomics and Perturbation Transcriptomics, *Jiayuan Chen, Ruoqi Liu, Zishan Gu, Ping Zhang*
526 ParaUni: Enhance Generation in Unified Multimodal Model with Reinforcement-driven Hierarchical Parallel Information Interaction, *Jiangtong Tan, Lin Liu, Jie Huang, Xiaopeng Zhang, Qi Tian, Feng Zhao*
527 PhysVid: Physics Aware Local Conditioning for Generative Video Models, *Saurabh Pathak, Elahe Arani, Mykola Pechenizkiy, Bahram Zonooz*
528 PromptLoop: Plug-and-Play Prompt Refinement via Latent Feedback for Diffusion Model Alignment, *Suhyeon Lee, Jong Chul Ye*
529 EvoID: Reinforced Evolution for Identity-Preserving Video Generation, *Yiheng Zhang, Zhaofan Qiu, Zunxu Liu, Yingwei Pan, Ting Yao, Tao Mei*
530 Masked Auto-Regressive Variational Acceleration: Fast Inference Makes Practical Reinforcement Learning, *Yuxuan Gu, Weimin Bai, Yifei Wang, Weijian Luo, He Sun*
531 PhyCo: Learning Controllable Physical Priors for Generative Motion, *Sriram Narayanan, Ziyu Jiang, Srinivasa Narasimhan, Manmohan Chandraker*
532 Unified Multimodal Models as Auto-Encoders, *Zhiyuan Yan, Kaiqing Lin, Zongjian Li, Junyan Ye, Hui Han, Haochen Wang, Zhendong Wang, Bin Lin, Hao Li, Xinyan Xiao, Jingdong Wang, Haifeng Wang, Li Yuan*
533 Expand and Prune: Maximizing Trajectory Diversity for Effective GRPO in Generative Models, *Shiran Ge, Chenyi Huang, Yuang Ai, Qihang Fan, Huaibo Huang, Ran He*
534 ThinkingViT: Matryoshka Thinking Vision Transformer for Elastic Inference, *Ali Hojjat, Janek Haberer, Sören Pirk, Olaf Landsiedel*
535 Drainage: A Unifying Framework for Addressing Class Uncertainty,
❊ *Yasser Taha, Grégoire Montavon, Nils Körber*
536 Neural Differentiation in Deep Networks: A Theoretical Framework for Expressivity and Representational Diversity, *Boyuan Wang, Richard Jiang*
537 DuetMerging: Synergizing Dynamic and Static Strategies for Mitigating Task Interference in Model Merging, *Yan Li, Guiping Cao, Yaguang Song, Ming Tao, Haoran Gong, Junhui Liu, Yaowei Wang, Dongmei Jiang*
538 SASNet: Spatially-Adaptive Sinusoidal Networks for INRs, *Haoan Feng, Diana Aldana, Tiago Novello, Leila De Floriani*
539 Generative Modeling of Weights: Generalization or Memorization?,
❊ *Boya Zeng, Yida Yin, Zhiqiu Xu, Zhuang Liu*
540 Vision-Oriented Lightweight Neural Architecture Search with Budget-Adaptive Evaluation, *Yi Fan, Yu-Bin Yang*
541 Improving Sparse Autoencoder with Dynamic Attention, *Dongsheng Wang, Jinsen Zhang, Dawei Su, Hui Huang*
542 Stepwise Credit Assignment for GRPO on Flow-Matching Models, *Yash Savani, Branislav Kveton, Yuchen Liu, Yilin Wang,*

Jing Shi, Subhojyoti Mukherjee, Nikos Vlassis, Krishna Kumar Singh

543 FINE: Factorizing Knowledge for Initialization of Variable-sized Diffusion Models, *Yucheng Xie, Fu Feng, Ruixiao Shi, Jianlu Shen, Jing Wang, Yong Rui, Xin Geng*

544 Hyperbolic Busemann Neural Networks, *Ziheng Chen, Bernhard Schölkopf, Nicu Sebe*

545 FlowDIS: Language-Guided Dichotomous Image Segmentation with Flow Matching, *Andranik Sargsyan, Shant Navasardyan*

546 Image-to-Point Cloud Feature Back-Projection for Multimodal Training of 3D Semantic Segmentation, *Jiawei Han, Matteo Poggi, Li Huan, Changshuo Wang, Kaiqi Liu, Wei Li*

547 ❊ NG-GS: NeRF-guided 3D Gaussian Splatting Segmentation, *Yi He, Tao Wang, Yi Jin, Congyan Lang, Yidong Li, Haibin Ling*

548 Teaching DINOv3 About Partial 3D Geometry: A Self-Supervised Geometry-Aware Approach, *Viktoria Ehm, Dongliang Cao, Riccardo Marin, Daniel Scholz, Weikang Wang, Florian Bernard, Daniel Cremers*

549 SemLayer: Semantic-aware Generative Segmentation and Layer Construction for Abstract Icons, *Haiyang Xu, Ronghuan Wu, Li-Yi Wei, Nanxuan Zhao, Chenxi Liu, Cuong Nguyen, Zhuowen Tu, Zhaowen Wang*

550 MatchED: Crisp Edge Detection Using End-to-End, Matching-based Supervision, *Bedrettin Cetinkaya, Sinan Kalkan, Emre Akbas*

551 SegGBC: Justifiable Coarse-to-Fine Granular-Ball Computing for Enhancing Clustering Image Segmentation, *Qianpeng Chong, Wenyi Zeng, Xiuxuan Shen, Jiajie Li, Qian Yin, Xin Zheng*

552 Seeing Beyond: Extrapolative Domain Adaptive Panoramic Segmentation, *Yuanfan Zheng, Kunyu Peng, Xu Zheng, Kailun Yang*

553 MatchMask: Mask-Centric Generative Data Augmentation for Label-Scarce Semantic Segmentation, *Yuqi Lin, Hao Zhang, Wenqi Shao, Shiqu Liu, Zhihong Gu, Wenxiao Wang, Xiaofei He, Kaipeng Zhang*

554 Boundary-Responsive Differentiable Gating for Superpixel-Based Segmentation, *Fatmaelzahraa Ahmed, Zhihe Lu, Gianni Caro, Diram Tabaa, Mohamed Hamdy, Muraam Abdel-Ghani, Abdulaziz Al-Ali, Muhammad Arsalan, Shidin Balakrishnan*

555 Task-Oriented Data Synthesis and Control-Rectify Sampling for Remote Sensing Semantic Segmentation, *Yunkai Yang, Yudong Zhang, Kunquan Zhang, Jinxiao Zhang, Xinying Chen, Haohuan Fu, Runmin Dong*

556 FUSAR-GPT: A Spatiotemporal Feature-Embedded and Two-Stage Decoupled Visual Language Model for SAR Imagery, *Xiaokun Zhang, Yi Yang, Ziqi Ye, Baiyun Baiyun, Xiaorong Guo, Qingchen Fang, Ruyi Zhang, Xinpeng Zhou, Haipeng Wang*

557 UniChange: Unifying Change Detection with Multimodal Large Language Model, *Xu Zhang, Danyang Li, Xiaohang Dong, Tianhao Wu, Hualong Yu, Jianye Wang, Qicheng Li, Xiang Li*

558 Spatiotemporal Pyramid Flow Matching for Climate Emulation, *Jeremy A. Irvin, Jiaqi Han, Zikui Wang, Abdulaziz Alharbi, Yufei Zhao, Nomin-Erdene Bayarsaikhan, Daniele Visioni, Andrew Y. Ng, Duncan Watson-Parris*

559 ❊ See What We Cannot See: A Geo-guided Reasoning Benchmark for Object Counting under Adverse Earth Observation Conditions, *Jiayi Wang, Zhihong Tan, Hongchen Wei, Daiqin Yang, Zhenzhong Chen*

560 MM-OVSeg: Multimodal Optical–SAR Fusion for Open-Vocabulary Segmentation in Remote Sensing, *Yimin Wei, Aoran Xiao, Hongruixuan Chen, Junshi Xia, Naoto Yokoya*

561 RECS4R: Bridging Semantics and Geometry for Referring Remote Sensing Interpretation, *Jinming Chai, Lingling Li, Licheng Jiao, Xiaoqiang Lu, Long Sun, Xu Liu, Wenping Ma, Weibin Li*

562 Fourier Angle Alignment for Oriented Object Detection in Remote Sensing, *Changyu Gu, Linwei Chen, Lin Gu, Ying Fu*

563 Learning to Infer Parameterized Representations of Plants from 3D Scans, *Samara Ghrer, Christophe Godin, Stefanie Wuhrer*

564 Good Can Sometimes be Bad: A Unified Attack against 3D Point Cloud Classifier by a Flexible Isotropic Resampling, *Linkun Fan, Jiahao Zhang, Juntao Zhang, Lei Zhang, Fazhi He, Daojun Han*

565 V-Attack: Targeting Disentangled Value Features for Controllable Adversarial Attacks on LVLMs, *Sen Nie, Jie Zhang, Jianxin Yan, Shiguang Shan, Xilin Chen*

566 FeatureFool: Zero-Query Fooling of Video Models via Feature Map, *Duoxun Tang, Xi Xiao, Guangwu Hu, Kangkang Sun, Xiao Yang, Dongyang Chen, Qing Li, Yong-jie Yin, Jiyao Wang*

567 RankOOD - Class Ranking-based Out-of-Distribution Detection, *Dishanika Denipitiyage, Naveen Karunanayake, Suranga Seneviratne, Sanjay Chawla*

568 AdvFM: Lookahead Flow-Matching Velocity-Field Attacks for Imperceptible and Transferable Adversarial Examples, *Runze Liu, Zeyue Wang, Fanghui Sun, Rui Liu, Yihan Yan, Shen Wang, Zhaoyang Zhang*

569 The Power of Decaying Steps: Enhancing Attack Stability and Transferability for Sign-based Optimizers, *Wei Tao, Yang Dai, Jincai Huang, Qing Tao*

570 Your Classifier Can Do More: Towards Balancing the Gaps in Classification, Robustness, and Generation, *Kaichao Jiang, He Wang, Xiaoshuai Hao, Xiulong Yang, Ajian Liu, Qi Chu, Yunfeng Diao, Richang Hong*

571 Learning Mutual View Information Graph for Adaptive Adversarial Collaborative Perception, *Yihang Tao, Senkang Hu, Haonan An, Zhengru Fang, Hangcheng Cao, Yuguang Fang*

572 Hierarchical Attacks for Multi-Modal Multi-Agent Reasoning, *Hao Zhou, Tiru Wu, Yan Jiang, Wanqi Zhou, Junxing Hu, Ai Han*

573 Omni-Attack: Adversarial Attacks on Open-Ended VQA in Black-Box Multimodal LLMs, *Kai Hu, Weichen Yu, Li Zhang, Alexander Robey, Andy Zou, Haoqi Hu, Chengming Xu, Matt Fredrikson*

574 CoMo: Learning Continuous Latent Motion from Internet Videos for Scalable Robot Learning, *Jiange Yang, Yansong Shi, Haoyi Zhu, Mingyu Liu, Kaijing Ma, Yating Wang, Gangshan Wu, Tong He, Limin Wang*

575 Δynamics: Language-Based Representation for Inferring Rigid-Body Dynamics From Videos, *Chia-Hsiang Kao, Cong Phuoc Huynh, Chien-Yi Wang, Noranart Vesdapunt, Stefan Stojanov, Bharath Hariharan, Oleksandr Obiednikov, Ning Zhou*

576 PvP: Data-Efficient Humanoid Robot Learning with Proprioceptive-Privileged Contrastive Representations, *Mingqi Yuan, Tao Yu, Haolin Song, Bo Li, Xin Jin, Hua Chen, Wenjun Zeng*

577 Diagnose, Correct, and Learn from Manipulation Failures via Visual Symbols, *Xianchao Zeng, Xinyu Zhou, Youcheng Li, Jiayou Shi, Tianle Li, Liangming Chen, Lei Ren, Yong-Lu Li*

578 RealVLG-R1: A Large-Scale Real-World Visual-Language Grounding Benchmark for Robotic Perception and Manipulation, *Linfei Li, Lin Zhang, Ying Shen*

579 GeCo-SRT: Geometry-aware Continual Adaptation for Cross-Task Sim-to-Real Transfer, *Wenbo Yu, Wenke Xia, Weitao Zhang, Di Hu*

580 ActiveGrasp: Information-Guided Active Grasping with Calibrated Energy-based Model, *Boshu Lei, Wen Jiang, Kostas Daniilidis*

581 ❊ BiPreManip: Learning Affordance-Based Bimanual Pre-Manipulation through Anticipatory Collaboration, *Yan Shen, Feng Jiang, Zichen He, Xiaoqi Li, Yuchen Liu, Zhiyu Li, Ruihai Wu, Hao Dong*

582 Learning Surgical Robotic Manipulation with 3D Spatial Priors, *Yu Sheng, Lidian Wang, Xiaomeng Chu, Jiajun Deng, Min Cheng, Yanyong Zhang, Bei Hua, Houqiang Li, Jianmin Ji*

583 ❊ SimRecon: SimReady Compositional Scene Reconstruction from Real Videos, *Chong Xia, Kai Zhu, Zizhuo Wang, Fangfu Liu, Zhizheng Zhang, Yueqi Duan*

584 ❊ STRNet: Visual Navigation with Spatio-Temporal Representation through Dynamic Graph Aggregation, *Hao Ren, Zetong Bi, Yiming Zeng, Zhaoliang Wan, Lu Qi, Hui Cheng*

585 ❊ RaUF: Learning the Spatial Uncertainty Field of Radar, *Shengpeng Wang, Kuangyu Wang, Wei Wang*

586 SIR: Structured Image Representations for Explainable Robot Learning, *Paul Mattes, Jan Schwab, Jens Bosch, Maximilian Xiling Li, Nils Blank, Minh-Trung Tang, Moritz Haberland, Rudolf Lioutikov*

587 Instance-level Visual Active Tracking with Occlusion-Aware Planning, *Haowei Sun, Kai Zhou, Hao Gao, Shiteng Zhang, Jinwu Hu, Xutao Wen, Qixiang Ye, Mingkui Tan*

588 Mantis: A Versatile Vision-Language-Action Model with Disentangled Visual Foresight, *Yi Yang, Xueqi Li, Yiyang Chen, Jin Song, Yihan Wang, Zipeng Xiao, Jiadi Su, You Qiaoben, Pengfei Liu, Zhijie Deng*

589 AnthroTAP: Learning Point Tracking with Real-World Motion, *Inès Hyeonsu Kim, Seokju Cho, Jahyeok Koo, Junghyun Park, Jiahui Huang, Honglak Lee, Joon-Young Lee, Seungryong Kim*

590 Tracking by Predicting 3-D Gaussians Over Time, *Tanish Baranwal, Himanshu Gaurav Singh, Jathushan Rajasegaran, Jitendra Malik*
591 Toward Low-Cost yet Effective Temporal Learning for UAV Tracking, *Chaocan Xue, Qihua Liang, Bineng Zhong, Yanting Zu, Yuanliang Xue, Haiying Xia, Shuxiang Song*
592 Rethinking Two-Stage Referring-by-Tracking in Referring Multi-Object Tracking: Make it Strong Again, *Weize Li, Yunhao Du, Qixiang Yin, Zhicheng Zhao, Fei Su*
593 Occlusion-Aware SORT: Observing Occlusion for Robust Multi-Object Tracking, *Chunjiang Li, Jianbo Ma, Li Shen, Yanru Chen, Liangyin Chen*
594 CoWTracker: Tracking by Warping instead of Correlation, *Zihang Lai, Eldar Insafutdinov, Edgar Sucar, Andrea Vedaldi*
595 Learning Long-term Motion Embeddings for Efficient Kinematics Generation, *Nick Stracke, Kolja Bauer, Stefan Andreas Baumann, Miguel Ángel Bautista, Josh Susskind, Björn Ommer*
596 SpatialVID: A Large-Scale Video Dataset with Spatial Annotations, *Jiahao Wang, Yufeng Yuan, Rujie Zheng, Youtian Lin, Jian Gao, Lin-Zhuo Chen, Yajie Bao, Chang Zeng, Yanxi Zhou, Xiao-Xiao Long, Hao Zhu, Zhaoxiang Zhang, Xun Cao, Yao Yao*
597 Beyond Explicit Language: Plug-and-Play Visual-to-Linguistic Modeling Toward General Object Tracking, *Kaiyang Lan, Ying Cui, Chenchen Jing, Jianwei Zheng, Dongyan Guo*
598 FairLLaVA: Fairness-Aware Parameter-Efficient Fine-Tuning for Large Vision-Language Assistants, *Mahesh Bhosale, Abdul Wasi, Shantam Srivastava, Shifa Latif, Tianyu Luan, Mingchen Gao, David Doermann, Xuan Gong*
599 InvCoSS: Inversion-driven Continual Self-supervised Learning in Medical Multi-modal Image Pre-training, *Zihao Luo, Shaohao Rui, Zhenyu Tang, Guotai Wang, Xiaosong Wang*
600 PETAR: Localized Findings Generation with Mask-Aware Vision-Language Modeling for PET Automated Reporting, *Danyal Maqbool, Changhee Lee, Zachary Huemann, Samuel D. Church, Matthew E. Larson, Scott B. Perlman, Tomas A. Romero, Joshua D. Warner, Meghan Lubner, Xin Tie, Jameson Merkow, Junjie Hu, Steve Y. Cho, Tyler J. Bradshaw*
601 From Panel to Pixel: Zoom-In Vision–Language Pretraining from Biomedical Scientific Literature, *Kun Yuan, Min Sun, Zhen Chen, Alejandro Lozano, Xiangteng He, Shi Li, Nassir Navab, Xiaoxiao Sun, Nicolas Padoy, Serena Yeung-Levy*
602 LEMON: A Large Endoscopic MONocular Dataset and Foundation Model for Perception in Surgical Settings, *Chengan Che, Chao Wang, Tom Vercauteren, Sophia Tsoka, Luis C. Garcia-Peraza-Herrera*
603 D2T2 - Multimodal Automated Planning for Brachytherapy, *Lance C. Moore, Aranyo Mitra, Ryan Truong, Karoline Kallis, Kelly Kisling, Sandra M. Meyers, Nuno Vasconcelos*
604 TopoCL: Topological Contrastive Learning for Medical Imaging, *Guangyu Meng, Pengfei Gu, Peixian Liang, John P. Lalor, Erin Wolf Chambers, Danny Z. Chen*
605 Diffusion with a Linguistic Compass: Steering the Generation of Clinically Plausible Future sMRI Representations for Early MCI Conversion Prediction, *Zhihao Tang, Chaozhuo Li, Litian Zhang, Xi Zhang*
606 Personalized Longitudinal Medical Report Generation via Temporally-Aware Federated Adaptation, *He Zhu, Ren Togo, Takahiro Ogawa, Kenji Hirata, Minghui Tang, Takaaki Yoshimura, Hiroyuki Sugimori, Noriko Nishioka, Yukie Shimizu, Kohsuke Kudo, Miki Haseyama*
607 Decoding 3D Perception via BrainSSD: Synergistic Fusion of EEG Representations from Static and Dynamic Visual Streams, *Yincheng Yao, Enze Shi, Shu Zhang*
608 Duala: Dual-Level Alignment of Subjects and Stimuli for Cross-Subject fMRI Decoding, *Shumeng Li, Jintao Guo, Jian Zhang, Yulin Zhou, Luyang Cao, Yinghuan Shi*
609 OmniBrainBench: A Comprehensive Multimodal Benchmark for Brain Imaging Analysis Across Multi-stage Clinical Tasks, *Zhihao Peng, Cheng Wang, Shengyuan Liu, Zhiying Liang, Zanting Ye, Min Jie Ju, Peter YM Woo, Yixuan Yuan*
610 Beyond Pixel Simulation: Pathology Image Generation via Diagnostic Semantic Tokens and Prototype Control, *Minghao Han, Yichen Liu, Yizhou Liu, Zizhi Chen, Jingqun Tang, Xuecheng Wu, Dingkang Yang, Lihua Zhang*
611 MedFG-VQA: Low-Frequency Memory and Graph Attention for Lightweight Medical VQA, *Haowen Gu, Gensheng Pei, Zeren Sun, Mingwu Ren, Xiangbo Shu, Yazhou Yao, Fumin Shen*
612 FISHuman: Fine-grained Single-image 3D Human Reconstruction via Multi-view 4D Remeshing, *Hanxi Liu, Yifang Men, Zhouhui Lian*
613 DuoMo: Dual Motion Diffusion for World-Space Human Reconstruction, *Yufu Wang, Evonne Ng, Soyong Shin, Rawal Khirodkar, Yuan Dong, Zhaoen Su, Jinhyung Park, Kris Kitani, Alexander Richard, Fabian Prada, Michael Zollhöfer*
614 RAM: Recover Any 3D Human Motion in-the-Wild, *Sen Jia, Ning Zhu, Jinqin Zhong, Jiale Zhou, Huaping Zhang, Jenq-Neng Hwang, Lei Li*
615 From 2D Alignment to 3D Plausibility: Unifying Heterogeneous 2D Priors and Penetration-Free Diffusion for Occlusion-Robust Two-Hand Reconstruction, *Gaoge Han, Yongkang Cheng, Zhe Chen, Shaoli Huang, Tongliang Liu*
616 MV-Fashion: Towards Enabling Virtual Try-On and Size Estimation with Multi-View Paired Data, *Hunor Laczkó, Libang Jia, Loc-Phat Truong, Diego Hernández, Sergio Escalera, Jordi Gonzalez, Meysam Madadi*
617 Forecasting 3D Scanpaths in Egocentric Video, *Fiona Ryan, Ishwarya Ananthabhotla, Yijun Qian, Judy Hoffman, James M. Rehg, Vamsi Krishna Ithapu, Calvin Murdock*
618 M4Human: A Large-Scale Multimodal mmWave Radar Benchmark for Human Mesh Reconstruction, *Junqiao Fan, Yunjiao Zhou, Yizhuo Yang, Xinyuan Cui, Jiarui Zhang, Lihua Xie, Jianfei Yang, Chris Xiaoxuan Lu, Fangqiang Ding*
619 ReGenHOI: Unifying Reconstruction and Generation for 3D Human–Object Interaction Understanding, *Miao Xu, Xiangyu Zhu, Zidu Wang, Xusheng Liang, Bao Li, Jinlin Wu, Zelin Zang, Zhen Lei*
620 Through the Frequency Lens: Cross-Domain Generalisable Gaze Estimation with Adaptive Modulation, *Yang Xu, Yiwei Bao, Feng Lu*
621 Mocap-2-to-3: Multi-view Lifting for Monocular Motion Recovery with 2D Pretraining, *Zhumei Wang, Zechen Hu, Ruoxi Guo, Huaijin Pi, Ziyong Feng, Liang Zhang, Mingtao Pei, Siyuan Huang*
622 SHands: A Multi-View Dataset and Benchmark for Surgical Hand-Gesture and Error Recognition Toward Medical Training, *Le Ma, Thiago Freitas dos Santos, Nadia Magnenat-Thalmann, Katarzyna Wac*
623 Beyond Static Frames: Temporal Aggregate-and-Restore Vision Transformer for Human Pose Estimation, *Hongwei Fang, Jiahang Cai, Xun Wang, Wenwu Yang*
624 IMU-HOI: A Symbiotic Framework for Coherent Human-Object Interaction and Motion Capture via Contact-Conscious Inertial Fusion, *Lizhou Lin, Songpengcheng Xia, Zengyuan Lai, Lan Sun, Jiarui Yang, Ling Pei*
625 Learning Forgery-Aware Lip Representations Without Forgery Priors, *Bofan Chen, Hongyu Zhu, Yi He, Sichu Liang, Shi-Lin Wang*
626 Beyond [CLS] Token: Query-Driven Token-Level Forgery Purification for Generalizable Deepfake Detection, *Changshuo Wang, Jiangming Wang, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, Shunli Wang, Ran Yi, Lizhuang Ma*
627 GEM-TFL: Bridging Weak and Full Supervision for Forgery Localization through EM-Guided Decomposition and Temporal Refinement, *Xiaodong Zhu, Yuanming Zheng, Suting Wang, Junqi Yang, Yuhong Yang, Weiping Tu, Zhongyuan Wang*
628 TokenTrace: Multi-Concept Attribution through Watermarked Token Recovery, *Li Zhang, Shruti Agarwal, John Collomosse, Pengtao Xie, Vishal Asnani*
629 Unleashing Vision-Language Semantics for Deepfake Video Detection, *Jiawen Zhu, Yunqi Miao, Xueyi Zhang, Jiankang Deng, Guansong Pang*
630 A Difference-in-Difference Approach to Detecting AI-Generated Images, *Xinyi Qi, Kai Ye, Chengchun Shi, Ying Yang, Jin Zhu, Hongyi Zhou*
631 RDFace: A Benchmark Dataset for Rare Disease Facial Image Analysis under Extreme Data Scarcity and Phenotype-Aware Synthetic Generation, *Ganlin Feng, Yuxi Long, Hafsa Ali, Erin Lou, Fahad Butt, Qian Liu, Yang Wang, Pingzhao Hu*
632 ActivityForensics: A Comprehensive Benchmark for Localizing Manipulated Activity in Videos, *Peijun Bao, Anwei Luo, Gang Pan, Alex C. Kot, Xudong Jiang*
633 Zero-shot Detection of AI-Generated Image via RAW-RGB

Alignment, *Haiwei Wu, Fengpeng Li, Zhilin Tu, Yuanman Li, Xiong Li, Jiantao Zhou*

634 Scaling Up AI-Generated Image Detection with Generator-Aware Prototypes, *Ziheng Qin, Yuheng Ji, Renshuai Tao, Yuxuan Tian, Yuyang Liu, Yipu Wang, Xiaolong Zheng*

635 Investigating Self-Supervised Representations for Audio-Visual Deepfake Detection, *Dragos-Alexandru Boldisor, Stefan Smeu, Dan Oneata, Elisabeta Oneata*

636 TIACam: Text-Anchored Invariant Feature Learning with Auto-Augmentation for Camera-Robust Zero-Watermarking, *Abdullah Tanvir, Agnibh Dasgupta, Xin Zhong*

637 FastRef: Fast Prototype Refinement for Few-shot Industrial Anomaly Detection, *Yufei Li, Long Tian, Yuyang Dai, Wenchao Chen, Liang Bao, Xiyang Liu*

638 RC-NF: Robot-Conditioned Normalizing Flow for Real-Time Anomaly Detection in Robotic Manipulation, *Shijie Zhou, Bin Zhu, Jiarui Yang, Xiangyu Zhao, Jingjing Chen, Yu-Gang Jiang*

639 Reasoning-Driven Anomaly Detection and Localization with Image-Level Supervision, *Yizhou Jin, Yuezhu Feng, Jinjin Zhang, Peng Wang, Qingjie Liu, Yunhong Wang*

640 MMR-AD: A Large-Scale Multimodal Dataset for Benchmarking General Anomaly Detection with Multimodal Large Language Models, *Xincheng Yao, Zefeng Qian, Chao Shi, Jiayang Song, Chongyang Zhang*

641 ❊ Wavelet-Driven 3D Anomaly Detection under Pose-Agnostic and Sparse-View, *Mingwen Shao, Qiao Zhang, Xinyuan Chen, Xiang Lv, Lingzhuang Meng, Chang Liu, Qinglin Zhan, Ling Jian*

642 Hunting Normality from Query Sample via Residual Learning for Generalist Anomaly Detection, *Xiaolei Wang, Yuexin Wang, Tianhong Dai, Huihui Bai, Yao Zhao, Jimin Xiao*

643 GPFlow: Gaussian Prototype Probability Flow for Unsupervised Multi-Modal Anomaly Detection, *Yiting Li, Xulei Yang, Jingyi Liao, Jing Zhang, Fayao Liu*

644 HP-Edit: A Human-Preference Post-Training Framework for Image Editing, *Fan Li, Chonghuinan Wang, Lina Lei, Yuping Qiu, Jiaqi Xu, Jiaxiu Jiang, Xinran Qin, Zhikai Chen, Fenglong Song, Zhixin Wang, Renjing Pei, Wangmeng Zuo*

645 It's Never Too Late: Noise Optimization for Collapse Recovery in Trained Diffusion Models, *Anne Harrington, A. Sophia Koepke, Shyamgopal Karthik, Trevor Darrell, Alexei A. Efros*

646 RebRL: Reinforcing Discrete Visual Diffusion Models with Rebalanced Timestep Credits, *Mu Zhang, Tianren Ma, Yunfan Liu, Kun Hu, Qixiang Ye*

647 Ego-InBetween: Generating Object State Transitions in Ego-Centric Videos, *Mengmeng Ge, Takashi Isobe, Xu Jia, Yanan Sun, Zetong Yang, Weinong Wang, Dong Zhou, Dong Li, Huchuan Lu, Emad Barsoum*

648 Towards Fine-Grained Attribution: Instance-Aware Preference Optimization for Aligning Diffusion Models, *Jiayang Sun, Pin Wang, Hongbo Wang, Xinyue Liu, Huaibo Huang, Ran He*

649 SketchRevive: Fine-Grained Pixel-to-Vector Sketch Completion with Diffusion-Prior-Guided Multimodal LLMs, *Ran Zuo, Haoxiang Hu, Chenxi Pei, Yanxuan Liu, Wenwen Qiang, Fang Liu, Xiaoming Deng, Cuixia Ma, Yong-Jin Liu*

650 UniPercept: A Unified Diffusion Model for Generalizable Visual Perception, *Zuyan Zhao, Zhenliang He, Meina Kan, Shiguang Shan, Xilin Chen*

651 ❊ Visual Diffusion Models are Geometric Solvers, *Nir Goren, Shai Yehezkel, Omer Dahary, Andrey Voynov, Or Patashnik, Daniel Cohen-Or*

652 You Only Erase Once: Erasing Anything without Bringing Unexpected Content, *Yixing Zhu, Qing Zhang, Wenju Xu, Wei-Shi Zheng*

653 Smoothing the Score Function to Enhance Generalization in Diffusion Models, *Xinyu Zhou, Jiawei Zhang, Stephen J. Wright*

654 NS-Diff: Fluid Navier–Stokes Guided Video Diffusion via Reinforcement Learning, *Zijun Deng, Yuxin Peng*

655 PropFly: Learning to Propagate via On-the-Fly Supervision from Pre-trained Video Diffusion Models, *Wonyong Seo, Jaeho Moon, Jaehyup Lee, Soo Ye Kim, Munchurl Kim*

656 Generative Neural Video Compression via Video Diffusion Prior, *Qi Mao, Hao Cheng, Tinghan Yang, Libiao Jin, Siwei Ma*

657 AdaCluster: Adaptive Query-Key Clustering for Sparse Attention in Video Generation, *Haoyue Tan, Shengnan Wang, Yulin Qiao, Juncheng Zhang, Youhui Bai, Ping Gong, Zewen Jin, Cheng Li*

658 Denoising, Fast and Slow: Difficulty-Aware Adaptive Sampling for Image Generation, *Johannes Schusterbauer, Ming Gui, Yusong Li, Pingchuan Ma, Felix Krause, Björn Ommer*

659 ❊ Image Diffusion Preview with Consistency Solver, *Fu-Yun Wang, Hao Zhou, Liangzhe Yuan, Sanghyun Woo, Boqing Gong, Bohyung Han, Ming-Hsuan Yang, Han Zhang, Yukun Zhu, Ting Liu, Long Zhao*

660 The Drift Kernel: Why Diffusion Models Change Even When Told Not To, *Gokul Srinath Seetha Ram, Rashmi Elavazhagan*

661 Interpretable Prompts made Edit-Friendly: Token-to-Token Similarity Reduction in dLLMs for Edit-Friendly Hard Prompt Inversion, *Naresh Kumar Devulapally, Shruti Agarwal, Vishal Asnani, Vishnu Suresh Lokhande*

662 LESA: Learnable Stage-Aware Predictors for Diffusion Model Acceleration, *Peiliang Cai, Jiacheng Liu, Haowen Xu, Xinyu Wang, Chang Zou, Linfeng Zhang*

663 Vision Foundation Models Can Be Good Tokenizers for Latent Diffusion Models, *Tianci Bi, Xiaoyi Zhang, Yan Lu, Nanning Zheng*

664 Adaptive Spectral Feature Forecasting for Diffusion Sampling Acceleration, *Jiaqi Han, Juntong Shi, Puheng Li, Haotian Ye, Qiushan Guo, Stefano Ermon*

665 Proxy-Tuning: Tailoring Multimodal Autoregressive Models for Subject-Driven Image Generation, *Yi Wu, Shengju Qian, Lingting Zhu, Lei Liu, Wandi Qiao, Ziqiang Li, Lequan Yu, Bin Li*

666 EasyOmnimatte: Taming Pretrained Inpainting Diffusion Models for End-to-End Video Layered Decompositio, *Yihan Hu, Xuelin Chen, Xiaodong Cun*

667 Hierarchical Codec Diffusion for Video-to-Speech Generation, *Jiaxin Ye, Gaoxiang Cong, Chenhui Wang, Xin-Cheng Wen, Zhaoyang Li, Boyuan Cao, Hongming Shan*

668 Semantic Alignment for Pose-Invariant Identity Preserving Diffusion, *Jiwon Kim, SeonHwa Kim, Soobin Park, Eunju Cha, Kyong Hwan Jin*

669 Causality in Video Diffusers is Separable from Denoising, *Xingjian Bai, Guande He, Zhengqi Li, Eli Shechtman, Xun Huang, Zongze Wu*

670 2ndMatch: Finetuning Pruned Diffusion Models via Second-Order Jacobian Matching, *Caleb Zheng, Eli Shlizerman*

671 Hear What You See: Video-to-Audio Generation with Diffusion Transformer and Semantic-Temporal Alignment-Ranked Direct Preference Optimization, *Kai Wang, Tao Zhou, Jiayi Lei, Jing Wang, Jinman Zhao, Weiguo Pian, Yuan Cheng, Yapeng Tian, Peng Gao, Bin Fu, Yihao Liu, Dimitrios Hatzinakos, Yuewen Cao*

672 ❊ MacTok: Robust Continuous Tokenization for Image Generation, *Hengyu Zeng, Xin Gao, Guanghao Li, Yuxiang Yan, Jiaoyang Ruan, Junpeng Ma, Haoyu Albert Wang, Jian Pu*

673 Group Editing: Edit Multiple Images in One Go, *Yue Ma, Xinyu Wang, Qianli Ma, Qinghe Wang, Mingzhe Zheng, Xiangpeng Yang, Hao Li, Chongbo Zhao, Jixuan Ying, Harry Yang, Hongyu Liu, Qifeng Chen*

674 Adaptive Video Distillation: Mitigating Oversaturation and Temporal Collapse in Few-Step Generation, *Yuyang You, Yongzhi Li, Jiahui Li, Yadong Mu, Quan Chen, Peng Jiang*

675 Beyond the Golden Data: Resolving the Motion-Vision Quality Dilemma via Timestep Selective Training, *Xiangyang Luo, Qingyu Li, Yuming Li, Guanbo Huang, Yongjie Zhu, Wenyu Qin, Meng Wang, Pengfei Wan, Shao-Lun Huang*

676 Toward Diffusible High-Dimensional Latent Spaces: A Frequency Perspective, *Bolin Lai, XuDong Wang, Saketh Rambhatla, James M. Rehg, Zsolt Kira, Rohit Girdhar, Ishan Misra*

677 Elucidating the SNR-t Bias of Diffusion Probabilistic Models, *Meng Yu, Lei Sun, Jianhao Zeng, Xiangxiang Chu, Kun Zhan*

678 What Is It Like to Be a Noise? An Entropy-based Gaussian Noise Regularization for Diffusion Models, *Pascal Chang, Kai Lascheit, Jingwei Tang, Markus Gross, Vinicius C. Azevedo*

679 FlashVSR: Towards Real-time Diffusion-Based Streaming Video Super Resolution, *Junhao Zhuang, Shi Guo, Xin Cai, Xiaohui Li, Yihao Liu, Chun Yuan, Tianfan Xue*

680 DiffusionHarmonizer: Bridging Neural Reconstruction and Photorealistic Simulation with Online Diffusion Enhancer, *Yuxuan Zhang, Katarína Tóthová, Zian Wang, Kangxue Yin, Haithem Turki, Riccardo de Lutio, Yen-Yu Chang, Or Litany, Sanja Fidler, Zan Gojcic*

681 GDRO: Group-level Reward Post-training Suitable for Diffusion Models, *Yiyang Wang, Xi Chen, Xiaogang Xu, Yu Liu, Hengshuang Zhao*
682 RFDM: Residual Flow Diffusion Models for Video Editing, *Mohammadreza Salehi, Mehdi Noroozi, Luca Morreale, Ruchika Chavhan, Malcolm Chadwick, Alberto Gil Couto Pimentel Ramos, Abhinav Mehrotra*
683 FreqEdit: Preserving High-Frequency Features for Robust Multi-Turn Image Editing, *Yucheng Liao, Jiajun Liang, Kaiqian Cui, Baoquan Zhao, Haoran Xie, Wei Liu, Qing Li, Xudong Mao*
684 ❊ Graph-Guided Online Concept Erasure for Text-to-Image Diffusion Models, *Ning Han, Zhenyu Ge, Feng Han, Yuhua Sun, Chengqing Li, Jingjing Chen*
685 HierEdit: Region-Aware Hierarchical Diffusion for Efficient High-Resolution Editing, *Yuyao Zhang, Alexander Huang-Menders, Yu-Wing Tai*
686 CTCal: Rethinking Text-to-Image Diffusion Models via Cross-Timestep Self-Calibration, *Xiefan Guo, Xinzhu Ma, Haiyu Zhang, Di Huang*
687 Edit2Perceive: Image Editing Diffusion Models Are Strong Dense Perceivers, *Yiqing Shi, Yiren Song, Mike Zheng Shou*
688 DeltaQuant: 4-bit Video Diffusion Models with Spatiotemporal Delta Smoothing, *Xingyang Li, Samuel Tesfai, Zhekai Zhang, Haocheng Xi, Shuo Yang, Lvmin Zhang, Yufei Sun, Kelly Peng, Maneesh Agrawala, Ion Stoica, Kurt Keutzer, Jun-Yan Zhu, Song Han, Yujun Lin, Muyang Li*
689 D2Cache: Second-Order Delta Caching for Higher Video Diffusion Acceleration, *Enhuai Liu, Yunke Wang, Changming Sun, Chang Xu*
690 ❊ DeCo: Frequency-Decoupled Pixel Diffusion for End-to-End Image Generation, *Zehong Ma, Longhui Wei, Shuai Wang, Shiliang Zhang, Qi Tian*
691 Test-Time Alignment of Text-to-Image Diffusion Models via Null-Text Embedding Optimisation, *Taehoon Kim, Henry Gouk, Timothy Hospedales*
692 Accelerating Diffusion Model Training under Minimal Budgets: A Condensation-Based Perspective, *Rui Huang, Shitong Shao, zikai zhou, Pukun Zhao, Hangyu Guo, Tian Ye, Lichen Bai, Shuo Yang, Zeke Xie*
693 Denoising as Path Planning: Training-Free Acceleration of Diffusion Models with DPCache, *Bowen Cui, Yuanbin Wang, Huajiang Xu, Biaolong Chen, Aixi Zhang, Hao Jiang, Zhengzheng Jin, Xu Liu, Pipei Huang*
694 Taming Sampling Perturbations with Variance Expansion Loss for Latent Diffusion Models, *Qifan Li, Xingyu Zhou, Jinhua Zhang, Weiyi You, Shuhang Gu*
695 Guiding Diffusion Models with Semantically Degraded Conditions, *Shilong Han, Yuming Zhang, Hongxia Wang*
696 Semantics Lead the Way: Harmonizing Semantic and Texture Modeling with Asynchronous Latent Diffusion, *Yueming Pan, Ruoyu Feng, Qi Dai, Yuqi Wang, Wenfeng Lin, Mingyu Guo, Chong Luo, Nanning Zheng*
697 Reviving ConvNeXt for Efficient Convolutional Diffusion Models, *Taesung Kwon, Lorenzo Bianchi, Lennart Wittke, Felix Watine, Fabio Carrara, Jong Chul Ye, Romann Weber, Vinicius Azevedo*
698 Coupled Diffusion Sampling for Training-Free Multi-View Image Editing, *Hadi Alzayer, Yunzhi Zhang, Chen Geng, Jia-Bin Huang, Jiajun Wu*
699 Improving Diffusion Generalization with Weak-to-Strong Segmented Guidance, *Liangyu Yuan, Yufei Huang, Mingkun Lei, Tong Zhao, Ruoyu Wang, Changxi Chi, Yiwei Wang, Chi Zhang*
700 Adaptive Auxiliary Prompt Blending for Target-Faithful Diffusion Generation, *Kwanyoung Lee, SeungJu Cha, Yebin Ahn, Hyunwoo Oh, Sungho Koh, Dong-Jin Kim*
701 SegQuant: A Semantics-Aware and Generalizable Quantization Framework for Diffusion Models, *Jiaji Zhang, Ruichao Sun, Hailiang Zhao, Jiaju Wu, Peng Chen, Hao Li, Yuying Liu, Kingsum Chow, Gang Xiong, Shuiguang Deng*
702 BAgger: Backwards Aggregation for Mitigating Drift in Autoregressive Video Diffusion Models, *Ryan Po, Eric Ryan Chan, Changan Chen, Gordon Wetzstein*
703 Accelerating Autoregressive Video Diffusion via History-Guided Cache and Residual Correction, *Kepan Nan, Wangbo Zhao, Penghao Zhou, Jun Li, Zhenheng Yang, Jian Yang, Ying Tai*
704 MusicInfuser: Making Video Diffusion Listen and Dance, *Susung Hong, Ira Kemelmacher-Shlizerman, Brian Curless, Steven M. Seitz*

Notes:

Notes:

Notes:

COLORADO CONVENTION CENTER - OVERALL FLOOR PLAN

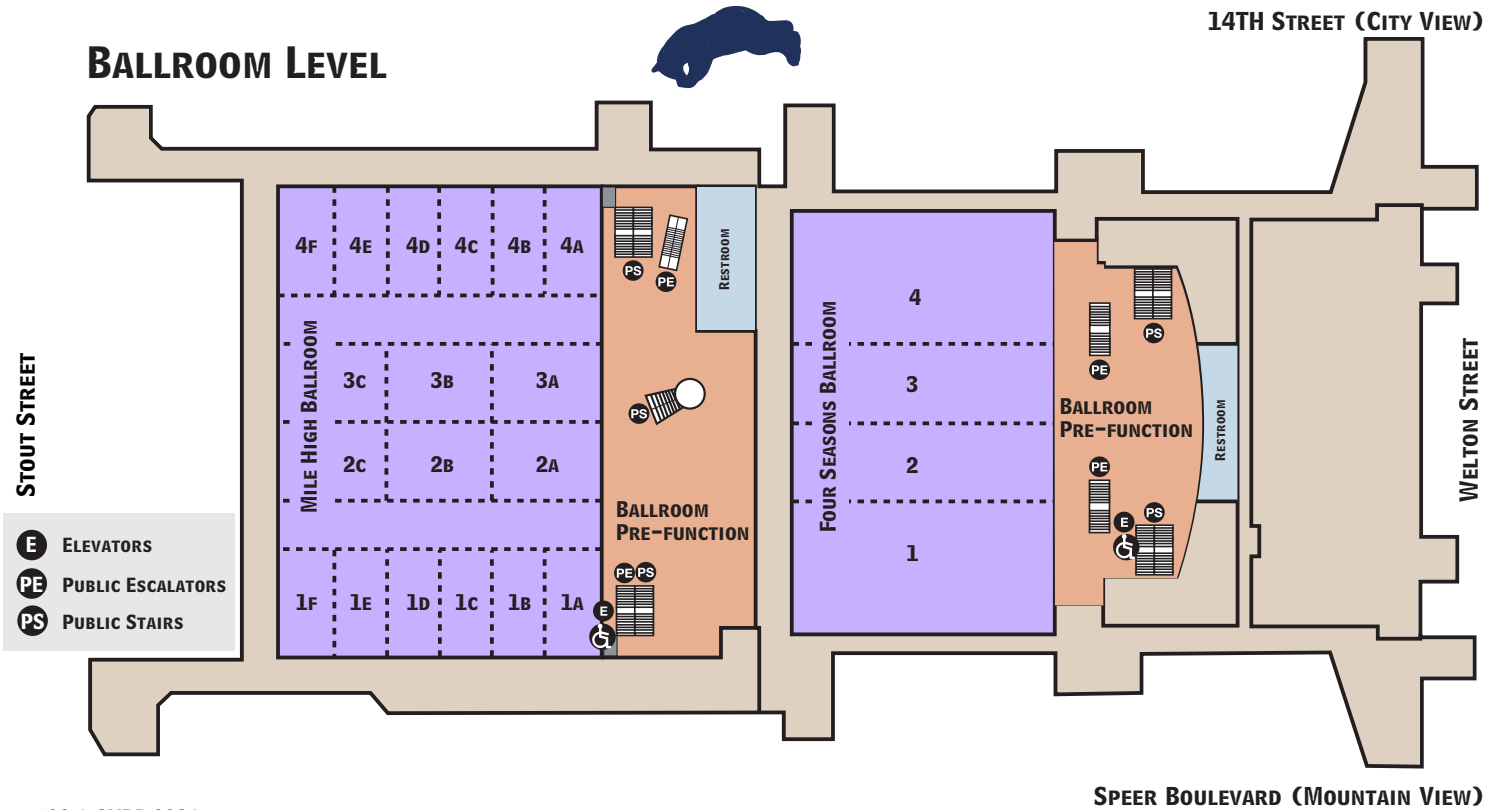
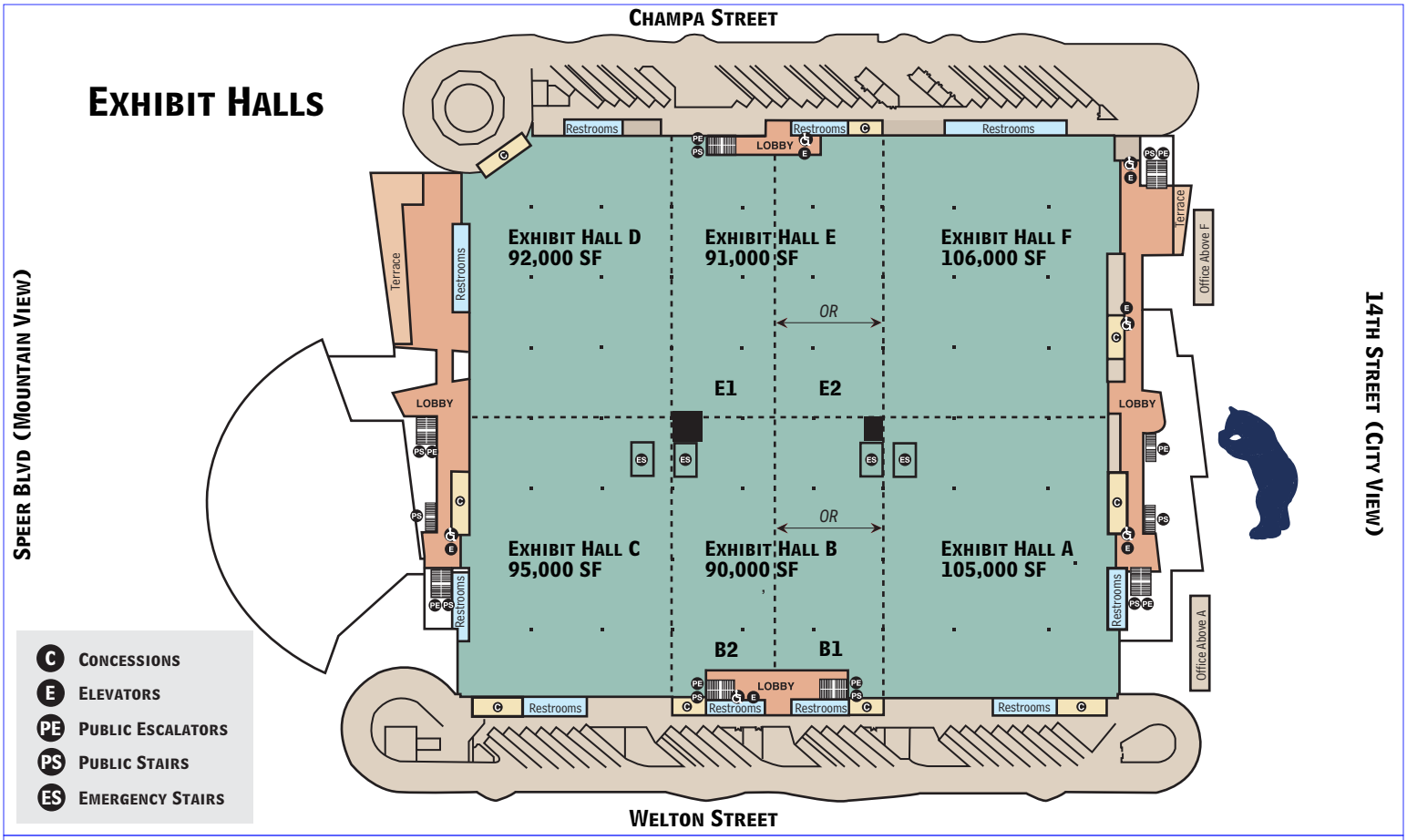
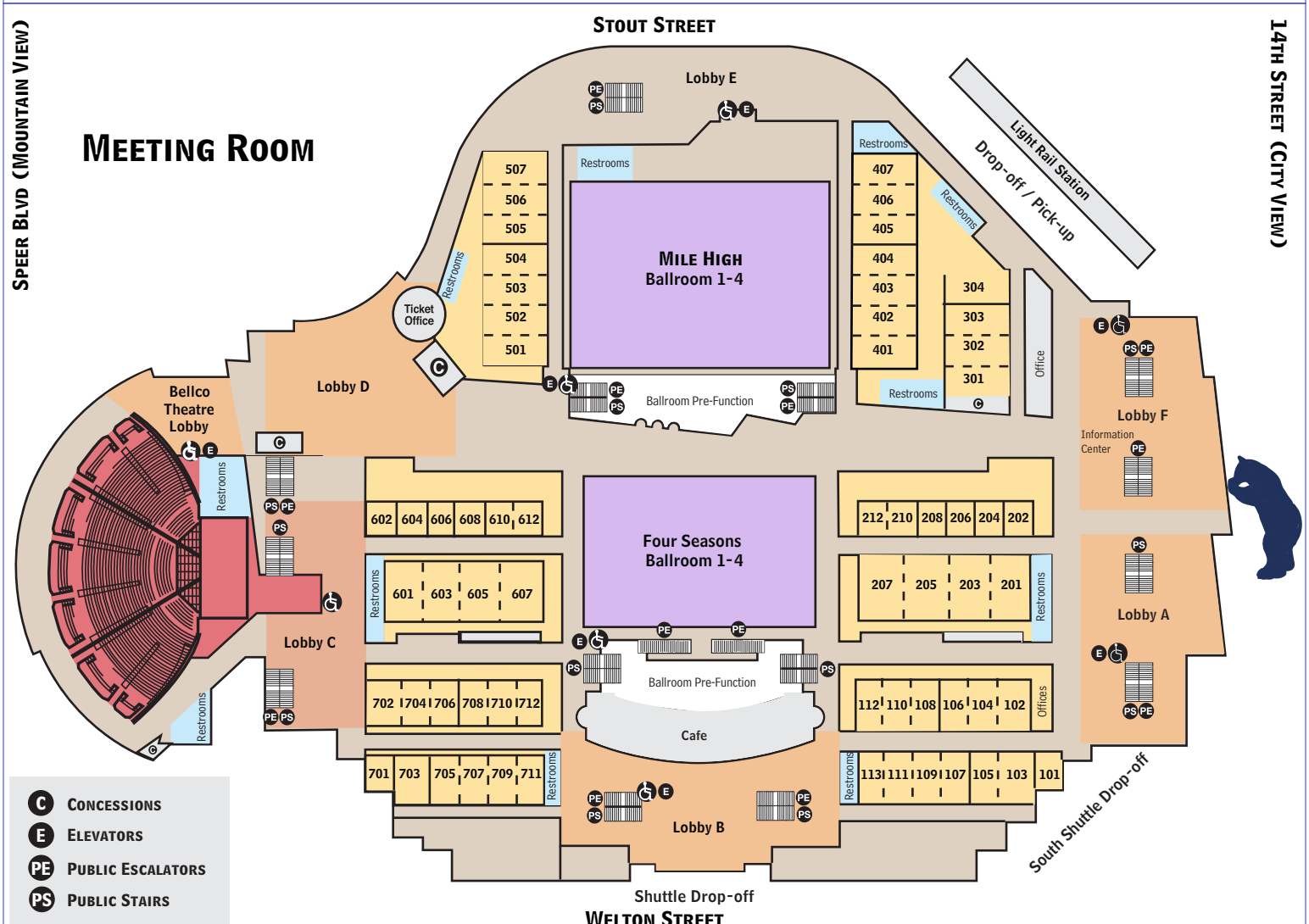


Exhibit Halls



Meeting Room



CVPR
JUNE 3-7, 2026
DENVER
COLORADO