

PROGRAM GUIDE

MAIN CONFERENCE

CVPR 2025

IEEE/CVF Conference on
Computer Vision and
Pattern Recognition



CVPR *Nashville* JUNE 11-15, 2025

PLATINUM SPONSORS



GOLD SPONSORS



SILVER SPONSORS



Welcome to the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Nashville, Tennessee! 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. Our program includes invited keynotes, oral and poster presentations, tutorials, workshops, demos, exhibitions, and social events, all aimed at providing attendees with an exciting and enriching experience. CVPR 2025 is primarily an in-person conference, but for those who are unable to join us physically, we are pleased to offer a virtual component that will provide access to conference papers, posters, videos, and talks.

CVPR 2025 received 13,008 valid paper submissions, a 13% increase from last CVPR. The review process was managed by the 6 Program Co-Chairs, 22 Senior Area Chairs, and 708 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 12,593 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, 2,878 papers were accepted, for a 22.1% 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 of the 2,878 accepted papers were invited to present posters at CVPR. In addition, 96 (3.3%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 387 (13%) were selected by ACs together with Senior ACs to be “highlights” because of their high quality and potential impact. The oral presentations will follow a three-track configuration. The highlights are flagged with a special annotation in the program. New this year, papers authored by outstanding reviewers are also flagged with a special annotation. ACs nominated 15 papers to be 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 the number of submissions and concerns about the deteriorating quality of reviews, the PCs decided to make several major changes to the reviewing process compared to the previous CVPR editions. These changes 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.

We would like to thank everyone involved in making CVPR 2025 a success. This includes the organizing committee, the Senior Area Chairs, Area Chairs, and reviewers, the authors, the demo session participants, contributing artists, and the donors and exhibitors. It is exciting to be part of a team of people around the world united in a common goal.

We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, Yoshitomo Matsubara for his 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. Last but not least, we thank all of you for attending CVPR 2025 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 Nashville during the conference.

Enjoy CVPR 2025. We look forward to meeting you in person!

CVPR Program Chairs

Phillip Isola (*Massachusetts Institute of Technology*)
Hedvig Kjellström (*KTH Royal Institute of Technology*)
Vincent Lepetit (*Ecole des Ponts ParisTech*)
Fuxin Li (*Oregon State University*)
Hao Su (*University of California, San Diego*)
Siyu Tang (*ETH Zürich*)

CVPR 2025 General Chairs

Ce Liu (*Meta*)
Bryan Morse (*Brigham Young University*)
Cristian Sminchisescu (*Google DeepMind and Lund University*)

CVPR 2025 QR CODES

CVPR 2025 Virtual Platform



Access schedules,
papers, workshops,
tutorials, etc

CVPR 2025 Slido Site



Attendee engagement
(Q&A, polls, etc.) for
plenary sessions:
keynotes, panels, orals

Wayfinding App



Find your way around
the Music Center

Exhibit Hall Floorplan



Access location of
exhibitors on the
main floor

Scholar Inbox Paper Recommender



Paper recommender to
locate the posters that
are most relevant to you

General Chairs

Ce Liu (*Meta*)
 Bryan Morse (*Brigham Young University*)
 Cristian Sminchisescu (*Google DeepMind and Lund University*)

Program Chairs

Phillip Isola (*Massachusetts Institute of Technology*)
 Hedvig Kjellström (*KTH Royal Institute of Technology*)
 Vincent Lepetit (*Ecole des Ponts ParisTech*)
 Fuxin Li (*Oregon State University*)
 Hao Su (*University of California, San Diego*)
 Siyu Tang (*ETH Zurich*)

Advisor to the Program Committee

David Forsyth (*University of Illinois at Urbana-Champaign*)

Workshop Chairs

Brian Clipp (*Kitware*)
 Forrester Cole (*Google*)
 Chen Sun (*Brown University*)
 Lijuan Wang (*Microsoft*)

Tutorial Chairs

Katarina Doctor (*U.S. Naval Research Laboratory*)
 Madhu Srinivasan (*AMD*)
 Bolei Zhou (*University of California, Los Angeles*)

Demonstration Chairs

Rahul Garg (*Netflix*)
 Neal Wadhwa (*Google*)

Finance Chair

Walter Scheirer (*University of Notre Dame*)

Publicity Chairs

Kosta Derpanis (*York University*)
 Deblina Bhattacharjee (*University of Bath*)
 Joakim Brunsund Haurum (*Aalborg University*)

Publications Chair

Eric Mortensen

Social Chairs

Vitor Albiero, (*Meta*)
 Yale Song (*Meta*)

Broadening Participation Chairs

Naresh Cuntoor (*BlueHalo Labs*)
 Christopher Funk (*Kitware*)
 Deepti Ghadiyaram (*Boston University*)
 Roni Sengupta (*University of North Carolina at Chapel Hill*)
 Saining Xie (*New York University*)

Senior PAMI-TC Ombud

David Forsyth (*University of Illinois at Urbana-Champaign*)
 Linda Shapiro (*University of Washington*)

AI Art Curator

Luba Elliott (*Independent Curator*)

Technical Chair

Yoshitomo Matsubara (*Yahoo!*)

Doctoral Consortium Chairs

Nathan Jacobs (*Washington University in St. Louis*)
 Abby Stylianou (*Saint Louis University*)

Web Developer

Lee Campbell (*Eventhosts*)

Conference Producer

Nicole Finn (*c to c events*)

CVPR 2025 SENIOR AREA CHAIRS

Andrea Tagliasacchi	Georgia Gkioxari	Kristen Grauman	Pascal Fua	Zeynep Akata
Aude Oliva	Gül Varol	Laura Leal-Taixé	Svetlana Lazebnik	Zhouchen Lin
Cordelia Schmid	Hongdong Li	Marc Niethammer	Tal Hassner	
Deva Ramanan	Katerina Fragkiadaki	Ming-Hsuan Yang	Tinne Tuytelaars	
Dima Damen	Ko Kishino	Minsu Cho	Vladlen Koltun	

CVPR 2025 AREA CHAIRS

A. Sophia Koepke	Aljosa Osep	Antonio Torralba	Boxin Shi	Chen Change Loy
Aayush Bansal	Ameesh Makadia	Anurag Arnab	Boyi Li	Chen Feng
Abby Stylianou	Amir Zamir	Arun Mallya	Brendan Tran Morris	Chen Wang
Abhinav Shrivastava	Anand Bhattad	Asako Kanezaki	Brian Clipp	Chengjiang Long
Abhishek Gupta	Anders Heyden	Atsuto Maki	Brian L. Price	Chenliang Xu
Adam Czajka	Andre Araujo	Baoyuan Wu	Bruce Allen Maxwell	Chenyou Fan
Adam Kortylewski	Andrea Fusiello	Bastian Wandt	Bryan A. Plummer	Chetan Arora
Adams Wai-Kin Kong	Andrei Bursuc	Behzad Bozorgtabar	Bryan Russell	Chi-Keung Tang
Adriana Kovashka	Andrei Zanfir	Benjamin Busam	Bumsub Ham	Chong Luo
Ailing Zeng	Andrew Owens	Benjamin Kimia	C.V. Jawahar	Christian Richardt
Aishwarya Agrawal	Angel X. Chang	Bernt Schiele	Carl Olsson	Christian Rupprecht
Akihiro Sugimoto	Angela Dai	Bharath Hariharan	Carsten Rother	Christian Theobalt
Akshay Gadi Patil	Angela Yao	Bin Fan	Cees G. M. Snoek	Christian Wolf
Alessio Del Bue	Anh Totti Nguyen	Bing Su	Chang Xu	Christoph Feichtenhofer
Alex Schwing	Anh Tuan Tran	Binh-Son Hua	Chanh Kim	Christopher Funk
Alexander C. Berg	Anna Rohrbach	Björn Ommer	Chao Chen	Chuang Gan
Alexandre Alahi	Anpei Chen	Bjorn Stenger	Chao Ma	Chuhang Zou
Alimoor Reza	Anthony Hoogs	Bohyung Han	Chaowei Xiao	Chunyu Wang
Alireza Fathi	Anton Milan	Bolei Zhou	Charless Fowlkes	Cihang Xie

Clement Mallet	Hajime Nagahara	Jonathan T. Barron	Lu Sheng	Or Litany
Cornelia Fermuller	Hamed Pirsiavash	Jong Chul Ye	Lu Yuan	Orazio Gallo
Cuiling Lan	Hamid Rezaatofighi	Jonghyun Choi	Luc Van Gool	Oriane Siméoni
Cuong V. Nguyen	Han-Jia Ye	Jongwoo Lim	Luca Weihs	Pablo Arbelaez
Dahun Kim	Hanbyul Joo	Joost van de Weijer	Luisa Verdoliva	Paola Cascante-Bonilla
Dan Xu	Hao Chen	Jordi Pont-Tuset	Luping Zhou	Paolo Favaro
Daniel Barath	Hazel Doughty	Joseph J. Lim	M. Salman Asif	Pascal Mettes
Daniel Zoran	Henghui Ding	Joseph Tighe	Magnus Oskarsson	Paul Hongsuck Seo
Danna Gurari	Hengshuang Zhao	Josephine Sullivan	Mahdi S. Hosseini	Pedro Morgado
David B. Lindell	Hideki Nakayama	Juan Carlos Nieves	Makarand Tapaswi	Peng Hu
David Fouhey	Hirokatsu Kataoka	Jufeng Yang	Mang Ye	Peng Wang
David J. Crandall	Hiroshi Kawasaki	Jun Liu	Manmohan Chandraker	Per-Erik Forssen
David J. Fleet	Holger Caesar	Jun-Yan Zhu	Manolis Savva	Peter Hedman
David Jacobs	Hong Chang	Jungseock Joo	Marc Pollefeys	Peter Vincent Gehler
David Picard	Hong Xuan	Junhwa Hur	Marcello Pelillo	Peter Wonka
Davide Modolo	Hongyang Li	Junmo Kim	Marco Cristani	Petia Radeva
Deng-Ping Fan	Hossein Azizpour	Junseok Kwon	Marcus Rohrbach	Peyman Milanfar
Derek Hoiem	Hossein Rahmani	Junsong Yuan	Margret Keuper	Philippos Mordohai
Devis Tuia	Hsin-Ying Lee	Junzhou Huang	Maria Vakalopoulou	Pietro Morerio
Di Huang	Hu Han	Jürgen Gall	Marius Leordeanu	Pinar Yanardag
Dim Papadopoulos	Huaizu Jiang	Justus Thies	Mark Yatskar	Ping Hu
Dimitrios Tzionas	Huan Fu	Kai Han	Martin Oswald	Ping Luo
Dimitris Samaras	Huanrui Yang	Kaichun Mo	Massimiliano Mancini	Ping Tan
Dimosthenis Karatzas	Hui Ji	Kaiyang Zhou	Matej Kristan	Piotr Bojanowski
Dingwen Zhang	Huijuan Xu	Kaleem Siddiqi	Matheus Gadelha	Piotr Koniusz
Dong Chen	Hung-Yu Tseng	Karteek Alahari	Mathieu Aubry	Pratul P. Srinivasan
Dong Gong	Hyun Soo Park	Kenneth Marino	Mathieu Salzmann	Qi Dai
Dong Xu	Hyunjung Shim	Kevin Smith	Matthew B. Blaschko	Qi Shan
Dongfang Liu	Hyunwoo J. Kim	Kevis-Kokitsi Maninis	Matthias Niessner	Qi Tian
Du Tran	Iasonas Kokkinos	Khoa Luu	Matthieu Cord	Qi Wu
Eddy Ilg	Ilke Demir	Kiana Ehsani	Matthijs Douze	Qi Yu
Edmond Boyer	Ioannis Gkioulekas	Kihyuk Sohn	Mehrdad Farajtabar	Qian Yu
Eduard Gabriel Bazavan	Iro Armeni	Konrad Schindler	Miaomiao Liu	Qianru Sun
Eduard Trulls	Ishan Misra	Konstantinos G. Derpanis	Michael Felsberg	Qifeng Chen
Efstratios Gavves	Ivan Laptev	Kostas Daniilidis	Michael Maire	Qilong Wang
Ehsan Adeli	Jaesik Choi	Kris M. Kitani	Michael Niemeyer	Qin Jin
Ehsan Elhamifar	Jaesik Park	Krishna Kumar Singh	Michael Rubinstein	Qixing Huang
Elisa Ricci	Jakob Verbeek	Krishna Murthy	Michael Wray	R Alex Colburn
Enric Corona	James Hays	Jatavallabhula	Mike Zheng Shou	Radu Timofte
Eric Brachmann	James Matthew Rehg	Kristin Dana	Min H. Kim	Rahaf Aljundi
Eshed Ohn-Bar	Jan C. van Gemert	Krystian Mikolajczyk	Min Sun	Ram Nevatia
Evan Shelhamer	Jan Eric Lenssen	Kwang In Kim	Minghua Liu	Rana Hanocka
Evangelos Kalogerakis	Jana Kosecka	Kwang Moo Yi	Mingkui Tan	Ranjay Krishna
Fabio Cuzzolin	Jason J. Corso	Kyros Kutulakos	Mingze Xu	Raoul de Charette
Fabio Galasso	Jasper Uijlings	Lam M. Nguyen	Minh Ha Quang	Ravi Ramamoorthi
Fartash Faghri	Javier Vazquez-Corral	Lamberto Ballan	Minh Hoai	Raymond A. Yeh
Federica Bogo	Jean-Francois Lalonde	Laszlo Attila Jeni	Minjie Cai	Rei Kawakami
Federico Tombari	Jean-Marc Odobez	Le Lu	Misha Andriluka	Renaud Marlet
Feng Liu	Jeany Son	Lei Wang	Mohit Gupta	Reza Sabzevari
Feng Lu	Jerome Revaud	Lei Zhang	Mohsen Ali	Richard Newcombe
Francesca Odone	Jia Deng	Lei Zhu	Muhammad Haris Khan	Richard Zhang
Francis Engelmann	Jia-Bin Huang	Leonid Sigal	Nalini K. Ratha	Rita Cucchiara
Fredrik Kahl	Jiachen Li	Li Erran Li	Nathan Jacobs	Robby T. Tan
Friedrich Fraundorfer	Jiahuan Zhou	Li Yi	Naveed Akhtar	Rogério Feris
Gang Hua	Jiajun Wu	Li Zhang	Nianyi Li	Rohit Girdhar
Gang Yu	Jian Sun	Liang Zheng	Nicolas Thome	Roman Pflugfelder
Gedas Bertasius	Jian Wang	Liang-Chieh Chen	Nicoletta Noceti	Ronen Basri
Georgios Pavlakos	Jianfei Cai	Liangyan Gui	Nicu Sebe	Roni Sengupta
Georgios Tzimiropoulos	Jiangxin Dong	Lianli Gao	Niki Martinel	Roozbeh Mottaghi
Gim Hee Lee	Jiankang Deng	Limin Wang	Nikita Arslanov	Rui Chen
Giovanni Maria Farinella	Jianlong Wu	Lin Gao	Nikos Kolotouros	Ryan Farrell
Gordon Wetzstein	Jianwen Xie	Linchao Bao	Niloy Mitra	Ryoma Bise
Greg Mori	Jianxin Wu	Linchao Zhu	Ning Yu	S. Kevin Zhou
Greg Shakhnarovich	Jiawei Ma	Lingjie Liu	Ning Zhang	Sagie Benaim
Guanbin Li	Jiaya Jia	Lingni Ma	Noah Snively	Sai-Kit Yeung
Guansong Pang	Jiaying Liu	Lingxi Xie	Norimichi Ukita	Saining Xie
Guillermo Gallego	Jimei Yang	Linjie Yang	Nuno Vasconcelos	Salman Khan
Gunhee Kim	Jingya Wang	Liwei Wang	Octavia Camps	Sangdoo Yun
Guosheng Lin	Jinwei Gu	Loic Landrieu	Oisín Mac Aodha	Sanjeev Jagannatha
Guoyu Lu	Jiri Matas	Long Chen	Olga Russakovsky	Koppal
Hadi Pouransari	Joao Carreira	Longyin Wen	Olga Veksler	Sara Beery
Hae-Gon Jeon	Joao F. Henriques	Lourdes Agapito	Oncel Tuzel	Sareh Rowlands
Haibin Ling	Joao P. Barreto	Lu Jiang	Ondrej Chum	Sathyanarayanan N. Aakur

Satoshi Tsutsui	Stefan Lee	Venkatesh Babu	Xin Yu	Yonglong Tian
Saurabh Gupta	Stefano Mattoccia	Radhakrishnan	Xinchao Wang	Yongqin Xian
Saurabh Singh	Stella X. Yu	Vicente Ordóñez	Xingang Pan	Yosi Keller
Sayna Ebrahimi	Stephan Alaniz	Vicky Kalogeiton	Xinggang Wang	Young Min Kim
Scott McCloskey	Stephan Richter	Victor Prisacariu	Xinlei Chen	Yu Cheng
Scott Workman	Stephane Lathuilière	Vikram V. Ramaswamy	Xiu-Shen Wei	Yu Kong
Sebastiano Vascon	Stephen Gould	Viktor Larsson	Xucong Zhang	Yu Li
Seon Joo Kim	Stephen James	Viktoriia Sharmanska	Xuefei Ning	Yu Wu
Seong Joon Oh	Stephen Lin	Vinay P. Namboodiri	Xueting Li	Yu-Chiang Frank Wang
Serge Belongie	Subhankar Roy	Vincent Sitzmann	Xuming He	Yu-Chuan Su
Sergey Prokudin	Subhashini Venugopalan	Vineeth N. N	Yadan Luo	Yu-Wei Chao
Sergey Tulyakov	Subhansu Maji	Balasubramanian	Yadong Mu	Yu-Xiong Wang
Seunghoon Hong	Suha Kwak	Vishal M. Patel	Yale Song	Yuchao Dai
Seungryong Kim	Sung Ju Hwang	Vittorio Murino	Yan Huang	Yue Wang
Seungyong Lee	Supasorn Suwajanakorn	Vlad I. Morariu	Yan Yan	Yueqi Duan
Shai Bagon	Sylvain Paris	Vladimir Kim	Yan Zhang	Yuki M. Asano
Shalini De Mello	Tae Hyun Kim	Vladimir Pavlovic	Yanchao Yang	Yulan Guo
Shang-Hong Lai	Tae-Hyun Oh	Vladislav Golyanik	Yang Wang	Yulun Zhang
Shangzhe Wu	Taesup Moon	Vladislav Golyanik	Yansong Tang	Yumin Suh
Shanshan Zhang	Takayuki Okatani	Wangmeng Zuo	Yao Qin	Yung-Yu Chuang
Shaodi You	Tammy Riklin Raviv	Waqas Sultani	Yaoyao Liu	Yunhui Guo
Shaohui Lin	Tat-Jen Cham	Wei Yang	Yapeng Tian	Yunzhu Li
Shayok Chakraborty	Tat-Jun Chin	Wei-Chih Hung	Yasutaka Furukawa	Yuri Boykov
Shengcai Liao	Tatiana Tommasi	Wei-Chiu Ma	Yasuyuki Matsushita	Yusuke Sugano
Shenghua Gao	Tatsuya Harada	Wei-Shi Zheng	Ye Yuan	Yuyin Zhou
Shenlong Wang	Theo Gevers	Weihong Deng	Yebin Liu	Zan Gojcic
Shiguang Shan	Theodora Kontogianni	Wen Li	Yedid Hoshen	Zexiang Xu
Shijian Lu	Thibault Groueix	Wenhai Wang	Yen-Yu Lin	Zhangyang Wang
Shin'ichi Satoh	Thibaut Durand	Wenqi Ren	Yezhou Yang	Zhaopeng Cui
Shiry Ginosar	Thiemo Alldieck	Wenzheng Chen	Yi Fang	Zhaowei Cai
Shizhe Chen	Thomas Kipf	William A. P. Smith	Yi Wu	Zhaoxiang Zhang
Shohei Nobuhara	Tian Han	William Robson Schwartz	Yi Yang	Zhe Gan
Shubham Tulsiani	Tianfan Xue	Xavier Giro-i-Nieto	Yi-Hsuan Tsai	Zhe Lin
Shuicheng YAN	Tianzhu Zhang	Xi Peng	Yi-Ting Chen	Zheng Zhang
Shunsuke Saito	Ting Yao	Xi Peng	Yibing Song	Zhengming Ding
Shuo Chen	Ting Zhang	Xi Yin	Yifan Wang	Zhengqi Li
Siavash Bigdeli	Ting-Chun Wang	Xiang Bai	Yifei Huang	Zhirong Wu
Sicheng Zhao	Tobias Ritschel	Xiangyu Xu	Yijun Li	Zhizhong Han
Sifei Liu	Tolga Birdal	Xiangyu Yue	Yiming Wang	Zhuang Liu
Silvia Cascianelli	Tomas Hodan	Xiaoguang Han	Yin Li	Zhun Zhong
Silvia Zuffi	Tomas Pajdla	Xiaojuan Qi	Ying Fu	Ziad Al-Halah
Simon Niklaus	Tomasz Trzcinski	Xiaojun Chang	Ying Wu	Zicheng Liu
Simone Calderara	Tony Tung	Xiaolin Huang	Yinghuan Shi	Ziwei Liu
Simone Schaub-Meyer	Torsten Sattler	Xiaolong Wang	Yingli Tian	Zongwei Zhou
Sinisa Todorovic	Toshihiko Yamasaki	Xiaoming Liu	Yinqiang Zheng	Zsolt Kira
Siwei Wang	Tushar Nagarajan	Xiaoqian Wang	Yisen Wang	Zuxuan Wu
Slobodan Ilic	Ulugbek S. Kamilov	Xiaoshuai Zhang	Yiyi Liao	Zuzana Kukelova
Song Bai	Umar Iqbal	Xiaowei Zhou	Yizhou Yu	
Spyros Gidaris	Unnat Jain	Xiaoyu Wang	Yogesh S. Rawat	
Srinath Sridhar	Varun Jampani	Xihui Liu	Yoichi Sato	
Srinivasa Narasimhan	Vasileios Belagiannis	Xilin Chen	Yong Jae Lee	

CVPR 2025 OUTSTANDING REVIEWERS

We are grateful to all of the 9,872 reviewers who helped make CVPR 2025 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.

Aaron Hertzmann	Akhil Perincherry	Ameya Prabhu	Arif Mahmood	Bingkun Bao
Abdelrahman Eldesokey	Akisato Kimura	Amine Bourki	Aritra Bhowmik	Björn Michele
Abdelwahed Khamis	Akshat Dave	Amrita Mazumdar	Arjun Karpur	Bolin Lai
Abhishek Badki	Alessio Sampieri	Ana Cristina Murillo	Armin Hadzic	Boris Meinardus
Abhishek Jha	Alessio Tonioni	Anastasios Roussos	Aron Monszpart	Boseung Jeong
Abir Das	Alex Costanzino	Andreas Bär	Auke J. Wiggers	Bowen Cai
Aditya Agarwal	Alex Hanson	Andrew Zhenhao Hou	Avinash Paliwal	Bowen Yin
Aditya Arun	Alexander Richard	Andrew Zisserman	Axel Barroso-Laguna	Boxiao Pan
Aditya Chinchure	Alexandre Boulch	Aniruddh Sikdar	Ayça Takmaz	Brais Martinez
Aditya Ganeshan	Alexandros Delitzas	Anjan Dutta	Barbara Caputo	Brian Chen
Aditya Prakash	Alexandru Condurache	Ankit Dhiman	Bartłomiej Twardowski	Byeonghu Na
Adrian Penate-Sanchez	Ali Athar	Antoine Guédon	Benedetta Liberatori	Byung Cheol Song
Ahmed A. A. Osman	Alina Kuznetsova	Anubhav Anubhav	Bharadwaj Ravichandran	Carlo Masone
Ahmed Taha	Amélie Royer	Anup Kumar Gupta	Bin Yang	Carlos Esteves

Carlos Rodríguez - Pardo	Elena Garces	Hannah Kerner	Juan Miguel Valverde	Mahmoud Afifi
Carolina Raposo	Elliot Vincent	Haodong Lu	Juan Rodriguez	Mannat Singh
Ce Zheng	Ellis Langham Brown	Haoming Cai	Juanwu Lu	Manuel Kaufmann
Chaerin Min	Emad Bahrami	Haoyang	Julian Kooij	Manuel Mucientes
Changjiang Cai	Emanuele Vivoli	Haoyu Wu	Julian Straub	Marcel Buehler
Chao Ren	Enguang Wang	Hareesh Ravi	Jun Dan	Marcella Cornia
Chao Wen	Enis Simsar	Heeseung Yun	Jun Hao Liew	Marco Cannici
Charles Saunders	Enzo Tartaglione	Henrique Morimitsu	Jun Liu	Marcus Klasson
Che-Tsung Lin	Eric Slyman	Henry Howard-Jenkins	Jun Xiao	Marie-Julie Rakotosaona
Chen Guo	Ernest Valveny	Hermann Blum	Jun Yue	Mark Coates
Chen Zhao	Etienne Vouga	Hidir Yesiltepe	Junfan Lin	Markus Worchel
Cheng Peng	Evangelos Kazakos	Hoin Jung	Jung Uk Kim	Martin Engilberge
Cheng-Kuan Lin	Evin Pinar Örnek	Hongjie Wang	Jungho Lee	Martin Nicolas Everaert
Chengan He	Ewelina Rupnik	Hoyoung Kim	Junho Kim	Matias Alejandro
Chengcheng Tang	Fabian David Caba	Hritam Basak	Junhyeong Cho	Valdenegro-Toro
Chengxuan Zhu	Heilbron	Huaian Chen	Junjie Ye	Matteo Dunnhofer
Chenhui Gou	Fabio Poiesi	Hui Yuan	Junoh Lee	Matteo Farina
Chenyang Wang	Fabio Quattrini	Hyeongjin Nam	Junyu Xie	Matteo Mosconi
Chi Xu	Fabio Tos	Hyeongseok Son	Kai Fong Ernest Chong	Matteo Poggi
Chiao An Yang	Fahad Shamshad	Hyeontae Son	Karan Desai	Matthew Fisher
Chong Mou	Fangyin Wei	Hyoung-Kyu Song	Karsten Roth	Matthew J. Muckley
Chris Careaga	Favyen Bastani	Hyunjun Jung	Karthik Nandakumar	Matthew Joseph Leotta
Christian Micheloni	Federico Bolelli	Idil Sülo	Katrin Renz	Matthieu Guillaumin
Christoph Reich	Federico Perazzi	Ignacio Alzugaray	Kavisha Vidanapathirana	Meina Kan
Christopher Wewer	Fei Xue	In Cho	Ke Xu	Mengping Yang
Chun-Hao Paul Huang	Felipe C. Chamone	Ioannis Kakogeorgiou	Kefan Chen	Michael Firman
Cigdem Beyan	Felipe Gutierrez-Barragan	Ioannis Siglidis	Kelsey R Allen	Michael S Brown
Clément Rambour	Felix Sebastian Fent	Ishan Khatri	Kent Gauen	Michel Antunes
Codruta Ancuti	Felix Wimbauer	Istvan Sarandi	Kevin J Liang	Minesh Mathew
Cong Wang	Ferda Ofli	Itai Lang	Kim Jun-Seong	Ming-Yang Ho
Constantin Patsch	Filippo Ziliotto	Ivan Lopes	Kimin Yun	Minghan Li
Corentin Dumery	Florent Lafarge	Jacob Chalk	Kohei Yamashita	Minghui Hu
Corentin Sautier	Florin-Alexandru	Jaeha Kim	Konpat Ta Preechakul	Mingming He
Cristiano Saltori	Vasluianu	Jaekoo Lee	Konstantinos Batsos	Mingxuan Liu
Daan De Geus	Francesco Giuliari	Jaeseok Byun	Konstantinos Panagiotis	Minhyeok Lee
Daechul Ahn	Francesco Taioli	Jamie Watson	Alexandridis	Minhyun Lee
Damian Borth	Franck Mamalet	Jan P Klopp	Konstantinos Vougioukas	Miroslav Purkrábek
Damien Teney	Frederick W. B. Li	Jan Zdenek	Kota Yamaguchi	Mitch Hill
Daniel Ritchie	Fu-Yun Wang	Jason Rambach	Kumail Alhamoud	Mohamed Sayed
Daniyar Turmukhambetov	Gabriele Goleto	Javier Civera	Kunal Swami	Mohammad Fahes
Danna Xue	Gabriele Trivigno	Je Hyeong Hong	Kunyu Peng	Moreno D'incà
Dario Serez	Gang Wu	Jenny Schmalfuss	Kuo-Hao Zeng	Mosam Dabhi
David Alexander Ross	Ge-Peng Ji	Jenny Seidenschwarz	Kwot Sin Lee	Muhammad Akhtar Munir
David Bourgin	Gemma Canet Tarrés	Jeong Gi Kwak	Kyungsu Lee	Munchurl Kim
David Chan	George Retsinas	Jeongseok Hyun	Le Zhang	Mustafa Shukor
David Charatan	George Toderici	Jia-Fong Yeh	Leander Girrback	Mutian Xu
David Hurych	Georgios Kopanas	Jiahao Xia	Lei Yang	Muyao Niu
David Schneider	Geri Skenderi	Jiahe Li	Lennart Bastian	Naoto Inoue
Davide Boscaini	Gianpaolo Bontempo	Jian Zheng	Leo Fillioux	Narendra Ahuja
Davide Caffagni	Gijds Dubbelman	Jiang-Jiang Liu	Leo Milecki	Nathan Piasco
Davis Rempe	Gilles Puy	Jianing Xi	Li Liu	Nermin Samet
Debaditya Roy	Giorgos Kordopatis-Zilos	Jianping Gou	Lin Geng Foo	Nesryne Mejri
Debi Dogra	Giorgos Tolas	Jianqing Zheng	Lin Sihao	Nhat-Tan Bui
Dhruv Srivastava	Giuseppe Lisanti	Jiequan Cui	Lingchen Sun	Nicholas Kolkin
Di Ming	Gnana Praveen	Jiho Choi	Lingdong Kong	Nicholas Moratelli
Dibyadip Chatterjee	Rajasekhar	Jin Zhang	Lingfeng Yang	Nicolas Audebert
Diego Martín Arroyo	Görkay Aydemir	Jing Wu	Linrui Gong	Nicolas Aziere
Digbalay Bose	Gregory Holste	Jingkang Yang	Linus Härenstam-Nielsen	Nicolas Gonthier
Dimitrios Kollias	Gregory Slabaugh	Jinglei Shi	Liujuan Cao	Nikhil Varma Keetha
Dingkang Liang	Guénolé Fiche	Jingyong Su	Liyuan Zhu	Niklas Hanselmann
Dipesh Tamboli	Guido Maria D'amely Di	Jinnan Chen	Lluis Castrejon	Nikola Popovic
Dmytro Mishkin	Melendugno	Jinpeng Liu	Lorenzo Baraldi	Nikolai Kalischek
Dominik Schnaus	Guilherme Potje	Jinpeng Wang	Lorenzo Bonicelli	Nikos Deligiannis
Donald Gilbert Dansereau	Guillaume Jeanneret	Jinseok Bae	Loris Bazzani	Nima Kalantari
Donald Shenaj	Guillermo Gomez-Trenado	Joachim Denzler	Luca Barsellotti	Nina Shvetsova
Dongkeun Oh	Gukyeong Kwon	John Richard Corring	Luca Bartolomei	Nithin C Babu
Dongyoon Wee	Guofeng Mei	Jonathan Prexl	Luca Zanella	Noa Garcia
Dongyoung Kim	Guoyuan An	Jordan Vice	Luigi Capogrosso	Octave Mariotti
Dor Verbin	Gwanhyeong Koo	Jose Dolz	Lukas Picek	Olga Zatsarynna
Dror Moran	Ha Min Son	Jose Oramas	Luxi Zhao	Oliver Hahn
Duo Zheng	Haicheng Wang	Josep Lladós	M. Saquib Sarfraz	Osman Ülger
Dylan Campbell	Hallee E Wong	Josip Šarić	M. Usman Rafique	Ozgur Kara
Eduardo Pérez-Pellitero	Hang Yang	Joy Hu	Mahdi Rad	Paolo Rota

Partha Das	Sangheum Hwang	Tamás Karácsony	Yann Labbé
Paul Albert	Sanghyuk Chun	Tassilo Wald	Yanran Li
Paul Henderson	Sanghyun Jo	Ted Lentsch	Yanyang Li
Paul Steven Scotti	Sangmin Woo	Thibaud Ehret	Yanyang Yan
Paul Strelli	Sanjay Hareesh	Thomas De Min	Yao Du
Paul Swoboda	Sanjoy Kundu	Thomas Mensink	Yao-Chih Lee
Pavel Suma	Sara Vicente	Tianchen Zhao	Yaqing Ding
Pavlo Molchanov	Sarah Rastegar	Tiancheng Sun	Yaroslava Lochman
Pei Wang	Sascha Hornauer	Tianshu Huang	Yasser Benigmim
Peihua Li	Saumya Gupta	Tianshuo Zhang	Yi Rong
Peiran Xu	Sayan Nag	Tim Meinhardt	Yiannis Aloimonos
Peiwu Qin	Sebastian Koch	Tim Seizinger	Yichang Shih
Pengcheng Xu	Sergio Izquierdo	Timm Haucke	Yifan Zhou
Peter Hirsch	Sethuraman T V	Timothy Duff	Yikai Wang
Petros Koutras	Seung Wook Kim	Timothy Redgrave	Yimu Wang
Philip Chikontwe	Seunggyu Chang	Tingting Xie	Ying He
Philipp Lindenberg	Seungju Cho	Tingyu Qu	Ying Zhao
Philipp Schröppel	Seyed Mahdi Hosseini	Tom Monnier	Yipeng Qin
Pierluigi Zama Ramirez	Miangoleh	Tomas Jakab	Yiran Wang
Pierre-Luc St-Charles	Shahira Abousamra	Tomas Vojir	Yiwu Zhong
Piyush Nitin Bagad	Shakiba Kheradmand	Tommie Keressies	Yixuan Ren
Prachi Garg	Shao-Yen Tseng	Tsung-Wei Ke	Yizhi Wang
Praneeth Chakravarthula	Shao-Yuan Lo	Tuan Duc Ngo	Yizhou Zhao
Priyank Pathak	Shaoxiong Zhang	Tuo Feng	Yongjae Lee
Qi Bi	Sharath Girish	Valentin Deschaintre	Yongjun Zhang
Qian Dr. Zheng	Shashank Agnihotri	Valerio Marsocci	Young-Jae Park
Qian Yang	Shashank Tripathi	Vishaal Udandara	Youngtaek Oh
Qiang Hu	Shenghai Yuan	Vishal Asnani	Youwei Lyu
Qianli Ma	Shengze Wang	Vito Paolo Pastore	Youyi Zheng
Qihang Fang	Sherwin Bahmani	Vitomer Struc	Yu Wang
Qimin Chen	Shihao Zhou	Vitus Benson	Yu Zhou
Qing Yu	Shijia Huang	Vladan Stojnić	Yuanbo Xiangli
Qingdong He	Shikui Wei	Walter Scheirer	Yuanzhi Zhu
Qingping Zheng	Shiyuan Huang	Weibin Wu	Yue Gao
Qingsheng Wang	Shrisudhan Govindarajan	Weijian Deng	Yuhui Yuan
Quan Anh Cao	Shuai Li	Weilai Xiang	Yujiao Shi
Quentin Bouniot	Shuai Li	Weisheng Dong	Yujie Wang
Rabiul Awal	Shuhuai Ren	Weiwei Sun	Yunhao Ba
Rahul Goel	Shuzhe Wang	Weiwei Wu	Yunzhong Hou
Rahul Sajani	Shyamal Buch	Wenhui Zhou	Yuping Duan
Rainer Stiefelwagen	Sicheng Li	Wenjing Li	Yuqing Wang
Raza Yunus	Sicheng Mo	Wenke Huang	Yusuf Dalva
Riccardo Salami	Siddharth Tourani	Wenqi Ouyang	Yusuke Hirota
Richard M. Jiang	Sifeng Shang	Wenrui Dai	Yuxiao Chen
Rinon Gal	Silvia Laura Pintea	Willi Menapace	Zehong Shen
Rishabh Shukla	Simon Schrod	Wojciech Masarczyk	Zeynep Sonat Baltaci
Rishabh Parihar	Sinisa Stekovic	Wonjae Kim	Zhen Zhu
Ritwik Gupta	Siqi Wang	Wooju Lee	Zhi-Yi Chin
Roberto Bigazzi	Siyu Lin	Woongoh Cho	Zhifan Zhu
Roei Herzig	Siyuan Li	Xi Wang	Zhikang Zhang
Rohan Choudhury	Sizhuo Ma	Xiang Hao	Zhitong Gao
Roman Bachmann	Songan Zhang	Xiang Xiang	Zhixuan Liang
Romarc Audigier	Soo Ye Kim	Xiao-Ming Wu	Zhuang Qian
Ronghui Li	Srijan Das	Xiaohan Zhang	Zhulin An
Rongqing Li	Sriram Yenamandra	Xiaohan Zhang	Zhuo Huang
Rongyuan Wu	Stamatios Georgoulis	Xiaoming Zhao	Zihui Wu
Rui Song	Stanislaw Szymanowicz	Xiaoqiang Li	Zike Yan
Rui Tian	Stefano Gasperini	Xiaoshan Yang	Zilin Gao
Ruiping Wang	Stergios Christodoulidis	Xiaoyu Kong	Zinan Lin
Ruiyang Liu	Steven Walton	Xin Ma	Ziyao Zeng
Rujie Wu	Stuart James	Xingqian Xu	Ziyi Wu
Ruozhen He	Suchen Wang	Xingrui Wang	Ziyun Wang
Ruxiao Duan	Sudarshan Sunil Harithas	Xinyu Lyu	Zuheng Ming
Ryosuke Furuta	Sungha Choi	Xinyu Zhang	Zunnan Xu
Sagnik Das	Sunghoon Im	Xinyue Wei	
Sagnik Majumder	Sunghyun Cho	Xiyao Liu	
Sai Kancheti Srinivas	Sunyeon Kim	Xuhang Chen	
Sai Kumar Dwivedi	Surbhi Mittal	Yael Vinker	
Sai Saketh Rambhatla	Symeon Papadopoulos	Yang An	
Saikiran Kumar Tedla	Taeckyoung Lee	Yang Liu	
Sammy Joe Christen	Tai-Yu Pan	Yang Zhou	
Samuel Albanie	Taishi Ono	Yangzhong Zhou	
Sandeep Mishra	Takafumi Taketomi	Yangyi Huang	

Friday, June 13

- 7:00 - 18:00 Registration / Badge Pickup** (Summit Lobby)
7:00 - 18:00 Registration / Badge Pickup (ExHall Concourse)
7:00 - 17:00 Press Room (203 B)
7:00 - 17:00 Mother's Room (Level 1 near Room 101 and on Level 3 near Exhibit Hall D)
7:00 - 17:00 Prayer or Quiet Room (203 A)
7:00 - 9:00 Breakfast (ExHall C)
8:00 - 8:30 Poster Setup (ExHall D)
8:30 - 9:00 Welcome & Awards (Karl Dean Ballroom)

9:00 - 10:15 Oral Session 1A: Image and Video Synthesis (Karl Dean Ballroom)

- 🏆 - Award candidate paper
- Motion Prompting: Controlling Video Generation with Motion Trajectories, *Daniel Geng, Charles Herrmann, Junhwa Hur, Forrester Cole, Serena Zhang, Tobias Pfaff, Tatiana Lopez-Guevara, Yusuf Aytar, Michael Rubinstein, Chen Sun, Oliver Wang, Andrew Owens, Deqing Sun*
 - Go-with-the-Flow: Motion-Controllable Video Diffusion Models Using Real-Time Warped Noise, *Ryan Burgert, Yuancheng Xu, Wenqi Xian, Oliver Pilarski, Pascal Clausen, Mingming He, Li Ma, Yitong Deng, Lingxiao Li, Mohsen Mousavi, Michael Ryoo, Paul Debevec, Ning Yu*
 - LookingGlass: Generative Anamorphoses via Laplacian Pyramid Warping, *Pascal Chang, Sergio Sancho, Jingwei Tang, Markus Gross, Vinicius Azevedo*
 - Alias-Free Latent Diffusion Models: Improving Fractional Shift Equivariance of Diffusion Latent Space, *Yifan Zhou, Zeqi Xiao, Shuai Yang, Xingang Pan*
 - RandAR: Decoder-only Autoregressive Visual Generation in Random Orders, *Ziqi Pang, Tianyuan Zhang, Fujun Luan, Yunze Man, Hao Tan, Kai Zhang, William T. Freeman, Yu-Xiong Wang*

9:00 - 10:15 Oral Session 1B: Interpretability and Evaluation (ExHall A2)

- OpenING: A Comprehensive Benchmark for Judging Open-ended Interleaved Image-Text Generation, *Pengfei Zhou, Xiaopeng Peng, Jiajun Song, Chuanhao Li, Zhaopan Xu, Yue Yang, Ziyao Guo, Hao Zhang, Yuqi Lin, Yefei He, Lirui Zhao, Shuo Liu, Tianhua Li, Yuxuan Xie, Xiaojun Chang, Yu Qiao, Wenqi Shao, Kaipeng Zhang*
- LibraGrad: Balancing Gradient Flow for Universally Better Vision Transformer Attributions, *Faridoun Mehri, Mahdieh Soleymani Baghshah, Mohammad Taher Pilehvar*
- Do We Always Need the Simplicity Bias? Looking for Optimal Inductive Biases in the Wild, *Damien Teney, Liangze Jiang, Florin Gogianu, Ehsan Abbasnejad*
- Molmo and PixMo: Open Weights and Open Data for State-of-the-Art Vision-Language Models, *Matt Deitke, Christopher Clark, Sangho Lee, Rohun Tripathi, Yue Yang, Jae Sung Park, Mohammadreza Salehi, Niklas Muennighoff, Kyle Lo, Luca Soldaini, Jiasen Lu, Taira Anderson, Erin Bransom, Kiana Ehsani, Huong Ngo, YenSung Chen, Ajay Patel, Mark Yatskar, Chris Callison-Burch, Andrew Head, Rose Hendrix, Favyen Bastani, Eli VanderBilt, Nathan Lambert, Yvonne Chou, Arnavi Chheda, Jenna Sparks, Sam Skjonsberg, Michael Schmitz, Aaron Sarnat, Byron Bischoff, Pete Walsh, Chris Newell, Piper Wolters, Tanmay Gupta, Kuo-Hao Zeng, Jon Borchardt, Dirk Groeneveld, Crystal Nam, Sophie Lebrecht, Caitlin Wittliff, Carissa Schoenick, Oscar Michel, Ranjay Krishna, Luca Weihs, Noah A. Smith, Hannaneh Hajishirzi, Ross Girshick, Ali Farhadi, Aniruddha Kembhavi*
- Rethinking Vision-Language Model in Face Forensics: Multi-Modal Interpretable Forged Face Detector, *Xiao Guo, Xiufeng Song, Yue Zhang, Xiaohong Liu, Xiaoming Liu*

9:00 - 10:15 Oral Session 1C: Image Processing and Deep Architectures (Davidson Ballroom)

- CleanDIFT: Diffusion Features without Noise, *Nick Stracke, Stefan Andreas Baumann, Kolja Bauer, Frank Fundel, Björn Ommer*

- OverLoCK: An Overview-first-Look-Closely-next ConvNet with Context-Mixing Dynamic Kernels, *Meng Lou, Yizhou Yu*
- Towards Explicit Geometry-Reflectance Collaboration for Generalized LiDAR Segmentation in Adverse Weather, *Longyu Yang, Ping Hu, Shangbo Yuan, Lu Zhang, Jun Liu, Hengtao Shen, Xiaofeng Zhu*
- DiffFNO: Diffusion Fourier Neural Operator, *Xiaoyi Liu, Hao Tang*
- Removing Reflections from RAW Photos, *Eric Kee, Adam Pikielny, Kevin Blackburn-Matzen, Marc Levoy*

- 10:00 - 10:30 Poster Setup** (ExHall D)
10:00 - 11:00 Coffee Break (ExHall D)

10:30 - 12:30 Poster Session 1 and Exhibit Hall (ExHall D)

- * - Highlight paper 🏆 - Award candidate paper
 ○ - Oral Paper ☆ - Outstanding Reviewer
- Prosody-Enhanced Acoustic Pre-training and Acoustic-Disentangled Prosody Adapting for Movie Dubbing, *Zhedong Zhang, Liang Li, Chenggang Yan, Chunshan Liu, Anton van den Hengel, Yuankai Qi*
 - Wav2Sem: Plug-and-Play Audio Semantic Decoupling for 3D Speech-Driven Facial Animation, *Hao Li, Ju Dai, Xin Zhao, Feng Zhou, Junjun Pan, Lei Li*
 - Sonic: Shifting Focus to Global Audio Perception in Portrait Animation, *Xiaozhong Ji, Xiaobin Hu, Zhihong Xu, Junwei Zhu, Chuming Lin, Qingdong He, Jiangning Zhang, Donghao Luo, Yi Chen, Qin Lin, Qinglin Lu, Chengjie Wang*
 - Towards High-fidelity 3D Talking Avatar with Personalized Dynamic Texture, *Xuanchen Li, Jianyu Wang, Yuhao Cheng, Yikun Zeng, Xingyu Ren, Wenhan Zhu, Weiming Zhao, Yichao Yan*
 - Electromyography-Informed Facial Expression Reconstruction for Physiological-Based Synthesis and Analysis, *Tim Büchner, Christoph Anders, Orlando Guntinas-Lichius, Joachim Denzler*
 - High-Fidelity Relightable Monocular Portrait Animation with Lighting-Controllable Video Diffusion Model, *Mingtao Guo, Guanyu Xing, Yanli Liu*
 - Quaffure: Real-Time Quasi-Static Neural Hair Simulation, *Tuur Stuyck, Gene Wei-Chin Lin, Egor Larionov, Hsiao-yu Chen, Aljaz Bozic, Nikolaos Sarafianos, Doug Roble*
 - GPAvatar: High-fidelity Head Avatars by Learning Efficient Gaussian Projections, *Wei-Qi Feng, Dong Han, Ze-Kang Zhou, Shunkai Li, Xiaoqiang Liu, Pengfei Wan, Di Zhang, Miao Wang*
 - HERA: Hybrid Explicit Representation for Ultra-Realistic Head Avatars, *Hongrui Cai, Yuting Xiao, Xuan Wang, Jiafei Li, Yudong Guo, Yanbo Fan, Shenghua Gao, Juyong Zhang*
 - GASP: Gaussian Avatars with Synthetic Priors, *Jack Saunders, Charlie Hewitt, Yanan Jian, Marek Kowalski, Tadas Baltrusaitis, Yiye Chen, Darren Cosker, Virginia Estellers, Nicholas Gydé, Vinay P. Namboodiri, Benjamin E. Lundell*
 - FRESA: Feedforward Reconstruction of Personalized Skinned Avatars from Few Images, *Rong Wang, Fabian Prada, Ziyang Wang, Zhongshi Jiang, Chengxiang Yin, Junxuan Li, Shunsuke Saito, Igor Santesteban, Javier Romero, Rohan Joshi, Hongdong Li, Jason Saragih, Yaser Sheikh*
 - DAGSM: Disentangled Avatar Generation with GS-enhanced Mesh, *Jingyu Zhuang, Di Kang, Linchao Bao, Liang Lin, Guanbin Li*
 - HumanRig: Learning Automatic Rigging for Humanoid Character in a Large Scale Dataset, *Zedong Chu, Feng Xiong, Meiduo Liu, Jinzhi Zhang, Mingqi Shao, Zhaoxu Sun, Di Wang, Mu Xu*
 - SKDream: Controllable Multi-view and 3D Generation with Arbitrary Skeletons, *Yuanyou Xu, Zongxin Yang, Yi Yang*
 - FreeUV: Ground-Truth-Free Realistic Facial UV Texture Recovery via Cross-Assembly Inference Strategy, *Xingchao Yang, Takafumi Taketomi, Yuki Endo, Yoshihiro Kanamori*
 - MultiGO: Towards Multi-level Geometry Learning for Monocular 3D Textured Human Reconstruction, *Gangjian Zhang, Nanjie Yao, Shunsi Zhang, Hanfeng Zhao, Guoliang Pang, Jian Shu, Hao Wang*
 - GaussianIP: Identity-Preserving Realistic 3D Human Generation via Human-Centric Diffusion Prior, *Zichen Tang, Yuan Yao, Miaomiao Cui, Liefeng Bo, Hongyu Yang*

- 18 Shining Yourself: High-Fidelity Ornaments Virtual Try-on with Diffusion Model, *Yingmao Miao, Zhanpeng Huang, Rui Han, Zibin Wang, Chenhao Lin, Chao Shen*
- 19 SynthLight: Portrait Relighting with Diffusion Model by Learning to Re-render Synthetic Faces, *Sumit Chaturvedi, Mengwei Ren, Yannick Hold-Geoffroy, Jingyuan Liu, Julie Dorsey, Zhixin Shu*
- 20 Comprehensive Relighting: Generalizable and Consistent Monocular Human Relighting and Harmonization, *Junying Wang, Jingyuan Liu, Xin Sun, Krishna Kumar Singh, Zhixin Shu, He Zhang, Jimei Yang, Nanxuan Zhao, Tuanfeng Y. Wang, Simon S. Chen, Ulrich Neumann, Jae Shin Yoon*
- 21 Polarized Color Screen Matting, *Kenji Enomoto, Scott Cohen, * Brian Price, TJ Rhodes*
- 22 SLVR: Super-Light Visual Reconstruction via Blueprint Controllable Convolutions and Exploring Feature Diversity Representation, *Ning Ni, Libao Zhang*
- 23 Proximal Algorithm Unrolling: Flexible and Efficient Reconstruction Networks for Single-Pixel Imaging, *Ping Wang, Lishun Wang, Gang Qu, Xiaodong Wang, Yulun Zhang, Xin Yuan*
- 24 Glossy Object Reconstruction with Cost-effective Polarized Acquisition, *Bojian Wu, Yifan Peng, Ruizhen Hu, Xiaowei Zhou*
- 25 Blurry-Edges: Photon-Limited Depth Estimation from Defocused Boundaries, *Wei Xu, Charles James Wagner, Junjie Luo, Qi Guo*
- 26 LumiNet: Latent Intrinsic Meets Diffusion Models for Indoor Scene Relighting, *Xiaoyan Xing, Konrad Groh, Sezer Karaoglu, Theo Gevers, Anand Bhattad*
- 27 LEDiff: Latent Exposure Diffusion for HDR Generation, *Chao Wang, Zhihao Xia, Thomas Leimkuhler, Karol Myszkowski, Xuaner Zhang*
- 28 IRIS: Inverse Rendering of Indoor Scenes from Low Dynamic Range Images, *Chih-Hao Lin, Jia-Bin Huang, Zhengqin Li, Zhao Dong, Christian Richardt, Tuotuo Li, Michael Zollhöfer, Johannes Kopf, Shenlong Wang, Changil Kim*
- 29 Differentiable Inverse Rendering with Interpretable Basis BRDFs, *Hoon-Gyu Chung, Seokjun Choi, Seung-Hwan Baek*
- 30 Hardware-Rasterized Ray-Based Gaussian Splatting, *Samuel Rota Bulò, * Nemanja Bartolovic, Lorenzo Porzi, Peter Kotschieder*
- 31 TensoFlow: Tensorial Flow-based Sampler for Inverse Rendering, *Chun Gu, Xiaofei Wei, Li Zhang, Xiatian Zhu*
- 32 LIRM: Large Inverse Rendering Model for Progressive Reconstruction of Shape, Materials and View-dependent Radiance Fields, *Zhengqin Li, Dilin Wang, Ka Chen, Zhaoyang Lv, Thu Nguyen-Phuoc, Milim Lee, Jia-Bin Huang, Lei Xiao, Yufeng Zhu, Carl S. Marshall, Yuheng Ren, Richard Newcombe, Zhao Dong*
- 33 Gaussian Splashing: Unified Particles for Versatile Motion Synthesis and Rendering, *Yutao Feng, Xiang Feng, Yintong Shang, Ying Jiang, Chang Yu, Zeshun Zong, Tianjia Shao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, Yin Yang*
- 34 Accurate Differential Operators for Hybrid Neural Fields, *Aditya Chetan, Guandao Yang, Zichen Wang, Steve Marschner, Bharath Hariharan*
- 35 Learning Extremely High Density Crowds as Active Matters, *Feixiang He, Jiangbei Yue, Jialin Zhu, Armin Seyfried, Dan Casas, Julien Pettré, He Wang*
- 36 TexGaussian: Generating High-quality PBR Material via Octree-based 3D Gaussian Splatting, *Bojun Xiong, Jialun Liu, Jiakui Hu, Chenming Wu, Jinbo Wu, Xing Liu, Chen Zhao, Errui Ding, Zhouhui Lian*
- 37 Real-time Free-view Human Rendering from Sparse-view RGB Videos using Double Unprojected Textures, *Guoxing Sun, * Rishabh Dabral, Heming Zhu, Pascal Fua, Christian Theobalt, Marc Habermann*
- 38 RoomPainter: View-Integrated Diffusion for Consistent Indoor Scene Texturing, *Zhipeng Huang, Wangbo Yu, Xinhua Cheng, Chengshu Zhao, Yunyang Ge, Mingyi Guo, Li Yuan, Yonghong Tian*
- 39 MVPaint: Synchronized Multi-View Diffusion for Painting Anything 3D, *Wei Cheng, Juncheng Mu, Xianfang Zeng, Xin Chen, Anqi Pang, Chi Zhang, Zhibin Wang, Bin Fu, Gang Yu, Ziwei Liu, Liang Pan*
- 40 Fancy123: One Image to High-Quality 3D Mesh Generation via Plug-and-Play Deformation, *Qiao Yu, Xianzhi Li, Yuan Tang, Xu Han, Long Hu, Yixue Hao, Min Chen*
- 41 ShapeShifter: 3D Variations Using Multiscale and Sparse Point-Voxel Diffusion, *Nissim Maruani, Wang Yifan, Matthew Fisher, Pierre Alliez, Mathieu Desbrun*
- 42 MeshArt: Generating Articulated Meshes with Structure-Guided Transformers, *Daoyi Gao, Yawar Siddiqui, Lei Li, Angela Dai*
- 43 SceneFactor: Factored Latent 3D Diffusion for Controllable 3D Scene Generation, *Aleksey Bokhovkin, Quan Meng, Shubham Tulsiani, Angela Dai*
- 44 PrEditor3D: Fast and Precise 3D Shape Editing, *Ziya Erkoç, Can Gümeli, Chaoyang Wang, Matthias Nießner, Angela Dai, Peter Wonka, Hsin-Ying Lee, Peiye Zhuang*
- 45 LT3SD: Latent Trees for 3D Scene Diffusion, *Quan Meng, Lei Li, Matthias Nießner, Angela Dai*
- 46 iSegMan: Interactive Segment-and-Manipulate 3D Gaussians, *Yian Zhao, Wanshi Xu, Ruochong Zheng, Pengchong Qiao, Chang Liu, Jie Chen*
- 47 LOD-GS: Achieving Levels of Detail using Scalable Gaussian Soup, *Jianxiang Shen, Yue Qian, Xiaohang Zhan*
- 48 MaskGaussian: Adaptive 3D Gaussian Representation from Probabilistic Masks, *Yifei Liu, Zhihang Zhong, Yifan Zhan, Sheng Xu, Xiao Sun*
- 49 NTR-Gaussian: Nighttime Dynamic Thermal Reconstruction with 4D Gaussian Splatting Based on Thermodynamics, *Kun Yang, Yuxiang Liu, Zeyu Cui, Yu Liu, Maojun Zhang, Shen Yan, Qing Wang*
- 50 DropoutGS: Dropping Out Gaussians for Better Sparse-view Rendering, *Yexing Xu, Longguang Wang, Minglin Chen, Sheng Ao, Li Li, Yulan Guo*
- 51 S2Gaussian: Sparse-View Super-Resolution 3D Gaussian Splatting, *Yecong Wan, Mingwen Shao, Yuanshuo Cheng, Wangmeng Zuo*
- 52 DeSplat: Decomposed Gaussian Splatting for Distractor-Free Rendering, *Yihao Wang, Marcus Klasson, Matias Turkulainen, * Shuzhe Wang, Juho Kannala, Arno Solin*
- 53 Neural Hierarchical Decomposition for Single Image Plant Modeling, *Zhihao Liu, Zhanglin Cheng, Naoto Yokoya*
- 54 Symmetry Strikes Back: From Single-Image Symmetry Detection to 3D Generation, *Xiang Li, Zixuan Huang, Anh Thai, James M. Rehg*
- 55 Digital Twin Catalog: A Large-Scale Photorealistic 3D Object Digital Twin Dataset, *Zhao Dong, Ka Chen, Zhaoyang Lv, Hong-Xing Yu, Yunzhi Zhang, Cheng Zhang, Yufeng Zhu, Stephen Tian, Zhengqin Li, Geordie Moffatt, Sean Christofferson, James Fort, Xiaqing Pan, Mingfei Yan, Jiajun Wu, Carl Yuheng Ren, Richard Newcombe*
- 56 Zero-Shot Novel View and Depth Synthesis with Multi-View Geometric Diffusion, *Vitor Guizilini, Muhammad Zubair Irshad, Dian Chen, Greg Shakhnarovich, Rares Ambrus*
- 57 NVComposer: Boosting Generative Novel View Synthesis with Multiple Sparse and Unposed Images, *Lingen Li, Zhaoyang Zhang, Yaowei Li, Jiale Xu, Wenbo Hu, Xiaoyu Li, Weihao Cheng, Jinwei Gu, Tianfan Xue, Ying Shan*
- 58 HybridGS: Decoupling Transients and Statics with 2D and 3D Gaussian Splatting, *Jingyu Lin, Jiaqi Gu, Lubin Fan, Bojian Wu, Yujing Lou, Renjie Chen, Ligang Liu, Jieping Ye*
- 59 Wonderland: Navigating 3D Scenes from a Single Image, *Hanwen Liang, Junli Cao, Vidit Goel, Guocheng Qian, Sergei Korolev, Demetri Terzopoulos, Konstantinos N. Plataniotis, Sergey Tulyakov, Jian Ren*
- 60 SpatialDreamer: Self-supervised Stereo Video Synthesis from Monocular Input, *Zhen Lv, Yangqi Long, Congzhentao Huang, Cao Li, Chengfei Lv, Hao Ren, Dian Zheng*
- 61 StreetCrafter: Street View Synthesis with Controllable Video Diffusion Models, *Yunzhi Yan, Zhen Xu, Haotong Lin, Haian Jin, Haoyu Guo, Yida Wang, Kun Zhan, Xianpeng Lang, Hujun Bao, Xiaowei Zhou, Sida Peng*
- 62 DroneSplat: 3D Gaussian Splatting for Robust 3D Reconstruction from In-the-Wild Drone Imagery, *Jiadong Tang, Yu Gao, * Dianyi Yang, Liqi Yan, Yufeng Yue, Yi Yang*
- 63 IndoorGS: Geometric Cues Guided Gaussian Splatting for Indoor Scene Reconstruction, *Cong Ruan, Yuesong Wang, Tao Guan, Bin Zhang, Lili Ju*
- 64 MAC-Ego3D: Multi-Agent Gaussian Consensus for Real-Time Collaborative Ego-Motion and Photorealistic 3D

- Reconstruction, *Xiaohao Xu, Feng Xue, Shibo Zhao, Yike Pan, Sebastian Scherer, Xiaonan Huang*
- 65 ShowMak3r: Compositional TV Show Reconstruction, *Sangmin Kim, Seungkuk Do, Jaesik Park*
- 66 4DGC: Rate-Aware 4D Gaussian Compression for Efficient
☆ Streamable Free-Viewpoint Video, *Qiang Hu, Zihan Zheng, Houqiang Zhong, Sihua Fu, Li Song, Xiaoyun Zhang, Guangtao Zhai, Yanfeng Wang*
- 67 HiMoR: Monocular Deformable Gaussian Reconstruction with Hierarchical Motion Representation, *Yiming Liang, Tianhan Xu, Yuta Kikuchi*
- 68 EnliveningGS: Active Locomotion of 3DGS, *Siyuan Shen, Tianjia Shao, Kun Zhou, Chenfanfu Jiang, Yin Yang*
- 69 HOP: Heterogeneous Topology-based Multimodal Entanglement for Co-Speech Gesture Generation, *Hongye Cheng, Tianyu Wang, Guangsi Shi, Zexing Zhao, Yanwei Fu*
- 70 Stable-SCore: A Stable Registration-based Framework for 3D Shape Correspondence, *Haolin Liu, Xiaohang Zhan, Zizheng Yan, Zhongjin Luo, Yuxin Wen, Xiaoguang Han*
- 71 Active Hyperspectral Imaging Using an Event Camera,
* *Bohan Yu, Jinxiu Liang, Zhuofeng Wang, Bin Fan, Art Subpa-asa, Boxin Shi, Imari Sato*
- 72 SphereUFormer: A U-Shaped Transformer for Spherical 360 Perception, *Yaniv Benny, Lior Wolf*
- 73 Decoupling Fine Detail and Global Geometry for Compressed Depth Map Super-Resolution, *Huan Zheng, Wencheng Han, Jianbing Shen*
- 74 A Unified Image-Dense Annotation Generation Model for Underwater Scenes, *Hongkai Lin, Dingkang Liang, Zhenghao Qi, Xiang Bai*
- 75 Active Event-based Stereo Vision, *Jianing Li, Yunjian Zhang, Haiqian Han, Xiangyang Ji*
- 76 PanDA: Towards Panoramic Depth Anything with Unlabeled Panoramas and Mobius Spatial Augmentation, *Zidong Cao, Jinjing Zhu, Weiming Zhang, Hao Ai, Haotian Bai, Hengshuang Zhao, Lin Wang*
- 77 Flow-NeRF: Joint Learning of Geometry, Poses, and Dense Flow within Unified Neural Representations, *Xunzhi Zheng, Dan Xu*
- 78 OmniStereo: Real-time Omnidirectional Depth Estimation with Multiview Fisheye Cameras, *Jiayi Deng, Yushen Wang, Haitao Meng, Zuoxun Hou, Yi Chang, Gang Chen*
- 79 Stereo Anywhere: Robust Zero-Shot Deep Stereo Matching Even
☆ Where Either Stereo or Mono Fail, *Luca Bartolomei, Fabio Tosi, Matteo Poggi, Stefano Mattoccia*
- 80 UniK3D: Universal Camera Monocular 3D Estimation,
☆ *Luigi Piccinelli, Christos Sakaridis, Mattia Segu, Yung-Hsu Yang, Siyuan Li, Wim Abbeloos, Luc Van Gool*
- 81 Structure-from-Motion with a Non-Parametric Camera Model,
* *Yihan Wang, Linfei Pan, Marc Pollefeys, Viktor Larsson*
- 82 MUST3R: Multi-view Network for Stereo 3D
* Reconstruction, *Yohann Cabon, Lucas Stofli, Leonid Antsfeld, Gabriela Csurka, Boris Chidlovskii, Jerome Revaud, Vincent Leroy*
- 83 Extreme Rotation Estimation in the Wild, *Hana Bezaele, Dotan Ankri, Ruojin Cai, Hadar Averbach-Elor*
- 84 Pow3R: Empowering Unconstrained 3D Reconstruction with Camera and Scene Priors, *Wonbong Jang, Philippe Weinzaepfel, Vincent Leroy, Lourdes Agapito, Jerome Revaud*
- 85 Gaussian Splatting Feature Fields for (Privacy-Preserving) Visual Localization, *Maxime Pietrantonio, Gabriela Csurka, Torsten Sattler*
- 86 Dense Match Summarization for Faster Two-view Estimation, *Jonathan Astermark, Anders Heyden, Viktor Larsson*
- 87 Cross-View Completion Models are Zero-shot Correspondence
* Estimators, *Honggyu An, Jin Hyeon Kim, Seonghoon Park, Jaewoo Jung, Jisang Han, Sunghwan Hong, Seungryong Kim*
- 88 Uni4D: Unifying Visual Foundation Models for 4D Modeling from
* a Single Video, *David Yifan Yao, Albert J. Zhai, Shenlong Wang*
- 89 EquiPose: Exploiting Permutation Equivariance for Relative Camera Pose Estimation, *Yuzhen Liu, Qiulei Dong*
- 90 SemAlign3D: Semantic Correspondence between RGB-Images through Aligning 3D Object-Class Representations, *Krispin Wandel, Hesheng Wang*
- 91 PromptHMR: Promptable Human Mesh Recovery, *Yufu Wang, Yu Sun, Priyanka Patel, Kostas Daniilidis, Michael J. Black, Muhammed Kocabas*
- 92 DynPose: Largely Improving the Efficiency of Human Pose Estimation by a Simple Dynamic Framework, *Yalong Xu, Lin Zhao, Chen Gong, Guangyu Li, Di Wang, Nannan Wang*
- 93 Rethinking Correspondence-based Category-Level Object Pose Estimation, *Huan Ren, Wenfei Yang, Shifeng Zhang, Tianzhu Zhang*
- 94 UA-Pose: Uncertainty-Aware 6D Object Pose Estimation and
☆ Online Object Completion with Partial References, *Ming-Feng Li, Xin Yang, Fu-En Wang, Hritam Basak, Yuyin Sun, Shreekanth Gayaka, Min Sun, Cheng-Hao Kuo*
- 95 PlanarSplatting: Accurate Planar Surface Reconstruction in
* 3 Minutes, *Bin Tan, Rui Yu, Yujun Shen, Nan Xue*
- 96 Prior-free 3D Object Tracking, *Xiuqiang Song, Li Jin, Zhengxian Zhang, Jiachen Li, Fan Zhong, Guofeng Zhang, Xueying Qin*
- 97 Progressive Correspondence Regenerator for Robust 3D Registration, *Guiyu Zhao, Sheng Ao, Ye Zhang, Kai Xu, Yulan Guo*
- 98 CaMuViD: Calibration-Free Multi-View Detection, *Amir Etefaghi Daryani, M. Usman Maqbool Bhutta, Byron Hernandez, Henry Medeiros*
- 99 A New Statistical Model of Star Speckles for Learning to Detect and Characterize Exoplanets in Direct Imaging Observations, *Théo Bodrito, Olivier Flasseur, Julien Mairal, Jean Ponce, Maud Langlois, Anne-Marie Lagrange*
- 100 AG-VPReID: A Challenging Large-Scale Benchmark for Aerial-Ground Video-based Person Re-Identification, *Huy Nguyen, Kien Nguyen, Akila Pemasiri, Feng Liu, Sridha Sridharan, Clinton Fookes*
- 101 MambaVO: Deep Visual Odometry Based on Sequential Matching Refinement and Training Smoothing, *Shuo Wang, Wanting Li, Yongcai Wang, Zhaoxin Fan, Zhe Huang, Xudong Cai, Jian Zhao, Deying Li*
- 102 Point-Cache: Test-time Dynamic and Hierarchical Cache for
☆ Robust and Generalizable Point Cloud Analysis, *Hongyu Sun, Qihong Ke, Ming Cheng, Yongcai Wang, Deying Li, Chenhui Gou, Jianfei Cai*
- 103 HotSpot: Signed Distance Function Optimization with an
* Asymptotically Sufficient Condition, *Zimo Wang, Cheng Wang, Taiki Yoshino, Sirui Tao, Ziyang Fu, Tzu-Mao Li*
- 104 High-quality Point Cloud Oriented Normal Estimation via Hybrid Angular and Euclidean Distance Encoding, *Yuanqi Li, Jingcheng Huang, Hongshen Wang, Peiyuan Lv, Yansong Liu, Jiuming Zheng, Jie Guo, Yanwen Guo*
- 105 A Lightweight UDF Learning Framework for 3D Reconstruction
☆ Based on Local Shape Functions, *Jiangbei Hu, Yanggeng Li, Fei Hou, Junhui Hou, Zhebin Zhang, Shengfa Wang, Na Lei, Ying He*
- 106 GenPC: Zero-shot Point Cloud Completion via 3D Generative Priors, *An Li, Zhe Zhu, Mingqiang Wei*
- 107 UniPre3D: Unified Pre-training of 3D Point Cloud Models with Cross-Modal Gaussian Splatting, *Ziyi Wang, Yanran Zhang, Jie Zhou, Jiwen Lu*
- 108 DeepLA-Net: Very Deep Local Aggregation Networks for Point Cloud Analysis, *Ziyin Zeng, Mingyue Dong, Jian Zhou, Huan Qiu, Zhen Dong, Man Luo, Bijun Li*
- 109 SAMBLE: Shape-Specific Point Cloud Sampling for an Optimal Trade-Off Between Local Detail and Global Uniformity, *Chengzhi Wu, Yuxin Wan, Hao Fu, Julius Pfommer, Zeyun Zhong, Junwei Zheng, Jiaming Zhang, Jürgen Beyerer*
- 110 PO3AD: Predicting Point Offsets toward Better 3D Point Cloud Anomaly Detection, *Jianan Ye, Weiguang Zhao, Xi Yang, Guangliang Cheng, Kaizhu Huang*
- 111 HeMoRa: Unsupervised Heuristic Consensus Sampling for Robust Point Cloud Registration, *Shaocheng Yan, Yiming Wang, Kaiyan Zhao, Pengcheng Shi, Zhenjun Zhao, Yongjun Zhang, Jiayuan Li*
- 112 LogoSP: Local-global Grouping of Superpoints for Unsupervised Semantic Segmentation of 3D Point Clouds, *Zihui Zhang, Weisheng Dai, Hongtao Wen, Bo Yang*
- 113 AirRoom: Objects Matter in Room Reidentification, *Runmao Yao, Yi Du, Zhuoqun Chen, Haoze Zheng, Chen Wang*

- 114 Open-Canopy: Towards Very High Resolution Forest Monitoring,
* *Fajwel Fogel, Yohann Perron, Nikola Besic, Laurent Saint-André, Agnès Pellissier-Tanon, Martin Schwartz, Thomas Boudras, Ibrahim Fayad, Alexandre d'Aspremont, Loic Landrieu, Philippe Ciais*
- 115 UniMamba: Unified Spatial-Channel Representation Learning with Group-Efficient Mamba for LiDAR-based 3D Object Detection,
Xin Jin, Haisheng Su, Kai Liu, Cong Ma, Wei Wu, Fei HUI, Junchi Yan
- 116 Learning to Detect Objects from Multi-Agent LiDAR Scans without Manual Labels,
Qiming Xia, Wenkai Lin, Haoen Xiang, Xun Huang, Siheng Chen, Zhen Dong, Cheng Wang, Chenglu Wen
- 117  Generalized LiDAR Segmentation in Adverse Weather,
Longyu Yang, Ping Hu, Shangbo Yuan, Lu Zhang, Jun Liu, Hengtao Shen, Xiaofeng Zhu
- 118 HiLoTs: High-Low Temporal Sensitive Representation Learning for Semi-Supervised LiDAR Segmentation in Autonomous Driving,
R.D. Lin, Pengcheng Weng, Yinqiao Wang, Han Ding, Jinsong Han, Fei Wang
- 119 A Dataset for Semantic Segmentation in the Presence of Unknowns,
☆ *Zakaria Laskar, Tomas Vojir, Matej Grcic, Iaroslav Melekhov, Shankar Gangisetty, Juho Kannala, Jiri Matas, Giorgos Tolias, C.V. Jawahar*
- 120 MAD: Memory-Augmented Detection of 3D Objects,
Ben Agro, Sergio Casas, Patrick Wang, Thomas Gilles, Raquel Urtasun
- 121 High Temporal Consistency through Semantic Similarity Propagation in Semi-Supervised Video Semantic Segmentation for Autonomous Flight,
Cédric Vincent, Taehyoung Kim, Henri Meeß
- 122 EventFly: Event Camera Perception from Ground to the Sky,
☆ *Lingdong Kong, Dongyue Lu, Xiang Xu, Lai Xing Ng, Wei Tsang Ooi, Benoit R. Cottreau*
- 123 MNE-SLAM: Multi-Agent Neural SLAM for Mobile Robots,
☆ *Tianchen Deng, Guole Shen, Chen Xun, Shenghai Yuan, Tongxin Jin, Hongming Shen, Yanbo Wang, Jingchuan Wang, Hesheng Wang, Danwei Wang, Weidong Chen*
- 124 BEVDiffuser: Plug-and-Play Diffusion Model for BEV Denoising with Ground-Truth Guidance,
* *Xin Ye, Burhaneddin Yaman, Sheng Cheng, Feng Tao, Abhirup Mallik, Liu Ren*
- 125 Rethinking Temporal Fusion with a Unified Gradient Descent View for 3D Semantic Occupancy Prediction,
Dubing Chen, Huan Zheng, Jin Fang, Xingping Dong, Xianfei Li, Wenlong Liao, Tao He, Pai Peng, Jianbing Shen
- 126 STCOcc: Sparse Spatial-Temporal Cascade Renovation for 3D Occupancy and Scene Flow Prediction,
Zhimin Liao, Ping Wei, Shuaijia Chen, Haoxuan Wang, Ziyang Ren
- 127 Towards Satellite Image Road Graph Extraction: A Global-Scale Dataset and A Novel Method,
Pan Yin, Kaiyu Li, Xiangyong Cao, Jing Yao, Lei Liu, Xueru Bai, Feng Zhou, Deyu Meng
- 128 LiDAR-RT: Gaussian-based Ray Tracing for Dynamic LiDAR Re-simulation,
Chenxu Zhou, Lvchang Fu, Sida Peng, Yunzhi Yan, Zhanhua Zhang, Yong Chen, Jiazhi Xia, Xiaowei Zhou
- 129 FlexDrive: Toward Trajectory Flexibility in Driving Scene Gaussian Splatting Reconstruction and Rendering,
Jingqiu Zhou, Lue Fan, Linjiang Huang, Xiaoyu Shi, Si Liu, Zhaoxiang Zhang, Hongsheng Li
- 130 ReConDreamer: Crafting World Models for Driving Scene Reconstruction via Online Restoration,
Chaojun Ni, Guosheng Zhao, Xiaofeng Wang, Zheng Zhu, Wenkang Qin, Guan Huang, Chen Liu, Yuyin Chen, Yida Wang, Xueyang Zhang, Yifei Zhan, Kun Zhan, Peng Jia, Xianpeng Lang, Xingang Wang, Wenjun Mei
- 131 SceneDiffuser++: City-Scale Traffic Simulation via a Generative World Model,
Shuhan Tan, John Lambert, Hong Jeon, Sakshum Kulshrestha, Yijing Bai, Jing Luo, Dragomir Anguelov, Mingxing Tan, Chiyu Max Jiang
- 132 Vid2Sim: Realistic and Interactive Simulation from Video for Urban Navigation,
Ziyang Xie, Zhizheng Liu, Zhenghao Peng, Wayne Wu, Bolei Zhou
- 133 One is Plenty: A Polymorphic Feature Interpreter for Immutable Heterogeneous Collaborative Perception,
Yuchen Xia, Quan Yuan, Guiyang Luo, Xiaoyuan Fu, Yang Li, Xuanhan Zhu, Tianyou Luo, Siheng Chen, Jinglin Li
- 134 GoalFlow: Goal-Driven Flow Matching for Multimodal Trajectories Generation in End-to-End Autonomous Driving,
Zebin Xing, Xingyu Zhang, Yang Hu, Bo Jiang, Tong He, Qian Zhang, Xiaoxiao Long, Wei Yin
- 135 ModeSeq: Taming Sparse Multimodal Motion Prediction with Sequential Mode Modeling,
Zikang Zhou, Hengjian Zhou, Haibo Hu, Zihao Wen, Jianping Wang, Yung-Hui Li, Yu-Kai Huang
- 136 S4-Driver: Scalable Self-Supervised Driving Multimodal Large Language Model with Spatio-Temporal Visual Representation,
Yichen Xie, Runsheng Xu, Tong He, Jyh-Jing Hwang, Katie Luo, Jingwei Ji, Hubert Lin, Letian Chen, Yiren Lu, Zhaoqi Leng, Dragomir Anguelov, Mingxing Tan
- 137 JTD-UAV: MLLM-Enhanced Joint Tracking and Description Framework for Anti-UAV Systems,
Yifan Wang, Jian Zhao, Zhaoxin Fan, Xin Zhang, Xuecheng Wu, Yudian Zhang, Lei Jin, Xinyue Li, Gang Wang, Mengxi Jia, Ping Hu, Zheng Zhu, Xuelong Li
- 138 Adapting to Observation Length of Trajectory Prediction via Contrastive Learning,
Ruiqi Qiu, Jun Gong, Xinyu Zhang, Siqi Luo, Bowen Zhang, Yi Cen
- 139 Asynchronous Collaborative Graph Representation for Frames and Events,
Dianze Li, Jianing Li, Xu Liu, Xiaopeng Fan, Yonghong Tian
- 140 METASCENES: Towards Automated Replica Creation for Real-world 3D Scans,
Huangyue Yu, Baoxiong Jia, Yixin Chen, Yandan Yang, Puhao Li, Rongpeng Su, Jiaxin Li, Qing Li, Wei Liang, Song-Chun Zhu, Tengyu Liu, Siyuan Huang
- 141 GEAL: Generalizable 3D Affordance Learning with Cross-Modal Consistency,
☆ *Dongyue Lu, Lingdong Kong, Tianxin Huang, Gim Hee Lee*
- 142 SeqAfford: Sequential 3D Affordance Reasoning via Multimodal Large Language Model,
Chunlin Yu, Hanqing Wang, Ye Shi, Haoyang Luo, Sibe Yang, Jingyi Yu, Jingya Wang
- 143 CoT-VLA: Visual Chain-of-Thought Reasoning for Vision-Language-Action Models,
☆ *Qingqing Zhao, Yao Lu, Moo Jin Kim, Zipeng Fu, Zhuoyang Zhang, Yecheng Wu, Zhaoshuo Li, Qianli Ma, Song Han, Chelsea Finn, Ankur Handa, Tsung-Yi Lin, Gordon Wetzstein, Ming-Yu Liu, Donglai Xiang*
- 144 MoManipVLA: Transferring Vision-language-action Models for General Mobile Manipulation,
Zhenyu Wu, Yuheng Zhou, Xiuwei Xu, Ziwei Wang, Haibin Yan
- 145 RoboBrain: A Unified Brain Model for Robotic Manipulation from Abstract to Concrete,
Yuheng Ji, Huajie Tan, Jiayu Shi, Xiaoshuai Hao, Yuan Zhang, Hengyuan Zhang, Pengwei Wang, Mengdi Zhao, Hao Mu, Pengju An, Xinda Xue, Qinghang Su, Huaihai Lyu, Xiaolong Zheng, Jiaming Liu, Zhongyuan Wang, Shanghang Zhang
- 146 G3Flow: Generative 3D Semantic Flow for Pose-aware and Generalizable Object Manipulation,
☆ *Tianxing Chen, Yao Mu, Zhixuan Liang, Zanxin Chen, Shijia Peng, Qiangyu Chen, Mingkun Xu, Ruizhen Hu, Hongyuan Zhang, Xuelong Li, Ping Luo*
- 147 DexHandDiff: Interaction-aware Diffusion Planning for Adaptive Dexterous Manipulation,
☆ *Zhixuan Liang, Yao Mu, Yixiao Wang, Tianxing Chen, Wenqi Shao, Wei Zhan, Masayoshi Tomizuka, Ping Luo, Mingyu Ding*
- 148 GraphMimic: Graph-to-Graphs Generative Modeling from Videos for Policy Learning,
Guangyan Chen, Te Cui, Meiling Wang, Chengcai Yang, Mengxiao Hu, Haoyang Lu, Yao Mu, Zicai Peng, Tianxing Zhou, Xinran Jiang, Yi Yang, Yufeng Yue
- 149 CORE4D: A 4D Human-Object-Human Interaction Dataset for Collaborative Object REarrangement,
Yun Liu, Chengwen Zhang, Ruofan Xing, Bingda Tang, Bowen Yang, Li Yi
- 150 PICO: Reconstructing 3D People In Contact with Objects,
☆ *Alpár Cseke, Shashank Tripathi, Sai Kumar Dwivedi, Arjun S. Lakshminpathy, Agniv Chatterjee, Michael J. Black, Dimitrios Tzionas*
- 151 Hearing Hands: Generating Sounds from Physical Interactions in 3D Scenes,
Yiming Dou, Wonseok Oh, Yuqing Luo, Antonio Loquercio, Andrew Owens
- 152 HaWoR: World-Space Hand Motion Reconstruction from Egocentric Videos,
* *Jinglei Zhang, Jiankang Deng, Chao Ma, Rolandos Alexandros Potamias*
- 153 ParaHome: Parameterizing Everyday Home Activities Towards 3D Generative Modeling of Human-Object Interactions,
Jeonghwan Kim, Jisoo Kim, Jeonghyeon Na, Hanbyul Joo

- 154 DiSRT-In-Bed: Diffusion-Based Sim-to-Real Transfer Framework
☆ for In-Bed Human Mesh Recovery, *Jing Gao, Ce Zheng, Laszlo A. Jeni, Zackory Erickson*
- 155 EnvPoser: Environment-aware Realistic Human Motion Estimation from Sparse Observations with Uncertainty Modeling, *Songpengcheng Xia, Yu Zhang, Zhuo Su, Xiaozheng Zheng, Zheng Lv, Guidong Wang, Yongjie Zhang, Qi Wu, Lei Chu, Ling Pei*
- 156 From Sparse Signal to Smooth Motion: Real-Time Motion Generation with Rolling Prediction Models, *German Barquero, Nadine Bertsch, Manojkumar Marramreddy, Carlos Chacón, Filippo Arcadu, Ferran Rigual, Nicky Sijia He, Cristina Palmero, Sergio Escalera, Yuting Ye, Robin Kips*
- 157 ALIEN: Implicit Neural Representations for Human Motion
* Prediction under Arbitrary Latency, *Dong Wei, Xiaoning Sun, Xizhan Gao, Shengxiang Hu, Huaijiang Sun*
- 158 Nonisotropic Gaussian Diffusion for Realistic 3D Human Motion Prediction, *Cecilia Curreli, Dominik Muhle, Abhishek Saroha, Zhenzhang Ye, Riccardo Marin, Daniel Cremers*
- 159 Stochastic Human Motion Prediction with Memory of Action Transition and Action Characteristic, *Jianwei Tang, Hong Yang, Tengyue Chen, Jian-Fang Hu*
- 160 ArtFormer: Controllable Generation of Diverse 3D Articulated Objects, *Jiayi Su, Youhe Feng, Zheng Li, Jinhua Song, Yangfan He, Botao Ren, Botian Xu*
- 161 FinePhys: Fine-grained Human Action Generation by Explicitly Incorporating Physical Laws for Effective Skeletal Guidance, *Dian Shao, Mingfei Shi, Shengda Xu, Haodong Chen, Yongle Huang, Binglu Wang*
- 162 Shape My Moves: Text-Driven Shape-Aware Synthesis of Human
☆ Motions, *Ting-Hsuan Liao, Yi Zhou, Yu Shen, Chun-Hao Paul Huang, Saayan Mitra, Jia-Bin Huang, Uttaran Bhattacharya*
- 163 AniMo: Species-Aware Model for Text-Driven Animal Motion Generation, *Xuan Wang, Kai Ruan, Xing Zhang, Gaoang Wang*
- 164 Exploring Timeline Control for Facial Motion Generation, *Yifeng Ma, Jinwei Qi, Chaonan Ji, Peng Zhang, Bang Zhang, Zhidong Deng, Liefeng Bo*
- 165 TokenMotion: Decoupled Motion Control via Token Disentanglement for Human-centric Video Generation, *Ruining Li, Daitao Xing, Huiming Sun, Yuanzhou Ha, Jinglin Shen, Chiuman Ho*
- 166 Exploring Temporally-Aware Features for Point Tracking, *Inès Hyeonsu Kim, Seokju Cho, Jiahui Huang, Jung Yi, Joon-Young Lee, Seungryong Kim*
- 167 HumanMM: Global Human Motion Recovery from Multi-shot Videos, *Yuhong Zhang, Guanlin Wu, Ling-Hao Chen, Zhuokai Zhao, Jing Lin, Xiaoke Jiang, Jiamin Wu, Zhuoheng Li, Hao Frank Yang, Haoqian Wang, Lei Zhang*
- 168 EDCFlow: Exploring Temporally Dense Difference Maps for Event-based Optical Flow Estimation, *Daikun Liu, Lei Cheng, Teng Wang, Changyin Sun*
- 169 Removing Reflections from RAW Photos, *Eric Kee, Adam Pikielny, Kevin Blackburn-Matzen, Marc Levoy*
- 170 Explicit Depth-Aware Blurry Video Frame Interpolation Guided by Differential Curves, *Zaoming Yan, Pengcheng Lei, Tingting Wang, Faming Fang, Junkang Zhang, Yaomin Huang, Haichuan Song*
- 171 DepthCrafter: Generating Consistent Long Depth Sequences for
* Open-world Videos, *Wenbo Hu, Xiangjun Gao, Xiaoyu Li, Sijie Zhao, Xiaodong Cun, Yong Zhang, Long Quan, Ying Shan*
- 172 You See it, You Got it: Learning 3D Creation on Pose-Free Videos at Scale, *Baorui Ma, Huachen Gao, Haoge Deng, Zhengxiong Luo, Tiejun Huang, Lulu Tang, Xinlong Wang*
- 173 Motion Prompting: Controlling Video Generation with Motion
○ Trajectories, *Daniel Geng, Charles Herrmann, Junhwa Hur, Forrester Cole, Serena Zhang, Tobias Pfaff, Tatiana Lopez-Guevara, Yusuf Aytar, Michael Rubinstein, Chen Sun, Oliver Wang, Andrew Owens, Deqing Sun*
- 174 Go-with-the-Flow: Motion-Controllable Video Diffusion Models
☆ Using Real-Time Warped Noise, *Ryan Burgert, Yuancheng Xu, Wenqi Xian, Oliver Pilarski, Pascal Clausen, Mingming He, Li Ma, Yitong Deng, Lingxiao Li, Mohsen Mousavi, Michael Ryoo, Paul Debevec, Ning Yu*
- 175 Motion Modes: What Could Happen Next?, *Karran Pandey, Yannick Hold-Geoffroy, Matheus Gadelha, Niloy J. Mitra, Karan Singh, Paul Guerrero*
- 176 FloVD: Optical Flow Meets Video Diffusion Model for Enhanced
☆ Camera-Controlled Video Synthesis, *Wonjoon Jin, Qi Dai, Chong Luo, Seung-Hwan Baek, Sunghyun Cho*
- 177 ReCapture: Generative Video Camera Controls for User-Provided Videos using Masked Video Fine-Tuning, *David Junhao Zhang, Roni Paiss, Shiran Zada, Nikhil Karnad, David E. Jacobs, Yael Pritch, Inbar Mosseri, Mike Zheng Shou, Neal Wadhwa, Nataniel Ruiz*
- 178 Tora: Trajectory-oriented Diffusion Transformer for Video Generation, *Zhenghao Zhang, Junchao Liao, Menghao Li, ZuoZhuo Dai, Bingxue Qiu, Siyu Zhu, Long Qin, Weizhi Wang*
- 179 Align-A-Video: Deterministic Reward Tuning of Image Diffusion Models for Consistent Video Editing, *Shengzhi Wang, Yingkang Zhong, Jiangchuan Mu, Kai Wu, Mingliang Xiong, Wen Fang, Mingqing Liu, Hao Deng, Bin He, Gang Li, Qingwen Liu*
- 180 Classic Video Denoising in a Machine Learning World: Robust, Fast, and Controllable, *Xin Jin, Simon Niklaus, Zhoutong Zhang, Zhihao Xia, Chunle Guo, Yuting Yang, Jiawen Chen, Chongyi Li*
- 181 Augmented Deep Contexts for Spatially Embedded Video Coding,
* *Yifan Bian, Chuanbo Tang, Li Li, Dong Liu*
- 182 EDEN: Enhanced Diffusion for High-quality Large-motion Video Frame Interpolation, *Zihao Zhang, Haoran Chen, Haoyu Zhao, Guansong Lu, Yanwei Fu, Hang Xu, Zuxuan Wu*
- 183 Continuous Space-Time Video Resampling with Invertible Motion Steganography, *Yuantong Zhang, Zhenzhong Chen*
- 184 Learning Phase Distortion with Selective State Space Models
* for Video Turbulence Mitigation, *Xingguang Zhang, Nicholas Chimitt, Xijun Wang, Yu Yuan, Stanley H. Chan*
- 185 VideoGigaGAN: Towards Detail-rich Video Super-Resolution, *Yiran Xu, Taesung Park, Richard Zhang, Yang Zhou, Eli Shechtman, Feng Liu, Jia-Bin Huang, Difan Liu*
- 186 KVQ: Boosting Video Quality Assessment via Saliency-guided Local Perception, *Yunpeng Qu, Kun Yuan, Qizhi Xie, Ming Sun, Chao Zhou, Jian Wang*
- 187 SeedVR: Seeding Infinity in Diffusion Transformer Towards
* Generic Video Restoration, *Jianyi Wang, Zhijie Lin, Meng Wei, Yang Zhao, Ceyuan Yang, Chen Change Loy, Lu Jiang*
- 188 Multi-Modal Synergistic Implicit Image Enhancement for Efficient Optical Flow Estimation, *Weichen Dai, Hexing Wu, Xiaoyang Weng, Yuxin Zheng, Yuhang Ming, Wanzeng Kong*
- 189 Efficient Video Face Enhancement with Enhanced Spatial-Temporal Consistency, *Yutong Wang, Jiajie Teng, Jiajiong Cao, Yuming Li, Chenguang Ma, Hongteng Xu, Dixin Luo*
- 190 Diffusion-based Event Generation for High-Quality Image Deblurring, *Xinan Xie, Qing Zhang, Wei-Shi Zheng*
- 191 The Change You Want To Detect: Semantic Change Detection
☆ In Earth Observation With Hybrid Data Generationf, *Yanis Benidir, Nicolas Gonthier, Clement Mallet*
- 192 Illumination Spectrum Estimation for Multispectral Images via Surface Reflectance Modeling and Spatial-Spectral Feature Generation, *Hyejin Oh, Woo-Shik Kim, Sangyoon Lee, Yungkyung Park, Je-Won Kang*
- 193 DCEvo: Discriminative Cross-Dimensional Evolutionary Learning for Infrared and Visible Image Fusion, *Jinyuan Liu, Bowei Zhang, Qingyun Mei, Xingyuan Li, Yang Zou, Zhiying Jiang, Long Ma, Risheng Liu, Xin Fan*
- 194 Binarized Neural Network for Multi-spectral Image Fusion, *Junming Hou, Xiaoyu Chen, Ran Ran, Xiaofeng Cong, Xinyang Liu, Jian Wei You, Liang-Jian Deng*
- 195 DiffFNO: Diffusion Fourier Neural Operator,
○ *Xiaoyi Liu, Hao Tang*
- 196 Bridging the Vision-Brain Gap with an Uncertainty-Aware Blur Prior, *Haitao Wu, Qing Li, Changqing Zhang, Zhen He, Xiaomin Ying*
- 197 Tokenize Image Patches: Global Context Fusion for Effective Haze Removal in Large Images, *Jiuchen Chen, Xinyu Yan, Qizhi Xu, Kaiqi Li*
- 198 Towards Lossless Implicit Neural Representation via Bit Plane
☆ Decomposition, *Woo Kyoung Han, Byeonghun Lee, Hyunmin Cho, Sunghoon Im, Kyong Hwan Jin*

- 199 Progressive Focused Transformer for Single Image Super-Resolution, *Wei Long, Xingyu Zhou, Leheng Zhang, Shuhang Gu*
- 200 HIF: Hierarchical Encoding based Implicit Image Function for Continuous Super-resolution, *Yuxuan Jiang, Ho Man Kwan, Tianhao Peng, Ge Gao, Fan Zhang, Xiaqing Zhu, Joel Sole, David Bull*
- 201 A Regularization-Guided Equivariant Approach for Image Restoration, *Yulu Bai, Jiahong Fu, Qi Xie, Deyu Meng*
- 202 Augmenting Perceptual Super-Resolution via Image Quality Predictors, *Fengjia Zhang, Samrudhdi B. Rangrej, Tristan Aumentado-Armstrong, Afsaneh Fazly, Alex Levinstein*
- 203 Rethinking Reconstruction and Denoising in the Dark: New Perspective, General Architecture and Beyond, *Tengyu Ma, Long Ma, Ziye Li, Yuetong Wang, Jinyuan Liu, Chengpei Xu, Risheng Liu*
- 204 Pixel-level and Semantic-level Adjustable Super-resolution: ☆ A Dual-LoRA Approach, *Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, Lei Zhang*
- 205 Distilling Spatially-Heterogeneous Distortion Perception for Blind ☆ Image Quality Assessment, *Xudong Li, Wenjie Nie, Yan Zhang, Runze Hu, Ke Li, Xianwu Zheng, Liujuan Cao*
- 206 Latent Space Super-Resolution for Higher-Resolution Image Generation with Diffusion Models, *Jinho Jeong, Sangmin Han, Jinwoo Kim, Seon Joo Kim*
- 207 Segment Any-Quality Images with Generative Latent Space Enhancement, *Guangqian Guo, Yong Guo, Xuehui Yu, Wenbo Li, Yaoxing Wang, Shan Gao*
- 208 Traversing Distortion-Perception Tradeoff using a Single Score-Based Generative Model, *Yuhan Wang, Suzhi Bi, Ying-Jun Angela Zhang, Xiaojun Yuan*
- 209 Sampling Innovation-Based Adaptive Compressive Sensing, *Zhifu Tian, Tao Hu, Chaoyang Niu, Di Wu, Shu Wang*
- 210 Zero-Shot Image Restoration Using Few-Step Guidance of Consistency Models (and Beyond), *Tomer Garber, Tom Tlir*
- 211 Hierarchical Adaptive Filtering Network for Text Image Specular Highlight Removal, *Zhi Jiang, Jingbo Hu, Ling Zhang, Gang Fu, Chunxia Xiao*
- 212 Erase Diffusion: Empowering Object Removal Through Calibrating ☆ Diffusion Pathways, *Yi Liu, Hao Zhou, Benlei Cui, Wenxiang Shang, Ran Lin*
- 213 Balanced Rate-Distortion Optimization in Learned Image ☆ Compression, *Yichi Zhang, Zhihao Duan, Yuning Huang, Fengqing Zhu*
- 214 Alias-Free Latent Diffusion Models: Improving Fractional Shift ☆ Equivariance of Diffusion Latent Space, *Yifan Zhou, Zeqi Xiao, Shuai Yang, Xingang Pan*
- 215 LookingGlass: Generative Anamorphoses via Laplacian Pyramid ☆ Warping, *Pascal Chang, Sergio Sancho, Jingwei Tang, Markus Gross, Vinicius Azevedo*
- 216 RAD: Region-Aware Diffusion Models for Image Inpainting, *Sora Kim, Sungho Suh, Minsik Lee*
- 217 Bridging the Gap between Gaussian Diffusion Models and Universal Quantization for Image Compression, *Lucas Relic, Roberto Azevedo, Yang Zhang, Markus Gross, Christopher Schroers*
- 218 CleanDIFT: Diffusion Features without Noise, *Nick Stracke, Stefan Andreas Baumann, Kolja Bauer, Frank Fundel, Björn Ommer*
- 219 FAM Diffusion: Frequency and Attention Modulation for ☆ High-Resolution Image Generation with Stable Diffusion, *Haosen Yang, Adrian Bulat, Isma Hadji, Hai X. Pham, Xiatian Zhu, Georgios Tzimiropoulos, Brais Martinez*
- 220 DiC: Rethinking Conv3x3 Designs in Diffusion Models, *Yuchuan Tian, Jing Han, Chengcheng Wang, Yuchen Liang, Chao Xu, Hanting Chen*
- 221 SnapGen-V: Generating a Five-Second Video within Five Seconds on a Mobile Device, *Yushu Wu, Zhixing Zhang, Yanyu Li, Yanwu Xu, Anil Kag, Yang Sui, Huseyin Coskun, Ke Ma, Aleksei Lebedev, Ju Hu, Dimitris N. Metaxas, Yanzhi Wang, Sergey Tulyakov, Jian Ren*
- 222 RandAR: Decoder-only Autoregressive Visual Generation in Random ☆ Orders, *Ziqi Pang, Tianyuan Zhang, Fujun Luan, Yunze Man, Hao Tan, Kai Zhang, William T. Freeman, Yu-Xiong Wang*
- 223 Learning Flow Fields in Attention for Controllable Person Image Generation, *Zijian Zhou, Shikun Liu, Xiao Han, Haozhe Liu, Kam Woh Ng, Tian Xie, Yuren Cong, Hang Li, Mengmeng Xu, Juan-Manuel Perez-Rua, Aditya Patel, Tao Xiang, Miaojing Shi, Sen He*
- 224 Nested Diffusion Models Using Hierarchical Latent Priors, *Xiao Zhang, Ruoxi Jiang, Rebecca Willett, Michael Maire*
- 225 Adaptive Non-Uniform Timestep Sampling for Accelerating Diffusion Model Training, *Myunsoo Kim, Donghyeon Ki, Seong-Woong Shim, Byung-Jun Lee*
- 226 Scaling Inference Time Compute for Diffusion Models, *Nanye Ma, ☆ Shangyuan Tong, Haolin Jia, Hexiang Hu, Yu-Chuan Su, Mingda Zhang, Xuan Yang, Yandong Li, Tommi Jaakkola, Xuhui Jia, Saining Xie*
- 227 HMAR: Efficient Hierarchical Masked Auto-Regressive Image Generation, *Hermann Kumbong, Xian Liu, Tsung-Yi Lin, Ming-Yu Liu, Xihui Liu, Ziwei Liu, Daniel Y. Fu, Christopher Re, David W. Romero*
- 228 TokenFlow: Unified Image Tokenizer for Multimodal Understanding and Generation, *Liao Qu, Huichao Zhang, Yiheng Liu, Xu Wang, Yi Jiang, Yiming Gao, Hu Ye, Daniel K. Du, Zehuan Yuan, Xinglong Wu*
- 229 SketchFusion: Learning Universal Sketch Features through Fusing Foundation Models, *Subhadeep Koley, Tapas Kumar Dutta, Aneeshan Sain, Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Yi-Zhe Song*
- 230 StreamingT2V: Consistent, Dynamic, and Extendable Long Video Generation from Text, *Roberto Henschel, Levon Khachatryan, Hayk Poghosyan, Daniil Hayrapetyan, Vahram Tadevosyan, Zhangyang Wang, Shant Navasardyan, Humphrey Shi*
- 231 LinGen: Towards High-Resolution Minute-Length Text-to-Video ☆ Generation with Linear Computational Complexity, *Hongjie Wang, Chih-Yao Ma, Yen-Cheng Liu, Ji Hou, Tao Xu, Jialiang Wang, Felix Juefei-Xu, Yaqiao Luo, Peizhao Zhang, Tingbo Hou, Peter Vajda, Niraj K. Jha, Xiaoliang Dai*
- 232 VideoDirector: Precise Video Editing via Text-to-Video Models, *Yukun Wang, Longguang Wang, Zhiyuan Ma, Qibin Hu, Kai Xu, Yulan Guo*
- 233 VideoGuide: Improving Video Diffusion Models without Training Through a Teacher's Guide, *Dohun Lee, Bryan Sangwoo Kim, Geon Yeong Park, Jong Chul Ye*
- 234 AKiRa: Augmentation Kit on Rays for Optical Video Generation, ☆ *Xi Wang, Robin Courant, Marc Christie, Vicky Kalogeiton*
- 235 TCFG: Tangential Damping Classifier-free Guidance, *Mingi Kwon, Shin seong Kim, Jaeseok Jeong, Yi Ting Hsiao, Youngjung Uh*
- 236 StyleMaster: Stylize Your Video with Artistic Generation and Translation, *Zixuan Ye, Huijuan Huang, Xintao Wang, Pengfei Wan, Di Zhang, Wenhan Luo*
- 237 Conditional Balance: Improving Multi-Conditioning Trade-Offs in Image Generation, *Nadav Z. Cohen, Oron Nir, Ariel Shamir*
- 238 FDS: Frequency-Aware Denoising Score for Text-Guided Latent Diffusion Image Editing, *Yufan Ren, Zicong Jiang, Tong Zhang, Søren Forchhammer, Sabine Süsstrunk*
- 239 FeedEdit: Text-Based Image Editing with Dynamic Feedback Regulation, *Fengyi Fu, Lei Zhang, Mengqi Huang, Zhendong Mao*
- 240 One Diffusion to Generate Them All, *Duong H. Le, Tuan Pham, Sangho Lee, Christopher Clark, Aniruddha Kembhavi, Stephan Mandt, Ranjay Krishna, Jiasen Lu*
- 241 MoEdit: On Learning Quantity Perception for Multi-object Image Editing, *Yanfeng Li, Kahou Chan, Yue Sun, Chantong Lam, Tong Tong, Zitong Yu, Keren Fu, Xiaohong Liu, Tao Tan*
- 242 InsightEdit: Towards Better Instruction Following for Image Editing, *Yingjing Xu, Jie Kong, Jiazhi Wang, Xiao Pan, Bo Lin, Qiang Liu*
- 243 Instruction-based Image Manipulation by Watching How Things ☆ Move, *Mingdeng Cao, Xuaner Zhang, Yinqiang Zheng, Zhihao Xia*
- 244 TFCustom: Customized Image Generation with Time-Aware ☆ Frequency Feature Guidance, *Mushui Liu, Dong She, Jingxuan Pang, Qihan Huang, Jiacheng Ying, Wanggui He, Yuanlei Hou, Siming Fu*
- 245 OpenING: A Comprehensive Benchmark for Judging Open-ended ☆ Interleaved Image-Text Generation, *Pengfei Zhou, Xiaopeng Peng, Jiajun Song, Chuanhao Li, Zhaopan Xu, Yue Yang, Ziyao Guo, Hao Zhang, Yuqi Lin, Yefei He, Lirui Zhao, Shuo Liu, Tianhua Li, Yuxuan Xie, Xiaojun Chang, Yu Qiao, Wenqi Shao, Kaipeng Zhang*
- 246 PreciseCam: Precise Camera Control for Text-to-Image Generation, *Edurne Bernal-Berdun, Ana Serrano, Belen Masia, Matheus Gadelha, Yannick Hold-Geoffroy, Xin Sun, Diego Gutierrez*
- 247 Science-T2I: Addressing Scientific Illusions in Image Synthesis, *Jialuo Li, Wenhao Chai, Xingyu Fu, Haiyang Xu, Saining Xie*

- 248 Type-R: Automatically Retouching Typos for Text-to-Image Generation,
* *Wataru Shimoda, Naoto Inoue, Daichi Haraguchi, Hayato Mitani,*
☆ *Seiichi Uchida, Kota Yamaguchi*
- 249 Flowing from Words to Pixels: A Noise-Free Framework for
* Cross-Modality Evolution, *Qihao Liu, Xi Yin, Alan Yuille,*
☆ *Andrew Brown, Mannat Singh*
- 250 GPS as a Control Signal for Image Generation, *Chao Feng,*
Ziyang Chen, Aleksander Holynski, Alexei A. Efros, Andrew Owens
- 251 Dual Diffusion for Unified Image Generation and
Understanding, *Zijie Li, Henry Li, Yichun Shi, Amir Barati*
Farimani, Yuval Kluger, Linjie Yang, Peng Wang
- 252 Compass Control: Multi Object Orientation Control for Text-to-
☆ Image Generation, *Rishubh Parihar, Vaibhav Agrawal,*
Sachidanand VS, Venkatesh Babu Radhakrishnan
- 253 MC²: Multi-concept Guidance for Customized Multi-concept
Generation, *Jiaxiu Jiang, Yabo Zhang, Kailai Feng, Xiaohe Wu,*
Wenbo Li, Renjing Pei, Fan Li, Wangmeng Zuo
- 254 Synthetic Data is an Elegant GIFT for Continual Vision-Language
Models, *Bin Wu, Wuxuan Shi, Jinqiao Wang, Mang Ye*
- 255 Curriculum Direct Preference Optimization for Diffusion and
Consistency Models, *Florinel-Alin Croitoru, Vlad Hondru,*
Radu Tudor Ionescu, Nicu Sebe, Mubarak Shah
- 256 DoraCycle: Domain-Oriented Adaptation of Unified Generative
Model in Multimodal Cycles, *Rui Zhao, Weijia Mao, Mike Zheng Shou*
- 257 SerialGen: Personalized Image Generation by First
Standardization Then Personalization, *Cong Xie, Han Zou,*
Ruiqi Yu, Yan Zhang, Zhenpeng Zhan
- 258 Prometheus: 3D-Aware Latent Diffusion Models for Feed-Forward
Text-to-3D Scene Generation, *Yuanbo Yang, Jiahao Shao,*
Xinyang Li, Yujun Shen, Andreas Geiger, Yiyi Liao
- 259 VinaBench: Benchmark for Faithful and Consistent Visual Narratives,
Silin Gao, Sheryl Mathew, Li Mi, Sepideh Mamooler, Mengjie Zhao,
Hiromi Wakaki, Yuki Mitsufuji, Syrielle Montariol, Antoine Bosselut
- 260 CoSER: Towards Consistent Dense Multiview Text-to-Image
* Generator for 3D Creation, *Bonan Li, Zicheng Zhang, Xingyi Yang,*
Xinchao Wang
- 261 ArtiScene: Language-Driven Artistic 3D Scene Generation
☆ Through Image Intermediary, *Zeqi Gu, Yin Cui, Zhao Shuo Li,*
Fangyin Wei, Yunhao Ge, Jinwei Gu, Ming-Yu Liu, Abe Davis, Yifan Ding
- 262 AutoPresent: Designing Structured Visuals from Scratch,
Jiaxin Ge, Zora Zhiruo Wang, Xuhui Zhou, Yi-Hao Peng, Sanjay
Subramanian, Qinyue Tan, Maarten Sap, Alane Suhr, Daniel Fried,
Graham Neubig, Trevor Darrell
- 263 LineArt: A Knowledge-guided Training-free High-quality
Appearance Transfer for Design Drawing with Diffusion Model,
Xi Wang, Hongzhen Li, Heng Fang, Yichen Peng, Haoran Xie,
Xi Yang, Chuntao Li
- 264 ChatGarment: Garment Estimation, Generation and Editing via
Large Language Models, *Siyuan Bian, Chenghao Xu, Yuliang Xiu,*
Artur Grigorev, Zhen Liu, Cewu Lu, Michael J. Black, Yao Feng
- 265 Rethinking Personalized Aesthetics Assessment: Employing
* Physique Aesthetics Assessment as An Exemplification,
Haobin Zhong, Shuai He, Anlong Ming, Huadong Ma
- 266 ConceptGuard: Continual Personalized Text-to-Image Generation
with Forgetting and Confusion Mitigation, *Zirun Guo, Tao Jin*
- 267 DKDM: Data-Free Knowledge Distillation for Diffusion Models
with Any Architecture, *Qianlong Xiang, Miao Zhang,*
Yuzhang Shang, Jianlong Wu, Yan Yan, Liqiang Nie
- 268 Memories of Forgotten Concepts, *Matan Rusanovsky,*
* *Shimon Malnick, Amir Jevnisek, Ohad Fried, Shai Avidan*
- 269 Plug-and-Play Interpretable Responsible Text-to-Image
Generation via Dual-Space Multi-facet Concept Control,
Basim Azam, Naveed Akhtar
- 270 ID-Patch: Robust ID Association for Group Photo Personalization,
Yimeng Zhang, Tiancheng Zhi, Jing Liu, Shen Sang, Liming Jiang,
Qing Yan, Sijia Liu, Linjie Luo
- 271 Not Just Text: Uncovering Vision Modality Typographic Threats
in Image Generation Models, *Hao Cheng, Erjia Xiao, Jiayan Yang,*
Jiahang Cao, Qiang Zhang, Jize Zhang, Kaidi Xu, Jindong Gu,
Renjing Xu
- 272 OmniGuard: Hybrid Manipulation Localization via Augmented
Versatile Deep Image Watermarking, *Xuanyu Zhang, Zecheng Tang,*
Zhipei Xu, Runyi Li, Youmin Xu, Bin Chen, Feng Gao, Jian Zhang
- 273 IDProtector: An Adversarial Noise Encoder to Protect Against ID-
Preserving Image Generation, *Yiren Song, Pei Yang, Hai Ci,*
Mike Zheng Shou
- 274 Image Generation Diversity Issues and How to Tame Them,
Mischa Dombrowski, Weitong Zhang, Sarah Cechnicka,
Hadrien Reynaud, Bernhard Kainz
- 275 Forensic Self-Descriptions Are All You Need for Zero-Shot Detection,
Open-Set Source Attribution, and Clustering of AI-generated Images,
Tai D. Nguyen, Aref Azizpour, Matthew C. Stamm
- 276 ORIDa: Object-centric Real-world Image Composition Dataset,
☆ *Jinwoo Kim, Sangmin Han, Jinho Jeong, Jiwoo Choi, Dongyeoung Kim,*
Seon Joo Kim
- 277 SINR: Sparsity Driven Compressed Implicit Neural
Representations, *Dhananjaya Jayasundara, Sudarshan*
Rajagopalan, Yasiru Ranasinghe, Trac D. Tran, Vishal M. Patel
- 278 Tuning the Frequencies: Robust Training for Sinusoidal Neural
* Networks, *Tiago Novello, Diana Aldana, Andre Araujo, Luiz Velho*
- 279 GA3CE: Unconstrained 3D Gaze Estimation with Gaze-Aware
3D Context Encoding, *Yuki Kawana, Shintaro Shiba, Quan Kong,*
Norimasa Kobori
- 280 De²Gaze: Deformable and Decoupled Representation Learning
for 3D Gaze Estimation, *Yunfeng Xiao, Xiaowei Bai, Baojun Chen,*
Hao Su, Hao He, Liang Xie, Erwei Yin
- 281 FADA: Fast Diffusion Avatar Synthesis with Mixed-Supervised
Multi-CFG Distillation, *Tianyun Zhong, Chao Liang, Jianwen Jiang,*
Gaojie Lin, Jiaqi Yang, Zhou Zhao
- 282 Synchronized Video-to-Audio Generation via Mel Quantization-
Continuum Decomposition, *Juncheng Wang, Chao Xu, Cheng Yu,*
Lei Shang, Zhe Hu, Shujun Wang, Liefeng Bo
- 283 Improving Sound Source Localization with Joint Slot Attention
on Image and Audio, *Inho Kim, Youngkil Song, Jicheol Park,*
Won Hwa Kim, Suha Kwak
- 284 Dynamic Derivation and Elimination: Audio Visual Segmentation
with Enhanced Audio Semantics, *Chen Liu, Liying Yang, Peike Li,*
Dadong Wang, Lincheng Li, Xin Yu
- 285 Adapting to the Unknown: Training-Free Audio-Visual Event
Perception with Dynamic Thresholds, *Eitan Shaar, Ariel Shaulov,*
Gal Chechik, Lior Wolf
- 286 HarmonySet: A Comprehensive Dataset for Understanding
Video-Music Semantic Alignment and Temporal
Synchronization, *Zitang Zhou, Ke Mei, Yu Lu, Tianyi Wang,*
Fengyun Rao
- 287 Precise Event Spotting in Sports Videos: Solving Long-Range
Dependency and Class Imbalance, *Sanchayan Santra,*
Vishal Chudasama, Pankaj Wasnik, Vineeth N Balasubramanian
- 288 The Devil is in the Prompts: Retrieval-Augmented Prompt
Optimization for Text-to-Video Generation, *Bingjie Gao,*
Xinyu Gao, Xiaoxue Wu, Yujie Zhou, Yu Qiao, Li Niu, Xinyuan Chen,
Yaohui Wang
- 289 Long Video Diffusion Generation with Segmented Cross-Attention
and Content-Rich Video Data Curation, *Xin Yan, Yuxuan Cai,*
Qiuyue Wang, Yuan Zhou, Wenhao Huang, Huan Yang
- 290 V-Stylist: Video Stylization via Collaboration and Reflection of
MLLM Agents, *Zhengrong Yue, Shaobin Zhuang, Kunchang Li,*
Yanbo Ding, Yali Wang
- 291 FineVQ: Fine-Grained User Generated Content Video Quality
* Assessment, *Huiyu Duan, Qiang Hu, Jiarui Wang, Liu Yang,*
☆ *Zitong Xu, Lu Liu, Xiongkuo Min, Chunlei Cai, Tianxiao Ye,*
Xiaoyun Zhang, Guangtao Zhai
- 292 VLog: Video-Language Models by Generative Retrieval of
Narration Vocabulary, *Kevin Qinghong Lin, Mike Zheng Shou*
- 293 Q-Bench-Video: Benchmark the Video Quality Understanding of
LMMs, *Zicheng Zhang, Ziheng Jia, Haoning Wu, Chunyi Li, Zijian*
Chen, Yingjie Zhou, Wei Sun, Xiaohong Liu, Xiongkuo Min, Weisi
Lin, Guangtao Zhai
- 294 LION-FS: Fast & Slow Video-Language Thinker as Online Video
Assistant, *Wei Li, Bing Hu, Rui Shao, Leyang Shen, Liqiang Nie*

- 295 AVQACL: A Novel Benchmark for Audio-Visual Question Answering Continual Learning, *Kaixuan Wu, Xinde Li, Xinling Li, Chuanfei Hu, Guoliang Wu*
- 296 Commonsense Video Question Answering through Video-Grounded Entailment Tree Reasoning, *Huabin Liu, Filip Ilievski, Cees G. M. Snoek*
- 297 VideoTree: Adaptive Tree-based Video Representation for LLM Reasoning on Long Videos, *Ziyang Wang, Shoubin Yu, Elias Stengel-Eskin, Jaehong Yoon, Feng Cheng, Gedas Bertasius, Mohit Bansal*
- 298 STEP: Enhancing Video-LLMs' Compositional Reasoning by Spatio-Temporal Graph-guided Self-Training, *Haiyi Qiu, Minghe Gao, Long Qian, Kaihang Pan, Qifan Yu, Juncheng Li, Wenjie Wang, Siliang Tang, Yueting Zhuang, Tat-Seng Chua*
- 299 VideoICL: Confidence-based Iterative In-context Learning for Out-of-Distribution Video Understanding, *Kangsan Kim, Geon Park, Youngwan Lee, Woongyeong Yeo, Sung Ju Hwang*
- 300 PAVE: Patching and Adapting Video Large Language Models, ☆ *Zhuoming Liu, Yiquan Li, Khoi Duc Nguyen, Yiwu Zhong, Yin Li*
- 301 BOLT: Boost Large Vision-Language Model Without Training for Long-form Video Understanding, *Shuming Liu, Chen Zhao, Tianqi Xu, Bernard Ghanem*
- 302 Online Video Understanding: OVBench and VideoChat-Online, *Zhenpeng Huang, Xinhao Li, Jiaqi Li, Jing Wang, Xiangyu Zeng, Cheng Liang, Tao Wu, Xi Chen, Liang Li, Limin Wang*
- 303 Localizing Events in Videos with Multimodal Queries, *Gengyuan Zhang, Mang Ling Ada Fok, Jialu Ma, Yan Xia, Daniel Cremers, Philip Torr, Volker Tresp, Jindong Gu*
- 304 SALOVA: Segment-Augmented Long Video Assistant for Targeted Retrieval and Routing in Long-Form Video Analysis, *Junho Kim, Hyunjun Kim, Hosu Lee, Yong Man Ro*
- 305 EgoTextVQA: Towards Egocentric Scene-Text Aware Video Question Answering, *Sheng Zhou, Junbin Xiao, Qingyun Li, Yicong Li, Xun Yang, Dan Guo, Meng Wang, Tat-Seng Chua, Angela Yao*
- 306 VideoGEM: Training-free Action Grounding in Videos, *Felix Vogel, ☆ Walid Bousselham, Anna Kukleva, Nina Shvetsova, Hilde Kuehne*
- 307 STPro: Spatial and Temporal Progressive Learning for Weakly Supervised Spatio-Temporal Grounding, *Aaryan Garg, Akash Kumar, Yogesh S Rawat*
- 308 SAMWISE: Infusing Wisdom in SAM2 for Text-Driven Video Segmentation, *Claudia Cuttano, Gabriele Trivigno, Gabriele Rosi, ☆ Carlo Masone, Giuseppe Averta*
- 309 Segment Any Motion in Videos, *Nan Huang, Wenzhao Zheng, Chenfeng Xu, Kurt Keutzer, Shanghang Zhang, Angjoo Kanazawa, Qianqian Wang*
- 310 SAM-I2V: Upgrading SAM to Support Promptable Video Segmentation with Less than 0.2% Training Cost, *Haiyang Mei, Pengyu Zhang, Mike Zheng Shou*
- 311 RipVIS: Rip Currents Video Instance Segmentation Benchmark for Beach Monitoring and Safety, *Andrei Dumitriu, Florin Tatui, Florin Miron, Aakash Ralhan, Radu Tudor Ionescu, Radu Timofte*
- 312 MANTA: Diffusion Mamba for Efficient and Effective Stochastic Long-Term Dense Action Anticipation, *Olga Zatsarynna, Emad Bahrami, Yazan Abu Farha, Gianpiero Francesca, Juergen Gall*
- 313 TAMT: Temporal-Aware Model Tuning for Cross-Domain Few-Shot Action Recognition, *Yilong Wang, Zilin Gao, Qilong Wang, Zhaofeng Chen, Peihua Li, Qinghua Hu*
- 314 Bridging Gait Recognition and Large Language Models Sequence Modeling, *Shaopeng Yang, Jilong Wang, Saihui Hou, Xu Liu, Chunshui Cao, Liang Wang, Yongzhen Huang*
- 315 DIV-FF: Dynamic Image-Video Feature Fields For Environment Understanding in Egocentric Videos, *Lorenzo Mur-Labadia, ☆ Josechu Guerrero, Ruben Martinez-Cantin*
- 316 Discrete to Continuous: Generating Smooth Transition Poses from Sign Language Observations, *Shengeng Tang, Jiayi He, Lechao Cheng, Jingjing Wu, Dan Guo, Richang Hong*
- 317 NoPaIn: No-box Point Cloud Attack via Optimal Transport Singular Boundary, *Zezeng Li, Xiaoyu Du, Na Lei, Liming Chen, Weimin Wang*
- 318 AI-Face: A Million-Scale Demographically Annotated AI-Generated Face Dataset and Fairness Benchmark, *Li Lin, Santosh Santosh, Mingyang Wu, Xin Wang, Shu Hu*
- 319 Improving the Transferability of Adversarial Attacks on Face Recognition with Diverse Parameters Augmentation, *Fengfan Zhou, Bangjie Yin, Hefei Ling, Qianyu Zhou, Wenxuan Wang*
- 320 GIF: Generative Inspiration for Face Recognition at Scale, *Saeed Ebrahimi, Sahar Rahimi, Ali Dabouei, Srinjoy Das, Jeremy M. Dawson, Nasser M. Nasrabadi*
- 321 Towards Effective and Sparse Adversarial Attack on Spiking Neural Networks via Breaking Invisible Surrogate Gradients, *Li Lun, Kunyu Feng, Qinglong Ni, Ling Liang, Yuan Wang, Ying Li, Dunshan Yu, Xiaoxin Cui*
- 322 Brain-Inspired Spiking Neural Networks for Energy-Efficient Object Detection, *Ziqi Li, Tao Gao, Yisheng An, Ting Chen, Jing Zhang, Yuanbo Wen, Mengkun Liu, Qianxi Zhang*
- 323 BHViT: Binarized Hybrid Vision Transformer, *Tian Gao, Yu Zhang, Zhiyuan Zhang, Huajun Liu, Kaijie Yin, Chengzhong Xu, Hui Kong*
- 324 DKC: Differentiated Knowledge Consolidation for Cloth-Hybrid Lifelong Person Re-identification, *Zhenyu Cui, Jiahuan Zhou, Yuxin Peng*
- 325 2DMamba: Efficient State Space Model for Image Representation with Applications on Giga-Pixel Whole Slide Image Classification, *Jingwei Zhang, Anh Tien Nguyen, Xi Han, Vincent Quoc-Huy Trinh, Hong Qin, Dimitris Samaras, Mahdi S. Hosseini*
- 326 GauCho: Gaussian Distributions with Cholesky Decomposition for Oriented Object Detection, *José Henrique Lima Marques, Jeffri Murrugarra-Llerena, Claudio R. Jung*
- 327 Camouflage Anything: Learning to Hide using Controlled Out-painting and Representation Engineering, *Biplab Das, Viswanath Gopalakrishnan*
- 328 AeroGen: Enhancing Remote Sensing Object Detection with Diffusion-Driven Data Generation, *Datao Tang, Xiangyong Cao, Xuan Wu, Jialin Li, Jing Yao, Xueru Bai, Dongsheng Jiang, Yin Li, Deyu Meng*
- 329 ROS-SAM: High-Quality Interactive Segmentation for Remote Sensing Moving Object, *Zhe Shan, Yang Liu, Lei Zhou, Cheng Yan, Heng Wang, Xia Xie*
- 330 Any3DIS: Class-Agnostic 3D Instance Segmentation by 2D Mask Tracking, *Phuc Nguyen, Minh Luu, Anh Tran, Cuong Pham, Khoi Nguyen*
- 331 POp-GS: Next Best View in 3D-Gaussian Splatting with P-Optimality, *Joey Wilson, Marcelino Almeida, Sachit Mahajan, Martin Labrie, Maani Ghaffari, Omid Ghasemalizadeh, Min Sun, Cheng-Hao Kuo, Arnab Sen*
- 332 Rethinking End-to-End 2D to 3D Scene Segmentation in Gaussian Splatting, *Runsong Zhu, Shi Qiu, Zhengzhe Liu, Ka-Hei Hui, Qianyi Wu, Pheng-Ann Heng, Chi-Wing Fu*
- 333 Text-guided Sparse Voxel Pruning for Efficient 3D Visual Grounding, ☆ *Wenxuan Guo, Xiuwei Xu, Ziwei Wang, Jianjiang Feng, Jie Zhou, Jiwen Lu*
- 334 OnlineAnySeg: Online Zero-Shot 3D Segmentation by Visual Foundation Model Guided 2D Mask Merging, *Yijie Tang, Jiazhao Zhang, Yuqing Lan, Yulan Guo, Dezun Dong, Chenyang Zhu, Kai Xu*
- 335 SegAgent: Exploring Pixel Understanding Capabilities in MLLMs by Imitating Human Annotator Trajectories, *Muzhi Zhu, Yuzhuo Tian, Hao Chen, Chunlun Zhou, Qingpei Guo, Yang Liu, Ming Yang, Chunhua Shen*
- 336 RELOCATE: A Simple Training-Free Baseline for Visual Query Localization Using Region-Based Representations, *Savya Khosla, ☆ Sethuraman T V, Alexander Schwing, Derek Hoiem*
- 337 SeeGround: See and Ground for Zero-Shot Open-Vocabulary 3D Visual Grounding, *Rong Li, Shijie Li, Lingdong Kong, Xulei Yang, Junwei Liang*
- 338 ReasonGrounder: LVLM-Guided Hierarchical Feature Splatting for Open-Vocabulary 3D Visual Grounding and Reasoning, *Zhenyang Liu, Yikai Wang, Sixiao Zheng, Tongying Pan, Longfei Liang, Yanwei Fu, Xiangyang Xue*
- 339 Cross-Modal 3D Representation with Multi-View Images and Point Clouds, *Ziyang Zhou, Pinghui Wang, Zi Liang, Haitao Bai, Ruofei Zhang*
- 340 Learning Visual Composition through Improved Semantic Guidance, *Austin Stone, Hagen Soltau, Robert Geirhos, Xi Yi, Ye Xia, Bingyi Cao, Kaifeng Chen, Abhijit Ogale, Jonathon Shlens*

- 341 Beyond Human Perception: Understanding Multi-Object World from Monocular View, *Keyu Guo, Yongle Huang, Shijie Sun, Xiangyu Song, Mingtao Feng, Zedong Liu, Huansheng Song, Tiantian Wang, Jianxin Li, Naveed Akhtar, Ajmal Saeed Mian*
- 342 LSceneLLM: Enhancing Large 3D Scene Understanding Using Adaptive Visual Preferences, *Hongyan Zhi, Peihao Chen, Junyan Li, Shuailei Ma, Xinyu Sun, Tianhang Xiang, Yinjie Lei, Minghui Tan, Chuang Gan*
- 343 3D-LLaVA: Towards Generalist 3D LMMs with Omni Superpoint Transformer, *Jiajun Deng, Tianyu He, Li Jiang, Tianyu Wang, Feras Dayoub, Ian Reid*
- 344 Coarse Correspondences Boost Spatial-Temporal Reasoning in Multimodal Language Model, *Benlin Liu, Yuhao Dong, Yiqin Wang, Zixian Ma, Yansong Tang, Luming Tang, Yongming Rao, Wei-Chiu Ma, Ranjay Krishna*
- 345 ☆ Advancing Semantic Future Prediction through Multimodal Visual Sequence Transformers, *Efstathios Karypidis, Ioannis Kakogeorgiou, Spyros Gidaris, Nikos Komodakis*
- 346 VISTA: Enhancing Long-Duration and High-Resolution Video Understanding by Video Spatiotemporal Augmentation, *Weiming Ren, Huan Yang, Jie Min, Cong Wei, Wenhu Chen*
- 347 GFlowVLM: Enhancing Multi-step Reasoning in Vision-Language Models with Generative Flow Networks, *Haoqiang Kang, Enna Sachdeva, Piyush Gupta, Sangjae Bae, Kwonjoon Lee*
- 348 Provoking Multi-modal Few-Shot LVLMM via Exploration-Exploitation In-Context Learning, *Cheng Chen, Yunpeng Zhai, Yifan Zhao, Jinyang Gao, Bolin Ding, Jia Li*
- 349 Perception Tokens Enhance Visual Reasoning in Multimodal Language Models, *Mahtab Bigverdi, Zelun Luo, Cheng-Yu Hsieh, Ethan Shen, Dongping Chen, Linda G. Shapiro, Ranjay Krishna*
- 350 ☆ Do Visual Imaginations Improve Vision-and-Language Navigation Agents?, *Akhil Perincherry, Jacob Krantz, Stefan Lee*
- 351 HEIE: MLLM-Based Hierarchical Explainable AIGC Image Implausibility Evaluator, *Fan Yang, Ru Zhen, Jianing Wang, Yanhao Zhang, Haoxiang Chen, Haonan Lu, Sicheng Zhao, Guiguang Ding*
- 352 Words or Vision: Do Vision-Language Models Have Blind Faith in Text?, *Ailin Deng, Tri Cao, Zhirui Chen, Bryan Hooi*
- 353 VisionArena: 230k Real World User-VLM Conversations with Preference Labels, *Christopher Chou, Lisa Dunlap, Koki Mashita, Krishna Mandal, Trevor Darrell, Ion Stoica, Joseph E. Gonzalez, Wei-Lin Chiang*
- 354 Knowledge-Aligned Counterfactual-Enhancement Diffusion Perception for Unsupervised Cross-Domain Visual Emotion Recognition, *Wen Yin, Yong Wang, Guiduo Duan, Dongyang Zhang, Xin Hu, Yuan-Fang Li, Tao He*
- 355 PEACE: Empowering Geologic Map Holistic Understanding with MLLMs, *Yangyu Huang, Tianyi Gao, Haoran Xu, Qihao Zhao, Yang Song, Zhipeng Gui, Tengchao Lv, Hao Chen, Lei Cui, Scarlett Li, Furu Wei*
- 356 FRAMES-VQA: Benchmarking Fine-Tuning Robustness across Multi-Modal Shifts in Visual Question Answering, *Chengyue Huang, Brisa Maneechotesuwan, Shivang Chopra, Zsolt Kira*
- 357 Omni-RGPT: Unifying Image and Video Region-level Understanding via Token Marks, *Miran Heo, Min-Hung Chen, De-An Huang, Sifei Liu, Subhashree Radhakrishnan, Seon Joo Kim, Yu-Chiang Frank Wang, Ryo Hachiuma*
- 358 Task-aware Cross-modal Feature Refinement Transformer with Large Language Models for Visual Grounding, *Wenbo Chen, Zhen Xu, Ruotao Xu, Si Wu, Hau-San Wong*
- 359 ☆ GroundingFace: Fine-grained Face Understanding via Pixel Grounding Multimodal Large Language Model, *Yue Han, Jiangning Zhang, Junwei Zhu, Runze Hou, Xiaozhong Ji, Chuming Lin, Xiaobin Hu, Zhucun Xue, Yong Liu*
- 360 Chat-based Person Retrieval via Dialogue-Refined Cross-Modal Alignment, *Yang Bai, Yucheng Ji, Min Cao, Jinqiao Wang, Mang Ye*
- 361 ☆ Patch Matters: Training-free Fine-grained Image Caption Enhancement via Local Perception, *Ruotian Peng, Haiying He, Yake Wei, Yandong Wen, Di Hu*
- 362 ☆ CCIN: Compositional Conflict Identification and Neutralization for Composed Image Retrieval, *Likai Tian, Jian Zhao, Zechao Hu, Zhengwei Yang, Hao Li, Lei Jin, Zheng Wang, Xuelong Li*
- 363 Imagine and Seek: Improving Composed Image Retrieval with an Imagined Proxy, *You Li, Fan Ma, Yi Yang*
- 364 CoLLM: A Large Language Model for Composed Image Retrieval, *Chuonng Huynh, Jinyu Yang, Ashish Tawari, Mubarak Shah, Son Tran, Raffay Hamid, Trishul Chilimbi, Abhinav Shrivastava*
- 365 ASAP: Advancing Semantic Alignment Promotes Multi-Modal Manipulation Detecting and Grounding, *Zhenxing Zhang, Yaxiong Wang, Lechao Cheng, Zhun Zhong, Dan Guo, Meng Wang*
- 366 LamRA: Large Multimodal Model as Your Advanced Retrieval Assistant, *Yikun Liu, Yajie Zhang, Jiayin Cai, Xiaolong Jiang, Yao Hu, Jiangchao Yao, Yanfeng Wang, Weidi Xie*
- 367 Docopilot: Improving Multimodal Models for Document-Level Understanding, *Yuchen Duan, Zhe Chen, Yusong Hu, Weiyun Wang, Shenglong Ye, Botian Shi, Lewei Lu, Qibin Hou, Tong Lu, Hongsheng Li, Jifeng Dai, Wenhui Wang*
- 368 DocLayLLM: An Efficient Multi-modal Extension of Large Language Models for Text-rich Document Understanding, *Wenhui Liao, Jiapeng Wang, Hongliang Li, Chengyu Wang, Jun Huang, Lianwen Jin*
- 369 Diffusion Bridge: Leveraging Diffusion Model to Reduce the Modality Gap Between Text and Vision for Zero-Shot Image Captioning, *Jeong Ryoung Lee, Yejee Shin, Geonhui Son, Dosik Hwang*
- 370 ☆ Molmo and PixMo: Open Weights and Open Data for State-of-the-Art Vision-Language Models, *Matt Deitke, Christopher Clark, Sangho Lee, Rohun Tripathi, Yue Yang, Jae Sung Park, Mohammadreza Salehi, Niklas Muennighoff, Kyle Lo, Luca Soldaini, Jiasen Lu, Taira Anderson, Erin Bransom, Kiana Ehsani, Huang Ngo, YenSung Chen, Ajay Patel, Mark Yatskar, Chris Callison-Burch, Andrew Head, Rose Hendrix, Favyen Bastani, Eli VanderBilt, Nathan Lambert, Yvonne Chou, Arnavi Chheda, Jenna Sparks, Sam Skjonsberg, Michael Schmitz, Aaron Sarnat, Byron Bischoff, Pete Walsh, Chris Newell, Piper Wolters, Tanmay Gupta, Kuo-Hao Zeng, Jon Borchardt, Dirk Groeneveld, Crystal Nam, Sophie Lebrecht, Caitlin Wittliff, Carissa Schoenick, Oscar Michel, Ranjay Krishna, Luca Weihs, Noah A. Smith, Hannaneh Hajishirzi, Ross Girshick, Ali Farhadi, Aniruddha Kembhavi*
- 371 ☆ Towards Improved Text-Aligned Codebook Learning: Multi-Hierarchical Codebook-Text Alignment with Long Text, *Guotao Liang, Baoquan Zhang, Zhiyuan Wen, Junteng Zhao, Yunming Ye, Kola Ye, Yao He*
- 372 GOAL: Global-local Object Alignment Learning, *Hyungyu Choi, Young Kyun Jang, Chanho Eom*
- 373 FLAME: Frozen Large Language Models Enable Data-Efficient Language-Image Pre-training, *Anjia Cao, Xing Wei, Zhiheng Ma*
- 374 Data-Free Group-Wise Fully Quantized Winograd Convolution via Learnable Scales, *Shuokai Pan, Gerti Tuzi, Sudarshan Sreeram, Dibakar Gope*
- 375 ☆ VladVA: Discriminative Fine-tuning of LVLMMs, *Yassine Ouali, Adrian Bulat, Alexandros Xenos, Anestis Zaganidis, Ioannis Maniadi, Metaxas, Brais Martinez, Georgios Tzimiropoulos*
- 376 ☆ Galaxy Walker: Geometry-aware VLMs For Galaxy-scale Understanding, *Tianyu Chen, Xingcheng Fu, Yisen Gao, Haodong Qian, Yuecen Wei, Kun Yan, Haoyi Zhou, Jianxin Li*
- 377 ☆ NVILA: Efficient Frontier Visual Language Models, *Zhijian Liu, Ligeng Zhu, Baifeng Shi, Zhuoyang Zhang, Yuming Lou, Shang Yang, Haocheng Xi, Shiyi Cao, Yuxian Gu, Dacheng Li, Xiuyu Li, Haotian Tang, Yunhao Fang, Yukang Chen, Cheng-Yu Hsieh, De-An Huang, An-Chieh Cheng, Jinyi Hu, Sifei Liu, Ranjay Krishna, Pavlo Molchanov, Jan Kautz, Hongxu Yin, Song Han, Yao Lu*
- 378 Unveiling Visual Perception in Language Models: An Attention Head Analysis Approach, *Jing Bi, Junjia Guo, Yunlong Tang, Lianggong Bruce Wen, Zhang Liu, Chenliang Xu, Bingjie Wang*
- 379 BlueLM-V-3B: Algorithm and System Co-Design for Multimodal Large Language Models on Mobile Devices, *Xudong Lu, Yinghao Chen, Cheng Chen, Hui Tan, Boheng Chen, Yina Xie, Rui Hu, Guanxin Tan, Renzhou Wu, Yan Hu, Yi Zeng, Lei Wu, Liuyang Bian, Zhaoxiong Wang, Long Liu, Yanzhou Yang, Han Xiao, Aojun Zhou, Yafei Wen, Xiaoxin Chen, Shuai Ren, Hongsheng Li*

- 380 Multi-Layer Visual Feature Fusion in Multimodal LLMs: Methods, Analysis, and Best Practices, *Junyan Lin, Haoran Chen, Yue Fan, Yingqi Fan, Xin Jin, Hui Su, Jinlan Fu, Xiaoyu Shen*
- 381 Rethinking Vision-Language Model in Face Forensics: Multi-
 ○ Modal Interpretable Forged Face Detector, *Xiao Guo, Xiufeng Song, Yue Zhang, Xiaohong Liu, Xiaoming Liu*
- 382 MBQ: Modality-Balanced Quantization for Large Vision-Language Models, *Shiyao Li, Yingchun Hu, Xuefei Ning, Xihui Liu, Ke Hong, Xiaotao Jia, Xiuhong Li, Yaqi Yan, Pei Ran, Guohao Dai, Shengen Yan, Huazhong Yang, Yu Wang*
- 383 Align-KD: Distilling Cross-Modal Alignment Knowledge for Mobile Vision-Language Large Model Enhancement, *Qianhan Feng, Wenshuo Li, Tong Lin, Xinghao Chen*
- 384 VASparse: Towards Efficient Visual Hallucination Mitigation via Visual-Aware Token Sparsification, *Xianwei Zhuang, Zhihong Zhu, Yuxin Xie, Liming Liang, Yuexian Zou*
- 385 Stop Learning it all to Mitigate Visual Hallucination, Focus on the Hallucination Target. *Dokyoon Yoon, Youngsook Song, Woomyoung Park*
- 386 ICT: Image-Object Cross-Level Trusted Intervention for Mitigating Object Hallucination in Large Vision-Language Models, *Junzhe Chen, Tianshu Zhang, Shiyu Huang, Yuwei Niu, Linfeng Zhang, Lijie Wen, Xuming Hu*
- 387 Hyperbolic Safety-Aware Vision-Language Models, *Tobia Poppi, ☆ Tejaswi Kasarla, Pascal Mettes, Lorenzo Baraldi, Rita Cucchiara*
- 388 Forensics-Bench: A Comprehensive Forgery Detection Benchmark Suite for Large Vision Language Models, *Jin Wang, Chenghui Lv, Xian Li, Shichao Dong, Huadong Li, Kelu Yao, Chao Li, Wenqi Shao, Ping Luo*
- 389 Joint Vision-Language Social Bias Removal for CLIP, *Haoyu Zhang, Yangyang Guo, Mohan Kankanhalli*
- 390 Post-pre-training for Modality Alignment in Vision-Language Foundation Models, *Shin'ya Yamaguchi, Dewei Feng, Sekitoshi Kanai, Kazuki Adachi, Daiki Chijiwa*
- 391 Context-Aware Multimodal Pretraining, *Karsten Roth, ☆ Zeynep Akata, Dima Damen, Ivana Balazevic, Olivier J. Henaff*
- 392 Adaptive Parameter Selection for Tuning Vision-Language Models, *Yi Zhang, Yi-Xuan Deng, Meng-Hao Guo, Shi-Min Hu*
- 393 OpenSDI: Spotting Diffusion-Generated Images in the Open World, *Yabin Wang, Zhiwu Huang, Xiaopeng Hong*
- 394 SnowMaster: Comprehensive Real-world Image Desnowing via MLLM with Multi-Model Feedback Optimization, *Jianyu Lai, Sixiang Chen, Yunlong Lin, Tian Ye, Yun Liu, Song Fei, Zhaohu Xing, Hongtao Wu, Weiming Wang, Lei Zhu*
- 395 OverLoCK: An Overview-first-Look-Closely-next ConvNet with
 ○ Context-Mixing Dynamic Kernels, *Meng Lou, Yizhou Yu*
- 396 Do We Always Need the Simplicity Bias? Looking for Optimal
 ☆ Inductive Biases in the Wild, *Damien Teney, Liangze Jiang, Florin Gogianu, Ehsan Abbasnejad*
- 397 LibraGrad: Balancing Gradient Flow for Universally Better
 ○ Vision Transformer Attributions, *Faridoun Mehri, Mahdieh Soleymani Baghshah, Mohammad Taher Pilehvar*
- 398 SPARC: Score Prompting and Adaptive Fusion for Zero-Shot Multi-Label Recognition in Vision-Language Models, *Kevin Miller, Aditya Gangrade, Samarth Mishra, Kate Saenko, Venkatesh Saligrama*
- 399 Query Efficient Black-Box Visual Prompting with Subspace Learning, *Zhaogeng Liu, Haozhen Zhang, Hualin Zhang, Xingchen Li, Wanli Shi, Bin Gu, Yi Chang*
- 400 Plug-and-Play PPO: An Adaptive Point Prompt Optimizer Making SAM Greater, *Xueyu Liu, Rui Wang, Yexin Lai, Guangze Shi, Feixue Shao, Fang Hao, Jianan Zhang, Jia Shen, Yongfei Wu, Wen Zheng*
- 401 Discovering Hidden Visual Concepts Beyond Linguistic Input in Infant Learning, *Xueyi Ke, Satoshi Tsutsui, Yayun Zhang, Bihan Wen*
- 402 DA-VPT: Semantic-Guided Visual Prompt Tuning for Vision Transformers, *Li Ren, Chen Chen, Liqiang Wang, Kien Hua*
- 403 CoE: Chain-of-Explanation via Automatic Visual Concept Circuit Description and Polysemanticity Quantification, *Wenlong Yu, Qilong Wang, Chuang Liu, Dong Li, Qinghua Hu*
- 404 Prompt-CAM: Making Vision Transformers Interpretable for Fine-Grained Analysis, *Arpita Chowdhury, Dipanjyoti Paul, Zheda Mai, Jianyang Gu, Ziheng Zhang, Kazi Sajeed Mehrab, Elizabeth G. Campolongo, Daniel Rubenstein, Charles V. Stewart, Anuj Karpatne, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao*
- 405 Attention IoU: Examining Biases in CelebA using Attention Maps, *Aaron Serianni, Tyler Zhu, Olga Russakovsky, Vikram V. Ramaswamy*
- 406 ARKit LabelMaker: A New Scale for Indoor 3D Scene Understanding, ☆ *Guangda Ji, Silvan Weder, Francis Engelmann, Marc Pollefeys, Hermann Blum*
- 407 Seeing More with Less: Human-like Representations in Vision
 * Models, *Andrey Gidzov, Shimon Ullman, Daniel Harari*
- 408 Argus: A Compact and Versatile Foundation Model for Vision, *Weiming Zhuang, Chen Chen, Zhizhong Li, Sina Sajadmanesh, Jingtao Li, Jiabo Huang, Vikash Sehraw, Vivek Sharma, Hirotaka Shinozaki, Felan Carlo Garcia, Yihao Zhan, Naohiro Adachi, Ryoji Eki, Michael Spranger, Peter Stone, Lingjuan Lyu*
- 409 Test-Time Fine-Tuning of Image Compression Models for Multi-Task Adaptability, *Unki Park, Seongmoon Jeong, Youngchan Jang, Gyeong-Moon Park, Jong Hwan Ko*
- 410 L-SWAG: Layer-Sample Wise Activation with Gradients Information for Zero-Shot NAS on Vision Transformers, *Sofia Casarin, Sergio Escalera, Oswald Lanz*
- 411 NADER: Neural Architecture Design via Multi-Agent Collaboration, *Zekang Yang, Wang Zeng, Sheng Jin, Chen Qian, Ping Luo, Wentao Liu*
- 412 Quantization without Tears, *Minghao Fu, Hao Yu, Jie Shao, Junjie Zhou, Ke Zhu, Jianxin Wu*
- 413 Parallel Sequence Modeling via Generalized Spatial Propagation Network, *Hongjun Wang, Wonmin Byeon, Jiarui Xu, Jinwei Gu, Ka Chun Cheung, Xiaolong Wang, Kai Han, Jan Kautz, Sifei Liu*
- 414 MambaOut: Do We Really Need Mamba for Vision?, *Weihaoyu Yu, Xinchao Wang*
- 415 MobileMamba: Lightweight Multi-Receptive Visual Mamba Network, *Haoyang He, Jiangning Zhang, Yuxuan Cai, Hongxu Chen, Xiaobin Hu, Zhenye Gan, Yabiao Wang, Chengjie Wang, Yunsheng Wu, Lei Xie*
- 416 ComRoPE: Scalable and Robust Rotary Position Embedding Parameterized by Trainable Commuting Angle Matrices, *Hao Yu, Tangyu Jiang, Shuning Jia, Shannan Yan, Shunning Liu, Haolong Qian, Guanghao Li, Shuting Dong, Chun Yuan*
- 417 Associative Transformer, *Yuwei Sun, Hideya Ochiai, Zhirong Wu, Stephen Lin, Ryota Kanai*
- 418 Rashomon Sets for Prototypical-Part Networks: Editing Interpretable Models in Real-Time, *Jon Donnelly, Zhicheng Guo, Alina Jade Barnett, Hayden McTavish, Chaofan Chen, Cynthia Rudin*
- 419 SGC-Net: Stratified Granular Comparison Network for Open-Vocabulary HOI Detection, *Xin Lin, Chong Shi, Zuopeng Yang, Haojin Tang, Zhili Zhou*
- 420 CALICO: Part-Focused Semantic Co-Segmentation with Large Vision-Language Models, *Kiet A. Nguyen, Adheesh Juvekar, Tianjiao Yu, Muntasir Wahed, Ismini Lourentzou*
- 421 Understanding Fine-tuning CLIP for Open-vocabulary Semantic Segmentation in Hyperbolic Space, *Zelin Peng, Zhengqin Xu, Zhilin Zeng, Changsong Wen, Yu Huang, Menglin Yang, Feilong Tang, Wei Shen*
- 422 Scaling up Image Segmentation across Data and Tasks, ☆ *Pei Wang, Zhaowei Cai, Hao Yang, Ashwin Swaminathan, R. Manmatha, Stefano Soatto*
- 423 DSV-LFS: Unifying LLM-Driven Semantic Cues with Visual Features for Robust Few-Shot Segmentation, *Amin Karimi, Charalambos Poullis*
- 424 Rethinking Query-based Transformer for Continual Image Segmentation, *Yuchen Zhu, Cheng Shi, Dingyou Wang, Jiajin Tang, Zhengxuan Wei, Yu Wu, Guanbin Li, Sibeil Yang*
- 425 Universal Domain Adaptation for Semantic Segmentation, *Seun-An Choe, Keon-Hee Park, Jinwoo Choi, Gyeong-Moon Park*
- 426 The Devil is in Low-Level Features for Cross-Domain Few-Shot Segmentation, *Yuhan Liu, Yixiong Zou, Yuhua Li, Ruixuan Li*
- 427 EZSR: Event-based Zero-Shot Recognition, *Yan Yang, Liyuan Pan, Dongxu Li, Liu Liu*
- 428 Single Domain Generalization for Few-Shot Counting via Universal Representation Matching, *Xianing Chen, Si Huo, Borui Jiang, Hailin Hu, Xinghao Chen*

- 429 Recover and Match: Open-Vocabulary Multi-Label Recognition through Knowledge-Constrained Optimal Transport, *Hao Tan, Zichang Tan, Jun Li, Aijian Liu, Jun Wan, Zhen Lei*
- 430 Classifier-guided CLIP Distillation for Unsupervised Multi-label Classification, *Dongseob Kim, Hyunjung Shim*
- 431 SimLTD: Simple Supervised and Semi-Supervised Long-Tailed Object Detection, *Phi Vu Tran*
- 432 Percept, Memory, and Imagine: World Feature Simulating for Open-Domain Unknown Object Detection, *Aming Wu, Cheng Deng*
- 433 Large Self-Supervised Models Bridge the Gap in Domain Adaptive Object Detection, *Marc-Antoine Lavoie, Anas Mahmoud, Steven L. Waslander*
- 434 MI-DETR: An Object Detection Model with Multi-time Inquiries Mechanism, *Zhixiong Nan, Xianghong Li, Jifeng Dai, Tao Xiang*
- 435 SET: Spectral Enhancement for Tiny Object Detection, *Huixin Sun, Runqi Wang, Yanjing Li, Linlin Yang, Shaohui Lin, Xianbin Cao, Baochang Zhang*
- 436 Leveraging Perturbation Robustness to Enhance Out-of-Distribution Detection, *Wenxi Chen, Raymond A. Yeh, Shaoshuai Mou, Yan Gu*
- 437 PIAD: Pose and Illumination agnostic Anomaly Detection, *Kaichen Yang, Junjie Cao, Zeyu Bai, Zhixun Su, Andrea Tagliasacchi*
- 438 AA-CLIP: Enhancing Zero-Shot Anomaly Detection via Anomaly-Aware CLIP, *Wenxin Ma, Xu Zhang, Qingsong Yao, Fenghe Tang, Chenxu Wu, Yingtai Li, Rui Yan, Zihang Jiang, S. Kevin Zhou*
- 439 AnomalyNCD: Towards Novel Anomaly Class Discovery in Industrial Scenarios, *Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, Yu Zhou*
- 440 One-for-More: Continual Diffusion Model for Anomaly Detection, *Xiaofan Li, Xin Tan, Zhuo Chen, Zhizhong Zhang, Ruixin Zhang, Rizen Guo, Guanna Jiang, Yulong Chen, Yanyun Qu, Lizhuang Ma, Yuan Xie*
- 441 GeoMM: On Geodesic Perspective for Multi-modal Learning, *Shibin Mei, Hang Wang, Bingbing Ni*
- 442 HOT: Hadamard-based Optimized Training, *Seonggon Kim, Juncheol Shin, Seung-taek Woo, Eunhyeok Park*
- 443 DELT: A Simple Diversity-driven EarlyLate Training for Dataset Distillation, *Zhiqiang Shen, Ammar Sherif, Zeyuan Yin, Shitong Shao*
- 444 Flexible Group Count Enables Hassle-Free Structured Pruning, *Jiamu Zhang, Shaochen Zhong, Andrew Ye, Zirui Liu, Sebastian Zhao, Kaixiong Zhou, Li Li, Soo-Hyun Choi, Rui Chen, Xia Hu, Shuai Xu, Vipin Chaudhary*
- 445 WAVE: Weight Templates for Adaptive Initialization of Variable-sized Models, *Fu Feng, Yucheng Xie, Jing Wang, Xin Geng*
- 446 IterIS: Iterative Inference-Solving Alignment for LoRA Merging, *Hongxu Chen, Zhen Wang, Runshi Li, Bowei Zhu, Long Chen*
- 447 Boosting Domain Incremental Learning: Selecting the Optimal Parameters is All You Need, *Qiang Wang, Xiang Song, Yuhang He, Jizhou Han, Chenhao Ding, Xinyuan Gao, Yihong Gong*
- 448 Ferret: An Efficient Online Continual Learning Framework under Varying Memory Constraints, *Yuhao Zhou, Yuxin Tian, Jindi Lv, Mingjia Shi, Yuanxi Li, Qing Ye, Shuhao Zhang, Jiancheng Lv*
- 449 Learning Conditional Space-Time Prompt Distributions for Video * Class-Incremental Learning, *Xiaohan Zou, Wenchao Ma, Shu Zhao*
- 450 Handling Spatial-Temporal Data Heterogeneity for Federated Continual Learning via Tail Anchor, *Hao Yu, Xin Yang, Le Zhang, Hanlin Gu, Tianrui Li, Lixin Fan, Qiang Yang*
- 451 Adapter Merging with Centroid Prototype Mapping for Scalable Class-Incremental Learning, *Takuma Fukuda, Hiroshi Kera, Kazuhiko Kawamoto*
- 452 Order-Robust Class Incremental Learning: Graph-Driven Dynamic Similarity Grouping, *Guannan Lai, Yujie Li, Xiangkun Wang, Junbo Zhang, Tianrui Li, Xin Yang*
- 453 When Domain Generalization meets Generalized Category Discovery: An Adaptive Task-Arithmetic Driven Approach, *Vaibhav Rathore, Shubhranil B, Saikat Dutta, Sarthak Mehrotra, Zsolt Kira, Biplob Banerjee*
- 454 Link-based Contrastive Learning for One-Shot Unsupervised Domain Adaptation, *Yue Zhang, Mingyue Bin, Yuyang Zhang, Zhongyuan Wang, Zhen Han, Chao Liang*
- 455 Distinguish Then Exploit: Source-free Open Set Domain * Adaptation via Weight Barcode Estimation and Sparse Label Assignment, *Weiming Liu, Jun Dan, Fan Wang, Xinting Liao, Junhao Dong, Hua Yu, Shunjie Dong, Lianying Qi*
- 456 Instance-wise Supervision-level Optimization in Active Learning, *Shinnosuke Matsuo, Riku Togashi, Ryoma Bise, Seichi Uchida, Masahiro Nomura*
- 457 Towards Source-Free Machine Unlearning, *Sk Miraj Ahmed, Umit Yigit Basaran, Dripta S. Raychaudhuri, Arindam Dutta, Rohit Kundu, Fahim Faisal Niloy, Basak Guler, Amit K. Roy-Chowdhury*
- 458 Sufficient Invariant Learning for Distribution Shift, *Taero Kim, Subeen Park, Sungjun Lim, Yonghan Jung, Krikamol Muandet, Kyungwoo Song*
- 459 CADRef: Robust Out-of-Distribution Detection via Class-Aware Decoupled Relative Feature Leveraging, *Zhiwei Ling, Yachen Chang, Hailiang Zhao, Xinkui Zhao, Kingsum Chow, Shuiguang Deng*
- 460 Federated Learning with Domain Shift Eraser, *Zheng Wang, Zihui Wang, Zheng Wang, Xiaoliang Fan, Cheng Wang*
- 461 AFL: A Single-Round Analytic Approach for Federated Learning with Pre-trained Models, *Run He, Kai Tong, Di Fang, Han Sun, Ziqian Zeng, Haoran Li, Tianyi Chen, Huiping Zhuang*
- 462 Fortifying Federated Learning Towards Trustworthiness via Auditable Data Valuation and Verifiable Client Contribution, *K Naveen Kumar, Ranjeet Ranjan Jha, C Krishna Mohan, Ravindra Babu Tallamraju*
- 463 ESC: Erasing Space Concept for Knowledge Deletion, *Tae-Young Lee, Sundong Park, Minwoo Jeon, Hyoseok Hwang, Gyeong-Moon Park*
- 464 Deterministic Certification of Graph Neural Networks against Graph Poisoning Attacks with Arbitrary Perturbations, *Jiate Li, Meng Pang, Yun Dong, Binghui Wang*
- 465 Simplification Is All You Need against Out-of-Distribution Overconfidence, *Keke Tang, Chao Hou, Weilong Peng, Xiang Fang, Zhize Wu, Yongwei Nie, Wenping Wang, Zhihong Tian*
- 466 MOS-Attack: A Scalable Multi-objective Adversarial Attack Framework, *Ping Guo, Cheng Gong, Xi Lin, Fei Liu, Zhichao Lu, Qingfu Zhang, Zhenkun Wang*
- 467 Automated Proof of Polynomial Inequalities via Reinforcement Learning, *Banglong Liu, Niuniu Qi, Xia Zeng, Lydia Dehbi, Zhengfeng Yang*
- 468 Deep Fair Multi-View Clustering with Attention KAN, *HaiMing Xu, Qianqian Wang, Boyue Wang, Quanxue Gao*
- 469 Imputation-free and Alignment-free: Incomplete Multi-view Clustering Driven by Consensus Semantic Learning, *Yuzhuo Dai, Jiaqi Jin, Zhibin Dong, Siwei Wang, Xinwang Liu, En Zhu, Xihong Yang, Xinbiao Gan, Yu Feng*
- 470 Improve Representation for Imbalanced Regression through Geometric Constraints, *Zijian Dong, Yilei Wu, Chongyao Chen, Yingtian Zou, Yichi Zhang, Juan Helen Zhou*
- 471 MODfinity: Unsupervised Domain Adaptation with Multimodal Information Flow Intertwining, *Shanglin Liu, Jianming Lv, Jingdan Kang, Huaidong Zhang, Zequan Liang, Shengfeng He*
- 472 Distilled Prompt Learning for Incomplete Multimodal Survival Prediction, *Yingxue Xu, Fengtao Zhou, Chenyu Zhao, Yihui Wang, Can Yang, Hao Chen*
- 473 LMO: Linear Mamba Operator for MRI Reconstruction, *Wei Li, Jiawei Jiang, Jie Wu, Kaihao Yu, Jianwei Zheng*
- 474 CXPMRG-Bench: Pre-training and Benchmarking for X-ray Medical Report Generation on CheXpert Plus Dataset, *Xiao Wang, Fuling Wang, Yuehang Li, Qingchuan Ma, Shiao Wang, Bo Jiang, Jin Tang*
- 475 SlideChat: A Large Vision-Language Assistant for Whole-Slide Pathology Image Understanding, *Ying Chen, Guoan Wang, Yuanfeng Ji, Yanjun Li, Jin Ye, Tianbin Li, Ming Hu, Rongshan Yu, Yu Qiao, Junjun He*
- 476 Learning Heterogeneous Tissues with Mixture of Experts for Gigapixel Whole Slide Images, *Junxian Wu, Minheng Chen, Xinyi Ke, Tianwang Xun, Xiaoming Jiang, Hongyu Zhou, Lizhi Shao, Youyong Kong*
- 477 Patient-Level Anatomy Meets Scanning-Level Physics: Personalized Federated Low-Dose CT Denoising Empowered by Large Language Model, *Ziyuan Yang, Yingyu Chen, Zhiwen Wang, Hongming Shan, Yang Chen, Yi Zhang*

- 478 Multi-modal Vision Pre-training for Medical Image Analysis,
* *Shaohao Rui, Lingzhi Chen, Zhenyu Tang, Lilong Wang, Mianxin Liu, Shaoting Zhang, Xiaosong Wang*
- 479 Steady Progress Beats Stagnation: Mutual Aid of Foundation and Conventional Models in Mixed Domain Semi-Supervised Medical Image Segmentation, *Qinghe Ma, Jian Zhang, Zekun Li, Lei Qi, Qian Yu, Yinghuan Shi*
- 480 Revisiting MAE Pre-training for 3D Medical Image Segmentation,
* *Tassilo Wald, Constantin Ulrich, Stanislav Lukyanenko, Andrei Goncharov, Alberto Paderno, Maximilian Miller, Leander Maerkisch, Paul Jaeger, Klaus Maier-Hein*
- 481 SuperLightNet: Lightweight Parameter Aggregation Network for Multimodal Brain Tumor Segmentation, *Feng Yu, Jiacheng Cao, Li Liu, Minghua Jiang*
- 482 EchoONE: Segmenting Multiple Echocardiography Planes in One Model, *Jiongtong Hu, Wufeng Xue, Jun Cheng, Yingying Liu, Wei Zhuo, Dong Ni*
- 483 AeSPa : Attention-guided Self-supervised Parallel Imaging for MRI Reconstruction, *Jinho Joo, Hyeseong Kim, Hyeyeon Won, Deukhee Lee, Taejoon Eo, Dosik Hwang*
- 484 SACB-Net: Spatial-awareness Convolutions for Medical Image Registration, *Xinxing Cheng, Tianyang Zhang, Wenqi Lu, Qingjie Meng, Alejandro F. Frangi, Jinming Duan*
- 485 Segmenting Maxillofacial Structures in CBCT Volumes,
☆ *Federico Bolelli, Kevin Marchesini, Niels van Nistelrooij, Luca Lumetti, Vittorio Pipoli, Elisa Ficarra, Shankeeth Vinayahalingam, Costantino Grana*

10:30 - 18:00 Art Program (ExHall A1)

10:30 - 12:30 Demos (ExHall B)

- EdgeVidSum: Real-Time Personalized Video Summarization at the Edge, *Ghulam Mujtaba, Eun-Seok Ryu*
- Virtual Try-Off: Your Garment, Reimagined, *Riza Veliloglu*
- DynaMem: Online Dynamic Spatio-Semantic Memory for Open World Mobile Manipulation, *Peiqi Liu, Zhanqiu Guo, Soumith Chintala, Chris Paxton, Nur Muhammad "Mahi" Shafiullah, Lerrel Pinto*
- Realtime Medical Imaging Algorithms, *Halid Ziya Yerebakan, Yoshihisa Shinagawa, Gerardo Hermosillo-Valadez*
- 4D-SAL: Zero-Shot 4D Panoptic Lidar Segmentation, *Aljosa Osep, Tim Meinhardt, Laura Leal-Taixe*
- Dynamic Gaussian Splatting in VR, *Jonathon Luiten*
- GeoPI@ntNet: Species and Habitat Distribution Modeling at Very-high Resolution, *Lukas Picek, Cesar Leblanc, Pierre Bonnet, Rémi Palard, Maximilien Servajean, Alexis Joly*
- DSAIL-Orthopedia: A computer vision-based software for automated measurements of flexion angle and lower limb alignment, *Antony Gitau, Clinton Mwangi, Kayuyu Mwaura, Gachathi Wanjema, Ciira Maina*
- Real-time Interactive Human Motion Diffusion Network, *Mingyi Shi, Taku Komura*
- Inference-time Steering of Rectified Flow Models in the Vector Field for Controlled Visual Generations, *Maitreya Patel, Song Wen, Dimitris N. Metaxas, Yezhou Yang*
- VinTAGE: Joint Video and Text Conditioning for Holistic Audio Generation, *Saksham Singh Kushwaha, Yapeng Tian*
- Pixel-aligned RGB-NIR Stereo Imaging and Dataset for Robot Vision, *Jinnyeong Kim, Seung-Hwan Baek*

11:00 / 17:00 Art Gallery Tour with Curator, Luba Elliott (30 mins each), (ExHall A1)

11:30 - 13:30 LUNCH (ExHall D)

13:00 - 14:30 Oral Session 2A: 3D Computer Vision (Karl Dean Ballroom)

📄 - Award candidate paper

- FoundationStereo: Zero-Shot Stereo Matching, *Bowen Wen, Matthew Trepte, Joseph Aribido, Jan Kautz, Orazio Gallo, Stan Birchfield*
- MoGe: Unlocking Accurate Monocular Geometry Estimation for Open-Domain Images with Optimal Training Supervision, *Ruicheng Wang, Sicheng Xu, Cassie Dai, Jianfeng Xiang, Yu Deng, Xin Tong, Jiaolong Yang*

- Multi-view Reconstruction via SfM-guided Monocular Depth Estimation, *Haoyu Guo, He Zhu, Sida Peng, Haotong Lin, Yunzhi Yan, Tao Xie, Wenguan Wang, Xiaowei Zhou, Hujun Bao*
- MV-DUST3R+: Single-Stage Scene Reconstruction from Sparse Views In 2 Seconds, *Zhenggang Tang, Yuchen Fan, Dilin Wang, Hongyu Xu, Rakesh Ranjan, Alexander Schwing, Zhicheng Yan*
- VGTT: Visual Geometry Grounded Transformer, *Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, David Novotny*
- CraftsMan3D: High-fidelity Mesh Generation with 3D Native Diffusion and Interactive Geometry Refiner, *Wei Yu Li, Jiarui Liu, Hongyu Yan, Rui Chen, Yixun Liang, Xuelin Chen, Ping Tan, Xiaoxiao Long*

13:00 - 14:30 Oral Session 2B: Human Motion (ExHall A2)

- CAP4D: Creating Animatable 4D Portrait Avatars with Morphable Multi-View Diffusion Models, *Felix Taubner, Ruihang Zhang, Mathieu Tuli, David B. Lindell*
- Reanimating Images using Neural Representations of Dynamic Stimuli, *Jacob Yeung, Andrew F. Luo, Gabriel Sarch, Margaret M. Henderson, Deva Ramanan, Michael J. Tarr*
- EgoLM: Multi-Modal Language Model of Egocentric Motions, *Fangzhou Hong, Vladimir Guzov, Hyo Jin Kim, Yuting Ye, Richard Newcombe, Ziwei Liu, Lingni Ma*
- Reconstructing Humans with a Biomechanically Accurate Skeleton, *Yan Xia, Xiaowei Zhou, Etienne Vouga, Qixing Huang, Georgios Pavlakos*
- MEGA: Masked Generative Autoencoder for Human Mesh Recovery, *Guénolé Fiche, Simon Leglaive, Xavier Alameda-Pineda, Francesc Moreno-Noguer*
- TokenHSI: Unified Synthesis of Physical Human-Scene Interactions through Task Tokenization, *Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, Jingbo Wang*

13:00 - 14:30 Oral Session 2C: Temporal Modeling and Action Recognition (Davidson Ballroom)

- Descriptor-In-Pixel : Point-Feature Tracking For Pixel Processor Arrays, *Laurie Bose, Jianing Chen, Piotr Dudek*
- Temporally Consistent Object-Centric Learning by Contrasting Slots, *Anna Manasyan, Maximilian Seitzer, Filip Radovic, Georg Martius, Andrii Zadaianchuk*
- Temporal Alignment-Free Video Matching for Few-shot Action Recognition, *SuBeen Lee, WonJun Moon, Hyun Seok Seong, Jae-Pil Heo*
- Closed-Loop Supervised Fine-Tuning of Tokenized Traffic Models, *Zhejun Zhang, Peter Karkus, Maximilian Igl, Wenhao Ding, Yuxiao Chen, Boris Ivanovic, Marco Pavone*
- The PanAf-FGBG Dataset: Understanding the Impact of Backgrounds in Wildlife Behaviour Recognition, *Otto Brookes, Maksim Kukushkin, Majid Mirmehdi, Colleen Stephens, Paula Dieguez, Thurston C. Hicks, Sorrel Jones, Kevin Lee, Maureen S. McCarthy, Amelia Meier, Emmanuelle Normand, Erin G. Wessling, Roman M. Wittig, Kevin Langergraber, Klaus Zuberbühler, Lukas Boesch, Thomas Schmid, Mimi Arandjelovic, Hjalmar Kühl, Tilo Burghardt*
- Rethinking Spiking Self-Attention Mechanism: Implementing α -XNOR Similarity Calculation in Spiking Transformers, *Yichen Xiao, Shuai Wang, Dehao Zhang, Wenjie Wei, Yimeng Shan, Xiaoli Liu, Yulin Jiang, Malu Zhang*

13:30 - 14:30 Art Panel (209 ABC)

14:30 - 14:45 Courtesy Break

14:45 - 15:45 KEYNOTE 1 - Harry Shum; Exploring the Low Altitude Airspace: From Natural Resource to Economic Engine (Karl Dean Ballroom)

15:30 - 16:00 Poster Setup (ExHall D)

16:00 - 18:00 Poster Session 2 & Exhibit Hall w/ Coffee Break (ExHall D)

- * - Highlight paper 📄 - Award candidate paper
○ - Oral Paper ☆ - Outstanding Reviewer

- 1 EMOVA: Empowering Language Models to See, Hear and Speak with Vivid Emotions, *Kai Chen, Yunhao Gou, Runhui Huang, Zhili Liu, Daxin Tan, Jing Xu, Chunwei Wang, Yi Zhu, Yihan Zeng, Kuo Yang, Dingdong Wang, Kun Xiang, Haoyuan Li, Haoli Bai, Jianhua Han, Xiaohui Li, Wei Ke, Nian Xie, Yu Zhang, James T. Kwok, Hengshuang Zhao, Xiaodan Liang, Dit-Yan Yeung, Xiao Chen, Zhenguo Li, Wei Zhang, Qun Liu, Lanqing Hong, Lu Hou, Hang Xu*
- 2 Let's Chorus: Partner-aware Hybrid Song-Driven 3D Head Animation, *Xiumei Xie, Zikai Huang, Wenhao Xu, Peng Xiao, Xuemiao Xu, Huaidong Zhang*
- 3 KeyFace: Expressive Audio-Driven Facial Animation for Long Sequences via KeyFrame Interpolation, *Antoni Bigata, Michał Stypułkowski, Rodrigo Mira, Stella Bounareli, Konstantinos Vougioukas, Zoe Landgraf, Nikita Drobyshev, Maciej Zieba, Stavros Petridis, Maja Pantic*
- 4 EchoMimicV2: Towards Striking, Simplified, and Semi-Body Human Animation, *Rang Meng, Xingyu Zhang, Yuming Li, Chenguang Ma*
- 5 X-Dyna: Expressive Dynamic Human Image Animation, *Di Chang, Hongyi Xu, You Xie, Yipeng Gao, Zhengfei Kuang, Shengqu Cai, Chenxu Zhang, Guoxian Song, Chao Wang, Yichun Shi, Zeyuan Chen, Shijie Zhou, Linjie Luo, Gordon Wetzstein, Mohammad Soleymani*
- 6 Lux Post Facto: Learning Portrait Performance Relighting with Conditional Video Diffusion and a Hybrid Dataset, *Yiqun Mei, Mingming He, Li Ma, Julien Philip, Wenqi Xian, David M George, Xueming Yu, Gabriel Dedic, Ahmet Levent Taşel, Ning Yu, Vishal M. Patel, Paul Debevec*
- 7 Monocular and Generalizable Gaussian Talking Head Animation, *Shengjie Gong, Haojie Li, Jiapeng Tang, Dongming Hu, Shuangping Huang, Hao Chen, Tianshui Chen, Zhuoman Liu*
- 8 FATE: Full-head Gaussian Avatar with Textural Editing from Monocular Video, *Jiawei Zhang, Zijian Wu, Zhiyang Liang, Yicheng Gong, Dongfang Hu, Yao Yao, Xun Cao, Hao Zhu*
- 9 CAP4D: Creating Animatable 4D Portrait Avatars with Morphable Multi-View Diffusion Models, *Felix Taubner, Ruihang Zhang, Mathieu Tuli, David B. Lindell*
- 10 GAF: Gaussian Avatar Reconstruction from Monocular Videos via Multi-view Diffusion, *Jiapeng Tang, Davide Davoli, Tobias Kirschstein, Liam Schoneveld, Matthias Nießner*
- 11 Vid2Avatar-Pro: Authentic Avatar from Videos in the Wild via Universal Prior, *Chen Guo, Junxuan Li, Yash Kant, Yaser Sheikh, Shunsuke Saito, Chen Cao*
- 12 SinGS: Animatable Single-Image Human Gaussian Splats with Kinematic Priors, *Yufan Wu, Xuanhong Chen, Wen Li, Shunran Jia, Hualiang Wei, Kairui Feng, Jialiang Chen, Yuhao Li, Ang He, Weimin Zhang, Bingbing Ni, Wenjun Zhang*
- 13 EasyCraft: A Robust and Efficient Framework for Automatic Avatar Crafting, *Suzhen Wang, Weijie Chen, Wei Zhang, Minda Zhao, Lincheng Li, Rongsheng Zhang, Zhipeng Hu, Xin Yu*
- 14 RigGS: Rigging of 3D Gaussians for Modeling Articulated Objects in Videos, *Yuxin Yao, Zhi Deng, Junhui Hou*
- 15 Learning Person-Specific Animatable Face Models from In-the-Wild Images via a Shared Base Model, *Yuxiang Mao, Zhenfeng Fan, Zhilie Zhang, Zhiheng Zhang, Shihong Xia*
- 16 ControlFace: Harnessing Facial Parametric Control for Face Rigging, *Wooseok Jang, Youngjun Hong, Geonho Cha, Seungryong Kim*
- 17 HiFi-Portrait: Zero-shot Identity-preserved Portrait Generation with High-fidelity Multi-face Fusion, *Yifang Xu, Benxiang Zhai, Yunzhuo Sun, Ming Li, Yang Li, Sidan Du*
- 18 DeCloTH: Decomposable 3D Cloth and Human Body Reconstruction from a Single Image, *Hyeongjin Nam, Donghwan Kim, Jeongtaek Oh, Kyoung Mu Lee*
- 19 Disentangled Pose and Appearance Guidance for Multi-Pose Generation, *Tengfei Xiao, Yue Wu, Yuelong Li, Can Qin, Maoguo Gong, Qiguang Miao, Wenping Ma*
- 20 Touch2Shape: Touch-Conditioned 3D Diffusion for Shape Exploration and Reconstruction, *Yuanbo Wang, Zhaoxuan Zhang, Jiajin Qiu, Dilong Sun, Zhengyu Meng, Xiaopeng Wei, Xin Yang*
- 21 MangaNinja: Line Art Colorization with Precise Reference Following, *Zhiheng Liu, Ka Leong Cheng, Xi Chen, Jie Xiao, Hao Ouyang, Kai Zhu, Yu Liu, Yujun Shen, Qifeng Chen, Ping Luo*
- 22 HVI: A New Color Space for Low-light Image Enhancement, *Qingsen Yan, Yixu Feng, Cheng Zhang, Guansong Pang, Kangbiao Shi, Peng Wu, Wei Dong, Jinqiu Sun, Yanning Zhang*
- 23 Flash-Split: 2D Reflection Removal with Flash Cues and Latent Diffusion Separation, *Tianfu Wang, Mingyang Xie, Haoming Cai, Sachin Shah, Christopher A. Metzler*
- 24 Noise Modeling in One Hour: Minimizing Preparation Efforts for Self-supervised Low-Light RAW Image Denoising, *Feiran Li, Haiyang Jiang, Daisuke Iso*
- 25 Quad-Pixel Image Defocus Deblurring: A New Benchmark and Model, *Hang Chen, Yin Xie, Xiaoxiu Peng, Lihu Sun, Wenkai Su, Xiaodong Yang, Chengming Liu*
- 26 ScribbleLight: Single Image Indoor Relighting with Scribbles, *Jun Myeong Choi, Annie Wang, Pieter Peers, Anand Bhattad, Roni Sengupta*
- 27 Hearing Anywhere in Any Environment, *Xiulong Liu, Anurag Kumar, Paul Calamia, Sebastia V. Amengual, Calvin Murdock, Ishwarya Ananthabhotla, Philip Robinson, Eli Shlizerman, Vamsi Krishna Ithapu, Ruohan Gao*
- 28 EnvGS: Modeling View-Dependent Appearance with Environment Gaussian, *Tao Xie, Xi Chen, Zhen Xu, Yiman Xie, Yudong Jin, Yujun Shen, Sida Peng, Hujun Bao, Xiaowei Zhou*
- 29 Geometry Field Splatting with Gaussian Surfels, *Kaiwen Jiang, Venkataram Sivaram, Cheng Peng, Ravi Ramamoorthi*
- 30 Locally Orderless Images for Optimization in Differentiable Rendering, *Ishit Mehta, Manmohan Chandraker, Ravi Ramamoorthi*
- 31 Channel-wise Noise Scheduled Diffusion for Inverse Rendering in Indoor Scenes, *JunYong Choi, Min-cheol Sagong, SeokYeong Lee, Seung-Won Jung, Ig-Jae Kim, Junghyun Cho*
- 32 Feature-Preserving Mesh Decimation for Normal Integration, *Moritz Heep, Sven Behnke, Eduard Zell*
- 33 SGCR: Spherical Gaussians for Efficient 3D Curve Reconstruction, *Xinran Yang, Donghao Ji, Yuanqi Li, Jie Guo, Yanwen Guo, Junyuan Xie*
- 34 AMR-Transformer: Enabling Efficient Long-range Interaction for Complex Neural Fluid Simulation, *Zeyi Xu, Jinfan Liu, Kuangxu Chen, Ye Chen, Zhangli Hu, Bingbing Ni*
- 35 MaRI: Material Retrieval Integration across Domains, *Jianhui Wang, Zhifei Yang, Yangfan He, Huixiong Zhang, Yuxuan Chen, Jingwei Huang*
- 36 Spherical Manifold Guided Diffusion Model for Panoramic Image Generation, *Xiancheng Sun, Mai Xu, Shengxi Li, Senmao Ma, Xin Deng, Lai Jiang, Gang Shen*
- 37 MeshGen: Generating PBR Textured Mesh with Render-Enhanced Auto-Encoder and Generative Data Augmentation, *Zilong Chen, Yikai Wang, Wenqiang Sun, Feng Wang, Yiwen Chen, Huaping Liu*
- 38 RASP: Revisiting 3D Anamorphic Art for Shadow-Guided Packing of Irregular Objects, *Soumyaratna Debnath, Ashish Tiwari, Kaustubh Sadekar, Shanmuganathan Raman*
- 39 Twiner: Shining Light on Digital Twins in a Few Snaps, *Jesus Zarzar, Tom Monnier, Roman Shapovalov, Andrea Vedaldi, David Novotny*
- 40 CraftsMan3D: High-fidelity Mesh Generation with 3D Native Diffusion and Interactive Geometry Refiner, *Wei Yu Li, Jiarui Liu, Hongyu Yan, Rui Chen, Yixun Liang, Xuelin Chen, Ping Tan, Xiaoxiao Long*
- 41 Kiss3DGen: Repurposing Image Diffusion Models for 3D Asset Generation, *Jiantao Lin, Xin Yang, Meixi Chen, Yingjie Xu, Dongyu Yan, Leyi Wu, Xinli Xu, Lie Xu, Shunsi Zhang, Ying-Cong Chen*
- 42 PartGen: Part-level 3D Generation and Reconstruction with Multi-view Diffusion Models, *Minghao Chen, Roman Shapovalov, Iro Laina, Tom Monnier, Jianyuan Wang, David Novotny, Andrea Vedaldi*
- 43 FreeScene: Mixed Graph Diffusion for 3D Scene Synthesis from Free Prompts, *Tongyuan Bai, Wangyuanfan Bai, Dong Chen, Tieru Wu, Manyi Li, Rui Ma*

- 44 Reference-Based 3D-Aware Image Editing with Triplanes,
* Bahri Batuhan Bilecen, Yigit Yalin, Ning Yu, Aysegul Dundar
- 45 WonderWorld: Interactive 3D Scene Generation from a Single
* Image, Hong-Xing Yu, Haoyi Duan, Charles Herrmann,
William T. Freeman, Jiajun Wu
- 46 UVGS: Reimagining Unstructured 3D Gaussian Splatting using
☆ UV Mapping, Aashish Rai, Dilin Wang, Mihir Jain,
Nikolaos Sarafianos, Kefan Chen, Srinath Sridhar, Aayush Prakash
- 47 3D-GSW: 3D Gaussian Splatting for Robust Watermarking,
Youngdong Jang, Hyunje Park, Feng Yang, Heeju Ko, Euijin Choo,
Sangpil Kim
- 48 PUP 3D-GS: Principled Uncertainty Pruning for 3D Gaussian
☆ Splatting, Alex Hanson, Allen Tu, Vasu Singla, Mayuka
Jayawardhana, Matthias Zwicker, Tom Goldstein
- 49 Gaussian Splatting for Efficient Satellite Image Photogrammetry,
☆ Luca Savant Aira, Gabriele Facciolo, Thibaud Ehret
- 50 HyperGS: Hyperspectral 3D Gaussian Splatting, Christopher
Thirgood, Oscar Mendez, Erin Ling, Jon Storey, Simon Hadfield
- 51 RoGSplat: Learning Robust Generalizable Human Gaussian
Splatting from Sparse Multi-View Images, Junjin Xiao, Qing
Zhang, Yonewei Nie, Lei Zhu, Wei-Shi Zheng
- 52 GaussHDR: High Dynamic Range Gaussian Splatting via
Learning Unified 3D and 2D Local Tone Mapping, Jinfeng Liu,
Lingtong Kong, Bo Li, Dan Xu
- 53 MATCha Gaussians: Atlas of Charts for High-Quality Geometry
* and Photorealism From Sparse Views, Antoine Guedon,
☆ Tomoki Ichikawa, Kohei Yamashita, Ko Nishino
- 54 Exploiting Deblurring Networks for Radiance Fields,
☆ Haeyun Choi, Heemin Yang, Janghyeok Han, Sunghyun Cho
- 55 Compositional Neural Scene Reconstruction with Generative
Diffusion Prior, Junfeng Ni, Yu Liu, Ruijie Lu, Zirui Zhou, Song-
Chun Zhu, Yixin Chen, Siyuan Huang
- 56 MET3R: Measuring Multi-View Consistency in Generated
☆ Images, Mohammad Asim, Christopher Wewer, Thomas Wimmer,
Berni Schiele, Jan Eric Lenssen
- 57 MV-DUST3R+: Single-Stage Scene Reconstruction from Sparse
○ Views In 2 Seconds, Zhenggang Tang, Yuchen Fan, Dilin Wang,
Hongyu Xu, Rakesh Ranjan, Alexander Schwing, Zhicheng Yan
- 58 MVGenMaster: Scaling Multi-View Generation from Any Image
via 3D Priors Enhanced Diffusion Model, Chenjie Cao, Chaohui
Yu, Shang Liu, Fan Wang, Xiangyang Xue, Yanwei Fu
- 59 ERUPT: Efficient Rendering with Unposed Patch
Transformer, Maxim V. Shugaev, Vincent Chen, Maxim
Karrenbach, Kyle Ashley, Bridget Kennedy, Naresh P. Cuntoor
- 60 Satellite to GroundScape - Large-scale Consistent Ground View
Generation from Satellite Views, Ningli Xu, Rongjun Qin
- 61 GenFusion: Closing the Loop between Reconstruction and
Generation via Videos, Sibow Wu, Congrong Xu, Binbin Huang,
Andreas Geiger, Anpei Chen
- 62 Scene Splat: Momentum 3D Scene Generation from Single
Image with Video Diffusion Model, Shengjun Zhang, Jinzhao Li,
Xin Fei, Hao Liu, Yueqi Duan
- 63 Multi-subject Open-set Personalization in Video Generation,
☆ Tsai-Shien Chen, Aliaksandr Siarohin, Willi Menapace, Yuwei Fang,
Kwot Sin Lee, Ivan Skorokhodov, Kfir Aberman, Jun-Yan Zhu,
Ming-Hsuan Yang, Sergey Tulyakov
- 64 Generative Gaussian Splatting for Unbounded 3D City
Generation, Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, Ziwei Liu
- 65 GEN3C: 3D-Informed World-Consistent Video Generation with
* Precise Camera Control, Xuanchi Ren, Tianchang Shen, Jiahui
Huang, Huan Ling, Yifan Lu, Merlin Nimier-David, Thomas Müller,
Alexander Keller, Sanja Fidler, Jun Gao
- 66 Taming Video Diffusion Prior with Scene-Grounding Guidance for
* 3D Gaussian Splatting from Sparse Inputs, Yingji Zhong,
Zhihao Li, Dave Zhenyu Chen, Lanqing Hong, Dan Xu
- 67 DynamicScaler: Seamless and Scalable Video Generation for
Panoramic Scenes, Jinxiu Liu, Shaoheng Lin, Yinxiao Li,
Ming-Hsuan Yang
- 68 LIM: Large Interpolator Model for Dynamic Reconstruction,
Remy Sabathier, Niloy J. Mitra, David Novotny
- 69 EgoLM: Multi-Modal Language Model of Egocentric
○ Motions, Fangzhou Hong, Vladimir Guzov, Hyo Jin Kim, Yuting Ye,
Richard Newcombe, Ziwei Liu, Lingni Ma
- 70 MoSca: Dynamic Gaussian Fusion from Casual Videos via 4D
* Motion Scaffolds, Jiahui Lei, Yijia Weng, Adam W. Harley,
Leonidas Guibas, Kostas Daniilidis
- 71 PhysGen3D: Crafting a Miniature Interactive World from a Single
Image, Boyuan Chen, Hanxiao Jiang, Shaowei Liu, Saurabh Gupta,
Yunzhu Li, Hao Zhao, Shenlong Wang
- 72 Link to the Past: Temporal Propagation for Fast 3D Human
Reconstruction from Monocular Video, Matthew Marchellus,
Nadhira Noor, In Kyu Park
- 73 The Language of Motion: Unifying Verbal and Non-verbal
Language of 3D Human Motion, Changan Chen, Juze Zhang,
Shrinidhi K. Lakshmikanth, Yusu Fang, Ruizhi Shao,
Gordon Wetzstein, Li Fei-Fei, Ehsan Adeli
- 74 Towards Explainable and Unprecedented Accuracy in Matching
* Challenging Finger Crease Patterns, Zhenyu Zhou,
Chengdong Dong, Ajay Kumar
- 75 One-Step Event-Driven High-Speed Autofocus, Yuhan Bao,
* Shaohua Gao, Wenying Li, Kaiwei Wang
- 76 Self-Supervised Learning for Color Spike Camera
Reconstruction, Yanchen Dong, Ruiqin Xiong, Xiaopeng Fan,
Zhaofei Yu, Yonghong Tian, Tiejun Huang
- 77 PS-EIP: Robust Photometric Stereo Based on Event Interval Profile,
Kazuma Kitazawa, Takahito Aoto, Satoshi Ikehata, Tsuyoshi Takatani
- 78 Efficient Depth Estimation for Unstable Stereo Camera Systems
on AR Glasses, Yongfan Liu, Hyoukjun Kwon
- 79 Scalable Autoregressive Monocular Depth Estimation,
Jinhong Wang, Jian Liu, Dongqi Tang, Weiqiang Wang,
Wentong Li, Danny Chen, Jintai Chen, Jian Wu
- 80 Multi-view Reconstruction via SfM-guided Monocular Depth
○ Estimation, Haoyu Guo, He Zhu, Sida Peng, Haotong Lin,
Yunzhi Yan, Tao Xie, Wenguan Wang, Xiaowei Zhou, Hujun Bao
- 81 FoundationStereo: Zero-Shot Stereo Matching, Bowen Wen,
Matthew Trepte, Joseph Aribido, Jan Kautz, Orazio Gallo,
Stan Birchfield
- 82 MonSter: Marry Monodepth to Stereo Unleashes Power,
* Junda Cheng, Longliang Liu, Gangwei Xu, Xianqi Wang,
Zhaoxing Zhang, Yong Deng, Jinliang Zang, Yurui Chen,
Zhipeng Cai, Xin Yang
- 83 Dual Exposure Stereo for Extended Dynamic Range 3D
Imaging, Juhyung Choi, Jinnyeong Kim, Seokjun Choi, Jinwoo Lee,
Samuel Brucker, Mario Bijelic, Felix Heide, Seung-Hwan Baek
- 84 Adapting Dense Matching for Homography Estimation with Grid-
based Acceleration, Kaining Zhang, Yuxin Deng, Jiayi Ma, Paolo Favaro
- 85 ProtoDepth: Unsupervised Continual Depth Completion with
☆ Prototypes, Patrick Rim, Hyounseob Park, S. Gangopadhyay,
Ziyao Zeng, Younjoon Chung, Alex Wong
- 86 VGGT: Visual Geometry Grounded Transformer, Jianyuan Wang,
☆ Minghao Chen, Nikita Karaev, Andrea Vedaldi,
Christian Rupprecht, David Novotny
- 87 DiffusionSfM: Predicting Structure and Motion via Ray Origin and
Endpoint Diffusion, Qitao Zhao, Amy Lin, Jeff Tan, Jason Y. Zhang,
Deva Ramanan, Shubham Tulsiani
- 88 Descriptor-In-Pixel : Point-Feature Tracking For Pixel Processor
☆ Arrays, Laurie Bose, Jianing Chen, Piotr Dudek
- 89 ScaleLSD: Scalable Deep Line Segment Detection
Streamlined, Zeran Ke, Bin Tan, Xianwei Zheng, Yujun Shen,
Tianfu Wu, Nan Xue
- 90 MEGA: Masked Generative Autoencoder for Human Mesh
○ Recovery, Guénolé Fiche, Simon Leglaive,
Xavier Alameda-Pineda, Francesc Moreno-Noguer
- 91 Reconstructing Humans with a Biomechanically Accurate Skeleton,
☆ Yan Xia, Xiaowei Zhou, Etienne Vouga, Qixing Huang,
○ Georgios Pavlakos
- 92 EDM: Equirectangular Projection-Oriented Dense Kernelized
Feature Matching, Dongki Jung, Jaehoon Choi, Yonghan Lee,
Somi Jeong, Taejae Lee, Dinesh Manocha, Suyong Yeon
- 93 Feat2GS: Probing Visual Foundation Models with Gaussian Splatting,

- Yue Chen, Xingyu Chen, Anpei Chen, Gerard Pons-Moll, Yuliang Xiu
 94 FG²: Fine-Grained Cross-View Localization by Fine-Grained Feature Matching, Zimin Xia, Alexandre Alahi
- 95 Stop Walking in Circles! Bailing Out Early in Projected Gradient Descent, Philip Doldo, Derek Everett, Amol Khanna, Andre T Nguyen, Edward Raff
- 96 Self-Supervised Spatial Correspondence Across Modalities, Ayush Shrivastava, Andrew Owens
- 97 RDD: Robust Feature Detector and Descriptor using Deformable Transformer, Gonglin Chen, Tianwen Fu, Haiwei Chen, Wenbin Teng, Hanyuan Xiao, Yajie Zhao
- 98 Dense-SfM: Structure from Motion with Dense Consistent Matching, JongMin Lee, Sungjoo Yoo
- 99 HeatFormer: A Neural Optimizer for Multiview Human Mesh Recovery, Yuto Matsubara, Ko Nishino
- 100 DualPM: Dual Posed-Canonical Point Maps for 3D Shape and Pose Reconstruction, Ben Kaye, Tomas Jakab, Shangzhe Wu, Christian Rupprecht, Andrea Vedaldi
- 101 iG-6DoF: Model-free 6DoF Pose Estimation for Unseen Object via Iterative 3D Gaussian Splatting, Tuo Cao, Fei Luo, Jiongming Qin, Yu Jiang, Yusen Wang, Chunxia Xiao
- 102 RefPose: Leveraging Reference Geometric Correspondences for Accurate 6D Pose Estimation of Unseen Objects, Jaeguk Kim, Jaewoo Park, Keuntek Lee, Nam Ik Cho
- 103 One2Any: One-Reference 6D Pose Estimation for Any Object, Mengya Liu, Siyuan Li, Ajad Chhatkuli, Prune Truong, Luc Van Gool, Federico Tombari
- 104 Common3D: Self-Supervised Learning of 3D Morphable Models for Common Objects in Neural Feature Space, Leonhard Sommer, Olaf Dünkel, Christian Theobalt, Adam Kortylewski
- 105 ESCAPE: Equivariant Shape Completion via Anchor Point Encoding, Burak Bekci, Nassir Navab, Federico Tombari, Mahdi Saleh
- 106 Open-World Amodal Appearance Completion, Jiayang Ao, Yanbei Jiang, Qihong Ke, Krista A. Ehinger
- 107 Exploring Historical Information for RGBE Visual Tracking with Mamba, Chuanyu Sun, Jiqing Zhang, Yang Wang, Huilin Ge, Qianchen Xia, Baocai Yin, Xin Yang
- 108 EBS-EKF: Accurate and High Frequency Event-based Star Tracking, Albert W. Reed, Connor Hashemi, Dennis Melamed, Nitesh Menon, Keigo Hirakawa, Scott McCloskey
- 109 MonoDGP: Monocular 3D Object Detection with Decoupled-Query and Geometry-Error Priors, Fanqi Pu, Yifan Wang, Jiru Deng, Wenming Yang
- 110 MonoPlace3D: Learning 3D-Aware Object Placement for 3D Monocular Detection, Rishabh Parihar, Srinjay Sarkar, Sarthak Vora, Jogendra Nath Kundu, R. Venkatesh Babu
- 111 4Deform: Neural Surface Deformation for Robust Shape Interpolation, Lu Sang, Zehranaz Canfes, Dongliang Cao, Riccardo Marin, Florian Bernard, Daniel Cremers
- 112 Toward Robust Neural Reconstruction from Sparse Point Sets, Amine Ousafi, Shubhendu Jena, Eric Marchand, Adnane Boukhayma
- 113 MoGe: Unlocking Accurate Monocular Geometry Estimation for Open-Domain Images with Optimal Training Supervision, Ruicheng Wang, Sicheng Xu, Cassie Dai, Jianfeng Xiang, Yu Deng, Xin Tong, Jiaolong Yang
- 114 ArcPro: Architectural Programs for Structured 3D Abstraction of Sparse Points, Qirui Huang, Runze Zhang, Kangjun Liu, Minglun Gong, Hao Zhang, Hui Huang
- 115 ColabSfM: Collaborative Structure-from-Motion by Point Cloud Registration, Johan Edstedt, André Mateus, Alberto Jaenal
- 116 MoST: Efficient Monarch Sparse Tuning for 3D Representation Learning, Xu Han, Yuan Tang, Jinfeng Xu, Xianzhi Li
- 117 Flash3D: Super-scaling Point Transformers through Joint Hardware-Geometry Locality, Liyan Chen, Gregory P. Meyer, Zaiwei Zhang, Eric M. Wolff, Paul Vernaza
- 118 PointLoRA: Low-Rank Adaptation with Token Selection for Point Cloud Learning, Song Wang, Xiaolu Liu, Lingdong Kong, Jianyun Xu, Chunyong Hu, Gongfan Fang, Wentong Li, Jianke Zhu, Xinchao Wang
- 119 MICAS: Multi-grained In-Context Adaptive Sampling for 3D Point Cloud Processing, Feifei Shao, Ping Liu, Zhao Wang, Yawei Luo, Hongwei Wang, Jun Xiao
- 120 LidarGait++: Learning Local Features and Size Awareness from LiDAR Point Clouds for 3D Gait Recognition, Chuanfu Shen, Rui Wang, Lixin Duan, Shiqi Yu
- 121 CMMLoc: Advancing Text-to-PointCloud Localization with Cauchy-Mixture-Model Based Framework, Yanlong Xu, Haoxuan Qu, Jun Liu, Wenxiao Zhang, Xun Yang
- 122 HOTFormerLoc: Hierarchical Octree Transformer for Versatile Lidar Place Recognition Across Ground and Aerial Views, Ethan Griffiths, Maryam Haghighat, Simon Denman, Clinton Fookes, Milad Ramezani
- 123 ForestLPR: LiDAR Place Recognition in Forests Attentioning Multiple BEV Density Images, Yanqing Shen, Turcan Tuna, Marco Hutter, Cesar Cadena, Nanning Zheng
- 124 PSA-SSL: Pose and Size-aware Self-Supervised Learning on LiDAR Point Clouds, Barza Nisar, Steven L. Waslander
- 125 LightLoc: Learning Outdoor LiDAR Localization at Light Speed, Wen Li, Chen Liu, Shangshu Yu, Dunqiang Liu, Yin Zhou, Siqi Shen, Chenglu Wen, Cheng Wang
- 126 No Thing, Nothing: Highlighting Safety-Critical Classes for Robust LiDAR Semantic Segmentation in Adverse Weather, Junsung Park, Hwiyeong Lee, Inha Kang, Hyunjung Shim
- 127 RC-AutoCalib: An End-to-End Radar-Camera Automatic Calibration Network, Van-Tin Luu, Yon-Lin Cai, Vu-Hoang Tran, Wei-Chen Chiu, Yi-Ting Chen, Ching-Chun Huang
- 128 Pseudo Visible Feature Fine-Grained Fusion for Thermal Object Detection, Ting Li, Mao Ye, Tianwen Wu, Nianxin Li, Shuaifeng Li, Song Tang, Luping Ji
- 129 Resilient Sensor Fusion Under Adverse Sensor Failures via Multi-Modal Expert Fusion, Konyul Park, Yecheol Kim, Daehun Kim, Jun Won Choi
- 130 Similarity-Guided Layer-Adaptive Vision Transformer for UAV Tracking, Chaocan Xue, Bineng Zhong, Qihua Liang, Yaozong Zheng, Ning Li, Yuanliang Xue, Shuxiang Song
- 131 MAGic-SLAM: Multi-Agent Gaussian Globally Consistent SLAM, Vladimir Yugay, Theo Gevers, Martin R. Oswald
- 132 SDGOCC: Semantic and Depth-Guided Bird's-Eye View Transformation for 3D Multimodal Occupancy Prediction, ZaiPeng Duan, ChenXu Dang, Xuzhong Hu, Pei An, Junfeng Ding, Jie Zhan, YunBiao Xu, Jie Ma
- 133 VoxelSplat: Dynamic Gaussian Splatting as an Effective Loss for Occupancy and Flow Prediction, Ziyue Zhu, Shenlong Wang, Jin Xie, Jiang-jiang Liu, Jingdong Wang, Jian Yang
- 134 GaussianWorld: Gaussian World Model for Streaming 3D Occupancy Prediction, Sicheng Zuo, Wenzhao Zheng, Yuanhui Huang, Jie Zhou, Jiwen Lu
- 135 DeSiRe-GS: 4D Street Gaussians for Static-Dynamic Decomposition and Surface Reconstruction for Urban Driving Scenes, Chensheng Peng, Chengwei Zhang, Yixiao Wang, Chenfeng Xu, Yichen Xie, Wenzhao Zheng, Kurt Keutzer, Masayoshi Tomizuka, Wei Zhan
- 136 JiSAM: Alleviate Labeling Burden and Corner Case Problems in Autonomous Driving via Minimal Real-World Data, Runjian Chen, Wenqi Shao, Bo Zhang, Shaoshuai Shi, Li Jiang, Ping Luo
- 137 Rethinking Lanes and Points in Complex Scenarios for Monocular 3D Lane Detection, Yifan Chang, Junjie Huang, Xiaofeng Wang, Yun Ye, Zhujiu Liang, Yi Shan, Dalong Du, Xingang Wang
- 138 SceneCrafter: Controllable Multi-View Driving Scene Editing, Zehao Zhu, Yuliang Zou, Chiyu Max Jiang, Bo Sun, Vincent Casser, Xiukun Huang, Jiahao Wang, Zhenpei Yang, Ruiqi Gao, Leonidas Guibas, Mingxing Tan, Dragomir Anguelov
- 139 Driving by the Rules: A Benchmark for Integrating Traffic Sign Regulations into Vectorized HD Map, Xinyuan Chang, Maixuan Xue, Xinran Liu, Zheng Pan, Xing Wei
- 140 CoSDH: Communication-Efficient Collaborative Perception via Supply-Demand Awareness and Intermediate-Late Hybridization, Junhao Xu, Yanan Zhang, Zhi Cai, Di Huang

- 141 Generating Multimodal Driving Scenes via Next-Scene Prediction, *Yanhao Wu, Haoyang Zhang, Tianwei Lin, Lichao Huang, Shujie Luo, Rui Wu, Congpei Qiu, Wei Ke, Tong Zhang*
- 142 Bridging Past and Future: End-to-End Autonomous Driving with Historical Prediction and Planning, *Bozhou Zhang, Nan Song, Xin Jin, Li Zhang*
- 143 MMTL-UniAD: A Unified Framework for Multimodal and Multi-Task Learning in Assistive Driving Perception, *Wenzhuo Liu, Wenshuo Wang, Yicheng Qiao, Qiannan Guo, Jiayin Zhu, Pengfei Li, Zilong Chen, Huiming Yang, Zhiwei Li, Lening Wang, Tiao Tan, Huaping Liu*
- 144 CityWalker: Learning Embodied Urban Navigation from Web-Scale Videos, *Xinhao Liu, Jintong Li, Yicheng Jiang, Niranjan Sujay, Zhicheng Yang, Juexiao Zhang, John Abanes, Jing Zhang, Chen Feng*
- 145 Evaluating Vision-Language Models as Evaluators in Path Planning, *Mohamed Aghzal, Xiang Yue, Erion Plaku, Ziyu Yao*
- 146 Scene Map-based Prompt Tuning for Navigation Instruction Generation, *Sheng Fan, Rui Liu, Wenguan Wang, Yi Yang*
- 147 Graph Neural Network Combining Event Stream and Periodic Aggregation for Low-Latency Event-based Vision, *Manon Dampfthoffer, Thomas Mesquida, Damien Joubert, Thomas Dalgaty, Pascal Vivet, Christoph Posch*
- 148 Code-as-Monitor: Constraint-aware Visual Programming for Reactive and Proactive Robotic Failure Detection, *Enshen Zhou, Qi Su, Cheng Chi, Zhizheng Zhang, Zhongyuan Wang, Tiejun Huang, Lu Sheng, He Wang*
- 149 RoboPEPP: Vision-Based Robot Pose and Joint Angle Estimation through Embedding Predictive Pre-Training, *Raktim Gautam Goswami, Prashanth Krishnamurthy, Yann LeCun, Farshad Khorrami*
- 150 PhysVLM: Enabling Visual Language Models to Understand Robotic Physical Reachability, *Weijie Zhou, Manli Tao, Chaoyang Zhao, Haiyun Guo, Honghui Dong, Ming Tang, Jinqiao Wang*
- 151 GarmentPile: Point-Level Visual Affordance Guided Retrieval and Adaptation for Cluttered Garments Manipulation, *Ruihai Wu, Ziyu Zhu, Yuran Wang, Yue Chen, Jiarui Wang, Hao Dong*
- 152 Tra-MoE: Learning Trajectory Prediction Model from Multiple Domains for Adaptive Policy Conditioning, *Jiange Yang, Haoyi Zhu, Yating Wang, Gangshan Wu, Tong He, Limin Wang*
- 153 AffordDP: Generalizable Diffusion Policy with Transferable Affordance, *Shijie Wu, Yihang Zhu, Yunao Huang, Kaizhen Zhu, Jiayuan Gu, Jingyi Yu, Ye Shi, Jingya Wang*
- 154 Phoenix: A Motion-based Self-Reflection Framework for Fine-grained Robotic Action Correction, *Wenke Xia, Ruoxuan Feng, Dong Wang, Di Hu*
- 155 ManipTrans: Efficient Dexterous Bimanual Manipulation Transfer via Residual Learning, *Kailin Li, Puhao Li, Tengyu Liu, Yuyang Li, Siyuan Huang*
- 156 PartRM: Modeling Part-Level Dynamics with Large Cross-State Reconstruction Model, *Mingju Gao, Yike Pan, Huan-ang Gao, Zongzheng Zhang, Wenyi Li, Hao Dong, Hao Tang, Li Yi, Hao Zhao*
- 157 InteractAnything: Zero-shot Human Object Interaction Synthesis via LLM Feedback and Object Affordance Parsing, *Jinlu Zhang, Yixin Chen, Zan Wang, Jie Yang, Yizhou Wang, Siyuan Huang*
- 158 How Do I Do That? Synthesizing 3D Hand Motion and Contacts for Everyday Interactions, *Aditya Prakash, Benjamin Lundell, Dmitry Andreychuk, David Forsyth, Saurabh Gupta, Harpreet Sawhney*
- 159 TokenHSL: Unified Synthesis of Physical Human-Scene Interactions through Task Tokenization, *Liang Pan, Zeshi Yang, Zhiyang Dou, Wenjia Wang, Buzhen Huang, Bo Dai, Taku Komura, Jingbo Wang*
- 160 EasyHOI: Unleashing the Power of Large Models for Reconstructing Hand-Object Interactions in the Wild, *Yumeng Liu, Xiaoxiao Long, Zemin Yang, Yuan Liu, Marc Habermann, Christian Theobalt, Yuexin Ma, Wenping Wang*
- 161 Temporally Consistent Object-Centric Learning by Contrasting Slots, *Anna Manasyan, Maximilian Seitzer, Filip Radovic, Georg Martius, Andrii Zadaianchuk*
- 162 InterAct: Advancing Large-Scale Versatile 3D Human-Object Interaction Generation, *Sirui Xu, Dongting Li, Yucheng Zhang, Xiyan Xu, Qi Long, Ziyin Wang, Yunzhi Lu, Shuchang Dong, Hezi Jiang, Akshat Gupta, Yu-Xiong Wang, Liang-Yan Gui*
- 163 HOT3D: Hand and Object Tracking in 3D from Egocentric Multi-View Videos, *Prithviraj Banerjee, Sindi Shkodrani, Pierre Moulon, Shreyas Hampali, Shangchen Han, Fan Zhang, Linguang Zhang, Jade Fountain, Edward Miller, Selen Basol, Richard Newcombe, Robert Wang, Jakob Julian Engel, Tomas Hodan*
- 164 Estimating Body and Hand Motion in an Ego-sensed World, *Brent Yi, Vickie Ye, Maya Zheng, Yunqi Li, Lea Müller, Georgios Pavlakos, Yi Ma, Jitendra Malik, Angjoo Kanazawa*
- 165 UMotion: Uncertainty-driven Human Motion Estimation from Inertial and Ultra-wideband Units, *Huakun Liu, Hiroki Ota, Xin Wei, Yutaro Hirao, Monica Perusquia-Hernandez, Hideaki Uchiyama, Kiyoshi Kiyokawa*
- 166 REWIND: Real-Time Egocentric Whole-Body Motion Diffusion with Exemplar-Based Identity Conditioning, *Jihyun Lee, Weipeng Xu, Alexander Richard, Shih-En Wei, Shunsuke Saito, Shaojie Bai, Te-Li Wang, Minhyuk Sung, Tae-Kyun Kim, Jason Saragih*
- 167 LAL: Enhancing 3D Human Motion Prediction with Latency-aware Auxiliary Learning, *Xiaoning Sun, Dong Wei, Huaijiang Sun, Shengxiang Hu*
- 168 LATTE-MV: Learning to Anticipate Table Tennis Hits from Monocular Videos, *Daniel Etaat, Dvij Kalaria, Nima Rahmianian, S. Shankar Sastry*
- 169 Pose Priors from Language Models, *Sanjay Subramanian, Evonne Ng, Lea Müller, Dan Klein, Shiry Ginosar, Trevor Darrell*
- 170 HOIGPT: Learning Long-Sequence Hand-Object Interaction with Language Models, *Mingzhen Huang, Fu-Jen Chu, Bugra Tekin, Kevin J. Liang, Haoyu Ma, Weiyao Wang, Xingyu Chen, Pierre Gleize, Hongfei Xue, Siwei Lyu, Kris Kitani, Matt Feiszli, Hao Tang*
- 171 HSL-GPT: A General-Purpose Large Scene-Motion-Language Model for Human Scene Interaction, *Yuan Wang, Yali Li, Xiang Li, Shengjin Wang*
- 172 SALAD: Skeleton-aware Latent Diffusion for Text-driven Motion Generation and Editing, *Seokhyeon Hong, Chaelin Kim, Serin Yoon, Junghyun Nam, Sihun Cha, Junyong Noh*
- 173 TIMotion: Temporal and Interactive Framework for Efficient Human-Human Motion Generation, *Yabiao Wang, Shuo Wang, Jiangning Zhang, Ke Fan, Jiafu Wu, Zhucun Xue, Yong Liu*
- 174 HuMoCon: Concept Discovery for Human Motion Understanding, *Qihang Fang, Chengcheng Tang, Bugra Tekin, Shugao Ma, Yanchao Yang*
- 175 ConMo: Controllable Motion Disentanglement and Recomposition for Zero-Shot Motion Transfer, *Jiayi Gao, Zijin Yin, Changcheng Hua, Yuxin Peng, Kongming Liang, Zhanyu Ma, Jun Guo, Yang Liu*
- 176 DreamTrack: Dreaming the Future for Multimodal Visual Object Tracking, *Mingzhe Guo, Weiping Tan, Wenyu Ran, Liping Jing, Zhipeng Zhang*
- 177 Seurat: From Moving Points to Depth, *Seokju Cho, Jiahui Huang, Seungryong Kim, Joon-Young Lee*
- 178 CH3Depth: Efficient and Flexible Depth Foundation Model with Flow Matching, *Jiaqi Li, Yiran Wang, Jinghong Zheng, Junrui Zhang, Liao Shen, Tianqi Liu, Zhiguo Cao*
- 179 Video Depth without Video Models, *Bingxin Ke, Dominik Narnhofer, Shengyu Huang, Lei Ke, Torben Peters, Katerina Fragkiadaki, Anton Obukhov, Konrad Schindler*
- 180 BiM-VFI: Bidirectional Motion Field-Guided Frame Interpolation for Video with Non-uniform Motions, *Wonyong Seo, Jihyong Oh, Munchurl Kim*
- 181 Autoregressive Sequential Pretraining for Visual Tracking, *Shiyi Liang, Yifan Bai, Yihong Gong, Xing Wei*
- 182 IM-Zero: Instance-level Motion Controllable Video Generation in a Zero-shot Manner, *Yuyang Huang, Yabo Chen, Li Ding, Xiaopeng Zhang, Wenrui Dai, Junni Zou, Hongkai Xiong, Qi Tian*
- 183 Track4Gen: Teaching Video Diffusion Models to Track Points Improves Video Generation, *Hyeonho Jeong, Chun-Hao P. Huang, Jong Chul Ye, Niloy J. Mitra, Duygu Ceylan*
- 184 Consistent and Controllable Image Animation with Motion Diffusion Models, *Xin Ma, Yaohui Wang, Gengyun Jia, Xinyuan Chen, Tien-Tsin Wong, Yuan-Fang Li, Cunjian Chen*
- 185 MatAnyone: Stable Video Matting with Consistent Memory Propagation, *Peiqing Yang, Shangchen Zhou, Jixin Zhao,*

- Qingyi Tao, Chen Change Loy*
- 186 Unboxed: Geometrically and Temporally Consistent Video Outpainting, *Zhongrui Yu, Martina Megaro-Boldini, Robert W. Sumner, Abdelaziz Djelouah*
- 187 High Dynamic Range Video Compression: A Large-Scale Benchmark Dataset and A Learned Bit-depth Scalable Compression Algorithm, *Zhaoyi Tian, Feifeng Wang, Shiwei Wang, Zihao Zhou, Yao Zhu, Liquan Shen*
- 188 ECV: Exploiting Non-Local Correlations in Multiple Frames for Contextual Video Compression, *Wei Jiang, Junru Li, Kai Zhang, Li Zhang*
- 189 RivuletMLP: An MLP-based Architecture for Efficient Compressed Video Quality Enhancement, *Gang He, Weiran Wang, Guancheng Quan, Shihao Wang, Dajiang Zhou, Yunsong Li*
- 190 Timestep Embedding Tells: It's Time to Cache for Video Diffusion
* Model, *Feng Liu, Shiwei Zhang, Xiaofeng Wang, Yujie Wei, Haonan Qiu, Yuzhong Zhao, Yingya Zhang, Qixiang Ye, Fang Wan*
- 191 AR-Diffusion: Asynchronous Video Generation with Auto-Regressive Diffusion, *Mingzhen Sun, Weining Wang, Gen Li, Jiawei Liu, Jiahui Sun, Wanquan Feng, Shanshan Lao, Siyu Zhou, Qian He, Jing Liu*
- 192 Taming Teacher Forcing for Masked Autoregressive Video
☆ Generation, *Deyu Zhou, Quan Sun, Yuang Peng, Kun Yan, Runpei Dong, Duomin Wang, Zheng Ge, Nan Duan, Xiangyu Zhang*
- 193 Self-supervised ControlNet with Spatio-Temporal Mamba for Real-world Video Super-resolution, *Shijun Shi, Jing Xu, Lijing Lu, Zhihang Li, Kai Hu*
- 194 Face Forgery Video Detection via Temporal Forgery Cue Unraveling, *Zonghui Guo, Yingjie Liu, Jie Zhang, Haiyong Zheng, Shiguang Shan*
- 195 SVFR: A Unified Framework for Generalized Video Face Restoration, *Zhiyao Wang, Xu Chen, Chengming Xu, Junwei Zhu, Xiaobin Hu, Jiangning Zhang, Chengjie Wang, Yuqi Liu, Yiyi Zhou, Rongrong Ji*
- 196 RSAR: Restricted State Angle Resolver and Rotated SAR Benchmark, *Xin Zhang, Xue Yang, Yuxuan Li, Jian Yang, Ming-Ming Cheng, Xiang Li*
- 197 RobSense: A Robust Multi-modal Foundation Model for Remote Sensing with Static, Temporal, and Incomplete Data Adaptability, *Minh Kha Do, Kang Han, Phu Lai, Khoa T. Phan, Wei Xiang*
- 198 A Selective Re-learning Mechanism for Hyperspectral Fusion Imaging, *Yuanye Liu, Jinyang Liu, Renwei Dian, Shutao Li*
- 199 A General Adaptive Dual-level Weighting Mechanism for Remote Sensing Pansharpening, *Jie Huang, Haorui Chen, Jiaxuan Ren, Siran Peng, Liangjian Deng*
- 200 Task-driven Image Fusion with Learnable Fusion Loss, *Haowen Bai, Jingshe Zhang, Zixiang Zhao, Yichen Wu, Lilun Deng, Yukun Cui, Tao Feng, Shuang Xu*
- 201 Channel Consistency Prior and Self-Reconstruction Strategy
☆ Based Unsupervised Image Deraining, *Guanglu Dong, Tianheng Zheng, Yuanzhouhan Cao, Linbo Qing, Chao Ren*
- 202 OSMamba: Omnidirectional Spectral Mamba with Dual-Domain Prior Generator for Exposure Correction, *Gehui Li, Bin Chen, Chen Zhao, Lei Zhang, Jian Zhang*
- 203 MaLR: A Locality- and Continuity-Preserving Mamba for Image Restoration, *Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, Xi Peng*
- 204 Zero-Shot Blind-spot Image Denoising via Implicit Neural Sampling, *Yuhui Quan, Tianxiang Zheng, Zhiyuan Ma, Hui Ji*
- 205 Adaptive Dropout: Unleashing Dropout across Layers for Generalizable Image Super-Resolution, *Hang Xu, Jie Huang, Wei Yu, Jiangtong Tan, Zhen Zou, Feng Zhao*
- 206 Vision-Language Gradient Descent-driven All-in-One Deep
☆ Unfolding Networks, *Haijin Zeng, Xiangming Wang, Yongyong Chen, Jingyong Su, Jie Liu*
- 207 DiffISR: A Diffusion Model with Gradient Guidance for Infrared Image Super-Resolution, *Xingyuan Li, Zirui Wang, Yang Zou, Zhixin Chen, Jun Ma, Zhiying Jiang, Long Ma, Jinyuan Liu*
- 208 Reversing Flow for Image Restoration, *Haina Qin, Wenyang Luo, Libin Wang, Dandan Zheng, Jingdong Chen, Ming Yang, Bing Li, Weiming Hu*
- 209 Navigating Image Restoration with VAR's Distribution Alignment Prior, *Siyang Wang, Naishan Zheng, Jie Huang, Feng Zhao*
- 210 Image Quality Assessment: From Human to Machine Preference,
* *Chunyi Li, Yuan Tian, Xiaoyue Ling, Zicheng Zhang, Haodong Duan, Haoning Wu, Ziheng Jia, Xiaohong Liu, Xiongkuo Min, Guo Lu, Weisi Lin, Guangtao Zhai*
- 211 DnLUT: Ultra-Efficient Color Image Denoising via Channel-Aware Lookup Tables, *Sidi Yang, Bin Xiao Huang, Yulun Zhang, Dahai Yu, Yujia Yang, Ngai Wong*
- 212 Detail-Preserving Latent Diffusion for Stable Shadow Removal, *Jiamin Xu, Yuxin Zheng, Zelong Li, Chi Wang, Renshu Gu, Weiwei Xu, Gang Xu*
- 213 Shadow Generation Using Diffusion Model with Geometry Prior, *Haonan Zhao, Qingyang Liu, Xinhao Tao, Li Niu, Guangtao Zhai*
- 214 TurboFill: Adapting Few-step Text-to-image Model for Fast Image Inpainting, *Liangbin Xie, Daniil Pakhomov, Zhonghao Wang, Zongze Wu, Ziyang Chen, Yuqian Zhou, Haitian Zheng, Zhifei Zhang, Zhe Lin, Jiantao Zhou, Chao Dong*
- 215 Linear Attention Modeling for Learned Image Compression, *Donghui Feng, Zhengxue Cheng, Shen Wang, Ronghua Wu, Hongwei Hu, Guo Lu, Li Song*
- 216 Multirate Neural Image Compression with Adaptive Lattice
* Vector Quantization, *Hao Xu, Xiaolin Wu, Xi Zhang*
- 217 Generative Image Layer Decomposition with Visual Effects,
☆ *Jinrui Yang, Qing Liu, Yijun Li, Soo Ye Kim, Daniil Pakhomov, Mengwei Ren, Jianming Zhang, Zhe Lin, Cihang Xie, Yuyin Zhou*
- 218 NitroFusion: High-Fidelity Single-Step Diffusion through Dynamic Adversarial Training, *Dar-Yen Chen, Hmishav Bandyopadhyay, Kai Zou, Yi-Zhe Song*
- 219 DiG: Scalable and Efficient Diffusion Models with Gated Linear
☆ Attention, *Lianghui Zhu, Zilong Huang, Bencheng Liao, Jun Hao Liew, Hanshu Yan, Jia Shi Feng, Xinggang Wang*
- 220 Reanimating Images using Neural Representations of Dynamic
○ Stimuli, *Jacob Yeung, Andrew F. Luo, Gabriel Sarch, Margaret M. Henderson, Deva Ramanan, Michael J. Tarr*
- 221 Early-Bird Diffusion: Investigating and Leveraging Timestep-Aware Early-Bird Tickets in Diffusion Models for Efficient Training, *Lexington Whalen, Zhenbang Du, Haoran You, Chaojian Li, Sixu Li, Yingyan Lin*
- 222 Latent Drifting in Diffusion Models for Counterfactual Medical
* Image Synthesis, *Yousef Yeganeh, Azade Farshad, Ioannis Charisiadis, Marta Hasny, Martin Hartenberger, Björn Ommer, Nassir Navab, Ehsan Adeli*
- 223 Style Quantization for Data-Efficient GAN Training, *Jian Wang, Xin Lan, Jizhe Zhou, Yuxin Tian, Jiancheng Lv*
- 224 Temporal Score Analysis for Understanding and Correcting Diffusion Artifacts, *Yu Cao, Zengqun Zhao, Ioannis Patras, Shaogang Gong*
- 225 Efficient Personalization of Quantized Diffusion Model without Backpropagation, *Hoigi Seo, Wongi Jeong, Kyungryeol Lee, Se Young Chun*
- 226 Layered Image Vectorization via Semantic Simplification, *Zhenyu Wang, Jianxi Huang, Zhida Sun, Yuanhao Gong, Daniel Cohen-Or, Min Lu*
- 227 JanusFlow: Harmonizing Autoregression and Rectified Flow for Unified Multimodal Understanding and Generation, *Yiyang Ma, Xingchao Liu, Xiaokang Chen, Wen Liu, Chengyue Wu, Zhiyu Wu, Zizheng Pan, Zhenda Xie, Haowei Zhang, Xingkai Yu, Liang Zhao, Yisong Wang, Jiaying Liu, Chong Ruan*
- 228 OpenHumanVid: A Large-Scale High-Quality Dataset for
* Enhancing Human-Centric Video Generation, *Hui Li, Mingwang Xu, Yun Zhan, Shan Mu, Jiaye Li, Kaihui Cheng, Yuxuan Chen, Tan Chen, Mao Ye, Jingdong Wang, Siyu Zhu*
- 229 DiTCtrl: Exploring Attention Control in Multi-Modal Diffusion Transformer for Tuning-Free Multi-Prompt Longer Video Generation, *Minghong Cai, Xiaodong Cun, Xiaoyu Li, Wenzhe Liu, Zhaoyang Zhang, Yong Zhang, Ying Shan, Xiangyu Yue*
- 230 MotiF: Making Text Count in Image Animation with Motion Focal
☆ Loss, *Shijie Wang, Samaneh Azadi, Rohit Girdhar, Saketh Rambhatla, Chen Sun, Xi Yin*

- 231 Visual Prompting for One-shot Controllable Video Editing without
☆ Inversion, *Zhengbo Zhang, Yuxi Zhou, Duo Peng, Joo-Hwee Lim, Zhigang Tu, De Wen Soh, Lin Geng Foo*
- 232 Tiled Diffusion, *Or Madar, Ohad Fried*
- 233 Evaluating Model Perception of Color Illusions in Photorealistic Scenes, *Lingjun Mao, Zineng Tang, Alane Suhr*
- 234 Charm: The Missing Piece in ViT Fine-Tuning for Image Aesthetic Assessment, *Fatemeh Behrad, Tinne Tuytelaars, Johan Wagemans*
- 235 Morpheus: Text-Driven 3D Gaussian Splat Shape and Color
☆ Stylization, *Jamie Wynn, Zawar Qureshi, Jakub Powierza, Jamie Watson, Mohamed Sayed*
- 236 Optical-Flow Guided Prompt Optimization for Coherent Video Generation, *Hyelin Nam, Jaemin Kim, Dohun Lee, Jong Chul Ye*
- 237 OmniStyle: Filtering High Quality Style Transfer Data at Scale, *Ye Wang, Ruiqi Liu, Jiang Lin, Fei Liu, Zili Yi, Yilin Wang, Rui Ma*
- 238 Pathways on the Image Manifold: Image Editing via Video Generation, *Noam Rotstein, Gal Yona, Daniel Silver, Roy Velich, David Bensaid, Ron Kimmel*
- 239 PhyS-EdiT: Physics-aware Semantic Image Editing with Text Description, *Ziqi Cai, Shuchen Weng, Yifei Xia, Boxin Shi*
- 240 Stable Flow: Vital Layers for Training-Free Image Editing, *Omri Avrahami, Or Patashnik, Ohad Fried, Egor Nemchinov, Kfir Aberman, Dani Lischinski, Daniel Cohen-Or*
- 241 Improving Editability in Image Generation with Layer-wise Memory, *Daneul Kim, Jaeh Lee, Jaesik Park*
- 242 EditAR: Unified Conditional Generation with Autoregressive Models, *Jiteng Mu, Nuno Vasconcelos, Xiaolong Wang*
- 243 Zero-Shot Styled Text Image Generation, but Make It
☆ Autoregressive, *Vittorio Pippi, Fabio Quattrini, Silvia Cascianelli, Alessio Tonioni, Rita Cucchiara*
- 244 Generative Photography: Scene-Consistent Camera Control for
* Realistic Text-to-Image Synthesis, *Yu Yuan, Xijun Wang, Yichen Sheng, Prateek Chennuri, Xingguang Zhang, Stanley Chan*
- 245 Generative Photomontage, *Sean J. Liu, Nupur Kumari, Ariel Shamir, Jun-Yan Zhu*
- 246 Multi-party Collaborative Attention Control for Image Customization,
☆ *Han Yang, Chuanguang Yang, Qiuli Wang, Zhulin An, Weilun Feng, Libo Huang, Yongjun Xu*
- 247 ART: Anonymous Region Transformer for Variable Multi-Layer
☆ Transparent Image Generation, *Yifan Pu, Yiming Zhao, Zhicong Tang, Ruihong Yin, Haoxing Ye, Yuhui Yuan, Dong Chen, Jianmin Bao, Sirui Zhang, Yanbin Wang, Lin Liang, Lijuan Wang, Ji Li, Xiu Li, Zhouhui Lian, Gao Huang, Baining Guo*
- 248 UNIC-Adapter: Unified Image-instruction Adapter with Multi-modal Transformer for Image Generation, *Lunhao Duan, Shanshan Zhao, Wenjun Yan, Yinglun Li, Qing-Guo Chen, Zhao Xu, Weihua Luo, Kaifu Zhang, Mingming Gong, Gui-Song Xia*
- 249 MMAR: Towards Lossless Multi-Modal Auto-Regressive Probabilistic Modeling, *Jian Yang, Dacheng Yin, Yizhou Zhou, Fengyun Rao, Wei Zhai, Yang Cao, Zheng-Jun Zha*
- 250 Large-Scale Text-to-Image Model with Inpainting is a Zero-Shot Subject-Driven Image Generator, *Chaehun Shin, Jooyoung Choi, Heeseung Kim, Sungroh Yoon*
- 251 SnapGen: Taming High-Resolution Text-to-Image Models for
* Mobile Devices with Efficient Architectures and Training, *Jierun Chen, Dongting Hu, Xijie Huang, Huseyin Coskun, Arpit Sahni, Aarush Gupta, Anujraaj Goyal, Dishani Lahiri, Rajesh Singh, Yerlan Idelbayev, Junli Cao, Yanyu Li, Kwang-Ting Cheng, S.-H. Gary Chan, Mingming Gong, Sergey Tulyakov, Anil Kag, Yanwu Xu, Jian Ren*
- 252 VideoDPO: Omni-Preference Alignment for Video Diffusion Generation, *Runtao Liu, Haoyu Wu, Ziqiang Zheng, Chen Wei, Yingqing He, Renjie Pi, Qifeng Chen*
- 253 Personalized Preference Fine-tuning of Diffusion Models, *Meihua Dang, Anikait Singh, Linqi Zhou, Stefano Ermon, Jiaming Song*
- 254 Text Embedding is Not All You Need: Attention Control for Text-to-Image Semantic Alignment with Text Self-Attention Maps, *Jeeyung Kim, Erfan Esmaeili, Qiang Qiu*
- 255 VerbDiff: Text-Only Diffusion Models with Enhanced Interaction Awareness, *Seungju Cha, Kwanyoung Lee, Ye-Chan Kim, Hyunwoo Oh, Dong-Jin Kim*
- 256 Learning Visual Generative Priors without Text, *Shuailei Ma, Kecheng Zheng, Ying Wei, Wei Wu, Fan Lu, Yifei Zhang, Chen-Wei Xie, Biao Gong, Jiapeng Zhu, Yujun Shen*
- 257 Towards Understanding and Quantifying Uncertainty for Text-to-Image Generation, *Gianni Franchi, Nacim Belkhir, Dat Nguyen Trong, Guoxuan Xia, Andrea Pilzer*
- 258 CoMM: A Coherent Interleaved Image-Text Dataset for Multimodal
* Understanding and Generation, *Wei Chen, Lin Li, Yongqi Yang, Bin Wen, Fan Yang, Tingting Gao, Yu Wu, Long Chen*
- 259 PosterMaker: Towards High-Quality Product Poster Generation with Accurate Text Rendering, *Yifan Gao, Zihang Lin, Chuanbin Liu, Min Zhou, Tiezheng Ge, Bo Zheng, Hongtao Xie*
- 260 Multitwine: Multi-Object Compositing with Text and Layout
* Control, *Gemma Canet Tarrés, Zhe Lin, Zhifei Zhang, He Zhang, Andrew Gilbert, John Collomosse, Soo Ye Kim*
- 261 MARVEL-40M+: Multi-Level Visual Elaboration for High-Fidelity Text-to-3D Content Creation, *Sankalp Sinha, Mohammad Sadil Khan, Muhammad Usama, Shino Sam, Didier Stricker, Sk Aziz Ali, Muhammad Zeshan Afzal*
- 262 PosterO: Structuring Layout Trees to Enable Language Models in Generalized Content-Aware Layout Generation, *HsiaoYuan Hsu, Yuxin Peng*
- 263 From Elements to Design: A Layered Approach for Automatic Graphic Design Composition, *Jiawei Lin, Shizhao Sun, Danqing Huang, Ting Liu, Ji Li, Jiang Bian*
- 264 Apparel: A Multimodal Foundation Model for Digital Garments,
* *Kiyohiro Nakayama, Jan Ackermann, Timur Levent Kesdogan, Yang Zheng, Maria Korosteleva, Olga Sorkine-Hornung, Leonidas J. Guibas, Guandao Yang, Gordon Wetzstein*
- 265 ChatHuman: Chatting about 3D Humans with Tools, *Jing Lin, Yao Feng, Weiyang Liu, Michael J. Black*
- 266 Interpretable Generative Models through Post-hoc Concept Bottlenecks, *Akshay Kulkarni, Ge Yan, Chung-En Sun, Tuomas Oikarinen, Tsui-Wei Weng*
- 267 Concept Replacer: Replacing Sensitive Concepts in Diffusion Models via Precision Localization, *Lingyun Zhang, Yu Xie, Yanwei Fu, Ping Chen*
- 268 Enhancing Privacy-Utility Trade-offs to Mitigate Memorization in Diffusion Models, *Chen Chen, Daochang Liu, Mubarak Shah, Chang Xu*
- 269 Dissecting and Mitigating Diffusion Bias via Mechanistic Interpretability, *Yingdong Shi, Changming Li, Yifan Wang, Yongxiang Zhao, Anqi Pang, Sibe Yang, Jingyi Yu, Kan Ren*
- 270 Silent Branding Attack: Trigger-free Data Poisoning Attack on Text-to-Image Diffusion Models, *Sangwon Jang, June Suk Choi, Jaehyeong Jo, Kimin Lee, Sung Ju Hwang*
- 271 SleeperMark: Towards Robust Watermark against Fine-Tuning Text-to-image Diffusion Models, *Zilan Wang, Junfeng Guo, Jiacheng Zhu, Yiming Li, Heng Huang, Muhao Chen, Zhengzhong Tu*
- 272 Watermarking One for All: A Robust Watermarking Scheme Against Partial Image Theft, *Gaozhi Liu, Silu Cao, Zhenxing Qian, Xinpeng Zhang, Sheng Li, Wanli Peng*
- 273 Enhancing Facial Privacy Protection via Weakening Diffusion Purification, *Ali Salar, Qing Liu, Yingli Tian, Guoying Zhao*
- 274 Community Forensics: Using Thousands of Generators to Train Fake Image Detectors, *Jeongsoo Park, Andrew Owens*
- 275 Beyond Generation: A Diffusion-based Low-level Feature Extractor for Detecting AI-generated Images, *Nan Zhong, Haoyu Chen, Yiran Xu, Zhenxing Qian, Xinpeng Zhang*
- 276 Unveiling Differences in Generative Models: A Scalable
* Differential Clustering Approach, *Jingwei Zhang, Mohammad Jalali, Cheuk Ting Li, Farzan Farnia*
- 277 The PanAf-FGBG Dataset: Understanding the Impact of
📄 Backgrounds in Wildlife Behaviour Recognition, *Otto Brookes, Maksim Kukushkin, Majid Mirmehdi, Colleen Stephens, Paula Dieguez, Thurston C. Hicks, Sorrel Jones, Kevin Lee, Maureen S. McCarthy, Amelia Meier, Emmanuelle Normand, Erin G. Wessling, Roman M. Wittig, Kevin Langergraber, Klaus Zuberbühler, Lukas Boesch, Thomas Schmid, Mimi Arandjelovic, Hjalmar Kühl, Tilo Burghardt*
- 278 MIRE: Matched Implicit Neural Representations, *Dhananjaya*

- Jayasundara, Heng Zhao, Demetrio Labate, Vishal M. Patel
 279 Learning to Normalize on the SPD Manifold under Bures-Wasserstein Geometry, Rui Wang, Shaocheng Jin, Ziheng Chen, Xiaoqing Luo, Xiao-Jun Wu
- 280 Benchmarking Object Detectors under Real-World Distribution Shifts in Satellite Imagery, Sara A. Al-Emadi, Yin Yang, Ferda Ofli
 ☆ Gazing Into Missteps: Leveraging Eye-Gaze for Unsupervised Mistake Detection in Egocentric Videos of Skilled Human Activities, Michele Mazzamuto, Antonino Furnari, Yoichi Sato, Giovanni Maria Farinella
- 282 Enhancing Dance-to-Music Generation via Negative Conditioning Latent Diffusion Model, Changchang Sun, Gaowen Liu, Charles Fleming, Yan Yan
- 283 SoundVista: Novel-View Ambient Sound Synthesis via Visual-Acoustic Binding, Mingfei Chen, Israel D. Gebru,
 ☆ Ishwarya Ananthabhotla, Christian Richardt, Dejan Markovic, Jake Sandakly, Steven Krenn, Todd Keebler, Eli Shlizerman, Alexander Richard
- 284 Object-aware Sound Source Localization via Audio-Visual Scene Understanding, Sung Jin Um, Dongjin Kim, Sangmin Lee, Jung Uk Kim
- 285 Revisiting Audio-Visual Segmentation with Vision-Centric Transformer, Shaofei Huang, Rui Ling, Tianrui Hui, Hongyu Li, Xu Zhou, Shifeng Zhang, Si Li, Richang Hong, Meng Wang
- 286 Towards Open-Vocabulary Audio-Visual Event Localization, Jinxing Zhou, Dan Guo, Ruohao Guo, Yuxin Mao, Jingjing Hu, Yiran Zhong, Xiaojun Chang, Meng Wang
- 287 Contextual AD Narration with Interleaved Multimodal Sequence, Hanlin Wang, Zhan Tong, Kecheng Zheng, Yujun Shen, Limin Wang
- 288 Towards Universal Soccer Video Understanding, Jiayuan Rao, Haoning Wu, Hao Jiang, Ya Zhang, Yanfeng Wang, Weidi Xie
- 289 Neuro-Symbolic Evaluation of Text-to-Video Models using Formal Verification, S P Sharan, Minkyu Choi, Sahil Shah, Harsh Goel, Mohammad Omama, Sandeep Chinchali
- 290 T2V-CompBench: A Comprehensive Benchmark for Compositional Text-to-video Generation, Kaiyue Sun, Kaiyi Huang, Xian Liu, Yue Wu, Zihan Xu, Zhenguo Li, Xihui Liu
- 291 Event-Equalized Dense Video Captioning, Kangyi Wu, Pengna Li, Jingwen Fu, Yizhe Li, Yang Wu, Yuhao Liu, Jinjun Wang, Sanping Zhou
- 292 Koala-36M: A Large-scale Video Dataset Improving Consistency between Fine-grained Conditions and Video Content, Qiheng Wang, Yukai Shi, Jiarong Ou, Rui Chen, Ke Lin, Jiahao Wang, Boyuan Jiang, Haotian Yang, Mingwu Zheng, Xin Tao, Fei Yang, Pengfei Wan, Di Zhang
- 293 SMILE: Infusing Spatial and Motion Semantics in Masked Video Learning, Fida Mohammad Thoker, Letian Jiang, Chen Zhao, Bernard Ghanem
- 294 MotionBench: Benchmarking and Improving Fine-grained Video Motion Understanding for Vision Language Models, Wenyi Hong, Yean Cheng, Zhuoyi Yang, Weihao Wang, Lefan Wang, Xiaotao Gu, Shiyu Huang, Yuxiao Dong, Jie Tang
- 295 VideoAutoArena: An Automated Arena for Evaluating Large Multimodal Models in Video Analysis through User Simulation, Ziyang Luo, Haoning Wu, Dongxu Li, Jing Ma, Mohan Kankanhalli, Junnan Li
- 296 MMVU: Measuring Expert-Level Multi-Discipline Video Understanding, Yilun Zhao, Haowei Zhang, Lujing Xie, Tongyan Hu, Guo Gan, Yitao Long, Zhiyuan Hu, Weiyan Chen, Chuhan Li, Zhijian Xu, Chengye Wang, Ziyao Shangguan, Zhenwen Liang, Yixin Liu, Chen Zhao, Arman Cohan
- 297 VideoComposition: Can MLLMs Analyze Compositions in Compiled Videos?, Yunlong Tang, Junjia Guo, Hang Hua, Susan Liang, Mingqian Feng, Xinyang Li, Rui Mao, Chao Huang, Jing Bi, Zeliang Zhang, Pooyan Fazli, Chenliang Xu
- 298 VITED: Video Temporal Evidence Distillation, Yujie Lu, Yale Song, William Wang, Lorenzo Torresani, Tushar Nagarajan
- 299 DynFocus: Dynamic Cooperative Network Empowers LLMs with Video Understanding, Yudong Han, Qingpei Guo, Liyuan Pan, Liu Liu, Yu Guan, Ming Yang
- 300 Enhancing Video-LLM Reasoning via Agent-of-Thoughts Distillation, Yudi Shi, Shangzhe Di, Qirui Chen, Weidi Xie
- 301 AdaCM^2: On Understanding Extremely Long-Term Video with Adaptive Cross-Modality Memory Reduction, Yuanbin Man, Ying Huang, Chengming Zhang, Bingzhe Li, Wei Niu, Miao Yin
- 302 Temporal Alignment-Free Video Matching for Few-shot Action Recognition, SuBeen Lee, WonJun Moon, Hyun Seok Seong, Jae-Pil Heo
- 303 HierarQ: Task-Aware Hierarchical Q-Former for Enhanced Video Understanding, Shehreen Azad, Vibhav Vineet, Yogesh Singh Rawat
- 304 Video Language Model Pretraining with Spatio-temporal Masking, Yue Wu, Zhaobo Qi, Junshu Sun, Yaowei Wang, Qingming Huang, Shuhui Wang
- 305 Hybrid-Level Instruction Injection for Video Token Compression in Multi-modal Large Language Models, Zhihang Liu, Chen-Wei Xie, Pandeng Li, Liming Zhao, Longxiang Tang, Yun Zheng, Chuanbin Liu, Hongtao Xie
- 306 Re-thinking Temporal Search for Long-Form Video Understanding, Jinhui Ye, Zihan Wang, Haosen Sun, Keshigeyan Chandrasegaran, Zane Durante, Cristobal Eyzaguirre, Yonatan Bisk, Juan Carlos Niebles, Ehsan Adeli, Li Fei-Fei, Jiajun Wu, Manling Li
- 307 LLaVA-ST: A Multimodal Large Language Model for Fine-Grained Spatial-Temporal Understanding, Hongyu Li, Jinyu Chen, Ziyu Wei, Shaofei Huang, Tianrui Hui, Jialin Gao, Xiaoming Wei, Si Liu
- 308 Scalable Video-to-Dataset Generation for Cross-Platform Mobile Agents, Yunseok Jang, Yeda Song, Sungryull Sohn, Lajanugen Logeswaran, Tiange Luo, Dong-Ki Kim, Kyunghoon Bae, Honglak Lee
- 309 PHGC: Procedural Heterogeneous Graph Completion for Natural Language Task Verification in Egocentric Videos, Xun Jiang, Zhiyi Huang, Xing Xu, Jingkuan Song, Fumin Shen, Heng Tao Shen
- 310 Rethinking Spiking Self-Attention Mechanism: Implementing α -XNOR Similarity Calculation in Spiking Transformers, Yichen Xiao, Shuai Wang, Dehao Zhang, Wenjie Wei, Yimeng Shan, Xiaoli Liu, Yulin Jiang, Malu Zhang
- 311 Motion-Grounded Video Reasoning: Understanding and Perceiving Motion at Pixel Level, Andong Deng, Tongjia Chen, Shoubin Yu, Taojiannan Yang, Lincoln Spencer, Yapeng Tian, Ajmal Saeed Mian, Mohit Bansal, Chen Chen
- 312 V^2Dial: Unification of Video and Visual Dialog via Multimodal Experts, Adnen Abdessaied, Anna Rohrbach, Marcus Rohrbach, Andreas Bulling
- 313 Towards Unbiased and Robust Spatio-Temporal Scene Graph Generation and Anticipation, Rohith Peddi, Saurabh Saurabh, Ayush Abhay Shrivastava, Parag Singla, Vibhav Gogate
- 314 GLUS: Global-Local Reasoning Unified into A Single Large Language Model for Video Segmentation, Lang Lin, Xueyang Yu, Ziqi Pang, Yu-Xiong Wang
- 315 LiVOS: Light Video Object Segmentation with Gated Linear Matching, Qin Liu, Jianfeng Wang, Zhengyuan Yang, Linjie Li, Kevin Lin, Marc Niethammer, Lijuan Wang
- 316 VERA: Explainable Video Anomaly Detection via Verbalized Learning of Vision-Language Models, Muchao Ye, Weiyang Liu, Pan He
- 317 Track Any Anomalous Object: A Granular Video Anomaly Detection Pipeline, Yuzhi Huang, Chenxin Li, Haitao Zhang, Zixu Lin, Yunlong Lin, Hengyu Liu, Wuyang Li, Xinyu Liu, Jiechao Gao, Yue Huang, Xinghao Ding, Yixuan Yuan
- 318 Context-Enhanced Memory-Refined Transformer for Online Action Detection, Zhanzhong Pang, Fadime Sener, Angela Yao
- 319 Bridge the Gap: From Weak to Full Supervision for Temporal Action Localization with PseudoFormer, Ziyi Liu, Yangcen Liu
- 320 Neuron: Learning Context-Aware Evolving Representations for Zero-Shot Skeleton Action Recognition, Yang Chen, Jingcai Guo, Song Guo, Dacheng Tao
- 321 MambaVLT: Time-Evolving Multimodal State Space Model for Vision-Language Tracking, Xinqi Liu, Li Zhou, Zikun Zhou, Jianqiu Chen, Zhenyu He
- 322 Lost in Translation, Found in Context: Sign Language Translation with Contextual Cues, Youngjoon Jang, Haran Raajesh, Liliane Momeni, Gül Varol, Andrew Zisserman

- 323 Theoretical Insights in Model Inversion Robustness and
 * Conditional Entropy Maximization for Collaborative Inference Systems, *Song Xia, Yi Yu, Wenhan Yang, Meiwen Ding, Zhuo Chen, Ling-Yu Duan, Alex C. Kot, Xudong Jiang*
- 324 Where the Devil Hides: Deepfake Detectors Can No Longer Be Trusted, *Shuaiwei Yuan, Junyu Dong, Yuezun Li*
- 325 FreqDebias: Towards Generalizable Deepfake Detection via Consistency-Driven Frequency Debiasing, *Hossein Kashiani, Niloufar Alipour Talemi, Fatemeh Afghah*
- 326 Omni-ID: Holistic Identity Representation Designed for Generative Tasks, *Guocheng Qian, Kuan-Chieh Wang, Or Patashnik, Negin Heravi, Daniil Ostashev, Sergey Tulyakov, Daniel Cohen-Or, Kfir Aberman*
- 327 VISTREAM: Improving Computation Efficiency of Visual Streaming Perception via Law-of-Charge-Conservation Inspired Spiking Neural Network, *Kang You, Ziling Wei, Jing Yan, Boning Zhang, Qinghai Guo, Yaoyu Zhang, Zhezhi He*
- 328 Temporal Separation with Entropy Regularization for Knowledge Distillation in Spiking Neural Networks, *Kairong Yu, Chengting Yu, Tianqing Zhang, Xiaochen Zhao, Shu Yang, Hongwei Wang, Qiang Zhang, Qi Xu*
- 329 Binarized Mamba-Transformer for Lightweight Quad Bayer
 ☆ HybridEVS Demosaicing, *Shiyang Zhou, Haijin Zeng, Yunfan Lu, Tong Shao, Ke Tang, Yongyong Chen, Jie Liu, Jingyong Su*
- 330 From Laboratory to Real World: A New Benchmark Towards Privacy-Preserved Visible-Infrared Person Re-Identification, *Yan Jiang, Hao Yu, Xu Cheng, Haoyu Chen, Zhaodong Sun, Guoying Zhao*
- 331 DefMamba: Deformable Visual State Space Model, *Leiyi Liu, Miao Zhang, Jihao Yin, Tingwei Liu, Wei Ji, Yongri Piao, Huchuan Lu*
- 332 ABBSP0: Adaptive Bounding Box Scaling and Symmetric
 ☆ Prior based Orientation Prediction for Detecting Aerial Image Objects, *Woojin Lee, Hyugjae Chang, Jaeho Moon, Jaehyup Lee, Munchul Kim*
- 333 Towards RAW Object Detection in Diverse Conditions, *Zhong-Yu Li, Xin Jin, Bo-Yuan Sun, Chun-Le Guo, Ming-Ming Cheng*
- 334 Closed-Loop Supervised Fine-Tuning of Tokenized Traffic
 ○ Models, *Zhejun Zhang, Peter Karkus, Maximilian Igl, Wenhao Ding, Yuxiao Chen, Boris Ivanovic, Marco Pavone*
- 335 Advancing Manga Analysis: Comprehensive Segmentation Annotations for the Manga109 Dataset, *Minshan Xie, Jian Lin, Hanyuan Liu, Chengze Li, Tien-Tsin Wong*
- 336 Sketchy Bounding-box Supervision for 3D Instance Segmentation, *Qian Deng, Le Hui, Jin Xie, Jian Yang*
- 337 Relation3D : Enhancing Relation Modeling for Point Cloud Instance Segmentation, *Jiahao Lu, Jiacheng Deng*
- 338 FSHNet: Fully Sparse Hybrid Network for 3D Object Detection, *Shuai Liu, Mingyue Cui, Boyang Li, Quanmin Liang, Tinghe Hong, Kai Huang, Yunxiao Shan, Kai Huang*
- 339 3D-AVS: LiDAR-based 3D Auto-Vocabulary Segmentation, *Weijie Wei, Osman Ülger, Fatemeh Karimi Nejadasi, Theo Gevers, Martin R. Oswald*
- 340 NTClick: Achieving Precise Interactive Segmentation With Noise-tolerant Clicks, *Chenyi Zhang, Ting Liu, Xiaochao Qu, Luoqi Liu, Yao Zhao, Yunchao Wei*
- 341 HyperSeg: Hybrid Segmentation Assistant with Fine-grained Visual Perceiver, *Cong Wei, Yujie Zhong, Haoxian Tan, Yong Liu, Jie Hu, Dengjie Li, Zheng Zhao, Yujiu Yang*
- 342 CholecTrack20: A Multi-Perspective Tracking Dataset for Surgical Tools, *Chinedu Innocent Nwoye, Kareem Elgohary, Anvita Srinivas, Fauzan Zaid, Joël L. Lavanchy, Nicolas Padoy*
- 343 Hybrid Reciprocal Transformer with Triplet Feature Alignment for Scene Graph Generation, *Jiawei Fu, Tiantian Zhang, Kai Chen, Qi Dou*
- 344 Textured Gaussians for Enhanced 3D Scene Appearance Modeling, *Brian Chao, Hung-Yu Tseng, Lorenzo Porzi, Chen Gao, Tuotuo Li, Qinbo Li, Ayush Saraf, Jia-Bin Huang, Johannes Kopf, Gordon Wetzstein, Changil Kim*
- 345 Global-Local Tree Search in VLMs for 3D Indoor Scene Generation, *Wei Deng, Mengshi Qi, Huadong Ma*
- 346 CrossOver: 3D Scene Cross-Modal Alignment, *Sayan Deb Sarkar, Ondrej Miksik, Marc Pollefeys, Daniel Barath, Iro Armeni*
- 347 Video-3D LLM: Learning Position-Aware Video Representation
 ☆ for 3D Scene Understanding, *Duo Zheng, Shijia Huang, Liwei Wang*
- 348 BIP3D: Bridging 2D Images and 3D Perception for Embodied Intelligence, *Xuewu Lin, Tianwei Lin, Lichao Huang, Hongyu Xie, Zhizhong Su*
- 349 Ges3ViG : Incorporating Pointing Gestures into Language-Based 3D Visual Grounding for Embodied Reference Understanding, *Atharv Mahesh Mane, Dulanga Weerakoon, Vigneshwaran Subbaraju, Sougata Sen, Sanjay E. Sarma, Archan Misra*
- 350 ANNEXE: Unified Analyzing, Answering, and Pixel Grounding for Egocentric Interaction, *Yuejiao Su, Yi Wang, Qiongyang Hu, Chuang Yang, Lap-Pui Chau*
- 351 Optimus-2: Multimodal Minecraft Agent with Goal-Observation-Action Conditioned Policy, *Zaijing Li, Yuquan Xie, Rui Shao, Gongwei Chen, Dongmei Jiang, Liqiang Nie*
- 352 Critic-V: VLM Critics Help Catch VLM Errors in Multimodal Reasoning, *Di Zhang, Jingdi Lei, Junxian Li, Xunzhi Wang, Yujie Liu, Zonglin Yang, Jiatong Li, Weida Wang, Suorong Yang, Jianbo Wu, Peng Ye, Wanli Ouyang, Dongzhan Zhou*
- 353 Insight-V: Exploring Long-Chain Visual Reasoning with
 * Multimodal Large Language Models, *Yuhao Dong, Zuyan Liu, Hai-Long Sun, Jingkang Yang, Winston Hu, Yongming Rao, Ziwei Liu*
- 354 Synthetic Visual Genome, *Jae Sung Park, Zixian Ma, Linjie Li, Chenhao Zheng, Cheng-Yu Hsieh, Ximing Lu, Khyathi Chandu, Quan Kong, Norimasa Kobori, Ali Farhadi, Yejin Choi, Ranjay Krishna*
- 355 Unveiling the Ignorance of MLLMs: Seeing Clearly, Answering Incorrectly, *Yexin Liu, Zhengyang Liang, Yueze Wang, Xianfeng Wu, Feilong Tang, Muyang He, Jian Li, Zheng Liu, Harry Yang, Sernam Lim, Bo Zhao*
- 356 DyFo: A Training-Free Dynamic Focus Visual Search for
 * Enhancing LLMs in Fine-Grained Visual Understanding, *Geng Li, Jinglin Xu, Yunzhen Zhao, Yuxin Peng*
- 357 Taxonomy-Aware Evaluation of Vision-Language Models, *Vésteinn Snæbjarnarson, Kevin Du, Niklas Stoehr, Serge Belongie, Ryan Cotterell, Nico Lang, Stella Frank*
- 358 Fine-Grained Erasure in Text-to-Image Diffusion-based Foundation Models, *Kartik Thakral, Tamar Glaser, Tal Hassner, Mayank Vatsa, Richa Singh*
- 359 K-Sort Arena: Efficient and Reliable Benchmarking for Generative Models via K-wise Human Preferences, *Zhikai Li, Xuewen Liu, Dongrong Joe Fu, Jianquan Li, Qingyi Gu, Kurt Keutzer, Zhen Dong*
- 360 AVF-MAE++: Scaling Affective Video Facial Masked Autoencoders via Efficient Audio-Visual Self-Supervised Learning, *Xuecheng Wu, Heli Sun, Yifan Wang, Jiayu Nie, Jie Zhang, Yabing Wang, Junxiao Xue, Liang He*
- 361 FaceBench: A Multi-View Multi-Level Facial Attribute VQA Dataset for Benchmarking Face Perception MLLMs, *Xiaoqin Wang, Xusen Ma, Xianxu Hou, Meidan Ding, Yudong Li, Junliang Chen, Wenting Chen, Xiaoyang Peng, Linlin Shen*
- 362 UPME: An Unsupervised Peer Review Framework for Multimodal Large Language Model Evaluation, *Qihui Zhang, Munan Ning, Zheyuan Liu, Yue Huang, Shuo Yang, Yanbo Wang, Jiayi Ye, Xiao Chen, Yibing Song, Li Yuan*
- 363 WeakMCN: Multi-task Collaborative Network for Weakly Supervised Referring Expression Comprehension and Segmentation, *Silin Cheng, Yang Liu, Xinwei He, Sebastien Ourselin, Lei Tan, Gen Luo*
- 364 COUNTS: Benchmarking Object Detectors and Multimodal Large
 * Language Models under Distribution Shifts, *Jiansheng Li, Xingxuan Zhang, Hao Zou, Yige Guo, Renzhe Xu, Yilong Liu, Chuzhao Zhu, Yue He, Peng Cui*
- 365 Augmenting Multimodal LLMs with Self-Reflective Tokens for
 ☆ Knowledge-based Visual Question Answering, *Federico Cocchi, Nicholas Moratelli, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara*
- 366 Variance-Based Membership Inference Attacks Against Large-Scale Image Captioning Models, *Daniel Samira, Edan Habler, Yuval Elovici, Asaf Shabtai*
- 367 Modeling Thousands of Human Annotators for Generalizable
 * Text-to-Image Person Re-identification, *Jiayu Jiang, Changxing Ding, Wentao Tan, Junhong Wang, Jin Tao, Xiangmin Xu*
- 368 Rethinking Noisy Video-Text Retrieval via Relation-aware Alignment, *Huakai Lai, Guoxin Xiong, Huayu Mai, Xiang Liu, Tianzhu Zhang*

- 369 Unleashing the Potential of Consistency Learning for Detecting and Grounding Multi-Modal Media Manipulation, *Yiheng Li, Yang Yang, Zichang Tan, Huan Liu, Weihua Chen, Xu Zhou, Zhen Lei*
- 370 Seeing the Abstract: Translating the Abstract Language for Vision Language Models, *Davide Talon, Federico Girella, Ziyue Liu, Marco Cristani, Yiming Wang*
- 371 NeighborRetr: Balancing Hub Centrality in Cross-Modal Retrieval, *Zengrong Lin, Zheng Wang, Tianwen Qian, Pan Mu, Sixian Chan, Cong Bai*
- 372 Bridging Modalities: Improving Universal Multimodal Retrieval by Multimodal Large Language Models, *Xin Zhang, Yanzhao Zhang, Wen Xie, Mingxin Li, Ziqi Dai, Dingkun Long, Pengjun Xie, Meishan Zhang, Wenjie Li, Min Zhang*
- 373 Recurrence-Enhanced Vision-and-Language Transformers for Robust Multimodal Document Retrieval, *Davide Caffagni, Sara Sarto, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara*
- 374 Img-Diff: Contrastive Data Synthesis for Multimodal Large Language Models, *Qirui Jiao, Daoyuan Chen, Yilun Huang, Bolin Ding, Yaliang Li, Ying Shen*
- 375 CLIP Under the Microscope: A Fine-Grained Analysis of Multi-Object Representation, *Reza Abbasi, Ali Nazari, Aminreza Sefid, Mohammadali Banayeezanade, Mohammad Hossein Rohban, Mahdieh Soleymani Baghsah*
- 376 Linguistics-aware Masked Image Modeling for Self-supervised Scene Text Recognition, *Yifei Zhang, Chang Liu, Jin Wei, Xiaomeng Yang, Yu Zhou, Can Ma, Xiangyang Ji*
- 377 SemiETS: Integrating Spatial and Content Consistencies for Semi-Supervised End-to-end Text Spotting, *Dongliang Luo, Hanshen Zhu, Ziyang Zhang, Dingkan Liang, Xudong Xie, Yuliang Liu, Xiang Bai*
- 378 Your Large Vision-Language Model Only Needs A Few Attention Heads For Visual Grounding, *Seil Kang, Jinyeong Kim, Junhyeok Kim, Seong Jae Hwang*
- 379 Improving Autoregressive Visual Generation with Cluster-Oriented Token Prediction, *Teng Hu, Jiangning Zhang, Ran Yi, Jieyu Weng, Yabiao Wang, Xianfang Zeng, Zhucun Xue, Lizhuang Ma*
- 380 SymDPO: Boosting In-Context Learning of Large Multimodal Models with Symbol Demonstration Direct Preference Optimization, *Hongrui Jia, Chaoya Jiang, Haiyang Xu, Wei Ye, Mengfan Dong, Ming Yan, Ji Zhang, Fei Huang, Shikun Zhang*
- 381 CASP: Compression of Large Multimodal Models Based on Attention Sparsity, *Mohsen Gholami, Mohammad Akbari, Kevin Cannons, Yong Zhang*
- 382 Lifting the Veil on Visual Information Flow in MLLMs: Unlocking Pathways to Faster Inference, *Hao Yin, Guangzong Si, Zilei Wang*
- 383 DivPrune: Diversity-based Visual Token Pruning for Large Multimodal Models, *Saeed Ranjbar Alvar, Gursimran Singh, Mohammad Akbari, Yong Zhang*
- 384 Libra-Merging: Importance-redundancy and Pruning-merging Trade-off for Acceleration Plug-in in Large Vision-Language Model, *Longrong Yang, Dong Shen, Chaoxiang Cai, Kaibing Chen, Fan Yang, Tingting Gao, Di Zhang, Xi Li*
- 385 AdaMMS: Model Merging for Heterogeneous Multimodal Large Language Models with Unsupervised Coefficient Optimization, *Yiyang Du, Xiaochen Wang, Chi Chen, Jiabo Ye, Yiru Wang, Peng Li, Ming Yan, Ji Zhang, Fei Huang, Zhifang Sui, Maosong Sun, Yang Liu*
- 386 Debiasing Multimodal Large Language Models via Noise-Aware Preference Optimization, *Zefeng Zhang, Hengzhu Tang, Jiawei Sheng, Zhenyu Zhang, Yiming Ren, Zhenyang Li, Dawei Yin, Duohe Ma, Tingwen Liu*
- 387 From Head to Tail: Towards Balanced Representation in Large Vision-Language Models through Adaptive Data Calibration, *Mingyang Song, Xiaoye Qu, Jiawei Zhou, Yu Cheng*
- 388 EfficientLLaVA: Generalizable Auto-Pruning for Large Vision-language Models, *Yinan Liang, Ziwei Wang, Xiuwei Xu, Jie Zhou, Jiwen Lu*
- 389 Lifelong Knowledge Editing for Vision Language Models with Low-Rank Mixture-of-Experts, *Qizhou Chen, Chengyu Wang, Dakan Wang, Taolin Zhang, Wangyue Li, Xiaofeng He*
- 390 Distraction is All You Need for Multimodal Large Language Model Jailbreaking, *Zuopeng Yang, Jiluan Fan, Anli Yan, Erdun Gao, Xin Lin, Tao Li, Kanghua Mo, Changyu Dong*
- 391 Revisiting Backdoor Attacks against Large Vision-Language Models from Domain Shift, *Siyuan Liang, Jiawei Liang, Tianyu Pang, Chao Du, Aishan Liu, Mingli Zhu, Xiaochun Cao, Dacheng Tao*
- 392 ICP: Immediate Compensation Pruning for Mid-to-high Sparsity, *Xin Luo, Xueming Fu, Zihang Jiang, S. Kevin Zhou*
- 393 Vision-Language Model IP Protection via Prompt-based Learning, *Lianyu Wang, Meng Wang, Huazhu Fu, Daoqiang Zhang*
- 394 A3: Few-shot Prompt Learning of Unlearnable Examples with Cross-Modal Adversarial Feature Alignment, *Xuan Wang, Xitong Gao, Dongping Liao, Tianrui Qin, Yu-liang Lu, Cheng-zhong Xu*
- 395 Explaining Domain Shifts in Language: Concept Erasing for Interpretable Image Classification, *Zequan Zeng, Yudi Su, Jianqiao Sun, Tiansheng Wen, Hao Zhang, Zhengjue Wang, Bo Chen, Hongwei Liu, Jiawei Ma*
- 396 VISCO: Benchmarking Fine-Grained Critique and Correction Towards Self-Improvement in Visual Reasoning, *Xueqing Wu, Yuheng Ding, Bingxuan Li, Pan Lu, Da Yin, Kai-Wei Chang, Nanyun Peng*
- 397 Free on the Fly: Enhancing Flexibility in Test-Time Adaptation with Online EM, *Qiyuan Dai, Sibe Yang*
- 398 SAIST: Segment Any Infrared Small Target Model Guided by Contrastive Language-Image Pretraining, *Mingjin Zhang, Xiaolong Li, Fei Gao, Jie Guo, Xinbo Gao, Jing Zhang*
- 399 Domain Generalization in CLIP via Learning with Diverse Text Prompts, *Changsong Wen, Zelin Peng, Yu Huang, Xiaokang Yang, Wei Shen*
- 400 Enhanced Visual-Semantic Interaction with Tailored Prompts for Pedestrian Attribute Recognition, *Junyi Wu, Yan Huang, Min Gao, Yuzhen Niu, Yuzhong Chen, Qiang Wu*
- 401 LOCORE: Image Re-ranking with Long-Context Sequence Modeling, *Zilin Xiao, Pavel Suma, Ayush Sachdeva, Hao-Jen Wang, Giorgos Kordopatis-Zilos, Giorgos Tolias, Vicente Ordonez*
- 402 Visual Consensus Prompting for Co-Salient Object Detection, *Jie Wang, Nana Yu, Zihao Zhang, Yahong Han*
- 403 Explainable Saliency: Articulating Reasoning with Contextual Prioritization, *Nuo Chen, Ming Jiang, Qi Zhao*
- 404 Finer-CAM: Spotting the Difference Reveals Finer Details for Visual Explanation, *Ziheng Zhang, Jianyang Gu, Arpita Chowdhury, Zheda Mai, David Carlyn, Tanya Berger-Wolf, Yu Su, Wei-Lun Chao*
- 405 Perceptual Inductive Bias Is What You Need Before Contrastive Learning, *Junru Zhao, Tianqin Li, Dunhan Jiang, Shenghao Wu, Alan Ramirez, Tai Sing Lee*
- 406 Scaling Vision Pre-Training to 4K Resolution, *Baifeng Shi, Boyi Li, Han Cai, Yao Lu, Sifei Liu, Marco Pavone, Jan Kautz, Song Han, Trevor Darrell, Pavlo Molchanov, Hongxu Yin*
- 407 Multimodal Autoregressive Pre-training of Large Vision Encoders, *Enrico Fini, Mustafa Shukor, Xiujun Li, Philipp Dufter, Michal Klein, David Haldimann, Sai Aitharaju, Victor G. Turrissi da Costa, Louis Béthune, Zhe Gan, Alexander Toshev, Marcin Eichner, Moin Nabi, Yinfei Yang, Joshua Susskind, Alaaeldin El-Nouby*
- 408 Sensitivity-Aware Efficient Fine-Tuning via Compact Dynamic-Rank Adaptation, *Tianran Chen, Jiarui Chen, Baoquan Zhang, Zhehao Yu, Shidong Chen, Rui Ye, Xutao Li, Yunming Ye*
- 409 UNEM: UNrolled Generalized EM for Transductive Few-Shot Learning, *Long Zhou, Fereshteh Shakeri, Aymen Sadraoui, Mounir Kaaniche, Jean-Christophe Pesquet, Ismail Ben Ayed*
- 410 MAP: Unleashing Hybrid Mamba-Transformer Vision Backbone's Potential with Masked Autoregressive Pretraining, *Yunze Liu, Li Yi*
- 411 APHQ-ViT: Post-Training Quantization with Average Perturbation Hessian Based Reconstruction for Vision Transformers, *Zhuguanyu Wu, Jiayi Zhang, Jiaxin Chen, Jinyang Guo, Di Huang, Yunhong Wang*
- 412 Two is Better than One: Efficient Ensemble Defense for Robust and Compact Models, *Yoojin Jung, Byung Cheol Song*
- 413 Building Vision Models upon Heat Conduction, *Zhaozhi Wang, Yue Liu, Yunjie Tian, Yunfan Liu, Yaowei Wang, Qixiang Ye*

- 414 LSNet: See Large, Focus Small, *Ao Wang, Hui Chen, Zijia Lin, Jungong Han, Guiguang Ding*
- 415 SATA: Spatial Autocorrelation Token Analysis for Enhancing the Robustness of Vision Transformers, *Nick Nikzad, Yi Liao, Yongsheng Gao, Jun Zhou*
- 416 Token Crop: Faster ViTs for Quite a Few Tasks, *Benjamin Bergner, Christoph Lippert, Aravindh Mahendran*
- 417 Hypergraph Vision Transformers: Images are More than Nodes, More than Edges, *Joshua Fixelle*
- 418 Interpretable Image Classification via Non-parametric Part Prototype Learning, *Zhijie Zhu, Lei Fan, Maurice Pagnucco, Yang Song*
- 419 COSMIC: Clique-Oriented Semantic Multi-space Integration for Robust CLIP Test-Time Adaptation, *Fanding Huang, Jingyan Jiang, Qinting Jiang, Hebei Li, Faisal Nadeem Khan, Zhi Wang*
- 420 Fine-Grained Image-Text Correspondence with Cost Aggregation for Open-Vocabulary Part Segmentation, *Jiho Choi, Seonho Lee, Minhyun Lee, Seungcho Lee, Hyunjung Shim*
- 421 LPOSS: Label Propagation Over Patches and Pixels for Open-vocabulary Semantic Segmentation, *Vladan Stojnić, Yannis Kalantidis, Jiří Matas, Giorgos Tolias*
- 422 Semantic Library Adaptation: LoRA Retrieval and Fusion for Open-Vocabulary Semantic Segmentation, *Reza Qorbani, Gianluca Villani, Theodoros Panagiotakopoulos, Marc Botet Colomer, Linus Härenstam-Nielsen, Mattia Segu, Pier Luigi Dovesi, Jussi Karlgren, Daniel Cremers, Federico Tombari, Matteo Poggi*
- 423 Guidance for Semantic Segmentation, *Hritam Basak, Zhaozheng Yin*
- 424 Weakly Supervised Semantic Segmentation via Progressive Confidence Region Expansion, *Xiangfeng Xu, Pinyi Zhang, Wenxuan Huang, Yunhang Shen, Haosheng Chen, Jingzhong Lin, Wei Li, Gaoqi He, Jiao Xie, Shaohui Lin*
- 425 Beyond Background Shift: Rethinking Instance Replay in Continual Semantic Segmentation, *Hongmei Yin, Tingliang Feng, Fan Lyu, Fanhua Shang, Hongying Liu, Wei Feng, Liang Wan*
- 426 Dual-Agent Optimization framework for Cross-Domain Few-Shot Segmentation, *Zhaoyang Li, Yuan Wang, Wangkai Li, Tianzhu Zhang, Xiang Liu*
- 427 Beyond Image Classification: A Video Benchmark and Dual-Branch Hybrid Discrimination Framework for Compositional Zero-Shot Learning, *Dongyao Jiang, Haodong Jing, Yongqiang Ma, Nanning Zheng*
- 428 Targeted Forgetting of Image Subgroups in CLIP Models, *Zeliang Zhang, Gaowen Liu, Charles Fleming, Ramana Rao Kompella, Chenliang Xu*
- 429 Enhancing Few-Shot Class-Incremental Learning via Training-Free Bi-Level Modality Calibration, *Yiyang Chen, Tianyu Ding, Lei Wang, Jing Huo, Yang Gao, Wenbin Li*
- 430 Hyperbolic Category Discovery, *Yuanpei Liu, Zhenqi He, Kai Han*
- 431 Solving Instance Detection from an Open-World Perspective, *Qianqian Shen, Yunhan Zhao, Nahyun Kwon, Jeeun Kim, Yanan Li, Shu Kong*
- 432 Learning Class Prototypes for Unified Sparse-Supervised 3D Object Detection, *Yun Zhu, Le Hui, Hang Yang, Jianjun Qian, Jin Xie, Jian Yang*
- 433 Generalized Diffusion Detector: Mining Robust Features from Diffusion Models for Domain-Generalized Detection, *Boyong He, Yuxiang Ji, Qianwen Ye, Zhuoyue Tan, Liaoni Wu*
- 434 Mr. DETR: Instructive Multi-Route Training for Detection Transformers, *Chang-Bin Zhang, Yujie Zhong, Kai Han*
- 435 PatchDEMUX: A Certifiably Robust Framework for Multi-label Classifiers Against Adversarial Patches, *Dennis Jacob, Chong Xiang, Prateek Mittal*
- 436 Disentangling Safe and Unsafe Image Corruptions via Anisotropy and Locality, *Ramchandran Muthukumar, Ambar Pal, Jeremias Sulam, Rene Vidal*
- 437 Beyond Single-Modal Boundary: Cross-Modal Anomaly Detection through Visual Prototype and Harmonization, *Kai Mao, Ping Wei, Yiyang Lian, Yangyang Wang, Nanning Zheng*
- 438 Exploring Intrinsic Normal Prototypes within a Single Image for Universal Anomaly Detection, *Wei Luo, Yunkang Cao, Haiming Yao, Xiaotian Zhang, Jianan Lou, Yuqi Cheng, Weiming Shen, Wenying Yu*
- 439 Multi-Sensor Object Anomaly Detection: Unifying Appearance, Geometry, and Internal Properties, *Wenqiao Li, Bozhong Zheng, Xiaohao Xu, Jinye Gan, Fading Lu, Xiang Li, Na Ni, Zheng Tian, Xiaonan Huang, Shenghua Gao, Yingna Wu*
- 440 UniNet: A Contrastive Learning-guided Unified Framework with Feature Selection for Anomaly Detection, *Shun Wei, Jieliang Jiang, Xiaolong Xu*
- 441 NN-Former: Rethinking Graph Structure in Neural Architecture Representation, *Ruihan Xu, Haokui Zhang, Yaowei Wang, Wei Zeng, Shiliang Zhang*
- 442 Enhancing Dataset Distillation via Non-Critical Region Refinement, *Minh-Tuan Tran, Trung Le, Xuan-May Le, Thanh-Toan Do, Dinh Phung*
- 443 Efficient ANN-Guided Distillation: Aligning Rate-based Features of Spiking Neural Networks through Hybrid Block-wise Replacement, *Shu Yang, Chengting Yu, Lei Liu, Hanzhi Ma, Aili Wang, Erping Li*
- 444 Breaking the Memory Barrier of Contrastive Loss via Tile-Based Strategy, *Zesen Cheng, Hang Zhang, Kehan Li, Sicong Leng, Zhiqiang Hu, Fei Wu, Deli Zhao, Xin Li, Lidong Bing*
- 445 LoTUS: Large-Scale Machine Unlearning with a Taste of Uncertainty, *Christoforos N. Spertalis, Theodoros Semertzidis, Efstratios Gavves, Petros Daras*
- 446 DeRS: Towards Extremely Efficient Upcycled Mixture-of-Experts Models, *Yongqi Huang, Peng Ye, Chenyu Huang, Jianjian Cao, Lin Zhang, Baopu Li, Gang Yu, Tao Chen*
- 447 Towards Consistent Multi-Task Learning: Unlocking the Potential of Task-Specific Parameters, *Xiaohan Qin, Xiaoxing Wang, Junchi Yan*
- 448 Do Your Best and Get Enough Rest for Continual Learning, *Hankyul Kang, Gregor Seifer, Donghyun Lee, Jongbin Ryu*
- 449 Self-Expansion of Pre-trained Models with Mixture of Adapters for Continual Learning, *Huiyi Wang, Haodong Lu, Lina Yao, Dong Gong*
- 450 Task-Agnostic Guided Feature Expansion for Class-Incremental Learning, *Bowen Zheng, Da-Wei Zhou, Han-Jia Ye, De-Chuan Zhan*
- 451 OSLoPrompt: Bridging Low-Supervision Challenges and Open-Set Domain Generalization in CLIP, *Mohamad Hassan N C, Divyam Gupta, Mainak Singha, Sai Bhargav Rongali, Ankit Jha, Muhammad Haris Khan, Biplab Banerjee*
- 452 Reducing Class-wise Confusion for Incremental Learning with Disentangled Manifolds, *Huitong Chen, Yu Wang, Yan Fan, Guosong Jiang, Qinghua Hu*
- 453 Unsupervised Continual Domain Shift Learning with Multi-Prototype Modeling, *Haopeng Sun, Yingwei Zhang, Lumin Xu, Sheng Jin, Ping Luo, Chen Qian, Wentao Liu, Yiqiang Chen*
- 454 A Theory of Learning Unified Model via Knowledge Integration from Label Space Varying Domains, *Dexuan Zhang, Thomas Westfechtel, Tatsuya Harada*
- 455 Rethinking Epistemic and Aleatoric Uncertainty for Active Open-Set Annotation: An Energy-Based Approach, *Chen-Chen Zong, Sheng-Jun Huang*
- 456 Towards Cost-Effective Learning: A Synergy of Semi-Supervised and Active Learning, *Tianxiang Yin, Ningzhong Liu, Han Sun*
- 457 Mind the Gap: Confidence Discrepancy Can Guide Federated Semi-Supervised Learning Across Pseudo-Mismatch, *Yijie Liu, Xinyi Shang, Yiqun Zhang, Yang Lu, Chen Gong, Jing-Hao Xue, Hanzhi Wang*
- 458 Beyond Clean Training Data: A Versatile and Model-Agnostic Framework for Out-of-Distribution Detection with Contaminated Training Data, *Yuchuan Li, Jae-Mo Kang, Il-Min Kim*
- 459 DPU: Dynamic Prototype Updating for Multimodal Out-of-Distribution Detection, *Shawn Li, Huixian Gong, Hao Dong, Tiankai Yang, Zhengzhong Tu, Yue Zhao*
- 460 dFLMoE: Decentralized Federated Learning via Mixture of Experts for Medical Data Analysis, *Luyuan Xie, Tianyu Luan, Wenyuan Cai, Guochen Yan, Zhaoyu Chen, Nan Xi, Yuejian Fang, Qingni Shen, Zhonghai Wu, Junsong Yuan*

- Randomization, *Peirong Liu, Ana Lawry Aguila, Juan E. Iglesias*
- 485 Blood Flow Speed Estimation with Optical Coherence Tomography Angiography Images, *Wensheng Cheng, Zhenghong Li, Jiaxiang Ren, Hyomin Jeong, Congwu Du, Yingtian Pan, Haibin Ling*
- 486 3D Dental Model Segmentation with Geometrical Boundary Preserving, *Shufan Xi, Zexian Liu, Junlin Chang, Hongyu Wu, Xiaogang Wang, Aimin Hao*
- 16:00 - 18:00 Demos (ExHall B)**
- 1 FotoBot, *Dawei Wang*
- 2 UltraFusion HDR, *Yujin Wang; Zixuan Chen; Xin Cai; Zhiyuan You; Zheming Lu; Fan Zhang; Shi Guo; Tianfan Xue*
- 3 DL3DV-10K, *Lu Ling*
- 4 R²-Tuning: Efficient Image-to-Video Transfer Learning for Video Temporal Grounding, *Ye Liu, Jixuan He, Chang Wen Chen*
- 5 Granite Vision: A Demo for Efficient Visual Document Understanding, *Pengyuan Li and Granite Vision Team*
- 6 Animating MetaGen Stickers, *Jesse Smith, Nicky He, David Geisert, Yuting Ye*
- 7 Efficient Depth Estimation for Unstable Stereo Camera Systems on AR Glasses, *Yongfan Liu, Hyoukjun Kwon*
- 8 Dense Dispersed Structured Light for Hyperspectral 3D Imaging of Dynamic Scenes, *Suhyun Shin*
- 9 MAST3R-SLAM Live Demo, *Riku Murai*, Eric Dexheimer*, Andrew J. Davison*
- 10 HD-EPIC Dataset, Annotations and Benchmark, *Dima Damen*
- 11 HVI+CIDNet: Real-Time Low-Light Enhancement with Less Artifacts!, *Yixu Feng, Qingsen Yan*
- 12 Interactive4D: Interactive 4D LiDAR Segmentation, *Idil Esen Zulfikar, Theodora Kontogianni, Ilya Fradlin, Kadir Yilmaz, Bastian Leibe*
- 13 Interpretable inverse rendering and its application, *Hoon-Gyu Chung*
- 14 ShowUI: A Multimodal Agent for Computer Use, *Kevin Qinghong Lin, Linjie Li, Difei Gao, Zhengyuan Yang, Shiwei Wu, Zechen Bai, Weixian Lei, Lijuan Wang, Mike Zheng Shou*

Notes:

Saturday, June 14

7:30 - 19:30 Registration / Badge Pickup (ExHall Concourse)
7:00 - 17:00 Press Room (203 B)
7:00 - 17:00 Mother's Room (Level 1 near Room 101 and on Level 3 near Exhibit Hall D)
7:00 - 17:00 Prayer or Quiet Room (203 A)
7:30 - 9:00 Breakfast (ExHall C)
8:00 - 8:30 Poster Setup (ExHall D)

9:00 - 10:15 Oral Session 3A: 3D Computer Vision (Karl Dean Ballroom)

🏆 - Award candidate paper

- 1 🏆 MegaSaM: Accurate, Fast and Robust Structure and Motion from Casual Dynamic Videos, *Zhengqi Li, Richard Tucker, Forrester Cole, Qianqian Wang, Linyi Jin, Vickie Ye, Angjoo Kanazawa, Aleksander Holynski, Noah Snavely*
- 2 Stereo4D: Learning How Things Move in 3D from Internet Stereo Videos, *Linyi Jin, Richard Tucker, Zhengqi Li, David Fouhey, Noah Snavely, Aleksander Holynski*
- 3 Continuous 3D Perception Model with Persistent State, *Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A. Efros, Angjoo Kanazawa*
- 4 🏆 TacoDepth: Towards Efficient Radar-Camera Depth Estimation with One-stage Fusion, *Yiran Wang, Jiaqi Li, Chaoyi Hong, Ruibo Li, Liusheng Sun, Xiao Song, Zhe Wang, Zhiguo Cao, Guosheng Lin*
- 5 Neural Inverse Rendering from Propagating Light, *Anagh Malik, Benjamin Attal, Andrew Xie, Matthew O'Toole, David B. Lindell*

9:00 - 10:15 Oral Session 3B: Multimodal Computer Vision (ExHall A2)

- 1 SegEarth-OV: Towards Training-Free Open-Vocabulary Segmentation for Remote Sensing Images, *Kaiyu Li, Ruixun Liu, Xiangyong Cao, Xueru Bai, Feng Zhou, Deyu Meng, Zhi Wang*
- 2 Towards Universal Dataset Distillation via Task-Driven Diffusion, *Ding Qi, Jian Li, Junyao Gao, Shuguang Dou, Ying Tai, Jianlong Hu, Bo Zhao, Yabiao Wang, Chengjie Wang, Cairong Zhao*
- 3 IceDiff: High Resolution and High-Quality Arctic Sea Ice Forecasting with Generative Diffusion Prior, *Jingyi Xu, Siwei Tu, Weidong Yang, Ben Fei, Shuhao Li, Keyi Liu, Yeqi Luo, Lipeng Ma, Lei Bai*
- 4 Efficient Test-time Adaptive Object Detection via Sensitivity-Guided Pruning, *Kunyu Wang, Xueyang Fu, Xin Lu, Chengjie Ge, Chengzhi Cao, Wei Zhai, Zheng-Jun Zha*
- 5 Keep the Balance: A Parameter-Efficient Symmetrical Framework for RGB+X Semantic Segmentation, *Jiaxin Cai, Jingze Su, Qi Li, Wenjie Yang, Shu Wang, Tiesong Zhao, Shengfeng He, Wenxi Liu*

9:00 - 10:15 Oral Session 3C: Vision and Language (Davidson Ballroom)

- 1 Identifying and Mitigating Position Bias of Multi-image Vision-Language Models, *Xinyu Tian, Shu Zou, Zhaoyuan Yang, Jing Zhang*
- 2 Mitigating Hallucinations in Large Vision-Language Models via DPO: On-Policy Data Hold the Key, *Zhihe Yang, Xufang Luo, Dongqi Han, Yunjian Xu, Dongsheng Li*
- 3 Q-Eval-100K: Evaluating Visual Quality and Alignment Level for Text-to-Vision Content, *Zicheng Zhang, Tengchuan Kou, Shushi Wang, Chunyi Li, Wei Sun, Wei Wang, Xiaoyu Li, Zongyu Wang, Xuezhi Cao, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai*
- 4 Thinking in Space: How Multimodal Large Language Models See, Remember, and Recall Spaces, *Jihan Yang, Shusheng Yang, Anjali W. Gupta, Rilyn Han, Li Fei-Fei, Saining Xie*
- 5 From Multimodal LLMs to Generalist Embodied Agents: Methods and Lessons, *Andrew Szot, Bogdan Mazouze, Omar Attia, Aleksei Timofeev, Harsh Agrawal, Devon Hjelm, Zhe Gan, Zsolt Kira, Alexander Toshev*

10:00 - 10:30 Poster Setup (ExHall D)

10:00 - 11:00 Coffee Break (ExHall D)

10:30 - 12:30 Poster Session 3 & Exhibit Hall (ExHall D)

✱ - Highlight paper

🏆 - Award candidate paper

○ - Oral Paper

☆ - Outstanding Reviewer

- 1 LLM-driven Multimodal and Multi-Identity Listening Head Generation, *Peiwen Lai, Weizhi Zhong, Yipeng Qin, Xiaohang Ren, Baoyuan Wang, Guanbin Li*
- 2 INFP: Audio-Driven Interactive Head Generation in Dyadic Conversations, *Yongming Zhu, Longhao Zhang, Zhengkun Rong, Tianshu Hu, Shuang Liang, Zhipeng Ge*
- 3 AudCast: Audio-Driven Human Video Generation by Cascaded Diffusion Transformers, *Jiazhi Guan, Kaisiyuan Wang, Zhiliang Xu, Quanwei Yang, Yasheng Sun, Shengyi He, Borong Liang, Yukang Cao, Yingying Li, Haocheng Feng, Errui Ding, Jingdong Wang, Youjian Zhao, Hang Zhou, Ziwei Liu*
- 4 InsTaG: Learning Personalized 3D Talking Head from Few-Second Video, *Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Jun Zhou, Lin Gu*
- 5 Dynamic Stereotype Theory Induced Micro-expression Recognition with Oriented Deformation, *Bohao Zhang, Xuejiao Wang, Changbo Wang, Gaoqi He*
- 6 Coherent 3D Portrait Video Reconstruction via Triplane Fusion, *Shengze Wang, Xueting Li, Chao Liu, Matthew Chan, Michael Stengel, Henry Fuchs, Shalini De Mello, Koki Nagano*
- 7 TaoAvatar: Real-Time Lifelike Full-Body Talking Avatars for Augmented Reality via 3D Gaussian Splatting, *Jianchuan Chen, Jingchuan Hu, Gaige Wang, Zhonghua Jiang, Tiansong Zhou, Zhiwen Chen, Chengfei Lv*
- 8 Synthetic Prior for Few-Shot Drivable Head Avatar Inversion, *Wojciech Zielonka, Stephan J. Garbin, Alexandros Lattas, George Kopanas, Paulo Gotardo, Thabo Beeler, Justus Thies, Timo Bolkart*
- 9 RGBAvatar: Reduced Gaussian Blendshapes for Online Modeling of Head Avatars, *Linzhou Li, Yumeng Li, Yanlin Weng, Youyi Zheng, Kun Zhou*
- 10 AvatarArtist: Open-Domain 4D Avatarization, *Hongyu Liu, Xuan Wang, Ziyu Wan, Yue Ma, Jingye Chen, Yanbo Fan, Yujun Shen, Yibing Song, Qifeng Chen*
- 11 Arc2Avatar: Generating Expressive 3D Avatars from a Single Image via ID Guidance, *Dimitrios Gerogiannis, Foivos Paraperas, Papantoniou, Rolandos Alexandros Potamias, Alexandros Lattas, Stefanos Zafeiriou*
- 12 Make-It-Animatable: An Efficient Framework for Authoring Animation-Ready 3D Characters, *Zhiyang Guo, Jinxu Xiang, Kai Ma, Wengang Zhou, Houqiang Li, Ran Zhang*
- 13 PhysAnimator: Physics-Guided Generative Cartoon Animation, *Tianyi Xie, Yiwei Zhao, Ying Jiang, Chenfanfu Jiang*
- 14 Zero-Shot Head Swapping in Real-World Scenarios, *Taewoong Kang, Sohyun Jeong, Hyojin Jang, Jaegul Choo*
- 15 CaricatureBooth: Data-Free Interactive Caricature Generation in a Photo Booth, *Zhiyu Qu, Yunqi Miao, Zhensong Zhang, Jifei Song, Jiankang Deng, Yi-Zhe Song*
- 16 FFaceNeRF: Few-shot Face Editing in Neural Radiance Fields, *Kwan Yun, Chaelin Kim, Hangeul Shin, Junyong Noh*
- 17 D³-Human: Dynamic Disentangled Digital Human from Monocular Video, *Honghu Chen, Bo Peng, Yunfan Tao, Juyong Zhang*
- 18 DiffLocks: Generating 3D Hair from a Single Image using Diffusion Models, *Radu Alexandru Rosu, Keyu Wu, Yao Feng, Youyi Zheng, Michael J. Black*
- 19 Remote Photoplethysmography in Real-World and Extreme Lighting Scenarios, *Hang Shao, Lei Luo, Jianjun Qian, Mengkai Yan, Shuo Chen, Jian Yang*
- 20 GCC: Generative Color Constancy via Diffusing a Color Checker, *Chen-Wei Chang, Cheng-De Fan, Chia-Che Chang, Yi-Chen Lo, Yu-Chee Tseng, Jiun-Long Huang, Yu-Lun Liu*
- 21 DarkIR: Robust Low-Light Image Restoration, *Daniel Feijoo, Juan C. Benito, Alvaro Garcia, Marcos V. Conde*
- 22 PolarFree: Polarization-based Reflection-Free Imaging, *Mingde Yao, Menglu Wang, King-Man Tam, Lingen Li, Tianfan Xue, Jinwei Gu*
- 23 OpticalNet: An Optical Imaging Dataset and Benchmark Beyond the Diffraction Limit, *Benquan Wang, Ruyi An, Jin-Kyu So, Sergei Kurdumov, Eng Aik Chan, Giorgio Adamo, Yuhang Peng*

- Yewen Li, Bo An
- 24 A Physics-Informed Blur Learning Framework for Imaging Systems, *Liquan Chen, Yuxuan Li, Jun Dai, Jinwei Gu, Tianfan Xue*
- 25 MaDCoW: Marginal Distortion Correction for Wide-Angle
☆ Photography with Arbitrary Objects, *Kevin Zhang, Jia-Bin Huang, Jose Ecivarria, Stephen DiVerdi, Aaron Hertzmann*
- 26 Generative Multiview Relighting for 3D Reconstruction under
* Extreme Illumination Variation, *Hadi Alzayer, Philipp Henzler, Jonathan T. Barron, Jia-Bin Huang, Pratul P. Srinivasan, Dor Verbin*
- 27 IRGS: Inter-Reflective Gaussian Splatting with 2D Gaussian Ray
Tracing, *Chun Gu, Xiaofei Wei, Zixuan Zeng, Yuxuan Yao, Li Zhang*
- 28 Volumetrically Consistent 3D Gaussian Rasterization,
* *Chinmay Talegaonkar, Yash Belhe, Ravi Ramamoorthi, Nicholas Antipa*
- 29 MultimodalStudio: A Heterogeneous Sensor Dataset and Framework
for Neural Rendering across Multiple Imaging Modalities, *Federico Lincetto, Gianluca Agresti, Mattia Rossi, Pietro Zanuttigh*
- 30 Neural Inverse Rendering from Propagating Light, *Anagh Malik, Benjamin Attal, Andrew Xie, Matthew O'Toole, David B. Lindell*
- 31 PBR-NeRF: Inverse Rendering with Physics-Based Neural Fields,
Sean Wu, Shamik Basu, Tim Broedermann, Luc Van Gool, Christos Sakaridis
- 32 MAGE : Single Image to Material-Aware 3D via the Multi-View
G-Buffer Estimation Model, *Haoyuan Wang, Zhenwei Wang, Xiaoxiao Long, Cheng Lin, Gerhard Hancke, Rynson W.H. Lau*
- 33 3D-HGS: 3D Half-Gaussian Splatting, *Haolin Li, Jinyang Liu, Mario Sznajder, Octavia Camps*
- 34 Spatiotemporal Skip Guidance for Enhanced Video Diffusion
Sampling, *Junha Hyung, Kinam Kim, Susung Hong, Min-Jung Kim, Jaegul Choo*
- 35 Unleashing the Potential of Multi-modal Foundation Models and
Video Diffusion for 4D Dynamic Physical Scene Simulation, *Zhuoman Liu, Weicai Ye, Yan Luximon, Pengfei Wan, Di Zhang*
- 36 ProbeSDF: Light Field Probes For Neural Surface Reconstruction,
Briac Toussaint, Diego Thomas, Jean-Sébastien Franco
- 37 Progressive Rendering Distillation: Adapting Stable Diffusion for
☆ Instant Text-to-Mesh Generation without 3D Data, *Zhiyuan Ma, Xinyue Liang, Rongyuan Wu, Xiangyu Zhu, Zhen Lei, Lei Zhang*
- 38 FruitNinja: 3D Object Interior Texture Generation with Gaussian
Splatting, *Fangyu Wu, Yuhao Chen*
- 39 DI-PCG: Diffusion-based Efficient Inverse Procedural Content
Generation for High-quality 3D Asset Creation, *Wang Zhao, Yan-Pei Cao, Jiale Xu, Yuejiang Dong, Ying Shan*
- 40 CADCrafter: Generating Computer-Aided Design Models from
Unconstrained Images, *Cheng Chen, Jiacheng Wei, Tianrun Chen, Chi Zhang, Xiaofeng Yang, Shangzhan Zhang, Bingchen Yang, Chuan-Sheng Foo, Guosheng Lin, Qixing Huang, Fayao Liu*
- 41 MAR-3D: Progressive Masked Auto-regressor for High-Resolution
* 3D Generation, *Jinnan Chen, Lingting Zhu, Zeyu Hu, Shengju Qian, Yugang Chen, Xin Wang, Gim Hee Lee*
- 42 Scaling Mesh Generation via Compressive Tokenization, *Haohan Weng, Zibo Zhao, Biwen Lei, Xianghui Yang, Jian Liu, Zeqiang Lai, Zhuo Chen, Yuhong Liu, Jie Jiang, Chunchao Guo, Tong Zhang, Shenghua Gao, C.L. Philip Chen*
- 43 Hierarchical Gaussian Mixture Model Splatting for Efficient and
☆ Part Controllable 3D Generation, *Qitong Yang, Mingtao Feng, Zijie Wu, Weisheng Dong, Fangfang Wu, Yaonan Wang, Ajmal Mian*
- 44 Identity-preserving Distillation Sampling by Fixed-Point
Iterator, *SeonHwa Kim, Jiwon Kim, Soobin Park, Donghoon Ahn, Jiwon Kang, Seungrong Kim, Kyong Hwan Jin, Eunju Cha*
- 45 PhysicsGen: Can Generative Models Learn from Images to
Predict Complex Physical Relations?, *Martin Spitznagel, Jan Vaillant, Janis Keuper*
- 46 EditSplat: Multi-View Fusion and Attention-Guided Optimization
for View-Consistent 3D Scene Editing with 3D Gaussian
Splatting, *Dong In Lee, Hyeoncheol Park, Jiyoung Seo, Eunbyung Park, Hyunje Park, Ha Dam Baek, Sangheon Shin, Sangmin Kim, Sangpil Kim*
- 47 DashGaussian: Optimizing 3D Gaussian Splatting in 200 Seconds,
* *Yoyu Chen, Junjun Jiang, Kui Jiang, Xiao Tang, Zhihao Li, Xianming Liu, Yinyu Nie*
- 48 Efficient Decoupled Feature 3D Gaussian Splatting via
Hierarchical Compression, *Zhenqi Dai, Ting Liu, Yanning Zhang*
- 49 SOGS: Second-Order Anchor for Advanced 3D Gaussian
Splatting, *Jiahui Zhang, Fangneng Zhan, Ling Shao, Shijian Lu*
- 50 RestorGS: Depth-aware Gaussian Splatting for Efficient 3D Scene
Restoration, *Yuanjian Qiao, Mingwen Shao, Lingzhuang Meng, Kai Xu*
- 51 Seeing A 3D World in A Grain of Sand, *Yufan Zhang, Yu Ji, Yu Guo, Jinwei Ye*
- 52 CoA: Towards Real Image Dehazing via Compression-and-
Adaptation, *Long Ma, Yuxin Feng, Yan Zhang, Jinyuan Liu, Weimin Wang, Guang-Yong Chen, Chengpei Xu, Zhuo Su*
- 53 S2D-LFE: Sparse-to-Dense Light Field Event Generation,
Yutong Liu, Wenming Weng, Yueyi Zhang, Zhiwei Xiong
- 54 Depth-Guided Bundle Sampling for Efficient Generalizable Neural
Radiance Field Reconstruction, *Li Fang, Hao Zhu, Longlong Chen, Fei Hu, Long Ye, Zhan Ma*
- 55 FrugalNeRF: Fast Convergence for Extreme Few-shot Novel View
Synthesis without Learned Priors, *Chin-Yang Lin, Chung-Ho Wu, Chang-Han Yeh, Shih-Han Yen, Cheng Sun, Yu-Lun Liu*
- 56 MirrorVerse: Pushing Diffusion Models to Realistically Reflect the
☆ World, *Ankit Dhiman, Manan Shah, R Venkatesh Babu*
- 57 Matrix3D: Large Photogrammetry Model All-in-One, *Yuanxun Lu, Jinyang Zhang, Tian Fang, Jean-Daniel Nahmias, Yanghai Tsin, Long Qian, Xun Cao, Yao Yao, Shiwei Li*
- 58 SPC-GS: Gaussian Splatting with Semantic-Prompt Consistency
for Indoor Open-World Free-view Synthesis from Sparse Inputs, *Guibiao Liao, Qing Li, Zhenyu Bao, Guoping Qiu, Kanglin Liu*
- 59 Geometry-guided Online 3D Video Synthesis with Multi-View
Temporal Consistency, *Hyunho Ha, Lei Xiao, Christian Richardt, Thu Nguyen-Phuoc, Changil Kim, Min H. Kim, Douglas Lanman, Numair Khan*
- 60 EVolSplat: Efficient Volume-based Gaussian Splatting for Urban
View Synthesis, *Sheng Miao, Jiaxin Huang, Dongfeng Bai, Xu Yan, Hongyu Zhou, Yue Wang, Bingbing Liu, Andreas Geiger, Yiyi Liao*
- 61 MEAT: Multiview Diffusion Model for Human Generation on
Megapixels with Mesh Attention, *Yuhan Wang, Fangzhou Hong, Shuai Yang, Liming Jiang, Wayne Wu, Chen Change Loy*
- 62 Sparse2DGS: Geometry-Prioritized Gaussian Splatting for Surface
Reconstruction from Sparse Views, *Jiang Wu, Rui Li, Yu Zhu, Rong Guo, Jinqiu Sun, Yanning Zhang*
- 63 NeRF-Prior: Learning Neural Radiance Field as a Prior for Indoor
* Scene Reconstruction, *Wenyuan Zhang, Emily Yue-ting Jia, Junsheng Zhou, Baorui Ma, Kanle Shi, Yu-Shen Liu, Zhizhong Han*
- 64 Efficient Video Super-Resolution for Real-time Rendering with
Decoupled G-buffer Guidance, *Mingjun Zheng, Long Sun, Jiangxin Dong, Jinshan Pan*
- 65 MoDec-GS: Global-to-Local Motion Decomposition and Temporal
☆ Interval Adjustment for Compact Dynamic 3D Gaussian Splatting, *Sangwoon Kwak, Joonsoo Kim, Jun Young Jeong, Won-Sik Cheong, Jihyong Oh, Munchurl Kim*
- 66 RePerformer: Immersive Human-centric Volumetric Videos from
Playback to Photoreal Reperformance, *Yuheng Jiang, Zhehao Shen, Chengcheng Guo, Yu Hong, Zhuo Su, Yingliang Zhang, Marc Habermann, Lan Xu*
- 67 DecoupledGaussian: Object-Scene Decoupling for Physics-
Based Interaction, *Miaowei Wang, Yibo Zhang, Weiwei Xu, Rui Ma, Changqing Zou, Daniel Morris*
- 68 Thin-Shell-SfT: Fine-Grained Monocular Non-rigid 3D Surface
Tracking with Neural Deformation Fields, *Navami Kairanda, Marc Habermann, Shanthika Naik, Christian Theobalt, Vladislav Golyanik*
- 69 Co-Speech Gesture Video Generation with Implicit Motion-Audio
Entanglement, *Xinjie Li, Ziyi Chen, Xinlu Yu, Iek-Heng Chu, Peng Chang, Jing Xiao*
- 70 QuCOOP: A Versatile Framework for Solving Composite and
* Binary-Parametrised Problems on Quantum Annealers, *Natacha Kuete Meli, Vladislav Golyanik, Marcel Seelbach Benkner, Michael Moeller*
- 71 Image Reconstruction from Readout-Multiplexed Single-Photon
* Detector Arrays, *Shashwath Bharadwaj, Ruangrawee Kitichotkul, Akshay Agarwal, Vivek K Goyal*

- 72 Spk2SRImgNet: Super-Resolve Dynamic Scene from Spike Stream via Motion Aligned Collaborative Filtering, *Yuanlin Wang, Yiyang Zhang, Ruiqin Xiong, Jing Zhao, Jian Zhang, Xiaopeng Fan, Tiejun Huang*
- 73 EventPSR: Surface Normal and Reflectance Estimation from
* Photometric Stereo Using an Event Camera, *Bohan Yu, Jin Han, Boxin Shi, Imari Sato*
- 74 PanSplat: 4K Panorama Synthesis with Feed-Forward Gaussian Splatting, *Cheng Zhang, Haoifei Xu, Qianyi Wu, Camilo Cruz Gambardella, Dinh Phung, Jianfei Cai*
- 75 QuartDepth: Post-Training Quantization for Real-Time Depth Estimation on the Edge, *Xuan Shen, Weize Ma, Jing Liu, Changdi Yang, Rui Ding, Quanyi Wang, Henghui Ding, Wei Niu, Yanzhi Wang, Pu Zhao, Jun Lin, Jiuxiang Gu*
- 76 WildGS-SLAM: Monocular Gaussian Splatting SLAM in Dynamic Environments, *Jianhao Zheng, Zihan Zhu, Valentin Bieri, Marc Pollefeys, Songyou Peng, Iro Armeni*
- 77 Continuous 3D Perception Model with Persistent State,
* Qianqian Wang, Yifei Zhang, Aleksander Holynski, Alexei A. Efros, Angjoo Kanazawa
- 78 MegaSaM: Accurate, Fast and Robust Structure and Motion from
Casual Dynamic Videos, *Zhengqi Li, Richard Tucker, Forrester Cole, Qianqian Wang, Linyi Jin, Vickie Ye, Angjoo Kanazawa, Aleksander Holynski, Noah Snavely*
- 79 Joint Optimization of Neural Radiance Fields and Continuous Camera Motion from a Monocular Video, *Hoang Chuong Nguyen, Wei Mao, Jose M. Alvarez, Miaomiao Liu*
- 80 Pixel-aligned RGB-NIR Stereo Imaging and Dataset for Robot Vision, *Jinnyeong Kim, Seung-Hwan Baek*
- 81 MVSAnywhere: Zero-Shot Multi-View Stereo, *Sergio Izquierdo, Mohamed Sayed, Michael Firman, Guillermo Garcia-Hernando, Daniyar Turmukhambetov, Javier Civera, Oisin Mac Aodha, Gabriel Brostow, Jamie Watson*
- 82 Three-view Focal Length Recovery From Homographies,
* Yaqing Ding, Viktor Kocur, Zuzana Berger Haladova, Qianliang Wu, Shen Cai, Jian Yang, Zuzana Kukelova
- 83 Full-DoF Egomotion Estimation for Event Cameras Using
* Geometric Solvers, *Ji Zhao, Banglei Guan, Zibin Liu, Laurent Kneip*
- 84 GeoDepth: From Point-to-Depth to Plane-to-Depth Modeling for Self-Supervised Monocular Depth Estimation, *Haifeng Wu, Shuhang Gu, Lixin Duan, Wen Li*
- 85 R-SCoRe: Revisiting Scene Coordinate Regression for Robust Large-Scale Visual Localization, *Xudong Jiang, Fangjinhua Wang, Silvano Galliani, Christoph Vogel, Marc Pollefeys*
- 86 HyperPose: Hypernetwork-Infused Camera Pose Localization and an Extended Cambridge Landmarks Dataset, *Ron Ferens, Yosi Keller*
- 87 Learning to Filter Outlier Edges in Global SfM, *Nicole Dambon, Marc Pollefeys, Daniel Barath*
- 88 Stereo4D: Learning How Things Move in 3D from Internet Stereo
* Videos, *Linyi Jin, Richard Tucker, Zhengqi Li, David Fouhey, Noah Snavely, Aleksander Holynski*
- 89 Towards Optimizing Large-Scale Multi-Graph Matching in Bioimaging, *Max Kahl, Sebastian Stricker, Lisa Hutschenreiter, Florian Bernard, Carsten Rother, Bogdan Savchynskyy*
- 90 Bridging Viewpoint Gaps: Geometric Reasoning Boosts Semantic Correspondence, *Qiyang Qian, Hansheng Chen, Masayoshi Tomizuka, Kurt Keutzer, Qianqian Wang, Chenfeng Xu*
- 91 MV-SSM: Multi-View State Space Modeling for 3D Human Pose Estimation, *Aviral Chharia, Wenbo Gou, Haoye Dong*
- 92 Multi-View Pose-Agnostic Change Localization with Zero Labels, *Chamuditha Jayanga Galappaththige, Jason Lai, Lloyd Windrim, Donald Dansereau, Niko Sunderhauf, Dimity Miller*
- 93 Structure-Aware Correspondence Learning for Relative Pose
* Estimation, *Yihan Chen, Wenfei Yang, Huan Ren, Shifeng Zhang, Tianzhu Zhang, Feng Wu*
- 94 Co-op: Correspondence-based Novel Object Pose Estimation,
* Sungphill Moon, Hyeontae Son, Dongcheol Hur, Sangwook Kim
- 95 Any6D: Model-free 6D Pose Estimation of Novel Object,
Taeyeop Lee, Bowen Wen, Minjun Kang, Gyuree Kang, In So Kweon, Kuk-Jin Yoon
- 96 CRISP: Object Pose and Shape Estimation with Test-Time Adaptation,
* Jingnan Shi, Rajat Talak, Harry Zhang, David Jin, Luca Carlone
- 97 CAP-Net: A Unified Network for 6D Pose and Size Estimation of
* Categorical Articulated Parts from a Single RGB-D Image, *Jingshun Huang, Haitao Lin, Tianyu Wang, Yanwei Fu, Xiangyang Xue, Yi Zhu*
- 98 EchoMatch: Partial-to-Partial Shape Matching via Correspondence Reflection, *Yizheng Xie, Viktoria Ehm, Paul Roetzer, Nafie El Amrani, Maolin Gao, Florian Bernard, Daniel Cremers*
- 99 Conformal Prediction and MLLM aided Uncertainty Quantification in Scene Graph Generation, *Sayak Nag, Udit Ghosh, Calvin-Khang Ta, Sarosij Bose, Jiachen Li, Amit K. Roy-Chowdhury*
- 100 Focusing on Tracks for Online Multi-Object Tracking, *Kyujin Shim, Kangwook Ko, Yujin Yang, Changick Kim*
- 101 GRAE-3DMOT: Geometry Relation-Aware Encoder for Online 3D Multi-Object Tracking, *Hyunseop Kim, Hyo-Jun Lee, Yonguk Lee, Jinu Lee, Hanul Kim, Yeong Jun Koh*
- 102 PointSR: Self-Regularized Point Supervision for Drone-View Object Detection, *Weizhuo Li, Yue Xi, Wenjing Jia, Zehao Zhang, Fei Li, Xiangzeng Liu, Qiguang Miao*
- 103 Multi-Modal Aerial-Ground Cross-View Place Recognition with Neural ODEs, *Sijie Wang, Rui She, Qiyu Kang, Siqi Li, Disheng Li, Tianyu Geng, Shangshu Yu, Wee Peng Tay*
- 104 OffsetOPT: Explicit Surface Reconstruction without Normals, *Huan Lei*
- 105 High-Fidelity Lightweight Mesh Reconstruction from Point Clouds,
* Chen Zhang, Wentao Wang, Ximeng Li, Xinyao Liao, Wanjuan Su, Wenbing Tao
- 106 Parametric Point Cloud Completion for Polygonal Surface Reconstruction, *Zhaiyu Chen, Yuqing Wang, Liangliang Nan, Xiao Xiang Zhu*
- 107 Self-Supervised Large Scale Point Cloud Completion for Archaeological Site Restoration, *Aocheng Li, James R. Zimmer-Dauphinee, Rajesh Kalyanam, Ian Lindsay, Parker VanValkenburgh, Steven Wernke, Daniel Aliaga*
- 108 Dual Focus-Attention Transformer for Robust Point Cloud Registration, *Kexue Fu, Mingzhi Yuan, Changwei Wang, Weiguang Pang, Jing Chi, Manning Wang, Longxiang Gao*
- 109 Generalized Gaussian Entropy Model for Point Cloud Attribute Compression with Dynamic Likelihood Intervals, *Changhao Peng*
- 110 TacoDepth: Towards Efficient Radar-Camera Depth Estimation with One-stage Fusion, *Yiran Wang, Jiaqi Li, Chaoyi Hong, Ruibo Li, Liusheng Sun, Xiao Song, Zhe Wang, Zhiguo Cao, Guosheng Lin*
- 111 SeaLion: Semantic Part-Aware Latent Point Diffusion Models for 3D Generation, *Dekai Zhu, Yan Di, Stefan Gavranovic, Slobodan Ilic*
- 112 Spectral Informed Mamba for Robust Point Cloud Processing, *Ali Bahri, Moslem Yazdanpanah, Mehrdad Noori, Sahar Dastani, Milad Cheraghali, Gustavo Adolfo Vargas Hakim, David Osowiecki, Farzad Beizadeh, Ismail Ben Ayed, Christian Desrosiers*
- 113 Hyperbolic Uncertainty-Aware Few-Shot Incremental Point Cloud Segmentation, *Tanuj Sur, Samrat Mukherjee, Kaizer Rahaman, Subhasis Chaudhuri, Muhammad Haris Khan, Biplab Banerjee*
- 114 CamPoint: Boosting Point Cloud Segmentation with Virtual Camera, *Jianhui Zhang, Yizhi Luo, Zicheng Zhang, Xuecheng Nie, Bonan Li*
- 115 ReRAW: RGB-to-RAW Image Reconstruction via Stratified Sampling for Efficient Object Detection on the Edge, *Radu Berdan, Beril Besbinar, Christoph Reinders, Junji Otsuka, Daisuke Iso*
- 116 ViKIENet: Towards Efficient 3D Object Detection with Virtual Key Instance Enhanced Network, *Zhuochen Yu, Bijie Qiu, Andy W. H. Khong*
- 117 ViiNeuS: Volumetric Initialization for Implicit Neural Surface
* Reconstruction of Urban Scenes with Limited Image Overlap, *Hala Djeghim, Nathan Piasco, Moussab Bennehar, Luis Roldao, Dmitry Tsishkou, Désiré Sidibé*
- 118 D³CTTA: Domain-Dependent Decorrelation for Continual Test-Time Adaption of 3D LiDAR Segmentation, *Jichun Zhao, Haiyong Jiang, Haoxuan Song, Jun Xiao, Dong Gong*

- 119 Spotting the Unexpected (STU): A 3D LiDAR Dataset for Anomaly Segmentation in Autonomous Driving, *Alexey Nekrasov, Malcolm Burdorf, Stewart Worrall, Bastian Leibe, Julie Stephany Berrio Perez*
- 120 Seeing is Not Believing: Adversarial Natural Object Optimization for Hard-Label 3D Scene Attacks, *Daizong Liu, Wei Hu*
- 121 Detection-Friendly Nonuniformity Correction: A Union Framework for Infrared UAV Target Detection, *Houzhong Fang, Xiaolin Wang, Zengyang Li, Lu Wang, Qingshan Li, Yi Chang, Luxin Yan*
- * 122 RCP-Bench: Benchmarking Robustness for Collaborative Perception Under Diverse Corruptions, *Shihang Du, Sanqing Qu, Tianhang Wang, Xudong Zhang, Yunwei Zhu, Jian Mao, Fan Lu, Qiao Lin, Guang Chen*
- 123 Generative Map Priors for Collaborative BEV Semantic Segmentation, *Jiahui Fu, Yue Gong, Luting Wang, Shifeng Zhang, Xu Zhou, Si Liu*
- 124 SGFormer: Satellite-Ground Fusion for 3D Semantic Scene Completion, *Xiyue Guo, Jiarui Hu, Junjie Hu, Hujun Bao, Guofeng Zhang*
- 125 Three Cars Approaching within 100m! Enhancing Distant Geometry by Tri-Axis Voxel Scanning for Camera-based Semantic Scene Completion, *Jongseong Bae, Junwoo Ha, Ha Young Kim*
- 126 OccMamba: Semantic Occupancy Prediction with State Space Models, *Heng Li, Yuenan Hou, Xiaohan Xing, Yuexin Ma, Xiao Sun, Yanyong Zhang*
- 127 GaussTR: Foundation Model-Aligned Gaussian Transformer for Self-Supervised 3D Spatial Understanding, *Haoyi Jiang, Liu Liu, Tianheng Cheng, Xinjie Wang, Tianwei Lin, Zhizhong Su, Wenyu Liu, Xinggang Wang*
- 128 UniScene: Unified Occupancy-centric Driving Scene Generation, *Bohan Li, Jiazhe Guo, Hongsi Liu, Yingshuang Zou, Yikang Ding, Xiwu Chen, Hu Zhu, Feiyang Tan, Chi Zhang, Tiancai Wang, Shuchang Zhou, Li Zhang, Xiaojuan Qi, Hao Zhao, Mu Yang, Wenjun Zeng, Xin Jin*
- 129 SplatAD: Real-Time Lidar and Camera Rendering with 3D Gaussian Splatting for Autonomous Driving, *Georg Hess, Carl Lindström, Maryam Fatemi, Christoffer Petersson, Lennart Svensson*
- 130 SimLingo: Vision-Only Closed-Loop Autonomous Driving with Language-Action Alignment, *Katrin Renz, Long Chen, Elahe Arani, Oleg Sinavski*
- * 131 FreeSim: Toward Free-viewpoint Camera Simulation in Driving Scenes, *Lue Fan, Hao Zhang, Qitai Wang, Hongsheng Li, Zhaoxiang Zhang*
- 132 DriveDreamer4D: World Models Are Effective Data Machines for 4D Driving Scene Representation, *Guosheng Zhao, Chaojun Ni, Xiaofeng Wang, Zheng Zhu, Xueyang Zhang, Yida Wang, Guan Huang, Xinze Chen, Boyuan Wang, Youyi Zhang, Wenjun Mei, Xingang Wang*
- 133 Transfer Your Perspective: Controllable 3D Generation from Any Viewpoint in a Driving Scene, *Tai-Yu Pan, Sooyoung Jeon, Mengdi Fan, Jinsu Yoo, Zhenyang Feng, Mark Campbell, Kilian Q. Weinberger, Bharath Hariharan, Wei-Lun Chao*
- * 134 DiffusionDrive: Truncated Diffusion Model for End-to-End Autonomous Driving, *Bencheng Liao, Shaoyu Chen, Haoran Yin, Bo Jiang, Cheng Wang, Sixu Yan, Xinbang Zhang, Xiangyu Li, Ying Zhang, Qian Zhang, Xinggang Wang*
- 135 TraF-Align: Trajectory-aware Feature Alignment for Asynchronous Multi-agent Perception, *Zhiying Song, Lei Yang, Fuxi Wen, Jun Li*
- 136 Trajectory Mamba: Efficient Attention-Mamba Forecasting Model Based on Selective SSM, *Yizhou Huang, Yihua Cheng, Kezhi Wang*
- 137 SOLVE: Synergy of Language-Vision and End-to-End Networks for Autonomous Driving, *Xuesong Chen, Linjiang Huang, Tao Ma, Rongyao Fang, Shaoshuai Shi, Hongsheng Li*
- 138 Towards Long-Horizon Vision-Language Navigation: Platform, Benchmark and Method, *Xinshuai Song, Weixing Chen, Yang Liu, Weikai Chen, Guanbin Li, Liang Lin*
- * 139 MPDrive: Improving Spatial Understanding with Marker-Based Prompt Learning for Autonomous Driving, *Zhiyuan Zhang, Xiaofan Li, Zhihao Xu, Wenjie Peng, Zijian Zhou, Miaoqing Shi, Shuangping Huang*
- 140 Prior Does Matter: Visual Navigation via Denoising Diffusion Bridge Models, *Hao Ren, Yiming Zeng, Zetong Bi, Zhaoliang Wan, Junlong Huang, Hui Cheng*
- 141 Reasoning in Visual Navigation of End-to-end Trained Agents: A Dynamical Systems Approach, *Steeven Janny, Hervé Poirier, Leonid Antsfeld, Guillaume Bono, Gianluca Monaci, Boris Chidlovskii, Francesco Giuliani, Alessio Del Bue, Christian Wolf*
- ☆ 142 ROCKET-1: Mastering Open-World Interaction with Visual-Temporal Context Prompting, *Shaofei Cai, Zihao Wang, Kewei Lian, Zhancun Mu, Xiaojian Ma, Anji Liu, Yitao Liang*
- 143 IAAO: Interactive Affordance Learning for Articulated Objects in 3D Environments, *Can Zhang, Gim Hee Lee*
- 144 A Data-Centric Revisit of Pre-Trained Vision Models for Robot Learning, *Xin Wen, Bingchen Zhao, Yilun Chen, Jiangmiao Pang, Xiaojuan Qi*
- 145 Robotic Visual Instruction, *Yanbang Li, Ziyang Gong, Haoyang Li, Xiaoqi Huang, Haolan Kang, Guangping Bai, Xianzheng Ma*
- 146 DynScene: Scalable Generation of Dynamic Robotic Manipulation Scenes for Embodied AI, *Sangmin Lee, Sungyong Park, Heewon Kim*
- 147 FlowRAM: Grounding Flow Matching Policy with Region-Aware Mamba Framework for Robotic Manipulation, *Sen Wang, Le Wang, Sanping Zhou, Jingyi Tian, Jiayi Li, Haowen Sun, Wei Tang*
- 148 GENMANIP: LLM-driven Simulation for Generalizable Instruction-Following Manipulation, *Ning Gao, Yilun Chen, Shuai Yang, Xinyi Chen, Yang Tian, Hao Li, Haifeng Huang, Hanqing Wang, Tai Wang, Jiangmiao Pang*
- 149 UniGraspTransformer: Simplified Policy Distillation for Scalable Dexterous Robotic Grasping, *Wenbo Wang, Fangyun Wei, Lei Zhou, Xi Chen, Lin Luo, Xiaohan Yi, Yizhong Zhang, Yaobo Liang, Chang Xu, Yan Lu, Jiaolong Yang, Baining Guo*
- 150 ManiVideo: Generating Hand-Object Manipulation Video with Dexterous and Generalizable Grasping, *Youxin Pang, Ruizhi Shao, Jiajun Zhang, Hanzhang Tu, Yun Liu, Boyao Zhou, Hongwen Zhang, Yebin Liu*
- * 151 Hand-held Object Reconstruction from RGB Video with Dynamic Interaction, *Shijian Jiang, Qi Ye, Rengan Xie, Yuchi Huo, Jiming Chen*
- 152 UniHOPE: A Unified Approach for Hand-Only and Hand-Object Pose Estimation, *Yinqiao Wang, Hao Xu, Pheng-Ann Heng, Chi-Wing Fu*
- 153 WiLoR: End-to-end 3D Hand Localization and Reconstruction in-the-wild, *Rolandos Alexandros Potamias, Jinglei Zhang, Jiankang Deng, Stefanos Zafeiriou*
- 154 Analyzing the Synthetic-to-Real Domain Gap in 3D Hand Pose Estimation, *Zhuoran Zhao, Linlin Yang, Pengzhan Sun, Pan Hui, Angela Yao*
- 155 InterMimic: Towards Universal Whole-Body Control for Physics-Based Human-Object Interactions, *Sirui Xu, Hung Yu Ling, Yu-Xiong Wang, Liang-Yan Gui*
- * 156 PoseBH: Prototypical Multi-Dataset Training Beyond Human Pose Estimation, *Uyoung Jeong, Jonathan Freer, Seungryul Baek, Hyung Jin Chang, Kwang In Kim*
- 157 M3GYM: A Large-Scale Multimodal Multi-view Multi-person Pose Dataset for Fitness Activity Understanding in Real-world Settings, *Qingzheng Xu, Ru Cao, Xin Shen, Heming Du, Sen Wang, Xin Yu*
- 158 Certified Human Trajectory Prediction, *Mohammadhossein Bahari, Saeed Saadatnejad, Amirhossein Askari Farsangi, Seyed-Mohsen Moosavi-Dezfooli, Alexandre Alahi*
- 159 ClimbingCap: Multi-Modal Dataset and Method for Rock Climbing in World Coordinate, *Ming Yan, Xincheng Lin, Yuhua Luo, Shuqi Fan, Yudi Dai, Qixin Zhong, Lincui Zhong, Yuexin Ma, Lan Xu, Chenglu Wen, Siqi Shen, Cheng Wang*
- 160 Physical Plausibility-aware Trajectory Prediction via Locomotion Embodiment, *Hiromu Taketsugu, Takeru Oba, Takahiro Maeda, Shohei Nobuhara, Norimichi Ukita*
- 161 Vision-Guided Action: Enhancing 3D Human Motion Prediction with Gaze-informed Affordance in 3D Scenes, *Ting Yu, Yi Lin, Jun Yu, Zhenyu Lou, Qiongjie Cui*
- 162 On Denoising Walking Videos for Gait Recognition, *Dongyang Jin, Chao Fan, Jingzhe Ma, Jingkai Zhou, Weihua Chen, Shiqi Yu*
- 163 ChainHOI: Joint-based Kinematic Chain Modeling for Human-Object Interaction Generation, *Ling-An Zeng, Guohong Huang, Yi-Lin Wei, Shengbo Gu, Yu-Ming Tang, Jingke Meng, Wei-Shi Zheng*

- 164 StickMotion: Generating 3D Human Motions by Drawing a Stickman, *Tao Wang, Zhihua Wu, Qiaozhi He, Jiaming Chu, Ling Qian, Yu Cheng, Junliang Xing, Jian Zhao, Lei Jin*
- 165 MixerMDM: Learnable Composition of Human Motion Diffusion Models, *Pablo Ruiz-Ponce, German Barquero, Cristina Palmero, Sergio Escalera, José García-Rodríguez*
- 166 HumanDreamer: Generating Controllable Human-Motion Videos via Decoupled Generation, *Boyuan Wang, Xiaofeng Wang, Chaojun Ni, Guosheng Zhao, Zhiqin Yang, Zheng Zhu, Muiyang Zhang, Yukun Zhou, Xinze Chen, Guan Huang, Lihong Liu, Xingang Wang*
- 167 Poly-Autoregressive Prediction for Modeling Interactions, *Neeraj Thakkar, Tara Sadjadpour, Jathushan Rajasegeran, Shiry Ginosar, Jitendra Malik*
- 168 Adapting Pre-trained 3D Models for Point Cloud Video Understanding via Cross-frame Spatio-temporal Perception, *Baixuan Lv, Yaohua Zha, Tao Dai, Xue Yuerong, Ke Chen, Shu-Tao Xia*
- 169 Recovering Dynamic 3D Sketches from Videos, *Jaeah Lee, Changwoon Choi, Young Min Kim, Jaesik Park*
- 170 FreeGave: 3D Physics Learning from Dynamic Videos by Gaussian Velocity, *Jinxi Li, Ziyang Song, Siyuan Zhou, Bo Yang*
- 171 Dynamic Camera Poses and Where to Find Them, *Chris Rockwell, Joseph Tung, Tsung-Yi Lin, Ming-Yu Liu, David F. Fouhey, Chen-Hsuan Lin*
- 172 Repurposing Pre-trained Video Diffusion Models for Event-based
☆ Video Interpolation, *Jingxi Chen, Brandon Y. Feng, Haoming Cai, Tianfu Wang, Levi Burner, Dehao Yuan, Cornelia Fermüller, Christopher A. Metzler, Yiannis Aloimonos*
- 173 InterDyn: Controllable Interactive Dynamics with Video Diffusion Models, *Rick Akkerman, Haiwen Feng, Michael J. Black, Dimitrios Tzionas, Victoria Fernández Abrevaya*
- 174 DreamCache: Finetuning-Free Lightweight Personalized Image Generation via Feature Caching, *Emanuele Aiello, Umberto Michieli, Diego Valsesia, Mete Ozay, Enrico Magli*
- 175 LeviTor: 3D Trajectory Oriented Image-to-Video Synthesis,
* *Hanlin Wang, Hao Ouyang, Qiuyu Wang, Wen Wang, Ka Leong Cheng, Qifeng Chen, Yujun Shen, Limin Wang*
- 176 UniReal: Universal Image Generation and Editing via Learning
* *Real-world Dynamics, Xi Chen, Zhifei Zhang, He Zhang, Yuqian Zhou, Soo Ye Kim, Qing Liu, Yijun Li, Jianming Zhang, Nanxuan Zhao, Yilin Wang, Hui Ding, Zhe Lin, Hengshuang Zhao*
- 177 Extrapolating and Decoupling Image-to-Video Generation Models:
* *Motion Modeling is Easier Than You Think, Jie Tian, Xiaoye Qu, Zhenyi Lu, Wei Wei, Sichen Liu, Yu Cheng*
- 178 Generative Omnimate: Learning to Decompose Video into Layers,
* *Yao-Chih Lee, Erika Lu, Sarah Rumbley, Michal Geyer, Jia-Bin Huang, Tali Dekel, Forrester Cole*
- 179 RL-RC-DoT: A Block-level RL agent for Task-Aware Video Compression, *Uri Gadot, Assaf Shocher, Shie Mannor, Gal Chechik, Assaf Hallak*
- 180 Towards Practical Real-Time Neural Video Compression, *Zhaoyang Jia, Bin Li, Jiahao Li, Wenxuan Xie, Linfeng Qi, Houqiang Li, Yan Lu*
- 181 Neural Video Compression with Context Modulation, *Chuanbo Tang, Zhuoyuan Li, Yifan Bian, Li Li, Dong Liu*
- 182 Event-based Video Super-Resolution via State Space Models, *Zeyu Xiao, Xinchao Wang*
- 183 STDD: Spatio-Temporal Dual Diffusion for Video Generation, *Shuaizhen Yao, Xiaoya Zhang, Xin Liu, Mengyi Liu, Zhen Cui*
- 184 IceDiff: High Resolution and High-Quality Arctic Sea Ice Forecasting with Generative Diffusion Prior, *Jingyi Xu, Siwei Tu, Weidong Yang, Ben Fei, Shuhao Li, Keyi Liu, Yeqi Luo, Lipeng Ma, Lei Bai*
- 185 OSV: One Step is Enough for High-Quality Image to Video Generation,
☆ *Xiaofeng Mao, Zhengkai Jiang, Fu-yun Wang, Jiangning Zhang, Hao Chen, Mingmin Chi, Yabiao Wang, Wenhan Luo*
- 186 I2VGuard: Safeguarding Images against Misuse in Diffusion-based Image-to-Video Models, *Dongnan Gui, Xun Guo, Wengang Zhou, Yan Lu*
- 187 CASP: Consistency-aware Audio-induced Saliency Prediction Model for Omnidirectional Video, *Zhaolin Wan, Han Qin, Zhiyang Li, Xiaopeng Fan, Wangmeng Zuo, Debin Zhao*
- 188 Generalizing Deepfake Video Detection with Plug-and-Play: Video-Level Blending and Spatiotemporal Adapter Tuning, *Zhiyuan Yan, Yandan Zhao, Shen Chen, Mingyi Guo, Xinghe Fu, Taiping Yao, Shouhong Ding, Yunsheng Wu, Li Yuan*
- 189 OSDFace: One-Step Diffusion Model for Face Restoration, *Jingkai Wang, Jue Gong, Lin Zhang, Zheng Chen, Xing Liu, Hong Gu, Yutong Liu, Yulun Zhang, Xiaokang Yang*
- 190 MFogHub: Bridging Multi-Regional and Multi-Satellite Data for Global Marine Fog Detection and Forecasting, *Mengqiu Xu, Kaixin Chen, Heng Guo, Yixiang Huang, Ming Wu, Zhenwei Shi, Chuang Zhang, Jun Guo*
- 191 Feature Spectrum Learning for Remote Sensing Change Detection, *Qi Zang, Dong Zhao, Shuang Wang, Dou Quan, Zhun Zhong*
- 192 Dual-Granularity Semantic Guided Sparse Routing Diffusion Model for General Pansharpening, *Yinghui Xing, Litao Qu, Shizhou Zhang, Di Xu, Yingkun Yang, Yanning Zhang*
- 193 Hyperspectral Pansharpening via Diffusion Models with Iteratively Zero-Shot Guidance, *Jin-Liang Xiao, Ting-Zhu Huang, Liang-Jian Deng, Guang Lin, Zihan Cao, Chao Li, Qibin Zhao*
- 194 Complementary Advantages: Exploiting Cross-Field Frequency Correlation for NIR-Assisted Image Denoising, *Yuchen Wang, Hongyuan Wang, Lizhi Wang, Xin Wang, Lin Zhu, Wanxuan Lu, Hua Huang*
- 195 Hazy Low-Quality Satellite Video Restoration Via Learning Optimal Joint Degradation Patterns and Continuous-Scale Super-Resolution Reconstruction, *Ning Ni, Libao Zhang*
- 196 Iterative Predictor-Critic Code Decoding for Real-World Image Dehazing, *Jiayi Fu, Siyu Liu, Zikun Liu, Chun-Le Guo, Hyunhee Park, Ruiqi Wu, Guoqing Wang, Chongyi Li*
- 197 Efficient Visual State Space Model for Image Deblurring, *Lingshun Kong, Jiangxin Dong, Jinhui Tang, Ming-Hsuan Yang, Jinshan Pan*
- 198 Rotation-Equivariant Self-Supervised Method in Image Denoising, *Hanze Liu, Jiahong Fu, Qi Xie, Deyu Meng*
- 199 A Universal Scale-Adaptive Deformable Transformer for Image Restoration across Diverse Artifacts, *Xuyi He, Yuhui Quan, Ruotao Xu, Hui Ji*
- 200 Toward Generalized Image Quality Assessment: Relaxing the Perfect Reference Quality Assumption, *Du Chen, Tianhe Wu, Kede Ma, Lei Zhang*
- 201 Complexity Experts are Task-Discriminative Learners for Any Image Restoration, *Eduard Zamfir, Zongwei Wu, Nancy Mehta, Yuedong Tan, Danda Pani Paudel, Yulun Zhang, Radu Timofte*
- 202 Visual-Instructed Degradation Diffusion for All-in-One Image Restoration, *Wenyang Luo, Haina Qin, Zewen Chen, Libin Wang, Dandan Zheng, Yuming Li, Yufan Liu, Bing Li, Weiming Hu*
- 203 PassionSR: Post-Training Quantization with Adaptive Scale in One-Step Diffusion based Image Super-Resolution, *Libo Zhu, Jianze Li, Haotong Qin, Wenbo Li, Yulun Zhang, Yong Guo, Xiaokang Yang*
- 204 Edge-SD-SR: Low Latency and Parameter Efficient On-device Super-Resolution with Stable Diffusion via Bidirectional Conditioning, *Isma Hadji, Mehdi Noroozi, Victor Escorcia, Anastis Zaganidis, Brais Martinez, Georgios Tzimiropoulos*
- 205 HUNet: Homotopy Unfolding Network for Image Compressive Sensing, *Feiyang Shen, Hongping Gan*
- 206 Dual Prompting Image Restoration with Diffusion Transformers, *Dehong Kong, Fan Li, Zhixin Wang, Jiaqi Xu, Renjing Pei, Wenbo Li, Wenqi Ren*
- 207 Frequency-Biased Synergistic Design for Image Compression and Compensation, *Jiaming Liu, Qi Zheng, Zihao Liu, Yilian Zhong, Peiye Liu, Tao Liu, Shusong Xu, Yanheng Lu, Sicheng Li, Dimin Niu, Yibo Fan*
- 208 FIRE: Robust Detection of Diffusion-Generated Images via Frequency-Guided Reconstruction Error, *Beilin Chu, Xuan Xu, Xin Wang, Yufei Zhang, Weiye You, Linna Zhou*
- 209 Robust Message Embedding via Attention Flow-Based Steganography, *Huayuan Ye, Shenzhuo Zhang, Shiqi Jiang, Jing Liao, Shuhang Gu, Dejun Zheng, Changbo Wang, Chenhui Li*
- 210 Learned Image Compression with Dictionary-based Entropy Model, *Jingbo Lu, Leheng Zhang, Xingyu Zhou, Mu Li, Wen Li, Shuhang Gu*

- 211 D²iT: Dynamic Diffusion Transformer for Accurate Image Generation, *Weinan Jia, Mengqi Huang, Nan Chen, Lei Zhang, Zhendong Mao*
- 212 Classifier-Free Guidance Inside the Attraction Basin May Cause Memorization, *Anubhav Jain, Yuya Kobayashi, Takashi Shibuya, Yuhta Takida, Nasir Memon, Julian Togelius, Yuki Mitsufuji*
- 213 Not All Parameters Matter: Masking Diffusion Models for Enhancing Generation Ability, *Lei Wang, Senmao Li, Fei Yang, Jianye Wang, Ziheng Zhang, Yuhang Liu, Yaxing Wang, Jian Yang*
- 214 BlockDance: Reuse Structurally Similar Spatio-Temporal Features to Accelerate Diffusion Transformers, *Hui Zhang, Tingwei Gao, Jie Shao, Zuxuan Wu*
- 215 Diffusion Model is Effectively Its Own Teacher, *Xinyin Ma, Runpeng Yu, Songhua Liu, Gongfan Fang, Xinchao Wang*
- 216 Reward Fine-Tuning Two-Step Diffusion Models via Learning Differentiable Latent-Space Surrogate Reward, *Zhiwei Jia, Yuesong Nan, Huixi Zhao, Gengdai Liu*
- 217 RaSS: Improving Denoising Diffusion Samplers with Reinforced Active Sampling Scheduler, *Xin Ding, Lei Yu, Xin Li, Zhijun Tu, Hanting Chen, Jie Hu, Zhibo Chen*
- 218 A Closer Look at Time Steps is Worthy of Triple Speed-Up for Diffusion Model Training, *Kai Wang, Mingjia Shi, Yukun Zhou, Zekai Li, Zhihang Yuan, Yuzhang Shang, Xiaojiang Peng, Hanwang Zhang, Yang You*
- 219 Scaling Properties of Diffusion Models For Perceptual Tasks, *Rahul Ravishankar, Zeeshan Patel, Jathushan Rajasegaran, Jitendra Malik*
- 220 Parallelized Autoregressive Visual Generation, *Yuqing Wang, * Shuhuai Ren, Zhijie Lin, Yujin Han, Haoyuan Guo, Zhenheng Yang, ☆ Difan Zou, Jiashi Feng, Xihui Liu*
- 221 Janus: Decoupling Visual Encoding for Unified Multimodal Understanding and Generation, *Chengyue Wu, Xiaokang Chen, Zhiyu Wu, Yiyang Ma, Xingchao Liu, Zizheng Pan, Wen Liu, Zhenda Xie, Xingkai Yu, Chong Ruan, Ping Luo*
- 222 Identity-Preserving Text-to-Video Generation by Frequency Decomposition, ** Shenghai Yuan, Jinfa Huang, Xianyi He, Yongyang Ge, Yujun Shi, Liuhan Chen, Jiebo Luo, Li Yuan*
- 223 BlobGEN-Vid: Compositional Text-to-Video Generation with Blob Video Representations, *Weixi Feng, Chao Liu, Sifei Liu, William Yang Wang, Arash Vahdat, Weili Nie*
- 224 ByTheWay: Boost Your Text-to-Video Generation Model to Higher Quality in a Training-free Way, *Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Dahua Lin, Jiaqi Wang*
- 225 Keyframe-Guided Creative Video Inpainting, *Yuwei Guo, Ceyuan Yang, Anyi Rao, Chenlin Meng, Omer Bar-Tal, Shuangrui Ding, Maneesh Agrawala, Dahua Lin, Bo Dai*
- 226 SemanticDraw: Towards Real-Time Interactive Content Creation from Image Diffusion Models, *Jaerin Lee, Daniel Sungho Jung, Kanggeon Lee, Kyoung Mu Lee*
- 227 TKG-DM: Training-free Chroma Key Content Generation Diffusion Model, ** Ryugo Morita, Stanislav Frolov, Brian Bernhard Moser, Takahiro Shirakawa, Ko Watanabe, Andreas Dengel, Jinjia Zhou*
- 228 K-LoRA: Unlocking Training-Free Fusion of Any Subject and Style LoRAs, *Ziheng Ouyang, Zhen Li, Qibin Hou*
- 229 SCSA: A Plug-and-Play Semantic Continuous-Sparse Attention for Arbitrary Semantic Style Transfer, ** Chunnan Shang, Zhizhong Wang, Hongwei Wang, Xiangming Meng*
- 230 MARBLE: Material Recomposition and Blending in CLIP-Space, *Ta Ying Cheng, Prafull Sharma, Mark Boss, Varun Jampani*
- 231 MagicQuill: An Intelligent Interactive Image Editing System, *Zichen Liu, Yue Yu, Hao Ouyang, Qiuyu Wang, Ka Leong Cheng, Wen Wang, Zhiheng Liu, Qifeng Chen, Yujun Shen*
- 232 FluxSpace: Disentangled Semantic Editing in Rectified Flow Models, *☆ Yusuf Dalva, Kavana Venkatesh, Pinar Yanardag*
- 233 FireEdit: Fine-grained Instruction-based Image Editing via Region-aware Vision Language Model, *☆ Jun Zhou, Jiahao Li, Zunnan Xu, Hanhui Li, Yiji Cheng, Fa-Ting Hong, Qin Lin, Qinglin Lu, Xiaodan Liang*
- 234 Recognition-Synergistic Scene Text Editing, *Zhengyao Fang, Pengyuan Lyu, Jingjing Wu, Chengquan Zhang, Jun Yu, Guangming Lu, Wenjie Pei*
- 235 HyperLoRA: Parameter-Efficient Adaptive Generation for Portrait Synthesis, ** Mengtian Li, Jinshu Chen, Wanquan Feng, Bingchuan Li, Fei Dai, Songtao Zhao, Qian He*
- 236 Self-Evolving Visual Concept Library using Vision-Language Critics, *Atharva Sehgal, Patrick Yuan, Ziniu Hu, Yisong Yue, Jennifer J. Sun, Swarat Chaudhuri*
- 237 Training-free Dense-Aligned Diffusion Guidance for Modular Conditional Image Synthesis, *Zixuan Wang, Duo Peng, Feng Chen, Yuwei Yang, Yinjie Lei*
- 238 Movie Weaver: Tuning-Free Multi-Concept Video Personalization with Anchored Prompts, *Feng Liang, Haoyu Ma, Zecheng He, Tingbo Hou, Ji Hou, Kunpeng Li, Xiaoliang Dai, Felix Juefei-Xu, Samaneh Azadi, Animesh Sinha, Peizhao Zhang, Peter Vajda, Diana Marculescu*
- 239 AMO Sampler: Enhancing Text Rendering with Overshooting, *Xixi Hu, Keyang Xu, Bo Liu, Qiang Liu, Hongliang Fei*
- 240 ArtiFade: Learning to Generate High-quality Subject from Blemished Images, *Shuya Yang, Shaozhe Hao, Yukang Cao, Kwan-Yee K. Wong*
- 241 OmniFlow: Any-to-Any Generation with Multi-Modal Rectified Flows, *Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Zichun Liao, Yusuke Kato, Kazuki Kozuka, Aditya Grover*
- 242 LoRACL: Contrastive Adaptation for Customization of Diffusion Models, *☆ Enis Simsar, Thomas Hofmann, Federico Tombari, Pinar Yanardag*
- 243 Aesthetic Post-Training Diffusion Models from Generic Preferences with Step-by-step Preference Optimization, *☆ Zhanhao Liang, Yuhui Yuan, Shuyang Gu, Bohan Chen, Tiankai Hang, Mingxi Cheng, Ji Li, Liang Zheng*
- 244 Composing Parts for Expressive Object Generation, *Harsh Rangwani, Aishwarya Agarwal, Kuldeep Kulkarni, R. Venkatesh Babu, Srikrishna Karanam*
- 245 DyMO: Training-Free Diffusion Model Alignment with Dynamic Multi-Objective Scheduling, *Xin Xie, Dong Gong*
- 246 Continuous, Subject-Specific Attribute Control in T2I Models by Identifying Semantic Directions, *Stefan Andreas Baumann, Felix Krause, Michael Neumayr, Nick Stracke, Melvin Sevi, Vincent Tao Hu, Björn Ommer*
- 247 Make It Count: Text-to-Image Generation with an Accurate Number of Objects, *Lital Binyamin, Yoad Tewel, Hilit Segev, Eran Hirsch, Royi Rassin, Gal Chechik*
- 248 Detect-and-Guide: Self-regulation of Diffusion Models for Safe Text-to-Image Generation via Guideline Token Optimization, *Feifei Li, Mi Zhang, Yiming Sun, Min Yang*
- 249 MCCD: Multi-Agent Collaboration-based Compositional Diffusion for Complex Text-to-Image Generation, *Mingcheng Li, Xiaolu Hou, Ziyang Liu, Dingkan Yang, Ziyun Qian, Jiawei Chen, Jinjie Wei, Yue Jiang, Qingyao Xu, Lihua Zhang*
- 250 StoryGPT-V: Large Language Models as Consistent Story Visualizers, *Xiaoqian Shen, Mohamed Elhoseiny*
- 251 ChatGen: Automatic Text-to-Image Generation From FreeStyle Chatting, *Chengyou Jia, Changliang Xia, Zhuohang Dang, Weijia Wu, Hangwei Qian, Minnan Luo*
- 252 OmniGen: Unified Image Generation, *Shitao Xiao, Yueze Wang, Junjie Zhou, Huaying Yuan, Xingrun Xing, Ruihan Yan, Chaofan Li, Shuting Wang, Tiejun Huang, Zheng Liu*
- 253 ShapeWords: Guiding Text-to-Image Synthesis with 3D Shape-Aware Prompts, *Dmitry Petrov, Pradyumn Goyal, Divyansh Shivashok, Yuanming Tao, Melinos Averkiou, Evangelos Kalogerakis*
- 254 From Words to Structured Visuals: A Benchmark and Framework for Text-to-Diagram Generation and Editing, ** Jingxuan Wei, Cheng Tan, Qi Chen, Gaowei Wu, Siyuan Li, Zhangyang Gao, Linzhuang Sun, Bihui Yu, Ruifeng Guo*
- 255 Eval3D: Interpretable and Fine-grained Evaluation for 3D Generation, *Shivam Duggal, Yushi Hu, Oscar Michel, Aniruddha Kembhavi, William T. Freeman, Noah A. Smith, Ranjay Krishna, Antonio Torralba, Ali Farhadi, Wei-Chiu Ma*

- 256 **EEE-Bench: A Comprehensive Multimodal Electrical And Electronics Engineering Benchmark**, *Ming Li, Jike Zhong, Tianle Chen, Yuxiang Lai, Konstantinos Psounis*
- 257 **Towards Precise Embodied Dialogue Localization via Causality Guided Diffusion**, *Haoyu Wang, Le Wang, Sanping Zhou, Jingyi Tian, Zheng Qin, Yabing Wang, Gang Hua, Wei Tang*
- 258 **Rethinking Training for De-biasing Text-to-Image Generation: Unlocking the Potential of Stable Diffusion**, *Eunji Kim, Siwon Kim, Minjun Park, Rahim Entezari, Sungroh Yoon*
- 259 **Rectified Diffusion Guidance for Conditional Generation**, *Mengfei Xia, Nan Xue, Yujun Shen, Ran Yi, Tieliang Gong, Yong-Jin Liu*
- 260 **T2ISafety: Benchmark for Assessing Fairness, Toxicity, and Privacy in Image Generation**, *Lijun Li, Zhelun Shi, Xuhao Hu, Bowen Dong, Yiran Qin, Xihui Liu, Lu Sheng, Jing Shao*
- 261 **The Illusion of Unlearning: The Unstable Nature of Machine Unlearning in Text-to-Image Diffusion Models**, *Naveen George, Karthik Nandan Dasaraju, Rutheesh Reddy Chittepuri, Konda Reddy Mopuri*
- 262 **Towards Universal Dataset Distillation via Task-Driven Diffusion**, *Ding Qi, Jian Li, Junyao Gao, Shuguang Dou, Ying Tai, Jianlong Hu, Bo Zhao, Yabiao Wang, Chengjie Wang, Cairong Zhao*
- 263 **RealEdit: Reddit Edits As a Large-scale Empirical Dataset for Image Transformations**, *Peter Sushko, Ayana Bharadwaj, Zhi Yang Lim, Vasily Ilin, Ben Caffee, Dongping Chen, Mohammadreza Salehi, Cheng-Yu Hsieh, Ranjay Krishna*
- 264 **Harnessing Global-Local Collaborative Adversarial Perturbation for Anti-Customization**, *Long Xu, Jiakai Wang, Haojie Hao, Haotong Qin, Jiejie Zhao, Xianglong Liu*
- 265 **Decoder Gradient Shield: Provable and High-Fidelity Prevention of Gradient-Based Box-Free Watermark Removal**, *Haonan An, Guang Hua, Zhengru Fang, Guowen Xu, Susanto Rahardja, Yuguang Fang*
- 266 **Silence is Golden: Leveraging Adversarial Examples to Nullify Audio Control in LDM-based Talking-Head Generation**, *Yuan Gan, Jiaxu Miao, Yunze Wang, Yi Yang*
- 267 **Secret Lies in Color: Enhancing AI-Generated Images Detection with Color Distribution Analysis**, *Zexi Jia, Chuanwei Huang, Yeshuang Zhu, Hongyan Fei, Xiaoyue Duan, Zhiqiang Yuan, Ying Deng, Jiawei Zhang, Jinchao Zhang, Jie Zhou*
- 268 **CO-SPY: Combining Semantic and Pixel Features to Detect Synthetic Images by AI**, *Siyan Cheng, Lingjuan Lyu, Zhenting Wang, Xiangyu Zhang, Vikash Sehwal*
- 269 **FirePlace: Geometric Refinements of LLM Common Sense Reasoning for 3D Object Placement**, *Ian Huang, Yanan Bao, Karen Truong, Howard Zhou, Cordelia Schmid, Leonidas Guibas, Alireza Fathi*
- 270 **VI³NR: Variance Informed Initialization for Implicit Neural Representations**, *Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, Sameera Ramasinghe, Stephen Gould*
- 271 **EigenGS Representation: From Eigenspace to Gaussian Image Space**, *Lo-Wei Tai, Ching-En Li, Cheng-Lin Chen, Chih-Jung Tsai, Hwann-Tzong Chen, Tyng-Luh Liu*
- 272 **Few-shot Personalized Scanpath Prediction**, *Ruoyu Xue, Jingyi Xu, Sounak Mondal, Hieu Le, Greg Zelinsky, Minh Hoai, Dimitris Samaras*
- 273 **Enhancing 3D Gaze Estimation in the Wild using Weak Supervision with Gaze Following Labels**, *Pierre Vuillecard, Jean-Marc Odobez*
- 274 **FilmComposer: LLM-Driven Music Production for Silent Film Clips**, *Zhifeng Xie, Qile He, Youjia Zhu, Qiwei He, Mengtian Li*
- 275 **VinTAGe: Joint Video and Text Conditioning for Holistic Audio Generation**, *Saksham Singh Kushwaha, Yapeng Tian*
- 276 **Seeing Speech and Sound: Distinguishing and Locating Audio Sources in Visual Scenes**, *Hyeonggon Ryu, Seongyu Kim, Joon Son Chung, Arda Senocak*
- 277 **Audio-Visual Instance Segmentation**, *Ruohao Guo, Xianghua Ying, Yaru Chen, Dantong Niu, Guangyao Li, Liao Qu, Yanyu Qi, Jinxing Zhou, Bowei Xing, Wenzhen Yue, Ji Shi, Qixun Wang, Peiliang Zhang, Buwen Liang*
- 278 **UWAV: Uncertainty-weighted Weakly-supervised Audio-Visual Video Parsing**, *Yung-Hsuan Lai, Janek Ebberts, Yu-Chiang Frank Wang, François Germain, Michael Jeffrey Jones, Moitrey Chatterjee*
- 279 **DistinctAD: Distinctive Audio Description Generation in Contexts**, *Bo Fang, Wenhao Wu, Qiangqiang Wu, Yuxin Song, Antoni B. Chan*
- 280 **ExpertAF: Expert Actionable Feedback from Video**, *Kumar Ashutosh, Tushar Nagarajan, Georgios Pavlakos, Kris Kitani, Kristen Grauman*
- 281 **FSBench: A Figure Skating Benchmark for Advancing Artistic Sports Understanding**, *Rong Gao, Xin Liu, Zhuozhao Hu, Bohao Xing, Baiqiang Xia, Zitong Yu, Heikki Kälviäinen*
- 282 **Divot: Diffusion Powers Video Tokenizer for Comprehension and Generation**, *Yuying Ge, Yizhuo Li, Yixiao Ge, Ying Shan*
- 283 **LLaVA-Critic: Learning to Evaluate Multimodal Models**, *Tianyi Xiong, Xiyao Wang, Dong Guo, Qinghao Ye, Haoqi Fan, Quanquan Gu, Heng Huang, Chunyuan Li*
- 284 **Is Your World Simulator a Good Story Presenter? A Consecutive Events-Based Benchmark for Future Long Video Generation**, *Yiping Wang, Xuehai He, Kuan Wang, Luyao Ma, Jianwei Yang, Shuohang Wang, Simon Shaolei Du, Yelong Shen*
- 285 **Progress-Aware Video Frame Captioning**, *Zihui Xue, Jounghbin An, Xitong Yang, Kristen Grauman*
- 286 **Learning from Streaming Video with Orthogonal Gradients**, *Tengda Han, Dilara Gokay, Joseph Heyward, Chuhan Zhang, Daniel Zoran, Viorica Patraucean, Joao Carreira, Dima Damen, Andrew Zisserman*
- 287 **Thinking in Space: How Multimodal Large Language Models See, Remember, and Recall Spaces**, *Jihan Yang, Shusheng Yang, Anjali W. Gupta, Rilyn Han, Li Fei-Fei, Saining Xie*
- 288 **Bootstrap Your Own Views: Masked Ego-Exo Modeling for Fine-grained View-invariant Video Representations**, *Jungin Park, Jiyoung Lee, Kwanghoon Sohn*
- 289 **VEU-Bench: Towards Comprehensive Understanding of Video Editing**, *Bozheng Li, Yongliang Wu, Yi Lu, Jiashuo Yu, Licheng Tang, Jiawang Cao, Wenqing Zhu, Yuyang Sun, Jay Wu, Wenbo Zhu*
- 290 **Question-Aware Gaussian Experts for Audio-Visual Question Answering**, *Hongyeob Kim, Inyoung Jung, Dayoon Suh, Youjia Zhang, Sangmin Lee, Sungeun Hong*
- 291 **MLVU: Benchmarking Multi-task Long Video Understanding**, *Junjie Zhou, Yan Shu, Bo Zhao, Boya Wu, Zhengyang Liang, Shitao Xiao, Minghao Qin, Xi Yang, Yongping Xiong, Bo Zhang, Tiejun Huang, Zheng Liu*
- 292 **M-LLM Based Video Frame Selection for Efficient Video Understanding**, *Kai Hu, Feng Gao, Xiaohan Nie, Peng Zhou, Son Tran, Tal Neiman, Lingyun Wang, Mubarak Shah, Raffay Hamid, Bing Yin, Trishul Chilimbi*
- 293 **On the Consistency of Video Large Language Models in Temporal Comprehension**, *Minjoon Jung, Junbin Xiao, Byoung-Tak Zhang, Angela Yao*
- 294 **VidHalluc: Evaluating Temporal Hallucinations in Multimodal Large Language Models for Video Understanding**, *Chaoyu Li, Eun Woo Im, Pooyan Fazli*
- 295 **ReWind: Understanding Long Videos with Instructed Learnable Memory**, *Anxhelo Diko, Tinghuai Wang, Wassim Swaileh, Shiyun Sun, Ioannis Patras*
- 296 **MASH-VLM: Mitigating Action-Scene Hallucination in Video-LLMs through Disentangled Spatial-Temporal Representations**, *Kyungho Bae, Jinhyung Kim, Sihaeng Lee, Soonyoung Lee, Gunhee Lee, Jinwoo Choi*
- 297 **Number it: Temporal Grounding Videos like Flipping Manga**, *Yongliang Wu, Xinting Hu, Yuyang Sun, Yizhou Zhou, Wenbo Zhu, Fengyun Rao, Bernt Schiele, Xu Yang*
- 298 **Seq2Time: Sequential Knowledge Transfer for Video LLM Temporal Grounding**, *Andong Deng, Zhongpai Gao, Anwesa Choudhuri, Benjamin Planche, Meng Zheng, Bin Wang, Terrence Chen, Chen Chen, Ziyang Wu*
- 299 **STOP: Integrated Spatial-Temporal Dynamic Prompting for Video Understanding**, *Zichen Liu, Kunlun Xu, Bing Su, Xu Zou, Yuxin Peng, Jiahuan Zhou*
- 300 **SyncVP: Joint Diffusion for Synchronous Multi-Modal Video Prediction**, *Enrico Pallotta, Sina Mokhtarzadeh Azar, Shuai Li, Olga Zatsarynna, Juergen Gall*
- 301 **SVLTA: Benchmarking Vision-Language Temporal Alignment via Synthetic Video Situation**, *Hao Du, Bo Wu, Yan Lu, Zhendong Mao*

- 302 DTOS: Dynamic Time Object Sensing with Large Multimodal
☆ Model, *Jirui Tian, Jinrong Zhang, Shenglan Liu, Luhao Xu, Zhixiong Huang, Gao Huang*
- 303 Decoupled Motion Expression Video Segmentation, *Hao Fang, Runmin Cong, Xiankai Lu, Xiaofei Zhou, Sam Kwong, Wei Zhang*
- 304 EdgeTAM: On-Device Track Anything Model, *Chong Zhou, Chenchen Zhu, Yunyang Xiong, Saksham Suri, Fanyu Xiao, Lemeng Wu, Raghuraman Krishnamoorthi, Bo Dai, Chen Change Loy, Vikas Chandra, Bilge Soran*
- 305 Holmes-VAU: Towards Long-term Video Anomaly Understanding at
* Any Granularity, *Huaxin Zhang, Xiaohao Xu, Xiang Wang, Jialong Zuo, Xiaonan Huang, Changxin Gao, Shanjun Zhang, Li Yu, Nong Sang*
- 306 MammAlps: A Multi-view Video Behavior Monitoring Dataset of
* Wild Mammals in the Swiss Alps, *Valentin Gabeff, Haozhe Qi, Brendan Flaherty, Gencer Sumbul, Alexander Mathis, Devis Tuia*
- 307 Boosting Point-Supervised Temporal Action Localization through Integrating Query Reformation and Optimal Transport, *Mengnan Liu, Le Wang, Sanping Zhou, Kun Xia, Xiaolong Sun, Gang Hua*
- 308 Semantic-guided Cross-Modal Prompt Learning for Skeleton-based Zero-shot Action Recognition, *Anqi Zhu, Jingmin Zhu, James Bailey, Mingming Gong, QiuHong Ke*
- 309 Mono3DVL: Monocular-Video-Based 3D Visual Language Tracking, *Hongkai Wei, Yang Yang, Shijie Sun, Mingtao Feng, Xiangyu Song, Qi Lei, Hongli Hu, Rong Wang, Huansheng Song, Naveed Akhtar, Ajmal Saeed Mian*
- 310 F5board: Over 3 Million Characters of ASL Fingerspelling Collected via Smartphones, *Manfred Georg, Garrett Tanzer, Esha Uboweja, Saad Hassan, Maximus Shengelia, Sam Sepah, Sean Forbes, Thad Starner*
- 311 Data-free Universal Adversarial Perturbation with Pseudo-semantic Prior, *Chanhui Lee, Yeonghwan Song, Jeany Son*
- 312 Detecting Adversarial Data Using Perturbation Forgery, *Qian Wang, Chen Li, Yuchen Luo, Hefei Ling, Shijuan Huang, Ruoxi Jia, Ning Yu*
- 313 Stacking Brick by Brick: Aligned Feature Isolation for Incremental Face Forgery Detection, *Jikang Cheng, Zhiyuan Yan, Ying Zhang, Li Hao, Jiaxin Ai, Qin Zou, Chen Li, Zhongyuan Wang*
- 314 SapiensID: Foundation for Human Recognition, *Minchul Kim, Dingqiang Ye, Yiyang Su, Feng Liu, Xiaoming Liu*
- 315 Spiking Transformer with Spatial-Temporal Attention, *Donghyun Lee, Yuhang Li, Youngeun Kim, Shiting Xiao, Priyadarshini Panda*
- 316 STAA-SNN: Spatial-Temporal Attention Aggregator for Spiking Neural Networks, *Tianqing Zhang, Kairong Yu, Xian Zhong, Hongwei Wang, Qi Xu, Qiang Zhang*
- 317 Efficient Event-Based Object Detection: A Hybrid Neural Network with Spatial and Temporal Attention, *Soikat Hasan Ahmed, Jan Finkbeiner, Emre Neftci*
- 318 DIFFER: Disentangling Identity Features via Semantic Cues for Clothes-Changing Person Re-ID, *Xin Liang, Yogesh S Rawat*
- 319 SegEarth-OV: Towards Training-Free Open-Vocabulary
○ Segmentation for Remote Sensing Images, *Kaiyu Li, Ruixun Liu, Xiangyong Cao, Xueru Bai, Feng Zhou, Deyu Meng, Zhi Wang*
- 320 Mixture of Submodules for Domain Adaptive Person Search, *Minsu Kim, Seungryong Kim, Kwanghoon Sohn*
- 321 An Image-like Diffusion Method for Human-Object Interaction Detection, *Xiaofei Hui, Haoxuan Qu, Hossein Rahmani, Jun Liu*
- 322 Free Lunch Enhancements for Multi-modal Crowd Counting, *Haoliang Meng, Xiaopeng Hong, Zhengqin Lai, Miao Shang*
- 323 RORem: Training a Robust Object Remover with Human-in-the-Loop, *Ruibin Li, Tao Yang, Song Guo, Lei Zhang*
- 324 Exact: Exploring Space-Time Perceptive Clues for Weakly Supervised
* Satellite Image Time Series Semantic Segmentation, *Hao Zhu, Yan Zhu, Jiayu Xiao, Tianxiang Xiao, Yike Ma, Yucheng Zhang, Feng Dai*
- 325 MaSS13K: A Matting-level Semantic Segmentation Benchmark, *Chenxi Xie, Minghan Li, Hui Zeng, Jun Luo, Lei Zhang*
- 326 Insightful Instance Features for 3D Instance Segmentation, *Wonseok Roh, Hwanhee Jung, Giljoo Nam, Dong In Lee, Hyeongcheol Park, Sang Ho Yoon, Jungseock Joo, Sangpil Kim*
- 327 Convex Combination Star Shape Prior for Data-driven Image Semantic Segmentation, *Xinyu Zhao, Jun Xie, Shengzhe Chen, Jun Liu*
- 328 InstanceGaussian: Appearance-Semantic Joint Gaussian Representation for 3D Instance-Level Perception, *Haijie Li, Yanmin Wu, Jiarui Meng, Qiankun Gao, Zhiyao Zhang, Ronggang Wang, Jian Zhang*
- 329 From Multimodal LLMs to Generalist Embodied Agents: Methods
○ and Lessons, *Andrew Szot, Bogdan Mazouze, Omar Attia, Aleksei Timofeev, Harsh Agrawal, Devon Hjelm, Zhe Gan, Zsolt Kira, Alexander Toshev*
- 330 Mosaic3D: Foundation Dataset and Model for Open-Vocabulary 3D Segmentation, *Junha Lee, Chunghyun Park, Jaesung Choe, Yu-Chiang Frank Wang, Jan Kautz, Minsu Cho, Chris Choy*
- 331 UnCommon Objects in 3D, *Xingchen Liu, Piyush Tayal, Jianyuan Wang, Jesus Zarzar, Tom Monnier, Konstantinos Tertikas, Jiali Duan, Antoine Toisoul, Jason Y. Zhang, Natalia Neverova, Andrea Vedaldi, Roman Shapovalov, David Novotny*
- 332 PanoGS: Gaussian-based Panoptic Segmentation for 3D Open Vocabulary Scene Understanding, *Hongjia Zhai, Hai Li, Zhenzhe Li, Xiaokun Pan, Yijia He, Guofeng Zhang*
- 333 Masked Point-Entity Contrast for Open-Vocabulary 3D Scene Understanding, *Yan Wang, Baoxiong Jia, Ziyu Zhu, Siyuan Huang*
- 334 Dr. Splat: Directly Referring 3D Gaussian Splatting via Direct
* Language Embedding Registration, *Kim Jun-Seong, GeonU Kim, Kim Yu-Ji, Yu-Chiang Frank Wang, Jaesung Choe, Tae-Hyun Oh*
- 335 Inst3D-LMM: Instance-Aware 3D Scene Understanding with
* Multi-modal Instruction Tuning, *Hanxun Yu, Wentong Li, Song Wang, Junbo Chen, Jianke Zhu*
- 336 Universal Scene Graph Generation, *Shengqiong Wu, Hao Fei, Tat-seng Chua*
- 337 DSPNet: Dual-vision Scene Perception for Robust 3D Question Answering, *Jingzhou Luo, Yang Liu, Weixing Chen, Zhen Li, Yaowei Wang, Guanbin Li, Liang Lin*
- 338 Feature4X: Bridging Any Monocular Video to 4D Agentic AI with Versatile Gaussian Feature Fields, *Shijie Zhou, Hui Ren, Yijia Weng, Shuwang Zhang, Zhen Wang, DeJia Xu, Zhiwen Fan, Suya You, Zhangyang Wang, Leonidas Guibas, Achuta Kadambi*
- 339 g3D-LF: Generalizable 3D-Language Feature Fields for Embodied Tasks, *Zihan Wang, Gim Hee Lee*
- 340 Magma: A Foundation Model for Multimodal AI Agents, *Jianwei Yang, Reuben Tan, Qianhui Wu, Ruijie Zheng, Baolin Peng, Yongyuan Liang, Yu Gu, Mu Cai, Seonghyeon Ye, Joel Jang, Yuquan Deng, Jianfeng Gao*
- 341 Mosaic of Modalities: A Comprehensive Benchmark for Multimodal Graph Learning, *Jing Zhu, Yuhang Zhou, Shengyi Qian, Zhongmou He, Tong Zhao, Neil Shah, Danai Koutra*
- 342 Style Evolving along Chain-of-Thought for Unknown-Domain
* Object Detection, *Zihao Zhang, Aming Wu, Yahong Han*
- 343 Olympus: A Universal Task Router for Computer Vision Tasks,
* Yuanze Lin, Yunsheng Li, Dongdong Chen, Weijian Xu, Ronald Clark, Philip Torr
- 344 Filter Images First, Generate Instructions Later: Pre-Instruction
* Data Selection for Visual Instruction Tuning, *Bardia Safaei, Faizan Siddiqui, Jiacong Xu, Vishal M. Patel, Shao-Yuan Lo*
- 345 Is 'Right' Right? Enhancing Object Orientation Understanding in Multimodal Large Language Models through Egocentric Instruction Tuning, *Ji Hyeok Jung, Eun Tae Kim, Seoyeon Kim, Joo Ho Lee, Bumsoo Kim, Buru Chang*
- 346 Argus: Vision-Centric Reasoning with Grounded Chain-of-Thought, *Yunze Man, De-An Huang, Guilan Liu, Shiwei Sheng, Shilong Liu, Liang-Yan Gui, Jan Kautz, Yu-Xiong Wang, Zhiding Yu*
- 347 Model Diagnosis and Correction via Linguistic and Implicit Attribute Editing, *Xuanbai Chen, Xiang Xu, Zhihua Li, Tianchen Zhao, Pietro Perona, Qin Zhang, Yifan Xing*
- 348 Foundations of the Theory of Performance-Based Ranking, *Sébastien Piérard, Anaïs Halin, Anthony Cioppa, Adrien Deliege, Marc Van Droogenbroeck*
- 349 EarthDial: Turning Multi-sensory Earth Observations to Interactive
☆ Dialogues, *Sagar Soni, Akshay Dudhane, Hiyam Debary, Mustansar Fiaz, Muhammad Akhtar Munir, Muhammad Sohail Danish, Paolo Fraccaro, Campbell D Watson, Levente J Klein, Fahad Shahbaz Khan, Salman Khan*

- 350 EMOE: Modality-Specific Enhanced Dynamic Emotion Experts,
☆ *Yiyang Fang, Wenke Huang, Guancheng Wan, Kehua Su, Mang Ye*
- 351 XLRs-Bench: Could Your Multimodal LLMs Understand Extremely
* Large Ultra-High-Resolution Remote Sensing Imagery?,
*Fengxiang Wang, Hongzhen Wang, Zonghao Guo, Di Wang,
Yulin Wang, Mingshuo Chen, Qiang Ma, Long Lan, Wenjing Yang,
Jing Zhang, Zhiyuan Liu, Maosong Sun*
- 352 DiN: Diffusion Model for Robust Medical VQA with Semantic
Noisy Labels, *Erjian Guo, Zhen Zhao, Zicheng Wang, Tong Chen,
Yunyi Liu, Luping Zhou*
- 353 DVIN: Dynamic Visual Routing Network for Weakly Supervised
Referring Expression Comprehension, *Xiaofu Chen, Yaxin Luo,
Gen Luo, Jiayi Ji, Henghui Ding, Yiyi Zhou*
- 354 ROD-MLLM: Towards More Reliable Object Detection in
Multimodal Large Language Models, *Heng Yin, Yuqiang Ren,
Ke Yan, Shouhong Ding, Yongtao Hao*
- 355 PerLA: Perceptive 3D Language Assistant, *Guofeng Mei, Wei Lin,
☆ Luigi Riz, Yujiao Wu, Fabio Poiesi, Yiming Wang*
- 356 BACON: Improving Clarity of Image Captions via Bag-of-Concept
Graphs, *Zhantao Yang, Ruili Feng, Keyu Yan, Huangji Wang,
Zhicai Wang, Shangwen Zhu, Han Zhang, Jie Xiao, Pingyu Wu,
Kai Zhu, Jixuan Chen, Chen-Wei Xie, Yue Yang, Hongyang Zhang,
Yu Liu, Fan Cheng*
- 357 Human-centered Interactive Learning via MLLMs for Text-to-
Image Person Re-identification, *Yang Qin, Chao Chen, Zhihang
Fu, Dezhong Peng, Xi Peng, Peng Hu*
- 358 Q-Eval-100K: Evaluating Visual Quality and Alignment Level for
○ Text-to-Vision Content, *Zicheng Zhang, Tengchuan Kou, Shushi
Wang, Chunyi Li, Wei Sun, Wei Wang, Xiaoyu Li, Zongyu Wang,
Xuezhi Cao, Xiongkuo Min, Xiaohong Liu, Guangtao Zhai*
- 359 Reason-before-Retrieve: One-Stage Reflective Chain-of-Thoughts
* for Training-Free Zero-Shot Composited Image Retrieval,
*Yuanmin Tang, Jue Zhang, Xiaoting Qin, Jing Yu, Gaopeng Gou,
Gang Xiong, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, Qi Wu*
- 360 Can Machines Understand Composition? Dataset and Benchmark
* for Photographic Image Composition Embedding and Understanding,
*Zhaoran Zhao, Peng Lu, Anran Zhang, Peipei Li, Xia Li,
Xuannan Liu, Yang Hu, Shiyi Chen, Liwei Wang, Wenhao Guo*
- 361 Active Data Curation Effectively Distills Large-Scale Multimodal
☆ Models, *Vishaal Udandara, Nikhil Parthasarathy,
Muhammad Ferjad Naeem, Talfan Evans, Samuel Albanie,
Federico Tombari, Yongqin Xian, Alessio Tonioni, Olivier J. Henaff*
- 362 Yo'Chameleon: Personalized Vision and Language Generation,
*Thao Nguyen, Krishna Kumar Singh, Jing Shi, Trung Bui,
Yong Jae Lee, Yuheng Li*
- 363 Relation-Rich Visual Document Generator for Visual Information
Extraction, *Zi-Han Jiang, Chien-Wei Lin, Wei-Hua Li,
Hsuan-Tung Liu, Yi-Ren Yeh, Chu-Song Chen*
- 364 Marten: Visual Question Answering with Mask Generation for
Multi-modal Document Understanding, *Zining Wang, Tongkun
Guan, Pei Fu, Chen Duan, Qianyi Jiang, Zhentao Guo, Shan Guo,
Junfeng Luo, Wei Shen, Xiaokang Yang*
- 365 A Simple yet Effective Layout Token in Large Language Models
for Document Understanding, *Zhaoqing Zhu, Chuwei Luo, Zirui
Shao, Feiyu Gao, Hangdi Xing, Qi Zheng, Ji Zhang*
- 366 Teaching Large Language Models to Regress Accurate Image
Quality Scores Using Score Distribution, *Zhiyuan You, Xin Cai,
Jinjin Gu, Tianfan Xue, Chao Dong*
- 367 FineLIP: Extending CLIP's Reach via Fine-Grained Alignment with
Longer Text Inputs, *Mothilal Asokan, Kebin Wu, Fatima Albreiki*
- 368 MarkushGrapher: Joint Visual and Textual Recognition of
Markush Structures, *Lucas Morin, Valery Weber, Ahmed Nassar,
Gerhard Ingmar Meijer, Luc Van Gool, Yawei Li, Peter Staar*
- 369 Accurate Scene Text Recognition with Efficient Model Scaling
and Cloze Self-Distillation, *Andrea Maracani, Savas Ozkan,
Sijun Cho, Hyowon Kim, Eunchung Noh, Jeongwon Min, Cho Jung
Min, Dookun Park, Mete Ozay*
- 370 Mamba as a Bridge: Where Vision Foundation Models Meet
* Vision Language Models for Domain-Generalized Semantic
Segmentation, *Xin Zhang, Robby T. Tan*
- 371 RAP: Retrieval-Augmented Personalization for Multimodal Large
Language Models, *Haoran Hao, Jiaming Han, Changsheng Li,
Yu-Feng Li, Xiangyu Yue*
- 372 What's in the Image? A Deep-Dive into the Vision of Vision
Language Models, *Omri Kaduri, Shai Bagon, Tali Dekel*
- 373 Mitigating Hallucinations in Large Vision-Language Models via
○ DPO: On-Policy Data Hold the Key, *Zhihe Yang, Xufang Luo,
Dongqi Han, Yunjian Xu, Dongsheng Li*
- 374 HoVLE: Unleashing the Power of Monolithic Vision-Language
Models with Holistic Vision-Language Embedding, *Chenxin Tao,
Shiqian Su, Xizhou Zhu, Chenyu Zhang, Zhe Chen, Jiawen Liu,
Wenhai Wang, Lewei Lu, Gao Huang, Yu Qiao, Jifeng Dai*
- 375 FlashSloth: Lightning Multimodal Large Language Models via
Embedded Visual Compression, *Bo Tong, Bokai Lai, Yiyi Zhou,
Gen Luo, Yunhang Shen, Ke Li, Xiaoshuai Sun, Rongrong Ji*
- 376 Identifying and Mitigating Position Bias of Multi-image Vision-
○ Language Models, *Xinyu Tian, Shu Zou, Zhaoyuan Yang, Jing Zhang*
- 377 PACT: Pruning and Clustering-Based Token Reduction for Faster
Visual Language Models, *Mohamed Dhouib, Davide Buscaldi,
Sonia Vanier, Aymen Shabou*
- 378 Conical Visual Concentration for Efficient Large Vision-Language
Models, *Long Xing, Qidong Huang, Xiaoyi Dong, Jiajie Lu, Pan Zhang,
Yuhang Zang, Yuhang Cao, Conghui He, Jiaqi Wang, Feng Wu, Dahua Lin*
- 379 Assessing and Learning Alignment of Unimodal Vision and
*☆ Language Models, *Le Zhang, Qian Yang, Aishwarya Agrawal*
- 380 Continual SFT Matches Multimodal RLHF with Negative
☆ Supervision, *Ke Zhu, Yu Wang, Yanpeng Sun, Qiang Chen,
Jiangjiang Liu, Gang Zhang, Jingdong Wang*
- 381 ClearSight: Visual Signal Enhancement for Object Hallucination
Mitigation in Multimodal Large Language Models, *Hao Yin,
Guangzong Si, Zilei Wang*
- 382 Nullu: Mitigating Object Hallucinations in Large Vision-Language
Models via HalluSpace Projection, *Le Yang, Ziwei Zheng,
Boxu Chen, Zhengyu Zhao, Chenhao Lin, Chao Shen*
- 383 Antidote: A Unified Framework for Mitigating LVL
☆ Hallucinations in Counterfactual Presupposition and Object
Perception, *Yuanchen Wu, Lu Zhang, Hang Yao, Junlong Du,
Ke Yan, Shouhong Ding, Yunsheng Wu, Xiaoqiang Li*
- 384 MLLM-as-a-Judge for Image Safety without Human Labeling,
* *Zhenting Wang, Shuming Hu, Shiyu Zhao, Xiaowen Lin,
Felix Juefei-Xu, Zhuowei Li, Ligong Han, Harihar Subramanyam,
Li Chen, Jianfa Chen, Nan Jiang, Lingjuan Lyu, Shiqing Ma,
Dimitris N. Metaxas, Ankit Jain*
- 385 Can Large Vision-Language Models Correct Semantic Grounding
Errors By Themselves?, *Yuan-Hong Liao, Rafid Mahmood,
Sanja Fidler, David Acuna*
- 386 Chain of Attack: On the Robustness of Vision-Language Models
Against Transfer-Based Adversarial Attacks, *Peng Xie, Yequan Bie,
Jianda Mao, Yangqiu Song, Yang Wang, Hao Chen, Kani Chen*
- 387 COSMOS: Cross-Modality Self-Distillation for Vision Language
Pre-training, *Sanghwan Kim, Rui Xiao, Mariana-Iuliana Georgescu,
Stephan Alaniz, Zeynep Akata*
- 388 Reproducible Vision-Language Models Meet Concepts Out of
Pre-Training, *Ziliang Chen, Xin Huang, Xiaoxuan Fan, Keze Wang,
Yuyu Zhou, Quanlong Guan, Liang Lin*
- 389 Once-Tuning-Multiple-Variants: Tuning Once and Expanded as
Multiple Vision-Language Model Variants, *Chong Yu, Tao Chen,
Zhongxue Gan*
- 390 Skip Tuning: Pre-trained Vision-Language Models are Effective
and Efficient Adapters Themselves, *Shihan Wu, Ji Zhang,
Pengpeng Zeng, Lianli Gao, Jingkuan Song, Heng Tao Shen*
- 391 SkySense-O: Towards Open-World Remote Sensing Interpretation
with Vision-Centric Visual-Language Modeling, *Qi Zhu, Jiangwei Lao,
Deyi Ji, Junwei Luo, Kang Wu, Yingying Zhang, Lixiang Ru, Jian Wang,
Jingdong Chen, Ming Yang, Dong Liu, Feng Zhao*
- 392 Task-Aware Clustering for Prompting Vision-Language
Models, *Fusheng Hao, Fengxiang He, Fuxiang Wu, Tichao Wang,
Chengqun Song, Jun Cheng*
- 393 Learning Textual Prompts for Open-World Semi-Supervised
Learning, *Yuxin Fan, Junbiao Cui, Jiye Liang*

- 394 BiomedCoOp: Learning to Prompt for Biomedical Vision-Language Models, *Taha Koleilat, Hojat Asgariandehkordi, Hassan Rivaz, Yiming Xiao*
- 395 ILIAS: Instance-Level Image retrieval At Scale, ☆ *Giorgos Kordopatis-Zilos, Vladan Stojnić, Anna Manko, Pavel Suma, Nikolaos-Antonios Ypsilantis, Nikos Efthymiadis, Zakaria Laskar, Jiri Matas, Ondrej Chum, Giorgos Tolias*
- 396 VILA-M3: Enhancing Vision-Language Models with Medical Expert Knowledge, ☆ *Vishwesh Nath, Wenqi Li, Dong Yang, Andriy Myronenko, Mingxin Zheng, Yao Lu, Zhijian Liu, Hongxu Yin, Yee Man Law, Yucheng Tang, Pengfei Guo, Can Zhao, Ziyue Xu, Yufan He, Stephanie Harmon, Benjamin Simon, Greg Heinrich, Stephen Aylward, Marc Edgar, Michael Zephyr, Pavlo Molchanov, Baris Turkbey, Holger Roth, Daguang Xu*
- 397 Explaining in Diffusion: Explaining a Classifier with Diffusion Semantics, *Tahira Kazimi, Ritika Allada, Pinar Yanardag*
- 398 Gazing at Rewards: Eye Movements as a Lens into Human and AI Decision-Making in Hybrid Visual Foraging, *Bo Wang, Dingwei Tan, Yen-Ling Kuo, Zhaowei Sun, Jeremy M. Wolfe, Tat-Jen Cham, Mengmi Zhang*
- 399 DeCLIP: Decoupled Learning for Open-Vocabulary Dense Perception, *Junjie Wang, Bin Chen, Yulin Li, Bin Kang, Yichi Chen, Zhuotao Tian*
- 400 Masked Scene Modeling: Narrowing the Gap Between Supervised and Self-Supervised Learning in 3D Scene Understanding, *Pedro Hermosilla, Christian Stippel, Leon Sick*
- 401 Lessons and Insights from a Unifying Study of Parameter-Efficient Fine-Tuning (PEFT) in Visual Recognition, ☆ *Zheda Mai, Ping Zhang, Cheng-Hao Tu, Hong-You Chen, Quang-Huy Nguyen, Li Zhang, Wei-Lun Chao*
- 402 TADFormer: Task-Adaptive Dynamic TransFormer for Efficient Multi-Task Learning, *Seungmin Baek, Soyul Lee, Hayeon Jo, Hyesong Choi, Dongbo Min*
- 403 LoKi: Low-dimensional KAN for Efficient Fine-tuning Image Models, *Xuan Cai, Renjie Pan, Hua Yang*
- 404 Training-free Neural Architecture Search through Variance of Knowledge of Deep Network Weights, *Ondrej Tybl, Lukas Neumann*
- 405 FIMA-Q: Post-Training Quantization for Vision Transformers by Fisher Information Matrix Approximation, ☆ *Zhuguanyu Wu, Shihe Wang, Jiayi Zhang, Jiaxin Chen, Yunhong Wang*
- 406 Transformers without Normalization, *Jiachen Zhu, Xinlei Chen, Kaiming He, Yann LeCun, Zhuang Liu*
- 407 GroupMamba: Efficient Group-Based Visual State Space Model, *Abdelrahman Shaker, Syed Talal Wasim, Salman Khan, Juergen Gall, Fahad Shahbaz Khan*
- 408 EfficientViM: Efficient Vision Mamba with Hidden State Mixer based State Space Duality, *Sanghyeok Lee, Joonmyung Choi, Hyunwoo J. Kim*
- 409 JamMa: Ultra-lightweight Local Feature Matching with Joint Mamba, *Xiaoyong Lu, Songlin Du*
- 410 Keep the Balance: A Parameter-Efficient Symmetrical Framework for RGB+X Semantic Segmentation, ○ *Jiaxin Cai, Jingze Su, Qi Li, Wenjie Yang, Shu Wang, Tiesong Zhao, Shengfeng He, Wenxi Liu*
- 411 Mamba-Reg: Vision Mamba Also Needs Registers, *Feng Wang, Jiahao Wang, Sucheng Ren, Guoyizhe Wei, Jieru Mei, Wei Shao, Yuyin Zhou, Alan Yuille, Cihang Xie*
- 412 Rethinking Token Reduction with Parameter-Efficient Fine-Tuning in ViT for Pixel-Level Tasks, *Cheng Lei, Ao Li, Hu Yao, Ce Zhu, Le Zhang*
- 413 No Pains, More Gains: Recycling Sub-Salient Patches for Efficient High-Resolution Image Recognition, ☆ *Rong Qin, Xin Liu, Xingyu Liu, Jiaxuan Liu, Jinglei Shi, Liang Lin, Jufeng Yang*
- 414 Language Guided Concept Bottleneck Models for Interpretable Continual Learning, *Lu Yu, Haoyu Han, Zhe Tao, Hantao Yao, Changsheng Xu*
- 415 LLMDeT: Learning Strong Open-Vocabulary Object Detectors under the Supervision of Large Language Models, ☆ *Shenghao Fu, Qize Yang, Qijie Mo, Junkai Yan, Xihan Wei, Jingke Meng, Xiaohua Xie, Wei-Shi Zheng*
- 416 Mask-Adapter: The Devil is in the Masks for Open-Vocabulary Segmentation, *Yongkang Li, Tianheng Cheng, Bin Feng, Wenyu Liu, Xinggang Wang*
- 417 Parameter-efficient Fine-tuning in Hyperspherical Space for Open-vocabulary Semantic Segmentation, *Zelin Peng, Zhengqin Xu, Zhilin Zeng, Yu Huang, Yaoming Wang, Wei Shen*
- 418 DocSAM: Unified Document Image Segmentation via Query Decomposition and Heterogeneous Mixed Learning, *Xiao-Hui Li, Fei Yin, Cheng-Lin Liu*
- 419 Efficient Test-time Adaptive Object Detection via Sensitivity-Guided Pruning, ○ *Kunyu Wang, Xueyang Fu, Xin Lu, Chengjie Ge, Chengzhi Cao, Wei Zhai, Zheng-Jun Zha*
- 420 Distilling Spectral Graph for Object-Context Aware Open-Vocabulary Semantic Segmentation, *Chanyoung Kim, Dayun Ju, Woojung Han, Ming-Hsuan Yang, Seong Jae Hwang*
- 421 FisherTune: Fisher-Guided Robust Tuning of Vision Foundation Models for Domain Generalized Segmentation, *Dong Zhao, Jinlong Li, Shuang Wang, Mengyao Wu, Qi Zang, Nicu Sebe, Zhun Zhong*
- 422 POT: Prototypical Optimal Transport for Weakly Supervised Semantic Segmentation, *Jian Wang, Tianhong Dai, Bingfeng Zhang, Siyue Yu, Eng Gee Lim, Jimin Xiao*
- 423 FALCON: Fairness Learning via Contrastive Attention Approach to Continual Semantic Scene Understanding, *Thanh-Dat Truong, Utsav Prabhu, Bhiksha Raj, Jackson Cothren, Khoa Luu*
- 424 WISNet: Pseudo Label Generation on Unbalanced and Patch Annotated Waste Images, *Shifan Zhang, Hongzi Zhu, Yinan He, Minyi Guo, Ziyang Lou, Shan Chang*
- 425 Few-Shot Recognition via Stage-Wise Retrieval-Augmented Finetuning, *Tian Liu, Huixin Zhang, Shubham Parashar, Shu Kong*
- 426 Compositional Caching for Training-free Open-vocabulary Attribute Detection, ☆ *Marco Garosi, Alessandro Conti, Gaowen Liu, Elisa Ricci, Massimiliano Mancini*
- 427 Open Ad-hoc Categorization with Contextualized Feature Learning, *Zilin Wang, Sangwoo Mo, Stella X. Yu, Sima Behpour, Liu Ren*
- 428 MOS: Modeling Object-Scene Associations in Generalized Category Discovery, *Zhengyuan Peng, Jinpeng Ma, Zhimin Sun, Ran Yi, Haichuan Song, Xin Tan, Lizhuang Ma*
- 429 Search and Detect: Training-Free Long Tail Object Detection via Web-Image Retrieval, *Mankeerat Sidhu, Hetarth Chopra, Ansel Blume, Jeonghwan Kim, Revanth Gangi Reddy, Heng Ji*
- 430 Fractal Calibration for Long-tailed Object Detection, ☆ *Konstantinos Panagiotis Alexandridis, Ismail Elezi, Jiankang Deng, Anh Nguyen, Shan Luo*
- 431 Classifier-to-Bias: Toward Unsupervised Automatic Bias Detection for Visual Classifiers, ☆ *Quentin Guimard, Moreno D'Incà, Massimiliano Mancini, Elisa Ricci*
- 432 DEIM: DETR with Improved Matching for Fast Convergence, *Shihua Huang, Zhichao Lu, Xiaodong Cun, Yongjun Yu, Xiao Zhou, Xi Shen*
- 433 CLIP is Strong Enough to Fight Back: Test-time Counterattacks towards Zero-shot Adversarial Robustness of CLIP, *Songlong Xing, Zhengyu Zhao, Nicu Sebe*
- 434 FlexUOD: The Answer to Real-world Unsupervised Image Outlier Detection, *Zhonghang Liu, Kun Zhou, Changshuo Wang, Wen-Yan Lin, Jiangbo Lu*
- 435 UniVAD: A Training-free Unified Model for Few-shot Visual Anomaly Detection, *Zhaopeng Gu, Bingke Zhu, Guibo Zhu, Yingying Chen, Ming Tang, Jinqiao Wang*
- 436 Towards Training-free Anomaly Detection with Vision and Language Foundation Models, *Jinjin Zhang, Guodong Wang, Yizhou Jin, Di Huang*
- 437 Real-IAD D³: A Real-World 2D/Pseudo-3D/3D Dataset for Industrial Anomaly Detection, *Wenbing Zhu, Lidong Wang, Ziqing Zhou, Chengjie Wang, Yurui Pan, Ruoyi Zhang, Zuhao Chen, Linjie Cheng, Bin-Bin Gao, Jiangning Zhang, Zhenye Gan, Yuxie Wang, Yulong Chen, Shuguang Qian, Mingmin Chi, Bo Peng, Lizhuang Ma*
- 438 DFM: Differentiable Feature Matching for Anomaly Detection, *Sheng Wu, Yimi Wang, Xudong Liu, Yuguang Yang, Runqi Wang, Guodong Guo, David Doermann, Baochang Zhang*

- 439 Automatic Joint Structured Pruning and Quantization for Efficient Neural Network Training and Compression, *Xiaoyi Qu, David Aponte, Colby Banbury, Daniel P. Robinson, Tianyu Ding, Kazuhito Koishida, Ilya Zharkov, Tianyi Chen*
- 440 OPTICAL: Leveraging Optimal Transport for Contribution Allocation in Dataset Distillation, *Xiao Cui, Yulei Qin, Wengang Zhou, Hongsheng Li, Houqiang Li*
- 441 Learning Compatible Multi-Prize Subnetworks for Asymmetric Retrieval, *Yushuai Sun, Zikun Zhou, Dongmei Jiang, Yaowei Wang, Jun Yu, Guangming Lu, Wenjie Pei*
- 442 Less is More: Efficient Model Merging with Binary Task Switch, *Biqing Qi, Fangyuan Li, Zhen Wang, Junqi Gao, Dong Li, Peng Ye, Bowen Zhou*
- 443 On the Generalization of Handwritten Text Recognition Models, *Carlos Garrido-Munoz, Jorge Calvo-Zaragoza*
- 444 Investigating the Role of Weight Decay in Enhancing Nonconvex SGD, *Tao Sun, Yuhao Huang, Li Shen, Kele Xu, Bao Wang*
- 445 KAC: Kolmogorov-Arnold Classifier for Continual Learning, *Yusong Hu, Zichen Liang, Fei Yang, Qibin Hou, Xialei Liu, Ming-Ming Cheng*
- 446 LoRA Subtraction for Drift-Resistant Space in Exemplar-Free Continual Learning, *Xuan Liu, Xiaobin Chang*
- 447 Maintaining Consistent Inter-Class Topology in Continual Test-Time Adaptation, *Chenggong Ni, Fan Lyu, Jiayao Tan, Fuyuan Hu, Rui Yao, Tao Zhou*
- 448 Tripartite Weight-Space Ensemble for Few-Shot Class-Incremental Learning, *Juntae Lee, Munawar Hayat, Sungrack Yun*
- 449 T-CIL: Temperature Scaling using Adversarial Perturbation for Calibration in Class-Incremental Learning, *Seong-Hyeon Hwang, Minsu Kim, Steven Euijong Whang*
- 450 Seeking Consistent Flat Minima for Better Domain Generalization via Refining Loss Landscapes, *Aodi Li, Liansheng Zhuang, Xiao Long, Minghong Yao, Shafei Wang*
- 451 PEER Pressure: Model-to-Model Regularization for Single Source Domain Generalization, *Dong Kyu Cho, Inwoo Hwang, Sanghack Lee*
- 452 A Unified Framework for Heterogeneous Semi-supervised Learning, *Marzi Heidari, Abdullah Alchihabi, Hao Yan, Yuhong Guo*
- 453 CGMatch: A Different Perspective of Semi-supervised Learning, *Bo Cheng, Jueqing Lu, Yuan Tian, Haifeng Zhao, Yi Chang, Lan Du*
- 454 Label Shift Meets Online Learning: Ensuring Consistent Adaptation with Universal Dynamic Regret, *Yucong Dai, Shilin Gu, Ruidong Fan, Chao Xu, Chenping Hou*
- 455 Overcoming Shortcut Problem in VLM for Robust Out-of-Distribution Detection, *Zhuo Xu, Xiang Xiang, Yifan Liang*
- 456 H2ST: Hierarchical Two-Sample Tests for Continual Out-of-Distribution Detection, *Yuhang Liu, Wenjie Zhao, Yunhui Guo*
- 457 Detecting Out-of-Distribution Through the Lens of Neural Collapse, *Litian Liu, Yao Qin*
- 458 FedCS: Coreset Selection for Federated Learning, *Chenhe Hao, Weiyang Xie, Daixun Li, Haonan Qin, Hangyu Ye, Leyuan Fang, Yunsong Li*
- 459 FedCALM: Conflict-aware Layer-wise Mitigation for Selective Aggregation in Deeper Personalized Federated Learning, *Hao Zheng, Zhigang Hu, Liu Yang, Meiguang Zheng, Aikun Xu, Boyu Wang*
- 460 Model Poisoning Attacks to Federated Learning via Multi-Round Consistency, *Yueqi Xie, Minghong Fang, Neil Zhenqiang Gong*
- 461 FedSPA: Generalizable Federated Graph Learning under Homophily Heterogeneity, *Zihan Tan, Guancheng Wan, Wenke Huang, He Li, Guibin Zhang, Carl Yang, Mang Ye*
- 462 TAET: Two-Stage Adversarial Equalization Training on Long-Tailed Distributions, *Wang Yu-Hang, Junkang Guo, Aolei Liu, Kaihao Wang, Zaitong Wu, Zhenyu Liu, Wenfei Yin, Jian Liu*
- 463 Let Samples Speak: Mitigating Spurious Correlation by Exploiting the Clusterness of Samples, *Weiwei Li, Junzhuo Liu, Yuanyuan Ren, Yuchen Zheng, Yahao Liu, Wen Li*
- 464 Uncertainty Weighted Gradients for Model Calibration, *Jinxu Lin, Linwei Tao, Minjing Dong, Chang Xu*
- 465 Enhancing Testing-Time Robustness for Trusted Multi-View Classification in the Wild, *Wei Liu, Yufei Chen, Xiaodong Yue*
- 466 Enhanced then Progressive Fusion with View Graph for Multi-View Clustering, *Zhibin Dong, Meng Liu, Siwei Wang, Ke Liang, Yi Zhang, Suyuan Liu, Jiaqi Jin, Xinwang Liu, En Zhu*
- 467 A Hubness Perspective on Representation Learning for Graph-Based Multi-View Clustering, *Zheming Xu, He Liu, Congyan Lang, Tao Wang, Yidong Li, Michael C. Kampffmeyer*
- 468 CLOC: Contrastive Learning for Ordinal Classification with Multi-Margin N-pair Loss, *Dileepa Pitawela, Gustavo Carneiro, Hsiang-Ting Chen*
- 469 STIL: Semi-supervised Tabular-Image Learning for Comprehensive Task-Relevant Information Exploration in Multimodal Classification, *Siyi Du, Xinzhe Luo, Declan P. O'Regan, Chen Qin*
- 470 Q-PART: Quasi-Periodic Adaptive Regression with Test-time Training for Pediatric Left Ventricular Ejection Fraction Regression, *Jie Liu, Tiexin Qin, Hui Liu, Yilei Shi, Lichao Mou, Xiao Xiang Zhu, Shiqi Wang, Haoliang Li*
- 471 OralXrays-9: Towards Hospital-Scale Panoramic X-ray Anomaly Detection via Personalized Multi-Object Query-Aware Mining, *Bingzhi Chen, Sisi Fu, Xiaocheng Fang, Jieyi Cai, Boya Zhang, Minhua Lu, Yishu Liu*
- 472 DART: Disease-aware Image-Text Alignment and Self-correcting Re-alignment for Trustworthy Radiology Report Generation, *Sang-Jun Park, Keun-Soo Heo, Dong-Hee Shin, Young-Han Son, Ji-Hye Oh, Tae-Eui Kam*
- 473 FOCUS: Knowledge-enhanced Adaptive Visual Compression for Few-shot Whole Slide Image Classification, *Zhengrui Guo, Conghao Xiong, Jiabo Ma, Qichen Sun, Lishuang Feng, Jinzhuo Wang, Hao Chen*
- 474 M3amba: Memory Mamba is All You Need for Whole Slide Image Classification, *Tingting Zheng, Kui Jiang, Yi Xiao, Sicheng Zhao, Hongxun Yao*
- 475 MERGE: Multi-faceted Hierarchical Graph-based GNN for Gene Expression Prediction from Whole Slide Histopathology Images, *Aniruddha Ganguly, Debolina Chatterjee, Wentao Huang, Jie Zhang, Alisa Yurovsky, Travis Steele Johnson, Chao Chen*
- 476 Test-Time Domain Generalization via Universe Learning: A Multi-Graph Matching Approach for Medical Image Segmentation, *Xingguo Lv, Xingbo Dong, Liwen Wang, Jiewen Yang, Lei Zhao, Bin Pu, Zhe Jin, Xuejun Li*
- 477 CSC-PA: Cross-image Semantic Correlation via Prototype Attentions for Single-network Semi-supervised Breast Tumor Segmentation, *Zhenhui Ding, Guilian Chen, Qin Zhang, Huisi Wu, Jing Qin*
- 478 Take the Bull by the Horns: Learning to Segment Hard Samples, *Yuan Guo, Jingyu Kong, Yu Wang, Yuping Duan*
- 479 Cross-Modal Interactive Perception Network with Mamba for Lung Tumor Segmentation in PET-CT Images, *Jie Mei, Chenyu Lin, Yu Qiu, Yaonan Wang, Hui Zhang, Ziyang Wang, Dong Dai*
- 480 KMD: Koopman Multi-modality Decomposition for Generalized Brain Tumor Segmentation under Incomplete Modalities, *Tianyi Liu, Haochuan Jiang, Kaizhu Huang*
- 481 Noise-Consistent Siamese-Diffusion for Medical Image Synthesis and Segmentation, *Kunpeng Qiu, Zhiqiang Gao, Zhiying Zhou, Mingjie Sun, Yongxin Guo*
- 482 DeNVer: Deformable Neural Vessel Representations for Unsupervised Video Vessel Segmentation, *Chun-Hung Wu, Shih-Hong Chen, Chih-Yao Hu, Hsin-Yu Wu, Kai-Hsin Chen, Yu-You Chen, Chih-Hai Su, Chih-Kuo Lee, Yu-Lun Liu*
- 483 VasTSD: Learning 3D Vascular Tree-state Space Diffusion Model for Angiography Synthesis, *Zhifeng Wang, Renjiao Yi, Xin Wen, Chenyang Zhu, Kai Xu*

10:30 - 12:30 Demos (ExHallD)

- 1 Event Ellipsometer: Event-based Mueller-Matrix Video Imaging, *Ryota Maeda, Yunseong Moon, Seung-Hwan Baek*
- 2 Mobile Diffusion for Video Editing, *Amirhossein Habibi*
- 3 3D-Pose-Based Evaluation of the Risk of Sarcopenia, *Yu-Hsuan Chiu, Gee-Sern Jison Hsu, Jiunn-Horng Kang, Jie-Syuan Wu*
- 4 Automated Video Clustering and Annotation Software (AVCAS), *Chukwuemeka Duru, George Awad*

- 5 GenECA: A Generalizable Framework for Real-Time Multimodal Embodied Conversational Agents with Emotion-Sensitive Interaction, *Santosh Patapati, Trisanth Srinivasan*
- 6 AR2D2: Training a Robot Without A Robot, *Abhimanyu Saigal, Jiafei Duan, Ranjay Krishna, Dieter Fox*
- 7 HiRISE: High-Resolution Image Scaling for Edge ML via In-Sensor Compression and Selective ROI, *Brendan Reidy, Peyton Chandarana, Ramtin Zand*
- 8 Event-Driven ASL Recognition: Building a DVS Dataset for Neuromorphic Systems, *Arshia Eslami, James (Blake) Seekings, Peyton Chandarana, Ramtin Zand*
- 9 Real-time Facial Expression Recognition For Intuitive Robot Coaches, *Peyton Chandarana, Mohammadreza Mohammadi, Hasti Zanganeh, Ramtin Zand*
- 10 TaoAvatar, *Jianchuan Chen, Jingchuan Hu, Gaige Wang, Zhonghua Jiang, Tiansong Zhou, Zhiwen Chen, Chengfei Lv*
- 11 Toward Provably Private Image Obfuscation with Diffusion Models, *Joseph Roberson, Tianbao Ma, Liyue Fan*
- 12 VP Lab: a PEFT-Enabled Visual Prompting Laboratory for Semantic Segmentation, *Thomas Frick, Niccolo Avogaro, Yagmur G. Cinar, Daniel Caraballo, Cezary Skura, Filip M. Janicki, Piotr Kluska, Brown Ebouky, Nicola Farronato, Florian Scheidegger, Cristiano Malossi, Konrad Schindler, Andrea Bartezzaghi, Roy Assaf, Mattia Rigotti*
- 13 Edge AI in Action: Deploying Multi-Modal Models in Edge AI Devices, *Fabricio Batista Narcizo, Elizabete Munzlinger, Anuj Dutt, Shan Ahmed Shaffi, Sai Narsi Reddy Donthi Reddy*
- 14 Morfis, *Dimitrios Mallis, Mohamed Adel Mohamed Ali, Ahmet Serdar Karadeniz, Anis Kacem, Djamila Aouada*

11:00 / 17:00 Art Gallery Tour with Curator, Luba Elliott
(30 mins each) (ExHall A1)

11:30 - 13:30 LUNCH (ExHall C)

11:00 - 13:00 Doctoral Consortium (201 A)

13:00 - 14:15 Oral Session 4A: Image and Video Synthesis
(Karl Dean Ballroom)

🏆 - Award candidate paper

- 1 Reconstruction vs. Generation: Taming Optimization Dilemma in Latent Diffusion Models *Jingfeng Yao, Bin Yang, Xinggang Wang*
- 2 Language-Guided Image Tokenization for Generation, *Kaiwen Zha, Lijun Yu, Alireza Fathi, David A. Ross, Cordelia Schmid, Dina Katabi, Xiuye Gu*
- 3 DreamRelation: Bridging Customization and Relation Generation, *Qingyu Shi, Lu Qi, Jianzong Wu, Jinbin Bai, Jingbo Wang, Yunhai Tong, Xiangtai Li*
- 4 Infinity ∞ : Scaling Bitwise AutoRegressive Modeling for High-Resolution Image Synthesis, *Jian Han, Jinlai Liu, Yi Jiang, Bin Yan, Yuqi Zhang, Zehuan Yuan, Bingyue Peng, Xiaobing Liu*
- 5 Autoregressive Distillation of Diffusion Transformers, *Yeongmin Kim, Sotiris Anagnostidis, Yuming Du, Edgar Schönfeld, Jonas Kohler, Markos Georgopoulos, Albert Pumarola, Ali Thabet, Arsiom Sanakoyeu*

13:00 - 14:15 Oral Session 4B: Embodied Computer Vision
(ExHall A2)

- 1 PDFactor: Learning Tri-Perspective View Policy Diffusion Field for Multi-Task Robotic Manipulation, *Jingyi Tian, Le Wang, Sanping Zhou, Sen Wang, Jiayi Li, Haowen Sun, Wei Tang*
- 2 RoboSpatial: Teaching Spatial Understanding to 2D and 3D Vision-Language Models for Robotics, *Chan Hee Song, Valts Blukis, Jonathan Tremblay, Stephen Tyree, Yu Su, Stan Birchfield*
- 3 GROVE: A Generalized Reward for Learning Open-Vocabulary Physical Skill *Jieming Cui, Tengyu Liu, Ziyu Meng, Jiale Yu, Ran Song, Wei Zhang, Yixin Zhu, Siyuan Huang*
- 4 Navigation World Models, *Amir Bar, Gaoyue Zhou, Danny Tran, Trevor Darrell, Yann LeCun*
- 5 Viewpoint Rosetta Stone: Unlocking Unpaired Ego-Exo Videos for View-invariant Representation Learning, *Mi Luo, Zihui Xue, Alex Dimakis, Kristen Grauman*

13:00 - 14:15 Oral Session 4C: 3D Computer Vision
(Davidson Ballroom)

- 1 DORNet: A Degradation Oriented and Regularized Network for Blind Depth Super-Resolution, *Zhengxue Wang, Zhiqiang Yan, Jinshan Pan, Guangwei Gao, Kai Zhang, Jian Yang*
- 2 Convex Relaxation for Robust Vanishing Point Estimation in Manhattan World, *Bangyan Liao, Zhenjun Zhao, Haoang Li, Yi Zhou, Yingping Zeng, Hao Li, Peidong Liu*
- 3 Learned Binocular-Encoding Optics for RGBD Imaging Using Joint Stereo and Focus Cues, *Yuhui Liu, Liangxun Ou, Qiang Fu, Hadi Amata, Wolfgang Heidrich, Yifan Peng*
- 4 Camera Resection from Known Line Pencils and a Radially Distorted Scanline, *Juan C. Dibene, Enrique Dunn*
- 5 Opportunistic Single-Photon Time of Flight, *Sotiris Nouisias, Mian Wei, Howard Xiao, Maxx Wu, Shahmeer Athar, Kevin J. Wang, Anagh Malik, David A. Barmherzig, David B. Lindell, Kyros N. Kutulakos*

13:30 - 14:30 Art Panel (209 ABC)

14:15 - 14:30 Courtesy Break

14:30 - 15:30 KEYNOTE 2 - Laurens van der Maaten; The Llama Herd of Models: System 1, 2, 3 Go! (Karl Dean Ballroom)

15:45 - 16:45 PAMI TC Meeting (ExHall A2)

16:30 - 17:00 Poster Setup (Hall D)

17:00 - 19:00 Poster Session 4 & Exhibit Hall w/ Coffee Break
(Hall D)

* - Highlight paper 🏆 - Award candidate paper

○ - Oral Paper ☆ - Outstanding Reviewer

- 1 EmoDubber: Towards High Quality and Emotion Controllable Movie Dubbing, *Gaoxiang Cong, Jiadong Pan, Liang Li, Yuankai Qi, Yuxin Peng, Anton van den Hengel, Jian Yang, Qingming Huang*
- 2 From Faces to Voices: Learning Hierarchical Representations for High-quality Video-to-Speech, *Ji-Hoon Kim, Jeongsoo Choi, Jaehun Kim, Chaeyoung Jung, Joon Son Chung*
- 3 Diffusion-based Realistic Listening Head Generation via Hybrid Motion Modeling, *Yinuo Wang, Yanbo Fan, Xuan Wang, Guo Yu, Fei Wang*
- 4 VLOGGER: Multimodal Diffusion for Embodied Avatar Synthesis, *Enric Corona, Andrei Zanfir, Eduard Gabriel Bazavan, Nikos Kolotouros, Thimo Alldieck, Cristian Sminchisescu*
- 5 HunyuanPortrait: Implicit Condition Control for Enhanced Portrait Animations, *Zunnan Xu, Zhentao Yu, Zixiang Zhou, Jun Zhou, Xiaoyu Jin, Fa-ting Hong, Xiaozhong Ji, Junwei Zhu, Chengfei Cai, Shiyu Tang, Qin Lin, Xiu Li, Qinglin Lu*
- 6 MobilePortrait: Real-Time One-Shot Neural Head Avatars on Mobile Devices, *Jianwen Jiang, Gaojie Lin, Zhengkun Rong, Chao Liang, Yongming Zhu, Jiaqi Yang, Tianyun Zhong*
- 7 Gaussian Eigen Models for Human Heads, *Wojciech Zielonka, Timo Bolkart, Thabo Beeler, Justus Thies*
- 8 Zero-1-to-A: Zero-Shot One Image to Animatable Head Avatars Using Video Diffusion, *Zhenglin Zhou, Fan Ma, Hehe Fan, Tat-Seng Chua*
- 9 PERSE: Personalized 3D Generative Avatars from A Single Portrait, *Hyunsoo Cha, Inhee Lee, Hanbyul Joo*
- 10 WildAvatar: Learning In-the-wild 3D Avatars from the Web, *Zihao Huang, Shoukang Hu, Guangcong Wang, Tianqi Liu, Yuhang Zang, Zhiguo Cao, Wei Li, Ziwei Liu*
- 11 Creating Your Editable 3D Photorealistic Avatar with Tetrahedron-constrained Gaussian Splatting, *Hanxi Liu, Yifan Men, Zhouhui Lian*
- 12 FreeCloth: Free-form Generation Enhances Challenging Clothed Human Modeling, *Hang Ye, Xiaoxuan Ma, Hai Ci, Wentao Zhu, Yizhou Wang*
- 13 MagicArticulate: Make Your 3D Models Articulation-Ready, *Chaoyue Song, Jianfeng Zhang, Xiu Li, Fan Yang, Yiwen Chen, Zhongcong Xu, Jun Hao Liew, Xiaoyang Guo, Fayao Liu, Jiashi Feng, Guosheng Lin*
- 14 PSHuman: Photorealistic Single-image 3D Human Reconstruction using Cross-Scale Multiview Diffusion and Explicit Remeshing, *Peng Li, Wangguandong Zheng, Yuan Liu, Tao Yu, Yangguang Li, Xingqun Qi, Xiaowei Chi, Siyu Xia, Yan-Pei Cao, Wei Xue, Wenhan Luo, Yike Guo*

- 15 Multi-focal Conditioned Latent Diffusion for Person Image
☆ Synthesis, *Jiaqi Liu, Jichao Zhang, Paolo Rota, Nicu Sebe*
- 16 Robust-MVTN: Learning Cross-Pose Feature Alignment and Fusion for Robust Multi-View Virtual Try-On, *Nannan Zhang, Yijiang Li, Dong Du, Zheng Chong, Zhengwentai Sun, Jianhao Zeng, Yusheng Dai, Zhengyu Xie, Hairui Zhu, Xiaoguang Han*
- 17 GroomLight: Hybrid Inverse Rendering for Relightable Human Hair Appearance Modeling, *Yang Zheng, Menglei Chai, Delio Vicini, Yuxiao Zhou, Yinghao Xu, Leonidas Guibas, Gordon Wetzstein, Thabo Beeler*
- 18 S*3-Face: SSS-Compliant Facial Reflectance Estimation via Diffusion Priors, *Xingyu Ren, Jiankang Deng, Yuhao Cheng, Wenhan Zhu, Yichao Yan, Xiaokang Yang, Stefanos Zafeiriou, Chao Ma*
- 19 DL2G: Degradation-guided Local-to-Global Restoration for Eyeglass Reflection Removal, *Zhilv Yi, Xiao Lu, Hong Ding, Jingbo Hu, Zhi Jiang, Chunxia Xiao*
- 20 Improving Visual and Downstream Performance of Low-Light
☆ Enhancer with Vision Foundation Models Collaboration, *Yuxuan Gu, Haoxuan Wang, Pengyang Ling, Zhixiang Wei, Huaian Chen, Yi Jin, Enhong Chen*
- 21 PIDSR: Complementary Polarized Image Demosaicing and
☆ Super-Resolution, *Shuangfan Zhou, Chu Zhou, Youwei Lyu, Heng Guo, Zhanyu Ma, Boxin Shi, Imari Sato*
- 22 Learned Binocular-Encoding Optics for RGBD Imaging Using
○ Joint Stereo and Focus Cues, *Yuhui Liu, Liangxun Ou, Qiang Fu, Hadi Amata, Wolfgang Heidrich, Yifan Peng*
- 23 Volume Tells: Dual Cycle-Consistent Diffusion for 3D Fluorescence
* Microscopy De-noising and Super-Resolution, *Zelin Li, Chenwei Wang, Zhaoke Huang, Yiming Ma, Cunming Zhao, Zhongying Zhao, Hong Yan*
- 24 CoCoGaussian: Leveraging Circle of Confusion for Gaussian
☆ Splatting from Defocused Images, *Jungho Lee, Suhwan Cho, Taeoh Kim, Ho-Deok Jang, Minhyeok Lee, Geonho Cha, Dongyoon Wee, Dogyoon Lee, Sangyoun Lee*
- 25 UltraFusion: Ultra High Dynamic Imaging using Exposure
* Fusion, *Zixuan Chen, Yujin Wang, Xin Cai, Zhiyuan You, Zheming Lu, Fan Zhang, Shi Guo, Tianfan Xue*
- 26 LookCloser: Frequency-aware Radiance Field for Tiny-Detail Scene, *Xiaoyu Zhang, Weihong Pan, Chong Bao, Xiyu Zhang, Xiaojun Xiang, Hanqing Jiang, Hujun Bao*
- 27 SpecTre-GS: Modeling Highly Specular Surfaces with Reflected
* Nearby Objects by Tracing Rays in 3D Gaussian Splatting, *Jiajun Tang, Fan Fei, Zhihao Li, Xiao Tang, Shiyong Liu, Youyu Chen, Binxiao Huang, Zhenyu Chen, Xiaofei Wu, Boxin Shi*
- 28 SVG-IR: Spatially-Varying Gaussian Splatting for Inverse Rendering, *Hanxiao Sun, Yupeng Gao, Jin Xie, Jian Yang, Beibei Wang*
- 29 RainyGS: Efficient Rain Synthesis with Physically-Based Gaussian Splatting, *Qiyu Dai, Xingyu Ni, Qianfan Shen, Wenzheng Chen, Baoquan Chen, Mengyu Chu*
- 30 Light Transport-aware Diffusion Posterior Sampling for Single-
* View Reconstruction of 3D Volumes, *Ludwic Leonard, Nils Thuerey, Rüdiger Westermann*
- 31 StarVector: Generating Scalable Vector Graphics Code from Images and Text, *Juan A. Rodriguez, Abhay Puri, Shubham Agarwal, Issam H. Laradji, Pau Rodriguez, Sai Rajeswar, David Vazquez, Christopher Pal, Marco Pedersoli*
- 32 Sparse Voxels Rasterization: Real-time High-fidelity Radiance Field Rendering, *Cheng Sun, Jaesung Choe, Charles Loop, Wei-Chiu Ma, Yu-Chiang Frank Wang*
- 33 BG-Triangle: Bézier Gaussian Triangle for 3D Vectorization and Rendering, *Minye Wu, Haizhao Dai, Kaixin Yao, Tinne Tuytelaars, Jingyi Yu*
- 34 UniPhy: Learning a Unified Constitutive Model for Inverse Physics Simulation, *Himangi Mittal, Peiye Zhuang, Hsin-Ying Lee, Shubham Tulsiani*
- 35 Mesh Mamba: A Unified State Space Model for Saliency Prediction in Non-Textured and Textured Meshes, *Kaiwei Zhang, Dandan Zhu, Xiongkuo Min, Guangtao Zhai*
- 36 DirectTriGS: Triplane-based Gaussian Splatting Field Representation for 3D Generation, *Xiaoliang Ju, Hongsheng Li*
- 37 SF3D: Stable Fast 3D Mesh Reconstruction with UV-unwrapping and Illumination Disentanglement, *Mark Boss, Zixuan Huang, Aaryaman Vasishtha, Varun Jampani*
- 38 Dora: Sampling and Benchmarking for 3D Shape Variational Auto-Encoders, *Rui Chen, Jianfeng Zhang, Yixun Liang, Guan Luo, Weiye Li, Jiarui Liu, Xiu Li, Xiaoxiao Long, Jiashi Feng, Ping Tan*
- 39 Few-shot Implicit Function Generation via Equivariance,
* Suizhi Huang, Xingyi Yang, Hongtao Lu, Xinchao Wang
- 40 Instant3dit: Multiview Inpainting for Fast Editing of 3D Objects, *Amir Barda, Matheus Gadelha, Vladimir G. Kim, Noam Aigerman, Amit H. Bermano, Thibault Groueix*
- 41 PyTorchGeoNodes: Enabling Differentiable Shape Programs for
☆ 3D Shape Reconstruction, *Sinisa Stekovic, Arslan Artykov, Stefan Ainetter, Mattia D'Urso, Friedrich Fraundorfer*
- 42 Perturb-and-Revise: Flexible 3D Editing with Generative Trajectories, *Susung Hong, Johanna Karras, Ricardo Martin-Brualla, Ira Kemelmacher-Shlizerman*
- 43 DaCapo: Score Distillation as Stacked Bridge for Fast and High-quality 3D Editing, *Yufei Huang, Bangyan Liao, Yuqi Hu, Haitao Lin, Lirong Wu, Siyuan Li, Cheng Tan, Zicheng Liu, Yunfan Liu, Zelin Zang, Chang Yu, Zhen Lei*
- 44 Structure from Collision,
* Takuhiro Kaneko
- 45 GuardSplat: Efficient and Robust Watermarking for 3D Gaussian Splatting, *Zixuan Chen, Guangcong Wang, Jiahao Zhu, Jianhuang Lai, Xiaohua Xie*
- 46 DORNet: A Degradation Oriented and Regularized Network for
○ Blind Depth Super-Resolution, *Zhengxue Wang, Zhiqiang Yan, Jinshan Pan, Guangwei Gao, Kai Zhang, Jian Yang*
- 47 FlexGS: Train Once, Deploy Everywhere with Many-in-One
☆ Flexible 3D Gaussian Splatting, *Hengyu Liu, Yuehao Wang, Chenxin Li, Ruisi Cai, Kevin Wang, Wuyang Li, Pavlo Molchanov, Peihao Wang, Zhangyang Wang*
- 48 Evolving High-Quality Rendering and Reconstruction in a Unified
☆ Framework with Contribution-Adaptive Regularization, *You Shen, Zhipeng Zhang, Xinyang Li, Yansong Qu, Yu Lin, Shengchuan Zhang, Liujuan Cao*
- 49 OmniSplat: Taming Feed-Forward 3D Gaussian Splatting for
* Omnidirectional Images with Editable Capabilities, *Suyoung Lee, Jaeyoung Chung, Kihoon Kim, Jaeyoo Huh, Gunhee Lee, Minsoo Lee, Kyoung Mu Lee*
- 50 AuraFusion360: Augmented Unseen Region Alignment for Reference-based 360° Unbounded Scene Inpainting, *Chung-Ho Wu, Yang-Jung Chen, Ying-Huan Chen, Jie-Ying Lee, Bo-Hsu Ke, Chun-Wei Tuan Mu, Yi-Chuan Huang, Chin-Yang Lin, Min-Hung Chen, Yen-Yu Lin, Yu-Lun Liu*
- 51 Free360: Layered Gaussian Splatting for Unbounded 360-Degree View Synthesis from Extremely Sparse and Unposed Views, *Chong Bao, Xiyu Zhang, Zehao Yu, Jiale Shi, Guofeng Zhang, Songyou Peng, Zhaopeng Cui*
- 52 Advancing Adversarial Robustness in GNeRFs: The IL2-NeRF Attack, *Nicole Meng, Caleb Manicke, Ronak Sahu, Caiwen Ding, Yingjie Lao*
- 53 EVPGS: Enhanced View Prior Guidance for Splatting-based Extrapolated View Synthesis, *Jiahe Li, Feiyu Wang, Xiaochao Qu, Chengjing Wu, Luoqi Liu, Ting Liu*
- 54 CamFreeDiff: Camera-free Image to Panorama Generation with Diffusion Model, *Xiaoding Yuan, Shitao Tang, Kejie Li, Peng Wang*
- 55 Pippo: High-Resolution Multi-View Humans from a Single Image,
* Yash Kant, Ethan Weber, Jin Kyu Kim, Rawal Khiradkar, Su Zhaoen, Julieta Martinez, Igor Gilitschenski, Shunsuke Saito, Timur Bagautdinov
- 56 3DEnhancer: Consistent Multi-View Diffusion for 3D Enhancement, *Yihang Luo, Shangchen Zhou, Yushi Lan, Xingang Pan, Chen Change Loy*
- 57 MegaSynth: Scaling Up 3D Scene Reconstruction with Synthesized Data, *Hanwen Jiang, Zexiang Xu, Desai Xie, Ziwen Chen, Haian Jin, Fujun Luan, Zhixin Shu, Kai Zhang, Sai Bi, Xin Sun, Jiuxiang Gu, Qixing Huang, Georgios Pavlakos, Hao Tan*
- 58 DepthSplat: Connecting Gaussian Splatting and Depth,

- ☆ *Haofei Xu, Songyou Peng, Fangjinhua Wang, Hermann Blum, Daniel Barath, Andreas Geiger, Marc Pollefeys*
- 59 Dense-To-Sparse Video Diffusion For High-fidelity Multi-View
- ☆ *Images Synthesis, Fan Yang, Jianfeng Zhang, Jun Hao Liew, Chaoyue Song, Zhongcong Xu, Xiu Li, Jiashi Feng, Guosheng Lin*
- 60 SimVS: Simulating World Inconsistencies for Robust View Synthesis,
- ☆ *Alex Trevithick, Roni Paiss, Philipp Henzler, Dor Verbin, Rundi Wu, Hadi Alzayer, Ruiqi Gao, Ben Poole, Jonathan T. Barron, Aleksander Holynski, Ravi Ramamoorthi, Pratul P. Srinivasan*
- 61 VideoScene: Distilling Video Diffusion Model to Generate 3D Scenes
- * *in One Step, Hanyang Wang, Fangfu Liu, Jiawei Chi, Yueqi Duan*
- 62 ActiveGAMER: Active Gaussian Mapping through Efficient
- ☆ *Rendering, Liyan Chen, Huangying Zhan, Kevin Chen, Xiangyu Xu, Qingan Yan, Changjiang Cai, Yi Xu*
- 63 EAP-GS: Efficient Augmentation of Pointcloud for 3D Gaussian
- Splatting in Few-shot Scene Reconstruction, *Dongrui Dai, Yuxiang Xing*
- 64 Shading Meets Motion: Self-supervised Indoor 3D
- Reconstruction Via Simultaneous Shape-from-Shading and
- Structure-from-Motion, *Guoyu Lu*
- 65 Instant Gaussian Stream: Fast and Generalizable Streaming of
- * *Dynamic Scene Reconstruction via Gaussian Splatting, Jinbo Yan, Rui Peng, Zhiyan Wang, Luyang Tang, Jiayu Yang, Jie Liang, Jiahao Wu, Ronggang Wang*
- 66 BARD-GS: Blur-Aware Reconstruction of Dynamic Scenes via
- Gaussian Splatting, *Yiren Lu, Yunlai Zhou, Disheng Liu, Tuo Liang, Yu Yin*
- 67 GauSTAR: Gaussian Surface Tracking and
- Reconstruction, *Chengwei Zheng, Lixin Xue, Juan Zarate, Jie Song*
- 68 Opportunistic Single-Photon Time of Flight, *Sotiris Nousias,*
- *Mian Wei, Howard Xiao, Maxx Wu, Shahmeer Athar, Kevin J. Wang, Anagh Malik, David A. Barmherzig, David B. Lindell, Kyros N. Kutulakos*
- 69 ImViD: Immersive Volumetric Videos for Enhanced VR Engagement,
- * *Zhengxian Yang, Shi Pan, Shengqi Wang, Haoxiang Wang, Li Lin, Guanjun Li, Zhengqi Wen, Borong Lin, Jianhua Tao, Tao Yu*
- 70 Reconstructing Animals and the Wild, *Peter Kulits, Michael J. Black, Silvia Zuffi*
- 71 Retrieving Semantics from the Deep: an RAG Solution for Gesture
- Synthesis, *M. Hamza Mughal, Rishabh Dabral, Merel C.J. Scholman, Vera Demberg, Christian Theobalt*
- 72 Dense Dispersed Structured Light for Hyperspectral 3D Imaging
- of Dynamic Scenes, *Suhyun Shin, Seungwoo Yoon, Ryota Maeda, Seung-Hwan Baek*
- 73 HUSH: Holistic Panoramic 3D Scene Understanding using
- Spherical Harmonics, *Jongsung Lee, Harin Park, Byeong-Uk Lee, Kyungdon Joo*
- 74 USP-Gaussian: Unifying Spike-based Image Reconstruction,
- * *Pose Correction and Gaussian Splatting, Kang Chen, Jiyuan Zhang, Zecheng Hao, Yajing Zheng, Tiejun Huang, Zhaoqi Yu*
- 75 SVDC: Consistent Direct Time-of-Flight Video Depth Completion
- with Frequency Selective Fusion, *Xuan Zhu, Jijun Xiang, Xianqi Wang, Longliang Liu, Yu Wang, Hong Zhang, Fei Guo, Xin Yang*
- 76 Sea-ing in Low-light, *Nisha Varghese, A. N. Rajagopalan*
- 77 Consistency-aware Self-Training for Iterative-based Stereo
- Matching, *Jingyi Zhou, Peng Ye, Haoyu Zhang, Jiakang Yuan, Rao Qiang, Liu YangChenXu, Wu Caillin, Feng Xu, Tao Chen*
- 78 SLAM3R: Real-Time Dense Scene Reconstruction from
- * *Monocular RGB Videos, Yuzheng Liu, Siyan Dong, Shuzhe Wang, Yingda Yin, Yanchao Yang, Qingnan Fan, Baoquan Chen*
- 79 4D-Fly: Fast 4D Reconstruction from a Single Monocular Video,
- Diankun Wu, Fangfu Liu, Yi-Hsin Hung, Yue Qian, Xiaohang Zhan, Yueqi Duan*
- 80 Camera Resection from Known Line Pencils and a Radially
- *Distorted Scanline, Juan C. Dibene, Enrique Dunn*
- 81 AnyMap: Learning a General Camera Model for Structure-from-
- Motion with Unknown Distortion in Dynamic Scenes, *Andrea Porfiri Dal Cin, Georgi Dikov, Jihong Ju, Mohsen Ghafoorian*
- 82 SSHNet: Unsupervised Cross-modal Homography Estimation via
- * *Problem Reformulation and Split Optimization, Junchen Yu, Si-*
- Yuan Cao, Runmin Zhang, Chenghao Zhang, Zhu Yu, Shujie Chen, Bailin Yang, Hui-Liang Shen*
- 83 MAST3R-SLAM: Real-Time Dense SLAM with 3D Reconstruction
- * *Priors, Riku Murai, Eric Dexheimer, Andrew J. Davison*
- 84 Relative Pose Estimation through Affine Corrections of Monocular
- * *Depth Priors, Yifan Yu, Shaohui Liu, Rémi Pautrat, Marc Pollefeys, Viktor Larsson*
- 85 AnyCam: Learning to Recover Camera Poses and Intrinsics
- ☆ *from Casual Videos, Felix Wimbauer, Weirong Chen, Dominik Muhle, Christian Rupprecht, Daniel Cremers*
- 86 GPVK-VL: Geometry-Preserving Virtual Keyframes for Visual
- Localization under Large Viewpoint Changes, *Yunxuan Li, Lei Fan, Xiaoying Xing, Jianxiong Zhou, Ying Wu*
- 87 Reloc3r: Large-Scale Training of Relative Camera Pose
- ☆ *Regression for Generalizable, Fast, and Accurate Visual Localization, Siyan Dong, Shuzhe Wang, Shaohui Liu, Lulu Cai, Qingnan Fan, Juho Kannala, Yanchao Yang*
- 88 Viewpoint Rosetta Stone: Unlocking Unpaired Ego-Exo Videos
- *for View-invariant Representation Learning, Mi Luo, Zihui Xue, Alex Dimakis, Kristen Grauman*
- 89 Self-Supervised Cross-View Correspondence with Predictive
- Cycle Consistency, *Alan Baade, Changan Chen*
- 90 Can Generative Video Models Help Pose Estimation?,
- * *Ruojin Cai, Jason Y. Zhang, Philipp Henzler, Zhengqi Li, Noah Snavely, Ricardo Martin-Brualla*
- 91 Light3R-SfM: Towards Feed-forward Structure-from-Motion,
- * *Sven Elflein, Qunjie Zhou, Laura Leal-Taixé*
- 92 BADGR: Bundle Adjustment Diffusion Conditioned by Gradients
- * *for Wide-Baseline Floor Plan Reconstruction, Yuguang Li, Ivaylo Boyadzhiev, Zixuan Liu, Linda Shapiro, Alex Colburn*
- 93 SAT-HMR: Real-Time Multi-Person 3D Mesh Estimation via Scale-
- Adaptive Tokens, *Chi Su, Xiaoxuan Ma, Jiajun Su, Yizhou Wang*
- 94 HiPART: Hierarchical Pose AutoRegressive Transformer for
- ☆ *Occluded 3D Human Pose Estimation, Hongwei Zheng, Han Li, Wenrui Dai, Ziyang Zheng, Chenglin Li, Junni Zou, Hongkai Xiong*
- 95 Pos3R: 6D Pose Estimation for Unseen Objects Made Easy,
- ☆ *Weijian Deng, Dylan Campbell, Chunyi Sun, Jiahao Zhang, Shubham Kanitkar, Matt E. Shaffer, Stephen Gould*
- 96 ONDA-Pose: Occlusion-Aware Neural Domain Adaptation for
- Self-Supervised 6D Object Pose Estimation, *Tao Tan, Qiulei Dong*
- 97 Leveraging Global Stereo Consistency for Category-Level Shape
- and 6D Pose Estimation from Stereo Images, *Junning Qiu, Minglei Lu, Fei Wang, Yu Guo, Yonggen Ling*
- 98 One-shot 3D Object Canonicalization based on Geometric and
- * *Semantic Consistency, Li Jin, Yujie Wang, Wenzheng Chen, Qiyu Dai, Qingzhe Gao, Xueying Qin, Baoquan Chen*
- ☆ *SPAR3D: Stable Point-Aware Reconstruction of 3D Objects from Single Images, Zixuan Huang, Mark Boss, Aaryaman Vasishtha, James M. Rehg, Varun Jampani*
- 100 SPMTrack: Spatio-Temporal Parameter-Efficient Fine-Tuning
- with Mixture of Experts for Scalable Visual Tracking, *Wenrui Cai, Qingjie Liu, Yunhong Wang*
- 101 MUST: The First Dataset and Unified Framework for Multispectral
- UAV Single Object Tracking, *Haolin Qin, Tingfa Xu, Tianhao Li, Zhenxiang Chen, Tao Feng, Jianan Li*
- 102 Convex Relaxation for Robust Vanishing Point Estimation in
- 🏠 *Manhattan World, Bangyan Liao, Zhenjun Zhao, Haoang Li, Yi Zhou, Yingping Zeng, Hao Li, Peidong Liu*
- 103 All-Day Multi-Camera Multi-Target Tracking, *Huijie Fan, Yu Qiao, Yihao Zhen, Tinghui Zhao, Baojie Fan, Qiang Wang*
- 104 Shape Abstraction via Marching Differentiable Support
- * *Functions, Sunkyoung Park, Jeongmin Lee, Dongjun Lee*
- 105 MES3D: Mining Effective Semantic Cues for 3D Reconstruction
- from a Single Image, *Shaoming Li, Qing Cai, Songqi Kong, Runqing Tan, Heng Tong, Shiji Qiu, Yongguo Jiang, Zhi Liu*
- 106 Implicit Correspondence Learning for Image-to-Point Cloud
- * *Registration, Xinjun Li, Wenfei Yang, Jiacheng Deng, Zhixin Cheng, Xu Zhou, Tianzhu Zhang*
- 107 Consistent Normal Orientation for 3D Point Clouds via Least
- Squares on Delaunay Graph, *Rao Fu, Jianmin Zheng, Liang Yu*

- 108 Zero-shot RGB-D Point Cloud Registration with Pre-trained Large Vision Model, *Haobo Jiang, Jin Xie, Jian Yang, Liang Yu, Jianmin Zheng*
- 109 SuperPC: A Single Diffusion Model for Point Cloud Completion, Upsampling, Denoising, and Colorization, *Yi Du, Zhipeng Zhao, Shaoshu Su, Sharath Golluri, Haoze Zheng, Runmao Yao, Chen Wang*
- 110 Occlusion-aware Text-Image-Point Cloud Pretraining for Open-World 3D Object Recognition, *Khanh Nguyen, Ghulam Mubashar Hassan, Ajmal Mian*
- 111 PMA: Towards Parameter-Efficient Point Cloud Understanding via
☆ Point Mamba Adapter, *Yaohua Zha, Yanzi Wang, Hang Guo, Jinpeng Wang, Tao Dai, Bin Chen, Zhihao Ouyang, Xue Yuerong, Ke Chen, Shu-Tao Xia*
- 112 Point Cloud Upsampling Using Conditional Diffusion Module with Adaptive Noise Suppression, *Boqian Zhang, Shen Yang, Hao Chen, Chao Yang, Jing Jia, Guang Jiang*
- 113 Generalized Few-shot 3D Point Cloud Segmentation with Vision-Language Models, *Zhaochong An, Guolei Sun, Yun Liu, Runjia Li, Junlin Han, Ender Konukoglu, Serge Belongie*
- 114 EdgeDiff: Edge-aware Diffusion Network for Building Reconstruction from Point Clouds, *Yujun Liu, Ruisheng Wang, Shangfeng Huang, Guorong Cai*
- 115 WeatherGen: A Unified Diverse Weather Generator for LiDAR Point Clouds via Spider Mamba Diffusion, *Yang Wu, Yun Zhu, Kaihua Zhang, Jianjun Qian, Jin Xie, Jian Yang*
- 116 FASTER: Focal token Acquiring-and-Scaling Transformer for Long-term 3D Objection Detection, *Chenxu Dang, ZaiPeng Duan, Pei An, Xinmin Zhang, Xuzhong Hu, Jie Ma*
- 117 LiSu: A Dataset and Method for LiDAR Surface Normal Estimation, *Dušan Malić, Christian Fruhwirth-Reisinger, Samuel Schuster, Horst Possegger*
- 118 DiffLO: Semantic-Aware LiDAR Odometry with Diffusion-Based Refinement, *Yongshu Huang, Chen Liu, Minghang Zhu, Sheng Ao, Chenglu Wen, Cheng Wang*
- 119 SharpDepth: Sharpening Metric Depth Predictions Using Diffusion Distillation, *Duc-Hai Pham, Tung Do, Phong Nguyen, Binh-Son Hua, Khoi Nguyen, Rang Nguyen*
- 120 Prompting Depth Anything for 4K Resolution Accurate Metric Depth Estimation, *Haocong Lin, Sida Peng, Jingxiao Chen, Songyou Peng, Jiaming Sun, Minghuan Liu, Hujun Bao, Jiashi Feng, Xiaowei Zhou, Bingyi Kang*
- 121 RaCFormer: Towards High-Quality 3D Object Detection via Query-based Radar-Camera Fusion, *Xiaomeng Chu, Jiajun Deng, Guoliang You, Yifan Duan, Houqiang Li, Yanyong Zhang*
- 122 ZeroVO: Visual Odometry with Minimal Assumptions, *Lei Lai, Zekai Yin, Eshed Ohn-Bar*
- 123 Learning Occlusion-Robust Vision Transformers for Real-Time UAV Tracking, *You Wu, Xucheng Wang, Xiangyang Yang, Mengyuan Liu, Dan Zeng, Hengzhou Ye, Shuiwang Li*
- 124 On-Device Self-Supervised Learning of Low-Latency Monocular Depth from Only Events, *Jesse J. Hagenaars, Yilun Wu, Federico Paredes-Valles, Stein Stroobants, Guido C.H.E. de Croon*
- 125 Toward Real-world BEV Perception: Depth Uncertainty Estimation via Gaussian Splatting, *Shu-Wei Lu, Yi-Hsuan Tsai, Yi-Ting Chen*
- 126 3D Occupancy Prediction with Low-Resolution Queries via Prototype-aware View Transformation, *Gyeongrok Oh, Sungjune Kim, Heeju Ko, Hyung-gun Chi, Jinkyu Kim, Dongwook Lee, Daehyun Ji, Sungjoon Choi, Sujin Jang, Sangpil Kim*
- 127 SOAP: Vision-Centric 3D Semantic Scene Completion with Scene-Adaptive Decoder and Occluded Region-Aware View Projection, *Hyo-Jun Lee, Yeong Jun Koh, Hanul Kim, Hyunseop Kim, Yonguk Lee, Jinu Lee*
- 128 VoteFlow: Enforcing Local Rigidity in Self-Supervised Scene Flow, *Yancong Lin, Shiming Wang, Liangliang Nan, Julian Kooij, Holger Caesar*
- 129 VisionPAD: A Vision-Centric Pre-training Paradigm for Autonomous Driving, *Haiming Zhang, Wending Zhou, Yiyao Zhu, Xu Yan, Jiantao Gao, Dongfeng Bai, Yingjie Cai, Bingbing Liu, Shuguang Cui, Zhen Li*
- 130 InteractionMap: Improving Online Vectorized HDMap Construction with Interaction, *Kuang Wu, Chuan Yang, Zhanbin Li*
- 131 DriveScape: High-Resolution Driving Video Generation by Multi-View Feature Fusion, *Wei Wu, Xi Guo, Weixuan Tang, Tingxuan Huang, Chiyu Wang, Chenjing Ding*
- 132 T2SG: Traffic Topology Scene Graph for Topology Reasoning in Autonomous Driving, *Changsheng Lv, Mengshi Qi, Liang Liu, Huadong Ma*
- 133 Scenario Dreamer: Vectorized Latent Diffusion for Generating Driving Simulation Environments, *Luke Rowe, Roger Girgis, Anthony Gosselin, Liam Paull, Christopher Pal, Felix Heide*
- 134 Leveraging SD Map to Augment HD Map-based Trajectory Prediction, *Zhiwei Dong, Ran Ding, Wei Li, Peng Zhang, Guobin Tang, Jia Guo*
- 135 Enduring, Efficient and Robust Trajectory Prediction Attack
* in Autonomous Driving via Optimization-Driven Multi-Frame Perturbation Framework, *Yi Yu, Weizhen Han, Libing Wu, Bingyi Liu, Enshu Wang, Zhuangzhuang Zhang*
- 136 CarPlanner: Consistent Auto-regressive Trajectory Planning for Large-Scale Reinforcement Learning in Autonomous Driving, *Dongkun Zhang, Jiaming Liang, Ke Guo, Sha Lu, Qi Wang, Rong Xiong, Zhenwei Miao, Yue Wang*
- 137 SpatialLLM: A Compound 3D-Informed Design towards Spatially-
* Intelligent Large Multimodal Models, *Wufei Ma, Luoxin Ye, Celso M de Melo, Alan Yuille, Jieneng Chen*
- 138 DriveGPT4-V2: Harnessing Large Language Model Capabilities
* for Enhanced Closed-Loop Autonomous Driving, *Zhenhua Xu, Yan Bai, Yujia Zhang, Zhuoling Li, Fei Xia, Kwan-Yee K. Wong, Jianqiang Wang, Hengshuang Zhao*
- 139 Sim-to-Real Causal Transfer: A Metric Learning Approach to Causally-Aware Interaction Representations, *Ahmad Rahimi, Po-Chien Luan, Yuejiang Liu, Frano Rajić, Alexandre Alahi*
- 140 MoFlow: One-Step Flow Matching for Human Trajectory Forecasting via Implicit Maximum Likelihood Estimation based Distillation, *Yuxiang Fu, Qi Yan, Lele Wang, Ke Li, Renjie Liao*
- 141 3D-Mem: 3D Scene Memory for Embodied Exploration and Reasoning, *Yuncong Yang, Han Yang, Jiachen Zhou, Peihao Chen, Hongxin Zhang, Yilun Du, Chuang Gan*
- 142 HandOS: 3D Hand Reconstruction in One Stage, *Xingyu Chen, Zhuoheng Song, Xiaoke Jiang, Yaoqing Hu, Junzhi Yu, Lei Zhang*
- 143 MobileH2R: Learning Generalizable Human to Mobile Robot Handover Exclusively from Scalable and Diverse Synthetic Data, *Zifan Wang, Ziqing Chen, Junyu Chen, Jilong Wang, Yuxin Yang, Yunze Liu, Xueyi Liu, He Wang, Li Yi*
- 144 PDFactor: Learning Tri-Perspective View Policy Diffusion Field for
○ Multi-Task Robotic Manipulation, *Jingyi Tian, Le Wang, Sanping Zhou, Sen Wang, Jiayi Li, Haowen Sun, Wei Tang*
- 145 GROVE: A Generalized Reward for Learning Open-Vocabulary
○ Physical Skill, *Jieming Cui, Tengyu Liu, Ziyu Meng, Jiale Yu, Ran Song, Wei Zhang, Yixin Zhu, Siyuan Huang*
- 146 RoboSpatial: Teaching Spatial Understanding to 2D and 3D
○ Vision-Language Models for Robotics, *Chan Hee Song, Valts Blukis, Jonathan Tremblay, Stephen Tyree, Yu Su, Stan Birchfield*
- 147 GREAT: Geometry-Intention Collaborative Inference for Open-Vocabulary 3D Object Affordance Grounding, *Yawen Shao, Wei Zhai, Yuhang Yang, Hongchen Luo, Yang Cao, Zheng-Jun Zha*
- 148 Grounding 3D Object Affordance with Language Instructions, Visual Observations and Interactions, *He Zhu, Quyu Kong, Kechun Xu, Xunlong Xia, Bing Deng, Jieping Ye, Rong Xiong, Yue Wang*
- 149 Lift3D Policy: Lifting 2D Foundation Models for Robust 3D Robotic Manipulation, *Yueru Jia, Jiaming Liu, Sixiang Chen, Chenyang Gu, Zhilve Wang, Longzan Luo, Xiaoqi Li, Pengwei Wang, Zhongyuan Wang, Renrui Zhang, Shanghang Zhang*
- 150 OmniManip: Towards General Robotic Manipulation via Object-
* Centric Interaction Primitives as Spatial Constraints, *Mingjie Pan, Jiyao Zhang, Tianshu Wu, Yinghao Zhao, Wenlong Gao, Hao Dong*
- 151 Generating 6DoF Object Manipulation Trajectories from Action
* Description in Egocentric Vision, *Tomoya Yoshida, Shuhei Kurita, Taichi Nishimura, Shinsuke Mori*
- 152 Two by Two: Learning Multi-Task Pairwise Objects Assembly for Generalizable Robot Manipulation, *Yu Qi, Yuanchen Ju, Tianming*

- Wei, Chi Chu, Lawson L.S. Wong, Huazhe Xu*
- 153 Spatial-Temporal Graph Diffusion Policy with Kinematic Modeling for Bimanual Robotic Manipulation, *Qi Lv, Hao Li, Xiang Deng, Rui Shao, Yinchuan Li, Jianye Hao, Longxiang Gao, Michael Yu Wang, Liqiang Nie*
- 154 ZeroGrasp: Zero-Shot Shape Reconstruction Enabled Robotic Grasping, *Shun Iwase, Muhammad Zubair Irshad, Katherine Liu, Vitor Guizilini, Robert Lee, Takuya Ikeda, Ayako Amma, Koichi Nishiwaki, Kris Kitani, Rares Ambrus, Sergey Zakharov*
- 155 LatentHOI: On the Generalizable Hand Object Motion Generation with Latent Hand Diffusion, *Muchen Li, Sammy Christen, Chengde Wan, Yujun Cai, Renjie Liao, Leonid Sigal, Shugao Ma*
- 156 Reconstructing In-the-Wild Open-Vocabulary Human-Object Interactions, *Boran Wen, Dingbang Huang, Zichen Zhang, Jiahong Zhou, Jianbin Deng, Jingyu Gong, Yulong Chen, Lizhuang Ma, Yong-Lu Li*
- 157 BIGS: Bimanual Category-agnostic Interaction Reconstruction from Monocular Videos via 3D Gaussian Splatting, *Jeongwan On, Kyeonghwan Gwak, Gunyoung Kang, Junuk Cha, Soohyun Hwang, Hyein Hwang, Seungryul Baek*
- 158 FoundHand: Large-Scale Domain-Specific Learning for Controllable Hand Image Generation, *Kefan Chen, Chaerin Min, Linguang Zhang, Shreyas Hampali, Cem Keskin, Srinath Sridhar*
- 159 GigaHands: A Massive Annotated Dataset of Bimanual Hand Activities, *Rao Fu, Dingxi Zhang, Alex Jiang, Wanxia Fu, Austin Funk, Daniel Ritchie, Srinath Sridhar*
- 160 Reconstructing Close Human Interaction with Appearance and Proxemics Reasoning, *Buzhen Huang, Chen Li, Chongyang Xu, Dongyue Lu, Jinnan Chen, Yangang Wang, Gim Hee Lee*
- 161 AniMer: Animal Pose and Shape Estimation Using Family Aware Transformer, *Jin Lyu, Tianyi Zhu, Yi Gu, Li Lin, Pujin Cheng, Yebin Liu, Xiaoying Tang, Liang An*
- 162 FRAME: Floor-aligned Representation for Avatar Motion from Egocentric Video, *Andrea Boscolo Camilletto, Jian Wang, Eduardo Alvarado, Rishabh Dabral, Thabo Beeler, Marc Habermann, Christian Theobalt*
- 163 SyncSDE: A Probabilistic Framework for Diffusion Synchronization, *Hyunjun Lee, Hyunsoo Lee, Sookwan Han*
- 164 Lifting Motion to the 3D World via 2D Diffusion, *Jiaman Li, C. Karen Liu, Jiajun Wu*
- 165 Motions as Queries: One-Stage Multi-Person Holistic Human Motion Capture, *Kenkun Liu, Yurong Fu, Weihao Yuan, Jing Lin, Peihao Li, Xiaodong Gu, Lingteng Qiu, Haoqian Wang, Zilong Dong, Xiaoguang Han*
- 166 SkillMimic: Learning Basketball Interaction Skills from Demonstrations, *Yinhuai Wang, Qihan Zhao, Runyi Yu, Hok Wai Tsui, Ailing Zeng, Jing Lin, Zhengyi Luo, Jiwen Yu, Xiu Li, Qifeng Chen, Jian Zhang, Lei Zhang, Ping Tan*
- 167 Re-HOLD: Video Hand Object Interaction Reenactment via adaptive Layout-instructed Diffusion Model, *Yingying Fan, Quanwei Yang, Kaisiyuan Wang, Hang Zhou, Yingying Li, Haocheng Feng, Errui Ding, Yu Wu, Jingdong Wang*
- 168 SemGeoMo: Dynamic Contextual Human Motion Generation with Semantic and Geometric Guidance, *Peishan Cong, Ziyi Wang, Yuexin Ma, Xiangyu Yue*
- 169 Articulated Kinematics Distillation from Video Diffusion Models, *Xuan Li, Qianli Ma, Tsung-Yi Lin, Yongxin Chen, Chenfanfu Jiang, Ming-Yu Liu, Donglai Xiang*
- 170 Human Motion Instruction Tuning, *Lei Li, Sen Jia, Jianhao Wang, Zhongyu Jiang, Feng Zhou, Ju Dai, Tianfang Zhang, Zongkai Wu, Jenq-Neng Hwang*
- 171 EnergyMoGen: Compositional Human Motion Generation with Energy-Based Diffusion Model in Latent Space, *Jianrong Zhang, Hehe Fan, Yi Yang*
- 172 VideoMage: Multi-Subject and Motion Customization of Text-to-Video Diffusion Models, *Chi-Pin Huang, Yen-Siang Wu, Hung-Kai Chung, Kai-Po Chang, Fu-En Yang, Yu-Chiang Frank Wang*
- 173 F1ction: 4D Future Interaction Prediction from Video, *Kumar Ashutosh, Georgios Pavlakos, Kristen Grauman*
- 174 Mamba4D: Efficient 4D Point Cloud Video Understanding with Disentangled Spatial-Temporal State Space Models, *Jiuming Liu, Jinru Han, Lihao Liu, Angelica I. Aviles-Rivero, Chaokang Jiang, Zhe Liu, Hesheng Wang*
- 175 Layered Motion Fusion: Lifting Motion Segmentation to 3D in Egocentric Videos, *Vadim Tschernezki, Diane Larlus, Iro Laina, Andrea Vedaldi*
- 176 TimeTracker: Event-based Continuous Point Tracking for Video Frame Interpolation with Non-linear Motion, *Haoyue Liu, Jinghan Xu, Yi Chang, Hanyu Zhou, Haozhi Zhao, Lin Wang, Luxin Yan*
- 177 Buffer Anytime: Zero-Shot Video Depth and Normal from Image Priors, *Zhengfei Kuang, Tianyuan Zhang, Kai Zhang, Hao Tan, Sai Bi, Yiwei Hu, Zexiang Xu, Milos Hasan, Gordon Wetzstein, Fujun Luan*
- 178 LC-Mamba: Local and Continuous Mamba with Shifted Windows for Frame Interpolation, *Min Wu Jeong, Chae Eun Rhee*
- 179 ObjectMover: Generative Object Movement with Video Prior, *Xin Yu, Tianyu Wang, Soo Ye Kim, Paul Guerrero, Xi Chen, Qing Liu, Zhe Lin, Xiaojuan Qi*
- 180 VideoHandles: Editing 3D Object Compositions in Videos Using Video Generative Priors, *Juil Koo, Paul Guerrero, Chun-Hao P. Huang, Duygu Ceylan, Minhyuk Sung*
- 181 One-Minute Video Generation with Test-Time Training, *Jiarui Xu, Shihao Han, Karan Dalal, Daniel Kocaja, Yue Zhao, Ka Chun Cheung, Yejin Choi, Jan Kautz, Yu Sun, Xiaolong Wang*
- 182 Generative Video Propagation, *Shaoteng Liu, Tianyu Wang, Jui-Hsien Wang, Qing Liu, Zhifei Zhang, Joon-Young Lee, Yijun Li, Bei Yu, Zhe Lin, Soo Ye Kim, Jiaya Jia*
- 183 4Real-Video: Learning Generalizable Photo-Realistic 4D Video Diffusion, *Chaoyang Wang, Peiye Zhuang, Tuan Duc Ngo, Willi Menapace, Aliaksandr Siarohin, Michael Vasilkovsky, Ivan Skorokhodov, Sergey Tulyakov, Peter Wonka, Hsin-Ying Lee*
- 184 Condensing Action Segmentation Datasets via Generative Network Inversion, *Guodong Ding, Rongyu Chen, Angela Yao*
- 185 Perceptual Video Compression with Neural Wrapping, *Muhammad Umar Karim Khan, Aaron Chadha, Mohammad Ashraf Anam, Yiannis Andreopoulos*
- 186 EvEnhancer: Empowering Effectiveness, Efficiency and Generalizability for Continuous Space-Time Video Super-Resolution with Events, *Shuoyan Wei, Feng Li, Shengeng Tang, Yao Zhao, Huihui Bai*
- 187 Plug-and-Play Versatile Compressed Video Enhancement, *Huimin Zeng, Jiacheng Li, Zhiwei Xiong*
- 188 WF-VAE: Enhancing Video VAE by Wavelet-Driven Energy Flow for Latent Video Diffusion Model, *Zongjian Li, Bin Lin, Yang Ye, Liuhan Chen, Xinhua Cheng, Shenghai Yuan, Li Yuan*
- 189 LongDiff: Training-Free Long Video Generation in One Go, *Zhuoling Li, Hossein Rahmani, Qihong Ke, Jun Liu*
- 190 PatchVSR: Breaking Video Diffusion Resolution Limits with Patch-wise Video Super-Resolution, *Shian Du, Menghan Xia, Chang Liu, Xintao Wang, Jing Wang, Pengfei Wan, Di Zhang, Xiangyang Ji*
- 191 DPFlow: Adaptive Optical Flow Estimation with a Dual-Pyramid Framework, *Henrique Morimitsu, Xiaobin Zhu, Roberto M. Cesar, Xiangyang Ji, Xu-Cheng Yin*
- 192 Dynamic Content Prediction with Motion-aware Priors for Blind Face Video Restoration, *Lianxin Xie, Bingbing Zheng, Si Wu, Hau San Wong*
- 193 LP-Diff: Towards Improved Restoration of Real-World Degraded License Plate, *Haoyan Gong, Zhenrong Zhang, Yuzheng Feng, Anh Nguyen, Hongbin Liu*
- 194 AlphaPre: Amplitude-Phase Disentanglement Model for Precipitation Nowcasting, *Kenghong Lin, Baoquan Zhang, Demin Yu, Wenzhi Feng, Shidong Chen, Feifan Gao, Xutao Li, Yunming Ye*
- 195 Effective Cloud Removal for Remote Sensing Images by an Improved Mean-Reverting Denoising Model with Elucidated Design Space, *Yi Liu, Wengen Li, Jihong Guan, Shuigeng Zhou, Yichao Zhang*
- 196 Self-Learning Hyperspectral and Multispectral Image Fusion via Adaptive Residual Guided Subspace Diffusion Model, *Jian Zhu, He Wang, Yang Xu, Zebin Wu, Zhihui Wei*
- 197 Adaptive Rectangular Convolution for Remote Sensing Pansharpening, *Xueyang Wang, Zhixin Zheng, Jiandong Shao, Yule Duan, Liang-Jian Deng*

- 198 Every SAM Drop Counts: Embracing Semantic Priors for Multi-Modality Image Fusion and Beyond, *Guanyao Wu, Haoyu Liu, Hongming Fu, Yichuan Peng, Jinyuan Liu, Xin Fan, Risheng Liu*
- 199 Exposure-slot: Exposure-centric Representations Learning with Slot-in-Slot Attention for Region-aware Exposure Correction, *Donggoo Jung, Daehyun Kim, Guanghui Wang, Tae Hyun Kim*
- 200 CATANet: Efficient Content-Aware Token Aggregation for Lightweight Image Super-Resolution, *Xin Liu, Jie Liu, Jie Tang, Gangshan Wu*
- 201 ACL: Activating Capability of Linear Attention for Image Restoration, *Yubin Gu, Yuan Meng, Jiayi Ji, Xiaoshuai Sun*
- 202 Positive2Negative: Breaking the Information-Lossy Barrier in Self-Supervised Single Image Denoising, *Tong Li, Lizhi Wang, Zhiyuan Xu, Lin Zhu, Wanxuan Lu, Hua Huang*
- 203 From Zero to Detail: Deconstructing Ultra-High-Definition Image Restoration from Progressive Spectral Perspective, *Chen Zhao, Zhizhou Chen, Yunzhe Xu, Enxuan Gu, Jian Li, Zili Yi, Qian Wang, Jian Yang, Ying Tai*
- 204 Multi-Modal Contrastive Masked Autoencoders: A Two-Stage Progressive Pre-training Approach for RGBD Datasets, *Muhammad Abdullah Jamal, Omid Mohareri*
- 205 Auto-Encoded Supervision for Perceptual Image Super-Resolution, *Minkyu Lee, Sangeek Hyun, Woojin Jun, Jae-Pil Heo*
- 206 UniRestore: Unified Perceptual and Task-Oriented Image Restoration Model Using Diffusion Prior, *I-Hsiang Chen, Wei-Ting Chen, Yu-Wei Liu, Yuan-Chun Chiang, Sy-Yen Kuo, Ming-Hsuan Yang*
- 207 Uncertainty-guided Perturbation for Image Super-Resolution Diffusion Model, *Leheng Zhang, Weiye You, Kexuan Shi, Shuhang Gu*
- 208 Image Quality Assessment: Investigating Causal Perceptual Effects with Abductive Counterfactual Inference, *Wenhao Shen, Mingliang Zhou, Yu Chen, Xuekai Wei, Yong Feng, Huayan Pu, Weijia Jia*
- 209 Using Powerful Prior Knowledge of Diffusion Model in Deep Unfolding Networks for Image Compressive Sensing, *Chen Liao, Yan Shen, Dan Li, Zhongli Wang*
- 210 Accelerating Diffusion Transformer via Increment-Calibrated Caching with Channel-Aware Singular Value Decomposition, *Zhiyuan Chen, Keyi Li, Yifan Jia, Le Ye, Yufei Ma*
- 211 Optimizing for the Shortest Path in Denoising Diffusion Model, *Ping Chen, Xingpeng Zhang, Zhaoxiang Liu, Huan Hu, Xiang Liu, Kai Wang, Min Wang, Yanlin Qian, Shiguo Lian*
- 212 Acc3D: Accelerating Single Image to 3D Diffusion Models via Edge Consistency Guided Score Distillation, *Kendong Liu, Zhiyu Zhu, Hui Liu, Junhui Hou*
- 213 MambalC: State Space Models for High-Performance Learned Image Compression, *Fanhu Zeng, Hao Tang, Yihua Shao, Siyu Chen, Ling Shao, Yan Wang*
- 214 Decouple Distortion from Perception: Region Adaptive Diffusion for Extreme-low Bitrate Perception Image Compression, *Jinchang Xu, Shaokang Wang, Jintao Chen, Zhe Li, Peidong Jia, Fei Zhao, Guoqing Xiang, Zhijian Hao, Shanghang Zhang, Xiaodong Xie*
- 215 Simpler Diffusion: 1.5 FID on ImageNet512 with Pixel-space Diffusion, *Emiel Hoogeboom, Thomas Mensink, Jonathan Heek, Kay Lamerigts, Ruiqi Gao, Tim Salimans*
- 216 Layer- and Timestep-Adaptive Differentiable Token Compression Ratios for Efficient Diffusion Transformers, *Haoran You, Connelly Barnes, Yuqian Zhou, Yan Kang, Zhenbang Du, Wei Zhou, Lingzhi Zhang, Yotam Nitzan, Xiaoyang Liu, Zhe Lin, Eli Shechtman, Sohrab Amirghodsi, Yingyan Celine Lin*
- 217 Attend to Not Attended: Structure-then-Detail Token Merging for Post-training DiT Acceleration, *Haipeng Fang, Sheng Tang, Juan Cao, Enshuo Zhang, Fan Tang, Tong-Yee Lee*
- 218 NoiseCtrl: A Sampling-Algorithm-Agnostic Conditional Generation Method for Diffusion Models, *Longquan Dai, He Wang, Jinhui Tang*
- 219 See Further When Clear: Curriculum Consistency Model, *Yunpeng Liu, Boxiao Liu, Yi Zhang, Xingzhong Hou, Guanglu Song, Yu Liu, Haihang You*
- 220 RayFlow: Instance-Aware Diffusion Acceleration via Adaptive Flow Trajectories, *Huiyang Shao, Xin Xia, Yuhong Yang, Yuxi Ren, Xing Wang, Xuefeng Xiao*
- 221 Improved Video VAE for Latent Video Diffusion Model, *Pingyu Wu, Kai Zhu, Yu Liu, Liming Zhao, Wei Zhai, Yang Cao, Zheng-Jun Zha*
- 222 Pioneering 4-Bit FP Quantization for Diffusion Models: Mixup-Sign Quantization and Timestep-Aware Fine-Tuning, *Maosen Zhao, Pengtao Chen, Chong Yu, Yan Wen, Xudong Tan, Tao Chen*
- 223 TinyFusion: Diffusion Transformers Learned Shallow, *Gongfan Fang, Kunjun Li, Xinyin Ma, Xinchao Wang*
- 224 Towards Precise Scaling Laws for Video Diffusion Transformers, *Yuanyang Yin, Yaqi Zhao, Mingwu Zheng, Ke Lin, Jiarong Ou, Rui Chen, Victor Shea-Jay Huang, Jiahao Wang, Xin Tao, Pengfei Wan, Di Zhang, Baoqun Yin, Wentao Zhang, Kun Gai*
- 225 Less is More: Efficient Image Vectorization with Adaptive Parameterization, *Kaibo Zhao, Liang Bao, Yufei Li, Xu Su, Ke Zhang, Xiaotian Qiao*
- 226 Sketchtopia: A Dataset and Foundational Agents for Benchmarking Asynchronous Multimodal Communication with Iconic Feedback, *Mohd Hozaifa Khan, Ravi Kiran Sarvadevabhatla*
- 227 AniDoc: Animation Creation Made Easier, *Yihao Meng, Hao Ouyang, Hanlin Wang, Qiuyu Wang, Wen Wang, Ka Leong Cheng, Zhiheng Liu, Yujun Shen, Huamin Qu*
- 228 Through-The-Mask: Mask-based Motion Trajectories for Image-to-Video Generation, *Guy Yariv, Yuval Kirstain, Amit Zohar, Shelly Sheynin, Yaniv Taigman, Yossi Adi, Sagie Benaim, Adam Polyak*
- 229 Encapsulated Composition of Text-to-Image and Text-to-Video Models for High-Quality Video Synthesis, *Tongtong Su, Chengyu Wang, Bingyan Liu, Jun Huang, Dongming Lu*
- 230 Autoregressive Distillation of Diffusion Transformers, *Yeongmin Kim, Sotiris Anagnostidis, Yuming Du, Edgar Schönfeld, Jonas Kohler, Markos Georgopoulos, Albert Pumarola, Ali Thabet, Arslan Sanakoyeu*
- 231 EIDT-V: Exploiting Intersections in Diffusion Trajectories for Model-Agnostic, Zero-Shot, Training-Free Text-to-Video Generation, *Diljeet Jagpal, Xi Chen, Vinay P. Namboodiri*
- 232 TransPixeler: Advancing Text-to-Video Generation with Transparency, *Luozhou Wang, Yijun Li, Zhifei Chen, Jui-Hsien Wang, Zhifei Zhang, He Zhang, Zhe Lin, Ying-Cong Chen*
- 233 PTDiffusion: Free Lunch for Generating Optical Illusion Hidden Pictures with Phase-Transferred Diffusion Model, *Xiang Gao, Shuai Yang, Jiaying Liu*
- 234 Difference Inversion: Interpolate and Isolate the Difference with Token Consistency for Image Analogy Generation, *Hyunsoo Kim, Donghyun Kim, Suhyun Kim*
- 235 StyleSSP: Sampling StartPoint Enhancement for Training-free Diffusion-based Method for Style Transfer, *Ruojun Xu, Weijie Xi, XiaoDi Wang, Yongbo Mao, Zach Cheng*
- 236 Attention Distillation: A Unified Approach to Visual Characteristics Transfer, *Yang Zhou, Xu Gao, Zichong Chen, Hui Huang*
- 237 Style-Editor: Text-driven Object-centric Style Editing, *Jihun Park, Jongmin Gim, Kyoungmin Lee, Seunghun Lee, Sunghoon Im*
- 238 Towards Scalable Human-aligned Benchmark for Text-guided Image Editing, *Suho Ryu, Kihyun Kim, Eugene Baek, Dongsoo Shin, Joonseok Lee*
- 239 PS-Diffusion: Photorealistic Subject-Driven Image Editing with Disentangled Control and Attention, *Weicheng Wang, Guoli Jia, Zhongqi Zhang, Liang Lin, Jufeng Yang*
- 240 Paint by Inpaint: Learning to Add Image Objects by Removing Them First, *Navve Wasserman, Noam Rotstein, Roy Ganz, Ron Kimmel*
- 241 MTADiffusion: Mask Text Alignment Diffusion Model for Object Inpainting, *Jun Huang, Ting Liu, Yihang Wu, Xiaochao Qu, Luoqi Liu, Xiaolin Hu*
- 242 ATA: Adaptive Transformation Agent for Text-Guided Subject-Position Variable Background Inpainting, *Yizhe Tang, Zhimin Sun, Yuzhen Du, Ran Yi, Guangben Lu, Teng Hu, Luying Li, Lizhuang Ma, Fangyuan Zou*
- 243 Unleashing In-context Learning of Autoregressive Models for Few-shot Image Manipulation, *Bolin Lai, Felix Juefei-Xu, Miao Liu, Xiaoliang Dai, Nikhil Mehta, Chenguang Zhu, Zeyi Huang, James M. Rehg, Sangmin Lee, Ning Zhang, Tong Xiao*
- 244 Image is All You Need to Empower Large-scale Diffusion Models for In-Domain Generation, *Pu Cao, Feng Zhou, Lu Yang, Tianrui Huang, Qing Song*

- 245 PatchDPO: Patch-level DPO for Finetuning-free Personalized Image Generation, *Qihan Huang, Long Chan, Jinlong Liu, Wanggui He, Hao Jiang, Mingli Song, Jie Song*
- 246 VODiff: Controlling Object Visibility Order in Text-to-Image Generation, *Dong Liang, Jinyuan Jia, Yuhao Liu, Zhanghan Ke, Hongbo Fu, Rynson W. H. Lau*
- 247 Z-Magic: Zero-shot Multiple Attributes Guided Image Creator, *Yingying Deng, Xiangyu He, Fan Tang, Weiming Dong*
- 248 Infinity ∞ : Scaling Bitwise AutoRegressive Modeling for High-Resolution Image Synthesis, *Jian Han, Jinlai Liu, Yi Jiang, Bin Yan, Yuqi Zhang, Zehuan Yuan, Bingyue Peng, Xiaobing Liu*
- 249 Spatial Transport Optimization by Repositioning Attention Map for Training-Free Text-to-Image Synthesis, *Woojung Han, Yeonkyung Lee, Chanyoung Kim, Kwanghyun Park, Seong Jae Hwang*
- 250 Exploring Sparse MoE in GANs for Text-conditioned Image Synthesis, *Jiapeng Zhu, Ceyuan Yang, Kecheng Zheng, Yinghao Xu, Zifan Shi, Yifei Zhang, Qifeng Chen, Yujun Shen*
- 251 DreamRelation: Bridging Customization and Relation Generation, *Qingyu Shi, Lu Qi, Jianzong Wu, Jinbin Bai, Jingbo Wang, Yunhai Tong, Xiangtai Li*
- 252 Language-Guided Image Tokenization for Generation, *Kaiwen Zha, Lijun Yu, Alireza Fathi, David A. Ross, Cordelia Schmid, Dina Katabi, Xiuye Gu*
- 253 Scaling Down Text Encoders of Text-to-Image Diffusion Models, *Lifu Wang, Daqing Liu, Xincheng Liu, Xiaodong He*
- 254 Diffusion Self-Distillation for Zero-Shot Customized Image Generation, *Shengqu Cai, Eric Ryan Chan, Yunzhi Zhang, Leonidas Guibas, Jiajun Wu, Gordon Wetzstein*
- 255 Redefining <Creative> in Dictionary: Towards an Enhanced Semantic Understanding of Creative Generation, *Fu Feng, Yucheng Xie, Xu Yang, Jing Wang, Xin Geng*
- 256 Towards Transformer-Based Aligned Generation with Self-Coherence Guidance, *Shulei Wang, Wang Lin, Hai Huang, Hanting Wang, Sihang Cai, WenKang Han, Tao Jin, Jingyuan Chen, Jiacheng Sun, Jieming Zhu, Zhou Zhao*
- 257 Calibrated Multi-Preference Optimization for Aligning Diffusion Models, *Kyungmin Lee, Xiahong Li, Qifei Wang, Junfeng He, Junjie Ke, Ming-Hsuan Yang, Irfan Essa, Jinwoo Shin, Feng Yang, Yinxiao Li*
- 258 A4A: Adapter for Adapter Transfer via All-for-All Mapping for Cross-Architecture Models, *Keyu Tu, Mengqi Huang, Zhuowei Chen, Zhendong Mao*
- 259 Focus-N-Fix: Region-Aware Fine-Tuning for Text-to-Image Generation, *Xiaoying Xing, Avinab Saha, Junfeng He, Susan Hao, Paul Vicol, Moonkyung Ryu, Gang Li, Sahil Singla, Sarah Young, Yinxiao Li, Feng Yang, Deepak Ramachandran*
- 260 SILMM: Self-Improving Large Multimodal Models for Compositional Text-to-Image Generation, *Leigang Qu, Haochuan Li, Wenjie Wang, Xiang Liu, Juncheng Li, Liqiang Nie, Tat-Seng Chua*
- 261 CompGS: Unleashing 2D Compositionality for Compositional Text-to-3D via Dynamically Optimizing 3D Gaussians, *Chongjian Ge, Chenfeng Xu, Yuanfeng Ji, Chensheng Peng, Masayoshi Tomizuka, Ping Luo, Mingyu Ding, Varun Jampani, Wei Zhan*
- 262 Apply Hierarchical-Chain-of-Generation to Complex Attributes Text-to-3D Generation, *Yiming Qin, Zhu Xu, Yang Liu*
- 263 Empowering Vector Graphics with Consistently Arbitrary Viewing and View-dependent Visibility, *Yidi Li, Jun Xiao, Zhengda Lu, Yiqun Wang, Haiyong Jiang*
- 264 IDEA-Bench: How Far are Generative Models from Professional Designing?, *Chen Liang, Lianghua Huang, Jingwu Fang, Huanzhang Dou, Wei Wang, Zhi-Fan Wu, Yupeng Shi, Junge Zhang, Xin Zhao, Yu Liu*
- 265 Gen3DEval: Using vLLMs for Automatic Evaluation of Generated 3D Objects, *Shalini Maiti, Lourdes Agapito, Filippas Kokkinos*
- 266 CAD-Llama: Leveraging Large Language Models for Computer-Aided Design Parametric 3D Model Generation, *Jiahao Li, Weijian Ma, Xueyang Li, Yunzhong Lou, Guichun Zhou, Xiangdong Zhou*
- 267 BlenderGym: Benchmarking Foundational Model Systems for Graphics Editing, *Yunqi Gu, Ian Huang, Jihyeon Je, Guandao Yang, Leonidas Guibas*
- 268 Adversarial Domain Prompt Tuning and Generation for Single Domain Generalization, *Zhipeng Xu, De Cheng, Xinyang Jiang, Nannan Wang, Dongsheng Li, Xinbo Gao*
- 269 Localized Concept Erasure for Text-to-Image Diffusion Models Using Training-Free Gated Low-Rank Adaptation, *Byung Hyun Lee, Sungjin Lim, Se Young Chun*
- 270 Random Conditioning with Distillation for Data-Efficient Diffusion Model Compression, *Dohyun Kim, Sehwan Park, Geonhee Han, Seung Wook Kim, Paul Hongsuck Seo*
- 271 Efficient Fine-Tuning and Concept Suppression for Pruned Diffusion Models, *Reza Shirkavand, Peiran Yu, Shangqian Gao, Gowthami Somepalli, Tom Goldstein, Heng Huang*
- 272 Visual Persona: Foundation Model for Full-Body Human Customization, *Jisu Nam, Soowon Son, Zhan Xu, Jing Shi, Difan Liu, Feng Liu, Seungryong Kim, Yang Zhou*
- 273 The Art of Deception: Color Visual Illusions and Diffusion Models, *Alexandra Gomez-Villa, Kai Wang, C.Alejandro Parraga, Bartomiej Twardowski, Jesus Malo, Javier Vazquez-Corral, Joost van den Weijer*
- 274 Harnessing Frequency Spectrum Insights for Image Copyright Protection Against Diffusion Models, *Zhenguang Liu, Chao Shuai, Shaojing Fan, Ziping Dong, Jinwu Hu, Zhongjie Ba, Kui Ren*
- 275 Hiding Images in Diffusion Models by Editing Learned Score Functions, *Haoyu Chen, Yunqiao Yang, Nan Zhong, Kede Ma*
- 276 CDI: Copyrighted Data Identification in Diffusion Models, *Jan Dubiński, Antoni Kowalczyk, Franziska Boenisch, Adam Dziedzic*
- 277 A Bias-Free Training Paradigm for More General AI-generated Image Detection, *Fabrizio Guillaro, Giada Zingarini, Ben Usman, Avneesh Sud, Davide Cozzolino, Luisa Verdoliva*
- 278 Task Singular Vectors: Reducing Task Interference in Model Merging, *Antonio Andrea Gargiulo, Donato Crisostomi, Maria Sofia Bucarelli, Simone Scardapane, Fabrizio Silvestri, Emanuele Rodolà*
- 279 Any-Resolution AI-Generated Image Detection by Spectral Learning, *Dimitrios Karageorgiou, Symeon Papadopoulos, Ioannis Kompatsiaris, Efstratios Gavves*
- 280 DefectFill: Realistic Defect Generation with Inpainting Diffusion Model for Visual Inspection, *Jaewoo Song, Daemin Park, Kanghyun Baek, Sangyub Lee, Jooyoung Choi, Eunji Kim, Sungroh Yoon*
- 281 End-to-End Implicit Neural Representations for Classification, *Alexander Gielisse, Jan van Gemert*
- 282 A Flag Decomposition for Hierarchical Datasets, *Nathan Mankovich, Ignacio Santamaria, Gustau Camps-Valls, Tolga Birdal*
- 283 GazeGene: Large-scale Synthetic Gaze Dataset with 3D Eyeball Annotations, *Yiwei Bao, Zhiming Wang, Feng Lu*
- 284 FIFA: Fine-grained Inter-frame Attention for Driver's Video Gaze Estimation, *Daosong Hu, Mingyue Cui, Kai Huang*
- 285 Video-Guided Foley Sound Generation with Multimodal Controls, *Ziyang Chen, Prem Seetharaman, Bryan Russell, Oriol Nieto, David Bourgin, Andrew Owens, Justin Salamon*
- 286 VidMuse: A Simple Video-to-Music Generation Framework with Long-Short-Term Modeling, *Zeyue Tian, Zhaoyang Liu, Ruibin Yuan, Jiahao Pan, Qifeng Liu, Xu Tan, Qifeng Chen, Wei Xue, Yike Guo*
- 287 CAV-MAE Sync: Improving Contrastive Audio-Visual Mask Autoencoders via Fine-Grained Alignment, *Edson Araujo, Andrew Rouditchenko, Yuan Gong, Saurabhchand Bhati, Samuel Thomas, Brian Kingsbury, Leonid Karlinsky, Rogerio Feris, James R. Glass, Hilde Kuehne*
- 288 Crab: A Unified Audio-Visual Scene Understanding Model with Explicit Cooperation, *Henghui Du, Guangyao Li, Chang Zhou, Chunjie Zhang, Alan Zhao, Di Hu*
- 289 Circumventing Shortcuts in Audio-visual Deepfake Detection Datasets with Unsupervised Learning, *Stefan Smeu, Dragos-Alexandru Boldisor, Dan Oneata, Elisabeta Oneata*
- 290 PhyT2V: LLM-Guided Iterative Self-Refinement for Physics-Grounded Text-to-Video Generation, *Qiyao Xue, Xiangyu Yin, Boyuan Yang, Wei Gao*
- 291 Mask*2DiT: Dual Mask-based Diffusion Transformer for Multi-Scene Long Video Generation, *Tianhao Qi, Jianlong Yuan, Wanquan Feng, Shancheng Fang, Jiawei Liu, Siyu Zhou, Qian He, Hongtao Xie, Yongdong Zhang*

- 292 SMTPD: A New Benchmark for Temporal Prediction of Social Media Popularity, *Yijie Xu, Bolun Zheng, Wei Zhu, Hangjia Pan, Yuchen Yao, Ning Xu, Anan Liu, Quan Zhang, Chenggang Yan*
- 293 Video-Bench: Human-Aligned Video Generation Benchmark, *Hui Han, Siyuan Li, Jiaqi Chen, Yiwen Yuan, Yuling Wu, Chak Tou Leong, Hanwen Du, Junchen Fu, Youhua Li, Jie Zhang, Chi Zhang, Li-jia Li, Yongxin Ni, Yufan Deng*
- 294 ALGV-Assessor: Benchmarking and Evaluating the Perceptual Quality of Text-to-Video Generation with LMM, *Jiarui Wang, Huiyu Duan, Guangtao Zhai, Juntong Wang, Xiongkuo Min*
- 295 AutoSSVH: Exploring Automated Frame Sampling for Efficient Self-Supervised Video Hashing, *Niu Lian, Jun Li, Jinpeng Wang, Ruisheng Luo, Yaowei Wang, Shu-Tao Xia, Bin Chen*
- 296 Apollo: An Exploration of Video Understanding in Large Multimodal Models, *Orr Zohar, Xiaohan Wang, Yann Dubois, Nikhil Mehta, Tong Xiao, Philippe Hansen-Estruch, Licheng Yu, Xiaofang Wang, Felix Juefei-Xu, Ning Zhang, Serena Yeung-Levy, Xide Xia*
- 297 OVO-Bench: How Far is Your Video-LLMs from Real-World Online Video Understanding?, *Junbo Niu, Yifei Li, Ziyang Miao, Chunjiang Ge, Yuanhang Zhou, Qihao He, Xiaoyi Dong, Haodong Duan, Shuangrui Ding, Rui Qian, Pan Zhang, Yuhang Zang, Yuhang Cao, Conghui He, Jiaqi Wang*
- 298 VELOCITI: Benchmarking Video-Language Compositional Reasoning with Strict Entailment, *Darshana Saravanan, Varun Gupta, Darshan Singh, Zeeshan Khan, Vineet Gandhi, Makarand Tapaswi*
- 299 OmniMMI: A Comprehensive Multi-modal Interaction Benchmark in Streaming Video Contexts, *Yuxuan Wang, Yueqian Wang, Bo Chen, Tong Wu, Dongyan Zhao, Zilong Zheng*
- 300 DrVideo: Document Retrieval Based Long Video Understanding, *Ziyu Ma, Chenhui Gou, Hengcan Shi, Bin Sun, Shutao Li, Hamid Rezatofighi, Jianfei Cai*
- 301 Chapter-Llama: Efficient Chaptering in Hour-Long Videos with LLMs, *Lucas Ventura, Antoine Yang, Cordelia Schmid, Gül Varol*
- 302 LongVALE: Vision-Audio-Language-Event Benchmark Towards Time-Aware Omni-Modal Perception of Long Videos, *Tiantian Geng, Jinrui Zhang, Qingni Wang, Teng Wang, Jinming Duan, Feng Zheng*
- 303 VideoRefer Suite: Advancing Spatial-Temporal Object Understanding with Video LLM, *Yuqian Yuan, Hang Zhang, Wentong Li, Zesen Cheng, Boqiang Zhang, Long Li, Xin Li, Deli Zhao, Wenqiao Zhang, Yueting Zhuang, Jianke Zhu, Lidong Bing*
- 304 Video Summarization with Large Language Models, *Min Jung Lee, Dayoung Gong, Minsu Cho*
- 305 DyCoke: Dynamic Compression of Tokens for Fast Video Large Language Models, *Keda Tao, Can Qin, Haoxuan You, Yang Sui, Huan Wang*
- 306 RoadSocial: A Diverse VideoQA Dataset and Benchmark for Road Event Understanding from Social Video Narratives, *Chirag Parikh, Deepti Rawat, Rakshitha R. T., Tathagata Ghosh, Ravi Kiran Sarvadevabhatla*
- 307 ReVisionLLM: Recursive Vision-Language Model for Temporal Grounding in Hour-Long Videos, *Tanveer Hannan, Md Mohaiminul Islam, Jindong Gu, Thomas Seidl, Gedas Bertasius*
- 308 ViCaS: A Dataset for Combining Holistic and Pixel-level Video Understanding using Captions with Grounded Segmentation, *Ali Athar, Xueqing Deng, Liang-Chieh Chen*
- 309 VideoGLaMM: A Large Multimodal Model for Pixel-Level Visual Grounding in Videos, *Shehan Munasinghe, Hanan Gani, Wenqi Zhu, Jiale Cao, Eric Xing, Fahad Shahbaz Khan, Salman Khan*
- 310 Unbiased Video Scene Graph Generation via Visual and Semantic Dual Debiasing, *YanJun Li, Zhaoyang Li, Honghui Chen, Lizhi Xu*
- 311 UniGoal: Towards Universal Zero-shot Goal-oriented Navigation, *Hang Yin, Xiuwei Xu, Linqing Zhao, Ziwei Wang, Jie Zhou, Jiwen Lu*
- 312 Semantic and Sequential Alignment for Referring Video Object Segmentation, *Feiyu Pan, Hao Fang, Fangkai Li, Yanyu Xu, Yawei Li, Luca Benini, Xiankai Lu*
- 313 SegMAN: Omni-scale Context Modeling with State Space Models and Local Attention for Semantic Segmentation, *Yunxiang Fu, Meng Lou, Yizhou Yu*
- 314 Correcting Deviations from Normality: A Reformulated Diffusion Model for Multi-Class Unsupervised Anomaly Detection, *Farzad Beizaei, Gregory A. Lodygensky, Christian Desrosiers, Jose Dolz*
- 315 EchoTraffic: Enhancing Traffic Anomaly Understanding with Audio-Visual Insights, *Zhenghao Xing, Hao Chen, Binzhu Xie, Jiaqi Xu, Ziyu Guo, Xuemiao Xu, Jianye Hao, Chi-Wing Fu, Xiaowei Hu, Pheng-Ann Heng*
- 316 Noise-Resistant Video Anomaly Detection via RGB Error-Guided Multiscale Predictive Coding and Dynamic Memory, *Han Hu, Wenli Du, Peng Liao, Bing Wang, Siyuan Fan*
- 317 Understanding Multi-Task Activities from Single-Task Videos, *Yuhan Shen, Ehsan Elhamifar*
- 318 Action Detail Matters: Refining Video Recognition with Local Action Queries, *Mengmeng Wang, Zeyi Huang, Xiangjie Kong, Guojiang Shen, Guang Dai, Jingdong Wang, Yong Liu*
- 319 CountLLM: Towards Generalizable Repetitive Action Counting via Large Language Model, *Ziyu Yao, Xuxin Cheng, Zhiqi Huang, Lei Li*
- 320 Heterogeneous Skeleton-Based Action Representation Learning, *Hongsong Wang, Xiaoyan Ma, Jidong Kuang, Jie Gui*
- 321 Dynamic Updates for Language Adaptation in Visual-Language Tracking, *Xiaohai Li, Bineng Zhong, Qihua Liang, Zhiyi Mo, Jian Nong, Shuxiang Song*
- 322 Boosting Adversarial Transferability through Augmentation in Hypothesis Space, *Yu Guo, Weiquan Liu, Qingshan Xu, Shijun Zheng, Shujun Huang, Yu Zang, Siqi Shen, Chenglu Wen, Cheng Wang*
- 323 UIBDiffusion: Universal Imperceptible Backdoor Attack for Diffusion Models, *Yuning Han, Bingyin Zhao, Rui Chu, Feng Luo, Biplob Sikdar, Yingjie Lao*
- 324 CryptoFace: End-to-End Encrypted Face Recognition, *Wei Ao, Vishnu Naresh Boddeti*
- 325 Forensics Adapter: Adapting CLIP for Generalizable Face Forgery Detection, *Xinjie Cui, Yuezun Li, Ao Luo, Jiaran Zhou, Junyu Dong*
- 326 D2SP: Dynamic Dual-Stage Purification Framework for Dual Noise Mitigation in Vision-based Affective Recognition, *Haoran Wang, Xinji Mai, Zeng Tao, Xuan Tong, Junxiong Lin, Yan Wang, Jiawen Yu, Shaoqi Yan, Ziheng Zhou, Wenqiang Zhang*
- 327 Can't Slow Me Down: Learning Robust and Hardware-Adaptive Object Detectors against Latency Attacks for Edge Devices, *Tianyi Wang, Zichen Wang, Cong Wang, Yuanhao Shu, Ruilong Deng, Peng Cheng, Jiming Chen*
- 328 Decision SpikeFormer: Spike-Driven Transformer for Decision Making, *Wei Huang, Qinying Gu, Nanyang Ye*
- 329 Identity-Clothing Similarity Modeling for Unsupervised Clothing Change Person Re-Identification, *Zhiqi Pang, Junjie Wang, Lingling Zhao, Chunyu Wang*
- 330 Cheb-GR: Rethinking K-nearest Neighbor Search in Re-ranking for Person Re-identification, *Jinxi Yang, He Li, Bo Du, Mang Ye*
- 331 Shift the Lens: Environment-Aware Unsupervised Camouflaged Object Detection, *Ji Du, Fangwei Hao, Mingyang Yu, Desheng Kong, Jiesheng Wu, Bin Wang, Jing Xu, Ping Li*
- 332 Point2RBox-v2: Rethinking Point-supervised Oriented Object Detection with Spatial Layout Among Instances, *Yi Yu, Botao Ren, Peiyuan Zhang, Mingxin Liu, Junwei Luo, Shaofeng Zhang, Feipeng Da, Junchi Yan, Xue Yang*
- 333 BOOTPLACE: Bootstrapped Object Placement with Detection Transformers, *Hang Zhou, Xinxin Zuo, Rui Ma, Li Cheng*
- 334 Minimizing Labeled, Maximizing Unlabeled: An Image-Driven Approach for Video Instance Segmentation, *Fangyun Wei, Jinjing Zhao, Kun Yan, Chang Xu*
- 335 PolarNeXt: Rethink Instance Segmentation with Polar Representation, *Jiacheng Sun, Xinghong Zhou, Yiqiang Wu, Bin Zhu, Jiaxuan Lu, Yu Qin, Xiaomao Li*
- 336 SAM2Object: Consolidating View Consistency via SAM2 for Zero-Shot 3D Instance Segmentation, *Jihuai Zhao, Junbao Zhuo, Jiansheng Chen, Huimin Ma*
- 337 COB-GS: Clear Object Boundaries in 3DGS Segmentation Based on Boundary-Adaptive Gaussian Splitting, *Jiaxin Zhang, Junjun Jiang, Youyu Chen, Kui Jiang, Xianming Liu*

- 338 DFormerv2: Geometry Self-Attention for RGBD Semantic Segmentation, *Bo-Wen Yin, Jiao-Long Cao, Ming-Ming Cheng, Qibin Hou*
- 339 SAM-REF: Introducing Image-Prompt Synergy during Interaction for Detail Enhancement in the Segment Anything Model, *Chongkai Yu, Ting Liu, Anqi Li, Xiaochao Qu, Chengjing Wu, Luoqi Liu, Xiaolin Hu*
- 340 Believing is Seeing: Unobserved Object Detection using Generative Models, *Subhransu S. Bhattacharjee, Dylan Campbell, Rahul Shome*
- 341 MM-OR: A Large Multimodal Operating Room Dataset for Semantic Understanding of High-Intensity Surgical Environments, *Ege Özsoy, Chantal Pellegrini, Tobias Czempel, Felix Tristram, Kun Yuan, David Bani-Harouni, Ulrich Eck, Benjamin Busam, Matthias Keicher, Nassir Navab*
- 342 Cross-Modal and Uncertainty-Aware Agglomeration for Open-Vocabulary 3D Scene Understanding, *Jinlong Li, Cristiano Saltori, Fabio Poiesi, Nicu Sebe*
- 343 Open-Vocabulary Functional 3D Scene Graphs for Real-World Indoor Spaces, *Chenyanguang Zhang, Alexandros Delitzas, Fangjinhua Wang, Ruida Zhang, Xiangyang Ji, Marc Pollefeys, Francis Engelmann*
- 344 SKE-Layout: Spatial Knowledge Enhanced Layout Generation with LLMs, *Junsheng Wang, Nieqing Cao, Yan Ding, Mengying Xie, Fuqiang Gu, Chao Chen*
- 345 Zero-shot 3D Question Answering via Voxel-based Dynamic Token Compression, *Hsiang-Wei Huang, Fu-Chen Chen, Wenhao Chai, Che-Chun Su, Lu Xia, Sanghun Jung, Cheng-Yen Yang, Jenq-Neng Hwang, Min Sun, Cheng-Hao Kuo*
- 346 Empowering Large Language Models with 3D Situation Awareness, *Zhihao Yuan, Yibo Peng, Jinke Ren, Yinghong Liao, Yatong Han, Chun-Mei Feng, Hengshuang Zhao, Guanbin Li, Shuguang Cui, Zhen Li*
- 347 Visual Agentic AI for Spatial Reasoning with a Dynamic API, *Damiano Marsili, Rohun Agrawal, Yisong Yue, Georgia Gkioxari*
- 348 R2C: Mapping Room to Chessboard to Unlock LLM As Low-Level Action Planner, *Ziyi Bai, Hanxuan Li, Bin Fu, Chuyan Xiong, Ruiping Wang, Xilin Chen*
- 349 GRAPHGPT-O: Synergistic Multimodal Comprehension and Generation on Graphs, *Yi Fang, Bowen Jin, Jiacheng Shen, Sirui Ding, Qiaoyu Tan, Jiawei Han*
- 350 GUI-Xplore: Empowering Generalizable GUI Agents with One Exploration, *Yuchen Sun, Shanhui Zhao, Tao Yu, Hao Wen, Samith Va, Mengwei Xu, Yuanchun Li, Chongyang Zhang*
- 351 Empowering LLMs to Understand and Generate Complex Vector Graphics, *Ximing Xing, Juncheng Hu, Guotao Liang, Jing Zhang, Dong Xu, Qian Yu*
- 352 ShowUI: One Vision-Language-Action Model for GUI Visual Agent, *Kevin Qinghong Lin, Linjie Li, Difei Gao, Zhengyuan Yang, Shiwei Wu, Zechen Bai, Stan Weixian Lei, Lijuan Wang, Mike Zheng Shou*
- 353 SocialGesture: Delving into Multi-person Gesture Understanding, *Xu Cao, Pranav Virupaksha, Wenqi Jia, Bolin Lai, Fiona Ryan, Sangmin Lee, James M. Rehg*
- 354 Interleaved-Modal Chain-of-Thought, *Jun Gao, Yongqi Li, Ziqiang Cao, Wenjie Li*
- 355 AnySat: One Earth Observation Model for Many Resolutions, Scales, and Modalities, *Guillaume Astruc, Nicolas Gonthier, Clément Mallet, Loïc Landrieu*
- 356 MV-MATH: Evaluating Multimodal Math Reasoning in Multi-Visual Contexts, *Peijie Wang, Zhong-Zhi Li, Fei Yin, Dekang Ran, Cheng-Lin Liu*
- 357 MicroVQA: A Multimodal Reasoning Benchmark for Microscopy-Based Scientific Research, *James Burgess, Jeffrey J Nirschl, Laura Bravo-Sánchez, Alejandro Lozano, Sanket Rajan Gupta, Jesus G. Galaz-Montoya, Yuhui Zhang, Yuchang Su, Disha Bhowmik, Zachary Coman, Sarina M Hasan, Alexandra Johannesson, William D. Leineweber, Malvika G Nair, Ridhi Yarlagadda, Connor Zuraski, Wah Chiu, Sarah Cohen, Jan N. Hansen, Manuel D Leonetti, Chad Liu, Emma Lundberg, Serena Yeung-Levy*
- 358 All Languages Matter: Evaluating LMMs on Culturally Diverse
* 100 Languages, *Ashmal Vayani, Dinura Dissanayake, Hasindri Watawana, Noor Ahsan, Nevasini Sasikumar, Omkar Thawakar, Henok Biadgign Ademteu, Yahya Hmaiti, Amandeep Kumar, Kartik Kukreja, Mykola Maslych, Wafa Al Ghallabi, Mihail Minkov Mihaylov, Chao Qin, Abdelrahman M. Shaker, Mike Zhang, Mahardika Krisna Ihsani, Amiel Gian Esplana, Monil Gokani, Shachar Mirkin, Harsh Singh, Ashay Srivastava, Endre Hamerlik, Fathinah Asma Izzati, Fadillah Adamsyah Maani, Sebastian Cavada, Jenny Chim, Rohit Gupta, Sanjay Manjunath, Kamila Zhumakhanova, Feno Heriniaina Rabevohitra, Azril Hafizi Amirudin, Muhammad Ridzuan, Daniya Najiha Abdul Kareem, Ketan Pravin More, Kunyang Li, Pramesh Shakya, Muhammad Saad, Amirpouya Ghasemaghahi, Amirbek Djanibekov, Dilshod Azizov, Branislava Jankovic, Naman Bhatia, Alvaro Cabrera, Johan Obando-Ceron, Olympiah Otieno, Febian Farestam, Muztoba Rabbani, Sanoojan Ballah, Santosh Sanjeev, Abduragim Shtanchaev, Maheen Fatima, Thao Nguyen, Amrin Kareem, Toluwani Aremu, Nathan Augusto Zacarias Xavier, Amit Bhatkal, Hawau Olamide Toyin, Aman Chadha, Hisham Cholakkal, Rao Muhammad Anwer, Michael Felsberg, Jorma Laaksonen, Tamar Solorio, Monojit Choudhury, Ivan Laptev, Mubarak Shah, Salman Khan, Fahad Shahbaz Khan*
- 359 Towards General Visual-Linguistic Face Forgery Detection, *Ke Sun, Shen Chen, Taiping Yao, Ziyin Zhou, Jiayi Ji, Xiaoshuai Sun, Chia-Wen Lin, Rongrong Ji*
- 360 Exploring Contextual Attribute Density in Referring Expression Counting, *Zhicheng Wang, Zhiyu Pan, Zhan Peng, Jian Cheng, Liwen Xiao, Wei Jiang, Zhiguo Cao*
- 361 Notes-guided MLLM Reasoning: Enhancing MLLM with Knowledge and Visual Notes for Visual Question Answering, *Wenlong Fang, Qiaofeng Wu, Jing Chen, Yun Xue*
- 362 CL-MoE: Enhancing Multimodal Large Language Model with Dual Momentum Mixture-of-Experts for Continual Visual Question Answering, *Tianyu Huai, Jie Zhou, Xingjiao Wu, Qin Chen, Qingchun Bai, Ze Zhou, Liang He*
- 363 Benchmarking Large Vision-Language Models via Directed Scene Graph for Comprehensive Image Captioning, *Fan Lu, Wei Wu, Kecheng Zheng, Shuailei Ma, Biao Gong, Jiawei Liu, Wei Zhai, Yang Cao, Yujun Shen, Zheng-Jun Zha*
- 364 Learning with Noisy Triplet Correspondence for Composed Image Retrieval, *Shuxian Li, Changhao He, Xiting Liu, Joey Tianyi Zhou, Xi Peng, Peng Hu*
- 365 ConText-CIR: Learning from Concepts in Text for Composed Image Retrieval, *Eric Xing, Pranavi Kolouju, Robert Pless, Abby Stylianou, Nathan Jacobs*
- 366 PromptHash: Affinity-Prompted Collaborative Cross-Modal Learning for Adaptive Hashing Retrieval, *Qiang Zou, Shuli Cheng, Jiayi Chen*
- 367 GENIUS: A Generative Framework for Universal Multimodal Search, *Sungyeon Kim, Xinliang Zhu, Xiaofan Lin, Muhammet Bastan, Douglas Gray, Suha Kwak*
- 368 Font-Agent: Enhancing Font Understanding with Large Language Models, *Yingxin Lai, Cuijie Xu, Haitian Shi, Guoqing Yang, Xiaoning Li, Zhiming Luo, Shaozi Li*
- 369 Image Over Text: Transforming Formula Recognition Evaluation with Character Detection Matching, *Bin Wang, Fan Wu, Linke Ouyang, Zhuangcheng Gu, Rui Zhang, Renqiu Xia, Botian Shi, Bo Zhang, Conghui He*
- 370 Video-CoBERT: Contextualized Late Interaction for Text-to-Video Retrieval, *Arun Reddy, Alexander Martin, Eugene Yang, Andrew Yates, Kate Sanders, Kenton Murray, Reno Kriz, Celso M. de Melo, Benjamin Van Durme, Rama Chellappa*
- 371 Reconstruction vs. Generation: Taming Optimization Dilemma in Latent Diffusion Models, *Jingfeng Yao, Bin Yang, Xinggang Wang*
- 372 DiscoVLA: Discrepancy Reduction in Vision, Language, and Alignment for Parameter-Efficient Video-Text Retrieval, *Leqi Shen, Guoqiang Gong, Tianxiang Hao, Tao He, Yifeng Zhang, Pengzhang Liu, Sicheng Zhao, Jungong Han, Guiguang Ding*

- 373 MergeVQ: A Unified Framework for Visual Generation and Representation with Disentangled Token Merging and Quantization, *Siyuan Li, Luyuan Zhang, Zedong Wang, Juanxi Tian, Cheng Tan, Zicheng Liu, Chang Yu, Qingsong Xie, Haonan Lu, Haoqian Wang, Zhen Lei*
- 374 BIOMEDICA: An Open Biomedical Image-Caption Archive, Dataset, and Vision-Language Models Derived from Scientific Literature, *Alejandro Lozano, Min Woo Sun, James Burgess, Liangyu Chen, Jeffrey J. Nirschl, Jeffrey Gu, Ivan Lopez, Josiah Aklilu, Anita Rau, Austin Wolfgang Katzer, Yuhui Zhang, Collin Chiu, Xiaohan Wang, Alfred Seunghoon Song, Robert Tibshirani, Serena Yeung-Levy*
- 375 Visual Lexicon: Rich Image Features in Language Space, *XuDong Wang, Xingyi Zhou, Alireza Fathi, Trevor Darrell, Cordelia Schmid*
- 376 Improving Personalized Search with Regularized Low-Rank
* Parameter Updates, *Fiona Ryan, Josef Sivic, Fabian Caba Heilbron, Judy Hoffman, James M. Rehg, Bryan Russell*
- 377 AdaDARE-gamma: Balancing Stability and Plasticity in Multi-modal LLMs through Efficient Adaptation, *Jingyi Xie, Jintao Yang, Zhunchen Luo, Yunbo Cao, Qiang Gao, Mengyuan Zhang, Wenpeng Hu*
- 378 FastVLM: Efficient Vision Encoding for Vision Language Models, *Pavan Kumar Anasosalu Vasu, Fartash Faghri, Chun-Liang Li, Cem Koc, Nate True, Albert Antony, Gokula Santhanam, James Gabriel, Peter Grasch, Oncel Tuzel, Hadi Pouransari*
- 379 Cross-modal Information Flow in Multimodal Large Language Models, *Zhi Zhang, Srishti Yadav, Fengze Han, Ekaterina Shutova*
- 380 VisionZip: Longer is Better but Not Necessary in Vision Language Models, *Senqiao Yang, Yukang Chen, Zhuotao Tian, Chengyao Wang, Jingyao Li, Bei Yu, Jiaya Jia*
- 381 TopV: Compatible Token Pruning with Inference Time Optimization for Fast and Low-Memory Multimodal Vision Language Model, *Cheng Yang, Yang Sui, Jinqi Xiao, Lingyi Huang, Yu Gong, Chendi Li, Jinghua Yan, Yu Bai, Ponnuswamy Sadayappan, Xia Hu, Bo Yuan*
- 382 A Stitch in Time Saves Nine: Small VLM is a Precise Guidance for Accelerating Large VLMs, *Wangbo Zhao, Yizeng Han, Jiasheng Tang, Zhikai Li, Yibing Song, Kai Wang, Zhangyang Wang, Yang You*
- 383 Escaping Plato's Cave: Towards the Alignment of 3D and Text Latent Spaces, *Souhail Hadgi, Luca Moschella, Andrea Santilli, Diego Gomez, Qixing Huang, Emanuele Rodolà, Simone Melzi, Maks Ovsjanikov*
- 384 ODE: Open-Set Evaluation of Hallucinations in Multimodal Large Language Models, *Yahan Tu, Rui Hu, Jitao Sang*
- 385 MoVE-KD: Knowledge Distillation for VLMs with Mixture of Visual Encoders, *Jiajun Cao, Yuan Zhang, Tao Huang, Ming Lu, Qizhe Zhang, Ruichuan An, Ningning Ma, Shanghang Zhang*
- 386 PhD: A ChatGPT-Prompted Visual Hallucination Evaluation Dataset,
* Jiazhen Liu, Yuhan Fu, Ruobing Xie, Runquan Xie, Xingwu Sun, Fengzong Lian, Zhanhui Kang, Xirong Li
- 387 SPA-VL: A Comprehensive Safety Preference Alignment Dataset for Vision Language Models, *Yongting Zhang, Lu Chen, Guodong Zheng, Yifeng Gao, Rui Zheng, Jinlan Fu, Zhenfei Yin, Senjie Jin, Yu Qiao, Xuanjing Huang, Feng Zhao, Tao Gui, Jing Shao*
- 388 Do We Really Need Curated Malicious Data for Safety Alignment in Multi-modal Large Language Models?, *Yanbo Wang, Jiyang Guan, Jian Liang, Ran He*
- 389 Exploring Visual Vulnerabilities via Multi-Loss Adversarial Search for Jailbreaking Vision-Language Models, *Shuyang Hao, Bryan Hooi, Jun Liu, Kai-Wei Chang, Zi Huang, Yujun Cai*
- 390 Anyattack: Towards Large-scale Self-supervised Adversarial Attacks on Vision-language Models, *Jiaming Zhang, Junhong Ye, Xingjun Ma, Yige Li, Yunfan Yang, Yunhao Chen, Jitao Sang, Dit-Yan Yeung*
- 391 TAPT: Test-Time Adversarial Prompt Tuning for Robust Inference in Vision-Language Models, *Xin Wang, Kai Chen, Jiaming Zhang, Jingjing Chen, Xingjun Ma*
- 392 On the Zero-shot Adversarial Robustness of Vision-Language Models: A Truly Zero-shot and Training-free Approach, *Baoshun Tong, Hanjiang Lai, Yan Pan, Jian Yin*
- 393 Conformal Prediction for Zero-Shot Models, *Julio Silva-Rodríguez, ☆ Ismail Ben Ayed, Jose Dolz*
- 394 O-TPT: Orthogonality Constraints for Calibrating Test-time
* Prompt Tuning in Vision-Language Models, *Ashshak Sharifdeen, ☆ †Muhammad Akhtar Munir, Sanojan Baliah, Salman Khan, Muhammad Haris Khan*
- 395 Auto Cherry-Picker: Learning from High-quality Generative Data Driven by Language, *Yicheng Chen, Xiangtai Li, Yining Li, Yanhong Zeng, Jianzong Wu, Xiangyu Zhao, Kai Chen*
- 396 Navigation World Models, *Amir Bar, Gaoyue Zhou, Danny Tran, Trevor Darrell, Yann LeCun*
- 397 NLPrompt: Noise-Label Prompt Learning for Vision-Language
* Models, *Bikang Pan, Qun Li, Xiaoying Tang, Wei Huang, Zhen Fang, Feng Liu, Jingya Wang, Jingyi Yu, Ye Shi*
- 398 Preserving Clusters in Prompt Learning for Unsupervised Domain Adaptation, *Tung-Long Vuong, Hoang Phan, Vy Vo, Anh Bui, Thanh-Toan Do, Trung Le, Dinh Phung*
- 399 RLAI-F: Open-Source AI Feedback Leads to Super GPT-4V
* Trustworthiness, *Tianyu Yu, Haoye Zhang, Qiming Li, Qixin Xu, Yuan Yao, Da Chen, Xiaoman Lu, Ganqu Cui, Yunkai Dang, Taiwen He, Xiaocheng Feng, Jun Song, Bo Zheng, Zhiyuan Liu, T at-Seng Chua, Maosong Sun*
- 400 Test-Time Visual In-Context Tuning, *Jiahao Xie, Alessio Tonioni, ☆ Nathalie Rauschmayr, Federico Tombari, Bernt Schiele*
- 401 F³OCUS - Federated Finetuning of Vision-Language Foundation
* Models with Optimal Client Layer Updating Strategy via Multi-objective Meta-Heuristics, *Pramit Saha, Felix Wagner, Divyanshu Mishra, Can Peng, Anshul Thakur, David A. Clifton, Konstantinos Kamnitsas, J. Alison Noble*
- 402 Towards Human-Understandable Multi-Dimensional Concept Discovery, *Arne Grobrügge, Niklas Kühl, Gerhard Satzger, Philipp Spitzer*
- 403 From Prototypes to General Distributions: An Efficient Curriculum for Masked Image Modeling, *Jinhong Lin, Cheng-En Wu, Huanran Li, Jifan Zhang, Yu Hen Hu, Pedro Morgado*
- 404 Do Computer Vision Foundation Models Learn the Low-level
* Characteristics of the Human Visual System?, *Yancheng Cai, Fei Yin, Dounia Hammou, Rafal Mantiuk*
- 405 DepthCues: Evaluating Monocular Depth Perception in Large Vision Models, *Duolikin Danier, Mehmet Aygün, Changjian Li, Hakan Bilen, Oisin Mac Aodha*
- 406 LaVin-DiT: Large Vision Diffusion Transformer, *Zhaoqing Wang, Xiaobo Xia, Runnan Chen, Dongdong Yu, Changhu Wang, Mingming Gong, Tongliang Liu*
- 407 5%>100%: Breaking Performance Shackles of Full Fine-Tuning on Visual Recognition Tasks, *Dongshuo Yin, Leiyi Hu, Bin Li, Youqun Zhang, Xue Yang*
- 408 Efficient Data Driven Mixture-of-Expert Extraction from Trained Networks, *Uranik Berisha, Jens Mehnert, Alexandru Paul Condurache*
- 409 Split Adaptation for Pre-trained Vision Transformers, *Lixu Wang, Bingqi Shang, Yi Li, Payal Mohapatra, Wei Dong, Xiao Wang, Qi Zhu*
- 410 Your Scale Factors are My Weapon: Targeted Bit-Flip Attacks on Vision Transformers via Scale Factor Manipulation, *Jialai Wang, Yuxiao Wu, Weiye Xu, Yating Huang, Chao Zhang, Zongpeng Li, Mingwei Xu, Zhenkai Liang*
- 411 MDP: Multidimensional Vision Model Pruning with Latency Constraint, *Xinglong Sun, Barath Lakshmanan, Maying Shen, Shiyi Lan, Jingde Chen, Jose M. Alvarez*
- 412 Mamba-Adaptor: State Space Model Adaptor for Visual Recognition, *Fei Xie, Jiahao Nie, Yujin Tang, Wenkang Zhang, Hongshen Zhao*
- 413 CARE Transformer: Mobile-Friendly Linear Visual Transformer via
* Decoupled Dual Interaction, *Yuan Zhou, Qingshan Xu, ☆ Jiequan Cui, Junbao Zhou, Jing Zhang, Richang Hong, Hanwang Zhang*
- 414 Coeff-Tuning: A Graph Filter Subspace View for Tuning Attention-Based Large Models, *Zichen Miao, Wei Chen, Qiang Qiu*
- 415 DVHGN: Multi-Scale Dilated Vision HGNN for Efficient Vision Recognition, *Caoshuo Li, Tanzhe Li, Xiaobin Hu, Donghao Luo, Taisong Jin*
- 416 Graph-Embedded Structure-Aware Perceptual Hashing for Neural Network Protection and Piracy Detection, *Ruiheng Liu, Haozhe Chen, Boyao Zhao, Kejiang Chen, Weiming Zhang*

- 417 Hybrid Concept Bottleneck Models, *Yang Liu, Tianwei Zhang, Shi Gu*
 418 Locality-Aware Zero-Shot Human-Object Interaction Detection, *Sanghyun Kim, Deunsol Jung, Minsu Cho*
 419 UNICL-SAM: Uncertainty-Driven In-Context Segmentation with Part Prototype Discovery, *Dianmo Sheng, Dongdong Chen, Zhentao Tan, Qiankun Liu, Qi Chu, Tao Gong, Bin Liu, Jing Han, Wenbin Tu, Shengwei Xu, Nenghai Yu*
 420 Dual Semantic Guidance for Open Vocabulary Semantic Segmentation, *Zhengyang Wang, Tingliang Feng, Fan Lyu, Fanhua Shang, Wei Feng, Liang Wan*
 421 Exploring CLIP's Dense Knowledge for Weakly Supervised Semantic Segmentation, *Zhiwei Yang, Yucong Meng, Kexue Fu, Feilong Tang, Shuo Wang, Zhijian Song*
 422 Improving Semi-Supervised Semantic Segmentation with Sliced-Wasserstein Feature Alignment and Uniformity, *Chen-Yi Lu, Kasra Derakhshandeh, Somali Chaterji*
 423 Soft Self-labeling and Potts Relaxations for Weakly-supervised Segmentation, *Zhongwen Zhang, Yuri Boykov*
 424 Towards Efficient Foundation Model for Zero-shot Amodal Segmentation, *Zhaochen Liu, Limeng Qiao, Xiangxiang Chu, Lin Ma, Tingting Jiang*
 425 Generalizable Object Keypoint Localization from Generative Priors, *Dongkai Wang, Jiang Duan, Liangjian Wen, Shiyu Xuan, Hao Chen, Shiliang Zhang*
 426 Visual and Semantic Prompt Collaboration for Generalized Zero-Shot Learning, *Huajie Jiang, Zhengxian Li, Xiaohan Yu, Yongli Hu, Baocai Yin, Jian Yang, Yuankai Qi*
 427 Generalized Zero-Shot Classification via Semantics-Free Inter-Class Feature Generation, *Libiao Chen, Dong Nie, Junjun Pan, Jing Yan, Zhenyu Tang*
 428 GET: Unlocking the Multi-modal Potential of CLIP for Generalized
 ☆ Category Discovery, *Enguang Wang, Zhimao Peng, Zhengyuan Xie, Fei Yang, Xialei Liu, Ming-Ming Cheng*
 429 v-CLR: View-Consistent Learning for Open-World Instance
 * Segmentation, *Chang-Bin Zhang, Jinhong Ni, Yujie Zhong, Kai Han*
 430 Detecting Open World Objects via Partial Attribute Assignment, *Muli Yang, Gabriel James Goenawan, Huaiyuan Qin, Kai Han, Xi Peng, Yanhua Yang, Hongyuan Zhu*
 431 Uncertainty Meets Diversity: A Comprehensive Active Learning Framework for Indoor 3D Object Detection, *Jiangyi Wang, Na Zhao*
 432 Revisiting Generative Replay for Class Incremental Object Detection, *Shizhou Zhang, Xueqiang Lv, Yinghui Xing, Qirui Wu, Di Xu, Yanning Zhang*
 433 Decoupled Distillation to Erase: A General Unlearning Method
 * for Any Class-centric Tasks, *Yu Zhou, Dian Zheng, Qijie Mo, Renjie Lu, Kun-Yu Lin, Wei-Shi Zheng*
 434 Saliutl: Ensemble Saliency Guided Recovery of Adversarial Patches against CNNs, *Mauricio Byrd Victorica, György Dán, Henrik Sandberg*
 435 Towards Zero-Shot Anomaly Detection and Reasoning with
 * Multimodal Large Language Models, *Jiacong Xu, Shao-Yuan Lo, †Bardia Safaei, Vishal M. Patel, Isht Dwivedi*
 436 PatchGuard: Adversarially Robust Anomaly Detection and Localization through Vision Transformers and Pseudo Anomalies, *Mojtaba Nafez, Amirhossein Koochakian, Arad Maleki, Jafar Habibi, Mohammad Hossein Rohban*
 437 Odd-One-Out: Anomaly Detection by Comparing with Neighbors, *Ankan Bhunia, Changjian Li, Hakan Bilen*
 438 Dinomaly: The Less Is More Philosophy in Multi-Class Unsupervised Anomaly Detection, *Jia Guo, Shuai Lu, Weihang Zhang, Fang Chen, Huiqi Li, Hongen Liao*
 439 Distribution Prototype Diffusion Learning for Open-set Supervised Anomaly Detection, *Fuyun Wang, Tong Zhang, Yuanzhi Wang, Yide Qiu, Xin Liu, Xu Guo, Zhen Cui*
 440 ATP: Adaptive Threshold Pruning for Efficient Data Encoding in Quantum Neural Networks, *Mohamed Afane, Gabrielle Ebbrecht, Ying Wang, Juntao Chen, Junaid Farooq*
 441 Curriculum Coarse-to-Fine Selection for High-IPC Dataset Distillation, *Yanda Chen, Gongwei Chen, Miao Zhang, Weili Guan, Liqiang Nie*
 442 Masking meets Supervision: A Strong Learning Alliance, *Byeongho Heo, Taekyung Kim, Sangdoo Yun, Dongyoon Han*
 443 Scale Efficient Training for Large Datasets, *Qing Zhou, Junyu Gao, Qi Wang*
 444 Learning on Model Weights using Tree Experts, *Eliahu Horwitz, Bar Cavia, Jonathan Kahana, Yedid Hoshen*
 445 How to Merge Your Multimodal Models Over Time?,
 ☆ *Sebastian Dziadzio, Vishaal Udandara, Karsten Roth, Ameya Prabhu, Zeynep Akata, Samuel Albanie, Matthias Bethge*
 446 Revisiting Fairness in Multitask Learning: A Performance-Driven Approach for Variance Reduction, *Xiaohan Qin, Xiaoxing Wang, Junchi Yan*
 447 Enhancing Online Continual Learning with Plug-and-Play State Space Model and Class-Conditional Mixture of Discretization, *Sihao Liu, Yibo Yang, Xiaojie Li, David A. Clifton, Bernard Ghanem*
 448 Online Task-Free Continual Learning via Dynamic Expandable Memory Distribution, *Fei Ye, Adrian G. Bors*
 449 Knowledge Memorization and Rumination for Pre-trained Model-based Class-Incremental Learning, *Zijian Gao, Wangwang Jia, Xingxing Zhang, Dulan Zhou, Kele Xu, Feng Dawei, Yong Dou, Xinjun Mao, Huaimin Wang*
 450 COBRA: COmBinatorial Retrieval Augmentation for Few-Shot Adaptation, *Arnav M. Das, Gantavya Bhatt, Lilly Kumari, Sahil Verma, Jeff Bilmes*
 451 Dual Consolidation for Pre-Trained Model-Based Domain-Incremental Learning, *Da-Wei Zhou, Zi-Wen Cai, Han-Jia Ye, Lijun Zhang, De-Chuan Zhan*
 452 Gradient-Guided Annealing for Domain Generalization,
 * *Aristotelis Ballas, Christos Diou*
 453 AdMiT: Adaptive Multi-Source Tuning in Dynamic Environments, *Xiangyu Chang, Fahim Faisal Niloy, Sk Miraj Ahmed, Srikanth V. Krishnamurthy, Basak Guler, Ananthram Swami, Samet Oymak, Amit Roy-Chowdhury*
 454 Compositional Targeted Multi-Label Universal Perturbations, *Hassan Mahmood, Ehsan Elhamifar*
 455 Forming Auxiliary High-confident Instance-level Loss to Promote Learning from Label Proportions, *Tianhao Ma, Han Chen, Juncheng Hu, Yungang Zhu, Ximing Li*
 456 Dynamic Pseudo Labeling via Gradient Cutting for High-Low
 ☆ Entropy Exploration, *Jae Hyeon Park, Joo Hyeon Jeon, Jae Yun Lee, Sangyeon Ahn, Min Hee Cha, Min Geol Kim, Hyeok Nam, Sung In Cho*
 457 ProHOC: Probabilistic Hierarchical Out-of-Distribution Classification via Multi-Depth Networks, *Erik Wallin, Fredrik Kahl, Lars Hammarstrand*
 458 Test-time Augmentation Improves Efficiency in Conformal Prediction, *Divya Shanmugam, Helen Lu, Swami Sankaranarayanan, John Guttag*
 459 Subspace Constraint and Contribution Estimation for Heterogeneous Federated Learning, *Xiangtao Zhang, Sheng Li, Ao Li, Yipeng Liu, Fan Zhang, Ce Zhu, Le Zhang*
 460 FedMIA: An Effective Membership Inference Attack Exploiting "All for One" Principle in Federated Learning, *Gongxi Zhu, Donghao Li, Hanlin Gu, Yuan Yao, Lixin Fan, Yuxing Han*
 461 Detecting Backdoor Attacks in Federated Learning via Direction
 * Alignment Inspection, *Jiahao Xu, Zikai Zhang, Rui Hu*
 462 RAEncoder: A Label-Free Reversible Adversarial Examples Encoder for Dataset Intellectual Property Protection, *Fan Xing, Zhuo Tian, Xuefeng Fan, Xiaoyi Zhou*
 463 DeDe: Detecting Backdoor Samples for SSL Encoders via Decoders, *Sizai Hou, Songze Li, Duanyi Yao*
 464 Enhancing Adversarial Transferability with Checkpoints of a Single Model's Training, *Shixin Li, Chaoxiang He, Xiaojing Ma, Bin Benjamin Zhu, Shuo Wang, Hongsheng Hu, Dongmei Zhang, Linchen Yu*
 465 Tightening Robustness Verification of MaxPool-based Neural Networks via Minimizing the Over-Approximation Zone, *Yuan Xiao, Yuchen Chen, Shiqing Ma, Chunrong Fang, Tongtong Bai, Mingzheng Gu, Yuxin Cheng, Yanwei Chen, Zhenyu Chen*
 466 Theory-Inspired Deep Multi-View Multi-Label Learning with Incomplete Views and Noisy Labels, *Quanjiang Li, Tingjin Luo, Jiahui Liao*

- Mariya Hendriksen, Abdelhak Lemkhenter, Shanzheng Tan, Linda Wen, Katja Hofmann, Sarah Parisot*
- 6 *Live freeway traffic state super-resolution, Junyi Ji, Alex Richardson, Derek Gloudemans, Gergely Zachár, Matthew Nice, William Barbour, Jonathan Sprinkle, Benedetto Piccoli, Daniel B. Work*
- 7 *Efficient Segmentation for Edge Devices, Xin Li, Shuai Zhang*
- 8 *GenEx: Generating an explorable world, TaiMing Lu, Jieneng Chen*
- 9 *VIZ: Virtual and Physical Navigation System for the Visually Impaired, Trisanth Srinivasan, Santosh Patapati*
- 10 *A Snapshot Low-Light Depth from Defocus System, Wei Xu, Charles James Wagner, Junjie Luo, Qi Guo*
- 11 *City-wide interactive image geo-localization with MegaLoc, Gabriele Berton, Carlo Masone*
- 12 *FruitNinja: 3d object interior texture generation with gaussian splatting, Yuhao Chen, Shahan Nedadahandeh, Fangyu Wu*
- 13 *Seeing Around Corners in Real-Time using Mobile LiDAR, Aaron Young, Siddharth Somasundaram, Nick Tsao, Nikhil Behari, Akshat Dave, Adithya Pediredla, Ramesh Raskar*
- 14 *Grounding Pixels in Facts: Distilled Knowledge Retrieval for Factual Text-to-Video Generation, Daniel Lee, Arjun Chandra, Yang Zhou, Yunyao Li, Simone Conia*
- 15 *Focal Split: Untethered Snapshot Depth from Differential Defocus, Junjie Luo, John Mamish, Alan Fu, Thomas Concannon, Josiah Hester, Emma Alexander, Qi Guo*

Notes:

54 | CVPR 2025 | PROGRAM GUIDE MAIN CONFERENCE

Sunday, June 15

7:30 - 14:30 Registration / Badge Pickup (ExHall Concourse)

7:00 - 17:00 Press Room (203 B)

7:00 - 17:00 Mother's Room (Level 1 near Room 101 and on Level 3 near Exhibit Hall D.)

7:00 - 17:00 Prayer or Quiet Room (203 A)

7:30 - 9:00 Breakfast (ExHall C)

8:00 - 8:30 Poster Setup (ExHall D)

9:00 - 10:15 Oral Session 5A: Generative AI (Karl Dean Ballroom)

🏆 - Award candidate paper

- 1 Improving Diffusion Inverse Problem Solving with Decoupled Noise Annealing, *Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, Yang Song*
- 2 DesignDiffusion: High-Quality Text-to-Design Image Generation with Diffusion Models, *Zhendong Wang, Jianmin Bao, Shuyang Gu, Dong Chen, Wengang Zhou, Houqiang Li*
- 3 CustAny: Customizing Anything from A Single Example, *Lingjie Kong, Kai Wu, Chengming Xu, Xiaobin Hu, Wenhui Han, Jinlong Peng, Donghao Luo, Mengtian Li, Jiangning Zhang, Chengjie Wang, Yanwei Fu*
- 4 Minoroty-Focused Text-to-Image Generation via Prompt Optimization, *Soobin Um, Jong Chul Ye*
- 5 Black-Box Forgery Attacks on Semantic Watermarks for Diffusion Models, *Andreas Müller, Denis Lukovnikov, Jonas Thietke, Asja Fischer, Erwin Quiring*

9:00 - 10:15 Oral Session 5B: Learning Systems and Medical Applications (ExHall A2)

- 1 UniAP: Unifying Inter- and Intra-Layer Automatic Parallelism by Mixed Integer Quadratic Programming, *Hao Lin, Ke Wu, Jie Li, Jun Li, Wu-Jun Li*
- 2 Geometric Knowledge-Guided Localized Global Distribution Alignment for Federated Learning, *Yanbiao Ma, Wei Dai, Wenke Huang, Jiayi Chen*
- 3 Enhancing Diversity for Data-free Quantization, *Kai Zhao, Zhihao Zhuang, Miao Zhang, Chenjuan Guo, Yang Shu, Bin Yang*
- 4 TopoCellGen: Generating Histopathology Cell Topology with a Diffusion Model, *Meilong Xu, Saumya Gupta, Xiaoling Hu, Chen Li, Shahira Abousamra, Dimitris Samaras, Prateek Prasanna, Chao Chen*
- 5 Enhancing SAM with Efficient Prompting and Preference Optimization for Semi-supervised Medical Image Segmentation, *Aishik Konwer, Zhijian Yang, Erhan Bas, Cao Xiao, Prateek Prasanna, Parminder Bhatia, Taha Kass-Hout*

9:00 - 10:15 Oral Session 5C: Visual and Spatial Computing (Davidson Ballroom)

- 1 Adv-CPG: A Customized Portrait Generation Framework with Facial Adversarial Attacks, *Junying Wang, Hongyuan Zhang, Yuan Yuan*
- 2 Gromov-Wasserstein Problem with Cyclic Symmetry, *Shoichiro Takeda, Yasunori Akagi*
- 3 Time of the Flight of the Gaussians: Optimizing Depth Indirectly in Dynamic Radiance Fields, *Runfeng Li, Mikhail Okunov, Zixuan Guo, Anh Ha Duong, Christian Richardt, Matthew O'Toole, James Tompkin*
- 4 Zero-Shot Monocular Scene Flow Estimation in the Wild, *Yiqing Liang, Abhishek Badki, Hang Su, James Tompkin, Orazio Gallo*
- 5 3D Student Splatting and Scooping, *Jialin Zhu, Jiangbei Yue, Feixiang He, He Wang*

10:00 - 11:00 Coffee Break (ExHall D)

10:00 - 10:30 Poster Setup (ExHall D)

10:30 - 12:30 Poster Session 5 & Exhibit Hall (ExHall D)

* - Highlight paper 🏆 - Award candidate paper

○ - Oral Paper ☆ - Outstanding Reviewer

- 1 DualTalk: Dual-Speaker Interaction for 3D Talking Head Conversations, *Ziqiao Peng, Yanbo Fan, Haoyu Wu, Xuan Wang, Hongyan Liu, Jun He, Zhaoxin Fan*

- 2 Perceptually Accurate 3D Talking Head Generation: New Definitions, * Speech-Mesh Representation, and Evaluation Metrics, *Lee Chae-Yeon, Oh Hyun-Bin, Han EunGi, Kim Sung-Bin, Suekyeong Nam, Tae-Hyun Oh*
- 3 Teller: Real-Time Streaming Audio-Driven Portrait Animation with Autoregressive Motion Generation, *Dingcheng Zhen, Shunshun Yin, Shiyang Qin, Hou Yi, Ziwei Zhang, Siyuan Liu, Gan Qi, Ming Tao*
- 4 Hallo3: Highly Dynamic and Realistic Portrait Image Animation with Video Diffusion Transformer, *Jiahao Cui, Hui Li, Yun Zhan, Hanlin Shang, Kaihui Cheng, Yuqi Ma, Shan Mu, Hang Zhou, Jingdong Wang, Siyu Zhu*
- 5 StableAnimator: High-Quality Identity-Preserving Human Image Animation, *Shuyuan Tu, Zhen Xing, Xintong Han, Zhi-Qi Cheng, Qi Dai, Chong Luo, Zuxuan Wu*
- 6 IM-Portrait: Learning 3D-aware Video Diffusion for Photorealistic Talking Heads from Monocular Videos, *C, Yuan Li, Ziqian Bai, Feitong Tan, Zhaopeng Cui, Sean Fanello, Yinda Zhang*
- 7 3D Gaussian Head Avatars with Expressive Dynamic Appearances by Compact Tensorial Representations, *Yating Wang, Xuan Wang, Ran Yi, Yanbo Fan, Jichen Hu, Jingcheng Zhu, Lizhuang Ma*
- 8 LUCAS: Layered Universal Codec Avatars, *Di Liu, Teng Deng, Giljoo Nam, Yu Rong, Stanislav Pidhorskyi, Junxuan Li, Jason Saragih, Dimitris N. Metaxas, Chen Cao*
- 9 GeoAvatar: Geometrically-Consistent Multi-Person Avatar Reconstruction from Sparse Multi-View Videos, *Soo Hyun Lee, Seoyeon Kim, HeeKyung Lee, Won-Sik Jeong, Joo Ho Lee*
- 10 AniGS: Animatable Gaussian Avatar from a Single Image with Inconsistent Gaussian Reconstruction, *Lingteng Qiu, Shenhao Zhu, Qi Zuo, Xiaodong Gu, Yuan Dong, Junfei Zhang, Chao Xu, Zhe Li, Weihao Yuan, Liefeng Bo, Guanying Chen, Zilong Dong*
- 11 TAGA: Self-supervised Learning for Template-free Animatable ☆ Gaussian Articulated Model, *Zhichao Zhai, Guikun Chen, Wenguan Wang, Dong Zheng, Jun Xiao*
- 12 DRiVE: Diffusion-based Rigging Empowers Generation of Versatile and Expressive Characters, *Mingze Sun, Junhao Chen, Junting Dong, Yurun Chen, Xinyu Jiang, Shiwei Mao, Puhua Jiang, Jingbo Wang, Bo Dai, Ruqi Huang*
- 13 MIMO: Controllable Character Video Synthesis with Spatial Decomposed Modeling, *Yifang Men, Yuan Yao, Miaomiao Cui, Liefeng Bo*
- 14 Unsupervised Discovery of Facial Landmarks and Head Pose, ☆ *Satyajit Tourani, Siddharth Tourani, Arif Mahmood, Muhammad Haris Khan*
- 15 Data Synthesis with Diverse Styles for Face Recognition via 3DMM-Guided Diffusion, *Yuxi Mi, Zhizhou Zhong, Yuge Huang, Qiuyang Yuan, Xuan Zhao, Jianqing Xu, Shouhong Ding, Shaoming Wang, Rizen Guo, Shuigeng Zhou*
- 16 PGC: Physics-Based Gaussian Cloth from a Single Pose, * *Michelle Guo, Matt Jen-Yuan Chiang, Igor Santesteban, Nikolaos Sarafianos, Hsiao-yu Chen, Oshri Halimi, Aljaž Božič, Shunsuke Saito, Jiajun Wu, C. Karen Liu, Tuur Stuyck, Egor Larionov*
- 17 Is this Generated Person Existed in Real-world? Fine-grained * Detecting and Calibrating Abnormal Human-body, *Zeqing Wang, Qingyang Ma, Wentao Wan, Haojie Li, Keze Wang, Yonghong Tian*
- 18 Enhancing Virtual Try-On with Synthetic Pairs and Error-Aware Noise Scheduling, *Nannan Li, Kevin J. Shih, Bryan A. Plummer*
- 19 ProjAttacker: A Configurable Physical Adversarial Attack for Face Recognition via Projector, *Yuanwei Liu, Hui Wei, Chengyu Jia, Ruqi Xiao, Weijian Ruan, Xingxing Wei, Joey Tianyi Zhou, Zheng Wang*
- 20 ABC-Former: Auxiliary Bimodal Cross-domain Transformer with Interactive Channel Attention for White Balance, *Yu-Cheng Chiu, Guan-Rong Chen, Zihao Chen, Yan-Tsung Peng*
- 21 URWKV: Unified RWKV Model with Multi-state Perspective for Low-light Image Restoration, *Rui Xu, Yuzhen Niu, Yuezhou Li, Huangbiao Xu, Wenxi Liu, Yuzhong Chen*
- 22 Efficient Diffusion as Low Light Enhancer, *Guanzhou Lan, Qianli Ma, Yuqi Yang, Zhigang Wang, Dong Wang, Xuelong Li, Bin Zhao*

- 23 Noise Calibration and Spatial-Frequency Interactive Network for STEM Image Enhancement, *Hesong Li, Ziqi Wu, Ruiwen Shao, Tao Zhang, Ying Fu*
- 24 DOF-GS: Adjustable Depth-of-Field 3D Gaussian Splatting for Post-Capture Refocusing, Defocus Rendering and Blur Removal, *Yujie Wang, Praneeth Chakravarthula, Baoquan Chen*
- 25 ReCap: Better Gaussian Relighting with Cross-Environment Captures, *Jingzhi Li, Zongwei Wu, Eduard Zamfir, Radu Timofte*
- 26 Factored-NeuS: Reconstructing Surfaces, Illumination, and Materials of Possibly Glossy Objects, *Yue Fan, Ningjing Fan, Ivan Skorokhodov, Oleg Voynov, Savva Ignatyev, Evgeny Burnaev, Peter Wonka, Yiqun Wang*
- 27 SpectroMotion: Dynamic 3D Reconstruction of Specular Scenes, *Cheng-De Fan, Chen-Wei Chang, Yi-Ruei Liu, Jie-Ying Lee, Jiun-Long Huang, Yu-Chee Tseng, Yu-Lun Liu*
- 28 Radio Frequency Ray Tracing with Neural Object Representation for Enhanced RF Modeling, *Xingyu Chen, Zihao Feng, Kun Qian, Xinyu Zhang*
- 29 GoLF-NRT: Integrating Global Context and Local Geometry for Few-Shot View Synthesis, *You Wang, Li Fang, Hao Zhu, Fei Hu, Long Ye, Zhan Ma*
- 30 3D Convex Splatting: Radiance Field Rendering with 3D Smooth Convexes, *Jan Held, Renaud Vandeghen, Abdullah Hamdi, Adrien Deliege, Anthony Cioppa, Silvio Giancola, Andrea Vedaldi, Bernard Ghanem, Marc Van Droogenbroeck*
- 31 Volumetric Surfaces: Representing Fuzzy Geometries with Layered Meshes, *Stefano Esposito, Anpei Chen, Christian Reiser, Samuel Rota Bulò, Lorenzo Porzi, Katja Schwarz, Christian Richardt, Michael Zollhöfer, Peter Kotschieder, Andreas Geiger*
- 32 MetricGrids: Arbitrary Nonlinear Approximation with Elementary Metric Grids based Implicit Neural Representation, *Shu Wang, Yanbo Gao, Shuai Li, Chong Lv, Xun Cai, Chuankun Li, Hui Yuan, Jinglin Zhang*
- 33 Mani-GS: Gaussian Splatting Manipulation with Triangular Mesh, *Xiangjun Gao, Xiaoyu Li, Yiyu Zhuang, Qi Zhang, Wenbo Hu, Chaopeng Zhang, Yao Yao, Ying Shan, Long Quan*
- 34 TriTex: Learning Texture from a Single Mesh via Triplane Semantic Features, *Dana Cohen-Bar, Daniel Cohen-Or, Gal Chechik, Yoni Kasten*
- 35 HybridMQA: Exploring Geometry-Texture Interactions for Colored Mesh Quality Assessment, *Armin Shafiee Sarvestani, Sheyang Tang, Zhou Wang*
- 36 ARM: Appearance Reconstruction Model for Relightable 3D Generation, *Xiang Feng, Chang Yu, Zoubin Bi, Yintong Shang, Feng Gao, Hongzhi Wu, Kun Zhou, Chenfanfu Jiang, Yin Yang*
- 37 DTGBrepGen: A Novel B-rep Generative Model through Decoupling Topology and Geometry, *Jing Li, Yihang Fu, Falai Chen*
- 38 CADDreamer: CAD Object Generation from Single-view Images, *Yuan Li, Cheng Lin, Yuan Liu, Xiaoxiao Long, Chenxu Zhang, Ningna Wang, Xin Li, Wenping Wang, Xiaohu Guo*
- 39 Sharp-It: A Multi-view to Multi-view Diffusion Model for 3D Synthesis and Manipulation, *Yiftach Edelstein, Or Patashnik, Dana Cohen-Bar, Lihi Zelnik-Manor*
- 40 Structured 3D Latents for Scalable and Versatile 3D Generation, *Jianfeng Xiang, Zelong Lv, Sicheng Xu, Yu Deng, Ruicheng Wang, Bowen Zhang, Dong Chen, Xin Tong, Jiaolong Yang*
- 41 Hash3D: Training-free Acceleration for 3D Generation, *Xingyi Yang, Songhua Liu, Xinchao Wang*
- 42 SwiftEdit: Lightning Fast Text-Guided Image Editing via One-Step Diffusion, *Trong-Tung Nguyen, Quang Nguyen, Khoi Nguyen, Anh Tran, Cuong Pham*
- 43 Dragin3D: Image Editing by Dragging in 3D Space, *Weiran Guang, Xiaoguang Gu, Mengqi Huang, Zhendong Mao*
- 44 Deformable Radial Kernel Splatting, *Yi-Hua Huang, Ming-Xian Lin, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, Xiaojuan Qi*
- 45 SplatFlow: Multi-View Rectified Flow Model for 3D Gaussian Splatting Synthesis, *Hyojun Go, Byeongjun Park, Jiho Jang, Jin-Young Kim, Soonwoo Kwon, Changick Kim*
- 46 Speedy-Splat: Fast 3D Gaussian Splatting with Sparse Pixels and Sparse Primitives, *Alex Hanson, Allen Tu, Geng Lin, Vasu Singla, Matthias Zwicker, Tom Goldstein*
- 47 GS-2DGS: Geometrically Supervised 2DGS for Reflective Object Reconstruction, *Jinguang Tong, Xuesong Li, Fahira Afzal Maken, Sundaram Muthu, Lars Petersson, Chuong Nguyen, Hongdong Li*
- 48 High-fidelity 3D Object Generation from Single Image with RGBN-Volume Gaussian Reconstruction Model, *Yiyang Shen, Kun Zhou, He Wang, Yin Yang, Tianjia Shao*
- 49 MonoSplat: Generalizable 3D Gaussian Splatting from Monocular Depth Foundation Models, *Yifan Liu, Keyu Fan, Weihao Yu, Chenxin Li, Hao Lu, Yixuan Yuan*
- 50 LITA-GS: Illumination-Agnostic Novel View Synthesis via Reference-Free 3D Gaussian Splatting and Physical Priors, *Han Zhou, Wei Dong, Jun Chen*
- 51 Splatter-360: Generalizable 360 Gaussian Splatting for Wide-baseline Panoramic Images, *Zheng Chen, Chenming Wu, Zhelun Shen, Chen Zhao, Weicai Ye, Haocheng Feng, Errui Ding, Song-Hai Zhang*
- 52 DropGaussian: Structural Regularization for Sparse-view Gaussian Splatting, *Hyunwoo Park, Gun Ryu, Wonjun Kim*
- 53 Panorama Generation From NFOV Image Done Right, *Dian Zheng, Cheng Zhang, Xiao-Ming Wu, Cao Li, Chengfei Lv, Jian-Fang Hu, Wei-Shi Zheng*
- 54 SIR-DIFF: Sparse Image Sets Restoration with Multi-View Diffusion Model, *Yucheng Mao, Boyang Wang, Nilesh Kulkarni, Jeong Joon Park*
- 55 Ouroboros3D: Image-to-3D Generation via 3D-aware Recursive Diffusion, *Hao Wen, Zehuan Huang, Yaohui Wang, Xinyuan Chen, Lu Sheng*
- 56 MonoInstance: Enhancing Monocular Priors via Multi-view Instance Alignment for Neural Rendering and Reconstruction, *Wenyuan Zhang, Yixiao Yang, Han Huang, Liang Han, Kanle Shi, Yu-Shen Liu, Zhizhong Han*
- 57 SfM-Free 3D Gaussian Splatting via Hierarchical Training, *Bo Ji, Angela Yao*
- 58 MVBoost: Boost 3D Reconstruction with Multi-View Refinement, *Xiangyu Liu, Xiaomei Zhang, Zhiyuan Ma, Xiangyu Zhu, Zhen Lei*
- 59 AerialMegaDepth: Learning Aerial-Ground Reconstruction and View Synthesis, *Khiem Vuong, Anurag Ghosh, Deva Ramanan, Srinivasa Narasimhan, Shubham Tulsiani*
- 60 World-consistent Video Diffusion with Explicit 3D Modeling, *Qihang Zhang, Shuangfei Zhai, Miguel Ángel Bautista Martín, Kevin Miao, Alexander Toshev, Joshua Susskind, Jiatao Gu*
- 61 Improving Gaussian Splatting with Localized Points Management, *Haosen Yang, Chenhao Zhang, Wenqing Wang, Marco Volino, Adrian Hilton, Li Zhang, Xiatian Zhu*
- 62 RelationField: Relate Anything in Radiance Fields, *Sebastian Koch, Johanna Wald, Mirco Colosi, Narunas Vaskevicius, Pedro Hermosilla, Federico Tombari, Timo Ropinski*
- 63 GS-DiT: Advancing Video Generation with Dynamic 3D Gaussian Fields through Efficient Dense 3D Point Tracking, *Weikang Bian, Zhaoyang Huang, Xiaoyu Shi, Yijin Li, Fu-Yun Wang, Hongsheng Li*
- 64 DynaMoDe-NeRF: Motion-aware Deblurring Neural Radiance Field for Dynamic Scenes, *Ashish Kumar, Rajagopalan A. N.*
- 65 DiET-GS: Diffusion Prior and Event Stream-Assisted Motion Deblurring 3D Gaussian Splatting, *Seungjun Lee, Gim Hee Lee*
- 66 FreeTimeGS: Free Gaussian Primitives at Anytime Anywhere for Dynamic Scene Reconstruction, *Yifan Wang, Peishan Yang, Zhen Xu, Jiaming Sun, Zhanhua Zhang, Yong Chen, Hujun Bao, Sida Peng, Xiaowei Zhou*
- 67 GIFStream: 4D Gaussian-based Immersive Video with Feature Stream, *Hao Li, Sicheng Li, Xiang Gao, Abudouaihati Batuer, Lu Yu, Yiyi Liao*
- 68 DRAWER: Digital Reconstruction and Articulation With Environment Realism, *Hongchi Xia, Entong Su, Marius Memmel, Arhan Jain, Raymond Yu, Numfor Mbiziwo-Tiapo, Ali Farhadi, Abhishek Gupta, Shenlong Wang, Wei-Chiu Ma*
- 69 Gromov-Wasserstein Problem with Cyclic Symmetry,

- *Shoichiro Takeda, Yasunori Akagi*
- 70 Dynamic Neural Surfaces for Elastic 4D Shape Representation and Analysis, *Awais Nizamani, Hamid Laga, Guanjin Wang, Farid Boussaid, Mohammed Bennamoun, Anuj Srivastava*
- 71 Higher-Order Ratio Cycles for Fast and Globally Optimal Shape Matching, *Paul Roetzer, Viktoria Ehm, Daniel Cremers, Zorah Löhner, Florian Bernard*
- 72 Event Ellipsometer: Event-based Mueller-Matrix Video Imaging, *Ryota Maeda, Yunseong Moon, Seung-Hwan Baek*
- * 73 AniGrad: Anisotropic Gradient-Adaptive Sampling for 3D Reconstruction From Monocular Video, *Noah Stier, Alex Rich, Pradeep Sen, Tobias Höllerer*
- 74 ODHSR: Online Dense 3D Reconstruction of Humans and Scenes from Monocular Videos, *Zetong Zhang, Manuel Kaufmann, Lixin Xue, Jie Song, Martin R. Oswald*
- ☆ 75 All-directional Disparity Estimation for Real-world QPD Images, *Hongtao Yu, Shaohui Song, Lihu Sun, Wenkai Su, Xiaodong Yang, Chengming Liu*
- * 76 Mono2Stereo: A Benchmark and Empirical Study for Stereo Conversion, *Songsong Yu, Yuxin Chen, Zhongang Qi, Zeke Xie, Yifan Wang, Lijun Wang, Ying Shan, Huchuan Lu*
- 77 DEFOM-Stereo: Depth Foundation Model Based Stereo Matching, *Hualie Jiang, Zhiqiang Lou, Laiyan Ding, Rui Xu, Minglang Tan, Wenjie Jiang, Rui Huang*
- 78 Improved Monocular Depth Prediction Using Distance Transform Over Pre-semantic Contours with Self-supervised Neural Networks, *Marwane Hariat, Antoine Manzanera, David Filliat*
- 79 Synthetic-to-Real Self-supervised Robust Depth Estimation via Learning with Motion and Structure Priors, *Weilong Yan, Ming Li, Haipeng Li, Shuwei Shao, Robby T. Tan*
- 80 MP-SfM: Monocular Surface Priors for Robust Structure-from-Motion, *Zador Pataki, Paul-Edouard Sarlin, Johannes L. Schönberger, Marc Pollefeys*
- 81 PRaDA: Projective Radial Distortion Averaging, *Daniil Sinitsyn, Linus Härenstam-Nielsen, Daniel Cremers*
- ☆ 82 Practical Solutions to the Relative Pose of Three Calibrated Cameras, *Charalambos Tzamos, Viktor Kocur, Yaqing Ding, Daniel Barath, Zuzana Berger Haladova, Torsten Sattler, Zuzana Kukelova*
- 83 Fast3R: Towards 3D Reconstruction of 1000+ Images in One Forward Pass, *Jianing Yang, Alexander Sax, Kevin J. Liang, Mikael Henaff, Hao Tang, Ang Cao, Joyce Chai, Franziska Meier, Matt Feiszli*
- ☆ 84 FLARE: Feed-forward Geometry, Appearance and Camera Estimation from Uncalibrated Sparse Views, *Shangzhan Zhang, Jianyuan Wang, Yinghao Xu, Nan Xue, Christian Rupprecht, Xiaowei Zhou, Yujun Shen, Gordon Wetzstein*
- 85 Time of the Flight of the Gaussians: Optimizing Depth Indirectly in Dynamic Radiance Fields, *Runfeng Li, Mikhail Okunev, Zixuan Guo, Anh Ha Duong, Christian Richardt, Matthew O'Toole, James Tompkin*
- 86 Reconstructing People, Places, and Cameras, *Lea Müller, Hongsuk Choi, Anthony Zhang, Brent Yi, Jitendra Malik, Angjoo Kanazawa*
- * 87 Omnidirectional Multi-Object Tracking, *Kai Luo, Hao Shi, Sheng Wu, Fei Teng, Mengfei Duan, Chang Huang, Yuhang Wang, Kaiwei Wang, Kailun Yang*
- 88 CoMatcher: Multi-View Collaborative Feature Matching, *Jintao Zhang, Zimin Xia, Mingyue Dong, Shuhan Shen, Linwei Yue, Xianwei Zheng*
- 89 PIDLoc: Cross-View Pose Optimization Network Inspired by PID Controllers, *Wooju Lee, Juhye Park, Dasol Hong, Changki Sung, Youngwoo Seo, DongWan Kang, Hyun Myung*
- 90 BLADE: Single-view Body Mesh Estimation through Accurate Depth Estimation, *Shengze Wang, Jiefeng Li, Tianye Li, Ye Yuan, Henry Fuchs, Koki Nagano, Shalini De Mello, Michael Stengel*
- ☆ 91 4D LangSplat: 4D Language Gaussian Splatting via Multimodal Large Language Models, *Wanhua Li, Renping Zhou, Jiawei Zhou, Yingwei Song, Johannes Herter, Minghan Qin, Gao Huang, Hanspeter Pfister*
- 92 SelfSplat: Pose-Free and 3D Prior-Free Generalizable 3D Gaussian Splatting, *Gyeongjin Kang, Jisang Yoo, Jihyeon Park, Seungtae Nam, Hyeonsoo Im, Sangheon Shin, Sangpil Kim, Eunbyung Park*
- 93 UNOPose: Unseen Object Pose Estimation with an Unposed RGB-D Reference Image, *Xingyu Liu, Gu Wang, Ruida Zhang, Chenyangguang Zhang, Federico Tombari, Xiangyang Ji*
- 94 Recurrent Feature Mining and Keypoint Mixup Padding for Category-Agnostic Pose Estimation, *Junjie Chen, Weilong Chen, Yifan Zuo, Yuming Fang*
- 95 SCFlow2: Plug-and-Play Object Pose Refiner with Shape-Constraint Scene Flow, *Qingyuan Wang, Rui Song, Jiaojiao Li, Kerui Cheng, David Ferstl, Yinlin Hu*
- 96 GIVEPose: Gradual Intra-class Variation Elimination for RGB-based Category-Level Object Pose Estimation, *Ziqin Huang, Gu Wang, Chenyangguang Zhang, Ruida Zhang, Xiu Li, Xiangyang Ji*
- 97 Robust Multi-Object 4D Generation for In-the-wild Videos, *Wen-Hsuan Chu, Lei Ke, Jianmeng Liu, Mingxiao Huo, Pavel Tokmakov, Katerina Fragkiadaki*
- 98 Category-Agnostic Neural Object Rigging, *Guangzhao He, Chen Geng, Shangzhe Wu, Jiajun Wu*
- 99 PURA: Parameter Update-Recovery Test-Time Adaption for RGB-T Tracking, *Zekai Shao, Yufan Hu, Bin Fan, Hongmin Liu*
- 100 ACAttack: Adaptive Cross Attacking RGB-T Tracker via Multi-Modal Response Decoupling, *Xinyu Xiang, Qinglong Yan, Hao Zhang, Jiayi Ma*
- 101 Leveraging 3D Geometric Priors in 2D Rotation Symmetry Detection, *Ahyun Seo, Minsu Cho*
- 102 SeCap: Self-Calibrating and Adaptive Prompts for Cross-view Person Re-identification in Aerial-Ground Networks, *Shining Wang, Yunlong Wang, Ruiqi Wu, Bingliang Jiao, Wenxuan Wang, Peng Wang*
- * 103 HyperNet Fields: Efficiently Training Hypernetworks without Ground Truth by Learning Weight Trajectories, *Eric Hedlin, Munawar Hayat, Fatih Porikli, Kwang Moo Yi, Shweta Mahajan*
- 104 Learning Bijective Surface Parameterization for Inferring Signed Distance Functions from Sparse Point Clouds with Grid Deformation, *Takeshi Noda, Chao Chen, Junsheng Zhou, Weiqi Zhang, Yu-Shen Liu, Zhizhong Han*
- 105 EdgeMovingNet: Edge-preserving Point Cloud Reconstruction via Joint Geometry Features, *Xinran Yang, Donghao Ji, Yuanqi Li, Junyuan Xie, Jie Guo, Yanwen Guo*
- 106 GraphI2P: Image-to-Point Cloud Registration with Exploring Pattern of Correspondence via Graph Learning, *Lin Bie, Shouan Pan, Siqi Li, Yining Zhao, Yue Gao*
- 107 RENO: Real-Time Neural Compression for 3D LiDAR Point Clouds, *Kang You, Tong Chen, Dandan Ding, M. Salman Asif, Zhan Ma*
- 108 Point Clouds Meets Physics: Dynamic Acoustic Field Fitting Network for Point Cloud Understanding, *Changshuo Wang, Shuting He, Xiang Fang, Jiawei Han, Zhonghang Liu, Xin Ning, Weijun Li, Prayag Tiwari*
- 109 Sonata: Self-Supervised Learning of Reliable Point Representations, *Xiaoyang Wu, Daniel DeTone, Duncan Frost, Tianwei Shen, Chris Xie, Nan Yang, Jakob Engel, Richard Newcombe, Hengshuang Zhao, Julian Straub*
- * 110 Generative Hard Example Augmentation for Semantic Point Cloud Segmentation, *Qi Zhang, Jibin Peng, Zhao Huang, Wei Feng, Di Lin*
- 111 BWFormer: Building Wireframe Reconstruction from Airborne LiDAR Point Cloud with Transformer, *Yuzhou Liu, Lingjie Zhu, Hanqiao Ye, Shangfeng Huang, Xiang Gao, Xianwei Zheng, Shuhan Shen*
- * 112 Cubify Anything: Scaling Indoor 3D Object Detection, *Justin Lazarow, David Griffiths, Gefen Kohavi, Francisco Crespo, Afshin Dehghan*
- 113 Multi-Scale Neighborhood Occupancy Masked Autoencoder for Self-Supervised Learning in LiDAR Point Clouds, *Mohamed Abdelsamad, Michael Ulrich, Claudius Glaeser, Abhinav Valada*
- 114 Unlocking Generalization Power in LiDAR Point Cloud Registration, *Zhenxuan Zeng, Qiao Wu, Xiyu Zhang, Lin Yuanbo Wu, Pei An, Jiaqi Yang, Ji Wang, Peng Wang*
- * 115 Distilling Monocular Foundation Model for Fine-grained Depth Completion, *Yingping Liang, Yutao Hu, Wenqi Shao, Ying Fu*

- 116 MonoTAKD: Teaching Assistant Knowledge Distillation for Monocular 3D Object Detection, *Hou-I Liu, Christine Wu, Jen-Hao Cheng, Wenhao Chai, Shian-Yun Wang, Gaowen Liu, Hugo Latapie, Jih-Ciang Wu, Jenq-Neng Hwang, Hong-Han Shuai, Wen-Huang Cheng*
- 117 RICCARDO: Radar Hit Prediction and Convolution for Camera-Radar 3D Object Detection, *Yunfei Long, Abhinav Kumar, Xiaoming Liu, Daniel Morris*
- 118 UCM-VeID V2: A Richer Dataset and A Pre-training Method for UAV Cross-Modality Vehicle Re-Identification, *Xingyue Liu, Jiahao Qi, Chen Chen, KangCheng Bin, Ping Zhong*
- 119 SparseAlign: a Fully Sparse Framework for Cooperative Object Detection, *Yunshuang Yuan, Yan Xia, Daniel Cremers, Monika Sester*
- 120 Hyperdimensional Uncertainty Quantification for Multimodal Uncertainty Fusion in Autonomous Vehicles Perception, *Luke Chen, Junyao Wang, Trier Mortlock, Pramod Khargonekar, Mohammad Abdullah Al Faruque*
- 121 Omni-Scene: Omni-Gaussian Representation for Ego-Centric Sparse-View Scene Reconstruction, *Dongxu Wei, Zhiqi Li, Peidong Liu*
- 122 Floxels: Fast Unsupervised Voxel Based Scene Flow Estimation, *David T. Hoffmann, Syed Haseeb Raza, Hanqiu Jiang, Denis Tananaev, Steffen Klingenhoefer, Martin Meinke*
- 123 Spatiotemporal Decoupling for Efficient Vision-Based Occupancy Forecasting, *Jingyi Xu, Xieyuanli Chen, Junyi Ma, Jiawei Huang, Jintao Xu, Yue Wang, Ling Pei*
- 124 Rectification-specific Supervision and Constrained Estimator for Online Stereo Rectification, *Rui Gong, Kim-Hui Yap, Weide Liu, Xulei Yang, Jun Cheng*
- 125 Uncertainty-Instructed Structure Injection for Generalizable HD Map Construction, *Xiaolu Liu, Ruizhi Yang, Song Wang, Wentong Li, Junbo Chen, Jianke Zhu*
- 126 JarvisIR: Elevating Autonomous Driving Perception with Intelligent Image Restoration, *Yunlong Lin, Zixu Lin, Haoyu Chen, Panwang Pan, Chenxin Li, Sixiang Chen, Kairun Wen, Yeying Jin, Wenbo Li, Xinghao Ding*
- 127 MaskGWM: A Generalizable Driving World Model with Video Mask Reconstruction, *Jingcheng Ni, Yuxin Guo, Yichen Liu, Rui Chen, Lewei Lu, Zehuan Wu*
- 128 GenAssets: Generating in-the-wild 3D Assets in Latent Space, *Ze Yang, Jingkan Wang, Haowei Zhang, Sivabalan Manivasagam, Yun Chen, Raquel Urtasun*
- 129 GEM: A Generalizable Ego-Vision Multimodal World Model for Fine-Grained Ego-Motion, Object Dynamics, and Scene Composition Control, *Mariam Hassan, Sebastian Stapf, Ahmad Rahimi, Pedro M B Rezende, Yasaman Haghighi, David Brüggemann, Isinsu Katircioglu, Lin Zhang, Xiaoran Chen, Suman Saha, Marco Cannici, Elie Aljalbout, Botao Ye, Xi Wang, Aram Davtyan, Mathieu Salzmann, Davide Scaramuzza, Marc Pollefeys, Paolo Favaro, Alexandre Alahi*
- 130 Continuous Locomotive Crowd Behavior Generation, *Inhwan Bae, Junoh Lee, Hae-Gon Jeon*
- 131 Don't Shake the Wheel: Momentum-Aware Planning in End-to-End Autonomous Driving, *Ziying Song, Caiyan Jia, Lin Liu, Hongyu Pan, Yongchang Zhang, Junming Wang, Xingyu Zhang, Shaoqing Xu, Lei Yang, Yadan Luo*
- 132 OmniDrive: A Holistic Vision-Language Dataset for Autonomous Driving with Counterfactual Reasoning, *Shihao Wang, Zhiding Yu, Xiaohui Jiang, Shiyi Lan, Min Shi, Nadine Chang, Jan Kautz, Ying Li, Jose M. Alvarez*
- 133 Embodied Scene Understanding for Vision Language Models via MetaVQA, *Weizhen Wang, Chenda Duan, Zhenghao Peng, Yuxin Liu, Bolei Zhou*
- 134 SocialMOIF: Multi-Order Intention Fusion for Pedestrian Trajectory Prediction, *Kai Chen, Xiaodong Zhao, Yujie Huang, Guoyu Fang, Xiao Song, Ruiping Wang, Ziyuan Wang*
- 135 Unified Uncertainty-Aware Diffusion for Multi-Agent Trajectory Modeling, *Guillem Capellera, Antonio Rubio, Luis Ferraz, Antonio Agudo*
- 136 RADIOv2.5: Improved Baselines for Agglomerative Vision Foundation Models, *Greg Heinrich, Mike Ranzinger, Hongxu Yin, Yao Lu, Jan Kautz, Andrew Tao, Bryan Catanzaro, Pavlo Molchanov*
- 137 Let Humanoids Hike! Integrative Skill Development on Complex Trails, *Kwan-Yee Lin, Stella X. Yu*
- 138 Universal Actions for Enhanced Embodied Foundation Models, *Jinliang Zheng, Jianxiong Li, Dongxiu Liu, Yinan Zheng, Zhihao Wang, Zhonghong Ou, Yu Liu, Jingjing Liu, Ya-Qin Zhang, Xianyu Zhan*
- 139 Tartan IMU: A Light Foundation Model for Inertial Positioning in Robotics, *Shibo Zhao, Sifan Zhou, Raphael Blanchard, Yuheng Qiu, Wenshan Wang, Sebastian Scherer*
- 140 3D-MVP: 3D Multiview Pretraining for Manipulation, *Shengyi Qian, Kaichun Mo, Valts Blukis, David F. Fouhey, Dieter Fox, Ankit Goyal*
- 141 RoboGround: Robotic Manipulation with Grounded Vision-Language Priors, *Haifeng Huang, Xinyi Chen, Yilun Chen, Hao Li, Xiaoshen Han, Zehan Wang, Tai Wang, Jiangmiao Pang, Zhou Zhao*
- 142 Mitigating the Human-Robot Domain Discrepancy in Visual Pre-training for Robotic Manipulation, *Jiaming Zhou, Teli Ma, Kun-Yu Lin, Zifan Wang, Ronghe Qiu, Junwei Liang*
- 143 Prof. Robot: Differentiable Robot Rendering Without Static and Self-Collisions, *Quanyuan Ruan, Jiabao Lei, Wenhao Yuan, Yanglin Zhang, Dekun Lu, Guiliang Liu, Kui Jia*
- 144 Think Small, Act Big: Primitive Prompt Learning for Lifelong Robot Manipulation, *Yuangqi Yao, Siao Liu, Haoming Song, Delin Qu, Qizhi Chen, Yan Ding, Bin Zhao, Zhigang Wang, Xuelong Li, Dong Wang*
- 145 DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness, *Yiming Zhong, Qi Jiang, Jingyi Yu, Yuexin Ma*
- 146 CheckManual: A New Challenge and Benchmark for Manual-based Appliance Manipulation, *Yuxing Long, Jiyao Zhang, Mingjie Pan, Tianshu Wu, Taewhan Kim, Hao Dong*
- 147 InteractVLM: 3D Interaction Reasoning from 2D Foundational Models, *Sai Kumar Dwivedi, Dimitrije Antić, Shashank Tripathi, Omid Taheri, Cordelia Schmid, Michael J. Black, Dimitrios Tzionas*
- 148 VTON-HandFit: Virtual Try-on for Arbitrary Hand Pose Guided by Hand Priors Embedding, *Yujie Liang, Xiaobin Hu, Boyuan Jiang, Donghao Luo, Xu Peng, Kai Wu, Chengming Xu, Wenhui Han, Taisong Jin, Chengjie Wang, Rongrong Ji*
- 149 Pose-Guided Temporal Enhancement for Robust Low-Resolution Hand Reconstruction, *Kaixin Fan, Pengfei Ren, Jingyu Wang, Haifeng Sun, Qi Qi, Zirui Zhuang, Jianxin Liao*
- 150 GaPT-DAR: Category-level Garments Pose Tracking via Integrated 2D Deformation and 3D Reconstruction, *Li Zhang, Mingliang Xu, Jianan Wang, Qiaojun Yu, Lixin Yang, Yonglu Li, Cewu Lu, Rujiang Wang, Liu Liu*
- 151 Pursuing Temporal-Consistent Video Virtual Try-On via Dynamic Pose Interaction, *Dong Li, Wenqi Zhong, Wei Yu, Yingwei Pan, Dingwen Zhang, Ting Yao, Junwei Han, Tao Mei*
- 152 A Focused Human Body Model for Accurate Anthropometric Measurements Extraction, *Shuhang Chen, Xianliang Huang, Zhizhou Zhong, Juhong Guan, Shuigeng Zhou*
- 153 Ego4o: Egocentric Human Motion Capture and Understanding from Multi-Modal Input, *Jian Wang, Rishabh Dabral, Diogo Luvizon, Zhe Cao, Lingjie Liu, Thabo Beeler, Christian Theobalt*
- 154 MotionMap: Representing Multimodality in Human Pose Forecasting, *Reyhaneh Hosseinienejad, Megh Shukla, Saeed Saadatnejad, Mathieu Salzmann, Alexandre Alahi*
- 155 POMP: Physics-constrainable Motion Generative Model through Phase Manifolds, *Bin Ji, Ye Pan, Zhimeng Liu, Shuai Tan, Xiaogang Jin, Xiaokang Yang*
- 156 H-MoRe: Learning Human-centric Motion Representation for Action Analysis, *Zhanbo Huang, Xiaoming Liu, Yu Kong*
- 157 Guiding Human-Object Interactions with Rich Geometry and Relations, *Mengqing Xue, Yifei Liu, Ling Guo, Shaoli Huang, Changxing Ding*
- 158 Deterministic-to-Stochastic Diverse Latent Feature Mapping for Human Motion Synthesis, *Yu Hua, Weiming Liu, Gui Xu, Yaqing Hou, Yew-Soon Ong, Qiang Zhang*

- 159 Dynamic Motion Blending for Versatile Motion Editing, *Nan Jiang, Hongjie Li, Ziyi Yuan, Zimo He, Yixin Chen, Tengyu Liu, Yixin Zhu, Siyuan Huang*
- 160 AToM: Aligning Text-to-Motion Model at Event-Level with
☆ GPT-4Vision Reward, *Haonan Han, Xiangzuo Wu, Huan Liao, Zunnan Xu, Zhongyuan Hu, Ronghui Li, Yachao Zhang, Xiu Li*
- 161 PersonaBooth: Personalized Text-to-Motion Generation, *Boeun Kim, Hea In Jeong, JungHoon Sung, Yihua Cheng, Jeongmin Lee, Ju Yong Chang, Sang-Il Choi, Younggeun Choi, Saim Shin, Jungho Kim, Hyung Jin Chang*
- 162 Move-in-2D: 2D-Conditioned Human Motion Generation, *Hsin-Ping Huang, Yang Zhou, Jui-Hsien Wang, Difan Liu, Feng Liu, Ming-Hsuan Yang, Zhan Xu*
- 163 PoseTraj: Pose-Aware Trajectory Control in Video
Diffusion, *Longbin Ji, Lei Zhong, Pengfei Wei, Changjian Li*
- 164 Robust 3D Shape Reconstruction in Zero-Shot from a Single
☆ Image in the Wild, *Junhyeong Cho, Kim Youwang, Hunmin Yang, Tae-Hyun Oh*
- 165 Zero-Shot Monocular Scene Flow Estimation in the Wild,
🔍 *Yiqing Liang, Abhishek Badki, Hang Su, James Tompkin,*
☆ *Orazio Gallo*
- 166 HuPerFlow: A Comprehensive Benchmark for Human vs. Machine
* Motion Estimation Comparison, *Yung-Hao Yang, Zitang Sun, Taiki Fukiage, Shin'ya Nishida*
- 167 Tracktention: Leveraging Point Tracking to Attend Videos Faster
* and Better, *Zihang Lai, Andrea Vedaldi*
- 168 Align3R: Aligned Monocular Depth Estimation for Dynamic Videos,
* *Jiahao Lu, Tianyu Huang, Peng Li, Zhiyang Dou, Cheng Lin, Zhiming Cui, Zhen Dong, Sai-Kit Yeung, Wenping Wang, Yuan Liu*
- 169 Video Depth Anything: Consistent Depth Estimation for Super-
* Long Videos, *Sili Chen, Hengkai Guo, Shengnan Zhu, Feihu Zhang, Zilong Huang, Jiashi Feng, Bingyi Kang*
- 170 Learning Temporally Consistent Video Depth from Video Diffusion
☆ Priors, *Jiahao Shao, Yuanbo Yang, Hongyu Zhou, Youmin Zhang, Yujun Shen, Vitor Guizilini, Yue Wang, Matteo Poggi, Yiyi Liao*
- 171 Efficient Long Video Tokenization via Coordinate-based Patch
Reconstruction, *Huiwon Jang, Sihyun Yu, Jinwoo Shin, Pieter Abbeel, Younggyo Seo*
- 172 MotionStone: Decoupled Motion Intensity Modulation with
Diffusion Transformer for Image-to-Video Generation, *Shuwei Shi, Biao Gong, Xi Chen, Dandan Zheng, Shuai Tan, Zizheng Yang, Yuyuan Li, Jingwen He, Kecheng Zheng, Jingdong Chen, Ming Yang, Yinqiang Zheng*
- 173 AC3D: Analyzing and Improving 3D Camera Control in Video
☆ Diffusion Transformers, *Sherwin Bahmani, Ivan Skorokhodov, Guocheng Qian, Aliaksandr Siarohin, Willi Menapace, Andrea Tagliasacchi, David B. Lindell, Sergey Tulyakov*
- 174 Using Diffusion Priors for Video Amodal Segmentation,
Kaihua Chen, Deva Ramanan, Tarasha Khurana
- 175 VideoSPatS: Video SPatiotemporal Splines for Disentangled
Occlusion, Appearance and Motion Modeling and Editing, *Juan Luis Gonzalez, Xu Yao, Alex Whelan, Kyle Olszewski, Hyeonwoo Kim, Pablo Garrido*
- 176 Video Motion Transfer with Diffusion Transformers,
Alexander Pondaven, Aliaksandr Siarohin, Sergey Tulyakov, Philip Torr, Fabio Pizzati
- 177 VidTwin: Video VAE with Decoupled Structure and Dynamics,
Yuchi Wang, Junliang Guo, Xinyi Xie, Tianyu He, Xu Sun, Jiang Bian
- 178 HyperNVD: Accelerating Neural Video Decomposition via
☆ Hypernetworks, *Maria Pilligua, Danna Xue, Javier Vazquez-Corral*
- 179 Hierarchical Flow Diffusion for Efficient Frame Interpolation,
Yang Hai, Guo Wang, Tan Su, Wenjie Jiang, Yinlin Hu
- 180 HomoGen: Enhanced Video Inpainting via Homography
Propagation and Diffusion, *Ding Ding, Yueming Pan, Ruoyu Feng, Qi Dai, Kai Qiu, Jianmin Bao, Chong Luo, Zhenzhong Chen*
- 181 From Slow Bidirectional to Fast Autoregressive Video Diffusion
Models, *Tianwei Yin, Qiang Zhang, Richard Zhang, William T. Freeman, Fredo Durand, Eli Shechtman, Xun Huang*
- 182 Blind Bitstream-corrupted Video Recovery via Metadata-guided
Diffusion Model, *Shuyun Wang, Hu Zhang, Xin Shen, Dadong Wang, Xin Yu*
- 183 VidSeg: Training-free Video Semantic Segmentation based
☆ on Diffusion Models, *Qian Wang, Abdelrahman Eldesokey, Mohit Mendiratta, Fangneng Zhan, Adam Kortylewski, Christian Theobalt, Peter Wonka*
- 184 Towards More General Video-based Deepfake Detection
through Facial Component Guided Adaptation for Foundation
Model, *Yue-Hua Han, Tai-Ming Huang, Kai-Lung Hua, Jun-Cheng Chen*
- 185 Parameterized Blur Kernel Prior Learning for Local Motion
☆ Deblurring, *Zhenxuan Fang, Fangfang Wu, Tao Huang, Le Dong, Weisheng Dong, Xin Li, Guangming Shi*
- 186 Around the World in 80 Timesteps: A Generative Approach to
Global Visual Geolocation, *Nicolas Dufour, Vicky Kalogeiton, David Picard, Loic Landrieu*
- 187 Cross-Rejective Open-Set SAR Image Registration, *Shasha Mao, Shiming Lu, Zhaolong Du, Licheng Jiao, Shuiping Gou, Luntian Mou, Xuequan Lu, Lin Xiong, Yimeng Zhang*
- 188 Meta-Learning Hyperparameters for Parameter Efficient Fine-
* Tuning, *Zichen Tian, Yaoyao Liu, Qianru Sun*
- 189 HyperFree: A Channel-adaptive and Tuning-free Foundation
Model for Hyperspectral Remote Sensing Imagery, *Jingtao Li, Yingyi Liu, Xinyu Wang, Yunning Peng, Chen Sun, Shaoyu Wang, Zhendong Sun, Tian Ke, Xiao Jiang, Tangwei Lu, Anran Zhao, Yanfei Zhong*
- 190 MINIMA: Modality Invariant Image Matching, *Jiangwei Ren, Xingyu Jiang, Zizhuo Li, Dingkang Liang, Xin Zhou, Xiang Bai*
- 191 U-Know-DiffPAN: An Uncertainty-aware Knowledge Distillation
☆ Diffusion Framework with Details Enhancement for PAN-
Sharpening, *Sungpyo Kim, Jeonghyeok Do, Jaehyup Lee, Munchurl Kim*
- 192 QMambaBSR: Burst Image Super-Resolution with Query State
Space Model, *Xin Di, Long Peng, Peizhe Xia, Wenbo Li, Renjing Pei, Yang Cao, Yang Wang, Zheng-Jun Zha*
- 193 Learning Hazing to Dehazing: Towards Realistic Haze Generation
for Real-World Image Dehazing, *Ruiyi Wang, Yushuo Zheng, Zicheng Zhang, Chunyi Li, Shuaicheng Liu, Guangtao Zhai, Xiaohong Liu*
- 194 ADD: Attribution-Driven Data Augmentation Framework for
Boosting Image Super-Resolution, *Ze-Yu Mi, Yu-Bin Yang*
- 195 Gyro-based Neural Single Image Deblurring, *Heemin Yang,*
☆ *Jaesung Rim, Seungyong Lee, Seung-Hwan Baek, Sunghyun Cho*
- 196 UHD-processor: Unified UHD Image Restoration with Progressive
Frequency Learning and Degradation-aware Prompts, *Yidi Liu, Dong Li, Xueyang Fu, Xin Lu, Jie Huang, Zheng-Jun Zha*
- 197 AutoLUT: LUT-Based Image Super-Resolution with Automatic
Sampling and Adaptive Residual Learning, *Yuheng Xu, Shijie Yang, Xin Liu, Jie Liu, Jie Tang, Gangshan Wu*
- 198 The Power of Context: How Multimodality Improves Image Super-
Resolution, *Kangfu Mei, Hossein Talebi, Mojtaba Ardakani, Vishal M. Patel, Peyman Milanfar, Mauricio Delbracio*
- 199 Arbitrary-steps Image Super-resolution via Diffusion
Inversion, *Zongsheng Yue, Kang Liao, Chen Change Loy*
- 200 Improving Diffusion Inverse Problem Solving with Decoupled
○ Noise Annealing, *Bingliang Zhang, Wenda Chu, Julius Berner, Chenlin Meng, Anima Anandkumar, Yang Song*
- 201 Understanding Multi-layered Transmission Matrices,
* *Anat Levin, Marina Alterman*
- 202 TSD-SR: One-Step Diffusion with Target Score Distillation for
Real-World Image Super-Resolution, *Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, Changqing Zou*
- 203 FiRe: Fixed-points of Restoration Priors for Solving Inverse
Problems, *Matthieu Terris, Ulugbek S. Kamilov, Thomas Moreau*
- 204 Acquire and then Adapt: Squeezing out Text-to-Image Model
for Image Restoration, *Junyuan Deng, Xinyi Wu, Yongxing Yang, Congchao Zhu, Song Wang, Zhenyao Wu*

- 205 Reconciling Stochastic and Deterministic Strategies for Zero-shot Image Restoration using Diffusion Model in Dual, *Chong Wang, Lanqing Guo, Zixuan Fu, Siyuan Yang, Hao Cheng, Alex C. Kot, Bihan Wen*
- 206 SoftShadow: Leveraging Soft Masks for Penumra-Aware Shadow Removal, *Xinrui Wang, Lanqing Guo, Xiyu Wang, Siyu Huang, Bihan Wen*
- 207 Finding Local Diffusion Schrödinger Bridge using Kolmogorov-Arnold Network, *Xingyu Qiu, Mengying Yang, Xinghua Ma, Fanding Li, Dong Liang, Gongning Luo, Wei Wang, Kuanquan Wang, Shuo Li*
- 208 Towards Enhanced Image Inpainting: Mitigating Unwanted
* Object Insertion and Preserving Color Consistency, *Yikai Wang, ☆ Chenjie Cao, Junqiu Yu, Ke Fan, Xiangyang Xue, Yanwei Fu*
- 209 Fitted Neural Lossless Image Compression, *Zhe Zhang, Zhenzhong Chen, Shan Liu*
- 210 Good, Cheap, and Fast: Overfitted Image Compression with
* Wasserstein Distortion, *Jona Ballé, Luca Versari, Emilien Dupont, Hyunjik Kim, Matthias Bauer*
- 211 CacheQuant: Comprehensively Accelerated Diffusion Models, *Xuwen Liu, Zhikai Li, Qingyi Gu*
- 212 Decouple-Then-Merge: Finetune Diffusion Models as Multi-Task Learning, *Qianli Ma, Xuefei Ning, Dongrui Liu, Li Niu, Linfeng Zhang*
- 213 Decoupling Training-Free Guided Diffusion by ADMM, *Yuyuan Zhang, Zehua Liu, Zenan Li, Zhaoyu Li, James J. Clark, Xujie Si*
- 214 UniAP: Unifying Inter- and Intra-Layer Automatic Parallelism by
☞ Mixed Integer Quadratic Programming, *Hao Lin, Ke Wu, Jie Li, Jun Li, Wu-Jun Li*
- 215 DiverseFlow: Sample-Efficient Diverse Mode Coverage in Flows, *Mashrur M. Morshed, Vishnu Boddeti*
- 216 PCM : Picard Consistency Model for Fast Parallel Sampling of Diffusion Models, *Junhyuk So, Jiwoong Shin, Chaeyeon Jang, Eunhyeok Park*
- 217 Decentralized Diffusion Models, *David McAllister, Matthew Tancik, Jiaming Song, Angjoo Kanazawa*
- 218 Collaborative Decoding Makes Visual Auto-Regressive Modeling Efficient, *Zigeng Chen, Xinyin Ma, Gongfan Fang, Xinchao Wang*
- 219 Easy-editable Image Vectorization with Multi-layer Multi-scale Distributed Visual Feature Embedding, *Ye Chen, Zhangli Hu, Zhongyin Zhao, Yupeng Zhu, Yue Shi, Yuxuan Xiong, Bingbing Ni*
- 220 SketchAgent: Language-Driven Sequential Sketch Generation,
☆ *Yael Vinker, Tamar Rott Shaham, Kristine Zheng, Alex Zhao, Judith E Fan, Antonio Torralba*
- 221 Animate and Sound an Image, *Xihua Wang, Ruihua Song, Chongxuan Li, Xin Cheng, Boyuan Li, Yihan Wu, Yuyue Wang, Hongteng Xu, Yunfeng Wang*
- 222 SketchVideo: Sketch-based Video Generation and Editing, *Feng-Lin Liu, Hongbo Fu, Xintao Wang, Weicai Ye, Pengfei Wan, Di Zhang, Lin Gao*
- 223 Image Referenced Sketch Colorization Based on Animation Creation Workflow, *Dingkun Yan, Xinrui Wang, Zhuoru Li, Suguru Saito, Yusuke Iwasawa, Yutaka Matsuo, Jiaxian Guo*
- 224 Unity in Diversity: Video Editing via Gradient-Latent Purification, *Junyu Gao, Kunlin Yang, Xuan Yao, Yufan Hu*
- 225 Schedule On the Fly: Diffusion Time Prediction for Faster and Better Image Generation, *Zilyu Ye, Zhiyang Chen, Tiancheng Li, Zemin Huang, Weijian Luo, Guo-Jun Qi*
- 226 Chebyshev Attention Depth Permutation Texture Network with Latent Texture Attribute Loss, *Ravishankar Evani, Deepu Rajan, Shangbo Mao*
- 227 HSI: A Holistic Style Injector for Arbitrary Style Transfer, *Shuhao Zhang, Hui Kang, Yang Liu, Fang Mei, Hongjuan Li*
- 228 StyleStudio: Text-Driven Style Transfer with Selective Control of Style Elements, *Mingkun Lei, Xue Song, Beier Zhu, Hao Wang, Chi Zhang*
- 229 ZoomLDM: Latent Diffusion Model for Multi-scale Image Generation, *Srikanth Yellapragada, Alexandros Graikos, Kostas Triaridis, Prateek Prasanna, Rajarsi Gupta, Joel Saltz, Dimitris Samaras*
- 230 Diffusion-4K: Ultra-High-Resolution Image Synthesis with Latent Diffusion Models, *Jinjin Zhang, Qiuyu Huang, Junjie Liu, Xiefan Guo, Di Huang*
- 231 Preserve or Modify? Context-Aware Evaluation for Balancing Preservation and Modification in Text-Guided Image Editing, *Yoonjeon Kim, Soohyun Ryu, Yeonsung Jung, Hyunkoo Lee, Joowon Kim, June Yong Yang, Jaeryong Hwang, Eunho Yang*
- 232 Visual Representation Learning through Causal Intervention for
* Controllable Image Editing, *Shanshan Huang, Haoxuan Li, Chunyuan Zheng, Lei Wang, Guorui Liao, Zhili Gong, Huayi Yang, Li Liu*
- 233 MetaWriter: Personalized Handwritten Text Recognition Using Meta-Learned Prompt Tuning, *Wenhao Gu, Li Gu, Chingyee Yee Suen, Yang Wang*
- 234 ACE: Anti-Editing Concept Erasure in Text-to-Image Models, *Zihao Wang, Yuxiang Wei, Fan Li, Renjing Pei, Hang Xu, Wangmeng Zuo*
- 235 Goku: Flow Based Video Generative Foundation Models,
* *Shoufa Chen, Chongjian Ge, Yuqi Zhang, Yida Zhang, Fengda Zhu, Hao Yang, Hongxiang Hao, Hui Wu, Zhichao Lai, Yifei Hu, Ting-Che Lin, Shilong Zhang, Fu Li, Chuan Li, Xing Wang, Yanghua Peng, Peize Sun, Ping Luo, Yi Jiang, Zehuan Yuan, Bingyue Peng, Xiaobing Liu*
- 236 Self-Cross Diffusion Guidance for Text-to-Image Synthesis of Similar Subjects, *Weimin Qiu, Jieke Wang, Meng Tang*
- 237 Adapting Text-to-Image Generation with Feature Difference Instruction for Generic Image Restoration, *Chao Wang, Hehe Fan, Huichen Yang, Sarvnaz Karimi, Lina Yao, Yi Yang*
- 238 Boost Your Human Image Generation Model via Direct Preference
* Optimization, *Sanghyeon Na, Yonggyu Kim, Hyunjoon Lee*
- 239 DesignDiffusion: High-Quality Text-to-Design Image Generation
○ with Diffusion Models, *Zhendong Wang, Jianmin Bao, Shuyang Gu, Dong Chen, Wengang Zhou, Houqiang Li*
- 240 One-Way Ticket: Time-Independent Unified Encoder for Distilling Text-to-Image Diffusion Models, *Senmao Li, Lei Wang, Kai Wang, Tao Liu, Jiehang Xie, Joost van de Weijer, Fahad Shahbaz Khan, Shiqi Yang, Yaxing Wang, Jian Yang*
- 241 Noise Diffusion for Enhancing Semantic Faithfulness in Text-to-Image Synthesis, *Boming Miao, Chunxiao Li, Xiaoxiao Wang, Andi Zhang, Rui Sun, Zizhe Wang, Yao Zhu*
- 242 LaTeXBlend: Scaling Multi-concept Customized Generation with
* Latent Textual Blending, *Jian Jin, Zhenbo Yu, Yang Shen, Zhenyong Fu, Jian Yang*
- 243 Minority-Focused Text-to-Image Generation via Prompt
○ Optimization, *Soobin Um, Jong Chul Ye*
- 244 Devil is in the Detail: Towards Injecting Fine Details of Image Prompt in Image Generation via Conflict-free Guidance and Stratified Attention, *Kyungmin Jo, Jooyeol Yun, Jaegul Choo*
- 245 Towards Better Alignment: Training Diffusion Models with
☆ Reinforcement Learning Against Sparse Rewards, *Zijing Hu, Fengda Zhang, Long Chen, Kun Kuang, Jiahui Li, Kaifeng Gao, Jun Xiao, Xin Wang, Wenwu Zhu*
- 246 CustAny: Customizing Anything from A Single Example,
○ *Lingjie Kong, Kai Wu, Chengming Xu, Xiaobin Hu, Wenhui Han, Jinlong Peng, Donghao Luo, Mengtian Li, Jiangning Zhang, Chengjie Wang, Yanwei Fu*
- 247 BizGen: Advancing Article-level Visual Text Rendering for
☆ Infographics Generation, *Yuyang Peng, Shishi Xiao, Keming Wu, Qisheng Liao, Bohan Chen, Kevin Lin, Danqing Huang, Ji Li, Yuhui Yuan*
- 248 Learning to Sample Effective and Diverse Prompts for Text-to-Image Generation, *Taeyoung Yun, Dinghuai Zhang, Jinkyoo Park, Ling Pan*
- 249 ReNeg: Learning Negative Embedding with Reward Guidance,
* *Xiaomin Li, Yixuan Liu, Takashi Isobe, Xu Jia, Qinpeng Cui, Dong Zhou, Dong Li, You He, Huchuan Lu, Zhongdao Wang, Emad Barsoum*
- 250 MIDI: Multi-Instance Diffusion for Single Image to 3D Scene Generation, *Zehuan Huang, Yuan-Chen Guo, Xingqiao An, Yunhan Yang, Yangguang Li, Zi-Xin Zou, Ding Liang, Xihui Liu, Yan-Pei Cao, Lu Sheng*

- 251 ROIControl: Boosting Instance Control for Visual Generation, *Yuchao Gu, Yipin Zhou, Yunfan Ye, Yixin Nie, Licheng Yu, Pingchuan Ma, Kevin Qinghong Lin, Mike Zheng Shou*
- 252 Turbo3D: Ultra-fast Text-to-3D Generation, *Hanzhe Hu, Tianwei Yin, Fujun Luan, Yiwei Hu, Hao Tan, Zexiang Xu, Sai Bi, Shubham Tulsiani, Kai Zhang*
- 253 WeGen: A Unified Model for Interactive Multimodal Generation as We Chat, *Zipeng Huang, Shaobin Zhuang, Canmiao Fu, Binxin Yang, Ying Zhang, Chong Sun, Zhizheng Zhang, Yali Wang, Chen Li, Zheng-Jun Zha*
- 254 Chat2SVG: Vector Graphics Generation with Large Language Models and Image Diffusion Models, *Ronghuan Wu, Wanchao Su, Jing Liao*
- 255 AesthetiQ: Enhancing Graphic Layout Design via Aesthetic-Aware Preference Alignment of Multi-modal Large Language Models, *Sohan Patnaik, Rishabh Jain, Balaji Krishnamurthy, Mausoom Sarkar*
- 256 Black-Box Forgery Attacks on Semantic Watermarks for Diffusion Models, *Andreas Müller, Denis Lukovnikov, Jonas Thietke, Asja Fischer, Erwin Quiring*
- 257 Design2GarmentCode: Turning Design Concepts to Tangible Garments Through Program Synthesis, *Feng Zhou, Ruiyang Liu, Chen Liu, Gaofeng He, Yong-Lu Li, Xiaogang Jin, Huamin Wang*
- 258 AnyDressing: Customizable Multi-Garment Virtual Dressing via Latent Diffusion Models, *Xinghui Li, Qichao Sun, Pengze Zhang, Fulong Ye, Zhichao Liao, Wanquan Feng, Songtao Zhao, Qian He*
- 259 Adv-CPG: A Customized Portrait Generation Framework with Facial Adversarial Attacks, *Junying Wang, Hongyuan Zhang, Yuan Yuan*
- 260 ICE: Intrinsic Concept Extraction from a Single Image via Diffusion Models, *Fernando Julio Cendra, Kai Han*
- 261 Multi-Group Proportional Representations for Text-to-Image Models, *Sangwon Jung, Alex Oesterling, Claudio Mayrink Verdun, Sajani Vithana, Taesup Moon, Flavio P. Calmon*
- 262 What Makes a Good Dataset for Knowledge Distillation?, *Logan Frank, Jim Davis*
- 263 STEREO: A Two-Stage Framework for Adversarially Robust Concept Erasing from Text-to-Image Diffusion Models, *Koushik Srivatsan, Fahad Shamshad, Muzammal Naseer, Vishal M. Patel, Karthik Nandakumar*
- 264 PersonaHOI: Effortlessly Improving Face Personalization in Human-Object Interaction Generation, *Xinting Hu, Haoran Wang, Jan Eric Lenssen, Bernt Schiele*
- 265 Mind the Trojan Horse: Image Prompt Adapter Enabling Scalable and Deceptive Jailbreaking, *Junxi Chen, Junhao Dong, Xiaohua Xie*
- 266 Derivative-Free Diffusion Manifold-Constrained Gradient for Unified XAI, *Won Jun Kim, Hyungjin Chung, Jaemin Kim, Sangmin Lee, Byeongsu Sim, Jong Chul Ye*
- 267 Edit Away and My Face Will not Stay: Personal Biometric Defense against Malicious Generative Editing, *Hanhui Wang, Yihua Zhang, Ruizheng Bai, Yue Zhao, Sijia Liu, Zhengzhong Tu*
- 268 Training Data Provenance Verification: Did Your Model Use Synthetic Data from My Generative Model for Training?, *Yuechen Xie, Jie Song, Huiqiong Wang, Mingli Song*
- 269 Towards Universal AI-Generated Image Detection by Variational Information Bottleneck Network, *Haifeng Zhang, Qinghui He, Xiuli Bi, Weisheng Li, Bo Liu, Bin Xiao*
- 270 NightAdapter: Learning a Frequency Adapter for Generalizable Night-time Scene Segmentation, *Qi Bi, Jingjun Yi, Huimin Huang, Hao Zheng, Haolan Zhan, Yawen Huang, Yuexiang Li, Xian Wu, Yefeng Zheng*
- 271 D³: Scaling Up Deepfake Detection by Learning from Discrepancy, *Yongqi Yang, Zhihao Qian, Ye Zhu, Olga Russakovsky, Yu Wu*
- 272 Wavelet and Prototype Augmented Query-based Transformer for Pixel-level Surface Defect Detection, *Feng Yan, Xiaoheng Jiang, Yang Lu, Jiale Cao, Dong Chen, Mingliang Xu*
- 273 Neuro-3D: Towards 3D Visual Decoding from EEG Signals, *Zhanqiang Guo, Jiamin Wu, Yonghao Song, Jiahui Bu, Weijian Mai, Qihao Zheng, Wanli Ouyang, Chunfeng Song*
- 274 Spectral State Space Model for Rotation-Invariant Visual Representation Learning, *Sahar Dastani, Ali Bahri, Moslem Yazdanpanah, Mehrdad Noori, David Osowiechi, Gustavo Adolfo Vargas Hakim, Farzad Beizae, Milad Cheraghalikhani, Arnab Kumar Mondal, Herve Lombaert, Christian Desrosiers*
- 275 3D Prior Is All You Need: Cross-Task Few-shot 2D Gaze Estimation, *Yihua Cheng, Hengfei Wang, Zhongqun Zhang, Yang Yue, Boeun Kim, Feng Lu, Hyung Jin Chang*
- 276 HD-EPIC: A Highly-Detailed Egocentric Video Dataset, *Toby Perrett, Ahmad Darkhalil, Saptarshi Sinha, Omar Emara, Sam Pollard, Kranti Kumar Parida, Kaiting Liu, Prajwal Gatti, Siddhant Bansal, Kevin Flanagan, Jacob Chalk, Zhifan Zhu, Rhodri Guerrier, Fahd Abdelazim, Bin Zhu, Davide Moltisanti, Michael Wray, Hazel Doughty, Dima Damen*
- 277 Customized Condition Controllable Generation for Video Soundtrack, *Fan Qi, Kunsheng Ma, Changsheng Xu*
- 278 Learning to Highlight Audio by Watching Movies, *Chao Huang, Ruohan Gao, J. M. F. Tsang, Jan Kurcius, Cagdas Bilen, Chenliang Xu, Anurag Kumar, Sanjeel Parekh*
- 279 Supervising Sound Localization by In-the-wild Egomotion, *Anna Min, Ziyang Chen, Hang Zhao, Andrew Owens*
- 280 TSAM: Temporal SAM Augmented with Multimodal Prompts for Referring Audio-Visual Segmentation, *Abduljalil Radman, Jorma Laaksonen*
- 281 Audio-Visual Semantic Graph Network for Audio-Visual Event Localization, *Liang Liu, Shuaiyong Li, Yongqiang Zhu*
- 282 Language-Guided Audio-Visual Learning for Long-Term Sports Assessment, *Huangbiao Xu, Xiao Ke, Huanqi Wu, Rui Xu, Yuezhou Li, Wenzhong Guo*
- 283 Mimic: Improving Video Diffusion Models for Precise Text Understanding, *Shuai Tan, Biao Gong, Yutong Feng, Kecheng Zheng, Dandan Zheng, Shuwei Shi, Yujun Shen, Jingdong Chen, Ming Yang*
- 284 Mind the Time: Temporally-Controlled Multi-Event Video Generation, *Ziyi Wu, Aliaksandr Siarohin, Willi Menapace, Ivan Skorokhodov, Yuwei Fang, Varnith Chordia, Igor Gilitschenski, Sergey Tulyakov*
- 285 HOI-Gen-1M: A Large-scale Dataset for Human-Object Interaction Video Generation, *Kun Liu, Qi Liu, Xinchun Liu, Jie Li, Yongdong Zhang, Jiebo Luo, Xiaodong He, Wu Liu*
- 286 Change3D: Revisiting Change Detection and Captioning from A Video Modeling Perspective, *Duowang Zhu, Xiaohu Huang, Haiyan Huang, Hao Zhou, Zhenfeng Shao*
- 287 DejaVid: Encoder-Agnostic Learned Temporal Matching for Video Classification, *Darryl Ho, Samuel Madden*
- 288 When the Future Becomes the Past: Taming Temporal Correspondence for Self-supervised Video Representation Learning, *Yang Liu, Qianqian Xu, Peisong Wen, Siran Dai, Qingming Huang*
- 289 Dispider: Enabling Video LLMs with Active Real-Time Interaction via Disentangled Perception, Decision, and Reaction, *Rui Qian, Shuangrui Ding, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Dahua Lin, Jiaqi Wang*
- 290 Protecting Your Video Content: Disrupting Automated Video-based LLM Annotations, *Haitong Liu, Kuofeng Gao, Yang Bai, Jinmin Li, Jinxiao Shan, Tao Dai, Shu-Tao Xia*
- 291 DeCafNet: Delegate and Conquer for Efficient Temporal Grounding in Long Videos, *Zijia Lu, A S M Iftekhar, Gaurav Mittal, Tianjian Meng, Xiawei Wang, Cheng Zhao, Rohith Kukkala, Ehsan Elhamifar, Mei Chen*
- 292 Narrating the Video: Boosting Text-Video Retrieval via Comprehensive Utilization of Frame-Level Captions, *Chan Hur, Jeong-hun Hong, Dong-hun Lee, Dabin Kang, Semin Myeong, Sang-hyo Park, Hyeyoung Park*
- 293 Cross-modal Causal Relation Alignment for Video Question Grounding, *Weixing Chen, Yang Liu, Binglin Chen, Jiandong Su, Yongsun Zheng, Liang Lin*
- 294 Can Text-to-Video Generation help Video-Language Alignment?, *Luca Zanella, Massimiliano Mancini, Willi Menapace, Sergey Tulyakov, Yiming Wang, Elisa Ricci*

- 295 Video-MME: The First-Ever Comprehensive Evaluation Benchmark
* of Multi-modal LLMs in Video Analysis, *Chaoyou Fu,*
☆ *Yuhan Dai, Yongdong Luo, Lei Li, Shuhuai Ren, Renrui Zhang, Zihan Wang, Chenyu Zhou, Yunhang Shen, Mengdan Zhang, Peixian Chen, Yanwei Li, Shaohui Lin, Sirui Zhao, Ke Li, Tong Xu, Xiwu Zheng, Enhong Chen, Caifeng Shan, Ran He, Xing Sun*
- 296 Video-Panda: Parameter-efficient Alignment for Encoder-free Video-Language Models, *Jinhui Yi, Syed Talal Wasim, Yanan Luo, Muzammal Naseer, Juergen Gall*
- 297 Omnia de EgoTempo: Benchmarking Temporal Understanding of Multi-Modal LLMs in Egocentric Videos, *Chiara Plizzari, Alessio Tonioni, Yongqin Xian, Achin Kulshrestha, Federico Tombari*
- 298 Weakly Supervised Temporal Action Localization via Dual-Prior Collaborative Learning Guided by Multimodal Large Language Models, *Quan Zhang, Jinwei Fang, Rui Yuan, Xi Tang, Yuxin Qi, Ke Zhang, Chun Yuan*
- 299 MultiVENT 2.0: A Massive Multilingual Benchmark for Event-Centric Video Retrieval, *Reno Kriz, Kate Sanders, David Etter, Kenton Murray, Cameron Carpenter, Hannah Recknor, Jimena Guallar-Blasco, Alexander Martin, Eugene Yang, Benjamin Van Durme*
- 300 Efficient Motion-Aware Video MLLM, *Zijia Zhao, Yuqi Huo,*
* *Tongtian Yue, Longteng Guo, Haoyu Lu, Bingning Wang, Weipeng Chen, Jing Liu*
- 301 Building a Mind Palace: Structuring Environment-Grounded Semantic Graphs for Effective Long Video Analysis with LLMs,
☆ *Zeyi Huang, Yuyang Ji, Xiaofang Wang, Nikhil Mehta, Tong Xiao, Donghyun Lee, Sigmund Vanvalkenburgh, Shengxin Zha, Bolin Lai, Licheng Yu, Ning Zhang, Yong Jae Lee, Miao Liu*
- 302 Anchor-Aware Similarity Cohesion in Target Frames Enables Predicting Temporal Moment Boundaries in 2D, *Jiawei Tan, Hongxing Wang, Junwu Weng, Jiaxin Li, Zhilong Ou, Kang Dang*
- 303 Object-Shot Enhanced Grounding Network for Egocentric Video, *Yisen Feng, Haoyu Zhang, Meng Liu, Weili Guan, Liqiang Nie*
- 304 Black Swan: Abductive and Defeasible Video Reasoning in Unpredictable Events, *Aditya Chinchure, Sahithya Ravi, Raymond Ng, Vered Shwartz, Boyang Li, Leonid Sigal*
- 305 Exploiting Temporal State Space Sharing for Video Semantic Segmentation, *Syed Arif Syed Hesham, Yun Liu, Guolei Sun, Henghui Ding, Jing Yang, Ender Konukoglu, Xue Geng, Xudong Jiang*
- 306 Multi-modal Knowledge Distillation-based Human Trajectory Forecasting, *Jaewoo Jeong, Seohee Lee, Daehee Park, Giwon Lee, Kuk-Jin Yoon*
- 307 EntitySAM: Segment Everything in Video, *Mingqiao Ye, Seoung Wug Oh, Lei Ke, Joon-Young Lee*
- 308 SLADE: Shielding against Dual Exploits in Large Vision-Language Models, *Md Zarif Hossain, Ahmed Imteaj*
- 309 A Distractor-Aware Memory for Visual Object Tracking with SAM2, *Jovana Videnovic, Alan Lukezic, Matej Kristan*
- 310 Just Dance with pi! A Poly-modal Inductor for Weakly-supervised Video Anomaly Detection, *Snehasis Majhi, Giacomo D'Amicantonio, Antitza Dantcheva, Quan Kong, Lorenzo Garattoni, Gianpiero Francesca, Egor Bondarev, Francois Bremond*
- 311 Fish-Vista: A Multi-Purpose Dataset for Understanding & Identification of Traits from Images, *Kazi Sajeed Mehrab, M. Maruf, Arka Daw, Abhilash Neog, Harish Babu Manogaran, Mridul Khurana, Zhenyang Feng, Bahadır Altintas, Yasin Bakis, Elizabeth G Campolongo, Matthew J Thompson, Xiaojun Wang, Hilmar Lapp, Tanya Berger-Wolf, Paula Mabee, Henry Bart, Wei-Lun Chao, Wasila M Dahdul, Anuj Karpatne*
- 312 DiGIT: Multi-Dilated Gated Encoder and Central-Adjacent Region Integrated Decoder for Temporal Action Detection Transformer, *Ho-Joong Kim, Yearang Lee, Jung-Ho Hong, Seong-Whan Lee*
- 313 LLAVIDAL: A Large LAnguage VIsion Model for Daily Activities of Living, *Dominick Reilly, Rajat Subhra Chakraborty, Arkaprava Sinha, Manish Kumar Govind, Pu Wang, Francois Bremond, Le Xue, Srijan Das*
- 314 Are Spatial-Temporal Graph Convolution Networks for Human Action Recognition Over-Parameterized?, *Jiayang Xie, Yitian Zhao, Yanda Meng, He Zhao, Anh Nguyen, Yalin Zheng*
- 315 VSNet: Focusing on the Linguistic Characteristics of Sign Language, *Yuhao Li, Xinyue Chen, Hongkai Li, Xiaorong Pu, Peng Jin, Yazhou Ren*
- 316 Instant Adversarial Purification with Adversarial Consistency Distillation, *Chun Tong Lei, Hon Ming Yam, Zhongliang Guo, Yifei Qian, Chun Pong Lau*
- 317 Low-Rank Adaptation in Multilinear Operator Networks for Security-Preserving Incremental Learning, *Huu Binh Ta, Duc Nguyen, Quyen Tran, Toan Tran, Tung Pham*
- 318 Optimal Transport-Guided Source-Free Adaptation for Face Anti-Spoofing, *Zhuowei Li, Tianchen Zhao, Xiang Xu, Zheng Zhang, Zhihua Li, Xuanbai Chen, Qin Zhang, Alessandro Bergamo, Anil K. Jain, Yifan Xing*
- 319 FSFM: A Generalizable Face Security Foundation Model via Self-Supervised Facial Representation Learning, *Gaojian Wang, Feng Lin, Tong Wu, Zhenguang Liu, Zhongjie Ba, Kui Ren*
- 320 Test-Time Backdoor Detection for Object Detection Models, *Hangtao Zhang, Yichen Wang, Shihui Yan, Chenyu Zhu, Ziqi Zhou, Linshan Hou, Shengshan Hu, Minghui Li, Yanjun Zhang, Leo Yu Zhang*
- 321 Inference-Scale Complexity in ANN-SNN Conversion for High-Performance and Low-Power Applications, *Tong Bu, Maohua Li, Zhaofei Yu*
- 322 Spiking Transformer: Introducing Accurate Addition-Only Spiking Self-Attention for Transformer, *Yufei Guo, Xiaode Liu, Yuanpei Chen, Weihang Peng, Yuhan Zhang, Zhe Ma*
- 323 From Poses to Identity: Training-Free Person Re-Identification via Feature Centralization, *Chao Yuan, Guiwei Zhang, Changxiao Ma, Tianyi Zhang, Guanglin Niu*
- 324 Learning from Synchronization: Self-Supervised Uncalibrated Multi-View Person Association in Challenging Scenes, *Keqi Chen, Vinkle Srivastav, Didier Mutter, Nicolas Padoy*
- 325 ReDiffDet: Rotation-equivariant Diffusion Model for Oriented Object Detection, *Jiaqi Zhao, Zeyu Ding, Yong Zhou, Hancheng Zhu, Wen-Liang Du, Rui Yao*
- 326 Taste More, Taste Better: Diverse Data and Strong Model Boost Semi-Supervised Crowd Counting, *Maochen Yang, Zekun Li, Jian Zhang, Lei Qi, Yinghuan Shi*
- 327 SmartEraser: Remove Anything from Images using Masked-Region Guidance, *Longtao Jiang, Zhendong Wang, Jianmin Bao, Wengang Zhou, Dongdong Chen, Lei Shi, Dong Chen, Houqiang Li*
- 328 Towards Generalizable Scene Change Detection, *Jae-Woo Kim, Ue-Hwan Kim*
- 329 SUM Parts: Benchmarking Part-Level Semantic Segmentation of Urban Meshes, *Weixiao Gao, Liangliang Nan, Hugo Ledoux*
- 330 Scene-Centric Unsupervised Panoptic Segmentation,
* *Oliver Hahn, Christoph Reich, Nikita Araslanov, Daniel Cremers,*
☆ *Christian Rupprecht, Stefan Roth*
- 331 Foveated Instance Segmentation, *Hongyi Zeng, Wenxuan Liu, Tianhua Xia, Jinhui Chen, Ziyun Li, Sai Qian Zhang*
- 332 Zero-Shot 4D Lidar Panoptic Segmentation, *Yushan Zhang,*
☆ *Aljoša Ošep, Laura Leal-Taixé, Tim Meinhardt*
- 333 Repurposing Stable Diffusion Attention for Training-Free Unsupervised Interactive Segmentation, *Markus Karmann, Onay Urfalioglu*
- 334 Cross-Modal Distillation for 2D/3D Multi-Object Discovery from 2D Motion, *Saad Lahlali, Sandra Kara, Hejer Ammar, Florian Chabot, Nicolas Granger, Hervé Le Borgne, Quoc-Cuong Pham*
- 335 Learning 4D Panoptic Scene Graph Generation from Rich 2D Visual Scene, *Shengqiong Wu, Hao Fei, Jingkang Yang,*
☆ *Xiangtai Li, Juncheng Li, Hanwang Zhang, Tat-seng Chua*
- 336 Functionality Understanding and Segmentation in 3D Scenes, *Jaime Corsetti, Francesco Giuliari, Alice Fasoli,*
☆ *Davide Boscaini, Fabio Poiesi*
- 337 3D Student Splatting and Scooping, *Jialin Zhu, Jiangbei Yue,*
🎨 *Feixiang He, He Wang*

- 338 Chain of Semantics Programming in 3D Gaussian Splatting Representation for 3D Vision Grounding, *Jiaxin Shi, Mingyue Xiang, Hao Sun, Yixuan Huang, Zhi Weng*
- 339 Unveiling the Mist over 3D Vision-Language Understanding: Object-centric Evaluation with Chain-of-Analysis, *Jiangyong Huang, Baoxiong Jia, Yan Wang, Ziyu Zhu, Xiongkun Linghu, Qing Li, Song-Chun Zhu, Siyuan Huang*
- 340 ProxyTransformation: Preshaping Point Cloud Manifold With Proxy Attention For 3D Visual Grounding, *Qihang Peng, Henry Zheng, Gao Huang*
- 341 ECBench: Can Multi-modal Foundation Models Understand the Egocentric World? A Holistic Embodied Cognition Benchmark, *Ronghao Dang, Yuqian Yuan, Wenqi Zhang, Yifei Xin, Boqiang Zhang, Long Li, Liuyi Wang, Qinyang Zeng, Xin Li, Lidong Bing*
- 342 TANGO: Training-free Embodied AI Agents for Open-world Tasks, ☆ *Filippo Ziliotto, Tommaso Campari, Luciano Serafini, Lamberto Ballan*
- 343 ComfyBench: Benchmarking LLM-based Agents in ComfyUI for Autonomously Designing Collaborative AI Systems, *Xiangyuan Xue, Zeyu Lu, Di Huang, Zidong Wang, Wanli Ouyang, Lei Bai*
- 344 The Scene Language: Representing Scenes with Programs, * *Words, and Embeddings, Yunzhi Zhang, Zizhang Li, Matt Zhou, Shangzhe Wu, Jiajun Wu*
- 345 Ground-V: Teaching VLMs to Ground Complex Instructions in Pixels, *Yongshuo Zong, Qin Zhang, Dongsheng An, Zhihua Li, Xiang Xu, Linghan Xu, Zhuowen Tu, Yifan Xing, Onkar Dabeer*
- 346 ViUNIT: Visual Unit Tests for More Robust Visual Programming, *Artemis Panagopoulou, Honglu Zhou, Silvio Savarese, Caiming Xiong, Chris Callison-Burch, Mark Yatskar, Juan Carlos Niebles*
- 347 VL-RewardBench: A Challenging Benchmark for Vision-Language * *Generative Reward Models, Lei Li, Yuancheng Wei, Zhihui Xie, Xuqing Yang, Yifan Song, Peiyi Wang, Chenxin An, Tianyu Liu, Sujian Li, Bill Yuchen Lin, Lingpeng Kong, Qi Liu*
- 348 Spatial457: A Diagnostic Benchmark for 6D Spatial Reasoning of * *Large Multimodal Models, Xingrui Wang, Wufei Ma, Tiezheng Zhang, Celso M de Melo, Jieneng Chen, Alan Yuille*
- 349 RANGE: Retrieval Augmented Neural Fields for Multi-Resolution Geo-Embeddings, *Aayush Dhakal, Srikumar Sastry, Subash Khanal, Adeel Ahmad, Eric Xing, Nathan Jacobs*
- 350 EmoEdit: Evoking Emotions through Image Manipulation, *Jingyuan Yang, Jiawei Feng, Weibin Luo, Dani Lischinski, Daniel Cohen-Or, Hui Huang*
- 351 Uncertain Multimodal Intention and Emotion Understanding in the Wild, *Qu Yang, Qinghongya Shi, Tongxin Wang, Mang Ye*
- 352 F-LMM: Grounding Frozen Large Multimodal Models, *Size Wu, Sheng Jin, Wenwei Zhang, Lumin Xu, Wentao Liu, Wei Li, Chen Change Loy*
- 353 Reasoning to Attend: Try to Understand How <SEG> Token Works, *Rui Qian, Xin Yin, Dejing Dou*
- 354 MIMO: A Medical Vision Language Model with Visual Referring Multimodal Input and Pixel Grounding Multimodal Output, *Yanyuan Chen, Dexuan Xu, Yu Huang, Songkun Zhan, Hanpin Wang, Dongxue Chen, Xueping Wang, Meikang Qiu, Hang Li*
- 355 DynRefer: Delving into Region-level Multimodal Tasks via Dynamic Resolution, *Yuzhong Zhao, Feng Liu, Yue Liu, Mingxiang Liao, Chen Gong, Qixiang Ye, Fang Wan*
- 356 Separation of Powers: On Segregating Knowledge from Observation in LLM-enabled Knowledge-based Visual Question Answering, *Zhen Yang, Zhuo Tao, Qi Chen, Liang Li, Yuankai Qi, Anton van den Hengel, Qingming Huang*
- 357 FINECAPTION: Compositional Image Captioning Focusing on ☆ *Wherever You Want at Any Granularity, Hang Hua, Qing Liu, Lingzhi Zhang, Jing Shi, Soo Ye Kim, Zhifei Zhang, Yilin Wang, Jianming Zhang, Zhe Lin, Jiebo Luo*
- 358 AlignMamba: Enhancing Multimodal Mamba with Local and Global Cross-modal Alignment, *Yan Li, Yifei Xing, Xiangyuan Lan, Xin Li, Haifeng Chen, Dongmei Jiang*
- 359 Missing Target-Relevant Information Prediction with World Model for Accurate Zero-Shot Composed Image Retrieval, *Yuanmin Tang, Jing Yu, Keke Gai, Jiamin Zhuang, Gang Xiong, Gaopeng Gou, Qi Wu*
- 360 SynTab-LLaVA: Enhancing Multimodal Table Understanding with Decoupled Synthesis, *Bangbang Zhou, Zuan Gao, Zixiao Wang, Boqiang Zhang, Yuxin Wang, Zhineng Chen, Hongtao Xie*
- 361 The Photographer's Eye: Teaching Multimodal Large Language Models to See, and Critique Like Photographers, *Daiqing Qi, Handong Zhao, Jing Shi, Simon Jenni, Yifei Fan, Franck Derroncourt, Scott Cohen, Sheng Li*
- 362 Document Haystacks: Vision-Language Reasoning Over Piles of 1000+ Documents, *Jun Chen, Dannong Xu, Junjie Fei, Chun-Mei Feng, Mohamed Elhoseiny*
- 363 VDocRAG: Retrieval-Augmented Generation over Visually-Rich Documents, *Ryota Tanaka, Taichi Iki, Taku Hasegawa, Kyosuke Nishida, Kuniko Saito, Jun Suzuki*
- 364 OmniDocBench: Benchmarking Diverse PDF Document Parsing with Comprehensive Annotations, *Linke Ouyang, Yuan Qu, Hongbin Zhou, Jiawei Zhu, Rui Zhang, Qunshu Lin, Bin Wang, Zhiyuan Zhao, Man Jiang, Xiaomeng Zhao, Jin Shi, Fan Wu, Pei Chu, Minghao Liu, Zhenxiang Li, Chao Xu, Bo Zhang, Botian Shi, Zhongying Tu, Conghui He*
- 365 Enhancing Vision-Language Compositional Understanding with Multimodal Synthetic Data, *Haoxin Li, Boyang Li*
- 366 Seeing What Matters: Empowering CLIP with Patch Generation-to-Selection, *Gensheng Pei, Tao Chen, Yujia Wang, Xinhao Cai, Xiangbo Shu, Tianfei Zhou, Yazhou Yao*
- 367 CLIP is Almost All You Need: Towards Parameter-Efficient Scene Text Retrieval without OCR, *Xugong Qin, Peng Zhang, Jun Jie Ou Yang, Gangyan Zeng, Yubo Li, Yuanyuan Wang, Wanqian Zhang, Pengwen Dai*
- 368 FLAIR: VLM with Fine-grained Language-informed Image Representations, *Rui Xiao, Sanghwan Kim, Mariana-Iuliana Georgescu, Zeynep Akata, Stephan Alaniz*
- 369 Retaining Knowledge and Enhancing Long-Text Representations in CLIP through Dual-Teacher Distillation, *Yuheng Feng, Changsong Wen, Zelin Peng, Li Jiaye, Siyu Zhu*
- 370 DINOv2 Meets Text: A Unified Framework for Image- and Pixel-Level Vision-Language Alignment, *Cijo Jose, Théo Moutakanni, Dahyun Kang, Federico Baldassarre, Timothée Darcet, Hu Xu, Daniel Li, Marc Szafraniec, Michaël Ramamonjisoa, Maxime Oquab, Oriane Siméoni, Huy V. Vo, Patrick Labatut, Piotr Bojanowski*
- 371 Not Only Text: Exploring Compositionality of Visual Representations * *in Vision-Language Models, Davide Berasi, Matteo Farina, ☆ Massimiliano Mancini, Elisa Ricci, Nicola Strisciuglio*
- 372 Florence-VL: Enhancing Vision-Language Models with Generative Vision Encoder and Depth-Breadth Fusion, *Jiahai Chen, Jianwei Yang, Haiping Wu, Dianqi Li, Jianfeng Gao, Tianyi Zhou, Bin Xiao*
- 373 PVC: Progressive Visual Token Compression for Unified Image and Video Processing in Large Vision-Language Models, *Chenyu Yang, Xuan Dong, Xizhou Zhu, Weijie Su, Jiahao Wang, Hao Tian, Zhe Chen, Wenhai Wang, Lewei Lu, Jifeng Dai*
- 374 Beyond Sight: Towards Cognitive Alignment in LVLM via Enriched Visual Knowledge, *Yaqi Zhao, Yuanyang Yin, Lin Li, Mingan Lin, Victor Shea-Jay Huang, Siwei Chen, Weipeng Chen, Baoqun Yin, Zenan Zhou, Wentao Zhang*
- 375 Mono-InternVL: Pushing the Boundaries of Monolithic Multimodal Large Language Models with Endogenous Visual Pre-training, *Gen Luo, Xue Yang, Wenhan Dou, Zhaokai Wang, Jiawen Liu, Jifeng Dai, Yu Qiao, Xizhou Zhu*
- 376 ATP-LLaVA: Adaptive Token Pruning for Large Vision Language Models, *Xubing Ye, Yukang Gan, Yixiao Ge, Xiao-Ping Zhang, Yansong Tang*
- 377 It's a (Blind) Match! Towards Vision-Language Correspondence without Parallel Data, ☆ *Dominik Schnaus, Nikita Araslanov, Daniel Cremers*
- 378 DH-Set: Improving Vision-Language Alignment with Diverse and Hybrid Set-Embeddings Learning, *Kun Zhang, Jingyu Li, Zhe Li, S. Kevin Zhou*

- 379 Devils in Middle Layers of Large Vision-Language Models: Interpreting, Detecting and Mitigating Object Hallucinations via Attention Lens, *Zhangqi Jiang, Junkai Chen, Beier Zhu, Tingjin Luo, Yankun Shen, Xu Yang*
- 380 MMRL: Multi-Modal Representation Learning for Vision-Language Models, *Yuncheng Guo, Xiaodong Gu*
- 381 LoRA Recycle: Unlocking Tuning-Free Few-Shot Adaptability in Visual Foundation Models by Recycling Pre-Tuned LoRAs, *Zixuan Hu, Yongxian Wei, Li Shen, Chun Yuan, Dacheng Tao*
- 382 Immune: Improving Safety Against Jailbreaks in Multi-modal LLMs via Inference-Time Alignment, *Soumya Suvra Ghosal, Souradip Chakraborty, Vaibhav Singh, Tianrui Guan, Mengdi Wang, Ahmad Beirami, Furong Huang, Alvaro Velasquez, Dinesh Manocha, Amrit Singh Bedi*
- 383 SceneTAP: Scene-Coherent Typographic Adversarial Planner against Vision-Language Models in Real-World Environments, *Yue Cao, Yun Xing, Jie Zhang, Di Lin, Tianwei Zhang, Ivor Tsang, Yang Liu, Qing Guo*
- 384 Stealthy Backdoor Attack in Self-Supervised Learning Vision Encoders for Large Vision Language Models, *Zhaoyi Liu, Huan Zhang*
- 385 Improving Adversarial Transferability on Vision Transformers via Forward Propagation Refinement, *Yuchen Ren, Zhengyu Zhao, Chenhao Lin, Bo Yang, Lu Zhou, Zhe Liu, Chao Shen*
- 386 PARC: A Quantitative Framework Uncovering the Symmetries within Vision Language Models, *Jenny Schmalfuss, Nadine Chang, Vibashan VS, Maying Shen, Andres Bruhn, Jose M. Alvarez*
- 387 ProKeR: A Kernel Perspective on Few-Shot Adaptation of Large Vision-Language Models, *Yassir Bendou, Amine Ouasfi, Vincent Gripon, Adnane Boukhayma*
- 388 Realistic Test-Time Adaptation of Vision-Language Models, *Maxime Zanella, Clément Fuchs, Christophe De Vleeschouwer, Ismail Ben Ayed*
- 389 Low-Biased General Annotated Dataset Generation, *Dengyang Jiang, Haoyu Wang, Lei Zhang, Wei Wei, Guang Dai, Mengmeng Wang, Jingdong Wang, Yanning Zhang*
- 390 Joint Scheduling of Causal Prompts and Tasks for Multi-Task Learning, *Chaoyang Li, Jianyang Qin, Jinhao Cui, Zeyu Liu, Ning Hu, Qing Liao*
- 391 Beyond Words: Augmenting Discriminative Richness via Diffusions in Unsupervised Prompt Learning, *Hairui Ren, Fan Tang, He Zhao, Zixuan Wang, Dandan Guo, Yi Chang*
- 392 ProAPO: Progressively Automatic Prompt Optimization for Visual Classification, *Xiangyan Qu, Gaopeng Gou, Jiamin Zhuang, Jing Yu, Kun Song, Qihao Wang, Yili Li, Gang Xiong*
- 393 Embracing Collaboration Over Competition: Condensing Multiple Prompts for Visual In-Context Learning, *Jinpeng Wang, Tianci Luo, Yaohua Zha, Yan Feng, Ruisheng Luo, Bin Chen, Tao Dai, Long Chen, Yaowei Wang, Shu-Tao Xia*
- 394 Comprehensive Information Bottleneck for Unveiling Universal Attribution to Interpret Vision Transformers, *Jung-Ho Hong, Ho-Joong Kim, Kyu-Sung Jeon, Seong-Whan Lee*
- 395 CustomKD: Customizing Large Vision Foundation for Edge Model Improvement via Knowledge Distillation, *Jungsoo Lee, Debasmit Das, Munawar Hayat, Sungha Choi, Kyuwoong Hwang, Fatih Porikli*
- 396 Beyond Local Sharpness: Communication-Efficient Global Sharpness-aware Minimization for Federated Learning, *Debora Caldarola, Pietro Cagnasso, Barbara Caputo, Marco Ciccone*
- 397 Do ImageNet-trained Models Learn Shortcuts? The Impact of Frequency Shortcuts on Generalization, *Shunxin Wang, Raymond Veldhuis, Nicola Strisciuglio*
- 398 Minimal Interaction Separated Tuning: A New Paradigm for Visual Adaptation, *Ningyuan Tang, Minghao Fu, Jianxin Wu*
- 399 DiTASK: Multi-Task Fine-Tuning with Diffeomorphic Transformations, *Krishna Sri Ipsit Mantri, Carola-Bibiane Schönlieb, Bruno Ribeiro, Chaim Baskin, Moshe Eliasof*
- 400 Closest Neighbors are Harmful for Lightweight Masked Auto-encoders, *Jian Meng, Ahmed Hasssan, Li Yang, Deliang Fan, Jinwoo Shin, Jae-sun Seo*
- 401 GilaNet: Adaptive Neural Network Structure Learning with Gila-Driven, *Mengqiao Han, Liyuan Pan, Xiabi Liu*
- 402 From Alexnet to Transformers: Measuring the Non-linearity of Deep Neural Networks with Affine Optimal Transport, *☆ Quentin Bouniot, Ievgen Redko, Anton Mallasto, Charlotte Laclau, Oliver Struckmeier, Karol Arndt, Markus Heinonen, Ville Kyrki, Samuel Kaski*
- 403 MambaVision: A Hybrid Mamba-Transformer Vision Backbone, *Ali Hatamizadeh, Jan Kautz*
- 404 Breaking the Low-Rank Dilemma of Linear Attention, *Qihang Fan, Huaibo Huang, Ran He*
- 405 ShiftwiseConv: Small Convolutional Kernel with Large Kernel Effect, *Dachong Li, Li Li, Zhuangzhuang Chen, Jianqiang Li*
- 406 Star with Bilinear Mapping, *Zelin Peng, Yu Huang, Zhengqin Xu, Feilong Tang, Ming Hu, Xiaokang Yang, Wei Shen*
- 407 Your ViT is Secretly an Image Segmentation Model, *☆ Tommie Kerssies, Niccolò Cavagnero, Alexander Hermans, Narges Norouzi, Giuseppe Averta, Bastian Leibe, Gijs Dubbelman, Daan de Geus*
- 408 Samba: A Unified Mamba-based Framework for General Salient Object Detection, *☆ Jiahao He, Keren Fu, Xiaohong Liu, Qijun Zhao*
- 409 HORP: Human-Object Relation Priors Guided HOI Detection, *Pei Geng, Jian Yang, Shanshan Zhang*
- 410 T2ICount: Enhancing Cross-modal Understanding for Zero-Shot Counting, *☆ Yifei Qian, Zhongliang Guo, Bowen Deng, Chun Tong Lei, Shuai Zhao, Chun Pong Lau, Xiaopeng Hong, Michael P. Pound*
- 411 DPseg: Dual-Prompt Cost Volume Learning for Open-Vocabulary Semantic Segmentation, *Ziyu Zhao, Xiaoguang Li, Lingjia Shi, Nasrin Imanpour, Song Wang*
- 412 Text Augmented Correlation Transformer For Few-shot Classification & Segmentation, *Srinivasa Rao Nandam, Sara Atito, Zhenhua Feng, Josef Kittler, Muhammad Awaiz*
- 413 Golden Cudgel Network for Real-Time Semantic Segmentation, *Guoyu Yang, Yuan Wang, Daming Shi, Yanzhong Wang*
- 414 WISH: Weakly Supervised Instance Segmentation using Heterogeneous Labels, *☆ Hyeokjun Kweon, Kuk-Jin Yoon*
- 415 Hierarchical Compact Clustering Attention (COCA) for Unsupervised Object-Centric Learning, *Can Kucuksozen, Yucel Yemez*
- 416 Incremental Object Keypoint Learning, *Mingfu Liang, Jiahuan Zhou, Xu Zou, Ying Wu*
- 417 Logits DeConfusion with CLIP for Few-Shot Learning, *Shuo Li, Fang Liu, Zehua Hao, Xinyi Wang, Lingling Li, Xu Liu, Puhua Chen, Wenping Ma*
- 418 OCRT: Boosting Foundation Models in the Open World with Object-Concept-Relation Triad, *Luyao Tang, Yuxuan Yuan, Chaoqi Chen, Zeyu Zhang, Yue Huang, Kun Zhang*
- 419 Correlative and Discriminative Label Grouping for Multi-Label Visual Prompt Tuning, *Lei-Lei Ma, Shuo Xu, Ming-Kun Xie, Lei Wang, Dengdi Sun, Haifeng Zhao*
- 420 Adaptive Part Learning for Fine-Grained Generalized Category Discovery: A Plug-and-Play Enhancement, *Qiyuan Dai, Hanzhuo Huang, Yu Wu, Sibe Yang*
- 421 OW-OVD: Unified Open World and Open Vocabulary Object Detection, *Xing Xi, Yangyang Huang, Ronghua Luo, Yu Qiu*
- 422 SEEN-DA: SEmantic ENtropy guided Domain-aware Attention for Domain Adaptive Object Detection, *Haochen Li, Rui Zhang, Hantao Yao, Xin Zhang, Yifan Hao, Xinkai Song, Shaohui Peng, Yongwei Zhao, Chen Zhao, Yanjun Wu, Ling Li*
- 423 Detect Any Mirrors: Boosting Learning Reliability on Large-Scale Unlabeled Data with an Iterative Data Engine, *Zhaohu Xing, Lihao Liu, Yijun Yang, Hongqiu Wang, Tian Ye, Sixiang Chen, Wenxue Li, Guang Liu, Lei Zhu*
- 424 Project-Probe-Aggregate: Efficient Fine-Tuning for Group Robustness, *☆ Beier Zhu, Jiequan Cui, Hanwang Zhang, Chi Zhang*
- 425 Enhancing Diversity for Data-free Quantization, *☆ Kai Zhao, Zhihao Zhuang, Miao Zhang, Chenjuan Guo, Yang Shu, Bin Yang*

- 426 Gain from Neighbors: Boosting Model Robustness in the Wild
☆ via Adversarial Perturbations Toward Neighboring Classes,
Zhou Yang, Mingtao Feng, Tao Huang, Fangfang Wu, Weisheng Dong, Xin Li, Guangming Shi
- 427 Unseen Visual Anomaly Generation, *Han Sun, Yunkang Cao, Hao Dong, Olga Fink*
- 428 MANTA: A Large-Scale Multi-View and Visual-Text Anomaly Detection Dataset for Tiny Objects, *Lei Fan, Dongdong Fan, Zhiguang Hu, Yiwon Ding, Donglin Di, Kai Yi, Maurice Pagnucco, Yang Song*
- 429 A Unified Latent Schrödinger Bridge Diffusion Model for Unsupervised Anomaly Detection and Localization,
Shilhora Akshay, Niveditha Lakshmi Narasimhan, Jacob George, Vineeth N Balasubramanian
- 430 TailedCore: Few-Shot Sampling for Unsupervised Long-Tail Noisy Anomaly Detection, *Yoon Gyo Jung, Jaewoo Park, Jaeho Yoon, Kuan-Chuan Peng, Wonchul Kim, Andrew Beng Jin Teoh, Octavia Camps*
- 431 Potential Field Based Deep Metric Learning,
☆ *Shubhang Bhatnagar, Narendra Ahuja*
- 432 Inversion Circle Interpolation: Diffusion-based Image Augmentation for Data-scarce Classification, *Yanghai Wang, Long Chen*
- 433 Dataset Distillation with Neural Characteristic Function:
* A Minmax Perspective, *Shaobo Wang, Yicun Yang, Zhiyuan Liu, Chenghao Sun, Xuming Hu, Conghui He, Linfeng Zhang*
- 434 Towards Stable and Storage-efficient Dataset Distillation: Matching Convexified Trajectory, *Wenliang Zhong, Haoyu Tang, Qinghai Zheng, Mingzhu Xu, Yupeng Hu, Weili Guan*
- 435 Data Distributional Properties As Inductive Bias for Systematic Generalization, *Felipe del Rio, Alain Raymond-Saez, Daniel Florea, Rodrigo Toro Icarte, Julio Hurtado, Cristian B. Calderon, Alvaro Soto*
- 436 SoMA: Singular Value Decomposed Minor Components
* Adaptation for Domain Generalizable Representation Learning,
Seokju Yun, Seunghye Chae, Dongheon Lee, Youngmin Ro
- 437 BiLoRA: Almost-Orthogonal Parameter Spaces for Continual Learning, *Hao Zhu, Yifei Zhang, Junhao Dong, Piotr Koniusz*
- 438 DPC: Dual-Prompt Collaboration for Tuning Vision-Language Models, *Haoyang Li, Liang Wang, Chao Wang, Jing Jiang, Yan Peng, Guodong Long*
- 439 Effortless Active Labeling for Long-Term Test-Time Adaptation, *Guowei Wang, Changxing Ding*
- 440 SEC-Prompt: Semantic Complementary Prompting for Few-Shot Class-Incremental Learning, *Ye Liu, Meng Yang*
- 441 Attraction Diminishing and Distributing for Few-Shot Class-Incremental Learning, *Li-Jun Zhao, Zhen-Duo Chen, Yongxin Wang, Xin Luo, Xin-Shun Xu*
- 442 CoMBO: Conflict Mitigation via Branched Optimization for Class Incremental Segmentation, *Kai Fang, Anqi Zhang, Guangyu Gao, Jianbo Jiao, Chi Harold Liu, Yunchao Wei*
- 443 Geometric Knowledge-Guided Localized Global Distribution
☆ Alignment for Federated Learning, *Yanbiao Ma, Wei Dai, Wenke Huang, Jiayi Chen*
- 444 Joint Out-of-Distribution Filtering and Data Discovery Active Learning, *Sebastian Schmidt, Leonard Schenk, Leo Schwinn, Stephan Günnemann*
- 445 Revisiting Source-Free Domain Adaptation: Insights into Representativeness, Generalization, and Variety, *Ronghang Zhu, Mengxuan Hu, Weiming Zhuang, Lingjuan Lyu, Xiang Yu, Sheng Li*
- 446 Identifying and Mitigating Spurious Correlation in Multi-Task Learning, *Junyi Chai, Shenyu Lu, Xiaoqian Wang*
- 447 Language-Assisted Debiasing and Smoothing for Foundation Model-Based Semi-Supervised Learning, *Na Zheng, Xuemeng Song, Xue Dong, Aashish Nikhil Ghosh, Liqiang Nie, Roger Zimmermann*
- 448 Weakly Supervised Contrastive Adversarial Training for Learning Robust Features from Semi-supervised Data, *Lilin Zhang, Chengpei Wu, Ning Yang*
- 449 Dual Energy-Based Model with Open-World Uncertainty Estimation for Out-of-distribution Detection, *Qi Chen, Hu Ding*
- 450 Directional Label Diffusion Model for Learning from Noisy Labels, *Senyu Hou, Gaoxia Jiang, Jia Zhang, Shangrong Yang, Husheng Guo, Yaqing Guo, Wenjian Wang*
- 451 A Simple Data Augmentation for Feature Distribution Skewed Federated Learning, *Yunlu Yan, Huazhu Fu, Yuexiang Li, Jinheng Xie, Jun Ma, Guang Yang, Lei Zhu*
- 452 NoT: Federated Unlearning via Weight Negation, *Yasser H. Khalil, Leo Brunswic, Soufiane Lamghari, Xu Li, Mahdi Beitollahi, Xi Chen*
- 453 Infighting in the Dark: Multi-Label Backdoor Attack in Federated Learning, *Ye Li, Yanchao Zhao, Chengcheng Zhu, Jiale Zhang*
- 454 TAROT: Towards Essentially Domain-Invariant Robustness with Theoretical Justification, *Dongyoon Yang, Jihu Lee, Yongdai Kim*
- 455 Invisible Backdoor Attack against Self-supervised Learning, *Hanrong Zhang, Zhenting Wang, Boheng Li, Fulin Lin, Tingxu Han, Mingyu Jin, Chenlu Zhan, Mengnan Du, Hongwei Wang, Shiqing Ma*
- 456 Enhancing SAM with Efficient Prompting and
○ Preference Optimization for Semi-supervised Medical Image Segmentation, *Aishik Konwer, Zhijian Yang, Erhan Bas, Cao Xiao, Prateek Prasanna, Parminder Bhatia, Taha Kass-Hout*
- 457 Improving Transferable Targeted Attacks with Feature Tuning Mixup, *Kaisheng Liang, Xuelong Dai, Yanjie Li, Dong Wang, Bin Xiao*
- 458 TopoCellGen: Generating Histopathology Cell Topology with a
☆ Diffusion Model, *Meilong Xu, Saumya Gupta, Xiaoling Hu, Chen Li, Shahira Abousamra, Dimitris Samaras, Prateek Prasanna, Chao Chen*
- 459 Improving Accuracy and Calibration via Differentiated Deep Mutual Learning, *Han Liu, Peng Cui, Bingning Wang, Weipeng Chen, Yupeng Zhang, Jun Zhu, Xiaolin Hu*
- 460 SeqMvRL: A Sequential Fusion Framework for Multi-view Representation Learning, *Ren Wang, Haoliang Sun, Yuxiu Lin, Chuanhui Zuo, Yongshun Gong, Yilong Yin, Wenjia Meng*
- 461 Attribute-Missing Multi-view Graph Clustering, *Bowen Zhao, Qianqian Wang, Zhengming Ding, Quanxue Gao*
- 462 Finsler Multi-Dimensional Scaling: Manifold Learning for Asymmetric Dimensionality Reduction and Embedding,
Thomas Dagès, Simon Weber, Ya-Wei Eileen Lin, Ronen Talmon, Daniel Cremers, Michael Lindenbaum, Alfred M. Bruckstein, Ron Kimmel
- 463 Adaptive Unimodal Regulation for Balanced Multimodal Information Acquisition, *Chengxiang Huang, Yake Wei, Zequn Yang, Di Hu*
- 464 Knowledge Bridger: Towards Training-Free Missing Modality Completion, *Guanzhou Ke, Shengfeng He, Xiaoli Wang, Bo Wang, Guoqing Chao, Yuanyang Zhang, Yi Xie, Hexing Su*
- 465 OpenMIBOOD: Open Medical Imaging Benchmarks for Out-Of-Distribution Detection, *Max Gutbrod, David Rauber, Danilo Weber Nunes, Christoph Palm*
- 466 CheXWhatsApp: A Dataset for Exploring Challenges in the Diagnosis of Chest X-rays through Mobile Devices,
Mariamanna Antony, Rajiv Porana, Sahil M Lathiya, Siva Teja Kakileti, Chiranjib Bhattacharyya
- 467 Bringing CLIP to the Clinic: Dynamic Soft Labels and Negation-Aware Learning for Medical Analysis, *Hanbin Ko, Chang-Min Park*
- 468 Multi-Resolution Pathology-Language Pre-training Model with
☆ Text-Guided Visual Representation, *Shahad Albastaki, Anabia Sohail, Iyyakutti Iyappan Ganapathi, Basit Alawode, Asim Khan, Sajid Javed, Naoufel Werghi, Mohammed Bennamoun, Arif Mahmood*
- 469 ODA-GAN: Orthogonal Decoupling Alignment GAN Assisted by Weakly-supervised Learning for Virtual Immunohistochemistry Staining, *Tong Wang, Mingkang Wang, Zhongze Wang, Hongkai Wang, Qi Xu, Fengyu Cong, Hongming Xu*
- 470 STINR: Deciphering Spatial Transcriptomics via Implicit Neural Representation, *Yisi Luo, Xile Zhao, Kai Ye, Deyu Meng*
- 471 A Semantic Knowledge Complementarity based Decoupling Framework for Semi-supervised Class-imbalanced Medical Image Segmentation, *Zheng Zhang, Guanchun Yin, Bo Zhang, Wu Liu, Xiuzhuang Zhou, Wendong Wang*

- 472 Boltzmann Attention Sampling for Image Analysis with Small Objects, *Theodore Zhao, Sid Kiblawi, Naoto Usuyama, Ho Hin Lee, Sam Preston, Hoifung Poon, Mu Wei*
- 473 Boosting the Dual-Stream Architecture in Ultra-High Resolution Segmentation with Resolution-Biased Uncertainty Estimation, *Rong Qin, Xingyu Liu, Jinglei Shi, Liang Lin, Jufeng Yang*
- 474 Advancing Generalizable Tumor Segmentation with Anomaly-Aware Open-Vocabulary Attention Maps and Frozen Foundation Diffusion Models, *Yankai Jiang, Peng Zhang, Donglin Yang, Yuan Tian, Hai Lin, Xiaosong Wang*
- 475 Incomplete Multi-modal Brain Tumor Segmentation via Learnable Sorting State Space Model, *Zheyu Zhang, Yayuan Lu, Feipeng Ma, Yueyi Zhang, Huanjing Yue, Xiaoyan Sun*
- 476 EchoWorld: Learning Motion-Aware World Models for Echocardiography Probe Guidance, *Yang Yue, Yulin Wang, Haojun Jiang, Pan Liu, Shiji Song, Gao Huang*
- 477 A Unified Model for Compressed Sensing MRI Across Undersampling Patterns, *Armeet Singh Jatyani, Jiayun Wang, Aditi Chandrashekar, Zihui Wu, Miguel Liu-Schiaffini, Bahareh Tolooshams, Anima Anandkumar*
- 478 CARL: A Framework for Equivariant Image Registration, *Hastings Greer, Lin Tian, François-Xavier Vialard, Roland Kwitt, Raul San Jose Estepa, Marc Niethammer*

10:30 - 12:30 Demos (ExHall D)


- 1 A 30FPS, milliWatt-scale, Passive Snapshot Depth Camera, *John Mamish, Marcos Ferreria, Josiah Hester, Qi Guo, Emma Alexander, Yaman Sangar*
- 2 Creating a Topographic Map of the Human Cornea with a Smartphone: Specular Surface Reconstruction from Reflections in a Noisy Environment, *Omer Siddiqui, Jacalyn Wasserman, Asef Dian, Nambi Nallasamy MD*
- 3 WildGS-SLAM: Monocular Gaussian Splatting SLAM in Dynamic Environments, *Jianhao Zheng, Zihan Zhu, Valentin Bieri, Marc Pollefeys, Songyou Peng, Iro Armeni*
- 4 GenAIM: A Multimodal AI Music Generation Web Tool, *Callie Liao, Ellie Zhang*
- 5 AI3D Render, *Yosun Chang, Clara Burgert*
- 6 Generative Photography, *Yu Yuan*
- 7 SimWorld: A World Simulator for Scaling Photorealistic Multi-Agent Interactions, *Yan Zhuang*, Jiawei Ren*, Xiaokang Ye*, Xuhong He, Zijun Gao, Ryan Wu, Mrinal Dogra, Xiyan Zhang, Kai Kim, Bertt Wolfinger, Ziqiao Ma, Tianmin Shu*, Zhiting Hu*, Lianhui Qin**
- 8 Interactive Egocentric 3D Object Reconstruction from Sparse Observations, *Yufeng Zhu, Zhaoyang Lv, Zhao Dong, Zhengqin Li*
- 9 StoryBuddy V2: AI-Powered Interactive Storytelling with Visual Context, *Sumeet Kumar, Dhananjay Suresh Daundkar*
- 10 Open Vocabulary Attribution of 3D Buildings in City-Scale Photogrammetric Meshes, *Hakeem Frank, Caleb Buffa, Justin Snider, Oktay Eker, Franziska Lippoldt, Justin Chae, Patrick Tutzauer, Dmitry Kudinov*
- 11 OpenIAI-SNIO: A Systematic AR-Based Assembly Guidance System for Small-Scale, High-Density Industrial Components, *Yuntao Wang, Yu Cheng, Junhao Geng*
- 12 MSCAN - Explainable Lumbar Spinal Stenosis Diagnosis, *Arnesh Batra, Arush Gumber, Anushk Kumar*
- 13 BreakBeat: AI-Powered Dance Movement Segmentation & Choreography Insight Platform, *Jingwen Yang, Siyuan Liu, Jiawen Duan*
- 14 Real-Time Multimodal Modeling for Vision-Guided Urban Analytics, *Mehmet Kerem Turkcan, Andrew W. Smyth*

10:30 - 18:00 Art Program (ExHall A1)

11:00 Art Gallery Tour with Curator, Luba Elliott (30 mins) (ExHall A1)

11:30 - 13:30 LUNCH (ExHall C)

13:00 - 14:30 Oral Session 6A: 3D from Single or Multi-View Sensors (Karl Dean Ballroom)

 - Award candidate paper

- 1 DIFIX3D+: Improving 3D Reconstructions with Single-Step Diffusion Models, *Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, Huan Ling*
- 2 3DGUT: Enabling Distorted Cameras and Secondary Rays in Gaussian Splatting, *Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moënné-Loccoz, Zan Gojcic*
- 3 DNF: Unconditional 4D Generation with Dictionary-based Neural Fields, *Xinyi Zhang, Naiqi Li, Angela Dai*
- 4 CAT4D: Create Anything in 4D with Multi-View Video Diffusion Models, *Rundi Wu, Ruiqi Gao, Ben Poole, Alex Trevithick, Changxi Zheng, Jonathan T. Barron, Aleksander Holynski*
- 5 Diffusion Renderer: Neural Inverse and Forward Rendering with Video Diffusion Models, *Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Chih-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, Zian Wang*

13:00 - 14:30 Oral Session 6B: Scene Understanding, Image Editing and Multimodal Learning (ExHall A2)

- 1 Effective SAM Combination for Open-Vocabulary Semantic Segmentation, *Minhyeok Lee, Suhwan Cho, Jungho Lee, Sunghun Yang, Heeseung Choi, Ig-Jae Kim, Sangyoun Lee*
- 2 FluidNexus: 3D Fluid Reconstruction and Prediction from a Single Video, *Yue Gao, Hong-Xing Yu, Bo Zhu, Jiajun Wu*
- 3 Birth and Death of a Rose, *Chen Geng, Yunzhi Zhang, Shangzhe Wu, Jiajun Wu*
- 4 Semi-Supervised State-Space Model with Dynamic Stacking Filter for Real-World Video Deraining, *Shangquan Sun, Wenqi Ren, Juxiang Zhou, Shu Wang, Jianhou Gan, Xiaochun Cao*
- 5 AnyEdit: Mastering Unified High-Quality Image Editing for Any Idea, *Qifan Yu, Wei Chow, Zhongqi Yue, Kaihang Pan, Yang Wu, Xiaoyang Wan, Juncheng Li, Siliang Tang, Hanwang Zhang, Yueting Zhuang*
- 6 Generative Multimodal Pretraining with Discrete Diffusion Timestep Tokens, *Kaihang Pan, Wang Lin, Zhongqi Yue, Tenglong Ao, Liyu Jia, Wei Zhao, Juncheng Li, Siliang Tang, Hanwang Zhang*

13:00 - 14:30 Oral Session 6C: Video, Action, and Language (Davidson Ballroom)

- 1 Seeing Far and Clearly: Mitigating Hallucinations in MLLMs with Attention Causal Decoding, *Feilong Tang, Chengzhi Liu, Zhongxing Xu, Ming Hu, Zile Huang, Haochen Xue, Ziyang Chen, Zelin Peng, Zhiwei Yang, Sijin Zhou, Wenxue Li, Yulong Li, Wenxuan Song, Shiyang Su, Wei Feng, Jionglong Su, Mingquan Lin, Yifan Peng, Xuelian Cheng, Imran Razzak, Zongyuan Ge*
- 2 Video-XL: Extra-Long Vision Language Model for Hour-Scale Video Understanding, *Yan Shu, Zheng Liu, Peitian Zhang, Minghao Qin, Junjie Zhou, Zhengyang Liang, Tiejun Huang, Bo Zhao*
- 3 LoRASculpt: Sculpting LoRA for Harmonizing General and Specialized Knowledge in Multimodal Large Language Models, *Jian Liang, Wenke Huang, Guancheng Wan, Qu Yang, Mang Ye*
- 4 VideoEspresso: A Large-Scale Chain-of-Thought Dataset for Fine-Grained Video Reasoning via Core Frame Selection, *Songhao Han, Wei Huang, Hairong Shi, Le Zhuo, Xiu Su, Shifeng Zhang, Xu Zhou, Xiaojuan Qi, Yue Liao, Si Liu*
- 5 SEAL: Semantic Attention Learning for Long Video Representation, *Lan Wang, Yujia Chen, Du Tran, Vishnu Naresh Boddeti, Wen-Sheng Chu*
- 6 Learning Audio-guided Video Representation with Gated Attention for Video-Text Retrieval, *Boseung Jeong, Jicheol Park, Sungyeon Kim, Suha Kwak*

14:30 - 14:45 Courtesy Break

14:45 - 15:45 KEYNOTE 3 - Carolina Parada - Gemini Robotics, Bringing AI to the Physical World (Grand Ballroom)

15:15 - 15:45 Poster Setup (ExHall D)

16:00 - 18:00 Poster Session 6 & Exhibit Hall (ExHall D)

* - Highlight paper 🏆 - Award candidate paper

● - Oral Paper ☆ - Outstanding Reviewer

- 1 EmotiveTalk: Expressive Talking Head Generation through Audio Information Decoupling and Emotional Video Diffusion, *Haotian Wang, Yuzhe Weng, Yueyan Li, Zilu Guo, Jun Du, Shutong Niu, Jiefeng Ma, Shan He, Xiaoyan Wu, Qiming Hu, Bing Yin, Cong Liu, Qingfeng Liu*
- 2 MoEE: Mixture of Emotion Experts for Audio-Driven Portrait Animation, *Huaize Liu, Wenzhang Sun, Donglin Di, Shibo Sun, Jiahui Yang, Changqing Zou, Hujun Bao*
- 3 Synergizing Motion and Appearance: Multi-Scale Compensatory Codebooks for Talking Head Video Generation, *Shuling Zhao, Fa-Ting Hong, Xiaoshui Huang, Dan Xu*
- 4 MVPportrait: Text-Guided Motion and Emotion Control for Multi-view Vivid Portrait Animation, *Yukang Lin, Hokit Fung, Jianjin Xu, Zeping Ren, Adela S.M. Lau, Guosheng Yin, Xiu Li*
- 5 Free-viewpoint Human Animation with Pose-correlated Reference Selection, *Fa-Ting Hong, Zhan Xu, Haiyang Liu, Qinjie Lin, Luchuan Song, Zhixin Shu, Yang Zhou, Duygu Ceylan, Dan Xu*
- 6 DiffPortrait360: Consistent Portrait Diffusion for 360 View Synthesis, *Yuming Gu, Phong Tran, Yujian Zheng, Hongyi Xu, Heyuan Li, Adilbek Karmanov, Hao Li*
- 7 MeGA: Hybrid Mesh-Gaussian Head Avatar for High-Fidelity Rendering and Head Editing, *Cong Wang, Di Kang, Heyi Sun, Shenhan Qian, Zixuan Wang, Linchao Bao, Song-Hai Zhang*
- 8 HRAvatar: High-Quality and Relightable Gaussian Head Avatar, *Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Kangjie Chen, Minghan Qin, Yu Li, Haoqian Wang*
- 9 Real-time High-fidelity Gaussian Human Avatars with Position-based Interpolation of Spatially Distributed MLPs, *Yuyi Zhan, Tianjia Shao, Yin Yang, Kun Zhou*
- 10 IDOL: Instant Photorealistic 3D Human Creation from a Single Image, *Yiyu Zhuang, Jiayi Lv, Hao Wen, Qing Shuai, Ailing Zeng, Hao Zhu, Shifeng Chen, Yujiu Yang, Xun Cao, Wei Liu*
- 11 Birth and Death of a Rose, *Chen Geng, Yunzhi Zhang, Shangzhe Wu, Jiajun Wu*
- 12 DNF: Unconditional 4D Generation with Dictionary-based Neural Fields, *Xinyi Zhang, Naiqi Li, Angela Dai*
- 13 SimAvatar: Simulation-Ready Avatars with Layered Hair and Clothing, *Xueting Li, Ye Yuan, Shalini De Mello, Gilles Daviet, Jonathan Leaf, Miles Macklin, Jan Kautz, Umar Iqbal*
- 14 Disco4D: Disentangled 4D Human Generation and Animation from a Single Image, *Hui En Pang, Shuai Liu, Zhongang Cai, Lei Yang, Tianwei Zhang, Ziwei Liu*
- 15 StdGEN: Semantic-Decomposed 3D Character Generation from Single Images, *Yuze He, Yanning Zhou, Wang Zhao, Zhongkai Wu, Kaiwen Xiao, Wei Yang, Yong-Jin Liu, Xiao Han*
- 16 T-FAKE: Synthesizing Thermal Images for Facial Landmarking, *Philipp Flotho, Moritz Piening, Anna Kukleva, Gabriele Steidl*
- 17 Diff-Palm: Realistic Palmprint Generation with Polynomial Creases and Intra-Class Variation Controllable Diffusion Models, *Jianlong Jin, Chenglong Zhao, Ruixin Zhang, Sheng Shang, Jianqing Xu, Jingyun Zhang, ShaoMing Wang, Yang Zhao, Shouhong Ding, Wei Jia, Yunsheng Wu*
- 18 GBC-Splat: Generalizable Gaussian-Based Clothed Human Digitalization under Sparse RGB Cameras, *Hanzhang Tu, Zhanfeng Liao, Boyao Zhou, Shunyuan Zheng, Xilong Zhou, Liuxin Zhang, QianYing Wang, Yebin Liu*
- 19 VTON 360: High-Fidelity Virtual Try-On from Any Viewing Direction, *Zijian He, Yuwei Ning, Yipeng Qin, Guangrun Wang, Sibe Yang, Liang Lin, Guanbin Li*
- 20 BooW-VTON: Boosting In-the-Wild Virtual Try-On via Mask-Free Pseudo Data Training, *Xuanpu Zhang, Dan Song, Pengxin Zhan, Tianyu Chang, Jianhao Zeng, Qingguo Chen, Weihua Luo, An-An Liu*
- 21 SFDm: Robust Decomposition of Geometry and Reflectance for Realistic Face Rendering from Sparse-view Images, *Daisheng Jin, Jiangbei Hu, Baixin Xu, Yuxin Dai, Chen Qian, Ying He*
- 22 Integral Fast Fourier Color Constancy, *Wenjun Wei, Yanlin Qian, Huaian Chen, Junkang Dai, Yi Jin*
- 23 Reversible Decoupling Network for Single Image Reflection Removal, *Hao Zhao, Mingjia Li, Qiming Hu, Xiaojie Guo*
- 24 Stabilizing and Accelerating Autofocus with Expert Trajectory Regularized Deep Reinforcement Learning, *Shouhang Zhu, Chenglin Li, Yuankun Jiang, Li Wei, Nuowen Kan, Ziyang Zheng, Wenrui Dai, Junni Zou, Hongkai Xiong*
- 25 V2V3D: View-to-View Denoised 3D Reconstruction for Light Field Microscopy, *Jiayin Zhao, Zhenqi Fu, Tao Yu, Hui Qiao*
- 26 DoF-Gaussian: Controllable Depth-of-Field for 3D Gaussian Splatting, *Liao Shen, Tianqi Liu, Huiqiang Sun, Jiaqi Li, Zhiguo Cao, Wei Li, Chen Change Loy*
- 27 Luminance-GS: Adapting 3D Gaussian Splatting to Challenging Lighting Conditions with View-Adaptive Curve Adjustment, *Ziteng Cui, Xuangeng Chu, Tatsuya Harada*
- 28 3DGUT: Enabling Distorted Cameras and Secondary Rays in Gaussian Splatting, *Qi Wu, Janick Martinez Esturo, Ashkan Mirzaei, Nicolas Moënné-Loccoz, Zan Gojcic*
- 29 Diffusion Renderer: Neural Inverse and Forward Rendering with Video Diffusion Models, *Ruofan Liang, Zan Gojcic, Huan Ling, Jacob Munkberg, Jon Hasselgren, Chih-Hao Lin, Jun Gao, Alexander Keller, Nandita Vijaykumar, Sanja Fidler, Zian Wang*
- 30 Ref-GS: Directional Factorization for 2D Gaussian Splatting, *Youjia Zhang, Anpei Chen, Yumin Wan, Zikai Song, Junqing Yu, Yawei Luo, Wei Yang*
- 31 NeISF++: Neural Incident Stokes Field for Polarized Inverse Rendering of Conductors and Dielectrics, *Chen hao Li, Taishi Ono, Takeshi Uemori, Sho Nitta, Hajime Mihara, Alexander Gatto, Hajime Nagahara, Yusuke Moriuchi*
- 32 FluidNexus: 3D Fluid Reconstruction and Prediction from a Single Video, *Yue Gao, Hong-Xing Yu, Bo Zhu, Jiajun Wu*
- 33 Uni-Renderer: Unifying Rendering and Inverse Rendering Via Dual Stream Diffusion, *Zhifei Chen, Tianshuo Xu, Wenhang Ge, Leyi Wu, Dongyu Yan, Jing He, Luozhou Wang, Lu Zeng, Shunsi Zhang, Ying-Cong Chen*
- 34 Neural LightRig: Unlocking Accurate Object Normal and Material Estimation with Multi-Light Diffusion, *Zexin He, Tengfei Wang, Xin Huang, Xingang Pan, Ziwei Liu*
- 35 RNG: Relightable Neural Gaussians, *Jiahui Fan, Fujun Luan, Jian Yang, Milos Hasan, Beibei Wang*
- 36 SGSST: Scaling Gaussian Splatting Style Transfer, *Bruno Galerne, Jianling Wang, Lara Raad, Jean-Michel Morel*
- 37 Vid2Sim: Generalizable, Video-based Reconstruction of Appearance, Geometry and Physics for Mesh-free Simulation, *Chuhao Chen, Zhiyang Dou, Chen Wang, Yiming Huang, Anjun Chen, Qiao Feng, Jiatao Gu, Lingjie Liu*
- 38 Material Anything: Generating Materials for Any 3D Object via Diffusion, *Xin Huang, Tengfei Wang, Ziwei Liu, Qing Wang*
- 39 TexGarment: Consistent Garment UV Texture Generation via Efficient 3D Structure-Guided Diffusion Transformer, *Jialun Liu, Jinbo Wu, Xiaobo Gao, Jiakui Hu, Bojun Xiong, Xing Liu, Chen Zhao, Hongbin Pei, Haocheng Feng, Yingying Li, Errui Ding, Jingdong Wang*
- 40 3DTopia-XL: Scaling High-quality 3D Asset Generation via Primitive Diffusion, *Zhaoxi Chen, Jiaxiang Tang, Yuhao Dong, Ziang Cao, Fangzhou Hong, Yushi Lan, Tengfei Wang, Haozhe Xie, Tong Wu, Shunsuke Saito, Liang Pan, Dahua Lin, Ziwei Liu*
- 41 BrepGiff: Lightweight Generation of Complex B-rep with 3D GAT Diffusion, *Hao Guo, Xiaoshui Huang, Hao jiacheng, Yunpeng Bai, Hongping Gan, Yilei Shi*
- 42 Towards Realistic Example-based Modeling via 3D Gaussian Stitching, *Xinyu Gao, Ziyi Yang, Bingchen Gong, Xiaoguang Han, Sipeng Yang, Xiaogang Jin*
- 43 TreeMeshGPT: Artistic Mesh Generation with Autoregressive Tree Sequencing, *Stefan Lionar, Jiabin Liang, Gim Hee Lee*
- 44 GenVDM: Generating Vector Displacement Maps From a Single Image, *Yuezhi Yang, Qimin Chen, Vladimir G. Kim, Siddhartha Chaudhuri, Qixing Huang, Zhiqin Chen*
- 45 CTRL-D: Controllable Dynamic 3D Scene Editing with Personalized 2D Diffusion, *Kai He, Chin-Hsuan Wu, Igor Gilitschenski*

- 46 LeanGaussian: Breaking Pixel or Point Cloud Correspondence in Modeling 3D Gaussians, *Jiamin Wu, Kenkun Liu, Han Gao, Xiaoke Jiang, Yuan Yao, Lei Zhang*
- 47 FlashGS: Efficient 3D Gaussian Splatting for Large-scale and High-resolution Rendering, *Guofeng Feng, Siyan Chen, Rong Fu, Zimu Liao, Yi Wang, Tao Liu, Boni Hu, Linning Xu, Zhilin Pei, Hengjie Li, Xiuhong Li, Ninghui Sun, Xingcheng Zhang, Bo Dai*
- 48 Steepest Descent Density Control for Compact 3D Gaussian Splatting, *Peihao Wang, Yuehao Wang, Dilin Wang, Sreyas Mohan, Zhiwen Fan, Lemeng Wu, Ruisi Cai, Yu-Ying Yeh, Zhangyang Wang, Qiang Liu, Rakesh Ranjan*
- 49 GaussianSpa: An "Optimizing-Sparsifying" Simplification Framework for Compact and High-Quality 3D Gaussian Splatting, *Yangming Zhang, Wenqi Jia, Wei Niu, Miao Yin*
- 50 Generative Densification: Learning to Densify Gaussians for
* High-Fidelity Generalizable 3D Reconstruction, *Seungtae Nam, Xiangyu Sun, Gyeongjin Kang, Younggeun Lee, Seungjun Oh, Eunbyung Park*
- 51 IMFine: 3D Inpainting via Geometry-guided Multi-view Refinement, *Zhihao Shi, Dong Huo, Yuhongze Zhou, Yan Min, Juwei Lu, Xinxin Zuo*
- 52 3D Gaussian Inpainting with Depth-Guided Cross-View Consistency, *Sheng-Yu Huang, Zi-Ting Chou, Yu-Chiang Frank Wang*
- 53 CAT4D: Create Anything in 4D with Multi-View Video Diffusion
O Models, *Rundi Wu, Ruiqi Gao, Ben Poole, Alex Trevithick, Changxi Zheng, Jonathan T. Barron, Aleksander Holynski*
- 54 HoGS: Unified Near and Far Object Reconstruction via Homogeneous Gaussian Splatting, *Xinpeng Liu, Zeyi Huang, Fumio Okura, Yasuyuki Matsushita*
- 55 Scene4U: Hierarchical Layered 3D Scene Reconstruction from Single Panoramic Image for Your Immerse Exploration, *Zilong Huang, Jun He, Junyan Ye, Lihan Jiang, Weijia Li, Yiping Chen, Ting Han*
- 56 Learning Partonomic 3D Reconstruction from Image Collections, *Xiaoqian Ruan, Pei Yu, Dian Jia, Hyeonjeong Park, Peixi Xiong, Wei Tang*
- 57 DIFIX3D+: Improving 3D Reconstructions with Single-Step
S Diffusion Models, *Jay Zhangjie Wu, Yuxuan Zhang, Haithem Turki, Xuanchi Ren, Jun Gao, Mike Zheng Shou, Sanja Fidler, Zan Gojcic, Huan Ling*
- 58 Generative Sparse-View Gaussian Splatting, *Hanyang Kong, Xingyi Yang, Xinchao Wang*
- 59 Novel View Synthesis with Pixel-Space Diffusion Models, *Noam Elata, Bahjat Kavar, Yaron Ostrovsky-Berman, Miriam Farber, Ron Sokolovsky*
- 60 MOVIS: Enhancing Multi-Object Novel View Synthesis for Indoor Scenes, *Ruijie Lu, Yixin Chen, Junfeng Ni, Baoxiong Jia, Yu Liu, Diwen Wan, Gang Zeng, Siyuan Huang*
- 61 CoMapGS: Covisibility Map-based Gaussian Splatting for Sparse
☆ Novel View Synthesis, *Youngkyoon Jang, Eduardo Pérez-Pellitero*
- 62 Horizon-GS: Unified 3D Gaussian Splatting for Large-Scale Aerial-to-Ground Scenes, *Lihan Jiang, Kerui Ren, Mulin Yu, Linning Xu, Junting Dong, Tao Lu, Feng Zhao, Dahua Lin, Bo Dai*
- 63 NexusGS: Sparse View Synthesis with Epipolar Depth Priors in
* 3D Gaussian Splatting, *Yulong Zheng, Zicheng Jiang, Shengfeng He, Yandu Sun, Junyu Dong, Huaidong Zhang, Yong Du*
- 64 SPARS3R: Semantic Prior Alignment and Regularization for
☆ Sparse 3D Reconstruction, *Yutao Tang, Yuxiang Guo, Deming Li, Cheng Peng*
- 65 StarGen: A Spatiotemporal Autoregression Framework with Video Diffusion Model for Scalable and Controllable Scene Generation, *Shangjin Zhai, Zhichao Ye, Jialin Liu, Weijian Xie, Jiaqi Hu, Zhen Peng, Hua Xue, Danpeng Chen, Xiaomeng Wang, Lei Yang, Nan Wang, Haomin Liu, Guofeng Zhang*
- 66 PMNI: Pose-free Multi-view Normal Integration for Reflective and Textureless Surface Reconstruction, *Mingzhi Pei, Xu Cao, Xiangyi Wang, Heng Guo, Zhanyu Ma*
- 67 Learnable Infinite Taylor Gaussian for Dynamic View Rendering, *Bingbing Hu, Yanyan Li, Rui Xie, Bo Xu, Haoye Dong, Junfeng Yao, Gim Hee Lee*
- 68 Efficient Dynamic Scene Editing via 4D Gaussian-based Static-Dynamic Separation, *Joohyun Kwon, Hanbyel Cho, Junmo Kim*
- 69 SplineGS: Robust Motion-Adaptive Spline for Real-Time Dynamic
☆ 3D Gaussians from Monocular Video, *Jongmin Park, Minh-Quan Viet Bui, Juan Luis Gonzalez Bello, Jaeho Moon, Jihyong Oh, Munchurl Kim*
- 70 EventSplat: 3D Gaussian Splatting from Moving Event Cameras for Real-time Rendering, *Toshiya Yura, Ashkan Mirzaei, Igor Gilitschenski*
- 71 SOLAMI: Social Vision-Language-Action Modeling for Immersive Interaction with 3D Autonomous Characters, *Jianping Jiang, Weiye Xiao, Zhengyu Lin, Huaizhong Zhang, Tianxiang Ren, Yang Gao, Zhiqian Lin, Zhongang Cai, Lei Yang, Ziwei Liu*
- 72 Denoising Functional Maps: Diffusion Models for Shape Correspondence, *Aleksei Zhuravlev, Zorah Löhner, Vladislav Golyanik*
- 73 Event Fields: Capturing Light Fields at High Speed, Resolution,
* and Dynamic Range, *Ziyuan Qu, Zihao Zou, Vivek Boominathan, Praneeth Chakravarthula, Adithya Pediredla*
- 74 4DTAM: Non-Rigid Tracking and Mapping via Dynamic Surface Gaussians, *Hidenobu Matsuki, Gwangbin Bae, Andrew J. Davison*
- 75 IncEventGS: Pose-Free Gaussian Splatting from a Single Event
* Camera, *Jian Huang, Chengrui Dong, Xuanhua Chen, Peidong Liu*
- 76 Completion as Enhancement: A Degradation-Aware Selective Image Guided Network for Depth Completion, *Zhiqiang Yan, Zhengxue Wang, Kun Wang, Jun Li, Jian Yang*
- 77 Blurred LiDAR for Sharper 3D: Robust Handheld 3D Scanning
* with Diffuse LiDAR and RGB, *Nikhil Behari, Aaron Young, Siddharth Somasundaram, Tzofi Klinghoffer, Akshat Dave, Ramesh Raskar*
- 78 Focal Split: Untethered Snapshot Depth from Differential Defocus, *Junjie Luo, John Mamish, Alan Fu, Thomas Concannon, Josiah Hester, Emma Alexander, Qi Guo*
- 79 HELVIPAD: A Real-World Dataset for Omnidirectional Stereo
* Depth Estimation, *Mehdi Zayene, Jannik Endres, Albias Havolli, Charles Corbière, Salim Cherkaoui, Alexandre Kontouli, Alexandre Alahi*
- 80 OFER: Occluded Face Expression Reconstruction, *Pratheba Selvaraju, Victoria Fernandez Abrevaya, Timo Bolkart, Rick Akkerman, Tianyu Ding, Faezeh Amjadi, Ilya Zharkov*
- 81 Depth Any Camera: Zero-Shot Metric Depth Estimation from Any Camera, *Yuliang Guo, Sparsh Garg, S. Mahdi H. Miangoleh, Xinyu Huang, Liu Ren*
- 82 Order-One Rolling Shutter Cameras, *Marvin Anas Hahn, Kathlén Kohn, Orlando Marigliano, Tomas Pajdla*
- 83 Matrix-Free Shared Intrinsic Bundle Adjustment, *Daniel Safari*
- 84 Towards In-the-wild 3D Plane Reconstruction from a Single Image,
* *Jiachen Liu, Rui Yu, Sili Chen, Sharon X. Huang, Hengkai Guo*
- 85 Learning Affine Correspondences by Integrating Geometric Constraints, *Pengju Sun, Banglei Guan, Zhenbao Yu, Yang Shang, Qifeng Yu, Daniel Barath*
- 86 DiskVPS: Vanishing Point Detector via Hough Transform in a Disk Region, *Jianping Wu*
- 87 From Sparse to Dense: Camera Relocalization with Scene-Specific Detector from Feature Gaussian Splatting, *Zhiwei Huang, Hailin Yu, Yichun Shentu, Jin Yuan, Guofeng Zhang*
- 88 RUBIK: A Structured Benchmark for Image Matching across Geometric Challenges, *Thibaut Loiseau, Guillaume Bourmaud*
- 89 MATCHA: Towards Matching Anything, *Fei Xue, Sven Elflein, Laura Leal-Taixé, Qunjie Zhou*
- 90 Scene-agnostic Pose Regression for Visual Localization,
☆ *Junwei Zheng, Ruiping Liu, Yufan Chen, Zhenfang Chen, Kailun Yang, Jiaming Zhang, Rainer Stiefel*
- 91 Simulator HC: Regression-based Online Simulation of Starting
* Problem-Solution Pairs for Homotopy Continuation in Geometric Vision, *Xinyue Zhang, Zijia Dai, Wanting Xu, Laurent Kneip*
- 92 GaussianUDF: Inferring Unsigned Distance Functions through 3D
* Gaussian Splatting, *Shujuan Li, Yu-Shen Liu, Zhizhong Han*
- 93 ProbPose: A Probabilistic Approach to 2D Human Pose
☆ Estimation, *Miroslav Purkrabek, Jiri Matas*
- 94 Floating No More: Object-Ground Reconstruction from a Single

- Image, *Yunze Man, Yichen Sheng, Jianming Zhang, Liang-Yan Gui, Yu-Xiong Wang*
- 95 ArticulatedGS: Self-supervised Digital Twin Modeling of Articulated Objects using 3D Gaussian Splatting, *Junfu Guo, Yu Xin, Gaoyi Liu, Kai Xu, Ligang Liu, Ruizhen Hu*
- 96 GCE-Pose: Global Context Enhancement for Category-level Object Pose Estimation, *Weihang Li, Hongli XU, Junwen Huang, Hyunjun Jung, Peter KT Yu, Nassir Navab, Benjamin Busam*
- 97 Doppelgangers++: Improved Visual Disambiguation with Geometric 3D Features, *Yuanbo Xiangli, Ruojin Cai, Hanyu Chen, Jeffrey Byrne, Noah Snaveley*
- ☆ MITracker: Multi-View Integration for Visual Object Tracking, *Mengjie Xu, Yitao Zhu, Haotian Jiang, Jiaming Li, Zhenrong Shen, Sheng Wang, Haolin Huang, Xinyu Wang, Han Zhang, Qing Yang, Qian Wang*
- 99 ETAP: Event-based Tracking of Any Point, *Friedhelm Hamann, Daniel Gehrig, Filbert Febryanto, Kostas Daniilidis, Guillermo Gallego*
- 100 Ev-3DOD: Pushing the Temporal Boundaries of 3D Object Detection with Event Cameras, *Hoonhee Cho, Jae-Young Kang, Youngho Kim, Kuk-Jin Yoon*
- 101 GO-N3RDet: Geometry Optimized NeRF-enhanced 3D Object Detector, *Zechuan Li, Hongshan Yu, Yihao Ding, Jinhao Qiao, Basim Azam, Naveed Akhtar*
- 102 Preconditioners for the Stochastic Training of Neural Fields, *Shin-Fang Chng, Hemanth Saratchandran, Simon Lucey*
- 103 3D-SLNR: A Super Lightweight Neural Representation for Large-scale 3D Mapping, *Chenhui Shi, Fulin Tang, Ning An, Yihong Wu*
- 104 PCDreamer: Point Cloud Completion Through Multi-view Diffusion Priors, *Guangshun Wei, Yuan Feng, Long Ma, Chen Wang, Yuanfeng Zhou, Changjian Li*
- 105 STAR-Edge: Structure-aware Local Spherical Curve Representation for Thin-walled Edge Extraction from Unstructured Point Clouds, *Zikuan Li, Honghua Chen, Yuecheng Wang, Sibao Wu, Mingqiang Wei, Jun Wang*
- 106 DV-Matcher: Deformation-based Non-rigid Point Cloud Matching Guided by Pre-trained Visual Features, *Zhangquan Chen, Puhua Jiang, Ruqi Huang*
- 107 Mitigating Ambiguities in 3D Classification with Gaussian Splatting, *Ruiqi Zhang, Hao Zhu, Jingyi Zhao, Qi Zhang, Xun Cao, Zhan Ma*
- 108 Sparse Point Cloud Patches Rendering via Splitting 2D Gaussians, *Changfeng Ma, Ran Bi, Jie Guo, Chongjun Wang, Yanwen Guo*
- 109 SASep: Saliency-Aware Structured Separation of Geometry and Feature for Open Set Learning on Point Clouds, *Jinfeng Xu, Xianzhi Li, Yuan Tang, Xu Han, Qiao Yu, Yixue Hao, Long Hu, Min Chen*
- 110 TopNet: Transformer-Efficient Occupancy Prediction Network for Octree-Structured Point Cloud Geometry Compression, *Xinjie Wang, Yifan Zhang, Ting Liu, Xipu Liu, Ke Xu, Jianwei Wan, Yulan Guo, Hanyun Wang*
- 111 A Unified Approach to Interpreting Self-supervised Pre-training Methods for 3D Point Clouds via Interactions, *Qiang Li, Jian Ruan, Fanghao Wu, Yuchi Chen, Zhihua Wei, Wen Shen*
- 112 An End-to-End Robust Point Cloud Semantic Segmentation Network with Single-Step Conditional Diffusion Models, *Wentao Qu, Jing Wang, YongShun Gong, Xiaoshui Huang, Liang Xiao*
- 113 PillarHist: A Quantization-aware Pillar Feature Encoder based on Height-aware Histogram, *Sifan Zhou, Zhihang Yuan, Dawei Yang, Xing Hu, Jian Qian, Ziyu Zhao*
- 114 Deep Change Monitoring: A Hyperbolic Representative Learning Framework and a Dataset for Long-term Fine-grained Tree Change Detection, *Yante Li, Hanwen Qi, Haoyu Chen, Xinlian Liang, Guoying Zhao*
- 115 GBlobs: Explicit Local Structure via Gaussian Blobs for Improved Cross-Domain LiDAR-based 3D Object Detection, *Dušan Malić, Christian Fruhwirth-Reisinger, Samuel Schuster, Horst Possegger*
- 116 LiMoE: Mixture of LiDAR Representation Learners from Automotive Scenes, *Xiang Xu, Lingdong Kong, Hui Shuai, Liang Pan, Ziwei Liu, Qingshan Liu*
- 117 Exploring Scene Affinity for Semi-Supervised LiDAR Semantic Segmentation, *Chuangdong Liu, Xingxing Weng, Shuguo Jiang, Pengcheng Li, Lei Yu, Gui-Song Xia*
- 118 V2X-R: Cooperative LiDAR-4D Radar Fusion with Denoising Diffusion for 3D Object Detection, *Xun Huang, Jinlong Wang, Qiming Xia, Siheng Chen, Bisheng Yang, Xin Li, Cheng Wang, Chenglu Wen*
- 119 Leveraging Temporal Cues for Semi-Supervised Multi-View 3D Object Detection, *Jinhyung Park, Navyata Sanghvi, Hiroki Adachi, Yoshihisa Shibata, Shawn Hunt, Shinya Tanaka, Hironobu Fujiyoshi, Kris Kitani*
- 120 CorrBEV: Multi-View 3D Object Detection by Correlation Learning with Multi-modal Prototypes, *Ziteng Xue, Mingzhe Guo, Heng Fan, Shihui Zhang, Zhipeng Zhang*
- 121 CroCoDL: Cross-device Collaborative Dataset for Localization, *Hermann Blum, Alessandro Mercurio, Joshua O'Reilly, Tim Engelbracht, Mihai Dusmanu, Marc Pollefeys, Zuria Bauer*
- ☆ ShowHowTo: Generating Scene-Conditioned Step-by-Step Visual Instructions, *Tomáš Souček, Prajwal Gatti, Michael Wray, Ivan Laptev, Dima Damen, Josef Sivic*
- 123 RoboSense: Large-scale Dataset and Benchmark for Egocentric Robot Perception and Navigation in Crowded and Unstructured Environments, *Haisheng Su, Feixiang Song, Cong Ma, Wei Wu, Junchi Yan*
- 124 DIO: Decomposable Implicit 4D Occupancy-Flow World Model, *Christopher Diehl, Quinlan Sykora, Ben Agro, Thomas Gilles, Sergio Casas, Raquel Urtasun*
- 125 EvOcc: Accurate Semantic Occupancy for Automated Driving Using Evidence Theory, *Jonas Kälble, Sascha Wirges, Maxim Tatarchenko, Eddy Ilg*
- 126 GaussianFormer-2: Probabilistic Gaussian Superposition for Efficient 3D Occupancy Prediction, *Yuanhui Huang, Amonnut Thammatadatrakoon, Wenzhao Zheng, Yunpeng Zhang, Dalong Du, Jiwen Lu*
- 127 SplatFlow: Self-Supervised Dynamic Gaussian Splatting in Neural Motion Flow Field for Autonomous Driving, *Su Sun, Cheng Zhao, Zhuoyang Sun, Yingjie Victor Chen, Mei Chen*
- 128 DriveGEN: Generalized and Robust 3D Detection in Driving via Controllable Text-to-Image Diffusion Generation, *Hongbin Lin, Zilu Guo, Yifan Zhang, Shuaicheng Niu, Yafeng Li, Ruimao Zhang, Shuguang Cui, Zhen Li*
- 129 Graph-Based 3D Lane Detection from Monocular Images, *Halil İbrahim Öztürk, Muhammet Esat Kalfaoğlu, Ozzel Kilinc*
- 130 UrbanCAD: Towards Highly Controllable and Photorealistic 3D Vehicles for Urban Scene Simulation, *Yichong Lu, Yichi Cai, Shangzhan Zhang, Hongyu Zhou, Haoji Hu, Huimin Yu, Andreas Geiger, Yiyi Liao*
- 131 DrivingSphere: Building a High-fidelity 4D World for Closed-loop Simulation, *Tianyi Yan, Dongming Wu, Wencheng Han, Junpeng Jiang, Xia Zhou, Kun Zhan, Cheng-zhong Xu, Jianbing Shen*
- 132 Causal Composition Diffusion Model for Closed-loop Traffic Generation, *Haohong Lin, Xin Huang, Tung Phan, David Hayden, Huan Zhang, Ding Zhao, Siddhartha Srinivasa, Eric Wolff, Hongge Chen*
- 133 Towards Autonomous Micromobility through Scalable Urban Simulation, *Wayne Wu, Honglin He, Chaoyuan Zhang, Jack He, Seth Z. Zhao, Ran Gong, Quanyi Li, Bolei Zhou*
- 134 Towards Generalizable Trajectory Prediction using Dual-Level Representation Learning and Adaptive Prompting, *Kaouther Messaoud, Matthieu Cord, Alexandre Alahi*
- 135 Distilling Multi-modal Large Language Models for Autonomous Driving, *Deepti Hegde, Rajeev Yasarla, Hong Cai, Shizhong Han, Apratim Bhattacharyya, Shweta Mahajan, Litian Liu, Risheek Garrepalli, Vishal M. Patel, Fatih Porikli*
- 136 RoomTour3D: Geometry-Aware Video-Instruction Tuning for Embodied Navigation, *Mingfei Han, Liang Ma, Kamila Zhumakhanova, Ekaterina Radionova, Jingyi Zhang, Xiaojun Chang, Xiaodan Liang, Ivan Laptev*
- 137 Exploration-Driven Generative Interactive Environments, *Nedko Savov, Naser Kazemi, Mohammad Mahdi, Danda Pani Paudel, Xi Wang, Luc Van Gool*

- 138 Neural Motion Simulator Pushing the Limit of World Models in Reinforcement Learning, *Chenjie Hao, Weyl Lu, Yifan Xu, Yubei Chen*
- 139 Reasoning Mamba: Hypergraph-Guided Region Relation Calculating for Weakly Supervised Affordance Grounding, *Yuxuan Wang, Aming Wu, Muli Yang, Yukuan Min, Yihang Zhu, Cheng Deng*
- 140 AutoURDF: Unsupervised Robot Modeling from Point Cloud Frames Using Cluster Registration, *Jiong Lin, Lechen Zhang, Kwansoo Lee, Jialong Ning, Judah Goldfeder, Hod Lipson*
- 141 Object-Centric Prompt-Driven Vision-Language-Action Model for Robotic Manipulation, *Xiaoqi Li, Jingyun Xu, Mingxu Zhang, Jiaming Liu, Yan Shen, Iaroslav Ponomarenko, Jiahui Xu, Liang Heng, Siyuan Huang, Shanghang Zhang, Hao Dong*
- 142 RoboTwin: Dual-Arm Robot Benchmark with Generative Digital
* Twins, *Yao Mu, Tianxing Chen, Zanzin Chen, Shijia Peng,*
☆ *Zhigian Lan, Zeyu Gao, Zhixuan Liang, Qiaojun Yu, Yude Zou, Mingkun Xu, Lunkai Lin, Zhiqiang Xie, Mingyu Ding, Ping Luo*
- 143 VidBot: Learning Generalizable 3D Actions from In-the-Wild 2D Human Videos for Zero-Shot Robotic Manipulation, *Hanzhi Chen, Boyang Sun, Anran Zhang, Marc Pollefeys, Stefan Leutenegger*
- 144 Learning Physics-Based Full-Body Human Reaching and Grasping from Brief Walking References, *Yitang Li, Mingxian Lin, Zhuo Lin, Yipeng Deng, Yue Cao, Li Yi*
- 145 TASTE-Rob: Advancing Video Generation of Task-Oriented Hand-
☆ Object Interaction for Generalizable Robotic Manipulation, *Hongxiang Zhao, Xingchen Liu, Mutian Xu, Yiming Hao, Weikai Chen, Xiaoguang Han*
- 146 BimArt: A Unified Approach for the Synthesis of 3D Bimanual Interaction with Articulated Objects, *Wanyue Zhang, Rishabh Dabral, Vladislav Golyanik, Vasileios Choutas, Eduardo Alvarado, Thabo Beeler, Marc Habermann, Christian Theobalt*
- 147 End-to-End HOI Reconstruction Transformer with Graph-based
* Encoding, *Zhenrong Wang, Qi Zheng, Sihan Ma, Maosheng Ye, Yibing Zhan, Dongjiang Li*
- 148 Dyn-HaMR: Recovering 4D Interacting Hand Motion from a
* Dynamic Camera, *Zhengdi Yu, Stefanos Zafeiriou, Tolga Birdal*
- 149 EgoPressure: A Dataset for Hand Pressure and Pose Estimation
* in Egocentric Vision, *Yiming Zhao, Taein Kwon, Paul Strel, Marc Pollefeys, Christian Holz*
- 150 PI-HMR: Towards Robust In-bed Temporal Human Shape Reconstruction with Contact Pressure Sensing, *Ziyu Wu, Yufan Xiong, Mengting Niu, Fangting Xie, Quan Wan, Qijun Ying, Boyan Liu, Xiaohui Cai*
- 151 MVDoppler-Pose: Multi-Modal Multi-View mmWave Sensing for Long-Distance Self-Occluded Human Walking Pose Estimation, *Jaeho Choi, Soheil Hor, Shubo Yang, Amin Arbabian*
- 152 MotionPRO: Exploring the Role of Pressure in Human MoCap and
* Beyond, *Shenghao Ren, Yi Lu, Jiayi Huang, Jiayi Zhao, He Zhang, Tao Yu, Qiu Shen, Xun Cao*
- 153 MODA: Motion-Drift Augmentation for Inertial Human Motion
☆ Analysis, *Yinghao Wu, Shihui Guo, Yipeng Qin*
- 154 Homogeneous Dynamics Space for Heterogeneous Humans, *Xinpeng Liu, Junxuan Liang, Chenshuo Zhang, Zixuan Cai, Cewu Lu, Yong-Lu Li*
- 155 Modeling Multiple Normal Action Representations for Error Detection in Procedural Tasks, *Wei-Jin Huang, Yuan-Ming Li, Zhi-Wei Xia, Yu-Ming Tang, Kun-Yu Lin, Jian-Fang Hu, Wei-Shi Zheng*
- 156 UniPose: A Unified Multimodal Framework for Human Pose
* Comprehension, Generation and Editing, *Yiheng Li, Ruibing Hou, Hong Chang, Shiguang Shan, Xilin Chen*
- 157 Symbolic Representation for Any-to-Any Generative Tasks, *Jiaqi Chen, Xiaoye Zhu, Yue Wang, Tianyang Liu, Xinhui Chen, Ying Chen, Chak Tou Leong, Yifei Ke, Joseph Liu, Yiwen Yuan, Julian McAuley, Li-jia Li*
- 158 SimMotionEdit: Text-Based Human Motion Editing with Motion Similarity Prediction, *Zhengyuan Li, Kai Cheng, Anindita Ghosh, Uttaran Bhattacharya, Liangyan Gui, Aniket Bera*
- 159 AnyMoLe: Any Character Motion In-betweening Leveraging Video Diffusion Models, *Kwan Yun, Seokhyeon Hong, Chaelin Kim, Junyong Noh*
- 160 MG-MotionLLM: A Unified Framework for Motion Comprehension and Generation across Multiple Granularities, *Bizhu Wu, Jinheng Xie, Keming Shen, Zhe Kong, Jianfeng Ren, Ruibin Bai, Rong Qu, Linlin Shen*
- 161 Rethinking Diffusion for Text-Driven Human Motion Generation: Redundant Representations, Evaluation, and Masked Autoregression, *Zichong Meng, Yiming Xie, Xiaogang Peng, Zeyu Han, Huaizu Jiang*
- 162 ScaMo: Exploring the Scaling Law in Autoregressive Motion Generation Model, *Shunlin Lu, Jingbo Wang, Zeyu Lu, Ling-Hao Chen, Wenxun Dai, Junting Dong, Zhiyang Dou, Bo Dai, Ruimao Zhang*
- 163 Multiple Object Tracking as ID Prediction, *Ruopeng Gao, Ji Qi, Limin Wang*
- 164 Shape and Texture: What Influences Reliable Optical Flow Estimation?, *Libo Long, Xiao Hu, Jochen Lang*
- 165 Bridge Frame and Event: Common Spatiotemporal Fusion for High-Dynamic Scene Optical Flow, *Hanyu Zhou, Haonan Wang, Haoyue Liu, Yuxing Duan, Yi Chang, Luxin Yan*
- 166 Unified Reconstruction of Static and Dynamic Scenes from Events,
* *Qiyao Gao, Peiqi Duan, Hanyue Lou, Minggui Teng, Ziqi Cai, Xu Chen, Boxin Shi*
- 167 Learning Physics From Video: Unsupervised Physical Parameter Estimation for Continuous Dynamical Systems, *Alejandro Castañeda Garcia, Jan Warchocki, Jan van Gemert, Daan Brinks, Nergis Tomen*
- 168 Generating 3D-Consistent Videos from Unposed Internet Photos, *Gene Chou, Kai Zhang, Sai Bi, Hao Tan, Zexiang Xu, Fujun Luan, Bharath Hariharan, Noah Snaveley*
- 169 AnimateAnything: Consistent and Controllable Animation for Video Generation, *Guojun Lei, Chi Wang, Rong Zhang, Yikai Wang, Hong Li, Weiwei Xu*
- 170 MotionPro: A Precise Motion Controller for Image-to-Video Generation, *Zhongwei Zhang, Fuchen Long, Zhaofan Qiu, Yingwei Pan, Wu Liu, Ting Yao, Tao Mei*
- 171 Generative Inbetweening through Frame-wise Conditions-Driven Video Generation, *Tianyi Zhu, Dongwei Ren, Qilong Wang, Xiaohe Wu, Wangmeng Zuo*
- 172 FreePCA: Integrating Consistency Information across Long-
* short Frames in Training-free Long Video Generation via Principal Component Analysis, *Jiangtong Tan, Hu Yu, Jie Huang, Jie Xiao, Feng Zhao*
- 173 Probability Density Geodesics in Image Diffusion Latent Space,
☆ *Qingtao Yu, Jaskirat Singh, Zhaoyuan Yang, Peter Henry Tu, Jing Zhang, Hongdong Li, Richard Hartley, Dylan Campbell*
- 174 Bias for Action: Video Implicit Neural Representations with Bias Modulation, *Alper Kayabasi, Anil Kumar Vadathya, Guha Balakrishnan, Vishwanath Saragadam*
- 175 BF-STVSR: B-Splines and Fourier—Best Friends for High Fidelity Spatial-Temporal Video Super-Resolution, *Eunjin Kim, Hyeonjin Kim, Kyong Hwan Jin, Jaedun Yoo*
- 176 FLAVC: Learned Video Compression with Feature Level Attention, *Chun Zhang, Heming Sun, Jiro Katto*
- 177 ProReflow: Progressive Reflow with Decomposed Velocity, *Lei Ke, Haohang Xu, Xuefei Ning, Yu Li, Jiajun Li, Haoling Li, Yuxuan Lin, Dongsheng Jiang, Yujiu Yang, Linfeng Zhang*
- 178 Making Old Film Great Again: Degradation-aware State Space Model for Old Film Restoration, *Yudong Mao, Hao Luo, Zhiwei Zhong, Peilin Chen, Zhijiang Zhang, Shiqi Wang*
- 179 Towards a Universal Synthetic Video Detector: From Face or Background Manipulations to Fully AI-Generated Content, *Rohit Kundu, Hao Xiong, Vishal Mohanty, Athula Balachandran, Amit K. Roy-Chowdhury*
- 180 A Polarization-Aided Transformer for Image Deblurring via Motion
* Vector Decomposition, *Duosheng Chen, Shihao Zhou,*
☆ *Jinshan Pan, Jinglei Shi, Lishen Qu, Jufeng Yang*
- 181 Satellite Observations Guided Diffusion Model for Accurate
* Meteorological States at Arbitrary Resolution, *Siwei Tu, Ben Fei,*

- Weidong Yang, Fenghua Ling, Hao Chen, Zili Liu, Kun Chen, Hang Fan, Wanli Ouyang, Lei Bai
- 182 Automatic Spectral Calibration of Hyperspectral Images: Method, ☆ Dataset and Benchmark, Zhuoran Du, Shaodi You, Cheng Cheng, Shikui Wei
- 183 VolFormer: Explore More Comprehensive Cube Interaction for Hyperspectral Image Restoration and Beyond, Dabing Yu, Zheng Gao
- 184 One Model for ALL: Low-Level Task Interaction Is a Key to Task-Agnostic Image Fusion, Chunyang Cheng, Tianyang Xu, Zhenhua Feng, Xiaojun Wu, Zhangyong Tang, Hui Li, Zeyang Zhang, Sara Atito, Muhammad Awais, Josef Kittler
- 185 Continuous Adverse Weather Removal via Degradation-Aware Distillation, Xin Lu, Jie Xiao, Yurui Zhu, Xueyang Fu
- 186 MambalRv2: Attentive State Space Restoration, Hang Guo, Yong Guo, Yahua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, Yawei Li
- 187 TSP-Mamba: The Travelling Salesman Problem Meets Mamba for Image Super-resolution and Beyond, Kun Zhou, Xinyu Lin, Jiangbo Lu
- 188 Semi-Supervised State-Space Model with Dynamic Stacking ○ Filter for Real-World Video Deraining, Shangquan Sun, Wenqi Ren, Juxiang Zhou, Shu Wang, Jianhou Gan, Xiaochun Cao
- 189 GenDeg: Diffusion-based Degradation Synthesis for Generalizable All-In-One Image Restoration, Sudarshan Rajagopalan, Nithin Gopalakrishnan Nair, Jay N. Paranjape, Vishal M. Patel
- 190 Generalized Recorrupted-to-Recorrupted: Self-Supervised Learning Beyond Gaussian Noise, Brayan Monroy, Jorge Bacca, Julián Tachella
- 191 Degradation-Aware Feature Perturbation for All-in-One Image ☆ Restoration, Xiangpeng Tian, Xiangyu Liao, Xiao Liu, Meng Li, Chao Ren
- 192 Exploring Semantic Feature Discrimination for Perceptual Image ☆ Super-Resolution and Opinion-Unaware No-Reference Image Quality Assessment, Guanglu Dong, Xiangyu Liao, Mingyang Li, Guihuan Guo, Chao Ren
- 193 FaithDiff: Unleashing Diffusion Priors for Faithful Image Super-resolution, Junyang Chen, Jinshan Pan, Jiangxin Dong
- 194 DEAL: Data-Efficient Adversarial Learning for High-Quality Infrared Imaging, Zhu Liu, Zijun Wang, Jinyuan Liu, Fanqi Meng, Long Ma, Risheng Liu
- 195 Adversarial Diffusion Compression for Real-World Image Super- ☆ Resolution, Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, Lei Zhang
- 196 All-Optical Nonlinear Diffractive Deep Network for Ultrafast * Image Denoising, Xiaoling Zhou, Zhemg Lee, Wei Ye, Rui Xie, Wenbo Zhang, Guanju Peng, Zongze Li, Shikun Zhang
- 197 Deterministic Image-to-Image Translation via Denoising Brownian Bridge Models with Dual Approximators, Bohan Xiao, Peiyong Wang, Qisheng He, Ming Dong
- 198 Towards Smart Point-and-Shoot Photography, Jiawan Li, Fei Zhou, Zhipeng Zhong, Jiongzhi Lin, Guoping Qiu
- 199 MetaShadow: Object-Centered Shadow Detection, Removal, ☆ and Synthesis, Tianyu Wang, Jianming Zhang, Haitian Zheng, Zhihong Ding, Scott Cohen, Zhe Lin, Wei Xiong, Chi-Wing Fu, Luis Figueroa, Soo Ye Kim
- 200 Erasing Undesirable Influence in Diffusion Models, Jing Wu, Trung Le, Munawar Hayat, Mehrtash Harandi
- 201 EntityErasure: Erasing Entity Cleanly via Amodal Entity Segmentation and Completion, Yixing Zhu, Qing Zhang, Yitong Wang, Yongwei Nie, Wei-Shi Zheng
- 202 ITA-MDT: Image-Timestep-Adaptive Masked Diffusion Transformer Framework for Image-Based Virtual Try-On, Ji Woo Hong, Tri Ton, Trung X. Pham, Gwanhyeong Koo, Sunjae Yoon, Chang D. Yoo
- 203 Latent Space Imaging, Matheus Souza, Yidan Zheng, Kaizhang Kang, Yogeshwar Nath Mishra, Qiang Fu, Wolfgang Heidrich
- 204 Q-DiT: Accurate Post-Training Quantization for Diffusion Transformers, Lei Chen, Yuan Meng, Chen Tang, Xinzhu Ma, Jingyan Jiang, Xin Wang, Zhi Wang, Wenwu Zhu
- 205 FlexiDiT: Your Diffusion Transformer Can Easily Generate High- * Quality Samples with Less Compute, Sotiris Anagnostidis, Gregor Bachmann, Yeongmin Kim, Jonas Kohler, Markos Georgopoulos, Artsiom Sanakoyeu, Yuming Du, Albert Pumarola, Ali Thabet, Edgar Schönfeld
- 206 Consistency Posterior Sampling for Diverse Image Synthesis, Vishal Purohit, Matthew Repasky, Jianfeng Lu, Qiang Qiu, Yao Xie, Xiuyuan Cheng
- 207 Domain Adaptive Diabetic Retinopathy Grading with Model Absence and Flowing Data, Wenxin Su, Song Tang, Xiaofeng Liu, Xiaojing Yi, Mao Ye, Chunxiao Zu, Jiahao Li, Xiatian Zhu
- 208 Diff2Flow: Training Flow Matching Models via Diffusion Model Alignment, Johannes Schusterbauer, Ming Gui, Frank Fundel, Björn Ommer
- 209 SoftVQ-VAE: Efficient 1-Dimensional Continuous Tokenizer, Hao Chen, Ze Wang, Xiang Li, Ximeng Sun, Fangyi Chen, Jiang Liu, Jindong Wang, Bhiksha Raj, Zicheng Liu, Emad Barsoum
- 210 SAR3D: Autoregressive 3D Object Generation and Understanding via Multi-scale 3D VQVAE, Yongwei Chen, Yushi Lan, Shangchen Zhou, Tengfei Wang, Xingang Pan
- 211 Sketch Down the FLOPs: Towards Efficient Networks for Human Sketch, Aneeshan Sain, Subhajit Maity, Pinaki Nath Chowdhury, Shubhadeep Koley, Ayan Kumar Bhunia, Yi-Zhe Song
- 212 FlipSketch: Flipping Static Drawings to Text-Guided Sketch Animations, Hmrishav Bandyopadhyay, Yi-Zhe Song
- 213 ShotAdapter: Text-to-Multi-Shot Video Generation with Diffusion ☆ Models, Ozgur Kara, Krishna Kumar Singh, Feng Liu, Duygu Ceylan, James M. Rehg, Tobias Hinz
- 214 AnyEdit: Mastering Unified High-Quality Image Editing for Any ○ Idea, Qifan Yu, Wei Chow, Zhongqi Yue, Kaihang Pan, Yang Wu, Xiaoyang Wan, Juncheng Li, Siliang Tang, Hanwang Zhang, Yueting Zhuang
- 215 VIREs: Video Instance Repainting via Sketch and Text Guided Generation, Shuchen Weng, Haojie Zheng, Peixuan Zhang, Yuchen Hong, Han Jiang, Si Li, Boxin Shi
- 216 FADE: Frequency-Aware Diffusion Model Factorization for Video Editing, Yixuan Zhu, Haolin Wang, Shilin Ma, Wenliang Zhao, Yansong Tang, Lei Chen, Jie Zhou
- 217 PICD: Versatile Perceptual Image Compression with Diffusion Rendering, Tongda Xu, Jiahao Li, Bin Li, Yan Wang, Ya-Qin Zhang, Yan Lu
- 218 Color Alignment in Diffusion, Ka Chun Shum, Binh-Son Hua, Duc Thanh Nguyen, Sai-Kit Yeung
- 219 Geometry in Style: 3D Stylization via Surface Normal Deformation, ☆ Nam Anh Dinh, Itai Lang, Hyunwoo Kim, Oded Stein, Rana Hanocka
- 220 SaMam: Style-aware State Space Model for Arbitrary Image Style * Transfer, Hongda Liu, Longguang Wang, Ye Zhang, Ziru Yu, Yulan Guo
- 221 Unveil Inversion and Invariance in Flow Transformer for Versatile ☆ Image Editing, Pengcheng Xu, Boyuan Jiang, Xiaobin Hu, Donghao Luo, Qingdong He, Jiangning Zhang, Chengjie Wang, Yunsheng Wu, Charles Ling, Boyu Wang
- 222 h-Edit: Effective and Flexible Diffusion-Based Editing via Doob's h-Transform, Toan Nguyen, Kien Do, Duc Kieu, Thin Nguyen
- 223 Concept Lancet: Image Editing with Compositional Representation Transplant, Jinqi Luo, Tianjiao Ding, Kwan Ho Ryan Chan, Hancheng Min, Chris Callison-Burch, Rene Vidal
- 224 Instruct-CLIP: Improving Instruction-Guided Image Editing with Automated Data Refinement Using Contrastive Learning, Sherry X. Chen, Misha Sra, Pradeep Sen
- 225 GlyphMastero: A Glyph Encoder for High-Fidelity Scene Text Editing, Tong Wang, Ting Liu, Xiaochao Qu, Chengjing Wu, Luoqi Liu, Xiaolin Hu
- 226 DreamOmni: Unified Image Generation and Editing, Bin Xia, Yuechen Zhang, Jingyao Li, Chengyao Wang, Yitong Wang, Xinglong Wu, Bei Yu, Jiaya Jia
- 227 Black Hole-Driven Identity Absorbing in Diffusion Models, Muhammad Shaheryar, Jong Taek Lee, Soon Ki Jung
- 228 DreamText: High Fidelity Scene Text Synthesis, Yibin Wang, Weizhong Zhang, Honghui Xu, Cheng Jin
- 229 Prompt2Perturb (P2P): Text-Guided Diffusion-Based Adversarial Attack on Breast Ultrasound Images, Yasamin Medghalchi, Moein Heidari, Clayton Allard, Leonid Sigal, Ilker Hacihaliloglu

- 230 A Comprehensive Study of Decoder-Only LLMs for Text-to-Image Generation, *Andrew Z. Wang, Songwei Ge, Tero Karras, Ming-Yu Liu, Yogesh Balaji*
- 231 Exploring the Deep Fusion of Large Language Models and Diffusion Transformers for Text-to-Image Synthesis, *Bingda Tang, Boyang Zheng, Sayak Paul, Saining Xie*
- 232 Stretching Each Dollar: Diffusion Training from Scratch on a Micro-Budget, *Vikash Sehwal, Xianghao Kong, Jingtao Li, Michael Spranger, Lingjuan Lyu*
- 233 Enhancing Creative Generation on Stable Diffusion-based Models, *Jiyeon Han, Dahee Kwon, Gayoung Lee, Junho Kim, Jaesik Choi*
- 234 APT: Adaptive Personalized Training for Diffusion Models with Limited Data, *JungWoo Chae, Jiyeon Kim, JaeWoong Choi, Kyungyul Kim, Sangheum Hwang*
- 235 InPO: Inversion Preference Optimization with Reparametrized DDIM for Efficient Diffusion Model Alignment, *Yunhong Lu, Qichao Wang, Hengyuan Cao, Xierui Wang, Xiaoyin Xu, Min Zhang*
- 236 STEPS: Sequential Probability Tensor Estimation for Text-to-Image Hard Prompt Search, *Yuning Qiu, Andong Wang, Chao Li, Haonan Huang, Guoxu Zhou, Qibin Zhao*
- 237 PQPP: A Joint Benchmark for Text-to-Image Prompt and Query Performance Prediction, *Eduard Poesina, Adriana Valentina Costache, Adrian-Gabriel Chifu, Josiane Mothe, Radu Tudor Ionescu*
- 238 Let's Verify and Reinforce Image Generation Step by Step, *Renrui Zhang, Chengzhuo Tong, Zhizheng Zhao, Ziyu Guo, Haoquan Zhang, Manyuan Zhang, Jiaming Liu, Peng Gao, Hongsheng Li*
- 239 GLASS: Guided Latent Slot Diffusion for Object-Centric Learning, *Krishnakant Singh, Simone Schaub-Meyer, Stefan Roth*
- 240 DiffSensei: Bridging Multi-Modal LLMs and Diffusion Models for Customized Manga Generation, *Jianzong Wu, Chao Tang, Jingbo Wang, Yanhong Zeng, Xiangtai Li, Yunhai Tong*
- 241 POSTA: A Go-to Framework for Customized Artistic Poster Generation, *Haoyu Chen, Xiaojie Xu, Wenbo Li, Jingjing Ren, Tian Ye, Songhua Liu, Ying-Cong Chen, Lei Zhu, Xinchao Wang*
- 242 StageDesigner: Artistic Stage Generation for Scenography via Theater Scripts, *Zhaoxing Gan, Mengtian Li, Ruhua Chen, Zhongxia Ji, Sichen Guo, Huanling Hu, Guangnan Ye, Zuo Hu*
- 243 Pattern Analogies: Learning to Perform Programmatic Image Edits by Analogy, *Aditya Ganeshan, Thibault Groueix, Paul Guerrero, Radomir Mech, Matthew Fisher, Daniel Ritchie*
- 244 Text-Driven Fashion Image Editing with Compositional Concept Learning and Counterfactual Abduction, *Shanshan Huang, Haoxuan Li, Chunyuan Zheng, Mingyuan Ge, Wei Gao, Lei Wang, Li Liu*
- 245 Controllable Human Image Generation with Personalized Multi-Garments, *Yisul Choi, Sangkyung Kwak, Sihyun Yu, Hyungwon Choi, Jinwoo Shin*
- 246 AIM-Fair: Advancing Algorithmic Fairness via Selectively Fine-Tuning Biased Models with Contextual Synthetic Data, *Zengqun Zhao, Ziquan Liu, Yu Cao, Shaogang Gong, Ioannis Patras*
- 247 Precise, Fast, and Low-cost Concept Erasure in Value Space: Orthogonal Complement Matters, *Yuan Wang, Ouxiang Li, Tingting Mu, Yanbin Hao, Kuien Liu, Xiang Wang, Xiangnan He*
- 248 Six-CD: Benchmarking Concept Removals for Text-to-image Diffusion Models, *Jie Ren, Kangrui Chen, Yingqian Cui, Shenglai Zeng, Hui Liu, Yue Xing, Jiliang Tang, Lingjuan Lyu*
- 249 Implicit Bias Injection Attacks against Text-to-Image Diffusion Models, *Huayang Huang, Xiangye Jin, Jiaxu Miao, Yu Wu*
- 250 Are Images Indistinguishable to Humans Also Indistinguishable to Classifiers?, *Zebin You, Xinyu Zhang, Hanzhong Guo, Jingdong Wang, Chongxuan Li*
- 251 Nearly Zero-Cost Protection Against Mimicry by Personalized Diffusion Models, *Namhyuk Ahn, KiYoon Yoo, Wonhyuk Ahn, Daesik Kim, Seung-Hun Nam*
- 252 Fingerprinting Denoising Diffusion Probabilistic Models, *Huan Teng, Yuhui Quan, Chengyu Wang, Jun Huang, Hui Ji*
- 253 Where's the Liability in the Generative Era? Recovery-based Black-Box Detection of AI-Generated Content, *Haoyue Bai, Yiyu Sun, Wei Cheng, Haifeng Chen*
- 254 SIDA: Social Media Image Deepfake Detection, Localization and Explanation with Large Multimodal Model, *Zhenglin Huang, Jinwei Hu, Xiangtai Li, Yiwei He, Xingyu Zhao, Bei Peng, Baoyuan Wu, Xiaowei Huang, Guangliang Cheng*
- 255 Be More Specific: Evaluating Object-centric Realism in Synthetic Images, *Anqi Liang, Ciprian Corneanu, Qianli Feng, Giorgio Giannone, Aleix Martinez*
- 256 NSD-Imagery: A Benchmark Dataset for Extending fMRI Vision Decoding Methods to Mental Imagery, *Reese Kneeland, Paul S. Scotti, Ghislain St-Yves, Jesse Breedlove, Kendrick Kay, Thomas Naselaris*
- 257 GG-SSMs: Graph-Generating State Space Models, *Nikola Zubic, Davide Scaramuzza*
- 258 Gaze-LLE: Gaze Target Estimation via Large-Scale Learned Encoders, *Fiona Ryan, Ajay Bati, Sangmin Lee, Daniel Bolya, Judy Hoffman, James M. Rehg*
- 259 EgoLife: Towards Egocentric Life Assistant, *Jingkang Yang, Shuai Liu, Hongming Guo, Yuhao Dong, Xiamengwei Zhang, Sicheng Zhang, Pengyun Wang, Zitang Zhou, Binzhu Xie, Ziyue Wang, Bei Ouyang, Zhengyu Lin, Marco Cominelli, Zhongang Cai, Bo Li, Yuanhan Zhang, Peiyuan Zhang, Fangzhou Hong, Joerg Widmer, Francesco Gringoli, Lei Yang, Ziwei Liu*
- 260 MMAudio: Taming Multimodal Joint Training for High-Quality Video-to-Audio Synthesis, *Ho Kei Cheng, Masato Ishii, Akio Hayakawa, Takashi Shibuya, Alexander Schwing, Yuki Mitsufuji*
- 261 Foley-Flow: Coordinated Video-to-Audio Generation with Masked Audio-Visual Alignment and Dynamic Conditional Flows, *Shentong Mo, Yibing Song*
- 262 Robust Audio-Visual Segmentation via Audio-Guided Visual Convergent Alignment, *Chen Liu, Peike Li, Liying Yang, Dadong Wang, Lincheng Li, Xin Yu*
- 263 Learning Audio-guided Video Representation with Gated Attention for Video-Text Retrieval, *Boseung Jeong, Jicheol Park, Sungyeon Kim, Suha Kwak*
- 264 SAM2-LOVE: Segment Anything Model 2 in Language-aided Audio-Visual Scenes, *Yuji Wang, Haoran Xu, Yong Liu, Jiaze Li, Yansong Tang*
- 265 Sound Bridge: Associating Egocentric and Exocentric Videos via Audio Cues, *Sihong Huang, Jiaxin Wu, Xiaoyong Wei, Yi Cai, Dongmei Jiang, Yaowei Wang*
- 266 VideoEspresso: A Large-Scale Chain-of-Thought Dataset for Fine-Grained Video Reasoning via Core Frame Selection, *Songhao Han, Wei Huang, Hairong Shi, Le Zhuo, Xiu Su, Shifeng Zhang, Xu Zhou, Xiaojuan Qi, Yue Liao, Si Liu*
- 267 BASKET: A Large-Scale Video Dataset for Fine-Grained Skill Estimation, *Yulu Pan, Ce Zhang, Gedas Bertasius*
- 268 SEAL: Semantic Attention Learning for Long Video Representation, *Lan Wang, Yujia Chen, Du Tran, Vishnu Naresh Boddeti, Wen-Sheng Chu*
- 269 Unified Dense Prediction of Video Diffusion, *Lehan Yang, Lu Qi, Xiangtai Li, Sheng Li, Varun Jampani, Ming-Hsuan Yang*
- 270 InstanceCap: Improving Text-to-Video Generation via Instance-aware Structured Caption, *Tiehan Fan, Kepan Nan, Rui Xie, Penghao Zhou, Zhenheng Yang, Chaoyou Fu, Xiang Li, Jian Yang, Ying Tai*
- 271 MovieBench: A Hierarchical Movie Level Dataset for Long Video Generation, *Weijia Wu, Mingyu Liu, Zeyu Zhu, Xi Xia, Haoen Feng, Wen Wang, Kevin Qinghong Lin, Chunhua Shen, Mike Zheng Shou*
- 272 Seeing Far and Clearly: Mitigating Hallucinations in MLLMs with Attention Causal Decoding, *Feilong Tang, Chengzhi Liu, Zhongxing Xu, Ming Hu, Zile Huang, Haochen Xue, Ziyang Chen, Zelin Peng, Zhiwei Yang, Sijin Zhou, Wenxue Li, Yulong Li, Wenxuan Song, Shiyao Su, Wei Feng, Jionglong Su, Mingquan Lin, Yifan Peng, Xuelian Cheng, Imran Razzak, Zongyuan Ge*
- 273 SeriesBench: A Benchmark for Narrative-Driven Drama Series Understanding, *Chenkai Zhang, Yiming Lei, Zeming Liu, Haitao Leng, ShaoGuo Liu, Tingting Gao, Qingjie Liu, Yunhong Wang*
- 274 DocVLM: Make Your VLM an Efficient Reader, *Mor Shpigel Nacson, Aviad Aberdam, Roy Ganz, Elad Ben Avraham, Alona Golts, Yair Kittenplon, Shai Mazon, Ron Litman*

- 275 Which Viewpoint Shows it Best? Language for Weakly Supervising
 * View Selection in Multi-view Instructional Videos,
 ☆ Sagnik Majumder, Tushar Nagarajan, Ziad Al-Halah, Reina Pradhan, Kristen Grauman
- 276 VideoWorld: Exploring Knowledge Learning from Unlabeled Videos, Zhongwei Ren, Yunchao Wei, Xun Guo, Yao Zhao, Bingyi Kang, Jiashi Feng, Xiaojie Jin
- 277 ReSpec: Relevance and Specificity Grounded Online Filtering for Learning on Video-Text Data Streams, Chris Dongjoo Kim, Jihwan Moon, Sangwoo Moon, Heeseung Yun, Sihaeng Lee, Aniruddha Kembhavi, Soonyoung Lee, Gunhee Kim, Sangho Lee, Christopher Clark
- 278 Unbiasing through Textual Descriptions: Mitigating Representation Bias in Video Benchmarks, Nina Shvetsova, Arsha Nagrani, Bernt Schiele, Hilde Kuehne, Christian Rupprecht
- 279 VideoComp: Advancing Fine-Grained Compositional and Temporal Alignment in Video-Text Models, Dahun Kim, AJ Piergiovanni, Ganesh Mallya, Anelia Angelova
- 280 Flexible Frame Selection for Efficient Video Reasoning, ☆ Shyamal Buch, Arsha Nagrani, Anurag Arnab, Cordelia Schmid
- 281 LiveCC: Learning Video LLM with Streaming Speech Transcription at Scale, Joya Chen, Ziyun Zeng, Yiqi Lin, Wei Li, Zejun Ma, Mike Zheng Shou
- 282 BIMBA: Selective-Scan Compression for Long-Range Video Question Answering, Md Mohaiminul Islam, Tushar Nagarajan, Huiyu Wang, Gedas Bertasius, Lorenzo Torresani
- 283 SF2T: Self-supervised Fragment Finetuning of Video-LLMs for Fine-Grained Understanding, Yangliu Hu, Zikai Song, Na Feng, Yawei Luo, Junqing Yu, Yi-Ping Phoebe Chen, Wei Yang
- 284 Adaptive Keyframe Sampling for Long Video Understanding, Xi Tang, Jihao Qiu, Lingxi Xie, Yunjie Tian, Jianbin Jiao, Qixiang Ye
- 285 Efficient Transfer Learning for Video-language Foundation Models, Haoxing Chen, Zizheng Huang, Yan Hong, Yanshuo Wang, Zhongcai Lyu, Zhuoer Xu, Jun Lan, Zhangxuan Gu
- 286 EventGPT: Event Stream Understanding with Multimodal Large Language Models, Shaoyu Liu, Jianing Li, Guanghui Zhao, Yunjian Zhang, Xin Meng, Fei Richard Yu, Xiangyang Ji, Ming Li
- 287 HyperGLM: HyperGraph for Video Scene Graph Generation and Anticipation, Trong-Thuan Nguyen, Pha Nguyen, Jackson Cothren, Alper Yilmaz, Khoa Luu
- 288 DiffVsgg: Diffusion-Driven Online Video Scene Graph Generation, Mu Chen, Liulei Li, Wenguan Wang, Yi Yang
- 289 CASAGPT: Cuboid Arrangement and Scene Assembly for Interior Design, Weitao Feng, Hang Zhou, Jing Liao, Li Cheng, Wenbo Zhou
- 290 The Devil is in Temporal Token: High Quality Video Reasoning Segmentation, Sitong Gong, Yunzhi Zhuge, Lu Zhang, Zongxin Yang, Pingping Zhang, Huchuan Lu
- 291 M³-VOS: Multi-Phase, Multi-Transition, and Multi-Scenery Video Object Segmentation, Zixuan Chen, Jiaxin Li, Junxuan Liang, Liming Tan, Yejie Guo, Cewu Lu, Yong-Lu Li
- 292 Anomize: Better Open Vocabulary Video Anomaly Detection, Fei Li, Wenxuan Liu, Jingjing Chen, Ruixu Zhang, Yuran Wang, Xian Zhong, Zheng Wang
- 293 UniSTD: Towards Unified Spatio-Temporal Learning across Diverse Disciplines, Chen Tang, Xinzhu Ma, Encheng Su, Xiufeng Song, Xiaohong Liu, Wei-Hong Li, Lei Bai, Wanli Ouyang, Xiangyu Yue
- 294 Temporal Action Detection Model Compression by Progressive Block Drop, Xiaoyong Chen, Yong Guo, Jiaming Liang, Sitong Zhuang, Runhao Zeng, Xiping Hu
- 295 Period-LLM: Extending the Periodic Capability of Multimodal Large Language Model, Yuting Zhang, Hao Lu, Qingyong Hu, Yin Wang, Kaishen Yuan, Xin Liu, Kaishun Wu
- 296 Revealing Key Details to See Differences: A Novel Prototypical Perspective for Skeleton-based Action Recognition, Hongda Liu, Yunfan Liu, Min Ren, Hao Wang, Yunlong Wang, Zhenan Sun
- 297 DiSciPLE: Learning Interpretable Programs for Scientific Visual Discovery, Utkarsh Mall, Cheng Perng Phoo, Mia Chiquier, Bharath Hariharan, Kavita Bala, Carl Vondrick
- 298 Divide and Conquer: Heterogeneous Noise Integration for Diffusion-based Adversarial Purification, Gaozheng Pei, Shaojie Lyu, Gong Chen, Ke Ma, Qianqian Xu, Yingfei Sun, Qingming Huang
- 299 SDBF: Steep-Decision-Boundary Fingerprinting for Hard-Label Tampering Detection of DNN Models, Xiaofan Bai, Shixin Li, Xiaojing Ma, Bin Benjamin Zhu, Dongmei Zhang, Linchen Yu
- 300 From Head to Tail: Efficient Black-box Model Inversion Attack via Long-tailed Learning, Ziang Li, Hongguang Zhang, Juan Wang, Meihui Chen, Hongxin Hu, Wenzhe Yi, Xiaoyang Xu, Mengda Yang, Chenjun Ma
- 301 UMFN: Unified Multi-Domain Face Normalization for Joint Cross-domain Prototype Learning and Heterogeneous Face Recognition, Meng Pang, Wenjun Zhang, Nanrun Zhou, Shengbo Chen, Hong Rao
- 302 MEET: Towards Memory-Efficient Temporal Sparse Deep Neural Networks, Zeqi Zhu, Ibrahim Batuhan Akkaya, Luc Waeijen, Egor Bondarev, Arash Pourtaherian, Orlando Moreira
- 303 Object Detection using Event Camera: A MoE Heat Conduction based Detector and A New Benchmark Dataset, Xiao Wang, Yu Jin, Wentao Wu, Wei Zhang, Lin Zhu, Bo Jiang, Yonghong Tian
- 304 Person De-reidentification: A Variation-guided Identity Shift Modeling, Yi-Xing Peng, Yu-Ming Tang, Kun-Yu Lin, Qize Yang, Jingke Meng, Xihan Wei, Wei-Shi Zheng
- 305 WISE: A Framework for Gigapixel Whole-Slide-Image Lossless Compression, Yu Mao, Jun Wang, Nan Guan, Chun Jason Xue
- 306 BOE-ViT: Boosting Orientation Estimation with Equivariance in Self-Supervised 3D Subtomogram Alignment, Runmin Jiang, Jackson Daggett, Shriya Pingulkar, Yizhou Zhao, Priyanshu Dhingra, Daniel Brown, Qifeng Wu, Xiangrui Zeng, Xingjian Li, Min Xu
- 307 Point-to-Region Loss for Semi-Supervised Point-Based Crowd Counting, Wei Lin, Chenyang Zhao, Antoni B. Chan
- 308 SP3D: Boosting Sparsely-Supervised 3D Object Detection via Accurate Cross-Modal Semantic Prompts, Shijia Zhao, Qiming Xia, Xusheng Guo, Pufan Zou, Maoji Zheng, Hai Wu, Chenglu Wen, Cheng Wang
- 309 Segment Anything, Even Occluded, Wei-En Tai, Yu-Lin Shih, Cheng Sun, Yu-Chiang Frank Wang, Hwann-Tzong Chen
- 310 BFANet: Revisiting 3D Semantic Segmentation with Boundary Feature Analysis, Weiguang Zhao, Rui Zhang, Qiufeng Wang, Guangliang Cheng, Kaizhu Huang
- 311 SCSegamba: Lightweight Structure-Aware Vision Mamba for Crack Segmentation in Structures, Hui Liu, Chen Jia, Fan Shi, Xu Cheng, Shengyong Chen
- 312 Towards Continual Universal Segmentation, Zihan Lin, Zilei Wang, Xu Wang
- 313 Segment This Thing: Foveated Tokenization for Efficient Point-Prompted Segmentation, Tanner Schmidt, Richard Newcombe
- 314 Probabilistic Prompt Distribution Learning for Animal Pose Estimation, Jiyong Rao, Brian Nlong Zhao, Yu Wang
- 315 Navigating the Unseen: Zero-shot Scene Graph Generation via Capsule-Based Equivariant Features, Wenhuan Huang, Yi Ji, Guiqian Zhu, Li Ying, Chunping Liu
- 316 ASHiTA: Automatic Scene-grounded Hierarchical Task Analysis, Yun Chang, Leonor Feroselle, Duy Ta, Bernadette Bucher, Luca Carlone, Jiuguang Wang
- 317 LayoutVLM: Differentiable Optimization of 3D Layout via Vision-Language Models, Fan-Yun Sun, Weiyu Liu, Siyi Gu, Dylan Lim, Goutam Bhat, Federico Tombari, Manling Li, Nick Haber, Jiajun Wu
- 318 Vision-Language Embodiment for Monocular Depth Estimation, Jinchang Zhang, Guoyu Lu
- 319 SpiritSight Agent: Advanced GUI Agent with One Look, Zhiyuan Huang, Ziming Cheng, Junting Pan, Zhaohui Hou, Mingjie Zhan
- 320 3D-GRAND: A Million-Scale Dataset for 3D-LLMs with Better Grounding and Less Hallucination, Jianing Yang, Xuwei Chen, Nikhil Madaan, Madhavan Iyengar, Shengyi Qian, David F. Fouhey, Joyce Chai
- 321 Collaborative Tree Search for Enhancing Embodied Multi-Agent Collaboration, Lizheng Zu, Lin Lin, Song Fu, Na Zhao, Pan Zhou
- 322 CTRL-O: Language-Controllable Object-Centric Visual Representation Learning, Aniket Didolkar, Andrii Zadaianchuk, Rabiul Awal, Maximilian Seitzer, Efstratios Gavves, Aishwarya Agrawal

- 323 VLMs-Guided Representation Distillation for Efficient Vision-Based Reinforcement Learning, *Haoran Xu, Peixi Peng, Guang Tan, Yiqian Chang, Luntong Li, Yonghong Tian*
- 324 VLsI: Verbalized Layers-to-Interactions from Large to Small Vision Language Models, *Byung-Kwan Lee, Ryo Hachiuma, Yu-Chiang Frank Wang, Yong Man Ro, Yueh-Hua Wu*
- 325 Adaptive Markup Language Generation for Contextually-Grounded Visual Document Understanding, *Han Xiao, Yina Xie, Guanxin Tan, Yinghao Chen, Rui Hu, Ke Wang, Aojun Zhou, Hao Li, Hao Shao, Xudong Lu, Peng Gao, Yafei Wen, Xiaoxin Chen, Shuai Ren, Hongsheng Li*
- 326 CoSpace: Benchmarking Continuous Space Perception Ability for Vision-Language Models, *Yiqi Zhu, Ziyue Wang, Can Zhang, Peng Li, Yang Liu*
- 327 Automated Generation of Challenging Multiple-Choice Questions for Vision Language Model Evaluation, *Yuhui Zhang, Yuchang Su, Yiming Liu, Xiaohan Wang, James Burgess, Elaine Sui, Chenyu Wang, Josiah Aklilu, Alejandro Lozano, Anjiang Wei, Ludwig Schmidt, Serena Yeung-Levy*
- 328 CocoER: Aligning Multi-Level Feature by Competition and Coordination for Emotion Recognition, *Xuli Shen, Hua Cai, Weilin Shen, Qing Xu, Dingding Yu, Weifeng Ge, Xiangyang Xue*
- 329 LoRASculpt: Sculpting LoRA for Harmonizing General and
☆ Specialized Knowledge in Multimodal Large Language Models, *Jian Liang, Wenke Huang, Guancheng Wan, Qu Yang, Mang Ye*
- 330 Seek Common Ground While Reserving Differences: Semi-Supervised Image-Text Sentiment Recognition, *Wuyou Xia, Guoli Jia, Sicheng Zhao, Jufeng Yang*
- 331 Vision-Language Models Do Not Understand Negation,
☆ *Kumail Alhamoud, Shaden Alshammari, Yonglong Tian, Guohao Li, Philip H.S. Torr, Yoon Kim, Marzyeh Ghassemi*
- 332 Alignment, Mining and Fusion: Representation Alignment with Hard Negative Mining and Selective Knowledge Fusion for Medical Visual Question Answering, *Yuanhao Zou, Zhaozheng Yin*
- 333 Hybrid Global-Local Representation with Augmented Spatial Guidance for Zero-Shot Referring Image Segmentation, *Ting Liu, Siyuan Li*
- 334 Generative Multimodal Pretraining with Discrete Diffusion Timestep
🔊 Tokens, *Kaihang Pan, Wang Lin, Zhongqi Yue, Tenglong Ao, Liyu Jia, Wei Zhao, Juncheng Li, Silian Tang, Hanwang Zhang*
- 335 UNIALIGN: Scaling Multimodal Alignment within One Unified Model, *Bo Zhou, Liulei Li, Yujia Wang, Huafeng Liu, Yazhou Yao, Wenguan Wang*
- 336 SpatialCLIP: Learning 3D-aware Image Representations from Spatially Discriminative Language, *Zehan Wang, Sashuai Zhou, Shaoxuan He, Haifeng Huang, Lihe Yang, Ziang Zhang, Xize Cheng, Shengpeng Ji, Tao Jin, Hengshuang Zhao, Zhou Zhao*
- 337 Semantic and Expressive Variations in Image Captions Across Languages, *Andre Ye, Sebastin Santy, Jena D. Hwang, Amy X. Zhang, Ranjay Krishna*
- 338 ReCon: Enhancing True Correspondence Discrimination through Relation Consistency for Robust Noisy Correspondence Learning, *Quanxing Zha, Xin Liu, Shu-Juan Peng, Yiu-ming Cheung, Xing Xu, Nannan Wang*
- 339 Video-XL: Extra-Long Vision Language Model for Hour-Scale
🔊 Video Understanding, *Yan Shu, Zheng Liu, Peitian Zhang, Minghao Qin, Junjie Zhou, Zhengyang Liang, Tiejun Huang, Bo Zhao*
- 340 Generative Zero-Shot Composed Image Retrieval, *Lan Wang, Wei Ao, Vishnu Naresh Boddeti, Ser-Nam Lim*
- 341 IDEA: Inverted Text with Cooperative Deformable Aggregation for Multi-modal Object Re-Identification, *Yuhao Wang, Yongfeng Lv, Pingping Zhang, Huchuan Lu*
- 342 MP-GUI: Modality Perception with MLLMs for GUI Understanding, *Ziwei Wang, Weizhi Chen, Leyang Yang, Sheng Zhou, Shengchu Zhao, Hanbei Zhan, Jiongchao Jin, Liangcheng Li, Zirui Shao, Jiajun Bu*
- 343 Towards Natural Language-Based Document Image Retrieval: New Dataset and Benchmark, *Hao Guo, Xugong Qin, Jun Jie Ou Yang, Peng Zhang, Gangyan Zeng, Yubo Li, Hailun Lin*
- 344 Incorporating Dense Knowledge Alignment into Unified Multimodal Representation Models, *Yuhao Cui, Xinxing Zu, Wenhua Zhang, Zhongzhou Zhao, Jinyang Gao*
- 345 MedUnifier: Unifying Vision-and-Language Pre-training on Medical Data with Vision Generation Task using Discrete Visual Representations, *Ziyang Zhang, Yang Yu, Yucheng Chen, Xulei Yang, Si Yong Yeo*
- 346 Non-Natural Image Understanding with Advancing Frequency-based Vision Encoders, *Wang Lin, QingSong Wang, Yueying Feng, Shulei Wang, Tao Jin, Zhou Zhao, Fei Wu, Chang Yao, Jingyuan Chen*
- 347 SynerGen-VL: Towards Synergistic Image Understanding and Generation with Vision Experts and Token Folding, *Hao Li, Changyao Tian, Jie Shao, Xizhou Zhu, Zhaokai Wang, Jinguo Zhu, Wenhan Dou, Xiaogang Wang, Hongsheng Li, Lewei Lu, Jifeng Dai*
- 348 SmartCLIP: Modular Vision-language Alignment with Identification
* Guarantees, *Shaoan Xie, Lingjing Lingjing, Yujia Zheng, Yu Yao, Zeyu Tang, Eric P. Xing, Guangyi Chen, Kun Zhang*
- 349 Advancing Myopia To Holism: Fully Contrastive Language-Image
☆ Pre-training, *Haicheng Wang, Chen Ju, Weixiong Lin, Shuai Xiao, Mengting Chen, Yixuan Huang, Chang Liu, Mingshuai Yao, Jinsong Lan, Ying Chen, Qingwen Liu, Yanfeng Wang*
- 350 Language-Guided Salient Object Ranking, *Fang Liu, Yuhao Liu, Ke Xu, Shuquan Ye, Gerhard Petrus Hancke, Rynson W. H. Lau*
- 351 HiRes-LLaVA: Restoring Fragmentation Input in High-Resolution Large Vision-Language Models, *Runhui Huang, Xinpeng Ding, Chunwei Wang, Jianhua Han, Yulong Liu, Hengshuang Zhao, Hang Xu, Lu Hou, Wei Zhang, Xiaodan Liang*
- 352 Mimic In-Context Learning for Multimodal Tasks, *Yuchu Jiang, Jiale Fu, Chenduo Hao, Xinting Hu, Yingzhe Peng, Xin Geng, Xu Yang*
- 353 VoCo-LLaMA: Towards Vision Compression with Large Language Models, *Xubing Ye, Yukang Gan, Xiaoke Huang, Yixiao Ge, Yansong Tang*
- 354 Harnessing Frozen Unimodal Encoders for Flexible Multimodal Alignment, *Mayug Maniparambil, Raiymbek Akshulakov, Yasser Abdelaziz Dahou Djilali, Sanath Narayan, Ankit Singh, Noel E. O'Connor*
- 355 Towards Understanding How Knowledge Evolves in Large Vision-Language Models, *Sudong Wang, Yunjian Zhang, Yao Zhu, Jianing Li, Zizhe Wang, Yanwei Liu, Xiangyang Ji*
- 356 Accelerating Multimodal Large Language Models by Searching Optimal Vision Token Reduction, *Shiyu Zhao, Zhenting Wang, Felix Juefei-Xu, Xide Xia, Miao Liu, Xiaofang Wang, Mingfu Liang, Ning Zhang, Dimitris N. Metaxas, Licheng Yu*
- 357 Task Preference Optimization: Improving Multimodal Large Language Models with Vision Task Alignment, *Ziang Yan, Zhilin Li, Yanan He, Chenting Wang, Kunchang Li, Xinhao Li, Xiangyu Zeng, Zilei Wang, Yali Wang, Yu Qiao, Limin Wang, Yi Wang*
- 358 HalLoc: Token-level Localization of Hallucinations for Vision Language Models, *Eunkyu Park, Minyeong Kim, Gunhee Kim*
- 359 Octopus: Alleviating Hallucination via Dynamic Contrastive
* Decoding, *Wei Suo, Lijun Zhang, Mengyang Sun, Lin Yuanbo Wu, Peng Wang, Yanning Zhang*
- 360 Mitigating Object Hallucinations in Large Vision-Language Models with Assembly of Global and Local Attention, *Wenbin An, Feng Tian, Sicong Leng, Jiahao Nie, Haonan Lin, Qianying Wang, Ping Chen, Xiaojin Zhang, Shijian Lu*
- 361 BadToken: Token-level Backdoor Attacks to Multi-modal
☆ Large Language Models, *Zenghui Yuan, Jiawen Shi, Pan Zhou, Neil Zhenqiang Gong, Lichao Sun*
- 362 Playing the Fool: Jailbreaking LLMs and Multimodal LLMs with Out-of-Distribution Strategy, *Joonhyun Jeong, Seyun Bae, Yeonsung Jung, Jaeryong Hwang, Eunho Yang*
- 363 Steering Away from Harm: An Adaptive Approach to Defending Vision Language Model Against Jailbreaks, *Han Wang, Gang Wang, Huan Zhang*
- 364 R-TPT: Improving Adversarial Robustness of Vision-Language

- Models through Test-Time Prompt Tuning, *Lijun Sheng, Jian Liang, Zilei Wang, Ran He*
- 365 ResCLIP: Residual Attention for Training-free Dense Vision-language Inference, *Yuhang Yang, Jinhong Deng, Wen Li, Lixin Duan*
- 366 Enhanced OoD Detection through Cross-Modal Alignment of Multi-Modal Representations, *Jeonghyeon Kim, Sangheum Hwang*
- 367 Rethinking Few-Shot Adaptation of Vision-Language Models in Two Stages, *Matteo Farina, Massimiliano Mancini, Giovanni Iacca, Elisa Ricci*
- 368 Bayesian Test-Time Adaptation for Vision-Language Models, *Lihua Zhou, Mao Ye, Shuaifeng Li, Nianxin Li, Xiatian Zhu, Lei Deng, Hongbin Liu, Zhen Lei*
- 369 Cropper: Vision-Language Model for Image Cropping through In-Context Learning, *Seung Hyun Lee, Jijun Jiang, Yiran Xu, Zhuofang Li, Junjie Ke, Yinxiao Li, Junfeng He, Steven Hickson, Katie Datsenko, Sangpil Kim, Ming-Hsuan Yang, Irfan Essa, Feng Yang*
- 370 ImagineFSL: Self-Supervised Pretraining Matters on Imagined Base Set for VLM-based Few-shot Learning, *Haoyuan Yang, Xiaou Li, Jiaming Lv, Xianjun Cheng, Qilong Wang, Peihua Li*
- 371 SCAP: Transductive Test-Time Adaptation via Supportive Clique-based Attribute Prompting, *Chenyu Zhang, Kunlun Xu, Zichen Liu, Yuxin Peng, Jiahuan Zhou*
- 372 Interpreting Object-level Foundation Models via Visual Precision Search, *Ruoyu Chen, Siyuan Liang, Jingzhi Li, Shiming Liu, Maosen Li, Zhen Huang, Hua Zhang, Xiaochun Cao*
- 373 Towards Fine-Grained Interpretability: Counterfactual Explanations for Misclassification with Saliency Partition, *Lintong Zhang, Kang Yin, Seong-Whan Lee*
- 374 Show and Tell: Visually Explainable Deep Neural Nets via Spatially-Aware Concept Bottleneck Models, *Itay Benou, Tammy Riklin Raviv*
- 375 VL2Lite: Task-Specific Knowledge Distillation from Large Vision-Language Models to Lightweight Networks, *Jinseong Jang, Chunfei Ma, Byeongwon Lee*
- 376 DUNE: Distilling a Universal Encoder from Heterogeneous 2D and 3D Teachers, *Mert Bülent Saryıldız, Philippe Weinzaepfel, Thomas Lucas, Pau de Jorge, Diane Larlus, Yannis Kalantidis*
- 377 Probing the Mid-level Vision Capabilities of Self-Supervised Learning, *Xuwei Chen, Markus Marks, Zezhou Cheng*
- 378 Parameter Efficient Mamba Tuning via Projector-targeted Diagonal-centric Linear Transformation, *Seokil Ham, Hee-Seon Kim, Sangmin Woo, Changick Kim*
- 379 COAP: Memory-Efficient Training with Correlation-Aware Gradient Projection, *Jinqi Xiao, Shen Sang, Tiancheng Zhi, Jing Liu, Qing Yan, Linjie Luo, Bo Yuan*
- 380 Sample- and Parameter-Efficient Auto-Regressive Image Models, *Elad Amrani, Leonid Karlinsky, Alex Bronstein*
- 381 Subnet-Aware Dynamic Supernet Training for Neural Architecture Search, *Jeimin Jeon, Youngmin Oh, Junghyup Lee, Donghyeon Baek, Dohyung Kim, Chanho Eom, Bumsu Ham*
- 382 DeepCompress-ViT: Rethinking Model Compression to Enhance Efficiency of Vision Transformers at the Edge, *Sabbir Ahmed, Abdullah Al Arafat, Deniz Najafi, Akhlak Mahmood, Mamshad Nayeem Rizve, Mohaiminul Al Nahian, Ranyang Zhou, Shaahin Angizi, Adnan Siraj Rakin*
- 383 Effective SAM Combination for Open-Vocabulary Semantic Segmentation, *Minhyeok Lee, Suhwan Cho, Jungho Lee, Sunghun Yang, Heeseung Choi, Ig-Jae Kim, Sangyoun Lee*
- 384 Adventurer: Optimizing Vision Mamba Architecture Designs for Efficiency, *Feng Wang, Timing Yang, Yaodong Yu, Sucheng Ren, Guoyizhe Wei, Angtian Wang, Wei Shao, Yuyin Zhou, Alan Yuille, Cihang Xie*
- 385 Dynamic Group Normalization: Spatio-Temporal Adaptation to Evolving Data Statistics, *Yair Smadar, Assaf Hoogi*
- 386 Frequency Dynamic Convolution for Dense Image Prediction, *Linwei Chen, Lin Gu, Liang Li, Chenggang Yan, Ying Fu*
- 387 Faster Parameter-Efficient Tuning with Token Redundancy Reduction, *Kwonyoung Kim, Jungin Park, Jin Kim, Hyeonjun Kwon, Kwanghoon Sohn*
- 388 Discovering Fine-Grained Visual-Concept Relations by Disentangled Optimal Transport Concept Bottleneck Models, *Yan Xie, Zequn Zeng, Hao Zhang, Yucheng Ding, Yi Wang, Zhengjue Wang, Bo Chen, Hongwei Liu*
- 389 TIDE: Training Locally Interpretable Domain Generalization Models Enables Test-time Correction, *Aishwarya Agarwal, Srikrishna Karanam, Vineet Gandhi*
- 390 Exploring Simple Open-Vocabulary Semantic Segmentation, *Zihang Lai*
- 391 POPEN: Preference-Based Optimization and Ensemble for LVLM-Based Reasoning Segmentation, *Lanyun Zhu, Tianrun Chen, Qianxiang Xu, Xuanyi Liu, Deyi Ji, Haiyang Wu, De Wen Soh, Jun Liu*
- 392 Multi-Label Prototype Visual Spatial Search for Weakly Supervised Semantic Segmentation, *Songsong Duan, Xi Yang, Nannan Wang*
- 393 HistoFS: Non-IID Histopathologic Whole Slide Image Classification via Federated Style Transfer with RoI-Preserving, *Farchan Hakim Raswa, Chun-Shien Lu, Jia-Ching Wang*
- 394 FFR: Frequency Feature Rectification for Weakly Supervised Semantic Segmentation, *Ziqian Yang, Xinqiao Zhao, Xiaolei Wang, Quan Zhang, Jimin Xiao*
- 395 Prototype-Based Image Prompting for Weakly Supervised Histopathological Image Segmentation, *Qingchen Tang, Lei Fan, Maurice Pagnucco, Yang Song*
- 396 Pay Attention to the Foreground in Object-Centric Learning, *Pinzhuo Tian, Shengjie Yang, Hang Yu, Alex Kot*
- 397 Attribute-formed Class-specific Concept Space: Endowing Language Bottleneck Model with Better Interpretability and Scalability, *Jiayang Zhang, Qianli Luo, Guowu Yang, Wenjing Yang, Weide Liu, Guosheng Lin, Fengmao Lv*
- 398 LOGICZSL: Exploring Logic-induced Representation for Compositional Zero-shot Learning, *Peng Wu, Xiankai Lu, Hao Hu, Yongqin Xian, Jianbing Shen, Wenguan Wang*
- 399 CLIP-driven Coarse-to-fine Semantic Guidance for Fine-grained Open-set Semi-supervised Learning, *Xiaokun Li, Yaping Huang, Qingji Guan*
- 400 Less Attention is More: Prompt Transformer for Generalized Category Discovery, *Wei Zhang, Baopeng Zhang, Zhu Teng, Wenxin Luo, Junnan Zou, Jianping Fan*
- 401 Open-World Objectness Modeling Unifies Novel Object Detection, *Shan Zhang, Yao Ni, Jinhao Du, Yuan Xue, Philip Torr, Piotr Koniusz, Anton van den Hengel*
- 402 Activating Sparse Part Concepts for 3D Class Incremental Learning, *Zhenya Tian, Jun Xiao, Lupeng Liu, Haiyong Jiang*
- 403 Learning Endogenous Attention for Incremental Object Detection, *Xiang Song, Yuhang He, Jingyuan Li, Qiang Wang, Yihong Gong*
- 404 UCOD-DPL: Unsupervised Camouflaged Object Detection via Dynamic Pseudo-label Learning, *Weiqi Yan, Lvhai Chen, Huaijia Kou, Shengchuan Zhang, Yan Zhang, Liujuan Cao*
- 405 Feature Information Driven Position Gaussian Distribution Estimation for Tiny Object Detection, *Jinghao Bian, Mingtao Feng, Weisheng Dong, Fangfang Wu, Jianqiao Luo, Yaonan Wang, Guangming Shi*
- 406 A Unified, Resilient, and Explainable Adversarial Patch Detector, *Vishesh Kumar, Akshay Agarwal*
- 407 Bayesian Prompt Flow Learning for Zero-Shot Anomaly Detection, *Zhen Qu, Xian Tao, Xinyi Gong, ShiChen Qu, Qiyu Chen, Zhengtao Zhang, Xingang Wang, Guiguang Ding*
- 408 Towards Visual Discrimination and Reasoning of Real-World Physical Dynamics: Physics-Grounded Anomaly Detection, *Wenqiao Li, Yao Gu, Xintao Chen, Xiaohao Xu, Ming Hu, Xiaonan Huang, Yingna Wu*
- 409 Dual-Interrelated Diffusion Model for Few-Shot Anomaly Image Generation, *Ying Jin, Jinlong Peng, Qingdong He, Teng Hu, Jiafu Wu, Hao Chen, Haoxuan Wang, Wenbing Zhu, Mingmin Chi, Jun Liu, Yabiao Wang*
- 410 LotusFilter: Fast Diverse Nearest Neighbor Search via a Learned Cutoff Table, *Yusuke Matsui*

- 411 FedBiP: Heterogeneous One-Shot Federated Learning with Personalized Latent Diffusion Models, *Haokun Chen, Hang Li, Yao Zhang, Jinhe Bi, Gengyuan Zhang, Yueqi Zhang, Philip Torr, Jindong Gu, Denis Krompass, Volker Tresp*
- 412 Emphasizing Discriminative Features for Dataset Distillation in Complex Scenarios, *Kai Wang, Zekai Li, Zhi-Qi Cheng, Samir Khaki, Ahmad Sajedi, Ramakrishna Vedantam, Konstantinos N Plataniotis, Alexander Hauptmann, Yang You*
- 413 Hierarchical Features Matter: A Deep Exploration of Progressive Parameterization Method for Dataset Distillation, *Xinhao Zhong, Hao Fang, Bin Chen, Xulin Gu, Meikang Qiu, Shuhan Qi, Shu-Tao Xia*
- 414 EVOS: Efficient Implicit Neural Training via EVOLUTIONary Selector, *Weixiang Zhang, Shuzhao Xie, Chengwei Ren, Siyi Xie, Chen Tang, Shijia Ge, Mingzi Wang, Zhi Wang*
- 415 Learning from Neighbors: Category Extrapolation for Long-Tail Learning, *Shizhen Zhao, Xin Wen, Jiahui Liu, Chuofan Ma, Chunfeng Yuan, Xiaojuan Qi*
- 416 PLeaS - Merging Models with Permutations and Least Squares, *Anshul Nasery, Jonathan Hayase, Pang Wei Koh, Sewoong Oh*
- 417 Everything to the Synthetic: Diffusion-driven Test-time Adaptation via Synthetic-Domain Alignment, *Jiayi Guo, Junhao Zhao, Chaogun Du, Yulin Wang, Chunjiang Ge, Zanlin Ni, Shiji Song, Humphrey Shi, Gao Huang*
- 418 SURGEON: Memory-Adaptive Fully Test-Time Adaptation via
* Dynamic Activation Sparsity, *Ke Ma, Jiaqi Tang, Bin Guo, Fan Dang, Sicong Liu, Zhui Zhu, Lei Wu, Cheng Fang, Ying-Cong Chen, Zhiwen Yu, Yunhao Liu*
- 419 Hierarchical Knowledge Prompt Tuning for Multi-task Test-Time Adaptation, *Qiang Zhang, Mengsheng Zhao, Jiawei Liu, Fanrui Zhang, Yongchao Xu, Zheng-Jun Zha*
- 420 CL-LoRA: Continual Low-Rank Adaptation for Rehearsal-Free Class-Incremental Learning, *Jiangpeng He, Zhihao Duan, Fengqing Zhu*
- 421 Dynamic Integration of Task-Specific Adapters for Class Incremental Learning, *Jiashuo Li, Shaokun Wang, Bo Qian, Yuhang He, Xing Wei, Qiang Wang, Yihong Gong*
- 422 Task-Specific Gradient Adaptation for Few-Shot One-Class Classification, *Yunlong Li, Xiabi Liu, Liyuan Pan, Yuchen Ren*
- 423 Multi-Granularity Class Prototype Topology Distillation for Class-Incremental Source-Free Unsupervised Domain Adaptation, *Peihua Deng, Jiehua Zhang, Xichun Sheng, Chenggang Yan, Yaoqi Sun, Ying Fu, Liang Li*
- 424 Balanced Direction from Multifarious Choices: Arithmetic Meta-Learning for Domain Generalization, *Xiran Wang, Jian Zhang, Lei Qi, Yinghuan Shi*
- 425 ADU: Adaptive Detection of Unknown Categories in Black-Box Domain Adaptation, *Yushan Lai, Guowen Li, Haoyuan Liang, Juepeng Zheng, Zhiyu Ye*
- 426 Unlocking the Potential of Unlabeled Data in Semi-Supervised Domain Generalization, *Dongkwan Lee, Kyomin Hwang, Nojun Kwak*
- 427 Distilling Long-tailed Datasets, *Zhenghao Zhao, Haoxuan Wang, Yuzhang Shang, Kai Wang, Yan Yan*
- 428 Open Set Label Shift with Test Time Out-of-Distribution Reference, *Changkun Ye, Russell Tsuchida, Lars Petersson, Nick Barnes*
- 429 OODD: Test-time Out-of-Distribution Detection with Dynamic Dictionary, *Yifeng Yang, Lin Zhu, Zewen Sun, Hengyu Liu, Qinying Gu, Nanyang Ye*
- 430 pFedMxF: Personalized Federated Class-Incremental Learning with Mixture of Frequency Aggregation, *Yifei Zhang, Hao Zhu, Alysa Ziyang Tan, Dianzhi Yu, Longtao Huang, Han Yu*
- 431 FedAWA: Adaptive Optimization of Aggregation Weights in Federated Learning Using Client Vectors, *Changlong Shi, He Zhao, Bingjie Zhang, Mingyuan Zhou, Dandan Guo, Yi Chang*
- 432 Unlearning through Knowledge Overwriting: Reversible Federated Unlearning via Selective Sparse Adapter, *Zhengyi Zhong, Weidong Bao, Ji Wang, Shuai Zhang, Jingxuan Zhou, Lingjuan Lyu, Wei Yang Bryan Lim*
- 433 Jailbreaking the Non-Transferable Barrier via Test-Time Data Disguising, *Yongli Xiang, Ziming Hong, Lina Yao, Dadong Wang, Tongliang Liu*
- 434 Improving the Training of Data-Efficient GANs via Quality Aware Dynamic Discriminator Rejection Sampling, *Zhaoyu Zhang, Yang Hua, Guanxiong Sun, Hui Wang, Seán McLoone*
- 435 EntropyMark: Towards More Harmless Backdoor Watermark via Entropy-based Constraint for Open-source Dataset Copyright Protection, *Ming Sun, Rui Wang, Zixuan Zhu, Lihua Jing, Yuanfang Guo*
- 436 Towards Million-Scale Adversarial Robustness Evaluation With Stronger Individual Attacks, *Yong Xie, Weijie Zheng, Hanxun Huang, Guangnan Ye, Xingjun Ma*
- 437 Balancing Two Classifiers via A Simplex ETF Structure for Model Calibration, *Jiani Ni, He Zhao, Jintong Gao, Dandan Guo, Hongyuan Zha*
- 438 Incomplete Multi-View Multi-label Learning via Disentangled Representation and Label Semantic Embedding, *Xu Yan, Jun Yin, Jie Wen*
- 439 ROLL: Robust Noisy Pseudo-label Learning for Multi-View
* Clustering with Noisy Correspondence, *Yuan Sun, Yongxiang Li, Zhenwen Ren, Guiduo Duan, Dezhong Peng, Peng Hu*
- 440 Feature Selection for Latent Factor Models, *Rittwika Kansabanik, Adrian Barbu*
- 441 Multi-modal Contrastive Learning with Negative Sampling Calibration for Phenotypic Drug Discovery, *Jiahua Rao, Hanjing Lin, Leyu Chen, Jiancong Xie, Shuangjia Zheng, Yuedong Yang*
- 442 Multi-modal Medical Diagnosis via Large-small Model Collaboration, *Wanyi Chen, Zihua Zhao, Jiangchao Yao, Ya Zhang, Jiajun Bu, Haishuai Wang*
- 443 Towards All-in-One Medical Image Re-Identification, *Yuan Tian, Kaiyuan Ji, Rongzhao Zhang, Yankai Jiang, Chunyi Li, Xiaosong Wang, Guangtao Zhai*
- 444 FactCheXcker: Mitigating Measurement Hallucinations in Chest X-ray Report Generation Models, *Alice Heiman, Xiaoman Zhang, Emma Chen, Sung Eun Kim, Pranav Rajpurkar*
- 445 Interactive Medical Image Analysis with Concept-based Similarity Reasoning, *Ta Duc Huy, Sen Kim Tran, Phan Nguyen, Nguyen Hoang Tran, Tran Bao Sam, Anton van den Hengel, Zhibin Liao, Johan W. Verjans, Minh-Son To, Vu Minh Hieu Phan*
- 446 Unsupervised Foundation Model-Agnostic Slide-Level Representation Learning, *Tim Lenz, Peter Neidlinger, Marta Ligerio, Georg Wölfflein, Marko van Treeck, Jakob N. Kather*
- 447 Fast and Accurate Gigapixel Pathological Image Classification with Hierarchical Distillation Multi-Instance Learning, *Jiuyang Dong, Junjun Jiang, Kui Jiang, Jiahua Li, Yongbing Zhang*
- 448 ASIGN: An Anatomy-aware Spatial Imputation Graphic Network for 3D Spatial Transcriptomics, *Junchao Zhu, Ruining Deng, Tianyuan Yao, Juming Xiong, Chongyu Qu, Junlin Guo, Siqi Lu, Mengmeng Yin, Yu Wang, Shilin Zhao, Haichun Yang, Yuankai Huo*
- 449 beta-FFT: Nonlinear Interpolation and Differentiated Training Strategies for Semi-Supervised Medical Image Segmentation, *Ming Hu, Jianfu Yin, Zhuangzhuang Ma, Jianheng Ma, Feiyu Zhu, Bingbing Wu, Ya Wen, Meng Wu, Cong Hu, Bingliang Hu, Quan Wang*
- 450 DyCON: Dynamic Uncertainty-aware Consistency and Contrastive Learning for Semi-supervised Medical Image Segmentation, *Maregu Assefa, Muzammal Naseer, Iyyakutti Iyappan Ganapathi, Syed Sadaf Ali, Mohamed L Seghier, Naoufel Werghi*
- 451 Rethinking Decoder Design: Improving Biomarker Segmentation Using Depth-to-Space Restoration and Residual Linear Attention, *Saad Wazir, Daeyoung Kim*
- 452 LesionLocator: Zero-Shot Universal Tumor Segmentation and
☆ Tracking in 3D Whole-Body Imaging, *Maximilian Rokuss, Yannick Kirchhoff, Seval Akbal, Balint Kovacs, Saikat Roy, Constantin Ulrich, Tassilo Wald, Lukas T. Rotkopf, Heinz-Peter Schlemmer, Klaus Maier-Hein*

- 453 **DAMM-Diffusion: Learning Divergence-Aware Multi-Modal**
 * **Diffusion Model for Nanoparticles Distribution Prediction,**
Junjie Zhou, Shouju Wang, Yuxia Tang, Qi Zhu, Daoqiang Zhang,
Wei Shao

454 **DeformCL: Learning Deformable Centerline Representation for**
Vessel Extraction in 3D Medical Image, *Ziwei Zhao,*
Zhixing Zhang, Yuhang Liu, Zhao Zhang, Haojun Yu, Dong Wang,
Liwei Wang

455 **MultiMorph: On-demand Atlas Construction,** *S. Mazdak Abulnaga,*
Andrew Hoopes, Neel Dey, Malte Hoffmann, Bruce Fischl,
John Guttag, Adrian Dalca

456 **Anatomical Consistency and Adaptive Prior-informed**
Transformation for Multi-contrast MR Image Synthesis via
Diffusion Model, *Yejee Shin, Yeeun Lee, Hanbyol Jang,*
Geonhui Son, Hyeongyu Kim, Dosik Hwang

457 **CrossSDF: 3D Reconstruction of Thin Structures From Cross-**
Sections, *Thomas Walker, Salvatore Esposito, Daniel Rebain,*
Amir Vaxman, Arno Onken, Changjian Li, Oisin Mac Aodha

Notes:

City Map

Entrance Points from Hotels



① Level 1

Rep. John Lewis Way South and Korean Veterans Blvd.

AC Hotel by Marriott
Drury Plaza Hotel
Hampton Inn & Suites
Holiday Inn & Suites
Hyatt Place
The Joseph Hotel
Margaritaville Hotel
Omni Nashville
Residence Inn by Marriott

② Level 1

Rep. John Lewis Way South and Demonbreun

Four Seasons Hotel
Hilton Nashville
Downtown
Renaissance Hotel

③ Level 2

6th Ave and Demonbreun

Hotel shuttle drop-offs
Parking Garage

④ Level 3

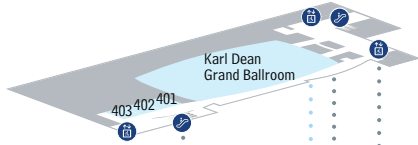
8th Ave and Demonbreun

1 Hotel Nashville
Cambria Hotel
Embassy Suites by Hilton
JW Marriott
The Westin
Renaissance Hotel

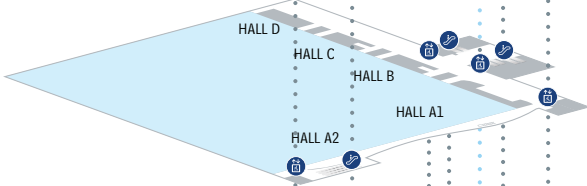
Music City Center.

201 Rep. John Lewis Way South
Nashville, TN 37203
615.401.1400

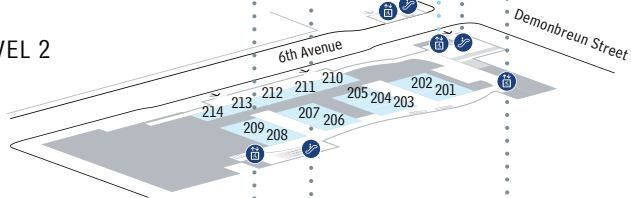
LEVEL 4



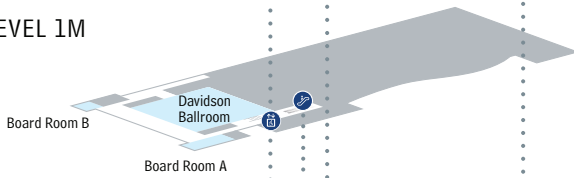
LEVEL 3



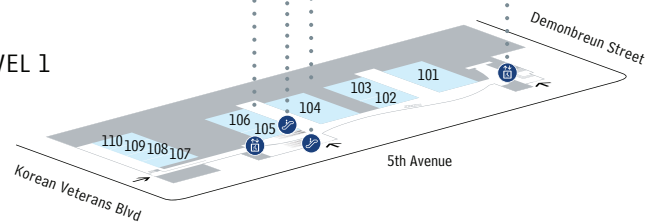
LEVEL 2



LEVEL 1M

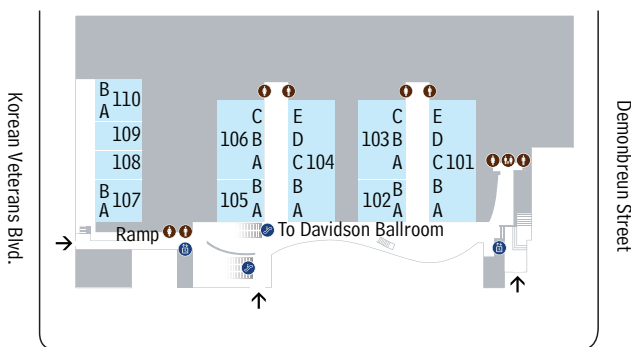


LEVEL 1

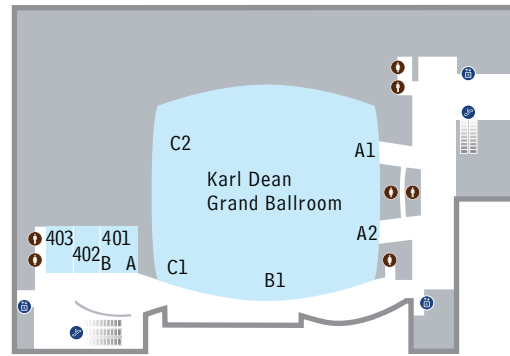


- Men's Restroom
- Family Restroom
- Escalator
- Women's Restroom
- Elevator

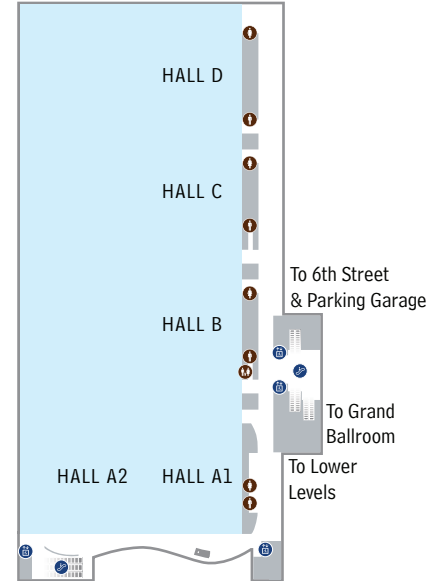
LEVEL 1



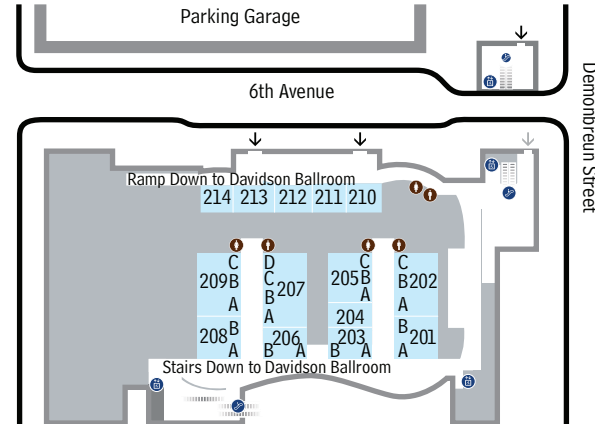
LEVEL 4



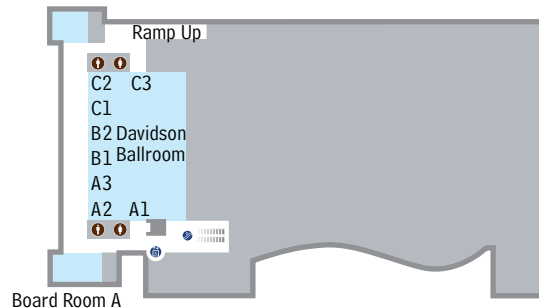
LEVEL 3



LEVEL 2



LEVEL 1M
Board Room B





CVPR *Nashville*  JUNE 11-15, 2025