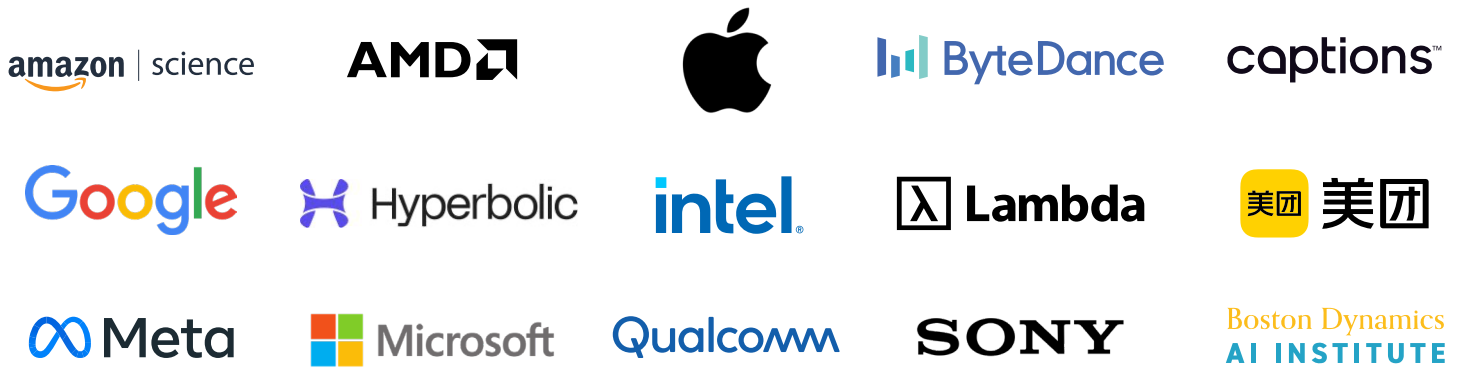


CVPR 2024

**IEEE/CVF Conference on
Computer Vision and
Pattern Recognition**



PLATINUM SPONSORS



GOLD SPONSORS



SILVER SPONSORS



Welcome to the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition in Seattle, Washington! 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, panels, tutorials, workshops, demos, exhibitions, and social events, all aimed at providing attendees with an exciting and enriching experience. CVPR 2024 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 2024 received 11,532 valid paper submissions, a 26% increase from CVPR 2023. The review process was managed by the 6 Program Co-Chairs, 24 Senior Area Chairs, and 477 Area Chairs. During the review phase, each paper received at least 3 reviews from a pool of 9,872 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 discussion 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, 2,719 papers were accepted, for a 23.6% 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 2,719 accepted papers were invited to present posters at CVPR. In addition, 90 (3.3%) papers were selected to be presented as oral talks, based on nominations from Area Chairs, and 324 (11.9%) were selected by ACs to be "highlights" because of their high quality and potential impact. The highlights are flagged with a special annotation in the program. ACs nominated 24 papers to be best paper award candidates, from which a committee convened by the PCs selected the award winners. The award winners will be announced during the conference.

CVPR 2024 brings back the tradition of oral presentations in a three-track configuration. Nevertheless, we kept many of the innovations of CVPR 2023, including single-track panel discussions, "highlights" to

indicate top-rated papers, the use of OpenReview for paper submission and management, and the role of Senior Area Chair to help oversee the review process.

We would like to thank everyone involved in making CVPR 2024 a success. This includes the organizing committee, the Senior Area Chairs, Area Chairs, and reviewers, the authors, the demo session participants, and the donors and exhibitors. David Forsyth's service as Senior Advisor to the Program Chairs was incredibly helpful. We also thank Nicole Finn and her C to C Events team for organizing the conference logistics, 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 inaugural AI Art Coordinator. Last but not least, we thank all of you for attending CVPR 2024 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 Seattle during the conference.

Enjoy CVPR 2024. We look forward to meeting you in person!

Program Chairs

Ali Farhadi, (*University of Washington*)
David Crandall, (*Indiana University*)
Imari Sato, (*National Institute of Informatics*)
Jianxin Wu, (*Nanjing University*)
Robert Pless, (*George Washington University*)
Zeynep Akata, (*University of Tübingen*)

General Chairs

Octavia Camps, (*Northeastern University*)
Ramin Zabih, (*Cornell University*)
Rita Cucchiara, (*University of Modena and Reggio Emilia*)
Sudeep Sarkar, (*University of South Florida*)
Walter Scheirer, (*University of Notre Dame*)

CVPR 2024 QR CODES

CVPR 2024 Virtual Platform



Access schedules, papers, workshops, tutorials, etc.

CVPR 2024 Slido Site



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

General Chairs

Octavia Camps (Northeastern University)
 Ramin Zabih (Cornell)
 Rita Cucchiara (University of Modena and Reggio Emilia)
 Sudeep Sarkar (University of South Florida)
 Walter Scheirer (University of Notre Dame)

Program Chairs

Ali Farhadi (University of Washington)
 David Crandall (Indiana University)
 Imari Sato (National Institute of Informatics)
 Jianxin Wu (Nanjing University)
 Robert Pless (George Washington University)
 Zeynep Akata (University of Tübingen)

Senior Advisor to the Program Chairs

David Forsyth (University of Illinois at Urbana-Champaign)

Technical Chair

Yoshitomo Matsubara (Spiffy AI)

Tutorial Chairs

Katarina Doctor (U.S. Naval Research Lab)
 Vitomir Struc (University of Ljubljana)

Workshop Chairs

Abhinav Shrivastava (University of Maryland)
 Andrew Owens (University of Michigan)
 Antitza Dantcheva (Inria)
 Luisa Verdoliva (University Federico II of Naples)

DEI Chairs

Adriana Kovashka (Pitt)
 CJ Taylor (Amazon AWS)
 Michael King (Fit)
 Roni Sengupta (UNC)
 Sara Beery (MIT)
 Shuran Song (Columbia University)
 Tamara L Berg

Conference Ombud

Angjoo Kanazawa (University of California Berkeley)
 Derek Hoiem (University of Illinois at Urbana-Champaign)

Demonstration Chairs

Sathyarayanan N. Aakur (Auburn University)
 Shu Kong (University of Macau, Texas A&M University)

Senior PAMI-TC Ombud

David Forsyth (University of Illinois at Urbana-Champaign)
 Linda Shapiro (UW Reality Lab University of Washington)

AI Art Curator

Luba Elliott (Independent Curator)

Conference Producer

Nicole Finn (c to c events)

Local Chairs

Ira Kemelmacher-Shlizerman (University of Washington)
 Ranjay Krishna (University of Washington)

Publications Chair

Eric Mortensen

Publicity Chairs

Abby Stylianou (Saint Louis University)
 Boqing Gong (Google)
 Jia-Bin Huang (University of Maryland, College Park)
 Kosta Derpanis (York University/Samsung)
 Shenghua Gao (ShanghaiTech University)

Finance Chair

Gerard Medioni (Amazon)

Accessibility Chair

Danna Gurari (University of Colorado, Boulder)

Doctoral Consortium Chairs

Aparna Bharati (Lehigh University)
 Nathan Jacobs (Washington University in St. Louis)

Website Chair

Mauricio Pamplona Segundo (University of South Florida)

Workflow Chair

Zhenyu (Sherry) Xue (CVPR)

Web Developer

Lee Campbell (Eventhosts)

Virtual Platform Chair

Andreas Geiger (University of Tübingen)

Corporate Relations Chairs

Brian Clipp (Kitware)
 Victor Fragoso (Microsoft)

Social Activities Chairs

Giovanni Maria Farinella (University of Catania, Italy)
 Vítor Albiero (Meta)
 Yale Song (Meta)

CVPR 2024 SENIOR AREA CHAIRS

Alex Schwing
 Alexei A Efros
 Aniruddha Kembhavi
 Anthony Hoogs
 Aude Oliva

Bernard Ghanem
 Bharath Hariharan
 Bohyung Han
 Fernando De la Torre
 Jingdong Wang

Jingyi Yu
 Juan Carlos Niebles
 Kyoung Mu Lee
 Nathan Jacobs
 Pascal Fua

Philipp Kraehenbuehl
 Phillip Isola
 Richard Souvenir
 Roozbeh Mottaghi
 Ruigang Yang

Sing Bing Kang
 Srinivasa Narasimhan
 Tal Hassner
 Yoichi Sato

Aayush Bansal	Cornelia Fermuller	Hung-Yu Tseng	Lu Sheng	Raffay Hamid
Abby Stylianou	Cristian Canton Ferrer	Hyun Soo Park	Lu Yuan	Rahaf Aljundi
Abhinav Shrivastava	Cuiling Lan	Hyunjung Shim	Luca Weihs	Ran He
Abhishek Gupta	Cuong V. Nguyen	Hyunwoo J. Kim	M. Salman Asif	Rana Hanocka
Adam Czakajka	Dahun Kim	Iacopo Masi	Mahdi S. Hosseini	Ranjay Krishna
Adam Kortylewski	Dan Xu	Ioannis Gkioulekas	Mang Ye	Ravi Ramamoorthi
Adriana Kovashka	Danna Gurari	Ishan Misra	Manohar Paluri	Raymond A. Yeh
Aishwarya Agrawal	David B. Lindell	Ishani Chakraborty	Manolis Savva	Rei Kawakami
Akihiro Sugimoto	David Fouhey	Jaesik Park	Marcello Pelillo	Richard Zhang
Alejandro F. Frangi	David Novotny	James Hays	Marco Lorenzi	Robby T. Tan
Alexander C. Berg	Deng-Ping Fan	Jason J Corso	Marcus Rohrbach	Rohit Girdhar
Alexander T Toshev	Devis Tuia	Jasper Uijlings	Margret Keuper	Roman Pflugfelder
Alexandre Alahi	Di He	Jean-Francois Lalonde	Maria Vakalopoulou	Ryoma Bise
Alimoor Reza	Di Huang	Jeany Son	Mark Yatskar	Sai-Kit Yeung
Alireza Fathi	Diane Larlus	Jiahuan Zhou	Massimiliano Mancini	Salman Khan
Aljosa Osep	Dim Papadopoulos	Jiajun Wu	Matej Kristan	Sanjeev Jagannatha Koppal
Ameesh Makadia	Dima Damen	Jian Wang	Matt Feiszli	Sara Beery
Amir Zamir	Dimosthenis Karatzas	Jianfei Cai	Matthew Brown	Sarah Adel Bargal
Andre Araujo	Dinesh Jayaraman	Jiankang Deng	Matthew O'Toole	Sareh Rowlands
Andrea Tagliasacchi	Dingwen Zhang	Jianwen Xie	Matthieu Cord	Sathyanarayanan N. Aakur
Andrew Owens	Dong Chen	Jimei Yang	Mehrdad Farajtabar	Satoshi Tsutsui
Angel X Chang	Dong Xu	Jingya Wang	Miaomiao Liu	Saurabh Gupta
Angela Yao	Dongfang Liu	Jinwei Gu	Michael Maire	Saurabh Singh
Anh Tuan Tran	Eddy Ilg	Joao P Barreto	Michael Wray	Scott McCloskey
Animesh Garg	Elisa Ricci	Jonathan T. Barron	Mike Zheng Shou	Scott Workman
Anna Rohrbach	Emily Morgan Hand	Jong Chul Ye	Min H. Kim	Seong Joon Oh
Anpei Chen	Enrique Dunn	Jonghyun Choi	Min Sun	Ser-Nam Lim
Anurag Arnab	Eshed Ohn-Bar	Jordi Pont-Tuset	Mingkui Tan	Sergey Tulyakov
Arsha Nagrani	Evan Shelhamer	Joseph Tighe	Mingsheng Long	Seunghoon Hong
Arun Mallya	Evangelos Kalogerakis	Judy Hoffman	Mingze Xu	Seungryong Kim
Asako Kanezaki	Fahad Khan	Jufeng Yang	Minh Hoai	Seungyong Lee
Ashok Veeraraghavan	Fartash Faghri	Jun Liu	Minjie Cai	Shalini De Mello
Aswin C.	Fatma Guney	Jun-Yan Zhu	Minsu Cho	Shang-Hong Lai
Sankaranarayanan	Federico Tombari	Jungseock Joo	Mohamed Elhoseiny	Shanshan Zhang
Baoyuan Wu	Feng Lu	Junseok Kwon	Mohit Gupta	Shaohui Lin
Bastian Leibe	Francesc Moreno-Noguer	Junsong Yuan	Mohsen Ali	Sharon X Huang
Ben Mildenhall	Gang Yu	Kai Han	Muhammad Haris Khan	Shayok Chakraborty
Bin Fan	Gedas Bertasius	Kaiyang Zhou	Negar Rostamzadeh	Shenghua Gao
Bing Su	Georgia Gkioxari	Kannappan Palaniappan	Neill D. F. Campbell	Shenlong Wang
Binh-Son Hua	Georgios Pavlakos	Karteek Aahari	Niany Li	Shijian Lu
Björn Ommer	Georgios Tzimiropoulos	Katerina Fragkiadaki	Niloy Mitra	Shin'ichi Satoh
Bo Chen	Gim Hee Lee	Keiji Yanai	Ning Yu	Shizhe Chen
Bolei Zhou	Giovanni Maria Farinella	Ken Sakurada	Noah Snavely	Shohei Nobuhara
Boqing Gong	Gordon Wetzstein	Kevin J Liang	Norimichi Ukita	Shubham Tulsiani
Boxin Shi	Guanbin Li	Khoa Luu	Nuno Vasconcelos	Shuicheng YAN
Boyi Li	Guansong Pang	Kiana Ehsani	Oana-Maria Camburu	Siavash Arjomand Bigdeli
Brendan Tran Morris	Gül Varol	Kihyuk Sohn	Oisin Mac Aodha	Siyu Tang
Bruce Allen Maxwell	Guosheng Lin	Kris M. Kitani	Oliver Wang	Song Bai
Bryan A. Plummer	Guoyu Lu	Kristin Dana	Oncel Tuzel	Srinath Sridhar
Bryan Morse	Hadi Pouransari	Kwang In Kim	Pablo Arbelaez	Stefan Lee
Bumsub Ham	Hae-Gon Jeon	Kwang Moo Yi	Paolo Favaro	Stefano Mattoccia
Carl Doersch	Hajime Nagahara	Lam M. Nguyen	Pascal Mettes	Stella X. Yu
Carl Vondrick	Hamid Rezaatofighi	Lamberto Ballan	Paul Hongsuck Seo	Stephan Alaniz
Chang Xu	Han-Jia Ye	Laura Leal-Taixé	Pedro Morgado	Stephan Richter
Chao Ma	Hanbyul Joo	Laurent Kneip	PENG WANG	Stephen James
Chao-Yuan Wu	Hao Chen	Le Lu	Peter Hedman	Stratis Gavves
Chaowei Xiao	Hao Dong	Lei Wang	Peter Vincent Gehler	Subhashini Venugopalan
Charless Fowlkes	Henghui Ding	Leif Kobbelt	Peter Wonka	Suha Kwak
Chen Change Loy	Hengshuang Zhao	Li Erran Li	Peyman Milanfar	Tae Hyun Kim
Chen Sun	Hideki Nakayama	Li Fuxin	Philippos Mordohai	Tae-Hyun Oh
Chen Wang	Hilde Kuehne	Li Yi	Ping Hu	Tae-Kyun Kim
Chengjiang Long	Hirokatsu Kataoka	Li Zhang	Ping Luo	Taesung Park
Chenliang Xu	Hiroshi Kawasaki	Liang Lin	Piotr Bojanowski	Taesup Moon
Chenyou Fan	Hong Chang	Liang Zheng	Piotr Koniusz	Takayuki Okatani
Chetan Arora	Hong Xuan	Liang-Chieh Chen	Pratul P. Srinivasan	Tanmay Gupta
Chong Luo	Hongyang Li	Limin Wang	Praveen Krishnan	Tat-Jen Cham
Christian Rupprecht	Hossein Rahmani	Lingjie Liu	Qi Dai	Tat-Jun Chin
Christopher Clark	Hsin-Ying Lee	Lingxi Xie	Qi Wu	Tatsuya Harada
Christopher Funk	Hu Han	Linjie Yang	Qian Yu	Thibaut Durand
Chuang Gan	Huaizu Jiang	Liwei Wang	Qifeng Chen	Thomas Kipf
Chuhang Zou	Hui Ji	Long Chen	Qilong Wang	Tian Han
Chunyu Wang	Huijuan Xu	Lourdes Agapito	Qixing Huang	Tianfan Xue
Cihang Xie	Huiwen Chang	Lu Jiang	Rafal Mantiuk	Ting Yao

Ting Zhang	Wei Yang	Xinlei Chen	Yifei Huang	Yumin Suh
Tinne Tuytelaars	Wei-Chih Hung	Xiu-Shen Wei	Yijun Li	Yunchao Wei
Todd Zickler	Wei-Chiu Ma	Xueting Li	Yin Li	Yung-Yu Chuang
Tolga Birdal	Weihong Deng	Xuming He	Ying Fu	Yunzhu Li
Tony Tung	Weisong Shi	Xun Cao	Ying Wu	Yusuke Sugano
Toshihiko Yamasaki	Wen Li	Yale Song	Yinghuan Shi	Yuyin Zhou
Tsung-Yi Lin	Wenguan Wang	Yan Huang	Yingli Tian	Zaid Harchaoui
Tushar Nagarajan	Wenqi Ren	Yan Zhang	Yinqiang Zheng	Zhangyang Wang
Ulugbek S. Kamilov	William Robson Schwartz	Yanchao Yang	Yisen Wang	Zhaowei Cai
Unnat Jain	Xi Peng	Yang Song	Yiyi Liao	Zhaoxiang Zhang
Varun Jampani	Xiangyu Xu	Yang Wang	Yonatan Bisk	Zhe Gan
Venkatesh Babu	Xiangyu Yue	Yang You	Yong Jae Lee	Zheng Zhang
Radhakrishnan	Xiaoguang Han	Yannis Kalantidis	Yosi Keller	Zhengming Ding
Viktoria Sharmanska	Xiaojuan Qi	Yaoyao Liu	Yu Cheng	Zhengqi Li
Vincent Lepetit	Xiaoming Liu	Yasushi Makihara	Yu Kong	Zhirong Wu
Vincent Sitzmann	Xiaowei Zhou	Yebin Liu	Yu Li	Zhizhong Han
Vishal M. Patel	Xiaoyu Wang	Yedid Hoshen	Yu Wu	Zhun Zhong
Vitor Albiero	Xihui Liu	Yen-Yu Lin	Yu-Chuan Su	Ziwei Liu
Vittorio Murino	Xin Wang	Yezhou Yang	Yu-Xiong Wang	Zongwei Zhou
Vlad I Morariu	Xin Yang	Yi Wu	Yuandong Tian	Zsolt Kira
Wangmeng Zuo	Xin Yu	Yi-Hsuan Tsai	Yuchao Dai	Zuxuan Wu
Waqas Sultani	Xingang Pan	Yi-Ting Chen	Yuki M Asano	Zuzana Kukelova
Wei Liu	Xinggang Wang	Yibing Song	Yulun Zhang	

CVPR 2024 OUTSTANDING REVIEWERS

We are grateful to all of the 9,872 reviewers who helped make CVPR 2024 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 2% of reviewers.

Aadarsh Sahoo	Davide Cozzolino	Jan Eric Lenssen	Ming Jiang	Torsten Sattler
Aaron Walsman	David Fan	Jangwon Lee	Mohamed E. Hussein	Touqeer Ahmad
Abhinav Shukla	Da-Wei Zhou	Jaskirat Singh	Mohammad Reza	Valerio Cambareri
Adnan Qayyum	Despoina Paschalidou	Jens Behley	Hosseinzadeh Taher	Vishaal Udandarao
Adrian Lopez	Di Hu	Jiangpeng He	Monica Hernandez	Volker Rodehorst
Adrian Penate-Sanchez	Dominique Ginjac	Jianyi Wang	Mo Zhou	Williem
Ahmed Taha	Donald G. Dansereau	Jieji Ren	Nan Pu	Wojciech Zielonka
Alessio Xompero	Dongbo Min	Jihyong Oh	Nicholas Kolkin	Wonwoong Cho
Alexander Hermans	Dong-Jun Han	Jingkang Yang	Paola Cascante-Bonilla	Xiang An
Alexandros Haliassos	Effrosyni Mavroudi	Jingyuan Yang	Patrick Ebel	Xiang Li
Alvaro Budria	Elisavet Konstantina	Joe Mathai	Paul-Edouard Sarlin	Xiangyu He
Amine Bourki	Stathopoulou	Joerg Stueckler	Philipp Lindenberger	Xiaohan Zhang
Ana C Murillo	Evangelos Kazakos	Jonghee Kim	Qiang Xu	Xingyi Zhou
Anastasia Antsiferova	Fabio Tosi	Juan Camilo Perez	Qian Zheng	Xiongwei Wu
Andrea Conti	Faisal Z Qureshi	Julian Straub	Qize Yang	Yanbin Liu
Andrew Brown	Forrester Cole	Junhwa Hur	Renato Martins	Yankai Jiang
Anh-Quan Cao	Francis Engelmann	Jun Xiao	Reuben Tan	Yannick Hold-Geoffroy
Anil Usumezbas	Francois Rameau	Junyao Hu	Riccardo Volpi	Yi Dong
Anna Kukleva	Fuyong Xing	Kevin Qinghong Lin	Richard Newcombe	Yihua Zhang
Arjun Karpur	Ganesh Subramanian Iyer	Khang Truong Giang	Ruicong Liu	Yinan Zhao
Armin Hadzic	Gilles Puy	Kiru Park	Ruihan Gao	Yongzhen Huang
Ashish Ramayee Asokan	Ginger Delmas	Konda Reddy Mopuri	Saihui Hou	Yuanhao Cai
Assia Benbihi	Goutam Bhat	Kongming Liang	Samaneh Azadi	Yuesong Wang
Ayush Jain	Guoli Jia	Konstantinos Batsos	Satwik Kottur	Yu-Lun Liu
Ayush Saraf	Gurumurthy Swaminathan	Kumar Ashutosh	Satya Krishna Gorti	Yunqi Miao
Benjamin Busam	Hakan Bilen	Lin Geng Foo	Shuhong Zheng	Yunze Man
Bicheng Xu	Hanchen Xie	Lin Zhu	Shu Zhang	Yurui Zhu
Bingliang Jiao	Hannah Kerner	Luisa Verdoliva	Shyamal Buch	Yusuke Matsui
Bumsoo Kim	Haomin Liu	M. Saquib Sarfraz	Siamul Karim Khan	Yu-Wei Chao
Carlo Masone	Hao Zhu	M. Usman Rafique	Sima Behpour	Yuyu Guo
Carlos Rodriguez-Pardo	Hawook Jeong	Mahmoud Afifi	Simon Stepputtis	Zan Gojcic
Changqing Zhang	Hazel Doughty	Mannat Singh	Siyang Dong	Zerong Zheng
Chenglin Yang	Hermann Blum	Martin R. Oswald	Sourav Garg	Zheheng Jiang
Chengze Miaomiao Li	Hongwei Bran Li	Matteo Poggi	Spencer Whitehead	Zhekun Luo
Cheston Tan	Huan Wang	Matthew Joseph Leotta	Stefan Leutenegger	Zheng Chen
Chiao An Yang	Hui Liang	Mattia Savardi	Stefano Berretti	Zhibin Liao
Chi Xu	Inhwan Bae	Mayu Otani	Sunghoon Im	Zhicheng Zhang
Chunghyun Park	Ioannis Papoutsis	Michaël Gharbi	Swathikiran Sudhakaran	Zhihe Lu
Cristian Rodriguez-Opazo	Iro Laina	Michael S. Brown	Terrance E. Boult	Zhou Yu
Dahyun Kang	Ishan Rajendrakumar Dave	Michael Weinmann	Thomas Hummel	
Daqi Liu	Jaeyoo Park	Michel Antunes	Toby Perrett	

Wednesday, June 19

7:00 - 17:00 Registration / Badge Pickup (Summit Lobby)

7:00 - 17:00 Press Room (Summit 340)

7:00 - 17:00 Mother's Room (Summit 341-adjacent and Summit 441-adjacent)

7:00 - 17:00 Prayer or Quiet Room (Summit 326)

7:00 - 9:00 Breakfast (Summit ExHall 1-2)

8:30 - 9:00 Welcome & Awards (Summit Flex Hall)

9:00 - 10:30 Oral Session 1A: Low-Level Vision

(Summit Ballroom)

- 1 Specularity Factorization for Low-Light Enhancement, *Saurabh Saini, P J Narayanan*
- 2 FlowIE: Efficient Image Enhancement via Rectified Flow, *Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, Jiwen Lu*
- 3 Towards Robust Event-guided Low-Light Image Enhancement: A Large-Scale Real-World Event-Image Dataset and Novel Approach, *Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, Lin Wang*
- 4 Bilateral Event Mining and Complementary for Event Stream Super-Resolution, *Zhilin Huang, Quanmin Liang, Yijie Yu, Chujun Qin, Xiawu Zheng, Kai Huang, Zikun Zhou, Wenming Yang*
- 5 FMA-Net: Flow-Guided Dynamic Filtering and Iterative Feature Refinement with Multi-Attention for Joint Video Super-Resolution and Deblurring, *Geunhyuk Youk, Jihyong Oh, Munchul Kim*

9:00 - 10:30 Oral Session 1B: Vision and Graphics

(Summit Flex Hall AB)

- 1 GPLD3D: Latent Diffusion of 3D Shape Generative Models by Enforcing Geometric and Physical Priors, *Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, Qixing Huang*
- 2 Retrieval-Augmented Layout Transformer for Content-Aware Layout Generation, *Daichi Horita, Naoto Inoue, Kotaro Kikuchi, Kota Yamaguchi, Kiyoharu Aizawa*
- 3 Eclipse: Disambiguating Illumination and Materials using Unintended Shadows, *Dor Verbin, Ben Mildenhall, Peter Hedman, Jonathan T. Barron, Todd Zickler, Pratul P. Srinivasan*
- 4 Objects as Volumes: A Stochastic Geometry View of Opaque Solids, *Bailey Miller, Hanyu Chen, Alice Lai, Ioannis Gkioulekas*
- 5 DiffusionLight: Light Probes for Free by Painting a Chrome Ball, *Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit Raj, Pramook Khungurn, Supasorn Suwajanakorn*

9:00 - 10:30 Oral Session 1C: Humans: Face, Body, Pose, Gesture, Movement

(Summit Flex Hall C)

- 1 MultiPly: Reconstruction of Multiple People from Monocular Video in the Wild, *Zeren Jiang, Chen Guo, Manuel Kaufmann, Tianjian Jiang, Julien Valentin, Otmar Hilliges, Jie Song*
- 2 URHand: Universal Relightable Hands, *Zhaoxi Chen, Gyeongsik Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, He Wen, Lucas Evans, Bo Peng, Julia Buffalini, Autumn Trimble, Kevyn McPhail, Melissa Schoeller, Shoou-I Yu, Javier Romero, Michael Zollhofer, Yaser Sheikh, Ziwei Liu, Shunsuke Saito*
- 3 Relightable Gaussian Codec Avatars, *Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam*
- 4 Semantic Human Mesh Reconstruction with Textures, *Xiaoyu Zhan, Jianxin Yang, Yuanqi Li, Jie Guo, Yanwen Guo, Wenping Wang*
- 5 Stratified Avatar Generation from Sparse Observations, *Han Feng, Wenchao Ma, Quankai Gao, Xianwei Zheng, Nan Xue, Huijuan Xu*

10:00 - 10:30 Poster Setup (Arch 4E)

10:30 - 12:00 Poster Session 1 & Exhibit Hall (Arch 4A-E)

* - Highlight paper (check it out)

🏆 - Award candidate paper (see award sessions)

- 1 SEAS: ShapE-Aligned Supervision for Person Re-Identification, *Haidong Zhu, Pranav Budhwant, Zhaoheng Zheng, Ram Nevatia*
- 2 Test-Time Domain Generalization for Face Anti-Spoofing, *Qianyu Zhou, Ke-Yue Zhang, Taiping Yao, Xuequan Lu, Shouhong Ding, Lizhuang Ma*
- 3 Gradient Alignment for Cross-Domain Face Anti-Spoofing, *Binh M. Le, Simon S. Woo*
- 4 BigGait: Learning Gait Representation You Want by Large Vision Models, *Dingqiang Ye, Chao Fan, Jingzhe Ma, Xiaoming Liu, Shiqi Yu*
- 5 Suppress and Rebalance: Towards Generalized Multi-Modal Face Anti-Spoofing, *Xun Lin, Shuai Wang, Rizhao Cai, Yizhong Liu, Ying Fu, Wenzhong Tang, Zitong Yu, Alex Kot*
- 6 CFPL-FAS: Class Free Prompt Learning for Generalizable Face Anti-spoofing, *Ajian Liu, Shuai Xue, Jianwen Gan, Jun Wan, Yanyan Liang, Jiankang Deng, Sergio Escalera, Zhen Lei*
- 7 Psychometry: An Omnifit Model for Image Reconstruction from Human Brain Activity, *Ruijie Quan, Wenguan Wang, Zhibo Tian, Fan Ma, Yi Yang*
- 8 KeyPoint Relative Position Encoding for Face Recognition, *Minchul Kim, Yiyang Su, Feng Liu, Anil Jain, Xiaoming Liu*
- 9 Distilling CLIP with Dual Guidance for Learning Discriminative Human Body Shape Representation, *Feng Liu, Minchul Kim, Zhiyuan Ren, Xiaoming Liu*
- 10 Flexible Biometrics Recognition: Bridging the Multimodality Gap through Attention Alignment and Prompt Tuning, *Leslie Ching Ow Tiong, Dick Sigmund, Chen-Hui Chan, Andrew Beng Jin Teoh*
- 11 One-Class Face Anti-spoofing via Spoof Cue Map-Guided Feature Learning, *Pei-Kai Huang, Cheng-Hsuan Chiang, Tzu-Hsien Chen, Jun-Xiong Chong, Tyng-Luh Liu, Chiou-Ting Hsu*
- 12 Activity-Biometrics: Person Identification from Daily Activities, *Shehreen Azad, Yogesh Singh Rawat*
- 13 Privacy-Preserving Face Recognition Using Trainable Feature Subtraction, *Yuxi Mi, Zhizhou Zhong, Yuge Huang, Jiazhen Ji, Jianqing Xu, Jun Wang, Shaoming Wang, Shouhong Ding, Shuigeng Zhou*
- 14 Molecular Data Programming: Towards Molecule Pseudo-labeling with Systematic Weak Supervision, *Xin Juan, Kaixiong Zhou, Ninghao Liu, Tianlong Chen, Xin Wang*
- 15 Clustering for Protein Representation Learning, *Ruijie Quan, Wenguan Wang, Fan Ma, Hehe Fan, Yi Yang*
- 16 Fun with Flags: Robust Principal Directions via Flag Manifolds, *Nathan Mankovich, Gustau Camps-Valls, Tolga Birdal*
- 17 CAM Back Again: Large Kernel CNNs from a Weakly Supervised Object Localization Perspective, *Shunsuke Yasuki, Masato Taki*
- 18 Confronting Ambiguity in 6D Object Pose Estimation via Score-Based Diffusion on SE(3), *Tsu-Ching Hsiao, Hao-Wei Chen, Hsuan-Kung Yang, Chun-Yi Lee*
- 19 Quantifying Task Priority for Multi-Task Optimization, *Woosong Jeong, Kuk-Jin Yoon*
- 20 Unbiased Estimator for Distorted Conics in Camera Calibration, *Chaehyeon Song, Jaeho Shin, Myung-Hwan Jeon, Jongwoo Lim, Ayoung Kim*
- 21 Multi-Object Tracking in the Dark, *Xinzhe Wang, Kang Ma, Qiankun Liu, Yunhao Zou, Ying Fu*
- 22 Implicit Discriminative Knowledge Learning for Visible-Infrared Person Re-Identification, *Kaijie Ren, Lei Zhang*
- 23 From Correspondences to Pose: Non-minimal Certifiably Optimal Relative Pose without Disambiguation, *Javier Tirado-Garín, Javier Civera*
- 24 From Activation to Initialization: Scaling Insights for Optimizing Neural Fields, *Hemanth Saratchandran, Sameera Ramasinghe, Simon Lucey*
- 25 PairDETR: Joint Detection and Association of Human Bodies and Faces, *Ammar Ali, Georgii Gaikov, Denis Rybalchenko, Alexander Chigorin, Ivan Laptev, Sergey Zagoruyko*
- 26 Move as You Say Interact as You Can: Language-guided Human Motion Generation with Scene Affordance, *Zan Wang, Yixin Chen, Baoxiong Jia, Puhao Li, Jinlu Zhang, Jingze Zhang, Tengyu Liu, Yixin Zhu, Wei Liang, Siyuan Huang*

- 27 OAKINK2: A Dataset of Bimanual Hands-Object Manipulation in Complex Task Completion, *Xinyu Zhan, Lixin Yang, Yifei Zhao, Kangrui Mao, Hanlin Xu, Zenan Lin, Kailin Li, Cewu Lu*
- 28 Seamless Human Motion Composition with Blended Positional Encodings, *German Barquero, Sergio Escalera, Cristina Palmero*
- 29 VideoRF: Rendering Dynamic Radiance Fields as 2D Feature Video Streams, *Liao Wang, Kaixin Yao, Chengcheng Guo, Zhirui Zhang, Qiang Hu, Jingyi Yu, Lan Xu, Minye Wu*
- 30 OMG: Towards Open-vocabulary Motion Generation via Mixture of Controllers, *Han Liang, Jiacheng Bao, Ruichi Zhang, Sihan Ren, Yuecheng Xu, Sibe Yang, Xin Chen, Jingyi Yu, Lan Xu*
- 31 HOLD: Category-agnostic 3D Reconstruction of Interacting Hands and Objects from Video, *Zicong Fan, Maria Parrelli, Maria Eleni Kadoglou, Xu Chen, Muhammed Kocabas, Michael J. Black, Otmar Hilliges*
- 32 HUGS: Human Gaussian Splats, *Muhammed Kocabas, Jen-Hao Rick Chang, James Gabriel, Oncel Tuzel, Anurag Ranjan*
- 33 HOI-M³: Capture Multiple Humans and Objects Interaction within Contextual Environment, *Juze Zhang, Jingyan Zhang, Zining Song, Zhanhe Shi, Chengfeng Zhao, Ye Shi, Jingyi Yu, Lan Xu, Jingya Wang*
- 34 InterHandGen: Two-Hand Interaction Generation via Cascaded Reverse Diffusion, *Jihyun Lee, Shunsuke Saito, Giljoon Nam, Minhyuk Sung, Tae-Kyun Kim*
- 35 SiTH: Single-view Textured Human Reconstruction with Image-Conditioned Diffusion, *Hsuan-I Ho, Jie Song, Otmar Hilliges*
- 36 4D-DRESS: A 4D Dataset of Real-World Human Clothing With Annotations, *Wenbo Wang, Hsuan-I Ho, Chen Guo, Boxiang Rong, Artur Grigorev, Jie Song, Juan Jose Zarate, Otmar Hilliges*
- 37 MultiPLY: Reconstruction of Multiple People from Monocular Video in the Wild, *Zeren Jiang, Chen Guo, Manuel Kaufmann, Tianjian Jiang, Julien Valentin, Otmar Hilliges, Jie Song*
- 38 FinePOSE: Fine-Grained Prompt-Driven 3D Human Pose Estimation via Diffusion Models, *Jinglin Xu, Yijie Guo, Yuxin Peng*
- 39 Real-Time Simulated Avatar from Head-Mounted Sensors, *Zhengyi Luo, Jinkun Cao, Rawal Khirodkar, Alexander Winkler, Kris Kitani, Weipeng Xu*
- 40 Digital Life Project: Autonomous 3D Characters with Social Intelligence, *Zhongang Cai, Jianping Jiang, Zhongfei Qing, Xinying Guo, Mingyuan Zhang, Zhengyu Lin, Haiyi Mei, Chen Wei, Ruisi Wang, Wanqi Yin, Liang Pan, Xiangyu Fan, Han Du, Peng Gao, Zhitao Yang, Yang Gao, Jiaqi Li, Tianxiang Ren, Yukun Wei, Xiaogang Wang, Chen Change Loy, Lei Yang, Ziwei Liu*
- 41 Learning Visual Prompt for Gait Recognition, *Kang Ma, Ying Fu, Chunshui Cao, Saihui Hou, Yongzhen Huang, Dezhi Zheng*
- 42 Hourglass Tokenizer for Efficient Transformer-Based 3D Human Pose Estimation, *Wenhao Li, Mengyuan Liu, Hong Liu, Pichao Wang, Jialun Cai, Nicu Sebe*
- 43 LocLLM: Exploiting Generalizable Human Keypoint Localization via Large Language Model, *Dongkai Wang, Shiyu Xuan, Shiliang Zhang*
- 44 Spatial-Aware Regression for Keypoint Localization, *Dongkai Wang, Shiliang Zhang*
- 45 GaussianAvatar: Towards Realistic Human Avatar Modeling from a Single Video via Animatable 3D Gaussians, *Liangxiao Hu, Hongwen Zhang, Yuxiang Zhang, Boyao Zhou, Boning Liu, Shengping Zhang, Liqiang Nie*
- 46 HHMR: Holistic Hand Mesh Recovery by Enhancing the Multimodal Controllability of Graph Diffusion Models, *Mengcheng Li, Hongwen Zhang, Yuxiang Zhang, Ruizhi Shao, Tao Yu, Yebin Liu*
- 47 Capturing Closely Interacted Two-Person Motions with Reaction Priors, *Qi Fang, Yinghui Fan, Yanjun Li, Junting Dong, Dingwei Wu, Weidong Zhang, Kang Chen*
- 48 SyncTalk: The Devil is in the Synchronization for Talking Head Synthesis, *Ziqiao Peng, Wentao Hu, Yue Shi, Xiangyu Zhu, Xiaomei Zhang, Hao Zhao, Jun He, Hongyan Liu, Zhaoxin Fan*
- 49 Single-to-Dual-View Adaptation for Egocentric 3D Hand Pose Estimation, *Ruicong Liu, Takehiko Ohkawa, Mingfang Zhang, Yoichi Sato*
- 50 Bidirectional Autoregressive Diffusion Model for Dance Generation, *Canyu Zhang, Youbao Tang, Ning Zhang, Rwei-Sung Lin, Mei Han, Jing Xiao, Song Wang*
- 51 High-Quality Facial Geometry and Appearance Capture at Home, *Yuxuan Han, Junfeng Lyu, Feng Xu*
- 52 Multiple View Geometry Transformers for 3D Human Pose Estimation, *Ziwei Liao, Jialiang Zhu, Chunyu Wang, Han Hu, Steven L. Waslander*
- 53 PACER+: On-Demand Pedestrian Animation Controller in Driving Scenarios, *Jingbo Wang, Zhengyi Luo, Ye Yuan, Yixuan Li, Bo Dai*
- 54 I'M HOI: Inertia-aware Monocular Capture of 3D Human-Object Interactions, *Chengfeng Zhao, Juze Zhang, Jiashen Du, Ziwei Shan, Junye Wang, Jingyi Yu, Jingya Wang, Lan Xu*
- 55 HAVE-FUN: Human Avatar Reconstruction from Few-Shot Unconstrained Images, *Xihe Yang, Xingyu Chen, Daiheng Gao, Shaohui Wang, Xiaoguang Han, Baoyuan Wang*
- 56 Can Language Beat Numerical Regression? Language-Based Multimodal Trajectory Prediction, *Inhwan Bae, Junoh Lee, Hae-Gon Jeon*
- 57 3D Human Pose Perception from Egocentric Stereo Videos, *Hiroyasu Akada, Jian Wang, Vladislav Golyanik, Christian Theobalt*
- 58 Egocentric Whole-Body Motion Capture with FisheyeViT and Diffusion-Based Motion Refinement, *Jian Wang, Zhe Cao, Diogo Luvizon, Lingjie Liu, Kripasindhu Sarkar, Danhang Tang, Thabo Beeler, Christian Theobalt*
- 59 Human Gaussian Splatting: Real-time Rendering of Animatable Avatars, *Arthur Moreau, Jifei Song, Helisa Dharmo, Richard Shaw, Yiren Zhou, Eduardo Pérez-Pellitero*
- 60 OHTA: One-shot Hand Avatar via Data-driven Implicit Priors, *Xiaozheng Zheng, Chao Wen, Zhuo Su, Zeran Xu, Zhaohu Li, Yang Zhao, Zhou Xue*
- 61 HOIAnimator: Generating Text-prompt Human-object Animations using Novel Perceptive Diffusion Models, *Wenfeng Song, Xinyu Zhang, Shuai Li, Yang Gao, Aimin Hao, Xia Hou, Chenglizhao Chen, Ning Li, Hong Qin*
- 62 Arbitrary Motion Style Transfer with Multi-condition Motion Latent Diffusion Model, *Wenfeng Song, Xingliang Jin, Shuai Li, Chenglizhao Chen, Aimin Hao, Xia Hou, Ning Li, Hong Qin*
- 63 Single-View Scene Point Cloud Human Grasp Generation, *Yan-Kang Wang, Chengyi Xing, Yi-Lin Wei, Xiao-Ming Wu, Wei-Shi Zheng*
- 64 Attention-Propagation Network for Egocentric Heatmap to 3D Pose Lifting, *Taeho Kang, Youngki Lee*
- 65 URHand: Universal Relightable Hands, *Zhaoxi Chen, Gyeongsook Moon, Kaiwen Guo, Chen Cao, Stanislav Pidhorskyi, Tomas Simon, Rohan Joshi, Yuan Dong, Yichen Xu, Bernardo Pires, He Wen, Lucas Evans, Bo Peng, Julia Buffalini, Autumn Trimble, Keyvyn McPhail, Melissa Schoeller, Shouo-I Yu, Javier Romero, Michael Zollhofer, Yaser Sheikh, Ziwei Liu, Shunsuke Saito*
- 66 AnySkill: Learning Open-Vocabulary Physical Skill for Interactive Agents, *Jieming Cui, Tengyu Liu, Nian Liu, Yaodong Yang, Yixin Zhu, Siyuan Huang*
- 67 From a Bird's Eye View to See: Joint Camera and Subject Registration without the Camera Calibration, *Zekun Qian, Ruize Han, Wei Feng, Song Wang*
- 68 HMD-Poser: On-Device Real-time Human Motion Tracking from Scalable Sparse Observations, *Peng Dai, Yang Zhang, Tao Liu, Zhen Fan, Tianyuan Du, Zhuo Su, Xiaozheng Zheng, Zeming Li*
- 69 Monocular Identity-Conditioned Facial Reflectance Reconstruction, *Xingyu Ren, Jiankang Deng, Yuhao Cheng, Jia Guo, Chao Ma, Yichao Yan, Wenhan Zhu, Xiaokang Yang*
- 70 GAvatar: Animatable 3D Gaussian Avatars with Implicit Mesh Learning, *Ye Yuan, Xueting Li, Yangyi Huang, Shalini De Mello, Koki Nagano, Jan Kautz, Umar Iqbal*
- 71 Score-Guided Diffusion for 3D Human Recovery, *Anastasis Stathopoulos, Ligong Han, Dimitris Metaxas*
- 72 3D-Aware Face Editing via Warping-Guided Latent Direction Learning, *Yuhao Cheng, Zhuo Chen, Xingyu Ren, Wenhan Zhu, Zhengqin Xu, Di Xu, Changpeng Yang, Yichao Yan*
- 73 WANDR: Intention-guided Human Motion Generation, *Markos Diomatari, Nikos Athanasiou, Omid Taheri, Xi Wang, Otmar Hilliges, Michael J. Black*

- 74 Exploring Vision Transformers for 3D Human Motion-Language Models with Motion Patches, *Qing Yu, Mikihiro Tanaka, Kent Fujiwara*
- 75 NIFTY: Neural Object Interaction Fields for Guided Human Motion Synthesis, *Nilesh Kulkarni, Davis Rempe, Kyle Genova, Abhijit Kundu, Justin Johnson, David Fouhey, Leonidas Guibas*
- 76 DreamAvatar: Text-and-Shape Guided 3D Human Avatar Generation via Diffusion Models, *Yukang Cao, Yan-Pei Cao, Kai Han, Ying Shan, Kwan-Yee K. Wong*
- 77 Person-in-WiFi 3D: End-to-End Multi-Person 3D Pose Estimation with Wi-Fi, *Kangwei Yan, Fei Wang, Bo Qian, Han Ding, Jinsong Han, Xing Wei*
- 78 ScoreHypo: Probabilistic Human Mesh Estimation with Hypothesis Scoring, *Yuan Xu, Xiaoxuan Ma, Jiajun Su, Wentao Zhu, Yu Qiao, Yizhou Wang*
- 79 Relightable and Animatable Neural Avatar from Sparse-View Video, *Zhen Xu, Sida Peng, Chen Geng, Linzhan Mou, Zihan Yan, Jiaming Sun, Hujun Bao, Xiaowei Zhou*
- 80 Relightable Gaussian Codec Avatars, *Shunsuke Saito, Gabriel Schwartz, Tomas Simon, Junxuan Li, Giljoo Nam*
- 81 From Audio to Photoreal Embodiment: Synthesizing Humans in Conversations, *Evonne Ng, Javier Romero, Timur Bagautdinov, Shaojie Bai, Trevor Darrell, Angjoo Kanazawa, Alexander Richard*
- 82 Closely Interactive Human Reconstruction with Proxemics and Physics-Guided Adaption, *Buzhen Huang, Chen Li, Chongyang Xu, Liang Pan, Yangang Wang, Gim Hee Lee*
- 83 Video-Based Human Pose Regression via Decoupled Space-Time Aggregation, *Jijie He, Wenwu Yang*
- 84 Rethinking Generalizable Face Anti-spoofing via Hierarchical Prototype-guided Distribution Refinement in Hyperbolic Space, *Chengyang Hu, Ke-Yue Zhang, Taiping Yao, Shouhong Ding, Lizhuang Ma*
- 85 MoML: Online Meta Adaptation for 3D Human Motion Prediction, *Xiaoning Sun, Huaijiang Sun, Bin Li, Dong Wei, Weiqing Li, Jianfeng Lu*
- 86 KITRO: Refining Human Mesh by 2D Clues and Kinematic-tree Rotation, *Fengyuan Yang, Kerui Gu, Angela Yao*
- 87 Guess The Unseen: Dynamic 3D Scene Reconstruction from Partial 2D Glimpses, *Inhee Lee, Byungjun Kim, Hanbyul Joo*
- 88 PEGASUS: Personalized Generative 3D Avatars with Composable Attributes, *Hyunsoo Cha, Byungjun Kim, Hanbyul Joo*
- 89 Semantic Human Mesh Reconstruction with Textures, *Xiaoyu Zhan, Jianxin Yang, Yuanqi Li, Jie Guo, Yanwen Guo, Wenping Wang*
- 90 SDPose: Tokenized Pose Estimation via Circulation-Guide Self-Distillation, *Sichen Chen, Yingyi Zhang, Siming Huang, Ran Yi, Ke Fan, Ruixin Zhang, Peixian Chen, Jun Wang, Shouhong Ding, Lizhuang Ma*
- 91 Mocap Everyone Everywhere: Lightweight Motion Capture With Smartwatches and a Head-Mounted Camera, *Jiye Lee, Hanbyul Joo*
- 92 DPMesh: Exploiting Diffusion Prior for Occluded Human Mesh Recovery, *Yixuan Zhu, Ao Li, Yansong Tang, Wenliang Zhao, Jie Zhou, Jiwen Lu*
- 93 DPHMs: Diffusion Parametric Head Models for Depth-based Tracking, *Jiapeng Tang, Angela Dai, Yinyu Nie, Lev Markhasin, Justus Thies, Matthias Nießner*
- 94 KTPFormer: Kinematics and Trajectory Prior Knowledge-Enhanced Transformer for 3D Human Pose Estimation, *Jihua Peng, Yanghong Zhou, P. Y. Mok*
- 95 Exploiting Style Latent Flows for Generalizing Deepfake Video Detection, *Jongwook Choi, Taehoon Kim, Yonghyun Jeong, Seungryul Baek, Jongwon Choi*
- 96 EMAGE: Towards Unified Holistic Co-Speech Gesture Generation via Expressive Masked Audio Gesture Modeling, *Haiyang Liu, Zihao Zhu, Giorgio Becherini, Yichen Peng, Mingyang Su, You Zhou, Xuefei Zhe, Naoya Iwamoto, Bo Zheng, Michael J. Black*
- 97 A Unified Framework for Human-centric Point Cloud Video Understanding, *Yiteng Xu, Kecheng Ye, Xiao Han, Yiming Ren, Xinge Zhu, Yuexin Ma*
- 98 ASH: Animatable Gaussian Splats for Efficient and Photoreal Human Rendering, *Haokai Pang, Heming Zhu, Adam Kortylewski, Christian Theobalt, Marc Habermann*
- 99 CLOAF: CoLLisiOn-Aware Human Flow, *Andrey Davydov, Martin Engilberge, Mathieu Salzmann, Pascal Fua*
- 100 EventEgo3D: 3D Human Motion Capture from Egocentric Event Streams, *Christen Millerdurai, Hiroyasu Akada, Jian Wang, Diogo Luvizon, Christian Theobalt, Vladislav Golyanik*
- 101 A Call to Reflect on Evaluation Practices for Age Estimation: Comparative Analysis of the State-of-the-Art and a Unified Benchmark, *Jakub Paplham, Vojtěch Franc*
- 102 Holoported Characters: Real-time Free-viewpoint Rendering of Humans from Sparse RGB Cameras, *Ashwath Shetty, Marc Habermann, Guoxing Sun, Diogo Luvizon, Vladislav Golyanik, Christian Theobalt*
- 103 Synergistic Global-space Camera and Human Reconstruction from Videos, *Yizhou Zhao, Tuanfeng Yang Wang, Bhiksha Raj, Min Xu, Jimei Yang, Chun-Hao Paul Huang*
- 104 3D Face Tracking from 2D Video through Iterative Dense UV to Image Flow, *Felix Taubner, Prashant Raina, Mathieu Tuli, Eu Wern Teh, Chul Lee, Jinmiao Huang*
- 105 UltraAvatar: A Realistic Animatable 3D Avatar Diffusion Model with Authenticity Guided Textures, *Mingyuan Zhou, Rakib Hyder, Ziwei Xuan, Guojun Qi*
- 106 OmniMotionGPT: Animal Motion Generation with Limited Data, *Zhangshihao Yang, Mingyuan Zhou, Mengyi Shan, Bingbing Wen, Ziwei Xuan, Mitch Hill, Junjie Bai, Guo-Jun Qi, Yalin Wang*
- 107 Text-Guided 3D Face Synthesis - From Generation to Editing, *Yunjie Wu, Yapeng Meng, Zhipeng Hu, Lincheng Li, Haoqian Wu, Kun Zhou, Weiwei Xu, Xin Yu*
- 108 Multi-scale Dynamic and Hierarchical Relationship Modeling for Facial Action Units Recognition, *Zihan Wang, Siyang Song, Cheng Luo, Songhe Deng, Weicheng Xie, Linlin Shen*
- 109 LiveHPS: LiDAR-based Scene-level Human Pose and Shape Estimation in Free Environment, *Yiming Ren, Xiao Han, Chengfeng Zhao, Jingya Wang, Lan Xu, Jingyi Yu, Yuexin Ma*
- 110 FaceChain-ImagineID: Freely Crafting High-Fidelity Diverse Talking Faces from Disentangled Audio, *Chao Xu, Yang Liu, Jiazhen Xing, Weida Wang, Mingze Sun, Jun Dan, Tianxin Huang, Siyuan Li, Zhi-Qi Cheng, Ying Tai, Baigui Sun*
- 111 OpticalDR: A Deep Optical Imaging Model for Privacy-Protective Depression Recognition, *Yuchen Pan, Junjun Jiang, Kui Jiang, Zhihao Wu, Keyuan Yu, Xianming Liu*
- 112 SCE-MAE: Selective Correspondence Enhancement with Masked Autoencoder for Self-Supervised Landmark Estimation, *Kejia Yin, Varshanth Rao, Ruowei Jiang, Xudong Liu, Parham Aarabi, David B. Lindell*
- 113 TokenHMR: Advancing Human Mesh Recovery with a Tokenized Pose Representation, *Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Yao Feng, Michael J. Black*
- 114 Optimizing Diffusion Noise Can Serve As Universal Motion Priors, *Korrawe Karunratanakul, Konpat Preechakul, Emre Aksan, Thabo Beeler, Supasorn Suwajanakorn, Siyu Tang*
- 115 M&M VTO: Multi-Garment Virtual Try-On and Editing, *Luyang Zhu, Yingwei Li, Nan Liu, Hao Peng, Dawei Yang, Ira Kemelmacher-Shlizerman*
- 116 AvatarGPT: All-in-One Framework for Motion Understanding Planning Generation and Beyond, *Zixiang Zhou, Yu Wan, Baoyuan Wang*
- 117 A Simple Baseline for Efficient Hand Mesh Reconstruction, *Zhishan Zhou, Shihao Zhou, Zhi Lv, Minqiang Zou, Yao Tang, Jiajun Liang*
- 118 VINECS: Video-based Neural Character Skinning, *Zhouyingcheng Liao, Vladislav Golyanik, Marc Habermann, Christian Theobalt*
- 119 ConvoFusion: Multi-Modal Conversational Diffusion for Co-Speech Gesture Synthesis, *Muhammad Hamza Mughal, Rishabh Dabral, Ikhsanul Habibie, Lucia Donatelli, Marc Habermann, Christian Theobalt*
- 120 Programmable Motion Generation for Open-Set Motion Control Tasks, *Hanchao Liu, Xiaohang Zhan, Shaoli Huang, Tai-Jiang Mu, Ying Shan*
- 121 From Feature to Gaze: A Generalizable Replacement of Linear Layer for Gaze Estimation, *Yiwei Bao, Feng Lu*

- 122 Unsupervised Gaze Representation Learning from Multi-view Face Images, *Yiwei Bao, Feng Lu*
- 123 Joint2Human: High-Quality 3D Human Generation via Compact Spherical Embedding of 3D Joints, *Muxin Zhang, Qiao Feng, Zhuo Su, Chao Wen, Zhou Xue, Kun Li*
- 124 DiffHuman: Probabilistic Photorealistic 3D Reconstruction of Humans, *Akash Sengupta, Thiemo Alldieck, Nikos Kolotouros, Enric Corona, Andrei Zanfir, Cristian Sminchisescu*
- 125 Bi-Causal: Group Activity Recognition via Bidirectional Causality, *Youliang Zhang, Wenxuan Liu, Danni Xu, Zhuo Zhou, Zheng Wang*
- 126 HumanNeRF-SE: A Simple yet Effective Approach to Animate HumanNeRF with Diverse Poses, *Caoyuan Ma, Yu-Lun Liu, Zhixiang Wang, Wu Liu, Xinchun Liu, Zheng Wang*
- 127 LPSNet: End-to-End Human Pose and Shape Estimation with Lensless Imaging, *Haoyang Ge, Qiao Feng, Hailong Jia, Xiongzheng Li, Xiangjun Yin, You Zhou, Jingyu Yang, Kun Li*
- 128 MagicAnimate: Temporally Consistent Human Image Animation using Diffusion Model, *Zhongcong Xu, Jianfeng Zhang, Jun Hao Liew, Hanshu Yan, Jia-Wei Liu, Chenxu Zhang, Jiashi Feng, Mike Zheng Shou*
- 129 RTMO: Towards High-Performance One-Stage Real-Time Multi-Person Pose Estimation, *Peng Lu, Tao Jiang, Yining Li, Xiangtai Li, Kai Chen, Wenming Yang*
- 130 Human Motion Prediction Under Unexpected Perturbation, *Jiangbei Yue, Baiyi Li, Julien Pettré, Armin Seyfried, He Wang*
- 131 Cross-view and Cross-pose Completion for 3D Human Understanding, *Matthieu Armando, Salma Galaaoui, Fabien Baradel, Thomas Lucas, Vincent Leroy, Romain Brégier, Philippe Weinzaepfel, Grégory Rogez*
- 132 Lodge: A Coarse to Fine Diffusion Network for Long Dance Generation Guided by the Characteristic Dance Primitives, *Ronghui Li, YuXiang Zhang, Yachao Zhang, Hongwen Zhang, Jie Guo, Yan Zhang, Yebin Liu, Xiu Li*
- 133 GALA: Generating Animatable Layered Assets from a Single Scan, *Taeksoo Kim, Byungjun Kim, Shunsuke Saito, Hanbyul Joo*
- 134 MMM: Generative Masked Motion Model, *Ekkasit Pinyoanuntapong, Pu Wang, Minwoo Lee, Chen Chen*
- 135 What Do You See in Vehicle? Comprehensive Vision Solution for In-Vehicle Gaze Estimation, *Yihua Cheng, Yanning Zhu, Zongji Wang, Hongquan Hao, Yongwei Liu, Shiqing Cheng, Xi Wang, Hyung Jin Chang*
- 136 Towards Variable and Coordinated Holistic Co-Speech Motion Generation, *Yifei Liu, Qiong Cao, Yandong Wen, Huaiguang Jiang, Changxing Ding*
- 137 Text2HOL: Text-guided 3D Motion Generation for Hand-Object Interaction, *Junuk Cha, Jihyeon Kim, Jae Shin Yoon, Seungryul Baek*
- 138 Garment Recovery with Shape and Deformation Priors, *Ren Li, Corentin Dumery, Benoît Guillard, Pascal Fua*
- 139 Tri-Modal Motion Retrieval by Learning a Joint Embedding Space, *Kangning Yin, Shihao Zou, Yuxuan Ge, Zheng Tian*
- 140 SplattingAvatar: Realistic Real-Time Human Avatars with Mesh-Embedded Gaussian Splatting, *Zhijing Shao, Zhaolong Wang, Zhuang Li, Duotun Wang, Xiangru Lin, Yu Zhang, Mingming Fan, Zeyu Wang*
- 141 Multi-agent Long-term 3D Human Pose Forecasting via Interaction-aware Trajectory Conditioning, *Jaewoo Jeong, Daehee Park, Kuk-Jin Yoon*
- 142 HardMo: A Large-Scale Hardcase Dataset for Motion Capture, *Jiaqi Liao, Chuanchen Luo, Yinuo Du, Yuxi Wang, Xucheng Yin, Man Zhang, Zhaoxiang Zhang, Junran Peng*
- 143 LAFS: Landmark-based Facial Self-supervised Learning for Face Recognition, *Zhonglin Sun, Chen Feng, Ioannis Patras, Georgios Tzimiropoulos*
- 144 Motion Diversification Networks, *Hee Jae Kim, Eshed Ohn-Bar*
- 145 NRDF: Neural Riemannian Distance Fields for Learning Articulated Pose Priors, *Yannan He, Garvita Tiwari, Tolga Birdal, Jan Eric Lenssen, Gerard Pons-Moll*
- 146 3D Face Reconstruction with the Geometric Guidance of Facial Part Segmentation, *Zidu Wang, Xiangyu Zhu, Tianshuo Zhang, Baiqin Wang, Zhen Lei*
- 147 Unifying Top-down and Bottom-up Scanpath Prediction Using Transformers, *Zhibo Yang, Sounak Mondal, Seoyoung Ahn, Ruoyu Xue, Gregory Zelinsky, Minh Hoai, Dimitris Samaras*
- 148 CLIB-FIQA: Face Image Quality Assessment with Confidence Calibration, *Fu-Zhao Ou, Chongyi Li, Shiqi Wang, Sam Kwong*
- 149 MoST: Motion Style Transformer Between Diverse Action Contents, *Boeun Kim, Jungho Kim, Hyung Jin Chang, Jin Young Choi*
- 150 TexVocab: Texture Vocabulary-conditioned Human Avatars, *Yuxiao Liu, Zhe Li, Yebin Liu, Haoqian Wang*
- 151 Forecasting of 3D Whole-body Human Poses with Grasping Objects, *Haitao Yan, Qiongjie Cui, Jiexin Xie, Shijie Guo*
- 152 Scaling Up Dynamic Human-Scene Interaction Modeling, *Nan Jiang, Zhiyuan Zhang, Hongjie Li, Xiaoxuan Ma, Zan Wang, Yixin Chen, Tengyu Liu, Yixin Zhu, Siyuan Huang*
- 153 Design2Cloth: 3D Cloth Generation from 2D Masks, *Jiali Zheng, Rolandos Alexandros Potamias, Stefanos Zafeiriou*
- 154 ReGenNet: Towards Human Action-Reaction Synthesis, *Liang Xu, Yizhou Zhou, Yichao Yan, Xin Jin, Wenhan Zhu, Fengyun Rao, Xiaokang Yang, Wenjun Zeng*
- 155 MoSAR: Monocular Semi-Supervised Model for Avatar Reconstruction using Differentiable Shading, *Abdallah Dib, Luiz Gustavo Hafemann, Emeline Got, Trevor Anderson, Amin Fadaeinejad, Rafael M. O. Cruz, Marc-André Carbonneau*
- 156 FaceLift: Semi-supervised 3D Facial Landmark Localization, *David Ferman, Pablo Garrido, Gaurav Bharaj*
- 157 Fast Adaptation for Human Pose Estimation via Meta-Optimization, *Shengxiang Hu, Huaijiang Sun, Bin Li, Dong Wei, Weiqing Li, Jianfeng Lu*
- 158 FlashAvatar: High-fidelity Head Avatar with Efficient Gaussian Embedding, *Jun Xiang, Xuan Gao, Yudong Guo, Juyong Zhang*
- 159 AAMDM: Accelerated Auto-regressive Motion Diffusion Model, *Tianyu Li, Calvin Qiao, Guanqiao Ren, KangKang Yin, Sehoon Ha*
- 160 SynSP: Synergy of Smoothness and Precision in Pose Sequences Refinement, *Tao Wang, Lei Jin, Zheng Wang, Jianshu Li, Liang Li, Fang Zhao, Yu Cheng, Li Yuan, Li Zhou, Junliang Xing, Jian Zhao*
- 161 AiOS: All-in-One-Stage Expressive Human Pose and Shape Estimation, *Qingping Sun, Yanjun Wang, Ailing Zeng, Wanqi Yin, Chen Wei, Wenjia Wang, Haiyi Mei, Chi-Sing Leung, Ziwei Liu, Lei Yang, Zhongang Cai*
- 162 HumanRef: Single Image to 3D Human Generation via Reference-Guided Diffusion, *Jingbo Zhang, Xiaoyu Li, Qi Zhang, Yanpei Cao, Ying Shan, Jing Liao*
- 163 Generating Human Motion in 3D Scenes from Text Descriptions, *Zhi Cen, Huaijin Pi, Sida Peng, Zehong Shen, Minghui Yang, Shuai Zhu, Hujun Bao, Xiaowei Zhou*
- 164 Stratified Avatar Generation from Sparse Observations
- 165 Locally Adaptive Neural 3D Morphable Models, *Michail Tarasiou, Rolandos Alexandros Potamias, Eimear O'Sullivan, Stylianos Ploumpis, Stefanos Zafeiriou*
- 166 IntrinsicAvatar: Physically Based Inverse Rendering of Dynamic Humans from Monocular Videos via Explicit Ray Tracing, *Shaofei Wang, Bozidar Antic, Andreas Geiger, Siyu Tang*
- 167 Dynamic Inertial Poser (DynaIP): Part-Based Motion Dynamics Learning for Enhanced Human Pose Estimation with Sparse Inertial Sensors, *Yu Zhang, Songpengcheng Xia, Lei Chu, Jiarui Yang, Qi Wu, Ling Pei*
- 168 MoMask: Generative Masked Modeling of 3D Human Motions, *Chuan Guo, Yuxuan Mu, Muhammad Gohar Javed, Sen Wang, Li Cheng*
- 169 G-HOP: Generative Hand-Object Prior for Interaction Reconstruction and Grasp Synthesis, *Yufei Ye, Abhinav Gupta, Kris Kitani, Shubham Tulsiani*
- 170 Dynamic Support Information Mining for Category-Agnostic Pose Estimation, *Pengfei Ren, Yuanyuan Gao, Haifeng Sun, Qi Qi, Jingyu Wang, Jianxin Liao*
- 171 Gaussian Head Avatar: Ultra High-fidelity Head Avatar via Dynamic Gaussians, *Yuelang Xu, Benwang Chen, Zhe Li, Hongwen Zhang, Lizhen Wang, Zerong Zheng, Yebin Liu*

- 172 Emotional Speech-driven 3D Body Animation via Disentangled Latent Diffusion, *Kiran Chhatre, Radek Daněček, Nikos Athanasiou, Giorgio Becherini, Christopher Peters, Michael J. Black, Timo Bolkart*
- 173 ProxyCap: Real-time Monocular Full-body Capture in World Space via Human-Centric Proxy-to-Motion Learning, *Yuxiang Zhang, Hongwen Zhang, Liangxiao Hu, Jiajun Zhang, Hongwei Yi, Shengping Zhang, Yebin Liu*
- 174 MAS: Multi-view Ancestral Sampling for 3D Motion Generation Using 2D Diffusion, *Roy Kapon, Guy Tevet, Daniel Cohen-Or, Amit H. Bermano*
- 175 Efficient 3D Implicit Head Avatar with Mesh-anchored Hash Table Blendshapes, *Ziqian Bai, Feitong Tan, Sean Fanello, Rohit Pandey, Mingsong Dou, Shichen Liu, Ping Tan, Yinda Zhang*
- 176 Neural Sign Actors: A Diffusion Model for 3D Sign Language Production from Text, *Vasileios Baltatzis, Rolandos Alexandros Potamias, Evangelos Ververas, Guanxiong Sun, Jiankang Deng, Stefanos Zafeiriou*
- 177 RAM-Avatar: Real-time Photo-Realistic Avatar from Monocular Videos with Full-body Control, *Xiang Deng, Zerong Zheng, Yuxiang Zhang, Jingxiang Sun, Chao Xu, Xiaodong Yang, Lizhen Wang, Yebin Liu*
- 178 Sharingan: A Transformer Architecture for Multi-Person Gaze Following, *Samy Tafasca, Anshul Gupta, Jean-Marc Odobez*
- 179 Degrees of Freedom Matter: Inferring Dynamics from Point Trajectories, *Yan Zhang, Sergey Prokudin, Marko Mihajlovic, Qianli Ma, Siyu Tang*
- 180 Authentic Hand Avatar from a Phone Scan via Universal Hand Model, *Gyeongik Moon, Weipeng Xu, Rohan Joshi, Chenglei Wu, Takaaki Shiratori*
- 181 UniHuman: A Unified Model For Editing Human Images in the Wild, *Nannan Li, Qing Liu, Krishna Kumar Singh, Yilin Wang, Jianming Zhang, Bryan A. Plummer, Zhe Lin*
- 182 BlockGCN: Redefine Topology Awareness for Skeleton-Based Action Recognition, *Yuxuan Zhou, Xudong Yan, Zhi-Qi Cheng, Yan Yan, Qi Dai, Xian-Sheng Hua*
- 183 GoMAvatar: Efficient Animatable Human Modeling from Monocular Video Using Gaussians-on-Mesh, *Jing Wen, Xiaoming Zhao, Zhongzheng Ren, Alexander G. Schwing, Shenlong Wang*
- 184 WHAM: Reconstructing World-grounded Humans with Accurate 3D Motion, *Soyong Shin, Juyong Kim, Eni Halilaj, Michael J. Black*
- 185 Self-Supervised Facial Representation Learning with Facial Region Awareness, *Zheng Gao, Ioannis Patras*
- 186 ChatPose: Chatting about 3D Human Pose, *Yao Feng, Jing Lin, Sai Kumar Dwivedi, Yu Sun, Priyanka Patel, Michael J. Black*
- 187 AUEditNet: Dual-Branch Facial Action Unit Intensity Manipulation with Implicit Disentanglement, *Shiwei Jin, Zhen Wang, Lei Wang, Peng Liu, Ning Bi, Truong Nguyen*
- 188 Towards a Simultaneous and Granular Identity-Expression
* Control in Personalized Face Generation, *Renshuai Liu, Bowen Ma, Wei Zhang, Zhipeng Hu, Changjie Fan, Tangjie Lv, Yu Ding, Xuan Cheng*
- 189 PoselRM: Enhance 3D Human Pose Estimation on Unseen Camera Settings via Invariant Risk Minimization, *Yanlu Cai, Weizhong Zhang, Yuan Wu, Cheng Jin*
- 190 Rethinking Human Motion Prediction with Symplectic Integral, *Haipeng Chen, Kedi Lyu, Zhenguang Liu, Yifang Yin, Xun Yang, Yingda Lyu*
- 191 Multimodal Sense-Informed Forecasting of 3D Human Motions, *Zhenyu Lou, Qiongjie Cui, Haofan Wang, Xu Tang, Hong Zhou*
- 192 Semantics-aware Motion Retargeting with Vision-Language Models, *Haodong Zhang, Zhike Chen, Haocheng Xu, Lei Hao, Xiaofei Wu, Songcen Xu, Zhensong Zhang, Yue Wang, Rong Xiong*
- 193 Makeup Prior Models for 3D Facial Makeup Estimation and Applications, *Xingchao Yang, Takafumi Taketomi, Yuki Endo, Yoshihiro Kanamori*
- 194 FaceCom: Towards High-fidelity 3D Facial Shape Completion via Optimization and inpainting Guidance, *Yinglong Li, Hongyu Wu, Xiaogang Wang, Qingzhao Qin, Yijiao Zhao, Yong Wang, Aimin Hao*
- 195 When StyleGAN Meets Stable Diffusion: a W+ Adapter for Personalized Image Generation, *Xiaoming Li, Xinyu Hou, Chen Change Loy*
- 196 MANUS: Markerless Grasp Capture using Articulated 3D Gaussians, *Chandradeep Pokhariya, Ishaan Nikhil Shah, Angela Xing, Zekun Li, Kefan Chen, Avinash Sharma, Srinath Sridhar*
- 197 Loose Inertial Poser: Motion Capture with IMU-attached Loose-Wear Jacket, *Chengxu Zuo, Yiming Wang, Lishuang Zhan, Shihui Guo, Xinyu Yi, Feng Xu, Yipeng Qin*
- 198 Anatomically Constrained Implicit Face Models, *Prashanth Chandran, Gaspard Zoss*
- 199 DiffusionRegPose: Enhancing Multi-Person Pose Estimation using a Diffusion-Based End-to-End Regression Approach, *Dayi Tan, Hansheng Chen, Wei Tian, Lu Xiong*
- 200 A Dual-Augmentor Framework for Domain Generalization in 3D Human Pose Estimation, *Qucheng Peng, Ce Zheng, Chen Chen*
- 201 RELI1D: A Comprehensive Multimodal Human Motion Dataset and Method, *Ming Yan, Yan Zhang, Shuqiang Cai, Shuqi Fan, Xincheng Lin, Yudi Dai, Siqi Shen, Chenglu Wen, Lan Xu, Yuexin Ma, Cheng Wang*
- 202 Co-Speech Gesture Video Generation via Motion-Decoupled Diffusion Model, *Xu He, Qiaochu Huang, Zhensong Zhang, Zhiwei Lin, Zhiyong Wu, Sicheng Yang, Minglei Li, Zhiyi Chen, Songcen Xu, Xiaofei Wu*
- 203 HandDiff: 3D Hand Pose Estimation with Diffusion on Image-Point Cloud, *Wencan Cheng, Hao Tang, Luc Van Gool, Jong Hwan Ko*
- 204 Normalizing Flows on the Product Space of SO(3) Manifolds for Probabilistic Human Pose Modeling, *Olaf Dünkel, Tim Salzmann, Florian Pfaff*
- 205 Towards Robust 3D Pose Transfer with Adversarial Learning, *Haoyu Chen, Hao Tang, Ehsan Adeli, Guoying Zhao*
- 206 PhysPT: Physics-aware Pretrained Transformer for Estimating Human Dynamics from Monocular Videos, *Yufei Zhang, Jeffrey O. Kephart, Zijun Cui, Qiang Ji*
- 207 HumMUSS: Human Motion Understanding using State Space Models, *Arnab Mondal, Stefano Alletto, Denis Tome*
- 208 MultiPhys: Multi-Person Physics-aware 3D Motion Estimation, *Nicolas Ugrinovic, Boxiao Pan, Georgios Pavlakos, Despoina Paschalidou, Bokui Shen, Jordi Sanchez-Riera, Francesc Moreno-Noguer, Leonidas Guibas*
- 209 Physics-Aware Hand-Object Interaction Denoising, *Haowen Luo, Yunze Liu, Li Yi*
- 210 HOIST-Former: Hand-held Objects Identification Segmentation and Tracking in the Wild, *Supreeth Narasimhaswamy, Huy Anh Nguyen, Lihan Huang, Minh Hoai*
- 211 SCULPT: Shape-Conditioned Unpaired Learning of Pose-dependent Clothed and Textured Human Meshes, *Soubhik Sanyal, Partha Ghosh, Jinlong Yang, Michael J. Black, Justus Thies, Timo Bolkart*
- 212 PFStorer: Personalized Face Restoration and Super-Resolution, *Tuomas Varanka, Tapani Toivonen, Soumya Tripathy, Guoying Zhao, Erman Acar*
- 213 MS-MANO: Enabling Hand Pose Tracking with Biomechanical Constraints, *Pengfei Xie, Wenqiang Xu, Tutian Tang, Zhenjun Yu, Cewu Lu*
- 214 BOTH2Hands: Inferring 3D Hands from Both Text Prompts and Body Dynamics, *Wenqian Zhang, Molin Huang, Yuxuan Zhou, Juzhe Zhang, Jingyi Yu, Jingya Wang, Lan Xu*
- 215 MeshPose: Unifying DensePose and 3D Body Mesh Reconstruction, *Eric-Tuan Le, Antonis Kakolyris, Petros Koutras, Himmy Tam, Efstratios Skordos, George Papandreou, Riza Alp Güler, Iasonas Kokkinos*
- 216 CustomListener: Text-guided Responsive Interaction for User-friendly Listening Head Generation, *Xi Liu, Ying Guo, Cheng Zhen, Tong Li, Yingying Ao, Pengfei Yan*
- 217 Generalizable Face Landmarking Guided by Conditional Face Warping, *Jiayi Liang, Haotian Liu, Hongteng Xu, Dixin Luo*
- 218 Skeleton-in-Context: Unified Skeleton Sequence Modeling with In-Context Learning, *Xinshun Wang, Zhongbin Fang, Xia Li, Xiangtai Li, Chen Chen, Mengyuan Liu*

- 219 A Unified and Interpretable Emotion Representation and Expression Generation, *Reni Paskaleva, Mykyta Holubakha, Andela Ilic, Saman Motamed, Luc Van Gool, Danda Paudel*
- 220 Artist-Friendly Relightable and Animatable Neural Heads, *Yingyan Xu, Prashanth Chandran, Sebastian Weiss, Markus Gross, Gaspard Zoss, Derek Bradley*
- 221 HanDiffuser: Text-to-Image Generation With Realistic Hand Appearances, *Supreeth Narasimhaswamy, Uttaran Bhattacharya, Xiang Chen, Ishita Dasgupta, Saayan Mitra, Minh Hoai*
- 222 BodyMAP - Jointly Predicting Body Mesh and 3D Applied Pressure Map for People in Bed, *Abhishek Tandon, Anujraaj Goyal, Henry M. Clever, Zackory Erickson*
- 223 3D Facial Expressions through Analysis-by-Neural-Synthesis, *George Retsinas, Panagiotis P. Filntisis, Radek Danecsek, Victoria F. Abrevaya, Anastasios Roussos, Timo Bolkart, Petros Maragos*
- 224 SelfPose3d: Self-Supervised Multi-Person Multi-View 3d Pose Estimation, *Vinkle Srivastav, Keqi Chen, Nicolas Padoy*
- 225 DiffusionPoser: Real-time Human Motion Reconstruction From Arbitrary Sparse Sensors Using Autoregressive Diffusion, *Tom Van Wouwe, Seunghwan Lee, Antoine Falisse, Scott Delp, C. Karen Liu*
- 226 Specularity Factorization for Low-Light Enhancement, * *Saurabh Saini, P J Narayanan*
- 227 Learning Diffusion Texture Priors for Image Restoration, * *Tian Ye, Sixiang Chen, Wenhao Chai, Zhaohu Xing, Jing Qin, Ge Lin, Lei Zhu*
- 228 Upscale-A-Video: Temporal-Consistent Diffusion Model for Real- * *World Video Super-Resolution, Shangchen Zhou, Peiqing Yang, Jianyi Wang, Yihang Luo, Chen Change Loy*
- 229 Enhancing Video Super-Resolution via Implicit Resampling- * *based Alignment, Kai Xu, Ziwei Yu, Xin Wang, Michael Bi Mi, Angela Yao*
- 230 Boosting Neural Representations for Videos with a Conditional * *Decoder, Xinjie Zhang, Ren Yang, Dailan He, Xingtong Ge, Tongda Xu, Yan Wang, Hongwei Qin, Jun Zhang*
- 231 FlowIE: Efficient Image Enhancement via Rectified Flow, * *Yixuan Zhu, Wenliang Zhao, Ao Li, Yansong Tang, Jie Zhou, Jiwen Lu*
- 232 Restoration by Generation with Constrained Priors, * *Zheng Ding, Xuaner Zhang, Zhuowen Tu, Zhihao Xia*
- 233 Towards Robust Event-guided Low-Light Image Enhancement: * *A Large-Scale Real-World Event-Image Dataset and Novel Approach, Guoqiang Liang, Kanghao Chen, Hangyu Li, Yunfan Lu, Lin Wang*
- 234 Bilateral Event Mining and Complementary for Event Stream * *Super-Resolution, Zhilin Huang, Quanmin Liang, Yijie Yu, Chujun Qin, Xiawu Zheng, Kai Huang, Zikun Zhou, Wenming Yang*
- 235 Fantastic Animals and Where to Find Them: Segment Any * *Marine Animal with Dual SAM, Pingping Zhang, Tianyu Yan, Yang Liu, Huchuan Lu*
- 236 Estimating Extreme 3D Image Rotations using Cascaded Attention, *Shay Dekel, Yosi Keller, Martin Cadik*
- 237 Learned Scanpaths Aid Blind Panoramic Video Quality Assessment, *Kanglong Fan, Wen Wen, Mu Li, Yifan Peng, Kede Ma*
- 238 Automatic Controllable Colorization via Imagination, *Xiaoyan Cong, Yue Wu, Qifeng Chen, Chenyang Lei*
- 239 Reconstruction-free Cascaded Adaptive Compressive Sensing, *Chenxi Qiu, Tao Yue, Xuemei Hu*
- 240 A Semi-supervised Nighttime Dehazing Baseline with Spatial-Frequency Aware and Realistic Brightness Constraint, *Xiaofeng Cong, Jie Gui, Jing Zhang, Junming Hou, Hao Shen*
- 241 AdaBM: On-the-Fly Adaptive Bit Mapping for Image Super-Resolution, *Cheeun Hong, Kyoung Mu Lee*
- 242 Beyond Image Super-Resolution for Image Recognition with Task-Driven Perceptual Loss, *Jaeha Kim, Junghun Oh, Kyoung Mu Lee*
- 243 Boosting Image Quality Assessment through Efficient Transformer Adaptation with Local Feature Enhancement, *Kangmin Xu, Liang Liao, Jing Xiao, Chaofeng Chen, Haoning Wu, Qiong Yan, Weisi Lin*
- 244 Blur-aware Spatio-temporal Sparse Transformer for Video Deblurring, *Huicong Zhang, Haozhe Xie, Hongxun Yao*
- 245 XFeat: Accelerated Features for Lightweight Image Matching, *Guilherme Potje, Felipe Cadar, André Araujo, Renato Martins, Erickson R. Nascimento*
- 246 RecDiffusion: Rectangling for Image Stitching with Diffusion Models, *Tianhao Zhou, Haipeng Li, Ziyi Wang, Ao Luo, Chen-Lin Zhang, Jiajun Li, Bing Zeng, Shuaicheng Liu*
- 247 Unsupervised Salient Instance Detection, *Xin Tian, Ke Xu, Rynson Lau*
- 248 FINER: Flexible Spectral-bias Tuning in Implicit NEural Representation by Variable-periodic Activation Functions, *Zhen Liu, Hao Zhu, Qi Zhang, Jingde Fu, Weibing Deng, Zhan Ma, Yanwen Guo, Xun Cao*
- 249 FMA-Net: Flow-Guided Dynamic Filtering and Iterative Feature * *Refinement with Multi-Attention for Joint Video Super-Resolution and Deblurring, Geunhyuk Youk, Jihyong Oh, Munchurl Kim*
- 250 Robust Image Denoising through Adversarial Frequency Mixup, *Donghun Ryou, Inju Ha, Hyewon Yoo, Dongwan Kim, Bohyung Han*
- 251 Efficient Multi-scale Network with Learnable Discrete Wavelet Transform for Blind Motion Deblurring, *Xin Gao, Tianheng Qiu, Xinyu Zhang, Hanlin Bai, Kang Liu, Xuan Huang, Hu Wei, Guoying Zhang, Huaping Liu*
- 252 Efficient Scene Recovery Using Luminous Flux Prior, *Zhongyu Li, Lei Zhang*
- 253 Perception-Oriented Video Frame Interpolation via Asymmetric Blending, *Guangyang Wu, Xin Tao, Changlin Li, Wenyi Wang, Xiaohong Liu, Qingqing Zheng*
- 254 Modular Blind Video Quality Assessment, *Wen Wen, Mu Li, Yabin Zhang, Yiting Liao, Junlin Li, Li Zhang, Kede Ma*
- 255 Residual Denoising Diffusion Models, *Jiawei Liu, Qiang Wang, Huijie Fan, Yinong Wang, Yandong Tang, Liangqiong Qu*
- 256 JDEC: JPEG Decoding via Enhanced Continuous Cosine Coefficients, *Woo Kyoung Han, Sunghoon Im, Jaedeok Kim, Kyong Hwan Jin*
- 257 On the Robustness of Language Guidance for Low-Level Vision Tasks: Findings from Depth Estimation, *Agneet Chatterjee, Tejas Gokhale, Chitta Baral, Yezhou Yang*
- 258 Blur2Blur: Blur Conversion for Unsupervised Image Deblurring on Unknown Domains, *Bang-Dang Pham, Phong Tran, Anh Tran, Cuong Pham, Rang Nguyen, Minh Hoai*
- 259 Exploring Efficient Asymmetric Blind-Spots for Self-Supervised Denoising in Real-World Scenarios, *Shiyan Chen, Jiyuan Zhang, Zhaofei Yu, Tiejun Huang*
- 260 Deep Equilibrium Diffusion Restoration with Parallel Sampling, *Jiezhong Cao, Yue Shi, Kai Zhang, Yulun Zhang, Radu Timofte, Luc Van Gool*
- 261 PTM-VQA: Efficient Video Quality Assessment Leveraging Diverse PreTrained Models from the Wild, *Kun Yuan, Hongbo Liu, Mading Li, Muyi Sun, Ming Sun, Jiachao Gong, Jinhua Hao, Chao Zhou, Yansong Tang*
- 262 Depth Information Assisted Collaborative Mutual Promotion Network for Single Image Dehazing, *Yafei Zhang, Shen Zhou, Huafeng Li*
- 263 Transcending the Limit of Local Window: Advanced Super-Resolution Transformer with Adaptive Token Dictionary, *Leheng Zhang, Yawei Li, Xingyu Zhou, Xiaorui Zhao, Shuhang Gu*
- 264 Improving Image Restoration through Removing Degradations in Textual Representations, *Jingbo Lin, Zhilu Zhang, Yuxiang Wei, Dongwei Ren, Dongsheng Jiang, Qi Tian, Wangmeng Zuo*
- 265 Towards Real-World HDR Video Reconstruction: A Large-Scale Benchmark Dataset and A Two-Stage Alignment Network, *Yong Shu, Liqian Shen, Xiangyu Hu, Mengyao Li, Zihao Zhou*
- 266 Spatio-Temporal Turbulence Mitigation: A Translational Perspective, *Xingguang Zhang, Nicholas Chimitt, Yiheng Chi, Zhiyuan Mao, Stanley H. Chan*
- 267 Boosting Image Restoration via Priors from Pre-trained Models, *Xiaogang Xu, Shu Kong, Tao Hu, Zhe Liu, Hujun Bao*
- 268 Misalignment-Robust Frequency Distribution Loss for Image Transformation, *Zhangkai Ni, Juncheng Wu, Zian Wang, Wenhan Yang, Hanli Wang, Lin Ma*
- 269 CoDe: An Explicit Content Decoupling Framework for Image Restoration, *Enxuan Gu, Hongwei Ge, Yong Guo*

- 270 DSL-FIQA: Assessing Facial Image Quality via Dual-Set Degradation Learning and Landmark-Guided Transformer, *Wei-Ting Chen, Gurunandan Krishnan, Qiang Gao, Sy-Yen Kuo, Sizhou Ma, Jian Wang*
- 271 CLIPtone: Unsupervised Learning for Text-based Image Tone Adjustment, *Hyeongmin Lee, Kyoungkook Kang, Jungseul Ok, Sunghyun Cho*
- 272 Adapt or Perish: Adaptive Sparse Transformer with Attentive Feature Refinement for Image Restoration, *Shihao Zhou, Duosheng Chen, Jinshan Pan, Jinglei Shi, Jufeng Yang*
- 273 CPGA: Coding Priors-Guided Aggregation Network for Compressed Video Quality Enhancement, *Qiang Zhu, Jinhua Hao, Yukang Ding, Yu Liu, Qiao Mo, Ming Sun, Chao Zhou, Shuyuan Zhu*
- 274 Learning to Control Camera Exposure via Reinforcement Learning, *Kyunghyun Lee, Ukcheol Shin, Byeong-Uk Lee*
- 275 Real-Time Exposure Correction via Collaborative Transformations and Adaptive Sampling, *Ziwen Li, Feng Zhang, Meng Cao, Jinpu Zhang, Yuanjie Shao, Yuehuan Wang, Nong Sang*
- 276 Towards Progressive Multi-Frequency Representation for Image Warping, *Jun Xiao, Zihang Lyu, Cong Zhang, Yakun Ju, Changjian Shui, Kin-Man Lam*
- 277 HIR-Diff: Unsupervised Hyperspectral Image Restoration Via Improved Diffusion Models, *Li Pang, Xiangyu Rui, Long Cui, Hongzhong Wang, Deyu Meng, Xiangyong Cao*
- 278 ZERO-IG: Zero-Shot Illumination-Guided Joint Denoising and Adaptive Enhancement for Low-Light Images, *Yiqi Shi, Duo Liu, Liguang Zhang, Ye Tian, Xuezhi Xia, Xiaojing Fu*
- 279 Masked and Shuffled Blind Spot Denoising for Real-World Images, *Hamadi Chihaoui, Paolo Favaro*
- 280 Continuous Optical Zooming: A Benchmark for Arbitrary-Scale Image Super-Resolution in Real World, *Huiyuan Fu, Fei Peng, Xianwei Li, Yejun Li, Xin Wang, Huadong Ma*
- 281 Laplacian-guided Entropy Model in Neural Codec with Blur-dissipated Synthesis, *Atefeh Khoshkhahtinat, Ali Zafari, Piyush M. Mehta, Nasser M. Nasrabadi*
- 282 SD2Event: Self-supervised Learning of Dynamic Detectors and Contextual Descriptors for Event Cameras, *Yuan Gao, Yuqing Zhu, Xinjun Li, Yimin Du, Tianzhu Zhang*
- 283 LLaFS: When Large Language Models Meet Few-Shot Segmentation, *Lanyun Zhu, Tianrun Chen, Deyi Ji, Jieping Ye, Jun Liu*
- 284 Telling Left from Right: Identifying Geometry-Aware Semantic Correspondence, *Junyi Zhang, Charles Herrmann, Junhwa Hur, Eric Chen, Varun Jampani, Deqing Sun, Ming-Hsuan Yang*
- 285 One-Shot Open Affordance Learning with Foundation Models, *Gen Li, Deqing Sun, Laura Sevilla-Lara, Varun Jampani*
- 286 CorrMatch: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation, *Boyuan Sun, Yuqi Yang, Le Zhang, Ming-Ming Cheng, Qibin Hou*
- 287 Collaborating Foundation Models for Domain Generalized Semantic Segmentation, *Yasser Benigim, Subhankar Roy, Slim Essid, Vicky Kalogeiton, Stéphane Lathuilière*
- 288 FocSAM: Delving Deeply into Focused Objects in Segmenting Anything, *You Huang, Zongyu Lan, Liujuan Cao, Xianming Lin, Shengchuan Zhang, Guannan Jiang, Rongrong Ji*
- 289 Finsler-Laplace-Beltrami Operators with Application to Shape Analysis, *Simon Weber, Thomas Dagès, Maolin Gao, Daniel Cremers*
- 290 Neural Implicit Representation for Building Digital Twins of Unknown Articulated Objects, *Yijia Weng, Bowen Wen, Jonathan Tremblay, Valts Blukis, Dieter Fox, Leonidas Guibas, Stan Birchfield*
- 291 Putting the Object Back into Video Object Segmentation, * *Ho Kei Cheng, Seoung Wug Oh, Brian Price, Joon-Young Lee, Alexander Schwing*
- 292 BA-SAM: Scalable Bias-Mode Attention Mask for Segment Anything Model, *Yiran Song, Qianyu Zhou, Xiangtai Li, Deng-Ping Fan, Xuequan Lu, Lizhuang Ma*
- 293 Task-aligned Part-aware Panoptic Segmentation through Joint Object-Part Representations, *Daan de Geus, Gijs Dubbelman*
- 294 Open-World Semantic Segmentation Including Class Similarity, *Matteo Sodano, Federico Magistri, Lucas Nunes, Jens Behley, Cyrill Stachniss*
- 295 Hierarchical Histogram Threshold Segmentation – Auto-terminating High-detail Oversegmentation, *Thomas V. Chang, Simon Seibt, Bartosz von Rymon Lipinski*
- 296 AlignSAM: Aligning Segment Anything Model to Open Context via Reinforcement Learning, *Duojun Huang, Xinyu Xiong, Jie Ma, Jichang Li, Zequn Jie, Lin Ma, Guanbin Li*
- 297 SANeRF-HQ: Segment Anything for NeRF in High Quality, *Yichen Liu, Benran Hu, Chi-Keung Tang, Yu-Wing Tai*
- 298 UniVS: Unified and Universal Video Segmentation with Prompts as Queries, *Minghan Li, Shuai Li, Xindong Zhang, Lei Zhang*
- 299 RankED: Addressing Imbalance and Uncertainty in Edge Detection Using Ranking-based Losses, *Bedrettin Cetinkaya, Sinan Kalkan, Emre Akbas*
- 300 Event-assisted Low-Light Video Object Segmentation, *Hebei Li, Jin Wang, Jiahui Yuan, Yue Li, Wenming Weng, Yansong Peng, Yueyi Zhang, Zhiwei Xiong, Xiaoyan Sun*
- 301 Density-Guided Semi-Supervised 3D Semantic Segmentation with Dual-Space Hardness Sampling, *Jianan Li, Qiulei Dong*
- 302 Exploring Regional Clues in CLIP for Zero-Shot Semantic Segmentation, *Yi Zhang, Meng-Hao Guo, Miao Wang, Shi-Min Hu*
- 303 Category-Level Multi-Part Multi-Joint 3D Shape Assembly, *Yichen Li, Kaichun Mo, Yueqi Duan, He Wang, Jiequan Zhang, Lin Shao*
- 304 SA13D: Segment Any Instance in 3D Scenes, *Yingda Yin, Yuzheng Liu, Yang Xiao, Daniel Cohen-Or, Jingwei Huang, Baoquan Chen*
- 305 Towards the Uncharted: Density-Descending Feature Perturbation for Semi-supervised Semantic Segmentation, *Xiaoyang Wang, Huihui Bai, Limin Yu, Yao Zhao, Jimin Xiao*
- 306 Hybrid Functional Maps for Crease-Aware Non-Isometric Shape Matching, *Lennart Bastian, Yizheng Xie, Nassir Navab, Zorah Lähner*
- 307 Hunting Attributes: Context Prototype-Aware Learning for Weakly Supervised Semantic Segmentation, *Feilong Tang, Zhongxing Xu, Zhaojun Qu, Wei Feng, Xingjian Jiang, Zongyuan Ge*
- 308 Self-Calibrating Vicinal Risk Minimisation for Model Calibration, *Jiawei Liu, Changkun Ye, Ruikai Cui, Nick Barnes*
- 309 ECLIPSE: Efficient Continual Learning in Panoptic Segmentation with Visual Prompt Tuning, *Beomyoung Kim, Joonsang Yu, Sung Ju Hwang*
- 310 Clustering Propagation for Universal Medical Image Segmentation, *Yuhang Ding, Liulei Li, Wenguan Wang, Yi Yang*
- 311 Addressing Background Context Bias in Few-Shot Segmentation through Iterative Modulation, *Lanyun Zhu, Tianrun Chen, Jianxiang Yin, Simon See, Jun Liu*
- 312 Cross-Domain Few-Shot Segmentation via Iterative Support-Query Correspondence Mining, *Jiahao Nie, Yun Xing, Gongjie Zhang, Pei Yan, Aoran Xiao, Yap-Peng Tan, Alex C. Kot, Shijian Lu*
- 313 RankMatch: Exploring the Better Consistency Regularization for Semi-supervised Semantic Segmentation, *Huayu Mai, Rui Sun, Tianzhu Zhang, Feng Wu*
- 314 QDFormer: Towards Robust Audiovisual Segmentation in Complex Environments with Quantization-based Semantic Decomposition, *Xiang Li, Jinglu Wang, Xiaohao Xu, Xiulian Peng, Rita Singh, Yan Lu, Bhiksha Raj*
- 315 Frequency-Adaptive Dilated Convolution for Semantic Segmentation, *Linwei Chen, Lin Gu, Dezhi Zheng, Ying Fu*
- 316 SED: A Simple Encoder-Decoder for Open-Vocabulary Semantic Segmentation, *Bin Xie, Jiale Cao, Jin Xie, Fahad Shahbaz Khan, Yanwei Pang*
- 317 PSDPM: Prototype-based Secondary Discriminative Pixels Mining for Weakly Supervised Semantic Segmentation, *Xinqiao Zhao, Ziqian Yang, Tianhong Dai, Bingfeng Zhang, Jimin Xiao*
- 318 Coupled Laplacian Eigenmaps for Locally-Aware 3D Rigid Point Cloud Matching, *Matteo Bascico, Etienne Decencière, Laurent Corté, Yannick Tillier, David Ryckelynck*
- 319 Universal Segmentation at Arbitrary Granularity with Language Instruction, *Yong Liu, Cairong Zhang, Yitong Wang, Jiahao Wang, Yujiu Yang, Yansong Tang*
- 320 PartDistill: 3D Shape Part Segmentation by Vision-Language Model Distillation, *Ardian Umam, Cheng-Kun Yang, Min-Hung Chen, Jen-Hui Chuang, Yen-Yu Lin*

- 321 HIT: Estimating Internal Human Implicit Tissues from the Body Surface, *Marilyn Keller, Vaibhav Arora, Abdelmoutaleb Dakri, Shivam Chandhok, Jürgen Machann, Andreas Fritsche, Michael J. Black, Sergi Pujades*
- 322 Open-Vocabulary Segmentation with Semantic-Assisted Calibration, *Yong Liu, Sule Bai, Guanbin Li, Yitong Wang, Yansong Tang*
- 323 GraCo: Granularity-Controllable Interactive Segmentation, * *Yian Zhao, Kehan Li, Zesen Cheng, Pengchong Qiao, Xiawu Zheng, Rongrong Ji, Chang Liu, Li Yuan, Jie Chen*
- 324 Unleashing the Potential of SAM for Medical Adaptation via Hierarchical Decoding, *Zhiheng Cheng, Qingyue Wei, Hongru Zhu, Yan Wang, Liangqiong Qu, Wei Shao, Yuyin Zhou*
- 325 EAGLE: Eigen Aggregation Learning for Object-Centric * *Unsupervised Semantic Segmentation, Chanyoung Kim, Woojung Han, Dayun Ju, Seong Jae Hwang*
- 326 DuPL: Dual Student with Trustworthy Progressive Learning for Robust Weakly Supervised Semantic Segmentation, *Yuanchen Wu, Xichen Ye, Kequan Yang, Jide Li, Xiaoqiang Li*
- 327 Vanishing-Point-Guided Video Semantic Segmentation of Driving * *Scenes, Diandian Guo, Deng-Ping Fan, Tongyu Lu, Christos Sakaridis, Luc Van Gool*
- 328 Diffuse Attend and Segment: Unsupervised Zero-Shot Segmentation using Stable Diffusion, *Junjiao Tian, Lavisha Aggarwal, Andrea Colaco, Zsolt Kira, Mar Gonzalez-Franco*
- 329 ODIN: A Single Model for 2D and 3D Segmentation, * *Ayush Jain, Pushkal Katara, Nikolaos Gkanatsios, Adam W. Harley, Gabriel Sarch, Kriti Aggarwal, Vishrav Chaudhary, Katerina Fragkiadaki*
- 330 Infer from What You Have Seen Before: Temporally-dependent Classifier for Semi-supervised Video Segmentation, *Jiafan Zhuang, Zilei Wang, Yixin Zhang, Zhun Fan*
- 331 Semantic-aware SAM for Point-Prompted Instance Segmentation, * *Zhaoyang Wei, Pengfei Chen, Xuehui Yu, Guorong Li, Jianbin Jiao, Zhenjun Han*
- 332 Class Tokens Infusion for Weakly Supervised Semantic Segmentation, *Sung-Hoon Yoon, Hoyong Kwon, Hyeonseong Kim, Kuk-Jin Yoon*
- 333 Separate and Conquer: Decoupling Co-occurrence via Decomposition and Representation for Weakly Supervised Semantic Segmentation, *Zhiwei Yang, Kexue Fu, Minghong Duan, Linhao Qu, Shuo Wang, Zhijian Song*
- 334 Style Blind Domain Generalized Semantic Segmentation via Covariance Alignment and Semantic Consistence Contrastive Learning, *Woo-Jin Ahn, Geun-Yeong Yang, Hyun-Duck Choi, Myo-Taeg Lim*
- 335 AllSpark: Reborn Labeled Features from Unlabeled in Transformer for Semi-Supervised Semantic Segmentation, *Haonan Wang, Qixiang Zhang, Yi Li, Xiaomeng Li*
- 336 Unsupervised Semantic Segmentation Through Depth-Guided Feature Correlation and Sampling, *Leon Sick, Dominik Engel, Pedro Hermosilla, Timo Ropinski*
- 337 PoNQu: a Neural QEM-based Mesh Representation, *Nissim Maruan, Maks Ovsjanikov, Pierre Alliez, Mathieu Desbrun*
- 338 Spectral Meets Spatial: Harmonising 3D Shape Matching and Interpolation, *Dongliang Cao, Marvin Eisenberger, Nafie El Amrani, Daniel Cremers, Florian Bernard*
- 339 CosalPure: Learning Concept from Group Images for Robust Co-Saliency Detection, *Jiayi Zhu, Qing Guo, Felix Juefei-Xu, Yihao Huang, Yang Liu, Geguang Pu*
- 340 ContextSeg: Sketch Semantic Segmentation by Querying the Context with Attention, *Jiawei Wang, Changjian Li*
- 341 Training-Free Open-Vocabulary Segmentation with Offline Diffusion-Augmented Prototype Generation, *Luca Barsellotti, Roberto Amoroso, Marcella Cornia, Lorenzo Baraldi, Rita Cucchiara*
- 342 ASAM: Boosting Segment Anything Model with Adversarial Tuning, *Bo Li, Haoke Xiao, Lv Tang*
- 343 In-Context Matting, * *He Guo, Zixuan Ye, Zhiguo Cao, Hao Lu*
- 344 Weakly Supervised Point Cloud Semantic Segmentation via Artificial Oracle, *Hyeokjun Kweon, Jihun Kim, Kuk-Jin Yoon*
- 345 Contextrast: Contextual Contrastive Learning for Semantic Segmentation, *Changki Sung, Wanhee Kim, Jungho An, Wooju Lee, Hyungtae Lim, Hyun Myung*
- 346 Parameter Efficient Fine-tuning via Cross Block Orchestration for Segment Anything Model, *Zelin Peng, Zhengqin Xu, Zhilin Zeng, Lingxi Xie, Qi Tian, Wei Shen*
- 347 CADTalk: An Algorithm and Benchmark for Semantic * *Commenting of CAD Programs, Haocheng Yuan, Jing Xu, Hao Pan, Adrien Bousseau, Niloy J. Mitra, Changjian Li*
- 348 Point2CAD: Reverse Engineering CAD Models from 3D Point * *Clouds, Yujia Liu, Anton Obukhov, Jan Dirk Wegner, Konrad Schindler*
- 349 Rethinking Interactive Image Segmentation with Low Latency High Quality and Diverse Prompts, *Qin Liu, Jaemin Cho, Mohit Bansal, Marc Niethammer*
- 350 General Object Foundation Model for Images and Videos at * *Scale, Junfeng Wu, Yi Jiang, Qihao Liu, Zehuan Yuan, Xiang Bai, Song Bai*
- 351 Frozen CLIP: A Strong Backbone for Weakly Supervised Semantic * *Segmentation, Bingfeng Zhang, Siyue Yu, Yunchao Wei, Yao Zhao, Jimin Xiao*
- 352 Guided Slot Attention for Unsupervised Video Object Segmentation, *Minhyeok Lee, Suhwan Cho, Dogyoon Lee, Chaewon Park, Jungho Lee, Sangyoun Lee*
- 353 Unlocking the Potential of Pre-trained Vision Transformers for Few-Shot Semantic Segmentation through Relationship Descriptors, *Ziqin Zhou, Hai-Ming Xu, Yangyang Shu, Lingqiao Liu*
- 354 Grounding Everything: Emerging Localization Properties in Vision-Language Transformers, *Walid Boussethem, Felix Petersen, Vittorio Ferrari, Hilde Kuehne*
- 355 No Time to Train: Empowering Non-Parametric Networks for * *Few-shot 3D Scene Segmentation, Xiangyang Zhu, Renrui Zhang, Bowei He, Ziyu Guo, Jiaming Liu, Han Xiao, Chaoyou Fu, Hao Dong, Peng Gao*
- 356 Continual Segmentation with Disentangled Objectness Learning and Class Recognition, *Yizheng Gong, Siyue Yu, Xiaoyang Wang, Jimin Xiao*
- 357 GSVA: Generalized Segmentation via Multimodal Large Language Models, *Zhuofan Xia, Dongchen Han, Yizeng Han, Xuran Pan, Shiji Song, Gao Huang*
- 358 MaGgle: Masked Guided Gradual Human Instance Matting, *Chuong Huynh, Seoung Wug Oh, Abhinav Shrivastava, Joon-Young Lee*
- 359 EFormer: Enhanced Transformer towards Semantic-Contour Features of Foreground for Portraits Matting, *Zitao Wang, Qiguang Miao, Yue Xi, Peipei Zhao*
- 360 Segment Any Event Streams via Weighted Adaptation of Pivotal Tokens, *Zhiwen Chen, Zhiyu Zhu, Yifan Zhang, Junhui Hou, Guangming Shi, Jinjian Wu*
- 361 PolarMatte: Fully Computational Ground-Truth-Quality Alpha Matte Extraction for Images and Video using Polarized Screen Matting, *Kenji Enomoto, TJ Rhodes, Brian Price, Gavin Miller*
- 362 Segment Every Out-of-Distribution Object, *Wenjie Zhao, Jia Li, Xin Dong, Yu Xiang, Yunhui Guo*
- 363 Multi-view Aggregation Network for Dichotomous Image * *Segmentation, Qian Yu, Xiaoqi Zhao, Youwei Pang, Lihe Zhang, Huchuan Lu*
- 364 pix2gestalt: Amodal Segmentation by Synthesizing Wholes, * *Ege Ozguroglu, Ruoshi Liu, Dídac Surís, Dian Chen, Achal Dave, Pavel Tokmakov, Carl Vondrick*
- 365 Rethinking Prior Information Generation with CLIP for Few-Shot Segmentation, *Jin Wang, Bingfeng Zhang, Jian Pang, Honglong Chen, Weifeng Liu*
- 366 Image-to-Image Matching via Foundation Models: A New Perspective for Open-Vocabulary Semantic Segmentation, *Yuan Wang, Rui Sun, Naisong Luo, Yuwen Pan, Tianzhu Zhang*
- 367 Domain Separation Graph Neural Networks for Saliency Object Ranking, *Zijian Wu, Jun Lu, Jing Han, Lianfa Bai, Yi Zhang, Zhuang Zhao, Siyang Song*
- 368 DIOD: Self-Distillation Meets Object Discovery, *Sandra Kara, Hejer Ammar, Julien Denize, Florian Chabot, Quoc-Cuong Pham*

- 369 DiverGen: Improving Instance Segmentation by Learning Wider Data Distribution with More Diverse Generative Data, *Chengxiang Fan, Muzhi Zhu, Hao Chen, Yang Liu, Weijia Wu, Huaqi Zhang, Chunhua Shen*
- 370 Rethinking Few-shot 3D Point Cloud Semantic Segmentation, *Zhaochong An, Guolei Sun, Yun Liu, Fayao Liu, Zongwei Wu, Dan Wang, Luc Van Gool, Serge Belongie*
- 371 Training Vision Transformers for Semi-Supervised Semantic Segmentation, *Xinting Hu, Li Jiang, Bernt Schiele*
- 372 Open3DIS: Open-Vocabulary 3D Instance Segmentation with 2D Mask Guidance, *Phuc Nguyen, Tuan Duc Ngo, Evangelos Kalogerakis, Chuang Gan, Anh Tran, Cuong Pham, Khoi Nguyen*
- 373 Emergent Open-Vocabulary Semantic Segmentation from Off-the-shelf Vision-Language Models, *Jiayun Luo, Siddhesh Khandelwal, Leonid Sigal, Boyang Li*
- 374 Memory-Scalable and Simplified Functional Map Learning, *Robin Magnet, Maks Ovsjanikov*
- 375 MFP: Making Full Use of Probability Maps for Interactive Image Segmentation, *Chaewon Lee, Seon-Ho Lee, Chang-Su Kim*
- 376 Spherical Mask: Coarse-to-Fine 3D Point Cloud Instance Segmentation with Spherical Representation, *Sangyun Shin, Kaichen Zhou, Madhu Vankadari, Andrew Markham, Niki Trigoni*
- 377 Adaptive Bidirectional Displacement for Semi-Supervised Medical Image Segmentation, *Hanyang Chi, Jian Pang, Bingfeng Zhang, Weifeng Liu*
- 378 RobustSAM: Segment Anything Robustly on Degraded Images, * *Wei-Ting Chen, Yu-Jiet Vong, Sy-Yen Kuo, Sizhou Ma, Jian Wang*
- 379 LAKE-RED: Camouflaged Images Generation by Latent Background Knowledge Retrieval-Augmented Diffusion, *Pancheng Zhao, Peng Xu, Pengda Qin, Deng-Ping Fan, Zhicheng Zhang, Guoli Jia, Bowen Zhou, Jufeng Yang*
- 380 Learn to Rectify the Bias of CLIP for Unsupervised Semantic Segmentation, *Jingyun Wang, Guoliang Kang*
- 381 CAT-Seg: Cost Aggregation for Open-Vocabulary Semantic Segmentation, * *Seokju Cho, Heeseong Shin, Sunghwan Hong, Anurag Arnab, Paul Hongsuck Seo, Seungryong Kim*
- 382 Prompt-Driven Referring Image Segmentation with Instance Contrasting, *Chao Shang, Zichen Song, Heqian Qiu, Lanxiao Wang, Fanman Meng, Hongliang Li*
- 383 Kandinsky Conformal Prediction: Efficient Calibration of Image Segmentation Algorithms, *Joren Brunekreef, Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, Jonas Teuwen*
- 384 OVFoodSeg: Elevating Open-Vocabulary Food Image Segmentation via Image-Informed Textual Representation, *Xiongwei Wu, Sicheng Yu, Ee-Peng Lim, Chong-Wah Ngo*
- 385 Back to 3D: Few-Shot 3D Keypoint Detection with Back-Projected 2D Features, *Thomas Wimmer, Peter Wonka, Maks Ovsjanikov*
- 386 Deciphering 'What' and 'Where' Visual Pathways from Spectral * *Clustering of Layer-Distributed Neural Representations, Xiao Zhang, David Yunis, Michael Maire*
- 387 Open Vocabulary Semantic Scene Sketch Understanding, *Ahmed Bourouis, Judith E. Fan, Yulia Gryaditskaya*
- 388 USE: Universal Segment Embeddings for Open-Vocabulary Image Segmentation, *Xiaoqi Wang, Wenbin He, Xiwei Xuan, Clint Sebastian, Jorge Piazentin Ono, Xin Li, Sima Behpour, Thang Doan, Liang Gou, Han-Wei Shen, Liu Ren*
- 389 Diff-Plugin: Revitalizing Details for Diffusion-based Low-level Tasks, *Yuhao Liu, Zhanghan Ke, Fang Liu, Nanxuan Zhao, Rynson W.H. Lau*
- 390 XCube: Large-Scale 3D Generative Modeling using Sparse Voxel * *Hierarchies, Xuanchi Ren, Jiahui Huang, Xiaohui Zeng, Ken Museth, Sanja Fidler, Francis Williams*
- 391 SC-GS: Sparse-Controlled Gaussian Splatting for Editable Dynamic Scenes, *Yi-Hua Huang, Yang-Tian Sun, Ziyi Yang, Xiaoyang Lyu, Yan-Pei Cao, Xiaojuan Qi*
- 392 StylLitGAN: Image-Based Relighting via Latent Control, *Anand Bhattad, James Soole, D.A. Forsyth*
- 393 GPLD3D: Latent Diffusion of 3D Shape Generative Models by * *Enforcing Geometric and Physical Priors, Yuan Dong, Qi Zuo, Xiaodong Gu, Weihao Yuan, Zhengyi Zhao, Zilong Dong, Liefeng Bo, Qixing Huang*
- 394 Image Sculpting: Precise Object Editing with 3D Geometry Control, *Jiraphon Yenphraphai, Xichen Pan, Sainan Liu, Daniele Panozzo, Saining Xie*
- 395 Paint3D: Paint Anything 3D with Lighting-Less Texture Diffusion Models, *Xianfang Zeng, Xin Chen, Zhongqi Qi, Wen Liu, Zibo Zhao, Zhibin Wang, Bin Fu, Yong Liu, Gang Yu*
- 396 Retrieval-Augmented Layout Transformer for Content-Aware * *Layout Generation, Daichi Horita, Naoto Inoue, Kotaro Kikuchi, Kota Yamaguchi, Kiyoharu Aizawa*
- 397 Holo-Relighting: Controllable Volumetric Portrait Relighting from a Single Image, *Yiqun Mei, Yu Zeng, He Zhang, Zhixin Shu, Xuaner Zhang, Sai Bi, Jianming Zhang, HyunJoon Jung, Vishal M. Patel*
- 398 Neural Fields as Distributions: Signal Processing Beyond Euclidean Space, *Daniel Rebain, Soroosh Yazdani, Kwang Moo Yi, Andrea Tagliasacchi*
- 399 Eclipse: Disambiguating Illumination and Materials using * *Unintended Shadows, Dor Verbin, Ben Mildenhall, Peter Hedman, Jonathan T. Barron, Todd Zickler, Pratul P. Srinivasan*
- 400 TexOct: Generating Textures of 3D Models with Octree-based Diffusion, *Jialun Liu, Chenming Wu, Xinqi Liu, Xing Liu, Jinbo Wu, Haotian Peng, Chen Zhao, Haocheng Feng, Jingtuo Liu, Errui Ding*
- 401 Differentiable Micro-Mesh Construction, *Yishun Dou, Zhong Zheng, Qiaoqiao Jin, Rui Shi, Yuhan Li, Bingbing Ni*
- 402 TextureDreamer: Image-Guided Texture Synthesis Through Geometry-Aware Diffusion, *Yu-Ying Yeh, Jia-Bin Huang, Changil Kim, Lei Xiao, Thu Nguyen-Phuoc, Numair Khan, Cheng Zhang, Manmohan Chandraker, Carl S Marshall, Zhao Dong, Zhengqin Li*
- 403 As-Plausible-As-Possible: Plausibility-Aware Mesh Deformation Using 2D Diffusion Priors, *Seungwoo Yoo, Kunho Kim, Vladimir G. Kim, Minhyuk Sung*
- 404 Breathing Life Into Sketches Using Text-to-Video Priors, * *Rinon Gal, Yael Vinker, Yuval Alaluf, Amit Bermano, Daniel Cohen-Or, Ariel Shamir, Gal Chechik*
- 405 Real-Time Neural BRDF with Spherically Distributed Primitives, *Yishun Dou, Zhong Zheng, Qiaoqiao Jin, Bingbing Ni, Yugang Chen, Junxiang Ke*
- 406 Paint-it: Text-to-Texture Synthesis via Deep Convolutional Texture Map Optimization and Physically-Based Rendering, *Kim Youwang, Tae-Hyun Oh, Gerard Pons-Moll*
- 407 Neural Super-Resolution for Real-time Rendering with Radiance Demodulation, *Jia Li, Ziling Chen, Xiaolong Wu, Lu Wang, Beibei Wang, Lei Zhang*
- 408 DiffAvatar: Simulation-Ready Garment Optimization with Differentiable Simulation, *Yifei Li, Hsiao-yu Chen, Egor Larionov, Nikolaos Sarafianos, Wojciech Matusik, Tuur Stuyck*
- 409 Material Palette: Extraction of Materials from a Single Image, *Ivan Lopes, Fabio Pizzati, Raoul de Charette*
- 410 PhysGaussian: Physics-Integrated 3D Gaussians for Generative * *Dynamics, Tianyi Xie, Zeshun Zong, Yuxing Qiu, Xuan Li, Yutao Feng, Yin Yang, Chenfanfu Jiang*
- 411 Differentiable Point-based Inverse Rendering, *Hoon-Gyu Chung, Seokjun Choi, Seung-Hwan Baek*
- 412 Objects as Volumes: A Stochastic Geometry View of Opaque * *Solids, Bailey Miller, Hanyu Chen, Alice Lai, Ioannis Gkioulekas*
- 413 Towards a Perceptual Evaluation Framework for Lighting Estimation, *Justine Giroux, Mohammad Reza Karimi Dastjerdi, Yannick Hold-Geoffroy, Javier Vazquez-Corral, Jean-François Lalonde*
- 414 Vector Graphics Generation via Mutually Impulsed Dual-domain Diffusion, *Zhongyin Zhao, Ye Chen, Zhangli Hu, Xuanhong Chen, Bingbing Ni*
- 415 MatFuse: Controllable Material Generation with Diffusion Models, *Giuseppe Vecchio, Renato Sortino, Simone Palazzo, Concetto Spampinato*
- 416 DiffusionLight: Light Probes for Free by Painting a Chrome Ball, * *Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit Raj, Pramook Khungurn, Supasorn Suwajanakorn*

- 417 Textile: A Differentiable Metric for Texture Tileability, *Carlos Rodriguez-Pardo, Dan Casas, Elena Garces, Jorge Lopez-Moreno*
- 418 PIE-NeRF: Physics-based Interactive Elastodynamics with NeRF, *Yutao Feng, Yintong Shang, Xuan Li, Tianjia Shao, Chenfanfu Jiang, Yin Yang*
- 419 HashPoint: Accelerated Point Searching and Sampling for Neural
* Rendering, *Jiahao Ma, Miaomiao Liu, David Ahmedt-Aristizabal, Chuong Nguyen*
- 420 3D Paintbrush: Local Stylization of 3D Shapes with Cascaded Score Distillation, *Dale Decatur, Itai Lang, Kfir Aberman, Rana Hanocka*
- 421 DUDF: Differentiable Unsigned Distance Fields with Hyperbolic Scaling, *Miguel Fainstein, Viviana Silless, Emmanuel Iarussi*
- 422 Diffusion 3D Features (Diff3F): Decorating Untextured Shapes with Distilled Semantic Features, *Niladri Shekhar Dutt, Sanjeev Muralikrishnan, Niloy J. Mitra*
- 423 LeGO: Leveraging a Surface Deformation Network for Animatable
* Stylized Face Generation with One Example, *Soyeon Yoon, Kwan Yun, Kwangyoon Seo, Sihun Cha, Jung Eun Yoo, Junyong Noh*
- 424 Dr. Bokeh: Differentiable Occlusion-aware Bokeh Rendering, *Yichen Sheng, Zixun Yu, Lu Ling, Zhiwen Cao, Xuaner Zhang, Xin Lu, Ke Xian, Haiting Lin, Bedrich Benes*
- 425 DiffInDScene: Diffusion-based High-Quality 3D Indoor Scene Generation, *Xiaoliang Ju, Zhaoyang Huang, Yijin Li, Guofeng Zhang, Yu Qiao, Hongsheng Li*
- 426 LightOctree: Lightweight 3D Spatially-Coherent Indoor Lighting Estimation, *Xuecan Wang, Shibang Xiao, Xiaohui Liang*
- 427 SVGDreamer: Text Guided SVG Generation with Diffusion Model, *Ximing Xing, Haitao Zhou, Chuang Wang, Jing Zhang, Dong Xu, Qian Yu*
- 428 Control4D: Efficient 4D Portrait Editing with Text, *Ruizhi Shao, Jingxiang Sun, Cheng Peng, Zerong Zheng, Boyao Zhou, Hongwen Zhang, Yebin Liu*
- 429 HumanNorm: Learning Normal Diffusion Model for High-quality and Realistic 3D Human Generation, *Xin Huang, Ruizhi Shao, Qi Zhang, Hongwen Zhang, Ying Feng, Yebin Liu, Qing Wang*
- 430 Video2Game: Real-time Interactive Realistic and Browser-Compatible Environment from a Single Video, *Hongchi Xia, Zhi-Hao Lin, Wei-Chiu Ma, Shenlong Wang*
- 431 NIVel: Neural Implicit Vector Layers for Text-to-Vector Generation, *Vikas Thamizharasan, Difan Liu, Matthew Fisher, Nanxuan Zhao, Evangelos Kalogerakis, Michal Lukac*
- 432 ESR-NeRF: Emissive Source Reconstruction Using LDR Multi-view Images, *Jinseo Jeong, Junseo Koo, Qimeng Zhang, Gunhee Kim*
- 433 DreamPropeller: Supercharge Text-to-3D Generation with Parallel
* Sampling, *Linqi Zhou, Andy Shih, Chenlin Meng, Stefano Ermon*
- 434 GenesisTex: Adapting Image Denoising Diffusion to Texture Space, *Chenjian Gao, Boyan Jiang, Xinghui Li, Yingpeng Zhang, Qian Yu*
- 435 Mosaic-SDF for 3D Generative Models, *Lior Yariv, Omri Puny, Oran Gafni, Yaron Lipman*
- 436 NeRF Analogies: Example-Based Visual Attribute Transfer for NeRFs, *Michael Fischer, Zhengqin Li, Thu Nguyen-Phuoc, Aljaz Bozic, Zhao Dong, Carl Marshall, Tobias Ritschel*
- 437 Hyper-MD: Mesh Denoising with Customized Parameters Aware of Noise Intensity and Geometric Characteristics, *Xingtao Wang, Hongliang Wei, Xiaopeng Fan, Debin Zhao*
- 438 QUADify: Extracting Meshes with Pixel-level Details and
* Materials from Images, *Maximilian Frühauf, Hayko Riemenschneider, Markus Gross, Christopher Schroers*
- 439 SfmCAD: Unsupervised CAD Reconstruction by Learning Sketch-based Feature Modeling Operations, *Pu Li, Jianwei Guo, Huibin Li, Bedrich Benes, Dong-Ming Yan*
- 440 Self-Supervised Dual Contouring,
* *Ramana Sundararaman, Roman Klokov, Maks Ovsjanikov*
- 441 SVDTree: Semantic Voxel Diffusion for Single Image Tree Reconstruction, *Yuan Li, Zhihao Liu, Bedrich Benes, Xiaopeng Zhang, Jianwei Guo*
- 442 Text-Conditioned Generative Model of 3D Strand-based Human Hairstyles, *Vanessa Sklyarova, Egor Zakharov, Otmar Hilliges, Michael J. Black, Justus Thies*
- 443 CAD-SIGNet: CAD Language Inference from Point Clouds using
* Layer-wise Sketch Instance Guided Attention, *Mohammad Sadiq Khan, Elona Dupont, Sk Aziz Ali, Kseniya Cherenkova, Anis Kacem, Djamilia Aouada*
- 444 Functional Diffusion, *Biao Zhang, Peter Wonka*
- 10:30 - 18:45 Art Program** (Arch 4CDE)
- 10:30 - 18:45 DEMOS** (Arch 4CDE)
- DECIMER.ai: An Open-Source Platform for Chemical Structure Recognition from Scientific Literature, *Dr. Kohulan Rajan*
 - Computer Vision algorithms for 100 skin shades in the beauty industry, *Emmanuel Acheampong, Raymond Okyere-Forson*
 - roboMUA: A predictive and generative AI approach to makeup for 100 skin shades, *Emmanuel Acheampong*
 - Magic Mapping: Interactive Segmentation of Satellite Imagery with Embedding Fields, *Chris Brown*, Sean Askay*, Michal Kazmierski, William Rucklidge, Valerie Pasquarella, Evan Shelhamer*
 - RITA: A Real-Time Interactive Talking Avatars Framework, *Cheng Wan, Wuxinlin Cheng, Yupeng Cao*
 - ANGEL - Egocentric, AR Task Monitoring, *Brian Clipp*
 - Anyone Can Direct: 3D Film Creation from Text, *Zhongang Cai, Lei Yang, Ziwei Liu*
 - Learning Humanoid Locomotion, *Ilija Radosavovic*
 - WildfireWatch.org: Harnessing AI-Powered Visual Data for Swift and Accurate Wildfire Detection, *Kit Merker*
 - Enhancing Accuracy and Efficiency of Livestock Monitoring with AI-Powered Visual Data from Plainsight Technologies, *Kit Merker*
 - Depth Anything, *Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao*
 - Custom cloth creation and virtual try-on for everyone, *Pei Chen, Heng Wang, Zhiyuan Chen, Zhenkun Liu, Shuhua Cao, Li Yang, Minghui Yang*
 - Controllable Neural Reconstruction for Autonomous Driving, *Péter Kovács, Máté Tóth, Zoltán Bendefy, Zoltán Hortsin, Tamás Matuszka*
 - Seeing the World Through Your Eyes, *Kevin Zhang, Hadi Alzayer, Sachin Shah*
 - Aigen-AI Robotic Solution to Weed Control, *Kayla Stringfellow, Sandra Phillips*
 - Clickable Objects - Shoppable Video, *Toshiro Ozawa*
 - Contactless Optical Vital Sign Monitoring for Elderly Care, *Svenja Nicola Kobel, Caroline Rensing, Hueseyin Oenel, Andre Kosfeld, Christian Wiede, Karsten Seidl*
- 11:00 - 13:00 SOCIAL - Challenges/Opportunities for ECRs in Fast Paced AI** (Summit Terrace Suite)
**Reservation Required
- 12:00 - 14:00 LUNCH** Summit ExHall 1-2
- 13:15 - 14:15 EXPO TRACK - Swami Sivasubramanian Ph.D., Vice President, AI and Data, Amazon Web Services (AWS)** (Arch 4F)
- 13:00 - 14:30 Orals 2A: Image & Video Synthesis** (Summit Ballroom)
- FreeU: Free Lunch in Diffusion U-Net, *Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu*
 - Ranni: Taming Text-to-Image Diffusion for Accurate Instruction Following, *Yutong Feng, Biao Gong, Di Chen, Yujun Shen, Yu Liu, Jingren Zhou*
 - Instruct-Imagen: Image Generation with Multi-modal Instruction, *Hexiang Hu, Kelvin C.K. Chan, Yu-Chuan Su, Wenhu Chen, Yandong Li, Kihyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William Cohen, Ming-Wei Chang, Xuhui Jia*
 - Attention Calibration for Disentangled Text-to-Image Personalization, *Yanbing Zhang, Mengping Yang, Qin Zhou, Zhe Wang*
 - Style Aligned Image Generation via Shared Attention, *Amir Hertz, Andrey Voynov, Shlomi Fruchter, Daniel Cohen-Or*

13:00 - 14:30 Orals 2B: Deep Learning Architectures and Techniques (Summit Flex Hall AB)

- 1 Neural Redshift: Random Networks are not Random Functions, *Damien Teney, Armand Mihai Nicolicioiu, Valentin Hartmann, Ehsan Abbasnejad*
- 2 Neural Lineage, *Runpeng Yu, Xinchao Wang*
- 3 Learning Structure-from-Motion with Graph Attention Networks, *Lucas Brynte, José Pedro Iglesias, Carl Olsson, Fredrik Kahl*
- 4 Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks, *Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, Lu Yuan*
- 5 In Search of a Data Transformation That Accelerates Neural Field Training, *Junwon Seo, Sangyoon Lee, Kwang In Kim, Jaeho Lee*

13:00 - 14:30 Orals 2C: 3D from Multiview and Sensors, (Summit Flex Hall C)

- 1 Point Transformer V3: Simpler Faster Stronger, *Xiaoyang Wu, Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao*
- 2 Matching 2D Images in 3D: Metric Relative Pose from Metric Correspondences, *Axel Barroso-Laguna, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann*
- 3 Seeing the World through Your Eyes, *Hadi Alzayer, Kevin Zhang, Brandon Feng, Christopher A. Metzler, Jia-Bin Huang*
- 4 Tri-Perspective View Decomposition for Geometry-Aware Depth Completion, *Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, Jian Yang*
- 5 Steerers: A Framework for Rotation Equivariant Keypoint Descriptors, *Georg Bökman, Johan Edstedt, Michael Felsberg, Fredrik Kahl*

14:30 - 14:45 Courtesy Break

14:45 - 15:45 KEYNOTE 1 - *Joshua Bongard, Veinott Professor of Computer Science, University of Vermont* (Summit Flex Hall ABC)

15:45 - 16:00 Courtesy Break

16:00 - 17:00 PANEL 1 (Summit Flex Hall ABC)

16:45 - 17:15 Poster Setup Arch 4E

17:00 - 19:00 SOCIAL - How to Balance Research Interests and Academic Tasks (Summit Elliott Bay Room)
**Reservation Required

17:15 - 18:45 Poster Session 2 & Exhibit Hall (Arch 4A-E)

✧ - Highlight paper (check it out)

🏆 - Award candidate paper (see award sessions)

- 1 Point Transformer V3: Simpler Faster Stronger, *Xiaoyang Wu, ✧ Li Jiang, Peng-Shuai Wang, Zhijian Liu, Xihui Liu, Yu Qiao, Wanli Ouyang, Tong He, Hengshuang Zhao*
- 2 Matching 2D Images in 3D: Metric Relative Pose from Metric ✧ Correspondences, *Axel Barroso-Laguna, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann*
- 3 Seeing the World through Your Eyes, *Hadi Alzayer, Kevin Zhang, ✧ Brandon Feng, Christopher A. Metzler, Jia-Bin Huang*
- 4 Tri-Perspective View Decomposition for Geometry-Aware Depth ✧ Completion, *Zhiqiang Yan, Yuankai Lin, Kun Wang, Yupeng Zheng, Yufei Wang, Zhenyu Zhang, Jun Li, Jian Yang*
- 5 Steerers: A Framework for Rotation Equivariant Keypoint ✧ Descriptors, *Georg Bökman, Johan Edstedt, Michael Felsberg, Fredrik Kahl*
- 6 VP3D: Unleashing 2D Visual Prompt for Text-to-3D Generation, *Yang Chen, Yingwei Pan, Haibo Yang, Ting Yao, Tao Mei*
- 7 Entangled View-Epipolar Information Aggregation for Generalizable Neural Radiance Fields, *Zhiyuan Min, Yawei Luo, Wei Yang, Yuesong Wang, Yi Yang*
- 8 GroupContrast: Semantic-aware Self-supervised Representation Learning for 3D Understanding, *Chengyao Wang, Li Jiang, Xiaoyang Wu, Zhuotao Tian, Bohao Peng, Hengshuang Zhao, Jiaya Jia*
- 9 iToF-flow-based High Frame Rate Depth Imaging, *Yu Meng, Zhou Xue, Xu Chang, Xuemei Hu, Tao Yue*

- 10 Generalizable Novel-View Synthesis using a Stereo Camera, *Haechan Lee, Wonjoon Jin, Seung-Hwan Baek, Sunghyun Cho*
- 11 EfficientDreamer: High-Fidelity and Robust 3D Creation via Orthogonal-view Diffusion Priors, *Zhipeng Hu, Minda Zhao, Chaoyi Zhao, Xinyue Liang, Lincheng Li, Zeng Zhao, Changjie Fan, Xiaowei Zhou, Xin Yu*
- 12 Leveraging Camera Triplets for Efficient and Accurate Structure-from-Motion, *Lalit Manam, Venu Madhav Govindu*
- 13 LAENeRF: Local Appearance Editing for Neural Radiance Fields, *Lukas Radl, Michael Steiner, Andreas Kurz, Markus Steinberger*
- 14 SuperPrimitive: Scene Reconstruction at a Primitive Level, *Kirill Mazur, Gwangbin Bae, Andrew J. Davison*
- 15 Revisiting Sampson Approximations for Geometric Estimation Problems, *Felix Rydell, Angélica Torres, Viktor Larsson*
- 16 Interactive3D: Create What You Want by Interactive 3D Generation, *Shaocong Dong, Lihe Ding, Zhanpeng Huang, Zibin Wang, Tianfan Xue, Dan Xu*
- 17 Multiplane Prior Guided Few-Shot Aerial Scene Rendering, *Zihan Gao, Licheng Jiao, Lingling Li, Xu Liu, Fang Liu, Puhua Chen, Yuwei Guo*
- 18 3DGS-Avatar: Animatable Avatars via Deformable 3D Gaussian Splatting, *Zhiyin Qian, Shaofei Wang, Marko Mihajlovic, Andreas Geiger, Siyu Tang*
- 19 DaReNeRF: Direction-aware Representation for Dynamic Scenes, *Ange Lou, Benjamin Planche, Zhongpai Gao, Yamin Li, Tianyu Luan, Hao Ding, Terrence Chen, Jack Noble, Ziyang Wu*
- 20 ViewDiff: 3D-Consistent Image Generation with Text-to-Image Models, *Lukas Höllein, Aljaž Božič, Norman Müller, David Novotny, Hung-Yu Tseng, Christian Richardt, Michael Zollhöfer, Matthias Nießner*
- 21 LTM: Lightweight Textured Mesh Extraction and Refinement of Large Unbounded Scenes for Efficient Storage and Real-time Rendering, *Jaehoon Choi, Rajvi Shah, Qinbo Li, Yipeng Wang, Ayush Saraf, Changil Kim, Jia-Bin Huang, Dinesh Manocha, Suhil Alsian, Johannes Kopf*
- 22 Minimal Perspective Autocalibration, *Andrea Porfiri Dal Cin, Timothy Duff, Luca Magri, Tomas Pajdla*
- 23 X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition, *Shuofeng Sun, Yongming Rao, Jiwen Lu, Haibin Yan*
- 24 2S-UDF: A Novel Two-stage UDF Learning Method for Robust Non-watertight Model Reconstruction from Multi-view Images, *Junkai Deng, Fei Hou, Xuhui Chen, Wencheng Wang, Ying He*
- 25 UFORecon: Generalizable Sparse-View Surface Reconstruction from Arbitrary and Unfavorable Sets, *Youngju Na, Woo Jae Kim, Kyu Beom Han, Suhyeon Ha, Sung-Eui Yoon*
- 26 GenN2N: Generative NeRF2NeRF Translation, *Xiangyue Liu, Han Xue, Kunming Luo, Ping Tan, Li Yi*
- 27 Text-to-3D Generation with Bidirectional Diffusion using both 2D and 3D priors, *Lihe Ding, Shaocong Dong, Zhanpeng Huang, Zibin Wang, Yiyuan Zhang, Kaixiong Gong, Dan Xu, Tianfan Xue*
- 28 Noisy One-point Homographies are Surprisingly Good, *Yaqing Ding, Jonathan Astermark, Magnus Oskarsson, Viktor Larsson*
- 29 Adaptive Multi-Modal Cross-Entropy Loss for Stereo Matching, *Peng Xu, Zhiyu Xiang, Chengyu Qiao, Jingyun Fu, Tianyu Pu*
- 30 LiDAR4D: Dynamic Neural Fields for Novel Space-time View LiDAR Synthesis, *Zehan Zheng, Fan Lu, Weiyi Xue, Guang Chen, Changjun Jiang*
- 31 NC-SDF: Enhancing Indoor Scene Reconstruction Using Neural SDFs with View-Dependent Normal Compensation, *Ziyi Chen, Xiaolong Wu, Yu Zhang*
- 32 VastGaussian: Vast 3D Gaussians for Large Scene Reconstruction, *Jiaqi Lin, Zhihao Li, Xiao Tang, Jianzhuang Liu, Shiyong Liu, Jiayue Liu, Yangdi Lu, Xiaofei Wu, Songcen Xu, Youliang Yan, Wenming Yang*
- 33 Language-driven Object Fusion into Neural Radiance Fields with Pose-Conditioned Dataset Updates, *Ka Chun Shum, Jaeyeon Kim, Binh-Son Hua, Duc Thanh Nguyen, Sai-Kit Yeung*
- 34 SPU-PMD: Self-Supervised Point Cloud Upsampling via Progressive Mesh Deformation, *Yanzhe Liu, Rong Chen, Yushi Li, Yixi Li, Xuehou Tan*
- 35 Intrinsic Image Diffusion for Indoor Single-view Material Estimation, *Peter Kocsis, Vincent Sitzmann, Matthias Nießner*

- 36 Learning Dynamic Tetrahedra for High-Quality Talking Head Synthesis, *Zicheng Zhang, Ruobing Zheng, Bonan Li, Congying Han, Tianqi Li, Meng Wang, Tiande Guo, Jingdong Chen, Ziwen Liu, Ming Yang*
- 37 Robust Self-calibration of Focal Lengths from the Fundamental Matrix, *Viktor Kocur, Daniel Kyselica, Zuzana Kukelova*
- 38 RNb-NeuS: Reflectance and Normal-based Multi-View 3D Reconstruction, *Baptiste Brument, Robin Bruneau, Yvain Quéau, Jean Mélou, François Bernard Lauze, Jean-Denis Durou, Lilian Calvet*
- 39 Neural 3D Strokes: Creating Stylized 3D Scenes with Vectorized 3D Strokes, *Hao-Bin Duan, Miao Wang, Yan-Xun Li, Yong-Liang Yang*
- 40 Unsupervised Template-assisted Point Cloud Shape Correspondence Network, *Jiacheng Deng, Jiahao Lu, Tianzhu Zhang*
- 41 Efficient Detection of Long Consistent Cycles and its Application to Distributed Synchronization, *Shaohan Li, Yunpeng Shi, Gilad Lerman*
- 42 AirPlanes: Accurate Plane Estimation via 3D-Consistent Embeddings, *Jamie Watson, Filippo Aleotti, Mohamed Sayed, Zawar Qureshi, Oisín Mac Aodha, Gabriel Brostow, Michael Firman, Sara Vicente*
- 43 Accurate Training Data for Occupancy Map Prediction in Automated Driving Using Evidence Theory, *Jonas Kälble, Sascha Wirges, Maxim Tatarchenko, Eddy Ilg*
- 44 Continuous Pose for Monocular Cameras in Neural Implicit Representation, *Qi Ma, Danda Pani Paudel, Ajad Chhatkuli, Luc Van Gool*
- 45 Towards 3D Vision with Low-Cost Single-Photon Cameras, *Fangzhou Mu, Carter Sifferman, Sacha Jungerman, Yiquan Li, Mark Han, Michael Gleicher, Mohit Gupta, Yin Li*
- 46 Inlier Confidence Calibration for Point Cloud Registration, *Yongzhe Yuan, Yue Wu, Xiaolong Fan, Maoguo Gong, Qiguang Miao, Wenping Ma*
- 47 GaussianShader: 3D Gaussian Splatting with Shading Functions for Reflective Surfaces, *Yingwenqi Jiang, Jiadong Tu, Yuan Liu, Xifeng Gao, Xiaoxiao Long, Wenping Wang, Yuexin Ma*
- 48 Language Embedded 3D Gaussians for Open-Vocabulary Scene Understanding, *Jin-Chuan Shi, Miao Wang, Hao-Bin Duan, Shao-Hua Guan*
- 49 MVIP-NeRF: Multi-view 3D Inpainting on NeRF Scenes via Diffusion Prior, *Honghua Chen, Chen Change Loy, Xingang Pan*
- 50 SuGaR: Surface-Aligned Gaussian Splatting for Efficient 3D Mesh Reconstruction and High-Quality Mesh Rendering, *Antoine Guédon, Vincent Lepetit*
- 51 DreamControl: Control-Based Text-to-3D Generation with 3D Self-Prior, *Tianyu Huang, Yihan Zeng, Zhilu Zhang, Wan Xu, Hang Xu, Songcen Xu, Rynson W.H. Lau, Wangmeng Zuo*
- 52 VAREN: Very Accurate and Realistic Equine Network, *Silvia Zuffi, Ylva Mellbin, Ci Li, Markus Hoeschle, Hedvig Kjellström, Senya Polikovsky, Elin Hernlund, Michael J. Black*
- 53 REACTO: Reconstructing Articulated Objects from a Single Video, *Chaoyue Song, Jiacheng Wei, Chuan Sheng Foo, Guosheng Lin, Fayao Liu*
- 54 DITTO: Dual and Integrated Latent Topologies for Implicit 3D Reconstruction, *Jaehyeok Shim, Kyungdon Joo*
- 55 ICON: Incremental Confidence for Joint Pose and Radiance Field Optimization, *Weiyao Wang, Pierre Gleize, Hao Tang, Xingyu Chen, Kevin J Liang, Matt Feiszli*
- 56 Local-consistent Transformation Learning for Rotation-invariant Point Cloud Analysis, *Yiyang Chen, Lunhao Duan, Shanshan Zhao, Changxing Ding, Dacheng Tao*
- 57 PaReNeRF: Toward Fast Large-scale Dynamic NeRF with Patch-based Reference, *Xiao Tang, Min Yang, Penghui Sun, Hui Li, Yuchao Dai, Feng Zhu, Hojae Lee*
- 58 Fitting Flats to Flats, *Gabriel Dogadov, Ugo Finnenhahl, Marc Alexa*
- 59 ANIM: Accurate Neural Implicit Model for Human Reconstruction from a single RGB-D Image, *Marco Pesavento, Yuanlu Xu, Nikolaos Sarafianos, Robert Maier, Ziyang Wang, Chun-Han Yao, Marco Volino, Edmond Boyer, Adrian Hilton, Tony Tung*
- 60 Neural Markov Random Field for Stereo Matching, *Tongfan Guan, Chen Wang, Yun-Hui Liu*
- 61 Improving Physics-Augmented Continuum Neural Radiance Field-Based Geometry-Agnostic System Identification with Lagrangian Particle Optimization, *Takuhiro Kaneko*
- 62 DiffusionAvatars: Deferred Diffusion for High-fidelity 3D Head Avatars, *Tobias Kirschstein, Simon Giebenhain, Matthias Nießner*
- 63 ViT-CoMer: Vision Transformer with Convolutional Multi-scale Feature Interaction for Dense Predictions, *Chunlong Xia, Xinliang Wang, Feng Lv, Xin Hao, Yifeng Shi*
- 64 Pose-Transformed Equivariant Network for 3D Point Trajectory Prediction, *Ruixuan Yu, Jian Sun*
- 65 UniReplKNet: A Universal Perception Large-Kernel ConvNet for Audio Video Point Cloud Time-Series and Image Recognition, *Xiaohan Ding, Yiyuan Zhang, Yixiao Ge, Sijie Zhao, Lin Song, Xiangyu Yue, Ying Shan*
- 66 KPConvX: Modernizing Kernel Point Convolution with Kernel Attention, *Hugues Thomas, Yao-Hung Hubert Tsai, Timothy D. Barfoot, Jian Zhang*
- 67 Time- Memory- and Parameter-Efficient Visual Adaptation, *Otniel-Bogdan Mercea, Alexey Gritsenko, Cordelia Schmid, Anurag Arnab*
- 68 Affine Equivariant Networks Based on Differential Invariants, *Yikang Li, Yeqing Qiu, Yuxuan Chen, Lingshen He, Zhouchen Lin*
- 69 PeLK: Parameter-efficient Large Kernel ConvNets with Peripheral Convolution, *Honghao Chen, Xiangxiang Chu, Yongjian Ren, Xin Zhao, Kaiqi Huang*
- 70 Making Vision Transformers Truly Shift-Equivariant, *Renan A. Rojas-Gomez, Teck-Yian Lim, Minh N. Do, Raymond A. Yeh*
- 71 Once for Both: Single Stage of Importance and Sparsity Search for Vision Transformer Compression, *Hancheng Ye, Chong Yu, Peng Ye, Renqiu Xia, Yansong Tang, Jiwen Lu, Tao Chen, Bo Zhang*
- 72 Data-Free Quantization via Pseudo-label Filtering, *Chunxiao Fan, Ziqi Wang, Dan Guo, Meng Wang*
- 73 FedHCA2: Towards Hetero-Client Federated Multi-Task Learning, *Yuxiang Lu, Suizhi Huang, Yuwen Yang, Shalayiding Sirejiding, Yue Ding, Hongtao Lu*
- 74 SpikingResformer: Bridging ResNet and Vision Transformer in Spiking Neural Networks, *Xinyu Shi, Zecheng Hao, Zhaofei Yu*
- 75 TetraSphere: A Neural Descriptor for O(3)-Invariant Point Cloud Analysis, *Pavlo Melnyk, Andreas Robinson, Michael Felsberg, Mårten Wadenbäck*
- 76 Friendly Sharpness-Aware Minimization, *Tao Li, Pan Zhou, Zhengbao He, Xinwen Cheng, Xiaolin Huang*
- 77 RMT: Retentive Networks Meet Vision Transformers, *Qihang Fan, Huaibo Huang, Mingrui Chen, Hongmin Liu, Ran He*
- 78 Efficient Deformable ConvNets: Rethinking Dynamic and Sparse Operator for Vision Applications, *Yuwen Xiong, Zhiqi Li, Yuntao Chen, Feng Wang, Xizhou Zhu, Jiapeng Luo, Wenhai Wang, Tong Lu, Hongsheng Li, Yu Qiao, Lewei Lu, Jie Zhou, Jifeng Dai*
- 79 Boosting Order-Preserving and Transferability for Neural Architecture Search: a Joint Architecture Refined Search and Fine-tuning Approach, *Beichen Zhang, Xiaoxing Wang, Xiaohan Qin, Junchi Yan*
- 80 Neural Redshift: Random Networks are not Random Functions, *Damien Teney, Armand Mihai Nicolicioiu, Valentin Hartmann, Ehsan Abbasnejad*
- 81 InceptionNeXt: When Inception Meets ConvNeXt, *Weihao Yu, Pan Zhou, Shuicheng Yan, Xinchao Wang*
- 82* Neural Lineage, *Runpeng Yu, Xinchao Wang*
- 83 BiPer: Binary Neural Networks using a Periodic Function, *Edwin Vargas, Claudia V. Correa, Carlos Hinojosa, Henry Arguello*
- 84 Rewrite the Stars, *Xu Ma, Xiyang Dai, Yue Bai, Yizhou Wang, Yun Fu*
- 85 A&B BNN: Add&Bit-Operation-Only Hardware-Friendly Binary Neural Network, *Ruichen Ma, Guanchao Qiao, Yian Liu, Liwei Meng, Ning Ning, Yang Liu, Shaogang Hu*
- 86 Neural Clustering based Visual Representation Learning, *Guikun Chen, Xia Li, Yi Yang, Wenguan Wang*
- 87 Building Optimal Neural Architectures using Interpretable Knowledge, *Keith G. Mills, Fred X. Han, Mohammad Salameh, Shengyao Lu, Chunhua Zhou, Jiao He, Fengyu Sun, Di Niu*
- 88 Towards More Accurate Diffusion Model Acceleration with A Timestep Tuner, *Mengfei Xia, Yujun Shen, Changsong Lei, Yu*

- Zhou, Deli Zhao, Ran Yi, Wenping Wang, Yong-Jin Liu
- 89 UniPTS: A Unified Framework for Proficient Post-Training Sparsity, *Jingjing Xie, Yuxin Zhang, Mingbao Lin, Liujuan Cao, Rongrong Ji*
- 90 Learning Structure-from-Motion with Graph Attention Networks, *Lucas Brynte, José Pedro Iglesias, Carl Olsson, Fredrik Kahl*
- 91 SHViT: Single-Head Vision Transformer with Memory Efficient Macro Design, *Seokju Yun, Youngmin Ro*
- 92 Denoising Point Clouds in Latent Space via Graph Convolution and Invertible Neural Network, *Aihua Mao, Biao Yan, Zijing Ma, Ying He*
- 93 JointSQ: Joint Sparsification-Quantization for Distributed Learning, *Weijing Xie, Haowei Li, Jitao Ma, Yunsong Li, Jie Lei, Donglai Liu, Leyuan Fang*
- 94 YoOOD: Utilizing Object Detection Concepts for Multi-Label Out-of-Distribution Detection, *Alon Zolfi, Guy Amit, Amit Baras, Satoru Koda, Ikuya Morikawa, Yuval Elovici, Asaf Shabtai*
- 95 RepAn: Enhanced Annealing through Re-parameterization, *Xiang Fei, Xiaowu Zheng, Yan Wang, Fei Chao, Chenglin Wu, Liujuan Cao*
- 96 D⁴: Dataset Distillation via Disentangled Diffusion Model, *Duo Su, Junjie Hou, Weizhi Gao, Yingjie Tian, Bowen Tang*
- 97 State Space Models for Event Cameras, *Nikola Zubic, Mathias Gehrig, Davide Scaramuzza*
- 98 Your Image is My Video: Reshaping the Receptive Field via Image-To-Video Differentiable AutoAugmentation and Fusion, *Sofia Casarin, Cynthia I. Ugwu, Sergio Escalera, Oswald Lanz*
- 99 Sparse Semi-DETR: Sparse Learnable Queries for Semi-Supervised Object Detection, *Tahira Shehzadi, Khurram Azeem Hashmi, Didier Stricker, Muhammad Zeshan Afzal*
- 100 MAPSeg: Unified Unsupervised Domain Adaptation for Heterogeneous Medical Image Segmentation Based on 3D Masked Autoencoding and Pseudo-Labeling, *Xuzhe Zhang, Yuhao Wu, Elsa Angelini, Ang Li, Jia Guo, Jerod M. Rasmussen, Thomas G. O'Connor, Pathik D. Wadhwa, Andrea Parolin Jackowski, Hai Li, Jonathan Posner, Andrew F. Laine, Yun Wang*
- 101 FedUV: Uniformity and Variance for Heterogeneous Federated Learning, *Ha Min Son, Moon-Hyun Kim, Tai-Myoung Chung, Chao Huang, Xin Liu*
- 102 Florence-2: Advancing a Unified Representation for a Variety of Vision Tasks, *Bin Xiao, Haiping Wu, Weijian Xu, Xiyang Dai, Houdong Hu, Yumao Lu, Michael Zeng, Ce Liu, Lu Yuan*
- 103 Pick-or-Mix: Dynamic Channel Sampling for ConvNets, *Ashish Kumar, Daneul Kim, Jaesik Park, Laxmidhar Behera*
- 104 Sheared Backpropagation for Fine-tuning Foundation Models, *Zhiyuan Yu, Li Shen, Liang Ding, Xinmei Tian, Yixin Chen, Dacheng Tao*
- 105 AZ-NAS: Assembling Zero-Cost Proxies for Network Architecture Search, *Junghyup Lee, Bumsub Ham*
- 106 MRFP: Learning Generalizable Semantic Segmentation from Sim-2-Real with Multi-Resolution Feature Perturbation, *Sumanth Udupa, Prajwal Gurunath, Aniruddh Sikdar, Suresh Sundaram*
- 107 Training-Free Pretrained Model Merging, *Zhengqi Xu, Ke Yuan, Huiqiong Wang, Yong Wang, Mingli Song, Jie Song*
- 108 Training Generative Image Super-Resolution Models by Wavelet-Domain Losses Enables Better Control of Artifacts, *Cansu Korkmaz, A. Murat Tekalp, Zafer Dogan*
- 109 IReNe: Instant Recoloring of Neural Radiance Fields, *Alessio Mazzucchelli, Adrian Garcia-Garcia, Elena Garces, Fernando Rivas-Manzanera, Francesc Moreno-Noguer, Adrian Penate-Sanchez*
- 110 AdaShift: Learning Discriminative Self-Gated Neural Feature Activation With an Adaptive Shift Factor, *Sudong Cai*
- 111 Kernel Adaptive Convolution for Scene Text Detection via Distance Map Prediction, *Jinzhong Zheng, Heng Fan, Libo Zhang*
- 112 Towards Accurate and Robust Architectures via Neural Architecture Search, *Yuwei Ou, Yuqi Feng, Yanan Sun*
- 113 PDF: A Probability-Driven Framework for Open World 3D Point Cloud Semantic Segmentation, *Jinfeng Xu, Siyuan Yang, Xianzhi Li, Yuan Tang, Yixue Hao, Long Hu, Min Chen*
- 114 Permutation Equivariance of Transformers and Its Applications, *Hengyuan Xu, Liyao Xiang, Hangyu Ye, Dixi Yao, Pengzhi Chu, Baochun Li*
- 115 MedBN: Robust Test-Time Adaptation against Malicious Test Samples, *Hyejin Park, Jeongyeon Hwang, Sunung Mun, Sangdon Park, Jungseul Ok*
- 116 Small Scale Data-Free Knowledge Distillation, *He Liu, Yikai Wang, Huaping Liu, Fuchun Sun, Anbang Yao*
- 117 Identifying Important Group of Pixels using Interactions, *Kosuke Sumiyasu, Kazuhiko Kawamoto, Hiroshi Kera*
- 118 Efficiently Assemble Normalization Layers and Regularization for Federated Domain Generalization, *Khiem Le, Long Ho, Cuong Do, Danh Le-Phuoc, Kok-Seng Wong*
- 119 OrthCaps: An Orthogonal CapsNet with Sparse Attention Routing and Pruning, *Xinyu Geng, Jiaming Wang, Jiawei Gong, Yuerong Xue, Jun Xu, Fanglin Chen, Xiaolin Huang*
- 120 Mean-Shift Feature Transformer, *Takumi Kobayashi*
- 121 You Only Need Less Attention at Each Stage in Vision Transformers, *Shuoxi Zhang, Hanpeng Liu, Stephen Lin, Kun He*
- 122 HEAL-SWIN: A Vision Transformer On The Sphere, *Oscar Carlsson, Jan E. Gerken, Hampus Linander, Heiner Spieß, Fredrik Ohlsson, Christoffer Petersson, Daniel Persson*
- 123 NC-TTT: A Noise Contrastive Approach for Test-Time Training, *David Osowiecki, Gustavo A. Vargas Hakim, Mehrdad Noori, Milad Cheraghalikhani, Ali Bahri, Moslem Yazdanpanah, Ismail Ben Ayed, Christian Desrosiers*
- 124 Unlocking the Potential of Prompt-Tuning in Bridging Generalized and Personalized Federated Learning, *Wenlong Deng, Christos Thrampoulidis, Xiaoxiao Li*
- 125 MR-VNet: Media Restoration using Volterra Networks, *Siddharth Roheda, Amit Unde, Loay Rashid*
- 126 Multimodal Pathway: Improve Transformers with Irrelevant Data from Other Modalities, *Yiyuan Zhang, Xiaohan Ding, Kaixiong Gong, Yixiao Ge, Ying Shan, Xiangyu Yue*
- 127 GreedyViG: Dynamic Axial Graph Construction for Efficient Vision GNNs, *Mustafa Munir, William Avery, Md Mostafijur Rahman, Radu Marculescu*
- 128 FlowerFormer: Empowering Neural Architecture Encoding using a Flow-aware Graph Transformer, *Dongyeon Hwang, Hyunju Kim, Sunwoo Kim, Kijung Shin*
- 129 Mixed-Precision Quantization for Federated Learning on Resource-Constrained Heterogeneous Devices, *Huan Cheng Chen, Haris Vikalo*
- 130 In Search of a Data Transformation That Accelerates Neural Field Training, *Junwon Seo, Sangyoon Lee, Kwang In Kim, Jaeho Lee*
- 131 Wired Perspectives: Multi-View Wire Art Embraces Generative AI, *Zhiyu Qu, Lan Yang, Honggang Zhang, Tao Xiang, Kaiyue Pang, Yi-Zhe Song*
- 132 DemoFusion: Democratizing High-Resolution Image Generation With No \$\$\$, *Ruoyi Du, Dongliang Chang, Timothy Hospedales, Yi-Zhe Song, Zhanyu Ma*
- 133 DiffPerformer: Iterative Learning of Consistent Latent Guidance for Diffusion-based Human Video Generation, *Chenyang Wang, Zerong Zheng, Tao Yu, Xiaoqian Lv, Bineng Zhong, Shengping Zhang, Liqiang Nie*
- 134 InteractDiffusion: Interaction Control in Text-to-Image Diffusion Models, *Jiun Tian Hoe, Xudong Jiang, Chee Seng Chan, Yap-Peng Tan, Weipeng Hu*
- 135 Intelligent Grimm - Open-ended Visual Storytelling via Latent Diffusion Models, *Chang Liu, Haoning Wu, Yujie Zhong, Xiaoyun Zhang, Yanfeng Wang, Weidi Xie*
- 136 ControlRoom3D: Room Generation using Semantic Proxy Rooms, *Jonas Schult, Sam Tsai, Lukas Höllein, Bichen Wu, Jialiang Wang, Chih-Yao Ma, Kunpeng Li, Xiaofang Wang, Felix Wimbauer, Zijian He, Peizhao Zhang, Bastian Leibe, Peter Vajda, Ji Hou*
- 137 Cache Me if You Can: Accelerating Diffusion Models through Block Caching, *Felix Wimbauer, Bichen Wu, Edgar Schoenfeld, Xiaoliang Dai, Ji Hou, Zijian He, Artsiom Sanakoyeu, Peizhao Zhang, Sam Tsai, Jonas Kohler, Christian Rupprecht, Daniel Cremers, Peter Vajda, Jialiang Wang*
- 138 Real-time 3D-aware Portrait Video Relighting, *Ziqi Cai, Kaiwen Jiang, Shu-Yu Chen, Yu-Kun Lai, Hongbo Fu, Boxin Shi, Lin Gao*

- 139 InstanceDiffusion: Instance-level Control for Image Generation, *Xudong Wang, Trevor Darrell, Sai Saketh Rambhatla, Rohit Girdhar, Ishan Misra*
- 140 Make-It-Vivid: Dressing Your Animatable Biped Cartoon Characters from Text, *Junshu Tang, Yanhong Zeng, Ke Fan, Xuheng Wang, Bo Dai, Kai Chen, Lizhuang Ma*
- 141 ZONE: Zero-Shot Instruction-Guided Local Editing, *Shanglin Li, Bohan Zeng, Yutang Feng, Sicheng Gao, Xiuhui Liu, Jiaming Liu, Lin Li, Xu Tang, Yao Hu, Jianzhuang Liu, Baochang Zhang*
- 142 Don't Drop Your Samples! Coherence-Aware Training Benefits
* Conditional Diffusion, *Nicolas Dufour, Victor Besnier, Vicky Kalogeiton, David Picard*
- 143 Generating Illustrated Instructions, *Sachit Menon, Ishan Misra, Rohit Girdhar*
- 144 SpikeNeRF: Learning Neural Radiance Fields from Continuous Spike Stream, *Lin Zhu, Kangmin Jia, Yifan Zhao, Yunshan Qi, Lizhi Wang, Hua Huang*
- 145 Dancing with Still Images: Video Distillation via Static-Dynamic Disentanglement, *Ziyu Wang, Yue Xu, Cewu Lu, Yong-Lu Li*
- 146 UniGS: Unified Representation for Image Generation and Segmentation, *Lu Qi, Lehan Yang, Weidong Guo, Yu Xu, Bo Du, Varun Jampani, Ming-Hsuan Yang*
- 147 Adversarial Text to Continuous Image Generation, *Kilichbek Haydarov, Aashiq Muhamed, Xiaoqian Shen, Jovana Lazarevic, Ivan Skorokhodov, Chamuditha Jayanga Galappaththige, Mohamed Elhoseiny*
- 148 Self-correcting LLM-controlled Diffusion Models, *Tsung-Han Wu, Long Lian, Joseph E. Gonzalez, Boyi Li, Trevor Darrell*
- 149 TiNO-Edit: Timestep and Noise Optimization for Robust Diffusion-Based Image Editing, *Sherry X Chen, Yaron Vaxman, Elad Ben Baruch, David Asulin, Aviad Moreshet, Kuo-Chin Lien, Misha Sra, Pradeep Sen*
- 150 Taming Stable Diffusion for Text to 360 Panorama Image Generation,
* *Cheng Zhang, Qianyi Wu, Camilo Cruz Gambardella, Xiaoshui Huang, Dinh Phung, Wanli Ouyang, Jianfei Cai*
- 151 EmoGen: Emotional Image Content Generation with Text-to-Image Diffusion Models, *Jingyuan Yang, Jiawei Feng, Hui Huang*
- 152 Carve3D: Improving Multi-view Reconstruction Consistency for Diffusion Models with RL Finetuning, *Desai Xie, Jiahao Li, Hao Tan, Xin Sun, Zhixin Shu, Yi Zhou, Sai Bi, Sören Pirk, Arie E. Kaufman*
- 153 FreeU: Free Lunch in Diffusion U-Net,
* *Chenyang Si, Ziqi Huang, Yuming Jiang, Ziwei Liu*
- 154 Move Anything with Layered Scene Diffusion, *Jiawei Ren, Mengmeng Xu, Jui-Chieh Wu, Ziwei Liu, Tao Xiang, Antoine Toisoul*
- 155 DiffAgent: Fast and Accurate Text-to-Image API Selection with Large Language Model, *Lirui Zhao, Yue Yang, Kaipeng Zhang, Wenqi Shao, Yuxin Zhang, Yu Qiao, Ping Luo, Rongrong Ji*
- 156 CapHuman: Capture Your Moments in Parallel Universes, *Chao Liang, Fan Ma, Linchao Zhu, Yingying Deng, Yi Yang*
- 157 IQ-VFI: Implicit Quadratic Motion Estimation for Video Frame Interpolation, *Mengshun Hu, Kui Jiang, Zhihang Zhong, Zheng Wang, Yinqiang Zheng*
- 158 Coarse-to-Fine Latent Diffusion for Pose-Guided Person Image
* Synthesis, *Yanzuo Lu, Manlin Zhang, Andy J Ma, Xiaohua Xie, Jianhuang Lai*
- 159 MACE: Mass Concept Erasure in Diffusion Models, *Shilin Lu, Zilan Wang, Leyang Li, Yanzhu Liu, Adams Wai-Kin Kong*
- 160 GenTron: Diffusion Transformers for Image and Video Generation, *Shoufa Chen, Mengmeng Xu, Jiawei Ren, Yuren Cong, Sen He, Yanping Xie, Animesh Sinha, Ping Luo, Tao Xiang, Juan-Manuel Perez-Rua*
- 161 Relightful Harmonization: Lighting-aware Portrait Background Replacement, *Mengwei Ren, Wei Xiong, Jae Shin Yoon, Zhixin Shu, Jianming Zhang, HyunJoon Jung, Guido Gerig, He Zhang*
- 162 InstructVideo: Instructing Video Diffusion Models with Human Feedback, *Hangjie Yuan, Shiwei Zhang, Xiang Wang, Yujie Wei, Tao Feng, Yining Pan, Yingya Zhang, Ziwei Liu, Samuel Albanie, Dong Ni*
- 163 SportsSloMo: A New Benchmark and Baselines for Human-centric Video Frame Interpolation, *Jiabao Chen, Huaizu Jiang*
- 164 TeTriRF: Temporal Tri-Plane Radiance Fields for Efficient Free-Viewpoint Video, *Minye Wu, Zehao Wang, Georgios Kouros, Tinne Tuytelaars*
- 165 SmartMask: Context Aware High-Fidelity Mask Generation for Fine-grained Object Insertion and Layout Control, *Jaskirat Singh, Jianming Zhang, Qing Liu, Cameron Smith, Zhe Lin, Liang Zheng*
- 166 RAVE: Randomized Noise Shuffling for Fast and Consistent Video
* Editing with Diffusion Models, *Ozgur Kara, Bariscan Kurtkaya, Hidir Yesiltepe, James M. Rehg, Pinar Yanardag*
- 167 LucidDreamer: Towards High-Fidelity Text-to-3D Generation via
* Interval Score Matching, *Yixun Liang, Xin Yang, Jiantao Lin, Haodong Li, Xiaogang Xu, Yingcong Chen*
- 168 HyperDreamBooth: HyperNetworks for Fast Personalization of Text-to-Image Models, *Nataniel Ruiz, Yuanzhen Li, Varun Jampani, Wei Wei, Tingbo Hou, Yael Pritch, Neal Wadhwa, Michael Rubinstein, Kfir Aberman*
- 169 DreamVideo: Composing Your Dream Videos with Customized Subject and Motion, *Yujie Wei, Shiwei Zhang, Zhiwu Qing, Hangjie Yuan, Zhiheng Liu, Yu Liu, Yingya Zhang, Jingren Zhou, Hongming Shan*
- 170 SurMo: Surface-based 4D Motion Modeling for Dynamic Human Rendering, *Tao Hu, Fangzhou Hong, Ziwei Liu*
- 171 Ranni: Taming Text-to-Image Diffusion for Accurate Instruction
* Following, *Yutong Feng, Biao Gong, Di Chen, Yujun Shen, Yu Liu, Jingren Zhou*
- 172 GenHowTo: Learning to Generate Actions and State Transformations from Instructional Videos, *Tomáš Souček, Dima Damen, Michael Wray, Ivan Laptev, Josef Sivic*
- 173 A Recipe for Scaling up Text-to-Video Generation with Text-free Videos, *Xiang Wang, Shiwei Zhang, Hangjie Yuan, Zhiwu Qing, Biao Gong, Yingya Zhang, Yujun Shen, Changxin Gao, Nong Sang*
- 174 WaveFace: Authentic Face Restoration with Efficient Frequency Recovery, *Yunqi Miao, Jiankang Deng, Jungong Han*
- 175 AnyDoor: Zero-shot Object-level Image Customization, *Xi Chen, Lianghua Huang, Yu Liu, Yujun Shen, Deli Zhao, Hengshuang Zhao*
- 176 ElasticDiffusion: Training-free Arbitrary Size Image Generation through Global-Local Content Separation, *Moayed Haji-Ali, Guha Balakrishnan, Vicente Ordonez*
- 177 One-step Diffusion with Distribution Matching Distillation, *Tianwei Yin, Michaël Gharbi, Richard Zhang, Eli Shechtman, Frédo Durand, William T. Freeman, Taesung Park*
- 178 Check Locate Rectify: A Training-Free Layout Calibration System for Text-to-Image Generation, *Biao Gong, Siteng Huang, Yutong Feng, Shiwei Zhang, Yuyuan Li, Yu Liu*
- 179 Hierarchical Spatio-temporal Decoupling for Text-to-Video Generation, *Zhiwu Qing, Shiwei Zhang, Jiayu Wang, Xiang Wang, Yujie Wei, Yingya Zhang, Changxin Gao, Nong Sang*
- 180 HumanGaussian: Text-Driven 3D Human Generation with Gaussian
* Splatting, *Xian Liu, Xiaohang Zhan, Jiayang Tang, Ying Shan, Gang Zeng, Dahua Lin, Xihui Liu, Ziwei Liu*
- 181 WonderJourney: Going from Anywhere to Everywhere, *Hong-Xing Yu, Haoyi Duan, Junhua Hur, Kyle Sargent, Michael Rubinstein, William T. Freeman, Forrester Cole, Deqing Sun, Noah Snively, Jiajun Wu, Charles Herrmann*
- 182 Balancing Act: Distribution-Guided Debiasing in Diffusion Models, *Rishubh Parihar, Abhijnya Bhat, Abhipsa Basu, Saswat Mallick, Jogendra Nath Kundu, R. Venkatesh Babu*
- 183 SIGNeRF: Scene Integrated Generation for Neural Radiance Fields, *Jan-Niklas Döhlmann, Andreas Engelhardt, Hendrik Lensch*
- 184 VideoBooth: Diffusion-based Video Generation with Image Prompts, *Yuming Jiang, Tianxing Wu, Shuai Yang, Chenyang Si, Dahua Lin, Yu Qiao, Chen Change Loy, Ziwei Liu*
- 185 Total Selfie: Generating Full-Body Selfies, *Bowei Chen, Brian Curless,
* Ira Kemelmacher-Shlizerman, Steven M. Seitz*
- 186 CCEdit: Creative and Controllable Video Editing via Diffusion Models, *Ruoyu Feng, Wenming Weng, Yanhui Wang, Yuhui Yuan, Jianmin Bao, Chong Luo, Zhibo Chen, Baining Guo*
- 187 Cinematic Behavior Transfer via NeRF-based Differentiable Filming, *Xuekun Jiang, Anyi Rao, Jingbo Wang, Dahua Lin, Bo Dai*
- 188 Improving Subject-Driven Image Synthesis with Subject-Agnostic Guidance, *Kelvin C.K. Chan, Yang Zhao, Xuhui Jia, Ming-Hsuan Yang, Huisheng Wang*
- 189 Drag Your Noise: Interactive Point-based Editing via Diffusion

- Semantic Propagation, *Haofeng Liu, Chenshu Xu, Yifei Yang, Lihua Zeng, Shengfeng He*
- 190 Learning Continuous 3D Words for Text-to-Image Generation, *Ta-Ying Cheng, Matheus Gadelha, Thibault Groueix, Matthew Fisher, Radomir Mech, Andrew Markham, Niki Triloni*
- 191 CHAIN: Enhancing Generalization in Data-Efficient GANs via Lipschitz continuity constrained Normalization, *Yao Ni, Piotr Koniusz*
- 192 ViVid-1-to-3: Novel View Synthesis with Video Diffusion Models, * *Jeong-gi Kwak, Erqun Dong, Yuhe Jin, Hanseok Ko, Shweta Mahajan, Kwang Moo Yi*
- 193 JeDi: Joint-Image Diffusion Models for Finetuning-Free Personalized Text-to-Image Generation, *Yu Zeng, Vishal M. Patel, Haochen Wang, Xun Huang, Ting-Chun Wang, Ming-Yu Liu, Yogesh Balaji*
- 194 GaussianDreamer: Fast Generation from Text to 3D Gaussians by Bridging 2D and 3D Diffusion Models, *Taoran Yi, Jiemin Fang, Junjie Wang, Guanjun Wu, Lingxi Xie, Xiaopeng Zhang, Wenyu Liu, Qi Tian, Xinggang Wang*
- 195 Prompting Hard or Hardly Prompting: Prompt Inversion for Text-to-Image Diffusion Models, *Shweta Mahajan, Tanzila Rahman, Kwang Moo Yi, Leonid Sigal*
- 196 MIGC: Multi-Instance Generation Controller for Text-to-Image * *Synthesis, Dewei Zhou, You Li, Fan Ma, Xiaoting Zhang, Yi Yang*
- 197 Towards Text-guided 3D Scene Composition, *Qihang Zhang, Chaoyang Wang, Aliaksandr Siarohin, Peiye Zhuang, Yinghao Xu, Ceyuan Yang, Dahua Lin, Bolei Zhou, Sergey Tulyakov, Hsin-Ying Lee*
- 198 BerfScene: Bev-conditioned Equivariant Radiance Fields for Infinite 3D Scene Generation, *Qihang Zhang, Yinghao Xu, Yujun Shen, Bo Dai, Bolei Zhou, Ceyuan Yang*
- 199 Face2Diffusion for Fast and Editable Face Personalization, *Kaede Shiohara, Toshihiko Yamasaki*
- 200 FreeDrag: Feature Dragging for Reliable Point-based Image Editing, *Pengyang Ling, Lin Chen, Pan Zhang, Huaian Chen, Yi Jin, Jinjin Zheng*
- 201 OmniLocalRF: Omnidirectional Local Radiance Fields from Dynamic Videos, *Dongyoung Choi, Hyeonjoong Jang, Min H. Kim*
- 202 DIRECT-3D: Learning Direct Text-to-3D Generation on Massive Noisy 3D Data, *Qihao Liu, Yi Zhang, Song Bai, Adam Kortylewski, Alan Yuille*
- 203 Generate Like Experts: Multi-Stage Font Generation by Incorporating Font Transfer Process into Diffusion Models, *Bin Fu, Fanghua Yu, Anran Liu, Zixuan Wang, Jie Wen, Junjun He, Yu Qiao*
- 204 Panacea: Panoramic and Controllable Video Generation for Autonomous Driving, *Yuqing Wen, Yucheng Zhao, Yingfei Liu, Fan Jia, Yanhui Wang, Chong Luo, Chi Zhang, Tiancai Wang, Xiaoyan Sun, Xiangyu Zhang*
- 205 360DVD: Controllable Panorama Video Generation with 360-Degree Video Diffusion Model, *Qian Wang, Weiqi Li, Chong Mou, Xinhua Cheng, Jian Zhang*
- 206 CLiC: Concept Learning in Context, *Mehdi Safaei, Aryan Mikaeili, * Or Patashnik, Daniel Cohen-Or, Ali Mahdavi-Amiri*
- 207 Z*: Zero-shot Style Transfer via Attention Reweighting, *Yingying Deng, Xiangyu He, Fan Tang, Weiming Dong*
- 208 Tackling the Singularities at the Endpoints of Time Intervals in * *Diffusion Models, Pengze Zhang, Hubery Yin, Chen Li, Xiaohua Xie*
- 209 CosmicMan: A Text-to-Image Foundation Model for Humans, * *Shikai Li, Jianglin Fu, Kaiyuan Liu, Wentao Wang, Kwan-Yee Lin, Wayne Wu*
- 210 Customize your NeRF: Adaptive Source Driven 3D Scene Editing via Local-Global Iterative Training, *Runze He, Shaofei Huang, Xuecheng Nie, Tianrui Hui, Luoqi Liu, Jiao Dai, Jizhong Han, Guanbin Li, Si Liu*
- 211 PICTURE: Photorealistic virtual Try-on from UnconstRAined dEsigns, *Shuliang Ning, Duomin Wang, Yipeng Qin, Zirong Jin, Baoyuan Wang, Xiaoguang Han*
- 212 Focus on Your Instruction: Fine-grained and Multi-instruction Image Editing by Attention Modulation, *Qin Guo, Tianwei Lin*
- 213 Make-Your-Anchor: A Diffusion-based 2D Avatar Generation Framework, *Ziyao Huang, Fan Tang, Yong Zhang, Xiaodong Cun, Juan Cao, Jintao Li, Tong-Yee Lee*
- 214 Revisiting Non-Autoregressive Transformers for Efficient Image Synthesis, *Zanlin Ni, Yulin Wang, Renping Zhou, Jiayi Guo, Jinyi Hu, Zhiyuan Liu, Shiji Song, Yuan Yao, Gao Huang*
- 215 Texture-Preserving Diffusion Models for High-Fidelity Virtual Try-On, *Xu Yang, Changxing Ding, Zhibin Hong, Junhao Huang, Jin Tao, Xiangmin Xu*
- 216 PromptCoT: Align Prompt Distribution via Adapted Chain-of-Thought, *Junyi Yao, Yijiang Liu, Zhen Dong, Mingfei Guo, Helan Hu, Kurt Keutzer, Li Du, Daquan Zhou, Shanghang Zhang*
- 217 Snap Video: Scaled Spatiotemporal Transformers for Text-to-Video * *Synthesis, Willi Menapace, Aliaksandr Siarohin, Ivan Skorokhodov, Ekaterina Deyneka, Tsai-Shien Chen, Anil Kag, Yuwei Fang, Aleksei Stoliar, Elisa Ricci, Jian Ren, Sergey Tulyakov*
- 218 L-MAGIC: Language Model Assisted Generation of Images with Coherence, *Zhipeng Cai, Matthias Mueller, Reiner Birkel, Diana Wofk, Shao-Yen Tseng, Junda Cheng, Gabriela Ben-Melech Stan, Vasudev Lai, Michael Paulitsch*
- 219 Text-Driven Image Editing via Learnable Regions, *Yuanze Lin, Yi-Wen Chen, Yi-Hsuan Tsai, Lu Jiang, Ming-Hsuan Yang*
- 220 On Exact Inversion of DPM-Solvers, *Seongmin Hong, Kyeonghyun Lee, Suh Yoon Jeon, Hyewon Bae, Se Young Chun*
- 221 Instruct-Imagen: Image Generation with Multi-modal Instruction, * *Hexiang Hu, Kelvin C.K. Chan, Yu-Chuan Su, Wenhu Chen, Yandong Li, Kihyuk Sohn, Yang Zhao, Xue Ben, Boqing Gong, William Cohen, Ming-Wei Chang, Xuhui Jia*
- 222 ConsistNet: Enforcing 3D Consistency for Multi-view Images Diffusion, *Jiayu Yang, Ziang Cheng, Yunfei Duan, Pan Ji, Hongdong Li*
- 223 LAMP: Learn A Motion Pattern for Few-Shot Video Generation, *Ruiqi Wu, Liangyu Chen, Tong Yang, Chunle Guo, Chongyi Li, Xiangyu Zhang*
- 224 Task-Customized Mixture of Adapters for General Image Fusion, *Pengfei Zhu, Yang Sun, Bing Cao, Qinghua Hu*
- 225 Beyond Textual Constraints: Learning Novel Diffusion Conditions with Fewer Examples, *Yuyang Yu, Bangzhen Liu, Chenxi Zheng, Xuemiao Xu, Huaidong Zhang, Shengfeng He*
- 226 Portrait4D: Learning One-Shot 4D Head Avatar Synthesis using Synthetic Data, *Yu Deng, Duomin Wang, Xiaohang Ren, Xingyu Chen, Baoyuan Wang*
- 227 Animating General Image with Large Visual Motion Model, *Dengsheng Chen, Xiaoming Wei, Xiaolin Wei*
- 228 Sat2Scene: 3D Urban Scene Generation from Satellite Images * *with Diffusion, Zuoyue Li, Zhenqiang Li, Zhaopeng Cui, Marc Pollefeys, Martin R. Oswald*
- 229 Seeing and Hearing: Open-domain Visual-Audio Generation with Diffusion Latent Aligners, *Yazhou Xing, Yingqing He, Zeyue Tian, Xintao Wang, Qifeng Chen*
- 230 AVID: Any-Length Video Inpainting with Diffusion Model, *Zhixing Zhang, Bichen Wu, Xiaoyan Wang, Yaqiao Luo, Luxin Zhang, Yinan Zhao, Peter Vajda, Dimitris Metaxas, Licheng Yu*
- 231 Generative Powers of Ten, *Xiaojuan Wang, Janne Kontkanen, * Brian Curless, Steven M. Seitz, Ira Kemelmacher-Shlizerman, Ben Mildenhall, Pratul Srinivasan, Dor Verbin, Aleksander Holynski*
- 232 DistriFusion: Distributed Parallel Inference for High-Resolution * *Diffusion Models, Muyang Li, Tianle Cai, Jiaxin Cao, Qinsheng Zhang, Han Cai, Junjie Bai, Yangqing Jia, Kai Li, Song Han*
- 233 Condition-Aware Neural Network for Controlled Image Generation, *Han Cai, Muyang Li, Qinsheng Zhang, Ming-Yu Liu, Song Han*
- 234 It's All About Your Sketch: Democratizing Sketch Control in Diffusion Models, *Subhadeep Koley, Ayan Kumar Bhunia, Deeptanshu Sekhri, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song*
- 235 FaceChain-SuDe: Building Derived Class to Inherit Category Attributes for One-shot Subject-Driven Generation, *Pengchong Qiao, Lei Shang, Chang Liu, Baigui Sun, Xiangyang Ji, Jie Chen*
- 236 In-N-Out: Faithful 3D GAN Inversion with Volumetric Decomposition for Face Editing, *Yiran Xu, Zhixin Shu, Cameron Smith, Seoung Wug Oh, Jia-Bin Huang*
- 237 Video Prediction by Modeling Videos as Continuous Multi-Dimensional Processes, *Gaurav Shrivastava, Abhinav Shrivastava*

- 238 DetDiffusion: Synergizing Generative and Perceptive Models for Enhanced Data Generation and Perception, *Yibo Wang, Ruiyuan Gao, Kai Chen, Kaiqiang Zhou, Yingjie Cai, Lanqing Hong, Zhenguo Li, Lihui Jiang, Di-Yan Yeung, Qiang Xu, Kai Zhang*
- 239 Structure-Guided Adversarial Training of Diffusion Models, *Ling Yang, Haotian Qian, Zhilong Zhang, Jingwei Liu, Bin Cui*
- 240 Learning Adaptive Spatial Coherent Correlations for Speech-
* Preserving Facial Expression Manipulation, *Tianshui Chen, Jianman Lin, Zhijing Yang, Chunmei Qing, Liang Lin*
- 241 On the Content Bias in Fréchet Video Distance, *Songwei Ge, Aniruddha Mahapatra, Gaurav Parmar, Jun-Yan Zhu, Jia-Bin Huang*
- 242 Residual Learning in Diffusion Models, *Junyu Zhang, Daochang Liu, **
Eunbyung Park, Shichao Zhang, Chang Xu
- 243 A Unified Approach for Text- and Image-guided 4D Scene Generation, *Yufeng Zheng, Xueting Li, Koki Nagano, Sifei Liu, Otmar Hilliges, Shalini De Mello*
- 244 VideoCrafter2: Overcoming Data Limitations for High-Quality Video Diffusion Models, *Haoxin Chen, Yong Zhang, Xiaodong Cun, Menghan Xia, Xintao Wang, Chao Weng, Ying Shan*
- 245 Neural Implicit Morphing of Face Images, *Guilherme Schardong, Tiago Novello, Hallison Paz, Iurii Medvedev, Vinicius da Silva, Luiz Velho, Nuno Gonçalves*
- 246 One More Step: A Versatile Plug-and-Play Module for Rectifying Diffusion Schedule Flaws and Enhancing Low-Frequency Controls, *Minghui Hu, Jianbin Zheng, Chuanxia Zheng, Chaoyue Wang, Dacheng Tao, Tat-Jen Cham*
- 247 Video Interpolation with Diffusion Models, *Siddhant Jain, Daniel Watson, Eric Tabellion, Aleksander Holynski, Ben Poole, Janne Kontkanen*
- 248 DiffSHEG: A Diffusion-Based Approach for Real-Time Speech-driven Holistic 3D Expression and Gesture Generation, *Junming Chen, Yunfei Liu, Jianan Wang, Ailing Zeng, Yu Li, Qifeng Chen*
- 249 TFMQ-DM: Temporal Feature Maintenance Quantization for
* Diffusion Models, *Yushi Huang, Ruihao Gong, Jing Liu, Tianlong Chen, Xianglong Liu*
- 250 Improving Training Efficiency of Diffusion Models via Multi-Stage Framework and Tailored Multi-Decoder Architecture, *Huijie Zhang, Yifu Lu, Ismail Alkhouri, Saiprasad Ravishankar, Dogyoon Song, Qing Qu*
- 251 Scaling Laws of Synthetic Images for Model Training ... for Now, *Lijie Fan, Kaifeng Chen, Dilip Krishnan, Dina Katabi, Phillip Isola, Yonglong Tian*
- 252 BIVDiff: A Training-Free Framework for General-Purpose Video Synthesis via Bridging Image and Video Diffusion Models, *Fengyuan Shi, Jiayi Gu, Hang Xu, Songcen Xu, Wei Zhang, Limin Wang*
- 253 MaskINT: Video Editing via Interpolative Non-autoregressive Masked Transformers, *Haoyu Ma, Shahin Mahdizadehaghdam, Bichen Wu, Zhipeng Fan, Yuchao Gu, Wenliang Zhao, Lior Shapira, Xiaohui Xie*
- 254 Pose Adapted Shape Learning for Large-Pose Face Reenactment, *Gee-Sern Jison Hsu, Jie-Ying Zhang, Huang Yu Hsiang, Wei-Jie Hong*
- 255 PRDP: Proximal Reward Difference Prediction for Large-Scale Reward Finetuning of Diffusion Models, *Fei Deng, Qifei Wang, Wei Wei, Tingbo Hou, Matthias Grundmann*
- 256 Discriminative Probing and Tuning for Text-to-Image Generation, *Leigang Qu, Wenjie Wang, Yongqi Li, Hanwang Zhang, Liqiang Nie, Tat-Seng Chua*
- 257 Towards Automated Movie Trailer Generation, *Dawit Mureja Argaw, Mattia Soldan, Alejandro Pardo, Chen Zhao, Fabian Caba Heilbron, Joon Son Chung, Bernard Ghanem*
- 258 CDFormer: When Degradation Prediction Embraces Diffusion Model for Blind Image Super-Resolution, *Qingguo Liu, Chenyi Zhuang, Pan Gao, Jie Qin*
- 259 FreeControl: Training-Free Spatial Control of Any Text-to-Image Diffusion Model with Any Condition, *Sicheng Mo, Fangzhou Mu, Kuan Heng Lin, Yanli Liu, Bochen Guan, Yin Li, Bolei Zhou*
- 260 RealCustom: Narrowing Real Text Word for Real-Time Open-Domain Text-to-Image Customization, *Mengqi Huang, Zhendong Mao, Mingcong Liu, Qian He, Yongdong Zhang*
- 261 VidToMe: Video Token Merging for Zero-Shot Video Editing, *Xirui Li, Chao Ma, Xiaokang Yang, Ming-Hsuan Yang*
- 262 Layout-Agnostic Scene Text Image Synthesis with Diffusion Models, *Qilong Zhangli, Jindong Jiang, Di Liu, Licheng Yu, Xiaoliang Dai, Ankit Ramchandani, Guan Pang, Dimitris N. Metaxas, Praveen Krishnan*
- 263 3D Multi-frame Fusion for Video Stabilization, *Zhan Peng, Xinyi Ye, Weiyue Zhao, Tianqi Liu, Huiqiang Sun, Baopu Li, Zhiguo Cao*
- 264 DyBLURF: Dynamic Neural Radiance Fields from Blurry Monocular Video, *Huiqiang Sun, Xingyi Li, Liao Shen, Xinyi Ye, Ke Xian, Zhiguo Cao*
- 265 A Video is Worth 256 Bases: Spatial-Temporal Expectation-Maximization Inversion for Zero-Shot Video Editing, *Maomao Li, Yu Li, Tianyu Yang, Yunfei Liu, Dongxu Yue, Zhihui Lin, Dong Xu*
- 266 StrokeFaceNeRF: Stroke-based Facial Appearance Editing in Neural Radiance Field, *Xiao-Juan Li, Dingxi Zhang, Shu-Yu Chen, Feng-Lin Liu*
- 267 Smooth Diffusion: Crafting Smooth Latent Spaces in Diffusion Models, *Jiayi Guo, Xingqian Xu, Yifan Pu, Zanlin Ni, Chaofei Wang, Manushree Vasu, Shiji Song, Gao Huang, Humphrey Shi*
- 268 One-dimensional Adapter to Rule Them All: Concepts Diffusion
* Models and Erasing Applications, *Mengyao Lyu, Yuhong Yang, Haiwen Hong, Hui Chen, Xuan Jin, Yuan He, Hui Xue, Jungong Han, Guiguang Ding*
- 269 Hierarchical Patch Diffusion Models for High-Resolution Video Generation, *Ivan Skorokhodov, Willi Menapace, Aliaksandr Siarohin, Sergey Tulyakov*
- 270 Taming the Tail in Class-Conditional GANs: Knowledge Sharing via Unconditional Training at Lower Resolutions, *Saeed Khorram, Mingqi Jiang, Mohamad Shahbazi, Mohamad H. Danesh, Li Fuxin*
- 271 Don't Look into the Dark: Latent Codes for Pluralistic Image Inpainting, *Haiwei Chen, Yajie Zhao*
- 272 Content-Style Decoupling for Unsupervised Makeup Transfer without Generating Pseudo Ground Truth, *Zhaoyang Sun, Shengwu Xiong, Yaxiong Chen, Yi Rong*
- 273 Generative Rendering: Controllable 4D-Guided Video Generation with 2D Diffusion Models, *Shenggu Cai, Duygu Ceylan, Matheus Gadelha, Chun-Hao Paul Huang, Tuanfeng Yang Wang, Gordon Wetzstein*
- 274 VideoSwap: Customized Video Subject Swapping with Interactive Semantic Point Correspondence, *Yuchao Gu, Yipin Zhou, Bichen Wu, Licheng Yu, Jia-Wei Liu, Rui Zhao, Jay Zhangjie Wu, David Junhao Zhang, Mike Zheng Shou, Kevin Tang*
- 275 Rethinking the Objectives of Vector-Quantized Tokenizers for Image Synthesis, *Yuchao Gu, Xintao Wang, Yixiao Ge, Ying Shan, Mike Zheng Shou*
- 276 Dysen-VDM: Empowering Dynamics-aware Text-to-Video Diffusion with LLMs, *Hao Fei, Shengqiong Wu, Wei Ji, Hanwang Zhang, Tat-Seng Chua*
- 277 Geometry-aware Reconstruction and Fusion-refined Rendering for Generalizable Neural Radiance Fields, *Tianqi Liu, Xinyi Ye, Min Shi, Zihao Huang, Zhiyu Pan, Zhan Peng, Zhiguo Cao*
- 278 DynVideo-E: Harnessing Dynamic NeRF for Large-Scale Motion- and View-Change Human-Centric Video Editing, *Jia-Wei Liu, Yan-Pei Cao, Jay Zhangjie Wu, Weijia Mao, Yuchao Gu, Rui Zhao, Jussi Keppo, Ying Shan, Mike Zheng Shou*
- 279 High-fidelity Person-centric Subject-to-Image Synthesis, *Yibin Wang, Weizhong Zhang, Jianwei Zheng, Cheng Jin*
- 280 Relation Rectification in Diffusion Model, *Yinwei Wu, Xingyi Yang, Xinchao Wang*
- 281 Diffusion Handles Enabling 3D Edits for Diffusion Models by
* Lifting Activations to 3D, *Karran Pandey, Paul Guerrero, Matheus Gadelha, Yannick Hold-Geoffroy, Karan Singh, Niloy J. Mitra*
- 282 LeftRefill: Filling Right Canvas based on Left Reference through Generalized Text-to-Image Diffusion Model, *Chenjie Cao, Yunuo Cai, Qiaole Dong, Yikai Wang, Yanwei Fu*
- 283 FSRT: Facial Scene Representation Transformer for Face Reenactment from Factorized Appearance Head-pose and Facial Expression Features, *Andre Rochow, Max Schwarz, Sven Behnke*
- 284 Tailored Visions: Enhancing Text-to-Image Generation with Personalized Prompt Rewriting, *Zijie Chen, Lichao Zhang, Fangsheng Weng, Lili Pan, Zhenzhong Lan*
- 285 MMA-Diffusion: MultiModal Attack on Diffusion Models, *Yijun*

- Yang, Ruiyuan Gao, Xiaosen Wang, Tsung-Yi Ho, Nan Xu, Qiang Xu
 286 PIA: Your Personalized Image Animator via Plug-and-Play Modules in Text-to-Image Models, *Yiming Zhang, Zhening Xing, Yanhong Zeng, Youqing Fang, Kai Chen*
- 287 Codebook Transfer with Part-of-Speech for Vector-Quantized Image Modeling, *Baoquan Zhang, Huaibin Wang, Chuyao Luo, Xutao Li, Guotao Liang, Yunming Ye, Xiao Chen Qi, Yao He*
- 288 Generating Non-Stationary Textures using Self-Rectification, *Yang Zhou, Rongjun Xiao, Dani Lischinski, Daniel Cohen-Or, Hui Huang*
- 289 Fast ODE-based Sampling for Diffusion Models in Around 5 Steps, * *Zhenyu Zhou, Defang Chen, Can Wang, Chun Chen*
- 290 Deformable One-shot Face Stylization via DINO Semantic Guidance, *Yang Zhou, Zichong Chen, Hui Huang*
- 291 Learning Disentangled Identifiers for Action-Customized Text-to-Image Generation, *Siteng Huang, Biao Gong, Yutong Feng, Xi Chen, Yuqian Fu, Yu Liu, Donglin Wang*
- 292 SwiftBrush: One-Step Text-to-Image Diffusion Model with Variational Score Distillation, *Thuan Hoang Nguyen, Anh Tran*
- 293 Towards Understanding Cross and Self-Attention in Stable Diffusion for Text-Guided Image Editing, *Bingyan Liu, Chengyu Wang, Tingfeng Cao, Kui Jia, Jun Huang*
- 294 SimDA: Simple Diffusion Adapter for Efficient Video Generation, *Zhen Xing, Qi Dai, Han Hu, Zuxuan Wu, Yu-Gang Jiang*
- 295 Unlocking Pre-trained Image Backbones for Semantic Image Synthesis, *Tariq Berrada Ifriqi, Jakob Verbeek, Camille Couprie, Karteek Alahari*
- 296 Shadow-Enlightened Image Outpainting, *Hang Yu, Ruilin Li, Shaorong Xie, Jiayan Qiu*
- 297 Exploiting Diffusion Prior for Generalizable Dense Prediction, *Hsin-Ying Lee, Hung-Yu Tseng, Hsin-Ying Lee, Ming-Hsuan Yang*
- 298 StyleCineGAN: Landscape Cinemagraph Generation using a Pre-trained StyleGAN, *Jongwoo Choi, Kwanggyoon Seo, Amirsaman Ashtari, Junyong Noh*
- 299 MotionEditor: Editing Video Motion via Content-Aware Diffusion, *Shuyuan Tu, Qi Dai, Zhi-Qi Cheng, Han Hu, Xintong Han, Zuxuan Wu, Yu-Gang Jiang*
- 300 DanceCamera3D: 3D Camera Movement Synthesis with Music and Dance, *Zixuan Wang, Jia Jia, Shikun Sun, Haozhe Wu, Rong Han, Zhenyu Li, Di Tang, Jiaqing Zhou, Jiebo Luo*
- 301 Diversity-aware Channel Pruning for StyleGAN Compression, *Jiwoo Chung, Sangeek Hyun, Sang-Heon Shim, Jae-Pil Heo*
- 302 DiffMorpher: Unleashing the Capability of Diffusion Models for Image Morphing, *Kaiwen Zhang, Yifan Zhou, Xudong Xu, Bo Dai, Xingang Pan*
- 303 StegoGAN: Leveraging Steganography for Non-Bijective Image-to-Image Translation, *Sidi Wu, Yizi Chen, Samuel Mermet, Lorenz Hurni, Konrad Schindler, Nicolas Gonthier, Loic Landrieu*
- 304 Grounded Text-to-Image Synthesis with Attention Refocusing, *Quynh Phung, Songwei Ge, Jia-Bin Huang*
- 305 VecFusion: Vector Font Generation with Diffusion, * *Vikas Thamizharasan, Difan Liu, Shantanu Agarwal, Matthew Fisher, Michael Gharbi, Oliver Wang, Alec Jacobson, Evangelos Kalogerakis*
- 306 Single Mesh Diffusion Models with Field Latents for Texture Generation, *Thomas W. Mitchel, Carlos Esteves, Ameesh Makadia*
- 307 Orthogonal Adaptation for Modular Customization of Diffusion Models, * *Ryan Po, Guandao Yang, Kfir Aberman, Gordon Wetzstein*
- 308 Low-Latency Neural Stereo Streaming, *Qiqi Hou, Farzad Farhadzadeh, Amir Said, Guillaume Sautiere, Hoang Le*
- 309 TextCraftor: Your Text Encoder Can be Image Quality Controller, *Yanyu Li, Xian Liu, Anil Kag, Ju Hu, Yerlan Idelbayev, Dhritiman Sagar, Yanzhi Wang, Sergey Tulyakov, Jian Ren*
- 310 4D-fy: Text-to-4D Generation Using Hybrid Score Distillation Sampling, *Sherwin Bahmani, Ivan Skorokhodov, Victor Rong, Gordon Wetzstein, Leonidas Guibas, Peter Wonka, Sergey Tulyakov, Jeong Joon Park, Andrea Tagliasacchi, David B. Lindell*
- 311 Image Neural Field Diffusion Models, * *Yinbo Chen, Oliver Wang, Richard Zhang, Eli Shechtman, Xiaolong Wang, Michael Gharbi*
- 312 Learning Multi-Dimensional Human Preference for Text-to-Image Generation, *Sixian Zhang, Bohan Wang, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, Zhongyuan Wang*
- 313 Dynamic Policy-Driven Adaptive Multi-Instance Learning for * Whole Slide Image Classification, *Tingting Zheng, Kui Jiang, Hongxun Yao*
- 314 Structure Matters: Tackling the Semantic Discrepancy in Diffusion Models for Image Inpainting, *Haipeng Liu, Yang Wang, Biao Qian, Meng Wang, Yong Rui*
- 315 IMPRINT: Generative Object Compositing by Learning Identity-Preserving Representation, *Yizhi Song, Zhifei Zhang, Zhe Lin, Scott Cohen, Brian Price, Jianming Zhang, Soo Ye Kim, He Zhang, Wei Xiong, Daniel Aliaga*
- 316 Puff-Net: Efficient Style Transfer with Pure Content and Style Feature Fusion Network, *Sizhe Zheng, Pan Gao, Peng Zhou, Jie Qin*
- 317 SSR-Encoder: Encoding Selective Subject Representation for Subject-Driven Generation, *Yuxuan Zhang, Yiren Song, Jiaming Liu, Rui Wang, Jinpeng Yu, Hao Tang, Huaxia Li, Xu Tang, Yao Hu, Han Pan, Zhongliang Jing*
- 318 PEEKABOO: Interactive Video Generation via Masked-Diffusion, *Yash Jain, Anshul Nasery, Vibhav Vineet, Harkirat Behl*
- 319 CoDef: Content Deformation Fields for Temporally Consistent * Video Processing, *Hao Ouyang, Qiuyu Wang, Yuxi Xiao, Qingyan Bai, Juntao Zhang, Kecheng Zheng, Xiaowei Zhou, Qifeng Chen, Yujun Shen*
- 320 DreamMatcher: Appearance Matching Self-Attention for Semantically-Consistent Text-to-Image Personalization, *Jisu Nam, Heesu Kim, Dongjae Lee, Siyoon Jin, Seungryong Kim, Seunggyu Chang*
- 321 DreamComposer: Controllable 3D Object Generation via Multi-View Conditions, *Yunhan Yang, Yukun Huang, Xiaoyang Wu, Yuan-Chen Guo, Song-Hai Zhang, Hengshuang Zhao, Tong He, Xihui Liu*
- 322 Shadow Generation for Composite Image Using Diffusion Model, *Qingyang Liu, Junqi You, Jianting Wang, Xinhao Tao, Bo Zhang, Li Niu*
- 323 Adversarial Score Distillation: When score distillation meets GAN, *Min Wei, Jingkai Zhou, Junyao Sun, Xuesong Zhang*
- 324 Uncertainty-Aware Source-Free Adaptive Image Super-Resolution with Wavelet Augmentation Transformer, *Yuang Ai, Xiaoqiang Zhou, Huaibo Huang, Lei Zhang, Ran He*
- 325 Animate Anyone: Consistent and Controllable Image-to-Video Synthesis for Character Animation, *Li Hu*
- 326 Person in Place: Generating Associative Skeleton-Guidance Maps for Human-Object Interaction Image Editing, *ChangHee Yang, ChanHee Kang, Kyeongbo Kong, Hanni Oh, Suk-Ju Kang*
- 327 StableVITON: Learning Semantic Correspondence with Latent Diffusion Model for Virtual Try-On, *Jeongho Kim, Guojung Gu, Minh Park, Sunghyun Park, Jaegul Choo*
- 328 Attention Calibration for Disentangled Text-to-Image Personalization, * *Yanbing Zhang, Mengping Yang, Qin Zhou, Zhe Wang*
- 329 Personalized Residuals for Concept-Driven Text-to-Image Generation, *Cusuh Ham, Matthew Fisher, James Hays, Nicholas Kolkin, Yuchen Liu, Richard Zhang, Tobias Hinz*
- 330 UFOGen: You Forward Once Large Scale Text-to-Image * Generation via Diffusion GANs, *Yanwu Xu, Yang Zhao, Zhisheng Xiao, Tingbo Hou*
- 331 FlowVid: Taming Imperfect Optical Flows for Consistent Video-to-Video Synthesis, *Feng Liang, Bichen Wu, Jialiang Wang, Licheng Yu, Kunpeng Li, Yinan Zhao, Ishan Misra, Jia-Bin Huang, Peizhao Zhang, Peter Vajda, Diana Marculescu*
- 332 Readout Guidance: Learning Control from Diffusion Features, * *Grace Luo, Trevor Darrell, Oliver Wang, Dan B Goldman, Aleksander Holynski*
- 333 Diffusion Model Alignment Using Direct Preference Optimization, *Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, Nikhil Naik*
- 334 Diffusion Models Without Attention, *Jing Nathan Yan, Jiatao Gu, Alexander M. Rush*

- 335 CommonCanvas: Open Diffusion Models Trained on Creative-Commons Images, *Aaron Gokaslan, A. Feder Cooper, Jasmine Collins, Landan Seguin, Austin Jacobson, Mihir Patel, Jonathan Frankle, Cory Stephenson, Volodymyr Kuleshov*
- 336 Fairy: Fast Parallelized Instruction-Guided Video-to-Video Synthesis, *Bichen Wu, Ching-Yao Chuang, Xiaoyan Wang, Yichen Jia, Kapil Krishnakumar, Tong Xiao, Feng Liang, Licheng Yu, Peter Vajda*
- 337 Edit One for All: Interactive Batch Image Editing, *Thao Nguyen, Utkarsh Ojha, Yuheng Li, Haotian Liu, Yong Jae Lee*
- 338 Wavelet-based Fourier Information Interaction with Frequency Diffusion Adjustment for Underwater Image Restoration, *Chen Zhao, Weiling Cai, Chenyu Dong, Chengwei Hu*
- 339 Accelerating Diffusion Sampling with Optimized Time Steps, *Shuchen Xue, Zhaoqiang Liu, Fei Chen, Shifeng Zhang, Tianyang Hu, Enze Xie, Zhenguo Li*
- 340 One-Shot Structure-Aware Stylized Image Synthesis, *Hansam Cho, Jonghyun Lee, Seunggyu Chang, Yonghyun Jeong*
- 341 Selectively Informative Description can Reduce Undesired Embedding Entanglements in Text-to-Image Personalization, *Jimyeong Kim, Jungwon Park, Wonjong Rhee*
- 342 Observation-Guided Diffusion Probabilistic Models, *Junoh Kang, Jinyoung Choi, Sungik Choi, Bohyung Han*
- 343 Scaling Up Video Summarization Pretraining with Large Language Models, *Dawit Mureja Argaw, Seunghyun Yoon, Fabian Caba Heilbron, Hanieh Deilamsalehy, Trung Bui, Zhaowen Wang, Franck Dernoncourt, Joon Son Chung*
- 344 DREAM: Diffusion Rectification and Estimation-Adaptive Models, *Jinxin Zhou, Tianyu Ding, Tianyi Chen, Jiachen Jiang, Ilya Zharkov, Zhihui Zhu, Luming Liang*
- 345 Clockwork Diffusion: Efficient Generation With Model-Step Distillation, ✱ *Amirhossein Habibi, Amir Ghodrati, Noor Fathima, Guillaume Sautiere, Risheek Garrepalli, Fatih Porikli, Jens Petersen*
- 346 SmartEdit: Exploring Complex Instruction-based Image Editing with Multimodal Large Language Models, ✱ *Yuzhou Huang, Liangbin Xie, Xintao Wang, Ziyang Yuan, Xiaodong Cun, Yixiao Ge, Jiantao Zhou, Chao Dong, Rui Huang, Ruimao Zhang, Ying Shan*
- 347 CAT-DM: Controllable Accelerated Virtual Try-on with Diffusion Model, *Jianhao Zeng, Dan Song, Weizhi Nie, Hongshuo Tian, Tongtong Wang, An-An Liu*
- 348 Exact Fusion via Feature Distribution Matching for Few-shot Image Generation, *Yingbo Zhou, Yutong Ye, Pengyu Zhang, Xian Wei, Mingsong Chen*
- 349 Cross Initialization for Face Personalization of Text-to-Image Models, *Lianyu Pang, Jian Yin, Haoran Xie, Qiping Wang, Qing Li, Xudong Mao*
- 350 EasyDrag: Efficient Point-based Manipulation on Diffusion Models, *Xingzhong Hou, Boxiao Liu, Yi Zhang, Jihao Liu, Yu Liu, Haihang You*
- 351 MicroCinema: A Divide-and-Conquer Approach for Text-to-Video Generation, ✱ *Yanhui Wang, Jianmin Bao, Wenming Weng, Ruoyu Feng, Dacheng Yin, Tao Yang, Jingxu Zhang, Qi Dai, Zhiyuan Zhao, Chunyu Wang, Kai Qiu, Yuhui Yuan, Xiaoyan Sun, Chong Luo, Baining Guo*
- 352 Towards Memorization-Free Diffusion Models, *Chen Chen, Daochang Liu, Chang Xu*
- 353 SD-DiT: Unleashing the Power of Self-supervised Discrimination in Diffusion Transformer, *Rui Zhu, Yingwei Pan, Yehao Li, Ting Yao, Zhenglong Sun, Tao Mei, Chang Wen Chen*
- 354 Towards Effective Usage of Human-Centric Priors in Diffusion Models for Text-based Human Image Generation, *Junyan Wang, Zhenhong Sun, Zhiyu Tan, Xuanbai Chen, Weihua Chen, Hao Li, Cheng Zhang, Yang Song*
- 355 Text2QR: Harmonizing Aesthetic Customization and Scanning Robustness for Text-Guided QR Code Generation, *Guangyang Wu, Xiaohong Liu, Jun Jia, Xuehao Cui, Guangtao Zhai*
- 356 Space-Time Diffusion Features for Zero-Shot Text-Driven Motion Transfer, *Danah Yatim, Rafail Fridman, Omer Bar-Tal, Yoni Kasten, Tali Dekel*
- 357 Video Frame Interpolation via Direct Synthesis with the Event-based Reference, *Yuhan Liu, Yongjian Deng, Hao Chen, Zhen Yang*
- 358 DiffEditor: Boosting Accuracy and Flexibility on Diffusion-based Image Editing, *Chong Mou, Xintao Wang, Jiechong Song, Ying Shan, Jian Zhang*
- 359 EMOPortraits: Emotion-enhanced Multimodal One-shot Head Avatars, *Nikita Drobyshev, Antoni Bigata Casademunt, Konstantinos Vougioukas, Zoe Landgraf, Stavros Petridis, Maja Pantic*
- 360 Spacetime Gaussian Feature Splatting for Real-Time Dynamic View Synthesis, *Zhan Li, Zhang Chen, Zhong Li, Yi Xu*
- 361 HOIDiffusion: Generating Realistic 3D Hand-Object Interaction Data, *Mengqi Zhang, Yang Fu, Zheng Ding, Sifei Liu, Zhuowen Tu, Xiaolong Wang*
- 362 Learned Representation-Guided Diffusion Models for Large-Image Generation, *Alexandros Graikos, Srikanth Yellapragada, Minh-Quan Le, Saarthak Kapse, Prateek Prasanna, Joel Saltz, Dimitris Samaras*
- 363 InstantBooth: Personalized Text-to-Image Generation without Test-Time Finetuning, *Jing Shi, Wei Xiong, Zhe Lin, Hyun Joon Jung*
- 364 TokenCompose: Text-to-Image Diffusion with Token-level Supervision, *Zirui Wang, Zhizhou Sha, Zheng Ding, Yilin Wang, Zhuowen Tu*
- 365 Geometry Transfer for Stylizing Radiance Fields, *Hyunyoung Jung, Seonghyeon Nam, Nikolaos Sarafianos, Sungjoo Yoo, Alexander Sorkine-Hornung, Rakesh Ranjan*
- 366 Align Your Gaussians: Text-to-4D with Dynamic 3D Gaussians and Composed Diffusion Models, *Huan Ling, Seung Wook Kim, Antonio Torralba, Sanja Fidler, Karsten Kreis*
- 367 DreamSalon: A Staged Diffusion Framework for Preserving Identity-Context in Editable Face Generation, *Haonan Lin*
- 368 Video-P2P: Video Editing with Cross-attention Control, *Shaoteng Liu, Yuechen Zhang, Wenbo Li, Zhe Lin, Jiaya Jia*
- 369 PAIR Diffusion: A Comprehensive Multimodal Object-Level Image Editor, *Vidit Goel, Elia Peruzzo, Yifan Jiang, Dejia Xu, Xingqian Xu, Nicu Sebe, Trevor Darrell, Zhangyang Wang, Humphrey Shi*
- 370 ArtAdapter: Text-to-Image Style Transfer using Multi-Level Style Encoder and Explicit Adaptation, *Dar-Yen Chen, Hamish Tennent, Ching-Wen Hsu*
- 371 DemoCaricature: Democratising Caricature Generation with a Rough Sketch, *Dar-Yen Chen, Ayan Kumar Bhunia, Subhadeep Koley, Aneeshan Sain, Pinaki Nath Chowdhury, Yi-Zhe Song*
- 372 PhotoMaker: Customizing Realistic Human Photos via Stacked ID Embedding, *Zhen Li, Mingdeng Cao, Xintao Wang, Zhongang Qi, Ming-Ming Cheng, Ying Shan*
- 373 Predicated Diffusion: Predicate Logic-Based Attention Guidance for Text-to-Image Diffusion Models, ✱ *Kota Sueyoshi, Takashi Matsubara*
- 374 SNED: Superposition Network Architecture Search for Efficient Video Diffusion Model, *Zhengang Li, Yan Kang, Yuchen Liu, Difan Liu, Tobias Hinz, Feng Liu, Yanzhi Wang*
- 375 TRIP: Temporal Residual Learning with Image Noise Prior for Image-to-Video Diffusion Models, *Zhongwei Zhang, Fuchen Long, Yingwei Pan, Zhaofan Qiu, Ting Yao, Yang Cao, Tao Mei*
- 376 Prompt-Free Diffusion: Taking "Text" out of Text-to-Image Diffusion Models, *Xingqian Xu, Jiayi Guo, Zhangyang Wang, Gao Huang, Irfan Essa, Humphrey Shi*
- 377 DEADiff: An Efficient Stylization Diffusion Model with ✱ *Disentangled Representations, Tianhao Qi, Shancheng Fang, Yanze Wu, Hongtao Xie, Jiawei Liu, Lang Chen, Qian He, Yongdong Zhang*
- 378 FRESCO: Spatial-Temporal Correspondence for Zero-Shot Video Translation, *Shuai Yang, Yifan Zhou, Ziwei Liu, Chen Change Loy*
- 379 Correcting Diffusion Generation through Resampling, ✱ *Yujian Liu, Yang Zhang, Tommi Jaakkola, Shiyu Chang*
- 380 AnyScene: Customized Image Synthesis with Composed Foreground, *Ruidong Chen, Lanjun Wang, Weizhi Nie, Yongdong Zhang, An-An Liu*
- 381 Grid Diffusion Models for Text-to-Video Generation, *Taegyeong Lee, Soyeong Kwon, Taehwan Kim*
- 382 Direct2.5: Diverse Text-to-3D Generation via Multi-view 2.5D Diffusion, *Yuanxun Lu, Jingyang Zhang, Shiwei Li, Tian Fang, David McKinnon, Yanghai Tsin, Long Quan, Xun Cao, Yao Yao*
- 383 Anomaly Score: Evaluating Generative Models and Individual Generated Images based on Complexity and Vulnerability, *Jaehui Hwang, Junghyuk Lee, Jong-Seok Lee*

- 384 Style Aligned Image Generation via Shared Attention,
* Amir Hertz, Andrey Voynov, Shlomi Fruchter, Daniel Cohen-Or
- 385 Zero-Painter: Training-Free Layout Control for Text-to-Image Synthesis, Marianna Ohanyan, Hayk Manukyan, Zhangyang Wang, Shant Navasardyan, Humphrey Shi
- 386 X-Adapter: Adding Universal Compatibility of Plugins for Upgraded Diffusion Model, Lingmin Ran, Xiaodong Cun, Jia-Wei Liu, Rui Zhao, Song Zijie, Xintao Wang, Jussi Keppo, Mike Zheng Shou
- 387 Neural Point Cloud Diffusion for Disentangled 3D Shape and Appearance Generation, Philipp Schröppel, Christopher Wewer, Jan Eric Lenssen, Eddy Ilg, Thomas Brox
- 388 Style Injection in Diffusion: A Training-free Approach for Adapting
* Large-scale Diffusion Models for Style Transfer, Jiwoo Chung, Sangeek Hyun, Jae-Pil Heo
- 389 Vlogger: Make Your Dream A Vlog, Shaobin Zhuang, Kunchang Li, Xinyuan Chen, Yaohui Wang, Ziwei Liu, Yu Qiao, Yali Wang
- 390 Faces that Speak: Jointly Synthesizing Talking Face and Speech from Text, Youngjoon Jang, Ji-Hoon Kim, Junseok Ahn, Doyeop Kwak, Hong-Sun Yang, Yoon-Cheol Ju, Il-Hwan Kim, Byeong-Yeol Kim, Joon Son Chung
- 391 Prompt Augmentation for Self-supervised Text-guided Image Manipulation, Rumeysa Bodur, Binod Bhattarai, Tae-Kyun Kim
- 392 DragDiffusion: Harnessing Diffusion Models for Interactive Point-based Image Editing, Yujun Shi, Chuhui Xue, Jun Hao Liew, Jiachun Pan, Hanshu Yan, Wenqing Zhang, Vincent Y. F. Tan, Song Bai
- 393 Make Pixels Dance: High-Dynamic Video Generation, Yan Zeng, Guoqiang Wei, Jiani Zheng, Jiaxin Zou, Yang Wei, Yuchen Zhang, Hang Li
- 394 LEDITS++: Limitless Image Editing using Text-to-Image Models, Manuel Brack, Felix Friedrich, Katharia Kornmeier, Linoy Tsaban, Patrick Schramowski, Kristian Kersting, Apolinario Passos
- 395 Emu Edit: Precise Image Editing via Recognition and Generation
* Tasks, Shelly Sheynin, Adam Polyak, Uriel Singer, Yuval Kirstain, Amit Zohar, Oron Ashual, Devi Parikh, Yaniv Taigman
- 396 Concept Weaver: Enabling Multi-Concept Fusion in Text-to-Image Models, Gihyun Kwon, Simon Jenni, Dingzeyu Li, Joon-Young Lee, Jong Chul Ye, Fabian Caba Heilbron
- 397 ACT-Diffusion: Efficient Adversarial Consistency Training for One-step Diffusion Models, Fei Kong, Jinhao Duan, Lichao Sun, Hao Cheng, Renjing Xu, Hengtao Shen, Xiaofeng Zhu, Xiaoshuang Shi, Kaidi Xu
- 398 3D Geometry-Aware Deformable Gaussian Splatting for Dynamic View Synthesis, Zhicheng Lu, Xiang Guo, Le Hui, Tianrui Chen, Min Yang, Xiao Tang, Feng Zhu, Yuchao Dai
- 399 Boosting Diffusion Models with Moving Average Sampling in Frequency Domain, Yurui Qian, Qi Cai, Yingwei Pan, Yehao Li, Ting Yao, Qibin Sun, Tao Mei
- 400 NoiseCollage: A Layout-Aware Text-to-Image Diffusion Model Based on Noise Cropping and Merging, Takahiro Shirakawa, Seiichi Uchida
- 401 NeRF On-the-go: Exploiting Uncertainty for Distractor-free NeRFs in the Wild, Weining Ren, Zihan Zhu, Boyang Sun, Jiaqi Chen, Marc Pollefeys, Songyou Peng
- 402 Using Human Feedback to Fine-tune Diffusion Models without Any Reward Model, Kai Yang, Jian Tao, Jiafei Lyu, Chunjiang Ge, Jiaxin Chen, Weihan Shen, Xiaolong Zhu, Xiu Li
- 403 GeneAvatar: Generic Expression-Aware Volumetric Head Avatar Editing from a Single Image, Chong Bao, Yinda Zhang, Yuan Li, Xiyu Zhang, Bangbang Yang, Hujun Bao, Marc Pollefeys, Guofeng Zhang, Zhaopeng Cui
- 404 MaskPLAN: Masked Generative Layout Planning from Partial Input, Hang Zhang, Anton Savov, Benjamin Dillenburger
- 405 WOUAF: Weight Modulation for User Attribution and Fingerprinting in Text-to-Image Diffusion Models, Changhoon Kim, Kyle Min, Maitreya Patel, Sheng Cheng, Yezhou Yang
- 406 Transcending Forgery Specificity with Latent Space Augmentation for Generalizable Deepfake Detection, Zhiyuan Yan, Yuhao Luo, Siwei Lyu, Qingshan Liu, Baoyuan Wu
- 407 SCedit: Efficient and Controllable Image Diffusion Generation via
* Skip Connection Editing, Zeyinzi Jiang, Chaojie Mao, Yulin Pan, Zhen Han, Jingfeng Zhang
- 408 CONFORM: Contrast is All You Need for High-Fidelity Text-to-Image Diffusion Models, Tuna Han Salih Meral, Enis Simsar, Federico Tombari, Pinar Yanardag
- 409 TI2V-Zero: Zero-Shot Image Conditioning for Text-to-Video Diffusion Models, Haomiao Ni, Bernhard Egger, Suhas Lohit, Anoop Cherian, Ye Wang, Toshiaki Koike-Akino, Sharon X. Huang, Tim K. Marks
- 410 HIVE: Harnessing Human Feedback for Instructional Visual Editing, Shu Zhang, Xinyi Yang, Yihao Feng, Can Qin, Chia-Chih Chen, Ning Yu, Zeyuan Chen, Huan Wang, Silvio Savarese, Stefano Ermon, Caiming Xiong, Ran Xu
- 411 Taming Mode Collapse in Score Distillation for Text-to-3D Generation, Peihao Wang, Dejia Xu, Zhiwen Fan, Dilin Wang, Sreyas Mohan, Forrest Iandola, Rakesh Ranjan, Yilei Li, Qiang Liu, Zhangyang Wang, Vikas Chandra
- 412 CoDi: Conditional Diffusion Distillation for Higher-Fidelity and Faster Image Generation, Kangfu Mei, Mauricio Delbracio, Hossein Talebi, Zhengzhong Tu, Vishal M. Patel, Peyman Milanfar
- 413 Universal Robustness via Median Randomized Smoothing for Real-World Super-Resolution, Zakariya Chaouai, Mohamed Tamaazousti
- 414 ECLIPSE: A Resource-Efficient Text-to-Image Prior for Image Generations, Maitreya Patel, Changhoon Kim, Sheng Cheng, Chitta Baral, Yezhou Yang
- 415 CAMEL: CAusal Motion Enhancement Tailored for Lifting Text-driven Video Editing, Guiwei Zhang, Tianyu Zhang, Guanglin Niu, Zichang Tan, Yalong Bai, Qing Yang
- 416 FreeCustom: Tuning-Free Customized Image Generation for Multi-Concept Composition, Ganggui Ding, Canyu Zhao, Wen Wang, Zhen Yang, Zide Liu, Hao Chen, Chunhua Shen
- 417 Amodal Completion via Progressive Mixed Context Diffusion,
* Katherine Xu, Lingzhi Zhang, Jianbo Shi
- 418 Named Entity Driven Zero-Shot Image Manipulation, Zhida Feng, Li Chen, Jing Tian, JiaXiang Liu, Shikun Feng
- 419 Learning Degradation-unaware Representation with Prior-based Latent Transformations for Blind Face Restoration, Lianxin Xie, Csbingbing Zheng, Wen Xue, Le Jiang, Cheng Liu, Si Wu, Hau San Wong
- 420 AEROBLADE: Training-Free Detection of Latent Diffusion Images Using Autoencoder Reconstruction Error, Jonas Ricker, Denis Lukovnikov, Asja Fischer
- 421 VRetouchEr: Learning Cross-frame Feature Interdependence with Imperfection Flow for Face Retouching in Videos, Wen Xue, Le Jiang, Lianxin Xie, Si Wu, Yong Xu, Hau San Wong
- 422 Generative Unlearning for Any Identity, Juwon Seo, Sung-Hoon Lee, Tae-Young Lee, Seungjun Moon, Gyeong-Moon Park
- 423 Doubly Abductive Counterfactual Inference for Text-based Image Editing, Xue Song, Jiequan Cui, Hanwang Zhang, Jingjing Chen, Richang Hong, Yu-Gang Jiang
- 424 Text-conditional Attribute Alignment across Latent Spaces for 3D Controllable Face Image Synthesis, Feifan Xu, Rui Li, Si Wu, Yong Xu, Hau San Wong
- 425 Customization Assistant for Text-to-Image Generation, Yufan Zhou, Ruiyi Zhang, Jiuxiang Gu, Tong Sun
- 426 Contrastive Denoising Score for Text-guided Latent Diffusion Image Editing, Hyelin Nam, Gihyun Kwon, Geon Yeong Park, Jong Chul Ye
- 427 Arbitrary-Scale Image Generation and Upsampling using Latent Diffusion Model and Implicit Neural Decoder, Jinseok Kim, Tae-Kyun Kim
- 428 VMC: Video Motion Customization using Temporal Attention Adaption for Text-to-Video Diffusion Models, Hyeonho Jeong, Geon Yeong Park, Jong Chul Ye
- 429 Visual Layout Composer: Image-Vector Dual Diffusion Model for Design Layout Generation, Mohammad Amin Shabani, Zhaowen Wang, Difan Liu, Nanxuan Zhao, Jimei Yang, Yasutaka Furukawa
- 430 Learning Spatial Adaptation and Temporal Coherence in Diffusion Models for Video Super-Resolution, Zhikai Chen, Fuchen Long, Zhaofan Qiu, Ting Yao, Wengang Zhou, Jiebo Luo, Tao Mei
- 431 Open-Vocabulary Attention Maps with Token Optimization for Semantic Segmentation in Diffusion Models, Pablo Marcos-Manchón, Roberto Alcover-Couso, Juan C. SanMiguel, José M. Martínez

Thursday, June 20

- 7:30 - 16:00 Registration / Badge Pickup** (Summit Lobby)
7:00 - 17:00 Press Room (Summit 340)
7:00 - 17:00 Mother's Room (Summit 341-adjacent and Summit 441-adjacent)
7:00 - 17:00 Prayer or Quiet Room (Summit 326)
7:30 - 9:00 Breakfast (Summit ExHall 1-2)
8:30 - 10:30 SOCIAL - How to Know Your True Market Value as an AI Researcher (Summit Terrace Suite)
 **Reservation Required

- 9:00 - 10:30 Orals 3A: 3D from Single View** (Summit Ballroom)
- 1 Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation, *Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, Konrad Schindler*
 - 2 EscherNet: A Generative Model for Scalable View Synthesis, *Xin Kong, Shikun Liu, Xiaoyang Lyu, Marwan Taher, Xiaojuan Qi, Andrew J. Davison*
 - 3 WALT3D: Generating Realistic Training Data from Time-Lapse Imagery for Reconstructing Dynamic Objects Under Occlusion, *Khiem Vuong, N Dinesh Reddy, Robert Tamburo, Srinivasa G. Narasimhan*
 - 4 Diffusion-FOF: Single-View Clothed Human Reconstruction via Diffusion-Based Fourier Occupancy Field, *Yuanzhen Li, Fei Luo, Chunxia Xiao*
 - 5 Rethinking Inductive Biases for Surface Normal Estimation, *Gwangbin Bae, Andrew J. Davison*

- 9:00 - 10:30 Orals 3B: Vision, Language, and Reasoning** (Summit Flex Hall AB)
- 1 Comparing the Decision-Making Mechanisms by Transformers and CNNs via Explanation Methods, *Mingqi Jiang, Saeed Khorram, Li Fuxin*
 - 2 MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI, *Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhui Chen*
 - 3 Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs, *Shengbang Tong, Zhuang Liu, Yuexiang Zhai, Yi Ma, Yann LeCun, Saining Xie*
 - 4 LISA: Reasoning Segmentation via Large Language Model, *Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhui Yuan, Shu Liu, Jiaya Jia*
 - 5 Visual Program Distillation: Distilling Tools and Programmatic Reasoning into Vision-Language Models, *Yushi Hu, Otilia Stretcu, Chun-Ta Lu, Krishnamurthy Viswanathan, Kenji Hata, Enming Luo, Ranjay Krishna, Ariel Fuxman*

- 9:00 - 10:30 Orals 3C: Medical and Physics-Based Vision** (Summit Flex Hall C)
- 1 EventPS: Real-Time Photometric Stereo Using an Event Camera, *Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, Boxin Shi*
 - 2 EvDiG: Event-guided Direct and Global Components Separation, *Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, Boxin Shi*
 - 3 MemSAM: Taming Segment Anything Model for Echocardiography Video Segmentation, *Xiaolong Deng, Huisi Wu, Runhao Zeng, Jing Qin*
 - 4 Transcriptomics-guided Slide Representation Learning in Computational Pathology, *Guillaume Jaume, Lukas Oldenburg, Anurag Vaidya, Richard J. Chen, Drew F.K. Williamson, Thomas Peeters, Andrew H. Song, Faisal Mahmood*
 - 5 Correlation-aware Coarse-to-fine MLPs for Deformable Medical Image Registration, *Mingyuan Meng, Dagan Feng, Lei Bi, Jinman Kim*

- 10:00 - 10:30 Poster Setup** (Arch 4E)

- 10:30 - 11:30 EXPO TRACK** Andrea Gagliano
 Senior Director, AI/ML at Getty Images (Arch 4F)

- 10:30 - 12:00 Poster Session 3 & Exhibit Hall** (Arch 4A-E)
- 1 G3DR: Generative 3D Reconstruction in ImageNet, *Pradyumna Reddy, Ismail Elezi, Jiankang Deng*
 - 2 CityDreamer: Compositional Generative Model of Unbounded 3D Cities, *Haozhe Xie, Zhaoxi Chen, Fangzhou Hong, Ziwei Liu*
 - 3 6D-Diff: A Keypoint Diffusion Framework for 6D Object Pose Estimation, *Li Xu, Haoxuan Qu, Yujun Cai, Jun Liu*
 - 4 Generative Proxemics: A Prior for 3D Social Interaction from Images, *Lea Müller, Vickie Ye, Georgios Pavlakos, Michael Black, Angjoo Kanazawa*
 - 5 Repurposing Diffusion-Based Image Generators for Monocular Depth Estimation, *Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, Konrad Schindler*
 - 6 MVD-Fusion: Single-view 3D via Depth-consistent Multi-view Generation, *Hanzhe Hu, Zhizhuo Zhou, Varun Jampani, Shubham Tulsiani*
 - 7 WorDepth: Variational Language Prior for Monocular Depth Estimation, *Ziyao Zeng, Daniel Wang, Fengyu Yang, Hyungseob Park, Stefano Soatto, Dong Lao, Alex Wong*
 - 8 Free3D: Consistent Novel View Synthesis without 3D Representation, *Chuanxia Zheng, Andrea Vedaldi*
 - 9 PostureHMR: Posture Transformation for 3D Human Mesh Recovery, *Yu-Pei Song, Xiao Wu, Zhaoquan Yuan, Jian-Jun Qiao, Qiang Peng*
 - 10 3DFIRES: Few Image 3D REconstruction for Scenes with Hidden Surfaces, *Linyi Jin, Nilesh Kulkarni, David F. Fouhey*
 - 11 Learning the 3D Fauna of the Web, *Zizhang Li, Dor Litvak, Ruining Li, Yunzhi Zhang, Tomas Jakab, Christian Rupprecht, Shangzhe Wu, Andrea Vedaldi, Jiajun Wu*
 - 12 Bilateral Propagation Network for Depth Completion, *Jie Tang, Fei-Peng Tian, Boshi An, Jian Li, Ping Tan*
 - 13 Efficient Privacy-Preserving Visual Localization Using 3D Ray Clouds, *Heejoon Moon, Chunghwan Lee, Je Hyeong Hong*
 - 14 EpiDiff: Enhancing Multi-View Synthesis via Localized Epipolar-Constrained Diffusion, *Zehuan Huang, Hao Wen, Junting Dong, Yaohui Wang, Yangguang Li, Xinyuan Chen, Yan-Pei Cao, Ding Liang, Yu Qiao, Bo Dai, Lu Sheng*
 - 15 Doodle Your 3D: From Abstract Freehand Sketches to Precise 3D Shapes, *Hmrishav Bandyopadhyay, Subhadeep Koley, Ayan Das, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song*
 - 16 LowRankOcc: Tensor Decomposition and Low-Rank Recovery for Vision-based 3D Semantic Occupancy Prediction, *Linqing Zhao, Xiuwei Xu, Ziwei Wang, Yunpeng Zhang, Borui Zhang, Wenzhao Zheng, Dalong Du, Jie Zhou, Jiwen Lu*
 - 17 CNC-Net: Self-Supervised Learning for CNC Machining Operations, *Mohsen Yavartanoo, Sangmin Hong, Reyhaneh Neshatavar, Kyoung Mu Lee*
 - 18 Reconstructing Hands in 3D with Transformers, *Georgios Pavlakos, Dandan Shan, Ilija Radosavovic, Angjoo Kanazawa, David Fouhey, Jitendra Malik*
 - 19 Boosting Self-Supervision for Single-View Scene Completion via Knowledge Distillation, *Keonhee Han, Dominik Muhle, Felix Wimbauer, Daniel Cremers*
 - 20 Know Your Neighbors: Improving Single-View Reconstruction via Spatial Vision-Language Reasoning, *Rui Li, Tobias Fischer, Mattia Segu, Marc Pollefeys, Luc Van Gool, Federico Tombari*
 - 21 Depth Prompting for Sensor-Agnostic Depth Estimation, *Jin-Hwi Park, Chanhwi Jeong, Junoh Lee, Hae-Gon Jeon*
 - 22 ViewFusion: Towards Multi-View Consistency via Interpolated Denoising, *Xianghui Yang, Yan Zuo, Sameera Ramasinghe, Loris Bazzani, Gil Avraham, Anton van den Hengel*
 - 23 Slice3D: Multi-Slice Occlusion-Revealing Single View 3D Reconstruction, *Yizhi Wang, Wallace Lira, Wenqi Wang, Ali Mahdavi-Amiri, Hao Zhang*
 - 24 Consistent3D: Towards Consistent High-Fidelity Text-to-3D Generation with Deterministic Sampling Prior, *Zike Wu, Pan Zhou, Xuanyu Yi, Xiaoding Yuan, Hanwang Zhang*

- 25 GigaPose: Fast and Robust Novel Object Pose Estimation via One Correspondence, *Van Nguyen Nguyen, Thibault Groueix, Mathieu Salzmann, Vincent Lepetit*
- 26 RichDreamer: A Generalizable Normal-Depth Diffusion Model for
* Detail Richness in Text-to-3D, *Lingteng Qiu, Guanying Chen, Xiaodong Gu, Qi Zuo, Mutian Xu, Yushuang Wu, Weihao Yuan, Zilong Dong, Liefeng Bo, Xiaoguang Han*
- 27 Elite360D: Towards Efficient 360 Depth Estimation via Semantic- and Distance-Aware Bi-Projection Fusion, *Hao Ai, Lin Wang*
- 28 SIFU: Side-view Conditioned Implicit Function for Real-world
* Usable Clothed Human Reconstruction, *Zechuan Zhang, Zongxin Yang, Yi Yang*
- 29 Diffusion Time-step Curriculum for One Image to 3D Generation, *Xuanyu Yi, Zike Wu, Qingshan Xu, Pan Zhou, Joo-Hwee Lim, Hanwang Zhang*
- 30 SecondPose: SE(3)-Consistent Dual-Stream Feature Fusion for Category-Level Pose Estimation, *Yamei Chen, Yan Di, Guangyao Zhai, Fabian Manhardt, Chenyangguang Zhang, Ruida Zhang, Federico Tombari, Nassir Navab, Benjamin Busam*
- 31 Wonder3D: Single Image to 3D using Cross-Domain Diffusion,
* *Xiaoxiao Long, Yuan-Chen Guo, Cheng Lin, Yuan Liu, Zhiyang Dou, Lingjie Liu, Yuexin Ma, Song-Hai Zhang, Marc Habermann, Christian Theobalt, Wenping Wang*
- 32 En3D: An Enhanced Generative Model for Sculpting 3D Humans from 2D Synthetic Data, *Yifang Men, Biwen Lei, Yuan Yao, Miaomiao Cui, Zhouhui Lian, Xuansong Xie*
- 33 MOHO: Learning Single-view Hand-held Object Reconstruction with Multi-view Occlusion-Aware Supervision, *Chenyanguang Zhang, Guanlong Jiao, Yan Di, Gu Wang, Ziqin Huang, Ruida Zhang, Fabian Manhardt, Bowen Fu, Federico Tombari, Xiangyang Ji*
- 34 Template Free Reconstruction of Human-object Interaction with
* Procedural Interaction Generation, *Xianghui Xie, Bharat Lal Bhatnagar, Jan Eric Lenssen, Gerard Pons-Moll*
- 35 PatchFusion: An End-to-End Tile-Based Framework for High-Resolution Monocular Metric Depth Estimation, *Zhenyu Li, Shariq Farooq Bhat, Peter Wonka*
- 36 SPAD: Spatially Aware Multi-View Diffusers, *Yash Kant, Aliaksandr Siarohin, Ziyi Wu, Michael Vasilkovsky, Guocheng Qian, Jian Ren, Riza Alp Guler, Bernard Ghanem, Sergey Tulyakov, Igor Gilitschenski*
- 37 GenFlow: Generalizable Recurrent Flow for 6D Pose Refinement of Novel Objects, *Sunghill Moon, Hyeontae Son, Dongcheol Hur, Sangwook Kim*
- 38 PointInfinity: Resolution-Invariant Point Diffusion Models, *Zixuan Huang, Justin Johnson, Shoubhik Debnath, James M. Rehg, Chao-Yuan Wu*
- 39 ZeroShape: Regression-based Zero-shot Shape Reconstruction, *Zixuan Huang, Stefan Stojanov, Anh Thai, Varun Jampani, James M. Rehg*
- 40 One-2-3-45++: Fast Single Image to 3D Objects with Consistent Multi-View Generation and 3D Diffusion, *Minghua Liu, Ruoxi Shi, Linghao Chen, Zhuoyang Zhang, Chao Xu, Xinyue Wei, Hansheng Chen, Chong Zeng, Jiayuan Gu, Hao Su*
- 41 ConTex-Human: Free-View Rendering of Human from a Single Image with Texture-Consistent Synthesis, *Xiangjun Gao, Xiaoyu Li, Chaopeng Zhang, Qi Zhang, Yanpei Cao, Ying Shan, Long Quan*
- 42 MatchU: Matching Unseen Objects for 6D Pose Estimation from RGB-D Images, *Junwen Huang, Hao Yu, Kuan-Ting Yu, Nassir Navab, Slobodan Ilic, Benjamin Busam*
- 43 UniDepth: Universal Monocular Metric Depth Estimation,
* *Luigi Piccinelli, Yung-Hsu Yang, Christos Sakaridis, Mattia Segu, Siyuan Li, Luc Van Gool, Fisher Yu*
- 44 G-NeRF: Geometry-enhanced Novel View Synthesis from Single-View Images, *Zixiong Huang, Qi Chen, Libo Sun, Yifan Yang, Naizhou Wang, Qi Wu, Mingkui Tan*
- 45 3DToonify: Creating Your High-Fidelity 3D Stylized Avatar Easily from 2D Portrait Images, *Yifang Men, Hanxi Liu, Yuan Yao, Miaomiao Cui, Xuansong Xie, Zhouhui Lian*
- 46 Adaptive Fusion of Single-View and Multi-View Depth for Autonomous Driving, *Junda Cheng, Wei Yin, Kaixuan Wang, Xiaozhi Chen, Shijie Wang, Xin Yang*
- 47 HiPose: Hierarchical Binary Surface Encoding and Correspondence Pruning for RGB-D 6DoF Object Pose Estimation, *Yongliang Lin, Yongzhi Su, Praveen Nathan, Sandeep Inuganti, Yan Di, Martin Sundermeyer, Fabian Manhardt, Didier Stricker, Jason Rambach, Yu Zhang*
- 48 HandBooster: Boosting 3D Hand-Mesh Reconstruction by Conditional Synthesis and Sampling of Hand-Object Interactions, *Hao Xu, Haipeng Li, Yingqiao Wang, Shuaicheng Liu, Chi-Wing Fu*
- 49 3D-SceneDreamer: Text-Driven 3D-Consistent Scene Generation, *Songchun Zhang, Yibo Zhang, Quan Zheng, Rui Ma, Wei Hua, Hujun Bao, Weiwei Xu, Changqing Zou*
- 50 NViST: In the Wild New View Synthesis from a Single Image with Transformers, *Wonbong Jang, Lourdes Agapito*
- 51 CAD: Photorealistic 3D Generation via Adversarial Distillation, *Ziyu Wan, Despoina Paschalidou, Ian Huang, Hongyu Liu, Bokui Shen, Xiaoyu Xiang, Jing Liao, Leonidas Guibas*
- 52 Splatter Image: Ultra-Fast Single-View 3D Reconstruction, *Stanislaw Szymanowicz, Christian Rupprecht, Andrea Vedaldi*
- 53 Joint Reconstruction of 3D Human and Object via Contact-Based Refinement Transformer, *Hyeongjin Nam, Daniel Sungho Jung, Gyeongsik Moon, Kyoung Mu Lee*
- 54 Sculpt3D: Multi-View Consistent Text-to-3D Generation with Sparse 3D Prior, *Cheng Chen, Xiaofeng Yang, Fan Yang, Chengzeng Feng, Zhoujie Fu, Chuan-Sheng Foo, Guosheng Lin, Fayao Liu*
- 55 Object Pose Estimation via the Aggregation of Diffusion Features,
* *Tianfu Wang, Guosheng Hu, Hongguang Wang*
- 56 MonoCD: Monocular 3D Object Detection with Complementary Depths, *Longfei Yan, Pei Yan, Shengzhou Xiong, Xuanyu Xiang, Yihua Tan*
- 57 MultiDiff: Consistent Novel View Synthesis from a Single Image, *Norman Müller, Katja Schwarz, Barbara Rössle, Lorenzo Porzi, Samuel Rota Bulò, Matthias Nießner, Peter Kotschieder*
- 58 SeaBird: Segmentation in Bird's View with Dice Loss Improves Monocular 3D Detection of Large Objects, *Abhinav Kumar, Yuliang Guo, Xinyu Huang, Liu Ren, Xiaoming Liu*
- 59 Learning Occupancy for Monocular 3D Object Detection, *Liang Peng, Junkai Xu, Haoran Cheng, Zheng Yang, Xiaopei Wu, Wei Qian, Wenxiao Wang, Boxi Wu, Deng Cai*
- 60 NeRFDeformer: NeRF Transformation from a Single View via 3D Scene Flows, *Zhenggang Tang, Zhongzheng Ren, Xiaoming Zhao, Bowen Wen, Jonathan Tremblay, Stan Birchfield, Alexander Schwing*
- 61 R-Cyclic Diffuser: Reductive and Cyclic Latent Diffusion for 3D Clothed Human Digitalization, *Kennard Yanting Chan, Fayao Liu, Guosheng Lin, Chuan Sheng Foo, Weisi Lin*
- 62 Unleashing Network Potentials for Semantic Scene Completion, *Fengyun Wang, Qianru Sun, Dong Zhang, Jinhui Tang*
- 63 Triplane Meets Gaussian Splatting: Fast and Generalizable Single-View 3D Reconstruction with Transformers, *Zi-Xin Zou, Zhipeng Yu, Yuan-Chen Guo, Yangguang Li, Ding Liang, Yan-Pei Cao, Song-Hai Zhang*
- 64 VOODOO 3D: Volumetric Portrait Disentanglement For One-Shot 3D Head Reenactment, *Phong Tran, Egor Zakharov, Long-Nhat Ho, Anh Tuan Tran, Liwen Hu, Hao Li*
- 65 Compressed 3D Gaussian Splatting for Accelerated Novel View Synthesis, *Simon Niedermayr, Josef Stumpfegger, Rüdiger Westermann*
- 66 Morphable Diffusion: 3D-Consistent Diffusion for Single-image Avatar Creation, *Xiyi Chen, Marko Mihajlovic, Shaofei Wang, Sergey Prokudin, Siyu Tang*
- 67 Depth Anything: Unleashing the Power of Large-Scale Unlabeled Data, *Lihe Yang, Bingyi Kang, Zilong Huang, Xiaogang Xu, Jiashi Feng, Hengshuang Zhao*
- 68 SAOR: Single-View Articulated Object Reconstruction, *Mehmet Aygun, Oisin Mac Aodha*
- 69 EscherNet: A Generative Model for Scalable View Synthesis,
* *Xin Kong, Shikun Liu, Xiaoyang Lyu, Marwan Taher, Xiaojuan Qi, Andrew J. Davison*
- 70 HOISDF: Constraining 3D Hand-Object Pose Estimation with Global Signed Distance Fields, *Haozhe Qi, Chen Zhao, Mathieu Salzmann, Alexander Mathis*
- 71 Diffusion-driven GAN Inversion for Multi-Modal Face Image

- 72 Generation, *Jihyun Kim, Changjae Oh, Hoseok Do, Soohyun Kim, Kwanghoon Sohn*
- 72 Novel View Synthesis with View-Dependent Effects from a Single Image, *Juan Luis Gonzalez Bello, Munchurl Kim*
- 73 Weakly-Supervised Emotion Transition Learning for Diverse 3D Co-speech Gesture Generation, *Xingqun Qi, Jiahao Pan, Peng Li, Ruibin Yuan, Xiaowei Chi, Mengfei Li, Wenhan Luo, Wei Xue, Shanghang Zhang, Qifeng Liu, Yike Guo*
- 74 Seg2Reg: Differentiable 2D Segmentation to 1D Regression Rendering for 360 Room Layout Reconstruction, *Cheng Sun, Wei-En Tai, Yu-Lin Shih, Kuan-Wei Chen, Yong-Jing Syu, Kent Selwyn The, Yu-Chiang Frank Wang, Hwann-Tzong Chen*
- 75 Mining Supervision for Dynamic Regions in Self-Supervised Monocular Depth Estimation, *Hoang Chuong Nguyen, Tianyu Wang, Jose M. Alvarez, Miaomiao Liu*
- 76 WALT3D: Generating Realistic Training Data from Time-Lapse
* Imagery for Reconstructing Dynamic Objects Under Occlusion, *Khiem Vuong, N Dinesh Reddy, Robert Tamburo, Srinivasa G. Narasimhan*
- 77 DiffPortrait3D: Controllable Diffusion for Zero-Shot Portrait View
* Synthesis, *Yuming Gu, Hongyi Xu, You Xie, Guoxian Song, Yichun Shi, Di Chang, Jing Yang, Linjie Luo*
- 78 3D-LFM: Lifting Foundation Model, *Mosam Dabhi, László A. Jeni, Simon Lucey*
- 79 MRC-Net: 6-DoF Pose Estimation with MultiScale Residual Correlation, *Yuelong Li, Yafei Mao, Raja Bala, Sunil Hadap*
- 80 DiffusionGAN3D: Boosting Text-guided 3D Generation and Domain Adaptation by Combining 3D GANs and Diffusion Priors, *Biwen Lei, Kai Yu, Mengyang Feng, Miaomiao Cui, Xuansong Xie*
- 81 VS: Reconstructing Clothed 3D Human from Single Image via Vertex Shift, *Leyuan Liu, Yuhan Li, Yunqi Gao, Changxin Gao, Yuanyuan Liu, Jingying Chen*
- 82 Weakly Supervised Monocular 3D Detection with a Single-View Image, *Xueying Jiang, Sheng Jin, Lewei Lu, Xiaoqin Zhang, Shijian Lu*
- 83 From-Ground-To-Objects: Coarse-to-Fine Self-supervised Monocular Depth Estimation of Dynamic Objects with Ground Contact Prior, *Jaeho Moon, Juan Luis Gonzalez Bello, Byeongjun Kwon, Munchurl Kim*
- 84 Gated Fields: Learning Scene Reconstruction from Gated Videos, *Andrea Ramazzina, Stefanie Walz, Pragyan Dahal, Mario Bijelic, Felix Heide*
- 85 SCINeRF: Neural Radiance Fields from a Snapshot Compressive
* Image, *Yunhao Li, Xiaodong Wang, Ping Wang, Xin Yuan, Peidong Liu*
- 86 Diffusion-FOF: Single-View Clothed Human Reconstruction via
* Diffusion-Based Fourier Occupancy Field, *Yuanzhen Li, Fei Luo, Chunxia Xiao*
- 87 Instance-aware Contrastive Learning for Occluded Human Mesh Reconstruction, *Mi-Gyeong Gwon, Gi-Mun Um, Won-Sik Cheong, Wonjun Kim*
- 88 IBD-SLAM: Learning Image-Based Depth Fusion for Generalizable SLAM, *Minghao Yin, Shangzhe Wu, Kai Han*
- 89 HarmonyView: Harmonizing Consistency and Diversity in One-Image-to-3D, *Sangmin Woo, Byeongjun Park, Hyojun Go, Jin-Young Kim, Changick Kim*
- 90 UV-IDM: Identity-Conditioned Latent Diffusion Model for Face UV-Texture Generation, *Hong Li, Yutang Feng, Song Xue, Xuhui Liu, Bohan Zeng, Shanglin Li, Boyu Liu, Jianzhuang Liu, Shumin Han, Baochang Zhang*
- 91 AttriHuman-3D: Editable 3D Human Avatar Generation with Attribute Decomposition and Indexing, *Fan Yang, Tianyi Chen, Xiaosheng He, Zhongang Cai, Lei Yang, Si Wu, Guosheng Lin*
- 92 Mind The Edge: Refining Depth Edges in Sparsely-Supervised Monocular Depth Estimation, *Lior Talker, Aviad Cohen, Erez Yosef, Alexandra Dana, Michael Dinerstein*
- 93 3DiffTecton: 3D Object Detection with Geometry-Aware Diffusion Features, *Chenfeng Xu, Huan Ling, Sanja Fidler, Or Litany*
- 94 Bayesian Diffusion Models for 3D Shape Reconstruction, *Haiyang Xu, Yu Lei, Zeyuan Chen, Xiang Zhang, Yue Zhao, Yilin Wang, Zhuowen Tu*
- 95 Rethinking Inductive Biases for Surface Normal Estimation,
* *Gwangbin Bae, Andrew J. Davison*
- 96 LaneCPP: Continuous 3D Lane Detection using Physical Priors, *Maximilian Pittner, Joel Janai, Alexandra P. Condurache*
- 97 Enhancing 3D Fidelity of Text-to-3D using Cross-View Correspondences, *Seungwook Kim, Kejie Li, Xueqing Deng, Yichun Shi, Minsu Cho, Peng Wang*
- 98 MonoDiff: Monocular 3D Object Detection and Pose Estimation with Diffusion Models, *Yasiru Ranasinghe, Deepti Hegde, Vishal M. Patel*
- 99 HiLo: Detailed and Robust 3D Clothed Human Reconstruction with High-and Low-Frequency Information of Parametric Models, *Yifan Yang, Dong Liu, Shuhai Zhang, Zeshuai Deng, Zixiong Huang, Mingkui Tan*
- 100 MPOD123: One Image to 3D Content Generation Using Mask-enhanced Progressive Outline-to-Detail Optimization, *Jimin Xu, Tianbao Wang, Tao Jin, Shengyu Zhang, Dongjie Fu, Zhe Wang, Jiangjing Lyu, Chengfei Lv, Chaoyue Niu, Zhou Yu, Zhou Zhao, Fei Wu*
- 101 GeoReF: Geometric Alignment Across Shape Variation for Category-level Object Pose Refinement, *Linfang Zheng, Tze Ho Elden Tse, Chen Wang, Yinghan Sun, Hua Chen, Ales Leonardis, Wei Zhang, Hyung Jin Chang*
- 102 Unsupervised 3D Structure Inference from Category-Specific Image Collections, *Weikang Wang, Dongliang Cao, Florian Bernard*
- 103 Neural Parametric Gaussians for Monocular Non-Rigid Object Reconstruction, *Devikalyan Das, Christopher Wewer, Raza Yunus, Eddy Ilg, Jan Eric Lenssen*
- 104 BiTT: Bi-directional Texture Reconstruction of Interacting Two Hands from a Single Image, *Minje Kim, Tae-Kyun Kim*
- 105 DeCoTR: Enhancing Depth Completion with 2D and 3D Attentions, *Yunxiao Shi, Manish Kumar Singh, Hong Cai, Fatih Porikli*
- 106 MonoNPHM: Dynamic Head Reconstruction from Monocular
* Videos, *Simon Giebenhain, Tobias Kirschstein, Markos Georgopoulos, Martin Rünz, Lourdes Agapito, Matthias Nießner*
- 107 FakeInversion: Learning to Detect Images from Unseen Text-to-Image Models by Inverting Stable Diffusion, *George Cazenavette, Avneesh Sud, Thomas Leung, Ben Usman*
- 108 Forgery-aware Adaptive Transformer for Generalizable Synthetic Image Detection, *Huan Liu, Zichang Tan, Chuangchuang Tan, Yunchao Wei, Jingdong Wang, Yao Zhao*
- 109 Towards Modern Image Manipulation Localization: A Large-Scale Dataset and Novel Methods, *Chenfan Qu, Yiwu Zhong, Chongyu Liu, Guitao Xu, Dezhi Peng, Fengjun Guo, Lianwen Jin*
- 110 ManiFPT: Defining and Analyzing Fingerprints of Generative Models, *Hae Jin Song, Mahyar Khayatkhoei, Wael AbdAlmageed*
- 111 ProMark: Proactive Diffusion Watermarking for Causal Attribution, *Vishal Asnani, John Collomosse, Tu Bui, Xiaoming Liu, Shruti Agarwal*
- 112 CGI-DM: Digital Copyright Authentication for Diffusion Models via Contrasting Gradient Inversion, *Xiaoyu Wu, Yang Hua, Chumeng Liang, Jiaru Zhang, Hao Wang, Tao Song, Haibing Guan*
- 113 SCoFT: Self-Contrastive Fine-Tuning for Equitable Image Generation, *Zhixuan Liu, Peter Schaldenbrand, Beverley-Claire Okogwu, Wenxuan Peng, Youngsik Yun, Andrew Hundt, Jihie Kim, Jean Oh*
- 114 Would Deep Generative Models Amplify Bias in Future Models?, *Tianwei Chen, Yusuke Hirota, Mayu Otani, Noa Garcia, Yuta Nakashima*
- 115 Training Diffusion Models Towards Diverse Image Generation with Reinforcement Learning, *Zichen Miao, Jiang Wang, Ze Wang, Zhengyuan Yang, Lijuan Wang, Qiang Qiu, Zicheng Liu*
- 116 Consistency and Uncertainty: Identifying Unreliable Responses From Black-Box Vision-Language Models for Selective Visual Question Answering, *Zaid Khan, Yun Fu*
- 117 Visual Objectification in Films: Towards a New AI Task for Video
* Interpretation, *Julie Tores, Lucile Sassatelli, Hui-Yin Wu, Clement Bergman, Léa Andolfi, Victor Ecrement, Frédéric Precioso, Thierry Devars, Magali Guaresi, Virginie Julliard, Sarah Lecossais*

- 118 ToonerGAN: Reinforcing GANs for Obfuscating Automated Facial Indexing, *Kartik Thakral, Shashikant Prasad, Stuti Aswani, Mayank Vatsa, Richa Singh*
- 119 MCPNet: An Interpretable Classifier via Multi-Level Concept Prototypes, *Bor-Shiun Wang, Chien-Yi Wang, Wei-Chen Chiu*
- 120 Visual Concept Connectome (VCC): Open World Concept
* Discovery and their Interlayer Connections in Deep Models, *Matthew Kowal, Richard P. Wildes, Konstantinos G. Derpanis*
- 121 Discover and Mitigate Multiple Biased Subgroups in Image Classifiers, *Zeliang Zhang, Mingqian Feng, Zhiheng Li, Chenliang Xu*
- 122 CORES: Convolutional Response-based Score for Out-of-distribution Detection, *Keke Tang, Chao Hou, Weilong Peng, Runnan Chen, Peican Zhu, Wenping Wang, Zhihong Tian*
- 123 Token Transformation Matters: Towards Faithful Post-hoc Explanation for Vision Transformer, *Junyi Wu, Bin Duan, Weitai Kang, Hao Tang, Yan Yan*
- 124 Comparing the Decision-Making Mechanisms by Transformers
* and CNNs via Explanation Methods, *Mingqi Jiang, Saeed Khorram, Li Fuxin*
- 125 On the Faithfulness of Vision Transformer Explanations, *Junyi Wu, Weitai Kang, Hao Tang, Yuan Hong, Yan Yan*
- 126 Understanding Video Transformers via Universal Concept
* Discovery, *Matthew Kowal, Achal Dave, Rares Ambrus, Adrien Gaidon, Konstantinos G. Derpanis, Pavel Tokmakov*
- 127 Explaining the Implicit Neural Canvas: Connecting Pixels to Neurons by Tracing their Contributions, *Namitha Padmanabhan, Matthew Gwilliam, Pulkit Kumar, Shishira R Maiya, Max Ehrlich, Abhinav Shrivastava*
- 128 WWW: A Unified Framework for Explaining What Where and Why of Neural Networks by Interpretation of Neuron Concepts, *Yong Hyun Ahn, Hyeon Bae Kim, Seong Tae Kim*
- 129 HDQMF: Holographic Feature Decomposition Using Quantum Algorithms, *Prathyush Prasanth Poduval, Zhuowen Zou, Mohsen Imani*
- 130 SLICE: Stabilized LIME for Consistent Explanations for Image
* Classification, *Revoti Prasad Bora, Philipp Terhörst, Raymond Veldhuis, Raghavendra Ramachandra, Kiran Raja*
- 131 What Sketch Explainability Really Means for Downstream Tasks?, *Hmrishav Bandyopadhyay, Pinaki Nath Chowdhury, Ayan Kumar Bhunia, Aneeshan Sain, Tao Xiang, Yi-Zhe Song*
- 132 Structured Gradient-based Interpretations via Norm-Regularized Adversarial Training, *Shizhan Gong, Qi Dou, Farzan Farnia*
- 133 Learning Triangular Distribution in Visual World, *Ping Chen, Xingpeng Zhang, Chengtao Zhou, Dichao Fan, Peng Tu, Le Zhang, Yanlin Qian*
- 134 Incremental Residual Concept Bottleneck Models, *Chenming Shang, Shiji Zhou, Hengyuan Zhang, Xinzhe Ni, Yujiu Yang, Yuwang Wang*
- 135 Uncertainty Visualization via Low-Dimensional Posterior Projections, *Omer Yair, Elias Nehme, Tomer Michaeli*
- 136 Epistemic Uncertainty Quantification For Pre-Trained Neural Networks, *Hanjing Wang, Qiang Ji*
- 137 Interpretable Measures of Conceptual Similarity by Complexity-Constrained Descriptive Auto-Encoding, *Alessandro Achille, Greg Ver Steeg, Tian Yu Liu, Matthew Trager, Carson Klingenberg, Stefano Soatto*
- 138 CAPE: CAM as a Probabilistic Ensemble for Enhanced DNN Interpretation, *Townim Faisal Chowdhury, Kewen Liao, Vu Minh Hieu Phan, Minh-Son To, Yutong Xie, Kevin Hung, David Ross, Anton van den Hengel, Johan W. Verjans, Zhibin Liao*
- 139 Discovering and Mitigating Visual Biases through Keyword
* Explanation, *Younghyun Kim, Sangwoo Mo, Minkyu Kim, Kyungmin Lee, Jaeho Lee, Jinwoo Shin*
- 140 DiG-IN: Diffusion Guidance for Investigating Networks - Uncovering Classifier Differences Neuron Visualisations and Visual Counterfactual Explanations, *Maximilian Augustin, Yannic Neuhaus, Matthias Hein*
- 141 Cross-Dimension Affinity Distillation for 3D EM Neuron Segmentation, *Xiaoyu Liu, Miaomiao Cai, Yinda Chen, Yueyi Zhang, Te Shi, Ruobing Zhang, Xuejin Chen, Zhiwei Xiong*
- 142 Continual Self-supervised Learning: Towards Universal Multi-modal
* Medical Data Representation Learning, *Yiwen Ye, Yutong Xie, Jianpeng Zhang, Ziyang Chen, Qi Wu, Yong Xia*
- 143 A Unified Framework for Microscopy Defocus Deblur with Multi-Pyramid Transformer and Contrastive Learning, *Yuelin Zhang, Pengyu Zheng, Wanquan Yan, Chengyu Fang, Shing Shin Cheng*
- 144 CARZero: Cross-Attention Alignment for Radiology Zero-Shot Classification, *Haoran Lai, Qingsong Yao, Zihang Jiang, Rongsheng Wang, Zhiyang He, Xiaodong Tao, S. Kevin Zhou*
- 145 Towards Generalizable Tumor Synthesis, *Qi Chen, Xiaoxi Chen, Haorui Song, Zhiwei Xiong, Alan Yuille, Chen Wei, Zongwei Zhou*
- 146 Tyche: Stochastic In-Context Learning for Medical Image
* Segmentation, *Marianne Rakic, Hallee E. Wong, Jose Javier Gonzalez Ortiz, Beth A. Cimini, John V. Guttag, Adrian V. Dalca*
- 147 Structure-Aware Sparse-View X-ray 3D Reconstruction, *Yuanhao Cai, Jiahao Wang, Alan Yuille, Zongwei Zhou, Angtian Wang*
- 148 Each Test Image Deserves A Specific Prompt: Continual Test-Time Adaptation for 2D Medical Image Segmentation, *Ziyang Chen, Yongsheng Pan, Yiwen Ye, Mengkang Lu, Yong Xia*
- 149 Training Like a Medical Resident: Context-Prior Learning Toward Universal Medical Image Segmentation, *Yunhe Gao*
- 150 C²RV: Cross-Regional and Cross-View Learning for Sparse-View CBCT Reconstruction, *Yiqun Lin, Jiewen Yang, Hualiang Wang, Xinpeng Ding, Wei Zhao, Xiaomeng Li*
- 151 Modality-Agnostic Structural Image Representation Learning for
* Deformable Multi-Modality Medical Image Registration, *Tony C. W. Mok, Zi Li, Yunhao Bai, Jianpeng Zhang, Wei Liu, Yan-Jie Zhou, Ke Yan, Dakai Jin, Yu Shi, Xiaoli Yin, Le Lu, Ling Zhang*
- 152 SI-MIL: Taming Deep MIL for Self-Interpretability in Gigapixel Histopathology, *Saarthak Kapse, Pushpak Pati, Srijan Das, Jingwei Zhang, Chao Chen, Maria Vakalopoulou, Joel Saltz, Dimitris Samaras, Rajarsi R. Gupta, Prateek Prasanna*
- 153 Bootstrapping Chest CT Image Understanding by Distilling Knowledge from X-ray Expert Models, *Weiwei Cao, Jianpeng Zhang, Yingda Xia, Tony C. W. Mok, Zi Li, Xianghua Ye, Le Lu, Jian Zheng, Yuxing Tang, Ling Zhang*
- 154 ViLa-MIL: Dual-scale Vision-Language Multiple Instance Learning for Whole Slide Image Classification, *Jiangbo Shi, Chen Li, Tieliang Gong, Yefeng Zheng, Huazhu Fu*
- 155 Virtual Immunohistochemistry Staining for Histological Images Assisted by Weakly-supervised Learning, *Jiahan Li, Jiuyang Dong, Shenjin Huang, Xi Li, Junjun Jiang, Xiaopeng Fan, Yongbing Zhang*
- 156 Representing Part-Whole Hierarchies in Foundation Models by Learning Localizability Composability and Decomposability from Anatomy via Self Supervision, *Mohammad Reza Hosseinzadeh Taher, Michael B. Gotway, Jianming Liang*
- 157 XFibrosis: Explicit Vessel-Fiber Modeling for Fibrosis Staging from Liver Pathology Images, *Chong Yin, Siqi Liu, Fei Lyu, Jiahao Lu, Sune Darkner, Vincent Wai-Sun Wong, Pong C. Yuen*
- 158 Prompting Vision Foundation Models for Pathology Image Analysis, *Chong Yin, Siqi Liu, Kaiyang Zhou, Vincent Wai-Sun Wong, Pong C. Yuen*
- 159 One-Prompt to Segment All Medical Images, *Junde Wu, Min Xu*
- 160 Learning Large-Factor EM Image Super-Resolution with Generative Priors, *Jiateng Shou, Zeyu Xiao, Shiyu Deng, Wei Huang, Peiyao Shi, Ruobing Zhang, Zhiwei Xiong, Feng Wu*
- 161 Dynamic Graph Representation with Knowledge-aware Attention for Histopathology Whole Slide Image Analysis, *Jiawen Li, Yuxuan Chen, Hongbo Chu, Qiehe Sun, Tian Guan, Anjia Han, Yonghong He*
- 162 MindBridge: A Cross-Subject Brain Decoding Framework,
* *Shizun Wang, Songhua Liu, Zhenxiong Tan, Xinchao Wang*
- 163 Feature Re-Embedding: Towards Foundation Model-Level Performance in Computational Pathology, *Wenhao Tang, Fengtao Zhou, Sheng Huang, Xiang Zhu, Yi Zhang, Bo Liu*
- 164 Data-Efficient Unsupervised Interpolation Without Any Intermediate Frame for 4D Medical Images, *JungEun Kim, Hangyul Yoon, Geondo Park, Kyungsu Kim, Eunho Yang*
- 165 Rethinking Diffusion Model for Multi-Contrast MRI Super-Resolution, *Guangyuan Li, Chen Rao, Juncheng Mo, Zhanjie Zhang, Wei Xing, Lei Zhao*
- 166 Adapting Visual-Language Models for Generalizable Anomaly
* Detection in Medical Images, *Chaoqin Huang, Aofan Jiang,*

- Jinghao Feng, Ya Zhang, Xinchao Wang, Yanfeng Wang
- 167 ZePT: Zero-Shot Pan-Tumor Segmentation via Query-Disentangling and Self-Prompting, *Yankai Jiang, Zhongzhen Huang, Rongzhao Zhang, Xiaofan Zhang, Shaoting Zhang*
- 168 MemSAM: Taming Segment Anything Model for Echocardiography
* Video Segmentation, *Xiaolong Deng, Huisi Wu, Runhao Zeng, Jing Qin*
- 169 Generalizable Whole Slide Image Classification with Fine-Grained Visual-Semantic Interaction, *Hao Li, Ying Chen, Yifei Chen, Rongshan Yu, Wenxian Yang, Liansheng Wang, Bowen Ding, Yuchen Han*
- 170 Incremental Nuclei Segmentation from Histopathological Images via Future-class Awareness and Compatibility-inspired Distillation, *Huyong Wang, Huisi Wu, Jing Qin*
- 171 PH-Net: Semi-Supervised Breast Lesion Segmentation via Patch-wise Hardness, *Siyao Jiang, Huisi Wu, Junyang Chen, Qin Zhang, Jing Qin*
- 172 ToNNO: Tomographic Reconstruction of a Neural Network's Output for Weakly Supervised Segmentation of 3D Medical Images, *Marius Schmidt-Mengin, Alexis Benichoux, Shibeshih Belachew, Nikos Komodakis, Nikos Paragios*
- 173 Think Twice Before Selection: Federated Evidential Active Learning for Medical Image Analysis with Domain Shifts, *Jiayi Chen, Benteng Ma, Hengfei Cui, Yong Xia*
- 174 CPLIP: Zero-Shot Learning for Histopathology with Comprehensive Vision-Language Alignment, *Sajid Javed, Arif Mahmood, Iyyakutti Iyappan Ganapathi, Fayaz Ali Dharejo, Naoufel Werghi, Mohammed Bennamoun*
- 175 Transcriptomics-guided Slide Representation Learning in
* Computational Pathology, *Guillaume Jaume, Lukas Oldenburg, Anurag Vaidya, Richard J. Chen, Drew F.K. Williamson, Thomas Peeters, Andrew H. Song, Faisal Mahmood*
- 176 MicroDiffusion: Implicit Representation-Guided Diffusion for 3D Reconstruction from Limited 2D Microscopy Projections, *Mude Hui, Zihao Wei, Hongru Zhu, Fei Xia, Yuyin Zhou*
- 177 Diversified and Personalized Multi-rater Medical Image
* Segmentation, *Yicheng Wu, Xiangde Luo, Zhe Xu, Xiaoqing Guo, Lie Ju, Zongyuan Ge, Wenjun Liao, Jianfei Cai*
- 178 Modality-agnostic Domain Generalizable Medical Image Segmentation by Multi-Frequency in Multi-Scale Attention, *Ju-Hyeon Nam, Nur Suriza Syazwany, Su Jung Kim, Sang-Chul Lee*
- 179 Decomposing Disease Descriptions for Enhanced Pathology Detection: A Multi-Aspect Vision-Language Pre-training Framework, *Vu Minh Hieu Phan, Yutong Xie, Yuankai Qi, Lingqiao Liu, Liyang Liu, Bowen Zhang, Zhibin Liao, Qi Wu, Minh-Son To, Johan W. Verjans*
- 180 MedM2G: Unifying Medical Multi-Modal Generation via Cross-Guided Diffusion with Visual Invariant, *Chenlu Zhan, Yu Lin, Gaoang Wang, Hongwei Wang, Jian Wu*
- 181 H-ViT: A Hierarchical Vision Transformer for Deformable Image
* Registration, *Morteza Ghahremani, Mohammad Khateri, Bailiang Jian, Benedikt Wiestler, Ehsan Adeli, Christian Wachinger*
- 182 Seeing Unseen: Discover Novel Biomedical Concepts via Geometry-Constrained Probabilistic Modeling, *Jianan Fan, Dongnan Liu, Hang Chang, Heng Huang, Mei Chen, Weidong Cai*
- 183 Fully Convolutional Slice-to-Volume Reconstruction for Single-Stack MRI, *Sean I. Young, Yael Balbastre, Bruce Fischl, Polina Golland, Juan Eugenio Iglesias*
- 184 IIRP-Net: Iterative Inference Residual Pyramid Network for Enhanced Image Registration, *Tai Ma, Suwei Zhang, Jiafeng Li, Ying Wen*
- 185 ChAda-ViT: Channel Adaptive Attention for Joint Representation Learning of Heterogeneous Microscopy Images, *Nicolas Bourriez, Ihab Bendi, Ethan Cohen, Gabriel Watkinson, Maxime Sanchez, Guillaume Bollot, Auguste Genovesio*
- 186 Morphological Prototyping for Unsupervised Slide Representation Learning in Computational Pathology, *Andrew H. Song, Richard J. Chen, Tong Ding, Drew F.K. Williamson, Guillaume Jaume, Faisal Mahmood*
- 187 Modeling Dense Multimodal Interactions Between Biological Pathways and Histology for Survival Prediction, *Guillaume Jaume, Anurag Vaidya, Richard J. Chen, Drew F.K. Williamson, Paul Pu Liang, Faisal Mahmood*
- 188 Accurate Spatial Gene Expression Prediction by Integrating Multi-Resolution Features, *Youngmin Chung, Ji Hun Ha, Kyeong Chan Im, Joo Sang Lee*
- 189 Teeth-SEG: An Efficient Instance Segmentation Framework for Orthodontic Treatment based on Multi-Scale Aggregation and Anthropomorphic Prior Knowledge, *Bo Zou, Shaofeng Wang, Hao Liu, Gaoyue Sun, Yajie Wang, Feifei Zuo, Chengbin Quan, Youjian Zhao*
- 190 Low-Rank Knowledge Decomposition for Medical Foundation Models, *Yuhang Zhou, Haolin Li, Siyuan Du, Jiangchao Yao, Ya Zhang, Yanfeng Wang*
- 191 M3-UDA: A New Benchmark for Unsupervised Domain Adaptive Fetal Cardiac Structure Detection, *Bin Pu, Liwen Wang, Jiwen Yang, Guannan He, Xingbo Dong, Shengli Li, Ying Tan, Ming Chen, Zhe Jin, Kenli Li, Xiaomeng Li*
- 192 CycleINR: Cycle Implicit Neural Representation for Arbitrary-Scale Volumetric Super-Resolution of Medical Data, *Wei Fang, Yuxing Tang, Heng Guo, Mingze Yuan, Tony C. W. Mok, Ke Yan, Jiawen Yao, Xin Chen, Zaiyi Liu, Le Lu, Ling Zhang, Minfeng Xu*
- 193 Constructing and Exploring Intermediate Domains in Mixed Domain Semi-supervised Medical Image Segmentation, *Qinghe Ma, Jian Zhang, Lei Qi, Qian Yu, Yinghuan Shi, Yang Gao*
- 194 PairAug: What Can Augmented Image-Text Pairs Do for Radiology?, *Yutong Xie, Qi Chen, Sinuo Wang, Minh-Son To, Iris Lee, Ee Win Khoo, Kerolos Hendy, Daniel Koh, Yong Xia, Qi Wu*
- 195 Intraoperative 2D/3D Image Registration via Differentiable X-ray Rendering, *Vivek Gopalakrishnan, Neel Dey, Polina Golland*
- 196 Mudslide: A Universal Nuclear Instance Segmentation Method,
* *Jun Wang*
- 197 Correlation-aware Coarse-to-fine MLPs for Deformable Medical
* Image Registration, *Mingyuan Meng, Dagan Feng, Lei Bi, Jinman Kim*
- 198 Rotation-Agnostic Image Representation Learning for Digital Pathology, *Saghir Alfasly, Abubakr Shafique, Peyman Nejat, Jibrán Khan, Areej Alsaafin, Ghazal Alabtah, H.R. Tizhoosh*
- 199 Tumor Micro-environment Interactions Guided Graph Learning for Survival Analysis of Human Cancers from Whole-slide Pathological Images, *Wei Shao, YangYang Shi, Daoqiang Zhang, JunJie Zhou, Peng Wan*
- 200 MLIP: Enhancing Medical Visual Representation with Divergence Encoder and Knowledge-guided Contrastive Learning, *Zhe Li, Laurence T. Yang, Bocheng Ren, Xin Nie, Zhangyang Gao, Cheng Tan, Stan Z. Li*
- 201 FocusMAE: Gallbladder Cancer Detection from Ultrasound Videos with Focused Masked Autoencoders, *Soumen Basu, Mayana Gupta, Chetan Madan, Pankaj Gupta, Chetan Arora*
- 202 Bi-level Learning of Task-Specific Decoders for Joint Registration and One-Shot Medical Image Segmentation, *Xin Fan, Xiaolin Wang, Jiaxin Gao, Jia Wang, Zhongxuan Luo, Risheng Liu*
- 203 PrPSeg: Universal Proposition Learning for Panoramic Renal Pathology Segmentation, *Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jialin Yue, Juming Xiong, Lining Yu, Yifei Wu, Mengmeng Yin, Yu Wang, Shilin Zhao, Yucheng Tang, Haichun Yang, Yuankai Huo*
- 204 Versatile Medical Image Segmentation Learned from Multi-Source Datasets via Model Self-Disambiguation, *Xiaoyang Chen, Hao Zheng, Yuemeng Li, Yuncong Ma, Liang Ma, Hongming Li, Yong Fan*
- 205 Masked Autoencoders for Microscopy are Scalable Learners of
* Cellular Biology, *Oren Kraus, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean, Jess Leung, Vasudev Sharma, Ayla Khan, Jia Balakrishnan, Safiye Celik, Dominique Beaini, Maciej Sypetkowski, Chi Vicky Cheng, Kristen Morse, Maureen Makes, Ben Mabey, Berton Earnshaw*
- 206 EMCAD: Efficient Multi-scale Convolutional Attention Decoding for Medical Image Segmentation, *Md Mostafijur Rahman, Mustafa Munir, Radu Marculescu*
- 207 Neural Underwater Scene Representation, *Yunkai Tang, Chengxuan Zhu, Renjie Wan, Chao Xu, Boxin Shi*

- 208 Hearing Anything Anywhere, *Mason Long Wang, Ryosuke Sawata, Samuel Clarke, Ruohan Gao, Shangzhe Wu, Jiajun Wu*
- 209 VMINer: Versatile Multi-view Inverse Rendering with Near- and
 ✱ Far-field Light Sources, *Fan Fei, Jiajun Tang, Ping Tan, Boxin Shi*
- 210 EventPS: Real-Time Photometric Stereo Using an Event Camera,
 ✱ *Bohan Yu, Jieji Ren, Jin Han, Feishi Wang, Jinxiu Liang, Boxin Shi*
- 211 DiLiGenRT: A Photometric Stereo Dataset with Quantified
 Roughness and Translucency, *Heng Guo, Jieji Ren, Feishi Wang, Boxin Shi, Mingjun Ren, Yasuyuki Matsushita*
- 212 NeRSP: Neural 3D Reconstruction for Reflective Objects with
 Sparse Polarized Images, *Yufei Han, Heng Guo, Koki Fukai, Hiroaki Santo, Boxin Shi, Fumio Okura, Zhanyu Ma, Yunpeng Jia*
- 213 EvDiG: Event-guided Direct and Global Components Separation,
 ✱ *Xinyu Zhou, Peiqi Duan, Boyu Li, Chu Zhou, Chao Xu, Boxin Shi*
- 214 Differentiable Display Photometric Stereo, *Seokjun Choi, Seungwoo Yoon, Giljoo Nam, Seungyong Lee, Seung-Hwan Baek*
- 215 Bayesian Differentiable Physics for Cloth Digitalization, *Deshan Gong, Ningtao Mao, He Wang*
- 216 Atlantis: Enabling Underwater Depth Estimation with Stable
 ✱ Diffusion, *Fan Zhang, Shaodi You, Yu Li, Ying Fu*
- 217 Sparse Views Near Light: A Practical Paradigm for Uncalibrated
 Point-light Photometric Stereo, *Mohammed Brahimi, Bjoern Haefner, Zhenzhang Ye, Bastian Goldluecke, Daniel Cremers*
- 218 Diffusion Reflectance Map: Single-Image Stochastic Inverse
 ✱ Rendering of Illumination and Reflectance, *Yuto Enyo, Ko Nishino*
- 219 Deep Single Image Camera Calibration by Heatmap
 Regression to Recover Fisheye Images Under Manhattan World
 Assumption, *Nobuhiko Wakai, Satoshi Sato, Yasunori Ishii, Takayoshi Yamashita*
- 220 Physics-guided Shape-from-Template: Monocular Video
 Perception through Neural Surrogate Models, *David Stotko, Nils Wandel, Reinhard Klein*
- 221 Spin-UP: Spin Light for Natural Light Uncalibrated Photometric
 Stereo, *Zongrui Li, Zhan Lu, Haojie Yan, Boxin Shi, Gang Pan, Qian Zheng, Xudong Jiang*
- 222 Discontinuity-preserving Normal Integration with Auxiliary
 Edges, *Hyomin Kim, Yuchool Jung, Seungyong Lee*
- 223 A Theory of Joint Light and Heat Transport for Lambertian
 Scenes, *Mani Ramanagopal, Sriram Narayanan, Aswin C. Sankaranarayanan, Srinivasa G. Narasimhan*
- 224 IDGuard: Robust General Identity-centric POI Proactive Defense
 Against Face Editing Abuse, *Yunshu Dai, Jianwei Fei, Fangjun Huang*
- 225 Ungeneralizable Examples, *Jingwen Ye, Xinchao Wang*
- 226 Distilled Datamodel with Reverse Gradient Matching, *Jingwen Ye, Ruonan Yu, Songhua Liu, Xinchao Wang*
- 227 EditGuard: Versatile Image Watermarking for Tamper Localization
 and Copyright Protection, *Xuanyu Zhang, Runyi Li, Jiwen Yu, Youmin Xu, Weiqi Li, Jian Zhang*
- 228 SocialCounterfactuals: Probing and Mitigating Intersectional
 Social Biases in Vision-Language Models with Counterfactual
 Examples, *Phillip Howard, Avinash Madasu, Tiep Le, Gustavo Lujan Moreno, Anahita Bhiwandiwala, Vasudev Lal*
- 229 FedAS: Bridging Inconsistency in Personalized Federated
 Learning, *Xiyuan Yang, Wenke Huang, Mang Ye*
- 230 FairRAG: Fair Human Generation via Fair Retrieval
 Augmentation, *Robik Shrestha, Yang Zou, Qiuyu Chen, Zhiheng Li, Yusheng Xie, Siqi Deng*
- 231 Self-Discovering Interpretable Diffusion Latent Directions for
 Responsible Text-to-Image Generation, *Hang Li, Chengzhi Shen, Philip Torr, Volker Tresp, Jindong Gu*
- 232 ExMap: Leveraging Explainability Heatmaps for Unsupervised
 Group Robustness to Spurious Correlations, *Rwiddhi Chakraborty, Adrian Sletten, Michael C. Kampffmeyer*
- 233 Data Valuation and Detections in Federated Learning, *Wenqian Li, Shuran Fu, Fengrui Zhang, Yan Pang*
- 234 Utility-Fairness Trade-Offs and How to Find Them, *Sepehr Dehdashtian, Bashir Sadeghi, Vishnu Naresh Boddeti*
- 235 SimAC: A Simple Anti-Customization Method for Protecting Face
 Privacy against Text-to-Image Synthesis of Diffusion Models, *Feifei Wang, Zhentao Tan, Tianyi Wei, Yue Wu, Qidong Huang*
- 236 GLOW: Global Layout Aware Attacks on Object Detection,
Jun Bao, Buyu Liu, Kui Ren, Jun Yu
- 237 FADES: Fair Disentanglement with Sensitive Relevance,
Taeuk Jang, Xiaoqian Wang
- 238 Fair Federated Learning under Domain Skew with Local
 Consistency and Domain Diversity, *Yuhang Chen, Wenke Huang, Mang Ye*
- 239 WaterF: Robust Watermarks in Radiance Fields for Protection of
 Copyrights, *Youngdong Jang, Dong In Lee, MinHyuk Jang, Jong Wook Kim, Feng Yang, Sangpil Kim*
- 240 FLHetBench: Benchmarking Device and State Heterogeneity
 in Federated Learning, *Junyuan Zhang, Shuang Zeng, Miao Zhang, Runxi Wang, Feifei Wang, Yuyin Zhou, Paul Pu Liang, Liangqiong Qu*
- 241 An Upload-Efficient Scheme for Transferring Knowledge From a
 Server-Side Pre-trained Generator to Clients in Heterogeneous
 Federated Learning, *Jianqing Zhang, Yang Liu, Yang Hua, Jian Cao*
- 242 Privacy-Preserving Optics for Enhancing Protection in Face
 De-Identification, *Jhon Lopez, Carlos Hinojosa, Henry Arguello, Bernard Ghanem*
- 243 A Stealthy Wrongdoer: Feature-Oriented Reconstruction Attack
 against Split Learning, *Xiaoyang Xu, Mengda Yang, Wenzhe Yi, Ziang Li, Juan Wang, Hongxin Hu, Yong Zhuang, Yaxin Liu*
- 244 RCL: Reliable Continual Learning for Unified Failure
 Detection, *Fei Zhu, Zhen Cheng, Xu-Yao Zhang, Cheng-Lin Liu, Zhaoxiang Zhang*
- 245 Global and Local Prompts Cooperation via Optimal Transport for
 Federated Learning, *Hongxia Li, Wei Huang, Jingya Wang, Ye Shi*
- 246 Gaussian Shading: Provable Performance-Lossless Image
 Watermarking for Diffusion Models, *Zijin Yang, Kai Zeng, Kejiang Chen, Han Fang, Weiming Zhang, Nenghai Yu*
- 247 Explaining CLIP's Performance Disparities on Data from Blind/
 Low Vision Users, *Daniela Massiceti, Camilla Longden, Agnieszka Slowik, Samuel Wills, Martin Grayson, Cecily Morrison*
- 248 Model Inversion Robustness: Can Transfer Learning Help?,
Sy-Tuyen Ho, Koh Jun Hao, Keshigeyan Chandrasegaran, Ngoc-Bao Nguyen, Ngai-Man Cheung
- 249 Make Me a BNN: A Simple Strategy for Estimating Bayesian
 Uncertainty from Pre-trained Models, *Gianni Franchi, Olivier Laurent, Maxence Leguery, Andrei Bursuc, Andrea Pilzer, Angela Yao*
- 250 Validating Privacy-Preserving Face Recognition under a
 Minimum Assumption, *Hui Zhang, Xingbo Dong, YenLung Lai, Ying Zhou, Xiaoyan Zhang, Xingguo Lv, Zhe Jin, Xuejun Li*
- 251 Re-thinking Data Availability Attacks Against Deep Neural
 Networks, *Bin Fang, Bo Li, Shuang Wu, Shouhong Ding, Ran Yi, Lizhuang Ma*
- 252 OpenBias: Open-set Bias Detection in Text-to-Image Generative
 ✱ Models, *Moreno D'Inca, Elia Peruzzo, Massimiliano Mancini, Dejia Xu, Vidit Goel, Xingqian Xu, Zhangyang Wang, Humphrey Shi, Nicu Sebe*
- 253 In-distribution Public Data Synthesis with Diffusion Models for
 Differentially Private Image Classification, *Jinseong Park, Yujin Choi, Jaewook Lee*
- 254 Leak and Learn: An Attacker's Cookbook to Train Using Leaked
 Data from Federated Learning, *Joshua C. Zhao, Ahaan Dabholkar, Atul Sharma, Saurabh Bagchi*
- 255 Countering Personalized Text-to-Image Generation with Influence
 Watermarks, *Hanwen Liu, Zhicheng Sun, Yadong Mu*
- 256 Fair-VPT: Fair Visual Prompt Tuning for Image
 Classification, *Sungho Park, Hyeran Byun*
- 257 Relaxed Contrastive Learning for Federated Learning, *Seonguk Seo, Jinkyu Kim, Geeho Kim, Bohyung Han*
- 258 FairCLIP: Harnessing Fairness in Vision-Language Learning, *Yan Luo, Min Shi, Muhammad Osama Khan, Muhammad Muneeb Afzal, Hao Huang, Shuaihang Yuan, Yu Tian, Luo Song, Ava Kouhana, Tobias Elze, Yi Fang, Mengyu Wang*
- 259 Steganographic Passport: An Owner and User Verifiable
 Credential for Deep Model IP Protection Without Retraining, *Qi Cui, Ruohan Meng, Chaohui Xu, Chip-Hong Chang*
- 260 Adaptive Hyper-graph Aggregation for Modality-Agnostic
 Federated Learning, *Fan Qi, Shuai Li*

- 261 Navigate Beyond Shortcuts: Debiased Learning Through the
* Lens of Neural Collapse, *Yining Wang, Junjie Sun, Chenyue Wang, Mi Zhang, Min Yang*
- 262 Enhancing Intrinsic Features for Debiasing via Investigating Class-Discerning Common Attributes in Bias-Contrastive Pair, *Jeonghoon Park, Chaeyeon Chung, Jaegul Choo*
- 263 Device-Wise Federated Network Pruning, *Shangqian Gao, Junyi Li, Zeyu Zhang, Yanfu Zhang, Weidong Cai, Heng Huang*
- 264 All Rivers Run to the Sea: Private Learning with Asymmetric Flows, *Yue Niu, Ramy E. Ali, Saurav Prakash, Salman Avestimehr*
- 265 VA3: Virtually Assured Amplification Attack on Probabilistic
* Copyright Protection for Text-to-Image Generative Models, *Xiang Li, Qianli Shen, Kenji Kawaguchi*
- 266 CPR: Retrieval Augmented Generation for Copyright Protection, *Aditya Golatkar, Alessandro Achille, Luca Zancato, Yu-Xiang Wang, Ashwin Swaminathan, Stefano Soatto*
- 267 Communication-Efficient Federated Learning with Accelerated Client Gradient, *Geeho Kim, Jinkyu Kim, Bohyung Han*
- 268 Self-supervised Debiasing Using Low Rank Regularization, *Geon Yeong Park, Chanyong Jung, Sangmin Lee, Jong Chul Ye, Sang Wan Lee*
- 269 Facial Identity Anonymization via Intrinsic and Extrinsic Attention Distraction, *Zhenzhong Kuang, Xiaochen Yang, Yingjie Shen, Chao Hu, Jun Yu*
- 270 Collaborative Learning of Anomalies with Privacy (CLAP) for Unsupervised Video Anomaly Detection: A New Baseline, *Anas Al-Jahham, Muhammad Zaigham Zaheer, Nurbek Tastan, Karthik Nandakumar*
- 271 Label-Efficient Group Robustness via Out-of-Distribution Concept Curation, *Yiwei Yang, Anthony Z. Liu, Robert Wolfe, Aylin Caliskan, Bill Howe*
- 272 Long-Tailed Anomaly Detection with Learnable Class Names, *Chih-Hui Ho, Kuan-Chuan Peng, Nuno Vasconcelos*
- 273 Robust Emotion Recognition in Context Debiasing, *Dingkang Yang, Kun Yang, Mingcheng Li, Shunli Wang, Shuaibing Wang, Lihua Zhang*
- 274 Correlation-Decoupled Knowledge Distillation for Multimodal Sentiment Analysis with Incomplete Modalities, *Mingcheng Li, Dingkan Yang, Xiao Zhao, Shuaibing Wang, Yan Wang, Kun Yang, Mingyang Sun, Dongliang Kou, Ziyun Qian, Lihua Zhang*
- 275 An Edit Friendly DDPM Noise Space: Inversion and Manipulations, *Inbar Huberman-Spiegelglas, Vladimir Kulikov, Tomer Michaeli*
- 276 SleepVST: Sleep Staging from Near-Infrared Video Signals using
* Pre-Trained Transformers, *Jonathan F. Carter, João Jorge, Oliver Gibson, Lionel Tarassenko*
- 277 AM-RADIO: Agglomerative Vision Foundation Model Reduce All Domains Into One, *Mike Ranzinger, Greg Heinrich, Jan Kautz, Pavlo Molchanov*
- 278 Towards Language-Driven Video Inpainting via Multimodal Large Language Models, *Jianzong Wu, Xiangtai Li, Chenyang Si, Shangchen Zhou, Jingkan Yang, Jiangning Zhang, Yining Li, Kai Chen, Yunhai Tong, Ziwei Liu, Chen Change Loy*
- 279 FedSOL: Stabilized Orthogonal Learning with Proximal Restrictions in Federated Learning, *Gihun Lee, Minchan Jeong, Sangmook Kim, Jaehoon Oh, Se-Young Yun*
- 280 UnionFormer: Unified-Learning Transformer with Multi-View Representation for Image Manipulation Detection and Localization, *Shuaibo Li, Wei Ma, Jianwei Guo, Shibiao Xu, Benchong Li, Xiaopeng Zhang*
- 281 Motion Blur Decomposition with Cross-shutter Guidance, *Xiang Ji, Haiyang Jiang, Yinqiang Zheng*
- 282 SNIDA: Unlocking Few-Shot Object Detection with Non-linear Semantic Decoupling Augmentation, *Yanjie Wang, Xu Zou, Luxin Yan, Sheng Zhong, Jiahuan Zhou*
- 283 Rapid 3D Model Generation with Intuitive 3D Input,
* *Tianrun Chen, Chaotao Ding, Shangzhan Zhang, Chunan Yu, Ying Zang, Zejian Li, Sida Peng, Lingyun Sun*
- 284 SketchINR: A First Look into Sketches as Implicit Neural Representations, *Hmrishav Bandyopadhyay, Ayan Kumar Bhunia, Pinaki Nath Chowdhury, Aneeshan Sain, Tao Xiang, Timothy Hospedales, Yi-Zhe Song*
- 285 ERMVP: Communication-Efficient and Collaboration-Robust Multi-Vehicle Perception in Challenging Environments, *Jingyu Zhang, Kun Yang, Yilei Wang, Hanqi Wang, Peng Sun, Liang Song*
- 286 DiaLoc: An Iterative Approach to Embodied Dialog Localization, *Chao Zhang, Mohan Li, Ignas Budvytis, Stephan Liwicki*
- 287 WildlifeMapper: Aerial Image Analysis for Multi-Species Detection and Identification, *Satish Kumar, Bowen Zhang, Chandrakanth Gudavalli, Connor Levenson, Lacey Hughey, Jared A. Stabach, Irene Amoke, Gordon Ojwang, Joseph Mukeka, Stephen Mwiu, Joseph Ogutu, Howard Frederick, B.S. Manjunath*
- 288 Harnessing Meta-Learning for Improving Full-Frame Video Stabilization, *Muhammad Kashif Ali, Eun Woo Im, Dongjin Kim, Tae Hyun Kim*
- 289 De-confounded Data-free Knowledge Distillation for Handling Distribution Shifts, *Yuzheng Wang, Dingkan Yang, Zhaoyu Chen, Yang Liu, Siao Liu, Wenqiang Zhang, Lihua Zhang, Lizhe Qi*
- 290 Day-Night Cross-domain Vehicle Re-identification, *Hongchao Li, Jingong Chen, Aihua Zheng, Yong Wu, Yonglong Luo*
- 291 Brush2Prompt: Contextual Prompt Generator for Object Inpainting, *Mang Tik Chiu, Yuqian Zhou, Lingzhi Zhang, Zhe Lin, Connelly Barnes, Sohrab Amirghodsi, Eli Shechtman, Humphrey Shi*
- 292 Cloud-Device Collaborative Learning for Multimodal Large Language Models, *Guanqun Wang, Jiaming Liu, Chenxuan Li, Yuan Zhang, Junpeng Ma, Xinyu Wei, Kevin Zhang, Maurice Chong, Renrui Zhang, Yijiang Liu, Shanghang Zhang*
- 293 Making Visual Sense of Oracle Bones for You and Me, *Runqi Qiao, Lan Yang, Kaiyue Pang, Honggang Zhang*
- 294 Boosting Object Detection with Zero-Shot Day-Night Domain Adaptation, *Zhipeng Du, Miaoqing Shi, Jiankan Deng*
- 295 InNeRF360: Text-Guided 3D-Consistent Object Inpainting on 360-degree Neural Radiance Fields, *Dongqing Wang, Tong Zhang, Alaa Abboud, Sabine Süssstrunk*
- 296 Language Models as Black-Box Optimizers for Vision-Language Models, *Shihong Liu, Samuel Yu, Zhiqiu Lin, Deepak Pathak, Deva Ramanan*
- 297 Mind Marginal Non-Crack Regions: Clustering-Inspired Representation Learning for Crack Segmentation, *Zhuangzhuang Chen, Zhuonan Lai, Jie Chen, Jianqiang Li*
- 298 InstructDiffusion: A Generalist Modeling Interface for Vision Tasks, *Zigang Geng, Binxin Yang, Tiankai Hang, Chen Li, Shuyang Gu, Ting Zhang, Jianmin Bao, Zheng Zhang, Houqiang Li, Han Hu, Dong Chen, Baining Guo*
- 299 Design: A Pipeline for Controllable Design Template Generation, *Haohan Weng, Danqing Huang, Yu Qiao, Zheng Hu, Chin-Yew Lin, Tong Zhang, C. L. Philip Chen*
- 300 Physical Backdoor: Towards Temperature-based Backdoor Attacks in the Physical World, *Wen Yin, Jian Lou, Pan Zhou, Yulai Xie, Dan Feng, Yuhua Sun, Tailai Zhang, Lichao Sun*
- 301 Behind the Veil: Enhanced Indoor 3D Scene Reconstruction with Occluded Surfaces Completion, *Su Sun, Cheng Zhao, Yuliang Guo, Ruoyu Wang, Xinyu Huang, Yingjie Victor Chen, Liu Ren*
- 302 EarthLoc: Astronaut Photography Localization by Indexing Earth from Space, *Gabriele Berton, Alex Stoken, Barbara Caputo, Carlo Masone*
- 303 DiffForensics: Leveraging Diffusion Prior to Image Forgery Detection and Localization, *Zeqin Yu, Jiangqun Ni, Yuzhen Lin, Haoyi Deng, Bin Li*
- 304 MuseChat: A Conversational Music Recommendation System for
* Videos, *Zhikang Dong, Xiulong Liu, Bin Chen, Pawel Polak, Peng Zhang*
- 305 The Unreasonable Effectiveness of Pre-Trained Features for
* Camera Pose Refinement, *Gabriele Trivigno, Carlo Masone, Barbara Caputo, Torsten Sattler*
- 306 Blind Image Quality Assessment Based on Geometric Order Learning, *Nyeong-Ho Shin, Seon-Ho Lee, Chang-Su Kim*
- 307 CrowdDiff: Multi-hypothesis Crowd Density Estimation using Diffusion Models, *Yasiru Ranasinghe, Nithin Gopalakrishnan Nair, Wele Gedara Chaminda Bandara, Vishal M. Patel*
- 308 Towards Efficient Replay in Federated Incremental Learning, *Yichen Li, Qunwei Li, Haozhao Wang, Ruixuan Li, Wenliang Zhong, Guannan Zhang*

- 309 MART: Masked Affective Representation Learning via Masked Temporal Distribution Distillation, *Zhicheng Zhang, Pancheng Zhao, Eunil Park, Jufeng Yang*
- 310 PolarRec: Improving Radio Interferometric Data Reconstruction Using Polar Coordinates, *Ruoqi Wang, Zhuoyang Chen, Jiayi Zhu, Qiong Luo, Feng Wang*
- 311 Constrained Layout Generation with Factor Graphs, *Mohammed Haroon Dupty, Yanfei Dong, Sicong Leng, Guoji Fu, Yong Liang Goh, Wei Lu, Wee Sun Lee*
- 312 Visual In-Context Prompting, *Feng Li, Qing Jiang, Hao Zhang, Tianhe Ren, Shilong Liu, Xueyan Zou, Huaizhe Xu, Hongyang Li, Jianwei Yang, Chunyuan Li, Lei Zhang, Jianfeng Gao*
- 313 Traceable Federated Continual Learning, *Qiang Wang, Bingyan Liu, Yawen Li*
- 314 Interactive Continual Learning: Fast and Slow Thinking, *Biqing Qi, Xinquan Chen, Junqi Gao, Dong Li, Jianxing Liu, Ligang Wu, Bowen Zhou*
- 315 PIGEON: Predicting Image Geolocations,
* *Lukas Haas, Michal Skreta, Silas Alberti, Chelsea Finn*
- 316 LQMFormer: Language-aware Query Mask Transformer for Referring Image Segmentation, *Nisarg A. Shah, Vibashan VS, Vishal M. Patel*
- 317 ViP-LLaVA: Making Large Multimodal Models Understand Arbitrary Visual Prompts, *Mu Cai, Haotian Liu, Siva Karthik Mustikovela, Gregory P. Meyer, Yuning Chai, Dennis Park, Yong Jae Lee*
- 318 DePT: Decoupled Prompt Tuning, *Ji Zhang, Shihan Wu, Lianli Gao, Heng Tao Shen, Jingkuan Song*
- 319 Grounded Question-Answering in Long Egocentric Videos, *Shangzhe Di, Weidi Xie*
- 320 HalluciDoctor: Mitigating Hallucinatory Toxicity in Visual Instruction Data, *Qifan Yu, Juncheng Li, Longhui Wei, Liang Pang, Wentao Ye, Bosheng Qin, Siliang Tang, Qi Tian, Yueting Zhuang*
- 321 ViTamin: Designing Scalable Vision Models in the Vision-Language Era, *Jieneng Chen, Qihang Yu, Xiaohui Shen, Alan Yuille, Liang-Chieh Chen*
- 322 The Manga Whisperer: Automatically Generating Transcriptions for Comics, *Ragav Sachdeva, Andrew Zisserman*
- 323 Learning to Localize Objects Improves Spatial Reasoning in Visual-LLMs, *Kanchana Ranasinghe, Satya Narayan Shukla, Omid Poursaeed, Michael S. Ryoo, Tsung-Yu Lin*
- 324 The Neglected Tails in Vision-Language Models, *Shubham Parashar, Zhiqiu Lin, Tian Liu, Xiangjue Dong, Yanan Li, Deva Ramanan, James Caverlee, Shu Kong*
- 325 Unveiling Parts Beyond Objects: Towards Finer-Granularity Referring Expression Segmentation, *Wenxuan Wang, Tongtian Yue, Yisi Zhang, Longteng Guo, Xingjian He, Xinlong Wang, Jing Liu*
- 326 GLaMM: Pixel Grounding Large Multimodal Model, *Hanoona Rasheed, Muhammad Maaz, Sahal Shaji, Abdelrahman Shaker, Salman Khan, Hisham Cholakkal, Rao M. Anwer, Eric Xing, Ming-Hsuan Yang, Fahad S. Khan*
- 327 Alpha-CLIP: A CLIP Model Focusing on Wherever You Want, *Zeyi Sun, Ye Fang, Tong Wu, Pan Zhang, Yuhang Zang, Shu Kong, Yuanjun Xiong, Dahua Lin, Jiaqi Wang*
- 328 Pixel-Aligned Language Model, *Jiarui Xu, Xingyi Zhou, Shen Yan, Xiuye Gu, Anurag Arnab, Chen Sun, Xiaolong Wang, Cordelia Schmid*
- 329 mPLUG-Owl2: Revolutionizing Multi-modal Large Language
* Model with Modality Collaboration, *Qinghao Ye, Haiyang Xu, Jiabo Ye, Ming Yan, Anwen Hu, Haowei Liu, Qi Qian, Ji Zhang, Fei Huang*
- 330 SNIFFER: Multimodal Large Language Model for Explainable Out-of-Context Misinformation Detection, *Peng Qi, Zehong Yan, Wynne Hsu, Mong Li Lee*
- 331 Towards CLIP-driven Language-free 3D Visual Grounding via 2D-3D Relational Enhancement and Consistency, *Yuqi Zhang, Han Luo, Yinjie Lei*
- 332 SC-Tune: Unleashing Self-Consistent Referential Comprehension in Large Vision Language Models, *Tongtian Yue, Jie Cheng, Longteng Guo, Xingyuan Dai, Zijia Zhao, Xingjian He, Gang Xiong, Yisheng Lv, Jing Liu*
- 333 V?: Guided Visual Search as a Core Mechanism in Multimodal LLMs, *Penghao Wu, Saining Xie*
- 334 Improved Visual Grounding through Self-Consistent Explanations, *Ruozhen He, Paola Cascante-Bonilla, Ziyang Yang, Alexander C. Berg, Vicente Ordóñez*
- 335 Distilling Vision-Language Models on Millions of Videos, *Yue Zhao, Long Zhao, Xingyi Zhou, Jialin Wu, Chun-Te Chu, Hui Miao, Florian Schroff, Hartwig Adam, Ting Liu, Boqing Gong, Philipp Krahenbuhl, Liangzhe Yuan*
- 336 Separating the "Chirp" from the "Chat": Self-supervised Visual Grounding of Sound and Language, *Mark Hamilton, Andrew Zisserman, John R. Hershey, William T. Freeman*
- 337 Referring Image Editing: Object-level Image Editing via Referring Expressions, *Chang Liu, Xiangtai Li, Henghui Ding*
- 338 Vision-and-Language Navigation via Causal Learning, *Liuyi Wang, Zongtao He, Ronghao Dang, Mengjiao Shen, Chengju Liu, Qijun Chen*
- 339 VISTA-LLAMA: Reducing Hallucination in Video Language Models via Equal Distance to Visual Tokens, *Fan Ma, Xiaojie Jin, Heng Wang, Yuchen Xian, Jiashi Feng, Yi Yang*
- 340 Ranking Distillation for Open-Ended Video Question Answering with Insufficient Labels, *Tianming Liang, Chaolei Tan, Beihao Xia, Wei-Shi Zheng, Jian-Fang Hu*
- 341 CLIP as RNN: Segment Countless Visual Concepts without Training Endeavor, *Shuyang Sun, Runjia Li, Philip Torr, Xiuye Gu, Siyang Li*
- 342 Quilt-LLaVA: Visual Instruction Tuning by Extracting Localized Narratives from Open-Source Histopathology Videos, *Mehmet Saygin Seyfioglu, Wisdom O. Ikezogwo, Fatemeh Ghezloo, Ranjay Krishna, Linda Shapiro*
- 343 Aligning and Prompting Everything All at Once for Universal Visual Perception, *Yunhang Shen, Chaoyou Fu, Peixian Chen, Mengdan Zhang, Ke Li, Xing Sun, Yunsheng Wu, Shaohui Lin, Rongrong Ji*
- 344 Can I Trust Your Answer? Visually Grounded Video Question
* Answering, *Junbin Xiao, Angela Yao, Yicong Li, Tat-Seng Chua*
- 345 Prompt Highlighter: Interactive Control for Multi-Modal LLMs, *Yuechen Zhang, Shengju Qian, Bohao Peng, Shu Liu, Jiaya Jia*
- 346 Language-only Training of Zero-shot Composed Image Retrieval, *Geonmo Gu, Sanghyuk Chun, Wonjae Kim, Yoohoon Kang, Sangdoon Yun*
- 347 MoReVQA: Exploring Modular Reasoning Models for Video Question Answering, *Juhong Min, Shyamal Buch, Arsha Nagrani, Minsu Cho, Cordelia Schmid*
- 348 Let's Think Outside the Box: Exploring Leap-of-Thought in Large Language Models with Creative Humor Generation, *Shanshan Zhong, Zhongzhan Huang, Shanghua Gao, Wushao Wen, Liang Lin, Marinka Zitnik, Pan Zhou*
- 349 CLOVA: A Closed-Loop Visual Assistant with Tool Usage and Update, *Zhi Gao, Yuntao Du, Xintong Zhang, Xiaojian Ma, Wenjuan Han, Song-Chun Zhu, Qing Li*
- 350 Naturally Supervised 3D Visual Grounding with Language-Regularized Concept Learners, *Chun Feng, Joy Hsu, Weiyu Liu, Jiajun Wu*
- 351 Synthesize Diagnose and Optimize: Towards Fine-Grained Vision-Language Understanding, *Wujian Peng, Sicheng Xie, Zuyao You, Shiyi Lan, Zuxuan Wu*
- 352 AssistGUI: Task-Oriented PC Graphical User Interface Automation, *Difei Gao, Lei Ji, Zechen Bai, Mingyu Ouyang, Peiran Li, Dongxing Mao, Qinchen Wu, Weichen Zhang, Peiyi Wang, Xiangwu Guo, Hengxu Wang, Luowei Zhou, Mike Zheng Shou*
- 353 SEED-Bench: Benchmarking Multimodal Large Language Models, *Bohao Li, Yuying Ge, Yixiao Ge, Guangzhi Wang, Rui Wang, Ruimao Zhang, Ying Shan*
- 354 Unknown Prompt the only Lacuna: Unveiling CLIP's Potential for Open Domain Generalization, *Mainak Singha, Ankit Jha, Shirsha Bose, Ashwin Nair, Moloud Abdar, Biplab Banerjee*
- 355 Panda-70M: Captioning 70M Videos with Multiple Cross-Modality Teachers, *Tsai-Shien Chen, Aliaksandr Siarohin, Willi Menapace, Ekaterina Deyneka, Hsiang-wei Chao, Byung Eun Jeon, Yuwei Fang, Hsin-Ying Lee, Jian Ren, Ming-Hsuan Yang, Sergey Tulyakov*
- 356 Decoupling Static and Hierarchical Motion Perception for Referring Video Segmentation, *Shuting He, Henghui Ding*
- 357 Causal-CoG: A Causal-Effect Look at Context Generation for
* Boosting Multi-modal Language Models, *Shitian Zhao, Zhuowan Li, Yadong Lu, Alan Yuille, Yan Wang*

- 358 Posterior Distillation Sampling, *Juil Koo, Chanho Park, Minhyuk Sung*
- 359 Towards More Unified In-context Visual Understanding, *Dianmo Sheng, Dongdong Chen, Zhentao Tan, Qiankun Liu, Qi Chu, Jianmin Bao, Tao Gong, Bin Liu, Shengwei Xu, Nenghai Yu*
- 360 Mask4Align: Aligned Entity Prompting with Color Masks for Multi-Entity Localization Problems, *Haoquan Zhang, Ronggang Huang, Yi Xie, Huidong Zhang*
- 361 SOK-Bench: A Situated Video Reasoning Benchmark with Aligned Open-World Knowledge, *Andong Wang, Bo Wu, Sunli Chen, Zhenfang Chen, Haotian Guan, Wei-Ning Lee, Li Erran Li, Chuang Gan*
- 362 Align and Aggregate: Compositional Reasoning with Video Alignment and Answer Aggregation for Video Question-Answering, *Zhaohe Liao, Jiangtong Li, Li Niu, Liqing Zhang*
- 363 Segment and Caption Anything, *Xiaoke Huang, Jianfeng Wang, Yansong Tang, Zheng Zhang, Han Hu, Jiwen Lu, Lijuan Wang, Zicheng Liu*
- 364 OPERA: Alleviating Hallucination in Multi-Modal Large Language Models via Over-Trust Penalty and Retrospection-Allocation, *Qidong Huang, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, Nenghai Yu*
- 365 Learning by Correction: Efficient Tuning Task for Zero-Shot Generative Vision-Language Reasoning, *Rongjie Li, Yu Wu, Xuming He*
- 366 Revisiting Counterfactual Problems in Referring Expression Comprehension, *Zhihan Yu, Ruifan Li*
- 367 ScanFormer: Referring Expression Comprehension by Iteratively Scanning, *Wei Su, Peihan Miao, Huanzhang Dou, Xi Li*
- 368 See Say and Segment: Teaching LMMs to Overcome False Premises, *Tsung-Han Wu, Giscard Biamby, David Chan, Lisa Dunlap, Ritwik Gupta, Xudong Wang, Joseph E. Gonzalez, Trevor Darrell*
- 369 SignGraph: A Sign Sequence is Worth Graphs of Nodes, *Shiwei Gan, Yafeng Yin, Zhiwei Jiang, Hongkai Wen, Lei Xie, Sanglu Lu*
- 370 Enhancing Vision-Language Pre-training with Rich Supervisions, *Yuan Gao, Kunyu Shi, Pengkai Zhu, Edouard Belval, Oren Nuriel, Srikar Appalaraju, Shabnam Ghadar, Zhuowen Tu, Vijay Mahadevan, Stefano Soatto*
- 371 De-Diffusion Makes Text a Strong Cross-Modal Interface, *Chen Wei, Chenxi Liu, Siyuan Qiao, Zhishuai Zhang, Alan Yuille, Jiahui Yu*
- 372 MA-LMM: Memory-Augmented Large Multimodal Model for Long-Term Video Understanding, *Bo He, Hengduo Li, Young Kyun Jang, Menglin Jia, Xuefei Cao, Ashish Shah, Abhinav Srivastava, Ser-Nam Lim*
- 373 Incorporating Geo-Diverse Knowledge into Prompting for Increased Geographical Robustness in Object Recognition, *Kyle Buettner, Sina Malakouti, Xiang Lorraine Li, Adriana Kovashka*
- 374 Retrieval-Augmented Egocentric Video Captioning, *Jilan Xu, Yifei Huang, Junlin Hou, Guo Chen, Yuejie Zhang, Rui Feng, Weidi Xie*
- 375 Towards Better Vision-Inspired Vision-Language Models, *Yun-Hao Cao, Kaixiang Ji, Ziyuan Huang, Chuanyang Zheng, Jiajia Liu, Jian Wang, Jingdong Chen, Ming Yang*
- 376 PIN: Positional Insert Unlocks Object Localisation Abilities in VLMs, *Michael Dorkenwald, Nimrod Barazani, Cees G. M. Snoek, Yuki M. Asano*
- 377 Polos: Multimodal Metric Learning from Human Feedback for Image Captioning, *Yuiga Wada, Kanta Kaneda, Daichi Saito, Komei Sugijura*
- 378 Siamese Learning with Joint Alignment and Regression for Weakly-Supervised Video Paragraph Grounding, *Chaolei Tan, Jianhuang Lai, Wei-Shi Zheng, Jian-Fang Hu*
- 379 Koala: Key Frame-Conditioned Long Video-LLM, *Reuben Tan, Ximeng Sun, Ping Hu, Jui-hsien Wang, Hanieh Deilamsalehy, Bryan A. Plummer, Bryan Russell, Kate Saenko*
- 380 Generating Enhanced Negatives for Training Language-Based Object Detectors, *Shiyu Zhao, Long Zhao, Vijay Kumar B G, Yumin Suh, Dimitris N. Metaxas, Manmohan Chandraker, Samuel Schulter*
- 381 Non-autoregressive Sequence-to-Sequence Vision-Language Models, *Kunyu Shi, Qi Dong, Luis Goncalves, Zhuowen Tu, Stefano Soatto*
- 382 MMMU: A Massive Multi-discipline Multimodal Understanding and Reasoning Benchmark for Expert AGI, *Xiang Yue, Yuansheng Ni, Kai Zhang, Tianyu Zheng, Ruoqi Liu, Ge Zhang, Samuel Stevens, Dongfu Jiang, Weiming Ren, Yuxuan Sun, Cong Wei, Botao Yu, Ruibin Yuan, Renliang Sun, Ming Yin, Boyuan Zheng, Zhenzhu Yang, Yibo Liu, Wenhao Huang, Huan Sun, Yu Su, Wenhu Chen*
- 383 Synthesize Step-by-Step: Tools Templates and LLMs as Data Generators for Reasoning-Based Chart VQA, *Zhuowan Li, Bhavan Jasani, Peng Tang, Shabnam Ghadar*
- 384 Towards Learning a Generalist Model for Embodied Navigation, *Duo Zheng, Shijia Huang, Lin Zhao, Yiwu Zhong, Liwei Wang*
- 385 Previously on ... From Recaps to Story Summarization, *Aditya Kumar Singh, Dhruv Srivastava, Makarand Tapaswi*
- 386 MM-Narrator: Narrating Long-form Videos with Multimodal In-Context Learning, *Chaoyi Zhang, Kevin Lin, Zhengyuan Yang, Jianfeng Wang, Linjie Li, Chung-Ching Lin, Zicheng Liu, Lijuan Wang*
- 387 BT-Adapter: Video Conversation is Feasible Without Video Instruction Tuning, *Ruyang Liu, Chen Li, Yixiao Ge, Thomas H. Li, Ying Shan, Ge Li*
- 388 Holistic Autonomous Driving Understanding by Bird's-Eye-View Injected Multi-Modal Large Models, *Xinpeng Ding, Jianhua Han, Hang Xu, Xiaodan Liang, Wei Zhang, Xiaomeng Li*
- 389 Situational Awareness Matters in 3D Vision Language Reasoning, *Yunze Man, Liang-Yan Gui, Yu-Xiong Wang*
- 390 Eyes Wide Shut? Exploring the Visual Shortcomings of Multimodal LLMs, *Shengbang Tong, Zhuang Liu, Yuexiang Zhai, Yi Ma, Yann LeCun, Saining Xie*
- 391 SRTube: Video-Language Pre-Training with Action-Centric Video Tube Features and Semantic Role Labeling, *Ju-Hee Lee, Je-Won Kang*
- 392 Chat-UniVi: Unified Visual Representation Empowers Large Language Models with Image and Video Understanding, *Peng Jin, Ryuichi Takanobu, Wancai Zhang, Xiaochun Cao, Li Yuan*
- 393 Curriculum Point Prompting for Weakly-Supervised Referring Image Segmentation, *Qiyuan Dai, Sibe Yang*
- 394 Tune-An-Ellipse: CLIP Has Potential to Find What You Want, *Jinheng Xie, Songhe Deng, Bing Li, Haozhe Liu, Yawen Huang, Yefeng Zheng, Jurgen Schmidhuber, Bernard Ghanem, Linlin Shen, Mike Zheng Shou*
- 395 EVCap: Retrieval-Augmented Image Captioning with External Visual-Name Memory for Open-World Comprehension, *Jiaxuan Li, Duc Minh Vo, Akihiro Sugimoto, Hideki Nakayama*
- 396 Plug-and-Play Diffusion Distillation, *Yi-Ting Hsiao, Siavash Khodadadeh, Kevin Duarte, Wei-An Lin, Hui Qu, Mingi Kwon, Rathaesh Kalarot*
- 397 Lookahead Exploration with Neural Radiance Representation for Continuous Vision-Language Navigation, *Zihan Wang, Xiangyang Li, Jiahao Yang, Yeqi Liu, Junjie Hu, Ming Jiang, Shuqiang Jiang*
- 398 Low-Rank Approximation for Sparse Attention in Multi-Modal LLMs, *Lin Song, Yukang Chen, Shuai Yang, Xiaohan Ding, Yixiao Ge, Ying-Cong Chen, Ying Shan*
- 399 Contrasting Intra-Modal and Ranking Cross-Modal Hard Negatives to Enhance Visio-Linguistic Compositional Understanding, *Le Zhang, Rabiul Awal, Aishwarya Agrawal*
- 400 Iterated Learning Improves Compositionality in Large Vision-Language Models, *Chenhao Zheng, Jieyu Zhang, Aniruddha Kembhavi, Ranjay Krishna*
- 401 RegionGPT: Towards Region Understanding Vision Language Model, *Qiushan Guo, Shalini De Mello, Hongxu Yin, Wonmin Byeon, Ka Chun Cheung, Yizhou Yu, Ping Luo, Sifei Liu*
- 402 RLHF-V: Towards Trustworthy MLLMs via Behavior Alignment from Fine-grained Correctional Human Feedback, *Tianyu Yu, Yuan Yao, Haoye Zhang, Taiwan He, Yifeng Han, Ganqu Cui, Jinyi Hu, Zhiyuan Liu, Hai-Tao Zheng, Maosong Sun, Tat-Seng Chua*
- 403 Honeybee: Locality-enhanced Projector for Multimodal LLM, *Junbum Cha, Wooyoung Kang, Jonghwan Mun, Byungseok Roh*
- 404 E-GPS: Explainable Geometry Problem Solving via Top-Down Solver and Bottom-Up Generator, *Wenjun Wu, Lingling Zhang, Jun Liu, Xi Tang, Yaxian Wang, Shaowei Wang, Qianying Wang*
- 405 Pink: Unveiling the Power of Referential Comprehension for Multimodal LLMs, *Shiyu Xuan, Qingpei Guo, Ming Yang, Shiliang Zhang*
- 406 Any-Shift Prompting for Generalization over Distributions, *Zehao Xiao, Jiayi Shen, Mohammad Mahdi Derakhshani, Shengcai Liao, Cees G. M. Snoek*

- 407 Question Aware Vision Transformer for Multimodal Reasoning,
* Roy Ganz, Yair Kittenplon, Aviad Aberdam, Elad Ben Avraham, Oren Nuriel, Shai Mazor, Ron Litman
- 408 Mitigating Object Hallucinations in Large Vision-Language
* Models through Visual Contrastive Decoding, Sicong Leng, Hang Zhang, Guanzheng Chen, Xin Li, Shijian Lu, Chunyan Miao, Lidong Bing
- 409 Text-Image Alignment for Diffusion-Based Perception, Neehar Kondapaneni, Markus Marks, Manuel Knott, Rogerio Guimaraes, Pietro Perona
- 410 Do You Remember? Dense Video Captioning with Cross-Modal Memory Retrieval, Minkuk Kim, Hyeon Bae Kim, Jinyoung Moon, Jinwoo Choi, Seong Tae Kim
- 411 FairDeDup: Detecting and Mitigating Vision-Language Fairness Disparities in Semantic Dataset Deduplication, Eric Slyman, Stefan Lee, Scott Cohen, Kushal Kafle
- 412 G³-LQ: Marrying Hyperbolic Alignment with Explicit Semantic-Geometric Modeling for 3D Visual Grounding, Yuan Wang, Yali Li, Shengjin Wang
- 413 LISA: Reasoning Segmentation via Large Language Model,
* Xin Lai, Zhuotao Tian, Yukang Chen, Yanwei Li, Yuhui Yuan, Shu Liu, Jiaya Jia
- 414 VideoCon: Robust Video-Language Alignment via Contrast Captions, Hritik Bansal, Yonatan Bitton, Idan Szepkter, Kai-Wei Chang, Aditya Grover
- 415 Taming Self-Training for Open-Vocabulary Object Detection, Shiyu Zhao, Samuel Schuster, Long Zhao, Zhixing Zhang, Vijay Kumar B G, Yumin Suh, Manmohan Chandraker, Dimitris N. Metaxas
- 416 SyncMask: Synchronized Attentional Masking for Fashion-centric Vision-Language Pretraining, Chull Hwan Song, Taebaek Hwang, Jooyoung Yoon, Shunghyun Choi, Yeong Hyeon Gu
- 417 Generative Region-Language Pretraining for Open-Ended Object Detection, Chuang Lin, Yi Jiang, Lizhen Qu, Zehuan Yuan, Jianfei Cai
- 418 CoG-DQA: Chain-of-Guiding Learning with Large Language Models for Diagram Question Answering, Shaowei Wang, Lingling Zhang, Longji Zhu, Tao Qin, Kim-Hui Yap, Xinyu Zhang, Jun Liu
- 419 Multi-modal Instruction Tuned LLMs with Fine-grained Visual
* Perception, Junwen He, Yifan Wang, Lijun Wang, Huchuan Lu, Jun-Yan He, Jin-Peng Lan, Bin Luo, Xuansong Xie
- 420 Generate Subgoal Images before Act: Unlocking the Chain-of-Thought Reasoning in Diffusion Model for Robot Manipulation with Multimodal Prompts, Fei Ni, Jianye Hao, Shiguang Wu, Longxin Kou, Jiashun Liu, Yan Zheng, Bin Wang, Yuzheng Zhuang
- 421 LoSh: Long-Short Text Joint Prediction Network for Referring Video Object Segmentation, Linfeng Yuan, Miaojing Shi, Zijie Yue, Qijun Chen
- 422 MICap: A Unified Model for Identity-Aware Movie Descriptions, Haran Raajesh, Naveen Reddy Desanur, Zeeshan Khan, Makarand Tapaswi
- 423 CapsFusion: Rethinking Image-Text Data at Scale, Qiyang Yu, Quan Sun, Xiaosong Zhang, Yufeng Cui, Fan Zhang, Yue Cao, Xinlong Wang, Jingjing Liu
- 424 Visual Fact Checker: Enabling High-Fidelity Detailed Caption Generation, Yunhao Ge, Xiaohui Zeng, Jacob Samuel Huffman, Tsung-Yi Lin, Ming-Yu Liu, Yin Cui
- 425 VidLA: Video-Language Alignment at Scale, Mamshad Nayeem Rizve, Fan Fei, Jayakrishnan Unnikrishnan, Son Tran, Benjamin Z. Yao, Belinda Zeng, Mubarak Shah, Trishul Chilimbi
- 426 Viewpoint-Aware Visual Grounding in 3D Scenes, Xiangxi Shi, Zhonghua Wu, Stefan Lee
- 427 Multi-Modal Proxy Learning Towards Personalized Visual Multiple Clustering, Jiawei Yao, Qi Qian, Juhua Hu
- 428 Jack of All Tasks Master of Many: Designing General-Purpose
* Coarse-to-Fine Vision-Language Model, Shraman Pramanick, Guangxing Han, Rui Hou, Sayan Nag, Ser-Nam Lim, Nicolas Ballas, Qifan Wang, Rama Chellappa, Amjad Almahairi
- 429 LLaMA-Excitor: General Instruction Tuning via Indirect Feature Interaction, Bo Zou, Chao Yang, Yu Qiao, Chengbin Qian, Youjian Zhao
- 430 MeaCap: Memory-Augmented Zero-shot Image Captioning, Zequn Zeng, Yan Xie, Hao Zhang, Chiyu Chen, Bo Chen, Zhengjue Wang
- 431 The STVChrono Dataset: Towards Continuous Change Recognition in Time, Yanjun Sun, Yue Qiu, Mariia Khan, Fumiya Matsuzawa, Kenji Iwata
- 432 InstaGen: Enhancing Object Detection by Training on Synthetic Dataset, Chengjian Feng, Yujie Zhong, Zequn Jie, Weidi Xie, Lin Ma
- 433 MiKASA: Multi-Key-Anchor & Scene-Aware Transformer for 3D Visual Grounding, Chun-Peng Chang, Shaoxiang Wang, Alain Pagan, Didier Stricker
- 434 Investigating Compositional Challenges in Vision-Language
* Models for Visual Grounding, Yunan Zeng, Yan Huang, Jinjin Zhang, Zequn Jie, Zhenhua Chai, Liang Wang
- 435 Masked AutoDecoder is Effective Multi-Task Vision Generalist, Han Qiu, Jiaying Huang, Peng Gao, Lewei Lu, Xiaoqin Zhang, Shijian Lu
- 436 Efficient Test-Time Adaptation of Vision-Language Models, Adilbek Karmanov, Dayan Guan, Shijian Lu, Abdulmotaleb El Saddik, Eric Xing
- 437 FFF: Fixing Flawed Foundations in Contrastive Pre-Training Results in Very Strong Vision-Language Models, Adrian Bulat, Yassine Ouali, Georgios Tzimiropoulos
- 438 Open3DSG: Open-Vocabulary 3D Scene Graphs from Point Clouds with Queryable Objects and Open-Set Relationships, Sebastian Koch, Narunas Vaskevicius, Mirco Colosi, Pedro Hermosilla, Timo Ropinski
- 439 Instance-level Expert Knowledge and Aggregate Discriminative Attention for Radiology Report Generation, Shenshen Bu, Taiji Li, Yuedong Yang, Zhiming Dai
- 440 Omni-SMoLA: Boosting Generalist Multimodal Models with Soft
* Mixture of Low-rank Experts, Jialin Wu, Xia Hu, Yaqing Wang, Bo Pang, Radu Soricut
- 441 Building Vision-Language Models on Solid Foundations with Masked Distillation, Sepehr Sameni, Kushal Kafle, Hao Tan, Simon Jenni
- 442 GROUNDHOG: Grounding Large Language Models to Holistic Segmentation, Yichi Zhang, Ziqiao Ma, Xiaofeng Gao, Suhaila Shakiah, Qiaozhi Gao, Joyce Chai
- 443 Visual Program Distillation: Distilling Tools and Programmatic
* Reasoning into Vision-Language Models, Yushi Hu, Otilia Stretcu, Chun-Ta Lu, Krishnamurthy Viswanathan, Kenji Hata, Enming Luo, Ranjay Krishna, Ariel Fuxman
- 444 DRESS: Instructing Large Vision-Language Models to Align and Interact with Humans via Natural Language Feedback, Yangyi Chen, Karan Sikka, Michael Cogswell, Heng Ji, Ajay Divakaran
- 445 LASO: Language-guided Affordance Segmentation on 3D Object, Yicong Li, Na Zhao, Junbin Xiao, Chun Feng, Xiang Wang, Tat-seng Chua
- 446 Omni-Q: Omni-Directional Scene Understanding for Unsupervised Visual Grounding, Sai Wang, Yutian Lin, Yu Wu
- 447 VTimeLLM: Empower LLM to Grasp Video Moments,
* Bin Huang, Xin Wang, Hong Chen, Zihan Song, Wenwu Zhu
- 448 CogAgent: A Visual Language Model for GUI Agents,
* Wenyi Hong, Weihang Wang, Qingsong Lv, Jiazheng Xu, Wenmeng Yu, Junhui Ji, Yan Wang, Zihan Wang, Yuxiao Dong, Ming Ding, Jie Tang
- 449 EgoThink: Evaluating First-Person Perspective Thinking
* Capability of Vision-Language Models, Sijie Cheng, Zhicheng Guo, Jingwen Wu, Kechen Fang, Peng Li, Huaping Liu, Yang Liu
- 450 Multi-Modal Hallucination Control by Visual Information Grounding, Alessandro Favero, Luca Zancato, Matthew Trager, Siddharth Choudhary, Pramuditha Perera, Alessandro Achille, Ashwin Swaminathan, Stefano Soatto
- 451 TimeChat: A Time-sensitive Multimodal Large Language Model for Long Video Understanding, Shuhuai Ren, Linli Yao, Shicheng Li, Xu Sun, Lu Hou
- 452 AHIVE: Anatomy-aware Hierarchical Vision Encoding for Interactive Radiology Report Retrieval, Sixing Yan, William K. Cheung, Ivor W. Tsang, Keith Chiu, Terence M. Tong, Ka Chun Chung, Simon See
- 453 Do Vision and Language Encoders Represent the World Similarly?, Mayug Maniparambil, Raiymbek Akshulakov, Yasser Abdelaziz Dahou Djilali, Mohamed El Amine Seddik, Sanath Narayan, Karttikeya Mangalam, Noel E. O'Connor
- 454 Self-Training Large Language Models for Improved Visual

- Program Synthesis With Visual Reinforcement, *Zaid Khan, Vijay Kumar BG, Samuel Schulter, Yun Fu, Manmohan Chandraker*
- 455 Composing Object Relations and Attributes for Image-Text Matching, *Khoi Pham, Chuong Huynh, Ser-Nam Lim, Abhinav Shrivastava*
- 456 Zero-shot Referring Expression Comprehension via Structural Similarity Between Images and Captions, *Zeyu Han, Fangrui Zhu, Qianru Lao, Huaizu Jiang*
- 457 HallusionBench: An Advanced Diagnostic Suite for Entangled Language Hallucination and Visual Illusion in Large Vision-Language Models, *Tianrui Guan, Fluxiao Liu, Xiyang Wu, Ruiqi Xian, Zongxia Li, Xiaoyu Liu, Xijun Wang, Lichang Chen, Furong Huang, Yaser Yacoob, Dinesh Manocha, Tianyi Zhou*
- 458 A Simple Recipe for Contrastively Pre-training Video-First Encoders Beyond 16 Frames, *Pinelopi Papalampidi, Skanda Koppula, Shreya Pathak, Justin Chiu, Joe Heyward, Viorica Patraucean, Jiajun Shen, Antoine Miech, Andrew Zisserman, Aida Nematzdeh*
- 459 Generative Multimodal Models are In-Context Learners, *Quan Sun, Yufeng Cui, Xiaosong Zhang, Fan Zhang, Qiying Yu, Yuezhe Wang, Yongming Rao, Jingjing Liu, Tiejun Huang, Xinlong Wang*
- 460 A Vision Check-up for Language Models, *Pratyusha Sharma, Tamar Rott Shaham, Manel Baradad, Stephanie Fu, Adrian Rodriguez-Munoz, Shivam Duggal, Phillip Isola, Antonio Torralba*
- 461 Compositional Chain-of-Thought Prompting for Large Multimodal Models, *Chancharik Mitra, Brandon Huang, Trevor Darrell, Roei Herzig*
- 462 On Scaling Up a Multilingual Vision and Language Model, *Xi Chen, Josip Djolonga, Piotr Padlewski, Basil Mustafa, Soravit Changpinyo, Jialin Wu, Carlos Riquelme Ruiz, Sebastian Goodman, Xiao Wang, Yi Tay, Siamak Shakeri, Mostafa Dehghani, Daniel Salz, Mario Lucic, Michael Tschannen, Arsha Nagrani, Hexiang Hu, Mandar Joshi, Bo Pang, Ceslee Montgomery, Paulina Pietrzyk, Marvin Ritter, AJ Piergiovanni, Matthias Minderer, Filip Pavetic, Austin Waters, Gang Li, Ibrahim Alabdulmohsin, Lucas Beyer, Julien Amelot, Kenton Lee, Andreas Peter Steiner, Yang Li, Daniel Keysers, Anurag Arnab, Yuanzhong Xu, Keran Rong, Alexander Kolesnikov, Mojtaba Seyedhosseini, Anelia Angelova, Xiaohua Zhai, Neil Houlsby, Radu Soricut*
- 463 Dual-View Visual Contextualization for Web Navigation, *Jihyung Kil, Chan Hee Song, Boyuan Zheng, Xiang Deng, Yu Su, Wei-Lun Chao*
- 464 SpatialVLM: Endowing Vision-Language Models with Spatial Reasoning Capabilities, *Boyuan Chen, Zhuo Xu, Sean Kirmani, Brain Ichter, Dorsa Sadigh, Leonidas Guibas, Fei Xia*
- 465 Beyond Seen Primitive Concepts and Attribute-Object Compositional Learning, *Nirat Saini, Khoi Pham, Abhinav Shrivastava*

10:30 - 18:45 Art Program (Arch 4CDE)**10:30 - 18:45 DEMOS** (Arch 4CDE)

- 1 Infinite-ISP - An Adaptive soft ISP solution for Diverse Image processing Applications, *Sohaib Imran Bhatti, Bilal Zafar, Muhammad Abdullah*
- 2 Interactive Image Segmentation Guided by Visual Prompting, *Thomas Frick, Cezary Skura, Filip M. Janicki, Roy Assaf, Niccolo Avogaro, Daniel Caraballo, Yagmur G. Cinar, Brown Ebouky, Ioana Giurgiu, Takayuki Katsuki, Piotr Kluska, Cristiano Malossi, Haoxiang Qiu, Tomoya Sakai, Florian Scheidegger, Andrej Simeski, Daniel Yang, Andrea Bartezzaghi, Mattia Rigotti*
- 3 LipDub The Inclusive Translation Canvas, *Amogh Subbakrishna Adishesha, Stanislaw Beliasau*
- 4 A Live Demo of Single-Photon Imaging and Applications, *Varun Sundar, Sacha Jungerman, Mohit Gupta*
- 5 High Speed In-Pixel Feature Tracking, *Laurie Bose*
- 6 CLIP-Interpret: An interpretability Tool for CLIP-like Models, *Avinash Madasu, Yossi Gandelsman, Vasudev Lal, Phillip Howard*
- 7 WildVision Arena: Benchmarking Multimodal LLMs in the Wild, *Yujie Lu, Dongfu Jiang, Wenhui Chen, William Wang, Yejin Choi, Bill Yuchen Lin*
- 8 PathChat, an interactive vision-language AI assistant for human pathology, *Ming Y. (Max) Lu, Richard J. Chen*
- 9 Multimodal Video Understanding System for ESCS Assessment Context, *Jitesh Jain*, Christy Yoon*, Kaori Terol Aveiro*, Hedda Meadan, Jinjun Xiong, Humphrey Shi*

- 10 Co-operate Sign: learning sign language with real-time feedback on laptop, *Yuting Peng, Yuecong Min, Xilin Chen*
- 11 AI3D Desktop, *Yosun Chang*
- 12 Gaussian Splatting SLAM, *Hidenobu Matsuki, Riku Murai*
- 13 Robust depth perception through Virtual Pattern Projection, *Luca Bartolomei, Matteo Poggi, Fabio Tosi, Andrea Conti, Stefano Mattoccia*
- 14 Generating Emotional 3D Talking Heads from Speech, *Federico Nocentini, Claudio Ferrari, Stefano Berretti*
- 15 MINTest Demo, *Daniel DeAlcala, Aythami Morales, Gonzalo Mancera, Julian Fierrez, Ruben Tolosana, Javier Ortega-Garcia*
- 16 Analysis of 3D Pathology Samples using Weakly Supervised AI, *Andrew H. Song, Faisal Mahmood*
- 17 PIGEON: Predicting Image Geolocations in GeoGuessr, *Lukas Haas, Michal Skreta, Silas Alberti*

11:30 - 13:30 Doctoral Consortium (Summit Terrace Suite)

- Yufei Ye, *Carnegie Mellon University*
- Tarun Kalluri, *UC San Diego*
- Zhen Dong, *University of California, Berkeley*
- Matthew Kowal, *York University*
- Ms. Yunhua Zhang, *University of Amsterdam*
- Hyung-gun Chi, *Purdue University*
- Zuoyue Li, *ETH Zurich*
- Van Nguyen Nguyen, *Ecole des Ponts ParisTech, France*

Supported by:

U.S. National
Science
Foundation

- Shivangi Aneja, *Technical University of Munich*
- Wentao Bao, *Michigan State University*
- Zhengfeng Lai, *University of California, Davis*
- Inhwon Bae, *Gwangju Institute of Science and Technology*
- Supreeth Narasimhaswamy, *Stony Brook University*
- Jiawei Ma, *Columbia University*
- Shuyang Sun, *University of Oxford*
- Anna Kukleva, *Max-Planck-Institute for Informatics*
- Gowthami Somepalli, *University of Maryland, College Park*
- Siwei Zhang, *ETH Zurich*
- Tz-Ying Wu, *University of California San Diego*
- Thanh-Dat Truong, *University of Arkansas*
- Changhoon Kim, *Arizona State University*
- Vaibhav Vavilala, *University of Illinois at Urbana-Champaign*
- Jason Y. Zhang, *Carnegie Mellon University*
- Harsh Rangwani, *Indian Institute of Science, Bangalore*
- Xizi Wang, *Indiana university bloomington*
- Chen Wei, *Johns Hopkins University*
- Tan Wang, *Nanyang Technological University*
- Yi-Wen Chen, *University of California, Merced*
- Seonguk Seo, *Seoul National University*
- Yining Hong, *UCLA*
- Hanjing Wang, *Rensselaer Polytechnic Institute*
- Sheng Cheng, *Arizona State University*
- Shraman Pramanick, *Johns Hopkins University*
- Navaneet K L, *University of California, Davis*
- Lahav Lipson, *Princeton University*
- Jihyung Kil, *The Ohio State University*

12:00 - 14:00 LUNCH (Summit ExHall 1-2)**12:00 - 13:00 SOCIAL - Student Speed Mentorship Session** (Summit Elliott Bay Room)

**Reservation Required

13:00 - 14:30 Orals 4A: Autonomous Navigation and Egocentric Vision (Summit Ballroom)

- 1 SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection, *Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, Xiaolin Hu*
- 2 UnO: Unsupervised Occupancy Fields for Perception and Forecasting, *Ben Agro, Quinlan Sykora, Sergio Casas, Thomas Gilles, Raquel Urtasun*

- 3 EgoGen: An Egocentric Synthetic Data Generator, *Gen Li, Kaifeng Zhao, Siwei Zhang, Xiaozhong Lyu, Mihai Dusmanu, Yan Zhang, Marc Pollefeys, Siyu Tang*
- 4 Learning to Segment Referred Objects from Narrated Egocentric Videos, *Yuhan Shen, Huiyu Wang, Xitong Yang, Matt Feiszli, Ehsan Elhamifar, Lorenzo Torresani, Effrosyni Mavroudi*
- 5 Producing and Leveraging Online Map Uncertainty in Trajectory Prediction, *Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, Boris Ivanovic*

13:00 - 14:30 Orals 4B: 3D Vision (Summit Flex Hall AB)

- 1 SceneFun3D: Fine-Grained Functionality and Affordance Understanding in 3D Scenes, *Alexandros Delitzas, Ayca Takmaz, Federico Tombari, Robert Sumner, Marc Pollefeys, Francis Engelmann*
- 2 SpiderMatch: 3D Shape Matching with Global Optimality and Geometric Consistency, *Paul Roetzer, Florian Bernard*
- 3 PaSCo: Urban 3D Panoptic Scene Completion with Uncertainty Awareness, *Anh-Quan Cao, Angela Dai, Raoul de Charette*
- 4 PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar, *Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan*
- 5 A Subspace-Constrained Tyler's Estimator and its Applications to Structure from Motion, *Feng Yu, Teng Zhang, Gilad Lerman*

13:00 - 14:30 Orals 4C: Action and Motion (Summit Flex Hall C)

- 1 Modeling Multimodal Social Interactions: New Challenges and Baselines with Densely Aligned Representations, *Sangmin Lee, Bolin Lai, Fiona Ryan, Bikram Boote, James M. Rehg*
- 2 An N-Point Linear Solver for Line and Motion Estimation with Event Cameras, *Ling Gao, Daniel Gehrig, Hang Su, Davide Scaramuzza, Laurent Kneip*
- 3 RoHM: Robust Human Motion Reconstruction via Diffusion, *Siwei Zhang, Bharat Lal Bhatnagar, Yuanlu Xu, Alexander Winkler, Petr Kadlecek, Siyu Tang, Federica Bogo*
- 4 Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation, *Ming Xu, Stephen Gould*
- 5 FineParser: A Fine-grained Spatio-temporal Action Parser for Human-centric Action Quality Assessment, *Jinglin Xu, Sibao Yin, Guohao Zhao, Zishuo Wang, Yuxin Peng*

14:30 - 14:45 Courtesy Break

- 14:45 - 15:45 **KEYNOTE 2 - David Baker, Director of Institute for Protein Design** Henrietta and Aubrey Davis Endowed Professor in Biochemistry, University of Washington (Summit Flex Hall ABC)

15:45 - 16:00 Courtesy Break

16:00 - 17:00 PAMI TC Meeting (Summit Flex Hall ABC)

16:45 - 17:15 Poster Setup (Arch 4E)

17:00 - 19:00 SOCIAL - CV Entrepreneurship – Founders, Freelancers & Friends (Summit Terrace Suite & Lounge)

17:15 - 18:45 Poster Session 4 & Exhibit Hall (Arch 4A-E)

- 1 PaSCo: Urban 3D Panoptic Scene Completion with Uncertainty Awareness, *Anh-Quan Cao, Angela Dai, Raoul de Charette*
- 2 PlatoNeRF: 3D Reconstruction in Plato's Cave via Single-View Two-Bounce Lidar, *Tzofi Klinghoffer, Xiaoyu Xiang, Siddharth Somasundaram, Yuchen Fan, Christian Richardt, Ramesh Raskar, Rakesh Ranjan*
- 3 Self-Supervised Class-Agnostic Motion Prediction with Spatial and Temporal Consistency Regularizations, *Kewei Wang, Yizheng Wu, Jun Cen, Zhiyu Pan, Xingyi Li, Zhe Wang, Zhiguo Cao, Guosheng Lin*
- 4 Multi-Space Alignments Towards Universal LiDAR Segmentation, *Youquan Liu, Lingdong Kong, Xiaoyang Wu, Runnan Chen, Xin Li, Liang Pan, Ziwei Liu, Yuexin Ma*
- 5 Generalized Predictive Model for Autonomous Driving, *Jiazhi Yang, Shenyuan Gao, Yihang Qiu, Li Chen, Tianyu Li, Bo Dai, Kashyap Chitta, Penghao Wu, Jia Zeng, Ping Luo, Jun Zhang, Andreas Geiger, Yu Qiao, Hongyang Li*
- 6 Visual Point Cloud Forecasting enables Scalable Autonomous Driving, *Zetong Yang, Li Chen, Yanan Sun, Hongyang Li*

- 7 SeMoLi: What Moves Together Belongs Together, *Jenny Seidenschwarz, Aljosa Osep, Francesco Ferroni, Simon Lucey, Laura Leal-Taixe*
- 8 AIDE: An Automatic Data Engine for Object Detection in Autonomous Driving, *Mingfu Liang, Jong-Chyi Su, Samuel Schuster, Sparsh Garg, Shiyu Zhao, Ying Wu, Manmohan Chandraker*
- 9 Dynamic Adapter Meets Prompt Tuning: Parameter-Efficient Transfer Learning for Point Cloud Analysis, *Xin Zhou, Dingkang Liang, Wei Xu, Xingkui Zhu, Yihan Xu, Zhikang Zou, Xiang Bai*
- 10 BEVspread: Spread Voxel Pooling for Bird's-Eye-View Representation in Vision-based Roadside 3D Object Detection, *Wenjie Wang, Yehao Lu, Guangcong Zheng, Shuigen Zhan, Xiaoqing Ye, Zichang Tan, Jingdong Wang, Gaoang Wang, Xi Li*
- 11 DualAD: Disentangling the Dynamic and Static World for End-to-End Driving, *Simon Doll, Niklas Hanselmann, Lukas Schneider, Richard Schulz, Marius Cordts, Markus Enzweiler, Hendrik P. A. Lensch*
- 12 Towards Realistic Scene Generation with LiDAR Diffusion Models, *Haoxi Ran, Vitor Guizilini, Yue Wang*
- 13 Driving into the Future: Multiview Visual Forecasting and Planning with World Model for Autonomous Driving, *Yuqi Wang, Jiawei He, Lue Fan, Hongxin Li, Yuntao Chen, Zhaoxiang Zhang*
- 14 VLP: Vision Language Planning for Autonomous Driving, *Chenbin Pan, Burhaneddin Yaman, Tommaso Nesti, Abhirup Mallik, Alessandro G Allievi, Senem Velipasalar, Liu Ren*
- 15 Scaling Diffusion Models to Real-World 3D LiDAR Scene Completion, *Lucas Nunes, Rodrigo Marcuzzi, Benedikt Mersch, Jens Behley, Cyrill Stachniss*
- 16 UniMix: Towards Domain Adaptive and Generalizable LiDAR Semantic Segmentation in Adverse Weather, *Haimei Zhao, Jing Zhang, Zhuo Chen, Shanshan Zhao, Dacheng Tao*
- 17 Not All Voxels Are Equal: Hardness-Aware Semantic Scene Completion with Self-Distillation, *Song Wang, Jiawei Yu, Wentong Li, Wenyu Liu, Xiaolu Liu, Junbo Chen, Jianke Zhu*
- 18 OOSTraj: Out-of-Sight Trajectory Prediction With Vision-Positioning Denoising, *Haichao Zhang, Yi Xu, Hongsheng Lu, Takayuki Shimizu, Yun Fu*
- 19 MGMap: Mask-Guided Learning for Online Vectorized HD Map Construction, *Xiaolu Liu, Song Wang, Wentong Li, Ruizi Yang, Junbo Chen, Jianke Zhu*
- 20 Density-Adaptive Model Based on Motif Matrix for Multi-Agent Trajectory Prediction, *Di Wen, Haoran Xu, Zhaocheng He, Zhe Wu, Guang Tan, Peixi Peng*
- 21 StreamingFlow: Streaming Occupancy Forecasting with Asynchronous Multi-modal Data Streams via Neural Ordinary Differential Equation, *Yining Shi, Kun Jiang, Ke Wang, Jiusi Li, Yunlong Wang, Mengmeng Yang, Diange Yang*
- 22 SAFDNet: A Simple and Effective Network for Fully Sparse 3D Object Detection, *Gang Zhang, Junnan Chen, Guohuan Gao, Jianmin Li, Si Liu, Xiaolin Hu*
- 23 View From Above: Orthogonal-View aware Cross-view Localization, *Shan Wang, Chuong Nguyen, Jiawei Liu, Yanhao Zhang, Sundaram Muthu, Fahira Afzal Maken, Kaihao Zhang, Hongdong Li*
- 24 Improving Distant 3D Object Detection Using 2D Box Supervision, *Zetong Yang, Zhiding Yu, Chris Choy, Renhao Wang, Anima Anandkumar, Jose M. Alvarez*
- 25 Is Ego Status All You Need for Open-Loop End-to-End Autonomous Driving?, *Zhiqi Li, Zhiding Yu, Shiyi Lan, Jiahao Li, Jan Kautz, Tong Lu, Jose M. Alvarez*
- 26 CaDeT: a Causal Disentanglement Approach for Robust Trajectory Prediction in Autonomous Driving, *Mozhgan Pourkeshavarz, Junrui Zhang, Amir Rasouli*
- 27 Adversarial Backdoor Attack by Naturalistic Data Poisoning on Trajectory Prediction in Autonomous Driving, *Mozhgan Pourkeshavarz, Mohammad Sabokrou, Amir Rasouli*
- 28 NeuRAD: Neural Rendering for Autonomous Driving, *Adam Tonderski, Carl Lindström, Georg Hess, William Ljungbergh, Lennart Svensson, Christoffer Petersson*
- 29 IS-Fusion: Instance-Scene Collaborative Fusion for Multimodal 3D Object Detection, *Junbo Yin, Jianbing Shen, Runnan Chen, Wei Li, Ruigang Yang, Pascal Frossard, Wenguan Wang*

- 30 LSK3DNet: Towards Effective and Efficient 3D Perception with Large Sparse Kernels, *Tuo Feng, Wenguan Wang, Fan Ma, Yi Yang*
- 31 RCBEVDet: Radar-camera Fusion in Bird's Eye View for 3D Object Detection, *Zhiwei Lin, Zhe Liu, Zhongyu Xia, Xinhao Wang, Yongtao Wang, Shengxiang Qi, Yang Dong, Nan Dong, Le Zhang, Ce Zhu*
- 32 PTT: Point-Trajectory Transformer for Efficient Temporal 3D Object Detection, *Kuan-Chih Huang, Weijie Lyu, Ming-Hsuan Yang, Yi-Hsuan Tsai*
- 33 Driving Everywhere with Large Language Model Policy Adaptation, *Boyi Li, Yue Wang, Jiageng Mao, Boris Ivanovic, Sushant Veer, Karen Leung, Marco Pavone*
- 34 Text2Loc: 3D Point Cloud Localization from Natural Language, *Yan Xia, Letian Shi, Zifeng Ding, Joao F. Henriques, Daniel Cremers*
- 35 Commonsense Prototype for Outdoor Unsupervised 3D Object Detection, *Hai Wu, Shijia Zhao, Xun Huang, Chenglu Wen, Xin Li, Cheng Wang*
- 36 A-Teacher: Asymmetric Network for 3D Semi-Supervised Object Detection, *Hanshi Wang, Zhipeng Zhang, Jin Gao, Weiming Hu*
- 37 MoST: Multi-Modality Scene Tokenization for Motion Prediction, *Norman Mu, Jingwei Ji, Zhenpei Yang, Nate Harada, Haotian Tang, Kan Chen, Charles R. Qi, Runzhou Ge, Kratharth Goel, Zoey Yang, Scott Ettinger, Rami Al-Rfou, Dragomir Anguelov, Yin Zhou*
- 38 Feedback-Guided Autonomous Driving,
* *Jimuyang Zhang, Zanming Huang, Arijit Ray, Eshed Ohn-Bar*
- 39 Bootstrapping Autonomous Driving Radars with Self-Supervised Learning, *Yiduo Hao, Sohrab Madani, Junfeng Guan, Mohammed Alloulah, Saurabh Gupta, Haitham Hassanieh*
- 40 UnO: Unsupervised Occupancy Fields for Perception and Forecasting,
* *Ben Agro, Quinlan Sykora, Sergio Casas, Thomas Gilles, Raquel Urtasun*
- 41 SIRA: Scalable Inter-frame Relation and Association for Radar Perception, *Ryoma Yataka, Pu Wang, Petros Boufounos, Ryuhei Takahashi*
- 42 SparseOcc: Rethinking Sparse Latent Representation for Vision-Based Semantic Occupancy Prediction, *Pin Tang, Zhongdao Wang, Guoqing Wang, Jilai Zheng, Xiangxuan Ren, Bailan Feng, Chao Ma*
- 43 DiffLoc: Diffusion Model for Outdoor LiDAR Localization, *Wen Li, Yuyang Yang, Shangshu Yu, Guosheng Hu, Chenglu Wen, Ming Cheng, Cheng Wang*
- 44 Weak-to-Strong 3D Object Detection with X-Ray Distillation, *Alexander Gambashidze, Aleksandr Dadukin, Maxim Golyadkin, Maria Razzhivina, Ilya Makarov*
- 45 T4P: Test-Time Training of Trajectory Prediction via Masked Autoencoder and Actor-specific Token Memory, *Daehee Park, Jaeseok Jeong, Sung-Hoon Yoon, Jaewoo Jeong, Kuk-Jin Yoon*
- 46 Editable Scene Simulation for Autonomous Driving via
* Collaborative LLM-Agents, *Yuxi Wei, Zi Wang, Yifan Lu, Chenxin Xu, Changxing Liu, Hao Zhao, Siheng Chen, Yanfeng Wang*
- 47 Uncertainty-Guided Never-Ending Learning to Drive, *Lei Lai, Eshed Ohn-Bar, Sanjay Arora, John Seon Keun Yi*
- 48 On the Road to Portability: Compressing End-to-End Motion Planner for Autonomous Driving, *Kaituo Feng, Changsheng Li, Dongchun Ren, Ye Yuan, Guoren Wang*
- 49 DiffFlow3D: Toward Robust Uncertainty-Aware Scene Flow Estimation with Iterative Diffusion-Based Refinement, *Jiuming Liu, Guangming Wang, Weicai Ye, Chaokang Jiang, Jinru Han, Zhe Liu, Guofeng Zhang, Dalong Du, Hesheng Wang*
- 50 LMDrive: Closed-Loop End-to-End Driving with Large Language Models, *Hao Shao, Yuxuan Hu, Letian Wang, Guanglu Song, Steven L. Waslander, Yu Liu, Hongsheng Li*
- 51 SOAC: Spatio-Temporal Overlap-Aware Multi-Sensor Calibration using Neural Radiance Fields, *Quentin Herau, Nathan Piasco, Moussab Bennehar, Luis Roldao, Dzmirty Tsishkou, Cyrille Migniot, Pascal Vasseur, Cédric Demonceaux*
- 52 LaMPilot: An Open Benchmark Dataset for Autonomous Driving with Language Model Programs, *Yunsheng Ma, Can Cui, Xu Cao, Wenqian Ye, Peiran Liu, Juanwu Lu, Amr Abdelraouf, Rohit Gupta, Kyungtae Han, Aniket Bera, James M. Rehg, Ziran Wang*
- 53 GLiDR: Topologically Regularized Graph Generative Network for Sparse LiDAR Point Clouds, *Prashant Kumar, Kshitij Madhav Bhat, Vedang Bhupesh Shenvi Nadkarni, Prem Kalra*
- 54 Towards Robust 3D Object Detection with LiDAR and 4D Radar Fusion in Various Weather Conditions, *Yujeong Chae, Hyeonseong Kim, Kuk-Jin Yoon*
- 55 3DSFLabelling: Boosting 3D Scene Flow Estimation by Pseudo Auto-labelling, *Chaokang Jiang, Guangming Wang, Jiuming Liu, Hesheng Wang, Zhuang Ma, Zhenqiang Liu, Zhujin Liang, Yi Shan, Dalong Du*
- 56 ADA-Track: End-to-End Multi-Camera 3D Multi-Object Tracking with Alternating Detection and Association, *Shuxiao Ding, Lukas Schneider, Marius Cordts, Juergen Gall*
- 57 PointBeV: A Sparse Approach for BeV Predictions, *Loick Chambon, Eloi Zablocki, Mickaël Chen, Florent Bartoccioni, Patrick Pérez, Matthieu Cord*
- 58 Light the Night: A Multi-Condition Diffusion Framework for Unpaired Low-Light Enhancement in Autonomous Driving, *Jinlong Li, Baolu Li, Zhengzhong Tu, Xinyu Liu, Qing Guo, Felix Juefei-Xu, Runsheng Xu, Hongkai Yu*
- 59 CLIP-BEVFormer: Enhancing Multi-View Image-Based BEV Detector with Ground Truth Flow, *Chenbin Pan, Burhaneddin Yaman, Senem Velipasalar, Liu Ren*
- 60 Adapting to Length Shift: FlexiLength Network for Trajectory Prediction, *Yi Xu, Yun Fu*
- 61 UniPAD: A Universal Pre-training Paradigm for Autonomous Driving, *Honghui Yang, Sha Zhang, Di Huang, Xiaoyang Wu, Haoyi Zhu, Tong He, Shixiang Tang, Hengshuang Zhao, Qibo Qiu, Binbin Lin, Xiaofei He, Wanli Ouyang*
- 62 Higher-order Relational Reasoning for Pedestrian Trajectory Prediction, *Sungjune Kim, Hyung-gun Chi, Hyerin Lim, Karthik Ramani, Jinkyu Kim, Sangpil Kim*
- 63 HPNet: Dynamic Trajectory Forecasting with Historical Prediction Attention, *Xiaolong Tang, Meina Kan, Shiguang Shan, Zhilong Ji, Jinfeng Bai, Xilin Chen*
- 64 LiSA: LiDAR Localization with Semantic Awareness,
* *Bochun Yang, Zijun Li, Wen Li, Zhipeng Cai, Chenglu Wen, Yu Zang, Matthias Muller, Cheng Wang*
- 65 SmartRefine: A Scenario-Adaptive Refinement Framework for Efficient Motion Prediction, *Yang Zhou, Hao Shao, Letian Wang, Steven L. Waslander, Hongsheng Li, Yu Liu*
- 66 Pseudo Label Refinery for Unsupervised Domain Adaptation on Cross-dataset 3D Object Detection, *Zhanwei Zhang, Minghao Chen, Shuai Xiao, Liang Peng, Hengjia Li, Binbin Lin, Ping Li, Wenxiao Wang, Boxi Wu, Deng Cai*
- 67 Multi-agent Collaborative Perception via Motion-aware Robust Communication Network, *Shixin Hong, Yu Liu, Zhi Li, Shaohui Li, You He*
- 68 TASEg: Temporal Aggregation Network for LiDAR Semantic Segmentation, *Xiaopei Wu, Yuenan Hou, Xiaoshui Huang, Binbin Lin, Tong He, Xinge Zhu, Yuexin Ma, Boxi Wu, Haifeng Liu, Deng Cai, Wanli Ouyang*
- 69 HINTED: Hard Instance Enhanced Detector with Mixed-Density Feature Fusion for Sparsely-Supervised 3D Object Detection, *Qiming Xia, Wei Ye, Hai Wu, Shijia Zhao, Leyuan Xing, Xun Huang, Jinhao Deng, Xin Li, Chenglu Wen, Cheng Wang*
- 70 CaKDP: Category-aware Knowledge Distillation and Pruning Framework for Lightweight 3D Object Detection, *Haonan Zhang, Longjun Liu, Yuqi Huang, Zhao Yang, Xinyu Lei, Bihan Wen*
- 71 Diffusion-ES: Gradient-free Planning with Diffusion for Autonomous and Instruction-guided Driving, *Brian Yang, Huangyuan Su, Nikolaos Gkanatsios, Tsung-Wei Ke, Ayush Jain, Jeff Schneider, Katerina Fragkiadaki*
- 72 TULIP: Transformer for Upsampling of LiDAR Point Clouds, *Bin Yang, Patrick Pfreundschuh, Roland Siegwart, Marco Hutter, Peyman Moghadam, Vaishakh Patil*
- 73 Bézier Everywhere All at Once: Learning Drivable Lanes as Bézier Graphs, *Hugh Blayney, Hanlin Tian, Hamish Scott, Nils Goldbeck, Chess Stetson, Panagiotis Angeloudis*
- 74 Flow-Guided Online Stereo Rectification for Wide Baseline Stereo, *Anush Kumar, Fahim Mannan, Omid Hosseini Jafari, Shile Li, Felix Heide*

- 75 LASIL: Learner-Aware Supervised Imitation Learning For Long-term Microscopic Traffic Simulation, *Ke Guo, Zhenwei Miao, Wei Jing, Weiwei Liu, Weizi Li, Dayang Hao, Jia Pan*
- 76 HlMap: Hybrid Representation Learning for End-to-end Vectorized HD Map Construction, *Yi Zhou, Hui Zhang, Jiaqian Yu, Yifan Yang, Sangil Jung, Seung-In Park, ByungIn Yoo*
- 77 RadSimReal: Bridging the Gap Between Synthetic and Real Data in Radar Object Detection With Simulation, *Oded Bialer, Yuval Haitman*
- 78 3D LiDAR Mapping in Dynamic Environments using a 4D Implicit Neural Representation, *Xingguang Zhong, Yue Pan, Cyrill Stachniss, Jens Behley*
- 79 Quantifying Uncertainty in Motion Prediction with Variational Bayesian Mixture, *Juanwu Lu, Can Cui, Yunsheng Ma, Aniket Bera, Ziran Wang*
- 80 Continual Learning for Motion Prediction Model via Meta-Representation Learning and Optimal Memory Buffer Retention Strategy, *DaeJun Kang, Dongsuk Kum, Sanmin Kim*
- 81 PARA-Drive: Parallelized Architecture for Real-time Autonomous Driving, *Xinshuo Weng, Boris Ivanovic, Yan Wang, Yue Wang, Marco Pavone*
- 82 ChatScene: Knowledge-Enabled Safety-Critical Scenario Generation for Autonomous Vehicles, *Jiawei Zhang, Chejian Xu, Bo Li*
- 83 CRKD: Enhanced Camera-Radar Object Detection with Cross-modality Knowledge Distillation, *Lingjun Zhao, Jingyu Song, Katherine A. Skinner*
- 84 Communication-Efficient Collaborative Perception via Information Filling with Codebook, *Yue Hu, Juntong Peng, Sifei Liu, Junhao Ge, Si Liu, Siheng Chen*
- 85 RadarDistill: Boosting Radar-based Object Detection Performance via Knowledge Distillation from LiDAR Features, *Geonho Bang, Kwangjin Choi, Jisong Kim, Dongsuk Kum, Jun Won Choi*
- 86 ICP-Flow: LiDAR Scene Flow Estimation with ICP, *Yancong Lin, Holger Caesar*
- 87 Improving Bird's Eye View Semantic Segmentation by Task Decomposition, *Tianhao Zhao, Yongcan Chen, Yu Wu, Tianyang Liu, Bo Du, Peilun Xiao, Shi Qiu, Hongda Yang, Guozhen Li, Yi Yang, Yutian Lin*
- 88 DriveWorld: 4D Pre-trained Scene Understanding via World Models for Autonomous Driving, *Chen Min, Dawei Zhao, Liang Xiao, Jian Zhao, Xinli Xu, Zheng Zhu, Lei Jin, Jianshu Li, Yulan Guo, Junliang Xing, Liping Jing, Yiming Nie, Bin Dai*
- 89 Producing and Leveraging Online Map Uncertainty in Trajectory Prediction, *Xunjiang Gu, Guanyu Song, Igor Gilitschenski, Marco Pavone, Boris Ivanovic*
- 90 HRVDA: High-Resolution Visual Document Assistant, *Chaohu Liu, Kun Yin, Haoyu Cao, Xinghua Jiang, Xin Li, Yinsong Liu, Deqiang Jiang, Xing Sun, Linli Xu*
- 91 Enhancing Visual Document Understanding with Contrastive Learning in Large Visual-Language Models, *Xin Li, Yunfei Wu, Xinghua Jiang, Zhihao Guo, Mingming Gong, Haoyu Cao, Yinsong Liu, Deqiang Jiang, Xing Sun*
- 92 RoDLA: Benchmarking the Robustness of Document Layout Analysis Models, *Yufan Chen, Jiaming Zhang, Kunyu Peng, Junwei Zheng, Ruiping Liu, Philip Torr, Rainer Stiefelhagen*
- 93 Multi-modal In-Context Learning Makes an Ego-evolving Scene Text Recognizer, *Zhen Zhao, Jingqun Tang, Chunhui Lin, Binghong Wu, Can Huang, Hao Liu, Xin Tan, Zhizhong Zhang, Yuan Xie*
- 94 CMA: A Chromaticity Map Adapter for Robust Detection of Screen-Recapture Document Images, *Changsheng Chen, Liangwei Lin, Yongqi Chen, Bin Li, Jishen Zeng, Jiwu Huang*
- 95 ODM: A Text-Image Further Alignment Pre-training Approach for Scene Text Detection and Spotting, *Chen Duan, Pei Fu, Shan Guo, Qianyi Jiang, Xiaoming Wei*
- 96 GRAM: Global Reasoning for Multi-Page VQA, *Tsachi Blau, Sharon Fogel, Roi Ronen, Alona Golts, Roy Ganz, Elad Ben Avraham, Aviad Aberdam, Shahar Tsiper, Ron Litman*
- 97 Bridging the Gap Between End-to-End and Two-Step Text Spotting, *Mingxin Huang, Hongliang Li, Yuliang Liu, Xiang Bai, Lianwen Jin*
- 98 An Empirical Study of Scaling Law for Scene Text Recognition, *Miao Rang, Zhenni Bi, Chuanjian Liu, Yunhe Wang, Kai Han*
- 99 LayoutLLM: Layout Instruction Tuning with Large Language Models for Document Understanding, *Chuwei Luo, Yufan Shen, Zhaoqing Zhu, Qi Zheng, Zhi Yu, Cong Yao*
- 100 OmniParser: A Unified Framework for Text Spotting Key Information Extraction and Table Recognition, *Jianqiang Wan, Sibao Song, Wenwen Yu, Yuliang Liu, Wenqing Cheng, Fei Huang, Xiang Bai, Cong Yao, Zhibo Yang*
- 101 DocRes: A Generalist Model Toward Unifying Document Image Restoration Tasks, *Jiaxin Zhang, Dezhi Peng, Chongyu Liu, Peirong Zhang, Lianwen Jin*
- 102 LayoutFormer: Hierarchical Text Detection Towards Scene Text Understanding, *Min Liang, Jia-Wei Ma, Xiaobin Zhu, Jingyan Qin, Xu-Cheng Yin*
- 103 Generating Handwritten Mathematical Expressions From Symbol Graphs: An End-to-End Pipeline, *Yu Chen, Fei Gao, Yanguang Zhang, Maoying Qiao, Nannan Wang*
- 104 OpenESS: Event-based Semantic Scene Understanding with Open Vocabularies, *Lingdong Kong, Youquan Liu, Lai Xing Ng, Benoit R. Cottureau, Wei Tsang Ooi*
- 105 PELA: Learning Parameter-Efficient Models with Low-Rank Approximation, *Yangyang Guo, Guangzhi Wang, Mohan Kankanhalli*
- 106 MADTP: Multimodal Alignment-Guided Dynamic Token Pruning for Accelerating Vision-Language Transformer, *Jianjian Cao, Peng Ye, Shengze Li, Chong Yu, Yansong Tang, Jiwen Lu, Tao Chen*
- 107 Vkd: Improving Knowledge Distillation using Orthogonal Projections, *Roy Miles, Ismail Elezi, Jiankang Deng*
- 108 Logit Standardization in Knowledge Distillation, *Shangquan Sun, Wenqi Ren, Jingzhi Li, Rui Wang, Xiaochun Cao*
- 109 Multi-criteria Token Fusion with One-step-ahead Attention for Efficient Vision Transformers, *Sanghyeok Lee, Joonmyung Choi, Hyunwoo J. Kim*
- 110 ParameterNet: Parameters Are All You Need for Large-scale Visual Pretraining of Mobile Networks, *Kai Han, Yunhe Wang, Jianyuan Guo, Enhua Wu*
- 111 DeepCache: Accelerating Diffusion Models for Free, *Xinyin Ma, Gongfan Fang, Xinchao Wang*
- 112 ALGM: Adaptive Local-then-Global Token Merging for Efficient Semantic Segmentation with Plain Vision Transformers, *Narges Norouzi, Svetlana Orlova, Daan de Geus, Gijs Dubbelman*
- 113 A General and Efficient Training for Transformer via Token Expansion, *Wenxuan Huang, Yunhang Shen, Jiao Xie, Baochang Zhang, Gaoqi He, Ke Li, Xing Sun, Shaohui Lin*
- 114 Efficient Dataset Distillation via Minimax Diffusion, *Jiayang Gu, Saeed Vahidian, Vyacheslav Kungurtsev, Haonan Wang, Wei Jiang, Yang You, Yiran Chen*
- 115 PEM: Prototype-based Efficient MaskFormer for Image Segmentation, *Niccolò Cavagnero, Gabriele Rosi, Claudia Cuttano, Francesca Pistilli, Marco Ciccone, Giuseppe Averta, Fabio Cermelli*
- 116 Transferable and Principled Efficiency for Open-Vocabulary Segmentation, *Jingxuan Xu, Wuyang Chen, Yao Zhao, Yunchao Wei*
- 117 Dense Vision Transformer Compression with Few Samples, *Hanxiao Zhang, Yifan Zhou, Guo-Hua Wang*
- 118 Dr2Net: Dynamic Reversible Dual-Residual Networks for Memory-Efficient Finetuning, *Chen Zhao, Shuming Liu, Karttikeya Mangalam, Guocheng Qian, Fatimah Zohra, Abdulmohsen Alghannam, Jitendra Malik, Bernard Ghanem*
- 119 MaxQ: Multi-Axis Query for N:M Sparsity Network, *Jingyang Xiang, Siqi Li, Junhao Chen, Zhuangzhi Chen, Tianxin Huang, Linpeng Peng, Yong Liu*
- 120 Retraining-Free Model Quantization via One-Shot Weight-Coupling Learning, *Chen Tang, Yuan Meng, Jiacheng Jiang, Shuzhao Xie, Rongwei Lu, Xinzhu Ma, Zhi Wang, Wenwu Zhu*
- 121 LORS: Low-rank Residual Structure for Parameter-Efficient JNetwork Stacking, *ialin Li, Qiang Nie, Weifu Fu, Yuhuan Lin, Guangpin Tao, Yong Liu, Chengjie Wang*
- 122 Towards High-fidelity Artistic Image Vectorization via Texture-Encapsulated Shape Parameterization, *Ye Chen, Bingbing Ni, Jinfan Liu, Xiaoyang Huang, Xuanhong Chen*
- 123 Learning Vision from Models Rivals Learning Vision from Data, *Yonglong Tian, Lijie Fan, Kaifeng Chen, Dina Katabi*

- Dilip Krishnan, Phillip Isola*
- 124 Efficient Multitask Dense Predictor via Binarization, *Yuzhang Shang, Dan Xu, Gaowen Liu, Ramana Rao Kompella, Yan Yan*
- 125 RepViT: Revisiting Mobile CNN From ViT Perspective, *Ao Wang, Hui Chen, Zijia Lin, Jungong Han, Guiguang Ding*
- 126 Enhancing Post-training Quantization Calibration through Contrastive Learning, *Yuzhang Shang, Gaowen Liu, Ramana Rao Kompella, Yan Yan*
- 127 FreeKD: Knowledge Distillation via Semantic Frequency Prompt, *Yuan Zhang, Tao Huang, Jiaming Liu, Tao Jiang, Kuan Cheng, Shanghang Zhang*
- 128 PTQ4SAM: Post-Training Quantization for Segment Anything, *Chengtao Lv, Hong Chen, Jinyang Guo, Yifu Ding, Xianglong Liu*
- 129 CLIP-KD: An Empirical Study of CLIP Model Distillation, *Chuangang Yang, Zhulin An, Libo Huang, Junyu Bi, Xinqiang Yu, Han Yang, Boyu Diao, Yongjun Xu*
- 130 MobileCLIP: Fast Image-Text Models through Multi-Modal Reinforced Training, *Pavan Kumar Anasosalu Vasu, Hadi Pouransari, Fartash Faghri, Raviteja Vemulapalli, Oncel Tuzel*
- 131 Scaled Decoupled Distillation, *Shicai Wei, Chunbo Luo, Yang Luo*
- 132 Self-Distilled Masked Auto-Encoders are Efficient Video Anomaly Detectors, *Nicolae-Cătălin Ristea, Florinel-Alin Croitoru, Radu Tudor Ionescu, Marius Popescu, Fahad Shahbaz Khan, Mubarak Shah*
- 133 PikeLPN: Mitigating Overlooked Inefficiencies of Low-Precision Neural Networks, *Marina Neseem, Conor McCullough, Randy Hsin, Chas Leichner, Shan Li, In Suk Chong, Andrew Howard, Lukasz Lew, Sherief Reda, Ville-Mikko Rautio, Daniele Moro*
- 134 C2KD: Bridging the Modality Gap for Cross-Modal Knowledge Distillation, *Fushuo Huo, Wenchao Xu, Jingcai Guo, Haozhao Wang, Song Guo*
- 135 KD-DETR: Knowledge Distillation for Detection Transformer with Consistent Distillation Points Sampling, *Yu Wang, Xin Li, Shengzhao Weng, Gang Zhang, Haixiao Yue, Haocheng Feng, Junyu Han, Errui Ding*
- 136 Towards Accurate Post-training Quantization for Diffusion Models, *Changyuan Wang, Ziwei Wang, Xiuwei Xu, Yansong Tang, Jie Zhou, Jiwen Lu*
- 137 CURSOR: Scalable Mixed-Order Hypergraph Matching with CUR Decomposition, *Qixuan Zheng, Ming Zhang, Hong Yan*
- 138 Frozen Feature Augmentation for Few-Shot Image Classification, *Andreas Bär, Neil Houlsby, Mostafa Dehghani, Manoj Kumar*
- 139 Jointly Training and Pruning CNNs via Learnable Agent Guidance and Alignment, *Alireza Ganjdanesh, Shangqian Gao, Heng Huang*
- 140 Zero-TPrune: Zero-Shot Token Pruning through Leveraging of the Attention Graph in Pre-Trained Transformers, *Hongjie Wang, Bhisma Dedhia, Niraj K. Jha*
- 141 Attention-Driven Training-Free Efficiency Enhancement of Diffusion Models, *Hongjie Wang, Difan Liu, Yan Kang, Yijun Li, Zhe Lin, Niraj K. Jha, Yuchen Liu*
- 142 BilevelPruning: Unified Dynamic and Static Channel Pruning for Convolutional Neural Networks, *Shangqian Gao, Yanfu Zhang, Feihu Huang, Heng Huang*
- 143 Low-Rank Rescaled Vision Transformer Fine-Tuning: A Residual Design Approach, *Wei Dong, Xing Zhang, Bihui Chen, Dawei Yan, Zhijun Lin, Qingsen Yan, Peng Wang, Yang Yang*
- 144 EfficientSAM: Leveraged Masked Image Pretraining for Efficient Segment Anything, *Yunyang Xiong, Bala Varadarajan, Lemeng Wu, Xiaoyu Xiang, Fanyi Xiao, Chenchen Zhu, Xiaoliang Dai, Dilin Wang, Fei Sun, Forrest Iandola, Raghuraman Krishnamoorthi, Vikas Chandra*
- 145 FlashEval: Towards Fast and Accurate Evaluation of Text-to-image Diffusion Generative Models, *Lin Zhao, Tianchen Zhao, Zinan Lin, Xuefei Ning, Guohao Dai, Huazhong Yang, Yu Wang*
- 146 Instance-Aware Group Quantization for Vision Transformers, *Jaehyeon Moon, Dohyung Kim, Junyong Cheon, Bumsu Ham*
- 147 Finding Lottery Tickets in Vision Models via Data-driven Spectral Foresight Pruning, *Leonardo Iurada, Marco Ciccone, Tatiana Tommasi*
- 148 Joint-Task Regularization for Partially Labeled Multi-Task Learning, *Kento Nishi, Junsik Kim, Wanhua Li, Hanspeter Pfister*
- 149 Auto-Train-Once: Controller Network Guided Automatic Network Pruning from Scratch, *Xidong Wu, Shangqian Gao, Zeyu Zhang, Zhenzhen Li, Runxue Bao, Yanfu Zhang, Xiaoqian Wang, Heng Huang*
- 150 Reg-PTQ: Regression-specialized Post-training Quantization for Fully Quantized Object Detector, *Yifu Ding, Weilun Feng, Chuyan Chen, Jinyang Guo, Xianglong Liu*
- 151 MULTIFLOW: Shifting Towards Task-Agnostic Vision-Language Pruning, *Matteo Farina, Massimiliano Mancini, Elia Cunegatti, Gaowen Liu, Giovanni Iacca, Elisa Ricci*
- 152 MTLora: Low-Rank Adaptation Approach for Efficient Multi-Task Learning, *Ahmed Agiza, Marina Neseem, Sherief Reda*
- 153 Resource-Efficient Transformer Pruning for Finetuning of Large Models, *Fatih Ilhan, Gong Su, Selim Furkan Tekin, Tiansheng Huang, Sihao Hu, Ling Liu*
- 154 Promptable Behaviors: Personalizing Multi-Objective Rewards from Human Preferences, *Minyoung Hwang, Luca Weihs, Chanwoo Park, Kimin Lee, Aniruddha Kembhavi, Kiana Ehsani*
- 155 Holodeck: Language Guided Generation of 3D Embodied AI Environments, *Yue Yang, Fan-Yun Sun, Luca Weihs, Eli VanderBilt, Alvaro Herrasti, Winson Han, Jiajun Wu, Nick Haber, Ranjay Krishna, Lingjie Liu, Chris Callison-Burch, Mark Yatskar, Aniruddha Kembhavi, Christopher Clark*
- 156 SPOC: Imitating Shortest Paths in Simulation Enables Effective Navigation and Manipulation in the Real World, *Kiana Ehsani, Tanmay Gupta, Rose Hendrix, Jordi Salvador, Luca Weihs, Kuo-Hao Zeng, Kunal Pratap Singh, Yejin Kim, Winson Han, Alvaro Herrasti, Ranjay Krishna, Dustin Schwenk, Eli VanderBilt, Aniruddha Kembhavi*
- 157 RILA: Reflective and Imaginative Language Agent for Zero-Shot Semantic Audio-Visual Navigation, *Zeyuan Yang, Jiageng Liu, Peihao Chen, Anoop Cherian, Tim K. Marks, Jonathan Le Roux, Chuang Gan*
- 158 PhyScene: Physically Interactable 3D Scene Synthesis for Embodied AI, *Yandan Yang, Baoxiong Jia, Peiyuan Zhi, Siyuan Huang*
- 159 Seeing the Unseen: Visual Common Sense for Semantic Placement, *Ram Ramrakhya, Aniruddha Kembhavi, Dhruv Batra, Zolt Kira, Kuo-Hao Zeng, Luca Weihs*
- 160 LEMON: Learning 3D Human-Object Interaction Relation from 2D Images, *Yuhang Yang, Wei Zhai, Hongchen Luo, Yang Cao, Zheng-Jun Zha*
- 161 OVER-NAV: Elevating Iterative Vision-and-Language Navigation with Open-Vocabulary Detection and StructurEd Representation, *Ganlong Zhao, Guanbin Li, Weikai Chen, Yizhou Yu*
- 162 MP5: A Multi-modal Open-ended Embodied System in Minecraft via Active Perception, *Yiran Qin, Enshen Zhou, Qichang Liu, Zhenfei Yin, Lu Sheng, Ruimao Zhang, Yu Qiao, Jing Shao*
- 163 Volumetric Environment Representation for Vision-Language Navigation, *Rui Liu, Wenguan Wang, Yi Yang*
- 164 Instance-aware Exploration-Verification-Exploitation for Instance ImageGoal Navigation, *Xiaohan Lei, Min Wang, Wengang Zhou, Li Li, Houqiang Li*
- 165 UniGarmentManip: A Unified Framework for Category-Level Garment Manipulation via Dense Visual Correspondence, *Ruihai Wu, Haoran Lu, Yiyang Wang, Yubo Wang, Hao Dong*
- 166 Evidential Active Recognition: Intelligent and Prudent Open-World Embodied Perception, *Lei Fan, Mingfu Liang, Yunxuan Li, Gang Hua, Ying Wu*
- 167 GenH2R: Learning Generalizable Human-to-Robot Handover via Scalable Simulation Demonstration and Imitation, *Zifan Wang, Junyu Chen, Ziqing Chen, Pengwei Xie, Rui Chen, Li Yi*
- 168 GOAT-Bench: A Benchmark for Multi-Modal Lifelong Navigation, *Mukul Khanna, Ram Ramrakhya, Gunjan Chhablani, Sriram Yenamandra, Theophile Gervet, Matthew Chang, Zolt Kira, Devendra Singh Chaplot, Dhruv Batra, Roozbeh Mottaghi*
- 169 Habitat Synthetic Scenes Dataset (HSSD-200): An Analysis of 3D Scene Scale and Realism Tradeoffs for ObjectGoal Navigation, *Mukul Khanna, Yongsun Mao, Hanxiao Jiang, Sanjay Haresh, Brennan Shacklett, Dhruv Batra, Alexander Clegg, Eric Undersander, Angel X. Chang, Manolis Savva*
- 170 Active Open-Vocabulary Recognition: Let Intelligent Moving Mitigate CLIP Limitations, *Lei Fan, Jianxiang Zhou, Xiaoying Xing, Ying Wu*

- 171 Rapid Motor Adaptation for Robotic Manipulator Arms, *Yichao Liang, Kevin Ellis, João Henriques*
- 172 Imagine Before Go: Self-Supervised Generative Map for Object Goal Navigation, *Sixian Zhang, Xinyao Yu, Xinhang Song, Xiaohan Wang, Shuqiang Jiang*
- 173 Auto MC-Reward: Automated Dense Reward Design with Large Language Models for Minecraft, *Hao Li, Xue Yang, Zhaokai Wang, Xizhou Zhu, Jie Zhou, Yu Qiao, Xiaogang Wang, Hongsheng Li, Lewei Lu, Jifeng Dai*
- 174 GenNBV: Generalizable Next-Best-View Policy for Active 3D Reconstruction, *Xiao Chen, Quanyi Li, Tai Wang, Tianfan Xue, Jiangmiao Pang*
- 175 An Interactive Navigation Method with Effect-oriented Affordance, *Xiaohan Wang, Yuehu Liu, Xinhang Song, Yuyi Liu, Sixian Zhang, Shuqiang Jiang*
- 176 A Category Agnostic Model for Visual Rearrangement, *Yuyi Liu, Xinhang Song, Weijie Li, Xiaohan Wang, Shuqiang Jiang*
- 177 SkillDiffuser: Interpretable Hierarchical Planning via Skill Abstractions in Diffusion-Based Task Execution, *Zhixuan Liang, Yao Mu, Hengbo Ma, Masayoshi Tomizuka, Mingyu Ding, Ping Luo*
- 178 Fusing Personal and Environmental Cues for Identification and Segmentation of First-Person Camera Wearers in Third-Person Views, *Ziwei Zhao, Yuchen Wang, Chuhua Wang*
- 179 OpenEQA: Embodied Question Answering in the Era of Foundation Models, *Arjun Majumdar, Anurag Ajay, Xiaohan Zhang, Pranav Putta, Sriram Yenamandra, Mikael Henaff, Sneha Silwal, Paul Mccvay, Aleksandr Maksymets, Sergio Arnaud, Karmesh Yadav, Qiyang Li, Ben Newman, Mohit Sharma, Vincent Berges, Shiqi Zhang, Pulkit Agrawal, Yonatan Bisk, Dhruv Batra, Mrinal Kalakrishnan, Franziska Meier, Chris Paxton, Alexander Sax, Aravind Rajeswaran*
- 180 Model Adaptation for Time Constrained Embodied Control, *Jaehyun Song, Minjong Yoo, Honguk Woo*
- 181 EgoGen: An Egocentric Synthetic Data Generator,
* *Gen Li, Kaifeng Zhao, Siwei Zhang, Xiaozhong Lyu, Mihai Dusmanu, Yan Zhang, Marc Pollefeys, Siyu Tang*
- 182 RoHM: Robust Human Motion Reconstruction via Diffusion,
* *Siwei Zhang, Bharat Lal Bhatnagar, Yuanlu Xu, Alexander Winkler, Petr Kadlecek, Siyu Tang, Federica Bogo*
- 183 An N-Point Linear Solver for Line and Motion Estimation with
* *Event Cameras, Ling Gao, Daniel Gehrig, Hang Su, Davide Scaramuzza, Laurent Kneip*
- 184 A Subspace-Constrained Tyler's Estimator and its Applications to
* *Structure from Motion, Feng Yu, Teng Zhang, Gilad Lerman*
- 185 SpiderMatch: 3D Shape Matching with Global Optimality and
* *Geometric Consistency, Paul Roetzer, Florian Bernard*
- 186 You'll Never Walk Alone: A Sketch and Text Duet for Fine-Grained Image Retrieval, *Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song*
- 187 CrossKD: Cross-Head Knowledge Distillation for Object Detection, *Jiabao Wang, Yuming Chen, Zhaohui Zheng, Xiang Li, Ming-Ming Cheng, Qibin Hou*
- 188 ProTeCt: Prompt Tuning for Taxonomic Open Set Classification, *Tz-Ying Wu, Chih-Hui Ho, Nuno Vasconcelos*
- 189 CAT: Exploiting Inter-Class Dynamics for Domain Adaptive Object Detection, *Mikhail Kennerley, Jian-Gang Wang, Bharadwaj Veeravalli, Robby T. Tan*
- 190 Text Is MASS: Modeling as Stochastic Embedding for Text-Video Retrieval, *Jiamian Wang, Guohao Sun, Pichao Wang, Dongfang Liu, Sohail Dianat, Majid Rabbani, Raghuveer Rao, Zhiqiang Tao*
- 191 UniMODE: Unified Monocular 3D Object Detection,
* *Zhuoling Li, Xiaogang Xu, SerNam Lim, Hengshuang Zhao*
- 192 OVMR: Open-Vocabulary Recognition with Multi-Modal References, *Zehong Ma, Shiliang Zhang, Longhui Wei, Qi Tian*
- 193 From Isolated Islands to Pangea: Unifying Semantic Space for Human Action Understanding, *Yong-Lu Li, Xiaoqian Wu, Xinpeng Liu, Zehao Wang, Yiming Dou, Yikun Ji, Junyi Zhang, Yixing Li, Xudong Lu, Jingru Tan, Cewu Lu*
- 194 Language-conditioned Detection Transformer, *Jang Hyun Cho, Philipp Krähenbühl*
- 195 Distribution-aware Knowledge Prototyping for Non-exemplar Lifelong Person Re-identification, *Kunlun Xu, Xu Zou, Yuxin Peng, Jiahuan Zhou*
- 196 Learning Continual Compatible Representation for Re-indexing Free Lifelong Person Re-identification, *Zhenyu Cui, Jiahuan Zhou, Xun Wang, Manyu Zhu, Yuxin Peng*
- 197 Active Object Detection with Knowledge Aggregation and Distillation from Large Models, *Dejie Yang, Yang Liu*
- 198 SHiNe: Semantic Hierarchy Nexus for Open-vocabulary Object
* *Detection, Mingxuan Liu, Tyler L. Hayes, Elisa Ricci, Gabriela Csurka, Riccardo Volpi*
- 199 Object Recognition as Next Token Prediction,
* *Kaiyu Yue, Bor-Chun Chen, Jonas Geiping, Hengduo Li, Tom Goldstein, Ser-Nam Lim*
- 200 Exploring the Potential of Large Foundation Models for Open-Vocabulary HOI Detection, *Ting Lei, Shaofeng Yin, Yang Liu*
- 201 Gradient Reweighting: Towards Imbalanced Class-Incremental Learning, *Jiangpeng He*
- 202 Learning Background Prompts to Discover Implicit Knowledge for Open Vocabulary Object Detection, *Jiaming Li, Jiacheng Zhang, Jichang Li, Ge Li, Si Liu, Liang Lin, Guanbin Li*
- 203 Multi-View Attentive Contextualization for Multi-View 3D Object Detection, *Xianpeng Liu, Ce Zheng, Ming Qian, Nan Xue, Chen Chen, Zhebin Zhang, Chen Li, Tianfu Wu*
- 204 RealNet: A Feature Selection Network with Realistic Synthetic Anomaly for Anomaly Detection, *Ximiao Zhang, Min Xu, Xiuzhuang Zhou*
- 205 Generalized Large-Scale Data Condensation via Various Backbone
* *and Statistical Matching, Shitong Shao, Zeyuan Yin, Muxin Zhou, Xindong Zhang, Zhiqiang Shen*
- 206 Unleashing Unlabeled Data: A Paradigm for Cross-View Geo-Localization, *Guopeng Li, Ming Qian, Gui-Song Xia*
- 207 PointOBB: Learning Oriented Object Detection via Single Point Supervision, *Junwei Luo, Xue Yang, Yi Yu, Qingyun Li, Junchi Yan, Yansheng Li*
- 208 Scene-adaptive and Region-aware Multi-modal Prompt for Open Vocabulary Object Detection, *Xiaowei Zhao, Xianglong Liu, Duorui Wang, Yajun Gao, Zhide Liu*
- 209 Revisiting the Domain Shift and Sample Uncertainty in Multi-source Active Domain Transfer, *Wenqiao Zhang, Zheqi Lv, Hao Zhou, Jia-Wei Liu, Juncheng Li, Mengze Li, Yunfei Li, Dongping Zhang, Yueting Zhuang, Siliang Tang*
- 210 Hyperbolic Learning with Synthetic Captions for Open-World Detection, *Fanjie Kong, Yanbei Chen, Jiarui Cai, Davide Modolo*
- 211 CricaVPR: Cross-image Correlation-aware Representation Learning for Visual Place Recognition, *Feng Lu, Xiangyuan Lan, Lijun Zhang, Dongmei Jiang, Yaowei Wang, Chun Yuan*
- 212 Point2RBox: Combine Knowledge from Synthetic Visual Patterns for End-to-end Oriented Object Detection with Single Point Supervision, *Yi Yu, Xue Yang, Qingyun Li, Feipeng Da, Jifeng Dai, Yu Qiao, Junchi Yan*
- 213 Scene Adaptive Sparse Transformer for Event-based Object Detection, *Yansong Peng, Hebei Li, Yueyi Zhang, Xiaoyan Sun, Feng Wu*
- 214 Visual Delta Generator with Large Multi-modal Models for Semi-supervised Composed Image Retrieval, *Young Kyun Jang, Donghyun Kim, Zihang Meng, Dat Huynh, Ser-Nam Lim*
- 215 Preserving Fairness Generalization in Deepfake Detection, *Li Lin, Xinan He, Yan Ju, Xin Wang, Feng Ding, Shu Hu*
- 216 Text-to-Image Diffusion Models are Great Sketch-Photo Matchmakers, *Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song*
- 217 PromptAD: Learning Prompts with only Normal Samples for Few-Shot Anomaly Detection, *Xiaofan Li, Zhizhong Zhang, Xin Tan, Chengwei Chen, Yanyun Qu, Yuan Xie, Lizhuang Ma*
- 218 Structured Model Probing: Empowering Efficient Transfer Learning by Structured Regularization, *Zhi-Fan Wu, Chaojie Mao, Wue Wang, Jianwen Jiang, Yiliang Lv, Rong Jin*
- 219 How to Handle Sketch-Abstraction in Sketch-Based Image Retrieval?, *Subhadeep Koley, Ayan Kumar Bhunia, Aneeshan Sain, Pinaki Nath Chowdhury, Tao Xiang, Yi-Zhe Song*

- 220 Shallow-Deep Collaborative Learning for Unsupervised Visible-Infrared Person Re-Identification, *Bin Yang, Jun Chen, Mang Ye*
- 221 Solving the Catastrophic Forgetting Problem in Generalized Category Discovery, *Xinzi Cao, Xiawu Zheng, Guan hong Wang, Weijiang Yu, Yunhang Shen, Ke Li, Yutong Lu, Yonghong Tian*
- 222 Active Generalized Category Discovery, *Shijie Ma, Fei Zhu, Zhun Zhong, Xu-Yao Zhang, Cheng-Lin Liu*
- 223 YOLO-World: Real-Time Open-Vocabulary Object Detection, *Tianheng Cheng, Lin Song, Yixiao Ge, Wenyu Liu, Xinggang Wang, Ying Shan*
- 224 Theoretically Achieving Continuous Representation of Oriented Bounding Boxes, *Zikai Xiao, Guoye Yang, Xue Yang, Taijiang Mu, Junchi Yan, Shimin Hu*
- 225 Decoupled Pseudo-labeling for Semi-Supervised Monocular 3D Object Detection, *Jiacheng Zhang, Jiaming Li, Xiangru Lin, Wei Zhang, Xiao Tan, Junyu Han, Errui Ding, Jingdong Wang, Guanbin Li*
- 226 LEOD: Label-Efficient Object Detection for Event Cameras, *Ziyi Wu, Mathias Gehrig, Qing Lyu, Xudong Liu, Igor Gilitschenski*
- 227 Lane2Seq: Towards Unified Lane Detection via Sequence Generation, *Kunyang Zhou*
- 228 Open-World Human-Object Interaction Detection via Multi-modal Prompts, *Jie Yang, Bingliang Li, Ailing Zeng, Lei Zhang, Ruimao Zhang*
- 229 DETRs Beat YOLOs on Real-time Object Detection, *Yian Zhao, Wenyu Lv, Shangliang Xu, Jinman Wei, Guanzhong Wang, Qingqing Dang, Yi Liu, Jie Chen*
- 230 Exploring Region-Word Alignment in Built-in Detector for Open-Vocabulary Object Detection, *Heng Zhang, Qiuyu Zhao, Linyu Zheng, Hao Zeng, Zhiwei Ge, Tianhao Li, Sulong Xu*
- 231 Referring Expression Counting,
* *Siyang Dai, Jun Liu, Ngai-Man Cheung*
- 232 ActiveDC: Distribution Calibration for Active Finetuning, *Wenshuai Xu, Zhenghui Hu, Yu Lu, Jinzhou Meng, Qingjie Liu, Yunhong Wang*
- 233 LaRE²: Latent Reconstruction Error Based Method for Diffusion-Generated Image Detection, *Yunpeng Luo, Junlong Du, Ke Yan, Shouhong Ding*
- 234 Fine-grained Prototypical Voting with Heterogeneous Mixup for Semi-supervised 2D-3D Cross-modal Retrieval, *Fan Zhang, Xian-Sheng Hua, Chong Chen, Xiao Luo*
- 235 MS-DETR: Efficient DETR Training with Mixed Supervision, *Chuyang Zhao, Yifan Sun, Wenhao Wang, Qiang Chen, Errui Ding, Yi Yang, Jingdong Wang*
- 236 Context-based and Diversity-driven Specificity in Compositional Zero-Shot Learning, *Yun Li, Zhe Liu, Hang Chen, Lina Yao*
- 237 Pixel-level Semantic Correspondence through Layout-aware Representation Learning and Multi-scale Matching Integration, *Yixuan Sun, Zhangyue Yin, Haibo Wang, Yan Wang, Xipeng Qiu, Weifeng Ge, Wenqiang Zhang*
- 238 Exploiting Inter-sample and Inter-feature Relations in Dataset Distillation, *Wenxiao Deng, Wenbin Li, Tianyu Ding, Lei Wang, Hongguang Zhang, Kuihua Huang, Jing Huo, Yang Gao*
- 239 Point Segment and Count: A Generalized Framework for Object Counting, *Zhizhong Huang, Mingliang Dai, Yi Zhang, Junping Zhang, Hongming Shan*
- 240 Dual Pose-invariant Embeddings: Learning Category and Object-specific Discriminative Representations for Recognition and Retrieval, *Rohan Sarkar, Avinash Kak*
- 241 Riemannian Multinomial Logistics Regression for SPD Neural Networks, *Ziheng Chen, Yue Song, Gaowen Liu, Ramana Rao Kompella, Xiao-Jun Wu, Nicu Sebe*
- 242 Learning for Transductive Threshold Calibration in Open-World Recognition, *Qin Zhang, Dongsheng An, Tianjun Xiao, Tong He, Qingming Tang, Ying Nian Wu, Joseph Tighe, Yifan Xing*
- 243 Region-Based Representations Revisited, *Michal Shlapentokh-Rothman, Ansel Blume, Yao Xiao, Yuqun Wu, Sethuraman TV, Heyi Tao, Jae Yong Lee, Wilfredo Torres, Yu-Xiong Wang, Derek Hoiem*
- 244 Magic Tokens: Select Diverse Tokens for Multi-modal Object Re-Identification, *Pingping Zhang, Yuhao Wang, Yang Liu, Zhengzheng Tu, Huchuan Lu*
- 245 Harnessing the Power of MLLMs for Transferable Text-to-Image Person ReID, *Wentan Tan, Changxing Ding, Jiayu Jiang, Fei Wang, Yibing Zhan, Dapeng Tao*
- 246 Holistic Features are almost Sufficient for Text-to-Video Retrieval, *Kaibin Tian, Ruixiang Zhao, Zijie Xin, Bangxiang Lan, Xirong Li*
- 247 Enhancing the Power of OOD Detection via Sample-Aware Model Selection, *Feng Xue, Zi He, Yuan Zhang, Chuanlong Xie, Zhenguo Li, Falong Tan*
- 248 PanoOcc: Unified Occupancy Representation for Camera-based 3D Panoptic Segmentation, *Yuqi Wang, Yuntao Chen, Xingyu Liao, Lue Fan, Zhaoxiang Zhang*
- 249 VSCode: General Visual Salient and Camouflaged Object Detection with 2D Prompt Learning, *Ziyang Luo, Nian Liu, Wangbo Zhao, Xuguang Yang, Dingwen Zhang, Deng-Ping Fan, Fahad Khan, Junwei Han*
- 250 D3still: Decoupled Differential Distillation for Asymmetric Image Retrieval, *Yi Xie, Yihong Lin, Wenjie Cai, Xuemiao Xu, Huaidong Zhang, Yong Du, Shengfeng He*
- 251 SFOD: Spiking Fusion Object Detector, *Yimeng Fan, Wei Zhang, Changsong Liu, Mingyang Li, Wenrui Lu*
- 252 Depth-Aware Concealed Crop Detection in Dense Agricultural Scenes, *Liqiong Wang, Jinyu Yang, Yanfu Zhang, Fangyi Wang, Feng Zheng*
- 253 Extreme Point Supervised Instance Segmentation, *Hyeonjun Lee, Sehyun Hwang, Suha Kwak*
- 254 Enhance Image Classification via Inter-Class Image Mixup with Diffusion Model, *Zhicai Wang, Longhui Wei, Tan Wang, Heyu Chen, Yanbin Hao, Xiang Wang, Xiangnan He, Qi Tian*
- 255 Multimodal Industrial Anomaly Detection by Crossmodal Feature Mapping, *Alex Costanzino, Pierluigi Zama Ramirez, Giuseppe Lisanti, Luigi Di Stefano*
- 256 Effective Video Mirror Detection with Inconsistent Motion Cues, *Alex Warren, Ke Xu, Jiaying Lin, Gary K.L. Tam, Rynson W.H. Lau*
- 257 Multi-Attribute Interactions Matter for 3D Visual Grounding, *Can Xu, Yuehui Han, Rui Xu, Le Hui, Jin Xie, Jian Yang*
- 258 Looking 3D: Anomaly Detection with 2D-3D Alignment, *Ankan Bhunia, Changjian Li, Hakan Bilen*
- 259 Characteristics Matching Based Hash Codes Generation for Efficient Fine-grained Image Retrieval, *Zhen-Duo Chen, Li-Jun Zhao, Zi-Chao Zhang, Xin Luo, Xin-Shun Xu*
- 260 EASE-DETR: Easing the Competition among Object Queries, *Yulu Gao, Yifan Sun, Xudong Ding, Chuyang Zhao, Si Liu*
- 261 ProS: Prompting-to-simulate Generalized knowledge for Universal Cross-Domain Retrieval, *Kaipeng Fang, Jingkuan Song, Lianli Gao, Pengpeng Zeng, Zhi-Qi Cheng, Xiyao Li, Heng Tao Shen*
- 262 Exploring Orthogonality in Open World Object Detection, *Zhicheng Sun, Jinghan Li, Yadong Mu*
- 263 A Generative Approach for Wikipedia-Scale Visual Entity Recognition, *Mathilde Caron, Ahmet Iscen, Alireza Fathi, Cordelia Schmid*
- 264 Unleashing Channel Potential: Space-Frequency Selection Convolution for SAR Object Detection, *Ke Li, Di Wang, Zhangyuan Hu, Wenxuan Zhu, Shaofeng Li, Quan Wang*
- 265 Hyperspherical Classification with Dynamic Label-to-Prototype Assignment, *Mohammad Saeed Ebrahimi Saadabadi, Ali Dabouei, Sahar Rahimi Malakshan, Nasser M. Nasrabadi*
- 266 A Pedestrian is Worth One Prompt: Towards Language Guidance
* Person Re-Identification, *Zexian Yang, Dayan Wu, Chenming Wu, Zheng Lin, Jingzi Gu, Weiping Wang*
- 267 VSRD: Instance-Aware Volumetric Silhouette Rendering for Weakly Supervised 3D Object Detection, *Zihua Liu, Hiroki Sakuma, Masatoshi Okutomi*
- 268 Improving Visual Recognition with Hyperbolic Visual Hierarchy Mapping, *Hyeonjun Kwon, Jinhyun Jang, Jin Kim, Kwonyoung Kim, Kwanghoon Sohn*
- 269 On Train-Test Class Overlap and Detection for Image Retrieval, *Chull Hwan Song, Jooyoung Yoon, Taebaek Hwang, Shunghyun Choi, Yeong Hyeon Gu, Yannis Avrithis*
- 270 Multi-Scale Video Anomaly Detection by Multi-Grained
* Spatio-Temporal Representation Learning, *Menghao Zhang, Jingyu Wang, Qi Qi, Haifeng Sun, Zirui Zhuang, Pengfei Ren, Ruilong Ma, Jianxin Liao*

- 271 LAA-Net: Localized Artifact Attention Network for Quality-Agnostic and Generalizable Deepfake Detection, *Dat Nguyen, Nesryne Mejri, Inder Pal Singh, Polina Kuleshova, Marcella Astrid, Anis Kacem, Enjie Ghorbel, Djamila Aouada*
- 272 Rethinking Boundary Discontinuity Problem for Oriented Object Detection, *Hang Xu, Xinyuan Liu, Haonan Xu, Yike Ma, Zunjie Zhu, Chenggang Yan, Feng Dai*
- 273 Hybrid Proposal Refiner: Revisiting DETR Series from the Faster R-CNN Perspective, *Jinjing Zhao, Fangyun Wei, Chang Xu*
- 274 Retrieval-Augmented Open-Vocabulary Object Detection, *Jooyeon Kim, Eulrang Cho, Sehyung Kim, Hyunwoo J. Kim*
- 275 LiDAR-based Person Re-identification, *Wenxuan Guo, Zhiyu Pan, Yingping Liang, Ziheng Xi, Zhicheng Zhong, Jianjiang Feng, Jie Zhou*
- 276 EventDance: Unsupervised Source-free Cross-modal Adaptation for Event-based Object Recognition, *Xu Zheng, Lin Wang*
- 277 All in One Framework for Multimodal Re-identification in the Wild, *He Li, Mang Ye, Ming Zhang, Bo Du*
- 278 Logarithmic Lenses: Exploring Log RGB Data for Image Classification, *Bruce A. Maxwell, Sumegha Singhania, Avnish Patel, Rahul Kumar, Heather Fryling, Sihan Li, Haonan Sun, Ping He, Zewen Li*
- 279 ID-like Prompt Learning for Few-Shot Out-of-Distribution Detection, *Yichen Bai, Zongbo Han, Bing Cao, Xiaoheng Jiang, Qinghua Hu, Changqing Zhang*
- 280 Infrared Small Target Detection with Scale and Location Sensitivity, *Qiankun Liu, Rui Liu, Bolun Zheng, Hongkui Wang, Ying Fu*
- 281 SURE: SURvey REcipes for building reliable and robust deep networks, *Yuting Li, Yingyi Chen, Xuanlong Yu, Dexiong Chen, Xi Shen*
- 282 Hyperbolic Anomaly Detection, *Huimin Li, Zhentao Chen, Yunhao Xu, Junlin Hu*
- 283 Instruct-RelD: A Multi-purpose Person Re-identification Task with Instructions, *Weizhen He, Yiheng Deng, Shixiang Tang, Qihao Chen, Qingsong Xie, Yizhou Wang, Lei Bai, Feng Zhu, Rui Zhao, Wanli Ouyang, Donglian Qi, Yunfeng Yan*
- 284 CA-Jaccard: Camera-aware Jaccard Distance for Person Re-identification, *Yiyu Chen, Zheyi Fan, Zhaoru Chen, Yixuan Zhu*
- 285 Improved Zero-Shot Classification by Adapting VLMs with Text Descriptions, *Oindrila Saha, Grant Van Horn, Subhansu Maji*
- 286 Modeling Collaborator: Enabling Subjective Vision Classification with Minimal Human Effort via LLM Tool-Use, *Imad Eddine Toubal, Aditya Avinash, Neil Gordon Alldrin, Jan Dlabal, Wenlei Zhou, Enming Luo, Otilia Stretcu, Hao Xiong, Chun-Ta Lu, Howard Zhou, Ranjay Krishna, Ariel Fuxman, Tom Duerig*
- 287 Neural Exposure Fusion for High-Dynamic Range Object Detection, *Emmanuel Onzon, Maximilian Bömer, Fahim Mannan, Felix Heide*
- 288 Saliency DETR: Enhancing Detection Transformer with Hierarchical Saliency Filtering Refinement, *Xiuquan Hou, Meiqin Liu, Senlin Zhang, Ping Wei, Badong Chen*
- 289 Learning Transferable Negative Prompts for Out-of-Distribution Detection, *Tianqi Li, Guansong Pang, Xiao Bai, Wenjun Miao, Jin Zheng*
- 290 TransLoc4D: Transformer-based 4D Radar Place Recognition, *Guohao Peng, Heshan Li, Yangyang Zhao, Jun Zhang, Zhenyu Wu, Pengyu Zheng, Danwei Wang*
- 291 Prompt-Driven Dynamic Object-Centric Learning for Single Domain Generalization, *Deng Li, Aming Wu, Yaowei Wang, Yahong Han*
- 292 Anomaly Heterogeneity Learning for Open-set Supervised Anomaly Detection, *Jiawen Zhu, Choubo Ding, Yu Tian, Guansong Pang*
- 293 Contrastive Learning for DeepFake Classification and Localization via Multi-Label Ranking, *Cheng-Yao Hong, Yen-Chi Hsu, Tyng-Luh Liu*
- 294 Adaptive Softassign via Hadamard-Equipped Sinkhorn, *Binrui Shen, Qiang Niu, Shengxin Zhu*
- 295 An Asymmetric Augmented Self-Supervised Learning Method for Unsupervised Fine-Grained Image Hashing, *Feiran Hu, Chenlin Zhang, Jiangliang Guo, Xiu-Shen Wei, Lin Zhao, Anqi Xu, Lingyan Gao*
- 296 Optimal Transport Aggregation for Visual Place Recognition, *Sergio Izquierdo, Javier Civera*
- 297 Atom-Level Optical Chemical Structure Recognition with Limited Supervision, *Martijn Oldenhof, Edward De Brouwer, Adam Arany, Yves Moreau*
- 298 Novel Class Discovery for Ultra-Fine-Grained Visual Categorization, *Yu Liu, Yaqi Cai, Qi Jia, Binglin Qiu, Weimin Wang, Nan Pu*
- 299 Attribute-Guided Pedestrian Retrieval: Bridging Person Re-ID with Internal Attribute Variability, *Yan Huang, Zhang Zhang, Qiang Wu, Yi Zhong, Liang Wang*
- 300 Robust Noisy Correspondence Learning with Equivariant Similarity Consistency, *Yuchen Yang, Likai Wang, Erkun Yang, Cheng Deng*
- 301 Bootstrapping SparseFormers from Vision Foundation Models, *Ziteng Gao, Zhan Tong, Kevin Qinghong Lin, Joya Chen, Mike Zheng Shou*
- 302 Not All Classes Stand on Same Embeddings: Calibrating a Semantic Distance with Metric Tensor, *Jae Hyeon Park, Gyoomin Lee, Seunggi Park, Sung In Cho*
- 303 Improving Single Domain-Generalized Object Detection: A Focus on Diversification and Alignment, *Muhammad Sohail Danish, Muhammad Haris Khan, Muhammad Akhtar Munir, M. Saquib Sarfraz, Mohsen Ali*
- 304 On the Estimation of Image-matching Uncertainty in Visual Place Recognition, *Mubariz Zaffar, Liangliang Nan, Julian F. P. Kooij*
- 305 Supervised Anomaly Detection for Complex Industrial Images, *Aimira Baitieva, David Hurych, Victor Besnier, Olivier Bernard*
- 306 Fourier-basis Functions to Bridge Augmentation Gap: Rethinking Frequency Augmentation in Image Classification, *Puru Vaish, Shunxin Wang, Nicola Strisciuglio*
- 307 TransNeXt: Robust Foveal Visual Perception for Vision Transformers, *Dai Shi*
- 308 Plug and Play Active Learning for Object Detection, *Chenhongyi Yang, Lichao Huang, Elliot J. Crowley*
- 309 BoQ: A Place is Worth a Bag of Learnable Queries, *Amar Ali-bey, Brahim Chaib-draa, Philippe Giguère*
- 310 From Coarse to Fine-Grained Open-Set Recognition, *Nico Lang, Vésteinn Snæbjarnarson, Elijah Cole, Oisín Mac Aodha, Christian Igel, Serge Belongie*
- 311 Exploring Pose-Aware Human-Object Interaction via Hybrid Learning, *Eastman Z Y Wu, Yali Li, Yuan Wang, Shengjin Wang*
- 312 Toward Generalist Anomaly Detection via In-context Residual Learning with Few-shot Sample Prompts, *Jiawen Zhu, Guansong Pang*
- 313 Learning to Navigate Efficiently and Precisely in Real Environments, *Guillaume Bono, Hervé Poirier, Leonid Antsfeld, Gianluca Monaci, Boris Chidlovskii, Christian Wolf*
- 314 Task-Conditioned Adaptation of Visual Features in Multi-Task Policy Learning, *Pierre Marza, Laetitia Matignon, Olivier Simonin, Christian Wolf*
- 315 FastMAC: Stochastic Spectral Sampling of Correspondence Graph, *Yifei Zhang, Hao Zhao, Hongyang Li, Siheng Chen*
- 316 FoundationPose: Unified 6D Pose Estimation and Tracking of Novel Objects, *Bowen Wen, Wei Yang, Jan Kautz, Stan Birchfield*
- 317 CAGE: Controllable Articulation GEneration, *Jiayi Liu, Hou In Ivan Tam, Ali Mahdavi-Amiri, Manolis Savva*
- 318 SingularTrajectory: Universal Trajectory Predictor Using Diffusion Model, *Inhwan Bae, Young-Jae Park, Hae-Gon Jeon*
- 319 Language-driven Grasp Detection, *An Dinh Vuong, Minh Nhat Vu, Baoru Huang, Nghia Nguyen, Hieu Le, Thieu Vo, Anh Nguyen*
- 320 MemoNav: Working Memory Model for Visual Navigation, *Hongxin Li, Zeyu Wang, Xu Yang, Yuran Yang, Shuqi Mei, Zhaoxiang Zhang*
- 321 NOPE: Novel Object Pose Estimation from a Single Image, *Van Nguyen Nguyen, Thibault Groueix, Georgy Ponimatkin, Yinlin Hu, Renaud Marlet, Mathieu Salzmann, Vincent Lepetit*
- 322 Dexterous Grasp Transformer, *Guo-Hao Xu, Yi-Lin Wei, Dian Zheng, Xiao-Ming Wu, Wei-Shi Zheng*
- 323 Versatile Navigation Under Partial Observability via Value-guided Diffusion Policy, *Gengyu Zhang, Hao Tang, Yan Yan*
- 324 CyberDemo: Augmenting Simulated Human Demonstration for Real-World Dexterous Manipulation, *Jun Wang, Yuzhe Qin, Kaiming Kuang, Yigit Korkmaz, Akhilan Gurumoorthy, Hao Su, Xiaolong Wang*

- 325 SchurVINS: Schur Complement-Based Lightweight Visual Inertial Navigation System, *Yunfei Fan, Tianyu Zhao, Guidong Wang*
- 326 READ: Retrieval-Enhanced Asymmetric Diffusion for Motion Planning, *Takeru Oba, Matthew Walter, Norimichi Ukita*
- 327 Retrieval-Augmented Embodied Agents, *Yichen Zhu, Zhicai Ou, Xiaofeng Mou, Jian Tang*
- 328 Collaborative Semantic Occupancy Prediction with Hybrid Feature Fusion in Connected Automated Vehicles, *Rui Song, Chenwei Liang, Hu Cao, Zhiran Yan, Walter Zimmer, Markus Gross, Andreas Festag, Alois Knoll*
- 329 Diffusion-EDFs: Bi-equivariant Denoising Generative Modeling on
* SE(3) for Visual Robotic Manipulation, *Hyunwoo Ryu, Jiwoo Kim, Hyunseok An, Junwoo Chang, Joohwan Seo, Taehan Kim, Yubin Kim, Chaewon Hwang, Jongeun Choi, Roberto Horowitz*
- 330 Adaptive VIO: Deep Visual-Inertial Odometry with Online Continual Learning, *Youqi Pan, Wugen Zhou, Yingdian Cao, Hongbin Zha*
- 331 F3Loc: Fusion and Filtering for Floorplan Localization,
* *Changan Chen, Rui Wang, Christoph Vogel, Marc Pollefeys*
- 332 Gaussian Splatting SLAM,
* *Hidenobu Matsuki, Riku Murai, Paul H.J. Kelly, Andrew J. Davison*
- 333 SUGAR: Pre-training 3D Visual Representations for Robotics, *Shizhe Chen, Ricardo Garcia, Ivan Laptev, Cordelia Schmid*
- 334 ManipLLM: Embodied Multimodal Large Language Model for Object-Centric Robotic Manipulation, *Xiaoqi Li, Mingxu Zhang, Yiran Geng, Haoran Geng, Yuxing Long, Yan Shen, Renrui Zhang, Jiaming Liu, Hao Dong*
- 335 Open-Vocabulary Object 6D Pose Estimation, *Jaime Corsetti, Davide Boscaini, Changjae Oh, Andrea Cavallaro, Fabio Poiesi*
- 336 Hierarchical Diffusion Policy for Kinematics-Aware Multi-Task Robotic Manipulation, *Xiao Ma, Sumit Patidar, Iain Haughton, Stephen James*
- 337 Smart Help: Strategic Opponent Modeling for Proactive and Adaptive Robot Assistance in Households, *Zhihao Cao, Zidong Wang, Siwen Xie, Anji Liu, Lifeng Fan*
- 338 Generalizing 6-DoF Grasp Detection via Domain Prior Knowledge, *Haoxiang Ma, Modi Shi, Boyang Gao, Di Huang*
- 339 A Simple and Effective Point-based Network for Event Camera 6-DOFs Pose Relocalization, *Hongwei Ren, Jiadong Zhu, Yue Zhou, Haotian Fu, Yulong Huang, Bojun Cheng*
- 340 Neural Visibility Field for Uncertainty-Driven Active Mapping, *Shangjie Xue, Jesse Dill, Pranay Mathur, Frank Dellaert, Panagiotis Tsiotra, Danfei Xu*
- 341 SPIN: Simultaneous Perception Interaction and Navigation, *Shagun Uppal, Ananye Agarwal, Haoyu Xiong, Kenneth Shaw, Deepak Pathak*
- 342 SceneFun3D: Fine-Grained Functionality and Affordance Understanding in 3D Scenes, *Alexandros Delitzas, Ayca Takmaz, Federico Tombari, Robert Sumner, Marc Pollefeys, Francis Engelmann*
- 343 PredToken: Predicting Unknown Tokens and Beyond with Coarse-to-Fine Iterative Decoding, *Xuesong Nie, Haoyuan Jin, Yunfeng Yan, Xi Chen, Zhihang Zhu, Donglian Qi*
- 344 TIM: A Time Interval Machine for Audio-Visual Action Recognition, *Jacob Chalk, Jaesung Huh, Evangelos Kazakos, Andrew Zisserman, Dima Damen*
- 345 AutoAD III: The Prequel – Back to the Pixels, *Tengda Han, Max Bain, Arsha Nagrani, Gül Varol, Weidi Xie, Andrew Zisserman*
- 346 FACT: Frame-Action Cross-Attention Temporal Modeling for Efficient Action Segmentation, *Zijia Lu, Ehsan Elhamifar*
- 347 Progress-Aware Online Action Segmentation for Egocentric Procedural Task Videos, *Yuhan Shen, Ehsan Elhamifar*
- 348 Video ReCap: Recursive Captioning of Hour-Long Videos, *Md Mohaiminul Islam, Ngan Ho, Xitong Yang, Tushar Nagarajan, Lorenzo Torresani, Gedas Bertasius*
- 349 OmniViD: A Generative Framework for Universal Video Understanding, *Junke Wang, Dongdong Chen, Chong Luo, Bo He, Lu Yuan, Zuxuan Wu, Yu-Gang Jiang*
- 350 MovieChat: From Dense Token to Sparse Memory for Long Video Understanding, *Enxin Song, Wenhao Chai, Guanhong Wang, Yucheng Zhang, Haoyang Zhou, Feiyang Wu, Haozhe Chi, Xun Guo, Tian Ye, Yanting Zhang, Yan Lu, Jenq-Neng Hwang, Gaoang Wang*
- 351 Learning Group Activity Features Through Person Attribute Prediction, *Chihiro Nakatani, Hiroaki Kawashima, Norimichi Ukita*
- 352 Streaming Dense Video Captioning, *Xingyi Zhou, Anurag Arnab, Shyamal Buch, Shen Yan, Austin Myers, Xuehan Xiong, Arsha Nagrani, Cordelia Schmid*
- 353 Efficient and Effective Weakly-Supervised Action Segmentation via Action-Transition-Aware Boundary Alignment, *Angchi Xu, Wei-Shi Zheng*
- 354 Benchmarking the Robustness of Temporal Action Detection Models Against Temporal Corruptions, *Runhao Zeng, Xiaoyong Chen, Jiaming Liang, Huisi Wu, Guangzhong Cao, Yong Guo*
- 355 A Backpack Full of Skills: Egocentric Video Understanding with Diverse Task Perspectives, *Simone Alberto Peirone, Francesca Pistilli, Antonio Alliegro, Giuseppe Averta*
- 356 Summarize the Past to Predict the Future: Natural Language Descriptions of Context Boost Multimodal Object Interaction Anticipation, *Razvan-George Pasca, Alexey Gavryushin, Muhammad Hamza, Yen-Ling Kuo, Kaichun Mo, Luc Van Gool, Otmar Hilliges, Xi Wang*
- 357 Open-Vocabulary Video Anomaly Detection, *Peng Wu, Xuerong Zhou, Guansong Pang, Yujia Sun, Jing Liu, Peng Wang, Yanning Zhang*
- 358 Task-Driven Exploration: Decoupling and Inter-Task Feedback for Joint Moment Retrieval and Highlight Detection, *Jin Yang, Ping Wei, Huan Li, Ziyang Ren*
- 359 Prompt-Enhanced Multiple Instance Learning for Weakly Supervised Video Anomaly Detection, *Junxi Chen, Liang Li, Li Su, Zheng-jun Zha, Qingming Huang*
- 360 Context-Guided Spatio-Temporal Video Grounding, *Xin Gu, Heng Fan, Yan Huang, Tiejian Luo, Libo Zhang*
- 361 Just Add ?! Pose Induced Video Transformers for Understanding Activities of Daily Living, *Dominick Reilly, Srijan Das*
- 362 Action Detection via an Image Diffusion Process, *Lin Geng Foo, Tianjiao Li, Hossein Rahmani, Jun Liu*
- 363 LLMs are Good Sign Language Translators, *Jia Gong, Lin Geng Foo, Yixuan He, Hossein Rahmani, Jun Liu*
- 364 End-to-End Spatio-Temporal Action Localisation with Video Transformers, *Alexey A. Gritsenko, Xuehan Xiong, Josip Djolonga, Mostafa Dehghani, Chen Sun, Mario Lucic, Cordelia Schmid, Anurag Arnab*
- 365 HIG: Hierarchical Interlacement Graph Approach to Scene Graph Generation in Video Understanding, *Trong-Thuan Nguyen, Pha Nguyen, Khoa Luu*
- 366 LLMs are Good Action Recognizers, *Haoxuan Qu, Yujun Cai, Jun Liu*
- 367 VideoLLM-online: Online Video Large Language Model for Streaming Video, *Joya Chen, Zhaoyang Lv, Shiwei Wu, Kevin Qinghong Lin, Chenan Song, Difei Gao, Jia-Wei Liu, Ziteng Gao, Dongxing Mao, Mike Zheng Shou*
- 368 What When and Where? Self-Supervised Spatio-Temporal Grounding in Untrimmed Multi-Action Videos from Narrated Instructions, *Brian Chen, Nina Shvetsova, Andrew Rouditchenko, Daniel Kondermann, Samuel Thomas, Shih-Fu Chang, Rogerio Feris, James Glass, Hilde Kuehne*
- 369 Narrative Action Evaluation with Prompt-Guided Multimodal Interaction, *Shiyi Zhang, Sule Bai, Guangyi Chen, Lei Chen, Jiwen Lu, Junle Wang, Yansong Tang*
- 370 Realigning Confidence with Temporal Saliency Information for Point-Level Weakly-Supervised Temporal Action Localization, *Ziyang Xia, Jian Cheng, Siyu Liu, Yongxiang Hu, Shiguang Wang, Yijie Zhang, Liwan Dang*
- 371 Action-slot: Visual Action-centric Representations for Multi-label Atomic Activity Recognition in Traffic Scenes, *Chi-Hsi Kung, Shu-Wei Lu, Yi-Hsuan Tsai, Yi-Ting Chen*
- 372 LoCoNet: Long-Short Context Network for Active Speaker Detection, *Xizi Wang, Feng Cheng, Gedas Bertasius*
- 373 Neighbor Relations Matter in Video Scene Detection, *Jiawei Tan, Hongxing Wang, Jiaxin Li, Zhilong Ou, Zhangbin Qian*
- 374 PREGO: Online Mistake Detection in PROcedural EGOcentric Videos, *Alessandro Flaborea, Guido Maria D'Amely di Melendugno, Leonardo Plini, Luca Scofano, Edoardo De Matteis, Antonino Furnari, Giovanni Maria Farinella, Fabio Galasso*

- 375 Learning Object State Changes in Videos: An Open-World Perspective, *Zihui Xue, Kumar Ashutosh, Kristen Grauman*
- 376 Enhanced Motion-Text Alignment for Image-to-Video Transfer Learning, *Wei Zhang, Chaoqun Wan, Tongliang Liu, Xinmei Tian, Xu Shen, Jieping Ye*
- 377 Asymmetric Masked Distillation for Pre-Training Small Foundation Models, *Zhiyu Zhao, Bingkun Huang, Sen Xing, Gangshan Wu, Yu Qiao, Limin Wang*
- 378 Harnessing Large Language Models for Training-free Video Anomaly Detection, *Luca Zanella, Willi Menapace, Massimiliano Mancini, Yiming Wang, Elisa Ricci*
- 379 SportsHHI: A Dataset for Human-Human Interaction Detection in Sports Videos, *Tao Wu, Runyu He, Gangshan Wu, Limin Wang*
- 380 VicTR: Video-conditioned Text Representations for Activity Recognition, *Kumara Kahatapitiya, Anurag Arnab, Arsha Nagrani, Michael S. Ryoo*
- 381 Dual DETRs for Multi-Label Temporal Action Detection, *Yuhan Zhu, Guozhen Zhang, Jing Tan, Gangshan Wu, Limin Wang*
- 382 Adapting Short-Term Transformers for Action Detection in Untrimmed Videos, *Min Yang, Huan Gao, Ping Guo, Limin Wang*
- 383 Can't Make an Omelette Without Breaking Some Eggs: Plausible Action Anticipation Using Large Video-Language Models, *Himangi Mittal, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee*
- 384 End-to-End Temporal Action Detection with 1B Parameters Across 1000 Frames, *Shuming Liu, Chen-Lin Zhang, Chen Zhao, Bernard Ghanem*
- 385 RMem: Restricted Memory Banks Improve Video Object Segmentation, *Junbao Zhou, Ziqi Pang, Yu-Xiong Wang*
- 386 Low-power Continuous Remote Behavioral Localization with Event Cameras, *Friedhelm Hamann, Suman Ghosh, Ignacio Juarez Martinez, Tom Hart, Alex Kacelnik, Guillermo Gallego*
- 387 Action Scene Graphs for Long-Form Understanding of Egocentric Videos, *Ivan Rodin, Antonino Furnari, Kyle Min, Subarna Tripathi, Giovanni Maria Farinella*
- 388 ExACT: Language-guided Conceptual Reasoning and Uncertainty Estimation for Event-based Action Recognition and More, *Jiazhou Zhou, Xu Zheng, Yuanhuiyi Lyu, Lin Wang*
- 389 Uncertainty-aware Action Decoupling Transformer for Action Anticipation, *Hongji Guo, Nakul Agarwal, Shao-Yuan Lo, Kwonjoon Lee, Qiang Ji*
- 390 Error Detection in Egocentric Procedural Task Videos, *Shih-Po Lee, Zijia Lu, Zekun Zhang, Minh Hoai, Ehsan Elhamifar*
- 391 Learning to Predict Activity Progress by Self-Supervised Video Alignment, *Gerard Donahue, Ehsan Elhamifar*
- 392 MaskCLR: Attention-Guided Contrastive Learning for Robust Action Representation Learning, *Mohamed Abdelfattah, Mariam Hassan, Alexandre Alahi*
- 393 Align Before Adapt: Leveraging Entity-to-Region Alignments for Generalizable Video Action Recognition, *Yifei Chen, Dapeng Chen, Ruijin Liu, Sai Zhou, Wenyuan Xue, Wei Peng*
- 394 DIBS: Enhancing Dense Video Captioning with Unlabeled Videos via Pseudo Boundary Enrichment and Online Refinement, *Hao Wu, Huabin Liu, Yu Qiao, Xiao Sun*
- 395 Bridging the Gap: A Unified Video Comprehension Framework for Moment Retrieval and Highlight Detection, *Yicheng Xiao, Zhuoyan Luo, Yong Liu, Yue Ma, Hengwei Bian, Yatai Ji, Yujia Yang, Xiu Li*
- 396 Test-Time Zero-Shot Temporal Action Localization, *Benedetta Liberatori, Alessandro Conti, Paolo Rota, Yiming Wang, Elisa Ricci*
- 397 Selective Interpretable and Motion Consistent Privacy Attribute Obfuscation for Action Recognition, *Filip Ilic, He Zhao, Thomas Pock, Richard P. Wildes*
- 398 Step Differences in Instructional Video, *Tushar Nagarajan, Lorenzo Torresani*
- 399 Compositional Video Understanding with Spatiotemporal Structure-based Transformers, *Hoyeoung Yun, Jinwoo Ahn, Minseo Kim, Eun-Sol Kim*
- 400 Temporally Consistent Unbalanced Optimal Transport for Unsupervised Action Segmentation, *Ming Xu, Stephen Gould*
- 401 FineParser: A Fine-grained Spatio-temporal Action Parser for Human-centric Action Quality Assessment, *Jinglin Xu, Sibao Yin, Guohao Zhao, Zishuo Wang, Yuxin Peng*
- 402 Part-aware Unified Representation of Language and Skeleton for Zero-shot Action Recognition, *Anqi Zhu, Qihong Ke, Mingming Gong, James Bailey*
- 403 vid-TLDR: Training Free Token Merging for Light-weight Video Transformer, *Joonmyung Choi, Sanghyeok Lee, Jaewon Chu, Minhyuk Choi, Hyunwoo J. Kim*
- 404 CPR-Coach: Recognizing Composite Error Actions based on Single-class Training, *Shunli Wang, Shuaibing Wang, Dingkang Yang, Mingcheng Li, Haopeng Kuang, Xiao Zhao, Liuzhen Su, Peng Zhai, Lihua Zhang*
- 405 Uncovering What Why and How: A Comprehensive Benchmark for Causation Understanding of Video Anomaly, *Hang Du, Sicheng Zhang, Binzhu Xie, Guoshun Nan, Jiayang Zhang, Junrui Xu, Hanyu Liu, Sicong Leng, Jiangming Liu, Hehe Fan, Dajiu Huang, Jing Feng, Linli Chen, Can Zhang, Xuhuan Li, Hao Zhang, Jianhang Chen, Qimei Cui, Xiaofeng Tao*
- 406 Detours for Navigating Instructional Videos, *Kumar Ashutosh, Zihui Xue, Tushar Nagarajan, Kristen Grauman*
- 407 Why Not Use Your Textbook? Knowledge-Enhanced Procedure Planning of Instructional Videos, *Kumaranage Ravindu Yasas Nagasinghe, Honglu Zhou, Malitha Gunawardhana, Martin Rengqiang Min, Daniel Harari, Muhammad Haris Khan*
- 408 Multiscale Vision Transformers Meet Bipartite Matching for Efficient Single-stage Action Localization, *Ioanna Ntinou, Enrique Sanchez, Georgios Tzimiropoulos*
- 409 TE-TAD: Towards Full End-to-End Temporal Action Detection via Time-Aligned Coordinate Expression, *Ho-Joong Kim, Jung-Ho Hong, Heejo Kong, Seong-Whan Lee*
- 410 CSTA: CNN-based Spatiotemporal Attention for Video Summarization, *Jaewon Son, Jaehun Park, Kwangsu Kim*
- 411 PeVL: Pose-Enhanced Vision-Language Model for Fine-Grained Human Action Recognition, *Haosong Zhang, Mei Chee Leong, Liyuan Li, Weisi Lin*
- 412 MULDE: Multiscale Log-Density Estimation via Denoising Score Matching for Video Anomaly Detection, *Jakub Micorek, Horst Possegger, Dominik Narnhofer, Horst Bischof, Mateusz Kozinski*
- 413 Language Model Guided Interpretable Video Action Reasoning, *Ning Wang, Guangming Zhu, HS Li, Liang Zhang, Syed Afaq Ali Shah, Mohammed Benamoun*
- 414 OST: Refining Text Knowledge with Optimal Spatio-Temporal Descriptor for General Video Recognition, *Tongjia Chen, Hongshan Yu, Zhengeng Yang, Zechuan Li, Wei Sun, Chen Chen*
- 415 Text Prompt with Normality Guidance for Weakly Supervised Video Anomaly Detection, *Zhiwei Yang, Jing Liu, Peng Wu*
- 416 VideoGrounding-DINO: Towards Open-Vocabulary Spatio-Temporal Video Grounding, *Syed Talal Wasim, Muzammal Naseer, Salman Khan, Ming-Hsuan Yang, Fahad Shahbaz Khan*
- 417 Unsupervised Video Domain Adaptation with Masked Pre-Training and Collaborative Self-Training, *Arun Reddy, William Paul, Corban Rivera, Ketul Shah, Celso M. de Melo, Rama Chellappa*
- 418 SnAG: Scalable and Accurate Video Grounding, *Fangzhou Mu, Sicheng Mo, Yin Li*
- 419 Learning Correlation Structures for Vision Transformers, *Manjin Kim, Paul Hongsuck Seo, Cordelia Schmid, Minsu Cho*
- 420 Weakly-Supervised Audio-Visual Video Parsing with Prototype-based Pseudo-Labeling, *Kranthi Kumar Rachavarapu, Kalyan Ramakrishnan, Rajagopalan A. N.*
- 421 Matching Anything by Segmenting Anything, *Siyuan Li, Lei Ke, Martin Danelljan, Luigi Piccinelli, Mattia Segu, Luc Van Gool, Fisher Yu*
- 422 3D Feature Tracking via Event Camera, *Siqi Li, Zhikuan Zhou, Zhou Xue, Yipeng Li, Shaoyi Du, Yue Gao*
- 423 Frequency Decoupling for Motion Magnification via Multi-Level Isomorphic Architecture, *Fei Wang, Dan Guo, Kun Li, Zhun Zhong, Meng Wang*
- 424 Towards Generalizable Multi-Object Tracking, *Zheng Qin, Le Wang, Sanping Zhou, Panpan Fu, Gang Hua, Wei Tang*
- 425 SocialCircle: Learning the Angle-based Social Interaction Representation for Pedestrian Trajectory Prediction, *Conghao*

Friday, June 21

- 8:00 - 14:00** **Registration / Badge Pickup** (Summit Lobby)
7:00 - 17:00 **Press Room** (Summit 340)
7:00 - 17:00 **Mother's Room** (Summit 341-adjacent and Summit 441-adjacent)
7:00 - 17:00 **Prayer or Quiet Room** (Summit 326)
8:00 - 9:30 **Breakfast** (Summit ExHall 1-2)
9:00 - 10:00 **EXPO TRACK - Ece Kamar, VP, Managing Director of AI Frontiers at Microsoft Research** (Arch 4F)

9:00 - 10:30 **Orals 5A: Datasets and Evaluation**
(Summit Ballroom)

- 1 Deep Generative Model based Rate-Distortion for Image Downscaling Assessment, *Yuanbang Liang, Bhavesh Garg, Paul Rosin, Yipeng Qin*
- 2 360+x: A Panoptic Multi-modal Scene Understanding Dataset, *Hao Chen, Yuqi Hou, Chenyuan Qu, Irene Testini, Xiaohan Hong, Jianbo Jiao*
- 3 Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives, *Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, Eugene Byrne, Zach Chavis, Joya Chen, Feng Cheng, Fu-Jen Chu, Sean Crane, Avijit Dasgupta, Jing Dong, Maria Escobar, Cristhian Forigua, Abrahm Gebreselasie, Sanjay Haresh, Jing Huang, Md Mohaiminul Islam, Suyog Jain, Rawal Khirodkar, Devansh Kukreja, Kevin J Liang, Jia-Wei Liu, Sagnik Majumder, Yongsen Mao, Miguel Martin, Effrosyni Mavroudi, Tushar Nagarajan, Francesco Ragusa, Santhosh Kumar Ramakrishnan, Luigi Seminara, Arjun Somayazulu, Yale Song, Shan Su, Zihui Xue, Edward Zhang, Jinxu Zhang, Angela Castillo, Changan Chen, Xinzhu Fu, Ryosuke Furuta, Cristina Gonzalez, Prince Gupta, Jiabo Hu, Yifei Huang, Yiming Huang, Weslie Khoo, Anush Kumar, Robert Kuo, Sach Lakhavani, Miao Liu, Mi Luo, Zhengyi Luo, Brigid Meredith, Austin Miller, Oluwatumininu Oguntola, Xiaqing Pan, Penny Peng, Shraman Pramanick, Merey Ramazanova, Fiona Ryan, Wei Shan, Kiran Somasundaram, Chenan Song, Audrey Southerland, Masatoshi Tateno, Huiyu Wang, Yuchen Wang, Takuma Yagi, Mingfei Yan, Xitong Yang, Zecheng Yu, Shengxin Cindy Zha, Chen Zhao, Ziwei Zhao, Zhifan Zhu, Jeff Zhuo, Pablo Arbelaez, Gedas Bertasius, Dima Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C.V. Jawahar, Richard Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, Michael Wray*
- 4 Rich Human Feedback for Text-to-Image Generation, *Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, Junjie Ke, Krishnamurthy Dj Dvijotham, Katherine M. Collins, Yiwen Luo, Yang Li, Kai J Kohlhoff, Deepak Ramachandran, Vidhya Navalpakkam*
- 5 BioCLIP: A Vision Foundation Model for the Tree of Life, *Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun Chao, Yu Su*

9:00 - 10:30 **Orals 5B: 3D from Multiview and Sensors**
(Summit Flex Hall AB)

- 1 Grounding and Enhancing Grid-based Models for Neural Fields, *Zelin Zhao, Fenglei Fan, Wenlong Liao, Junchi Yan*
- 2 NeRF-HuGS: Improved Neural Radiance Fields in Non-static Scenes Using Heuristics-Guided Segmentation, *Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, Guanbin Li*
- 3 Mip-Splatting: Alias-free 3D Gaussian Splatting, *Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger*
- 4 pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable Generalizable 3D Reconstruction, *David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, Vincent Sitzmann*
- 5 Learning to Produce Semi-dense Correspondences for Visual

Localization, *Khong Truong Giang, Soohwan Song, Sungho Jo*

9:00 - 10:30 **Orals 5C: Low-Shot, Self-Supervised, Semi-Supervised Learning** (Summit Flex Hall C)

- 1 CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning, *Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng*
- 2 MLP Can Be A Good Transformer Learner, *Sihao Lin, Pumeng Lyu, Dongrui Liu, Tao Tang, Xiaodan Liang, Andy Song, Xiaojun Chang*
- 3 From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation, *Hyeokjun Kweon, Kuk-Jin Yoon*
- 4 LTGC: Long-tail Recognition via Leveraging LLMs-driven Generated Content, *Qihao Zhao, Yalun Dai, Hao Li, Wei Hu, Fan Zhang, Jun Liu*
- 5 Improving Semantic Correspondence with Viewpoint-Guided Spherical Maps, *Octave Mariotti, Oisin Mac Aodha, Hakan Bilen*

10:00 - 10:30 **Poster Setup** (Arch 4E)

10:30 - 12:00 **Poster Session 5 & Exhibit Hall** (Arch 4A-E)

- 1 TeMO: Towards Text-Driven 3D Stylization for Multi-Object Meshes, *Xuying Zhang, Bo-Wen Yin, Yuming Chen, Zheng Lin, Yunheng Li, Qibin Hou, Ming-Ming Cheng*
- 2 Event-based Structure-from-Orbit, *Ethan Elms, Yasir Latif, Tae Ha Park, Tat-Jun Chin*
- 3 Towards Large-scale 3D Representation Learning with Multi-dataset Point Prompt Training, *Xiaoyang Wu, Zhuotao Tian, Xin Wen, Bohao Peng, Xihui Liu, Kaicheng Yu, Hengshuang Zhao*
- 4 LidaRF: Delving into Lidar for Neural Radiance Field on Street Scenes, *Shanlin Sun, Bingbing Zhuang, Ziyu Jiang, Buyu Liu, Xiaohui Xie, Manmohan Chandraker*
- 5 Instantaneous Perception of Moving Objects in 3D, *Di Liu, Bingbing Zhuang, Dimitris N. Metaxas, Manmohan Chandraker*
- 6 Implicit Event-RGBD Neural SLAM, *Delin Qu, Chi Yan, Dong Wang, Jie Yin, Qizhi Chen, Dan Xu, Yiting Zhang, Bin Zhao, Xuelong Li*
- 7 GS-SLAM: Dense Visual SLAM with 3D Gaussian Splatting, *Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, Xuelong Li*
- 8 Learning Instance-Aware Correspondences for Robust Multi-Instance Point Cloud Registration in Cluttered Scenes, *Zhiyuan Yu, Zheng Qin, Lintao Zheng, Kai Xu*
- 9 MeshGPT: Generating Triangle Meshes with Decoder-Only Transformers, *Yawar Siddiqui, Antonio Allegro, Alexey Artemov, Tatiana Tommasi, Daniele Sirigatti, Vladislav Rosov, Angela Dai, Matthias Nießner*
- 10 Multi-Session SLAM with Differentiable Wide-Baseline Pose Optimization, *Lahav Lipson, Jia Deng*
- 11 SHINOBI: Shape and Illumination using Neural Object Decomposition via BRDF Optimization In-the-wild, *Andreas Engelhardt, Amit Raj, Mark Boss, Yunzhi Zhang, Abhishek Kar, Yuanzhen Li, Deqing Sun, Ricardo Martin Brualla, Jonathan T. Barron, Hendrik P. A. Lensch, Varun Jampani*
- 12 HybridNeRF: Efficient Neural Rendering via Adaptive Volumetric Surfaces, *Haithem Turki, Vasu Agrawal, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder, Deva Ramanan, Michael Zollhöfer, Christian Richardt*
- 13 PLGSLAM: Progressive Neural Scene Representation with Local to Global Bundle Adjustment, *Tianchen Deng, Guole Shen, Tong Qin, Jianyu Wang, Wentao Zhao, Jingchuan Wang, Danwei Wang, Weidong Chen*
- 14 Gear-NeRF: Free-Viewpoint Rendering and Tracking with Motion-aware Spatio-Temporal Sampling, *Xinhang Liu, Yu-Wing Tai, Chi-Keung Tang, Pedro Miraldo, Suhas Lohit, Moitreyee Chatterjee*
- 15 GPS-Gaussian: Generalizable Pixel-wise 3D Gaussian Splatting for Real-time Human Novel View Synthesis, *Shunyuan Zheng, Boyao Zhou, Ruizhi Shao, Boning Liu, Shengping Zhang, Liqiang Nie, Yebin Liu*
- 16 HyperSDFusion: Bridging Hierarchical Structures in Language and Geometry for Enhanced 3D Text2Shape Generation, *Zhiying Leng, Tolga Birdal, Xiaohui Liang, Federico Tombari*
- 17 Selective-Stereo: Adaptive Frequency Information Selection for Stereo Matching, *Xianqi Wang, Gangwei Xu, Hao Jia, Xin Yang*

- 18 Animatable Gaussians: Learning Pose-dependent Gaussian Maps for High-fidelity Human Avatar Modeling, *Zhe Li, Zerong Zheng, Lizhen Wang, Yebin Liu*
- 19 Global Latent Neural Rendering, *Thomas Tanay, Matteo Maggioni*
- 20 HiFi4G: High-Fidelity Human Performance Rendering via Compact Gaussian Splatting, *Yuheng Jiang, Zhehao Shen, Penghao Wang, Zhuo Su, Yu Hong, Yingliang Zhang, Jingyi Yu, Lan Xu*
- 21 LoS: Local Structure-Guided Stereo Matching, *Kunhong Li, Longguang Wang, Ye Zhang, Kaiwen Xue, Shunbo Zhou, Yulan Guo*
- 22 EmbodiedScan: A Holistic Multi-Modal 3D Perception Suite Towards Embodied AI, *Tai Wang, Xiaohan Mao, Chenming Zhu, Runsen Xu, Ruiyuan Lyu, Peisen Li, Xiao Chen, Wenwei Zhang, Kai Chen, Tianfan Xue, Xihui Liu, Cewu Lu, Dahua Lin, Jiangmiao Pang*
- 23 Masked Spatial Propagation Network for Sparsity-Adaptive Depth Refinement, *Jinyoung Jun, Jae-Han Lee, Chang-Su Kim*
- 24 CausalPC: Improving the Robustness of Point Cloud Classification by Causal Effect Identification, *Yuanmin Huang, Mi Zhang, Daizong Ding, Erling Jiang, Zhaoxiang Wang, Min Yang*
- 25 RoMa: Robust Dense Feature Matching, *Johan Edstedt, Qiyu Sun, Georg Bökman, Mårten Wadenbäck, Michael Felsberg*
- 26 MVHumanNet: A Large-scale Dataset of Multi-view Daily Dressing Human Captures, *Zhangyang Xiong, Chenghong Li, Kenkun Liu, Hongjie Liao, Jianqiao Hu, Junyi Zhu, Shuliang Ning, Lingteng Qiu, Chongjie Wang, Shijie Wang, Shuguang Cui, Xiaoguang Han*
- 27 GES: Generalized Exponential Splatting for Efficient Radiance Field Rendering, *Abdullah Hamdi, Luke Melas-Kyriazi, Jinjie Mai, Guocheng Qian, Ruoshi Liu, Carl Vondrick, Bernard Ghanem, Andrea Vedaldi*
- 28 RegionPLC: Regional Point-Language Contrastive Learning for Open-World 3D Scene Understanding, *Jihan Yang, Runyu Ding, Weipeng Deng, Zhe Wang, Xiaojuan Qi*
- 29 NeLF-Pro: Neural Light Field Probes for Multi-Scale Novel View Synthesis, *Zinuo You, Andreas Geiger, Anpei Chen*
- 30 LEAP-VO: Long-term Effective Any Point Tracking for Visual Odometry, *Weirong Chen, Le Chen, Rui Wang, Marc Pollefeys*
- 31 FAR: Flexible Accurate and Robust 6DoF Relative Camera Pose Estimation, *Chris Rockwell, Nilesh Kulkarni, Linyi Jin, Jeong Joon Park, Justin Johnson, David F. Fouhey*
- 32 OmniGlue: Generalizable Feature Matching with Foundation Model Guidance, *Hanwen Jiang, Arjun Karpur, Bingyi Cao, Qixing Huang, André Araujo*
- 33 GART: Gaussian Articulated Template Models,
* *Jiahui Lei, Yufu Wang, Georgios Pavlakos, Lingjie Liu, Kostas Daniilidis*
- 34 CG-HOI: Contact-Guided 3D Human-Object Interaction Generation, *Christian Diller, Angela Dai*
- 35 FutureHuman3D: Forecasting Complex Long-Term 3D Human Behavior from Video Observations, *Christian Diller, Thomas Funkhouser, Angela Dai*
- 36 PI3D: Efficient Text-to-3D Generation with Pseudo-Image Diffusion, *Ying-Tian Liu, Yuan-Chen Guo, Guan Luo, Heyi Sun, Wei Yin, Song-Hai Zhang*
- 37 Building a Strong Pre-Training Baseline for Universal 3D Large-Scale Perception, *Haoming Chen, Zhizhong Zhang, Yanyun Qu, Ruixin Zhang, Xin Tan, Yuan Xie*
- 38 COTR: Compact Occupancy TRansformer for Vision-based 3D Occupancy Prediction, *Qihang Ma, Xin Tan, Yanyun Qu, Lizhuang Ma, Zhizhong Zhang, Yuan Xie*
- 39 SelfOcc: Self-Supervised Vision-Based 3D Occupancy Prediction, *Yuanhui Huang, Wenzhao Zheng, Borui Zhang, Jie Zhou, Jiwen Lu*
- 40 UnScene3D: Unsupervised 3D Instance Segmentation for Indoor Scenes, *David Rozenberszki, Or Litany, Angela Dai*
- 41 NEAT: Distilling 3D Wireframes from Neural Attraction Fields, *Nan Xue, Bin Tan, Yuxi Xiao, Liang Dong, Gui-Song Xia, Tianfu Wu, Yujun Shen*
- 42 NeRF-HuGS: Improved Neural Radiance Fields in Non-static Scenes Using Heuristics-Guided Segmentation,
* *Jiahao Chen, Yipeng Qin, Lingjie Liu, Jiangbo Lu, Guanbin Li*
- 43 3DInAction: Understanding Human Actions in 3D Point Clouds,
* *Yizhak Ben-Shabat, Oren Shtrout, Stephen Gould*
- 44 Dynamic LiDAR Re-simulation using Compositional Neural Fields,
* *Hanfeng Wu, Xingxing Zuo, Stefan Leutenegger, Or Litany, Konrad Schindler, Shengyu Huang*
- 45 Inverse Rendering of Glossy Objects via the Neural Plenoptic Function and Radiance Fields, *Haoyuan Wang, Wenbo Hu, Lei Zhu, Rynson W.H. Lau*
- 46 PanoPose: Self-supervised Relative Pose Estimation for Panoramic Images, *Diantao Tu, Hainan Cui, Xianwei Zheng, Shuhan Shen*
- 47 GeoAuxNet: Towards Universal 3D Representation Learning for Multi-sensor Point Clouds, *Shengjun Zhang, Xin Fei, Yueqi Duan*
- 48 4K4D: Real-Time 4D View Synthesis at 4K Resolution, *Zhen Xu, Sida Peng, Haotong Lin, Guangzhao He, Jiaming Sun, Yujun Shen, Hujun Bao, Xiaowei Zhou*
- 49 MuRF: Multi-Baseline Radiance Fields, *Haofei Xu, Anpei Chen, Yuedong Chen, Christos Sakaridis, Yulun Zhang, Marc Pollefeys, Andreas Geiger, Fisher Yu*
- 50 LangSplat: 3D Language Gaussian Splatting,
* *Minghan Qin, Wanhua Li, Jiawei Zhou, Haoqian Wang, Hanspeter Pfister*
- 51 Bays' Rays: Uncertainty Quantification for Neural Radiance Fields,
* *Lily Goli, Cody Reading, Silvia Sellán, Alec Jacobson, Andrea Tagliasacchi*
- 52 Accelerating Neural Field Training via Soft Mining, *Shakiba Kheradmand, Daniel Rebain, Gopal Sharma, Hossam Isack, Abhishek Kar, Andrea Tagliasacchi, Kwang Moo Yi*
- 53 CORE-MPI: Consistency Object Removal with Embedding MultiPlane Image, *Donggeun Yoon, Donghyeon Cho*
- 54 NECA: Neural Customizable Human Avatar, *Junjin Xiao, Qing Zhang, Zhan Xu, Wei-Shi Zheng*
- 55 S-DyRF: Reference-Based Stylized Radiance Fields for Dynamic Scenes, *Xingyi Li, Zhiguo Cao, Yizheng Wu, Kewei Wang, Ke Xian, Zhe Wang, Guosheng Lin*
- 56 BEVNeXt: Reviving Dense BEV Frameworks for 3D Object Detection, *Zhenxin Li, Shiyi Lan, Jose M. Alvarez, Zuxuan Wu*
- 57 Bi-SSC: Geometric-Semantic Bidirectional Fusion for Camera-based 3D Semantic Scene Completion, *Yujie Xue, Ruihui Li, Fan Wu, Zhuo Tang, Kenli Li, Mingxing Duan*
- 58 Learning to Select Views for Efficient Multi-View Understanding, *Yunzhong Hou, Stephen Gould, Liang Zheng*
- 59 Outdoor Scene Extrapolation with Hierarchical Generative Cellular Automata, *Dongsu Zhang, Francis Williams, Zan Gojcic, Karsten Kreis, Sanja Fidler, Young Min Kim, Amlan Kar*
- 60 Spectrum AUC Difference (SAUCD): Human-aligned 3D Shape Evaluation, *Tianyu Luan, Zhong Li, Lele Chen, Xuan Gong, Lichang Chen, Yi Xu, Junsong Yuan*
- 61 Federated Online Adaptation for Deep Stereo, *Matteo Poggi, Fabio Tosi*
- 62 Instruct 4D-to-4D: Editing 4D Scenes as Pseudo-3D Scenes Using 2D Diffusion, *Linzhan Mou, Jun-Kun Chen, Yu-Xiong Wang*
- 63 Real-time Acquisition and Reconstruction of Dynamic Volumes with Neural Structured Illumination, *Yixin Zeng, Zoubin Bi, Mingrui Yin, Xiang Feng, Kun Zhou, Hongzhi Wu*
- 64 Unifying Correspondence Pose and NeRF for Generalized Pose-Free Novel View Synthesis, *Sunghwan Hong, Jaewoo Jung, Heeseong Shin, Jiaolong Yang, Seungryong Kim, Chong Luo*
- 65 GoMVS: Geometrically Consistent Cost Aggregation for Multi-View Stereo, *Jiang Wu, Rui Li, Haofei Xu, Wenxun Zhao, Yu Zhu, Jinqiu Sun, Yanning Zhang*
- 66 MESA: Matching Everything by Segmenting Anything, *Yesheng Zhang, Xu Zhao*
- 67 OmniSDF: Scene Reconstruction using Omnidirectional Signed Distance Functions and Adaptive Binotrees, *Hakyeon Kim, Andreas Meuleman, Hyeonjoong Jang, James Tompkin, Min H. Kim*
- 68 MirageRoom: 3D Scene Segmentation with 2D Pre-trained Models by Mirage Projection, *Haowen Sun, Yueqi Duan, Juncheng Yan, Yifan Liu, Jiwen Lu*
- 69 Robust Synthetic-to-Real Transfer for Stereo Matching, *Jiawei Zhang, Jiahe Li, Lei Huang, Xiaohan Yu, Lin Gu, Jin Zheng, Xiao Bai*

- 70 Symphonize 3D Semantic Scene Completion with Contextual Instance Queries, *Haoyi Jiang, Tianheng Cheng, Naiyu Gao, Haoyang Zhang, Tianwei Lin, Wenyu Liu, Xinggang Wang*
- 71 Differentiable Neural Surface Refinement for Modeling Transparent Objects, *Weijian Deng, Dylan Campbell, Chunyi Sun, Shubham Kanitkar, Matthew E. Shaffer, Stephen Gould*
- 72 DeMatch: Deep Decomposition of Motion Field for Two-View Correspondence Learning, *Shihua Zhang, Zizhuo Li, Yuan Gao, Jiayi Ma*
- 73 Is Vanilla MLP in Neural Radiance Field Enough for Few-shot View Synthesis?, *Hanxin Zhu, Tianyu He, Xin Li, Bingchen Li, Zhibo Chen*
- 74 GaussianAvatars: Photorealistic Head Avatars with Rigged 3D Gaussians, *Shenhan Qian, Tobias Kirschstein, Liam Schoneveld, Davide Davoli, Simon Giebenhain, Matthias Nießner*
- 75 4D Gaussian Splatting for Real-Time Dynamic Scene Rendering, *Guanjun Wu, Taoran Yi, Jiemin Fang, Lingxi Xie, Xiaopeng Zhang, Wei Wei, Wenyu Liu, Qi Tian, Xinggang Wang*
- 76 How Far Can We Compress Instant-NGP-Based NeRF?, *Yihang Chen, Qianyi Wu, Mehrtash Harandi, Jianfei Cai*
- 77 Deformable 3D Gaussians for High-Fidelity Monocular Dynamic Scene Reconstruction, *Ziyi Yang, Xinyu Gao, Wen Zhou, Shaohui Jiao, Yuqing Zhang, Xiaogang Jin*
- 78 Learning with Unreliability: Fast Few-shot Voxel Radiance Fields with Relative Geometric Consistency, *Yingjie Xu, Bangzhen Liu, Hao Tang, Bailin Deng, Shengfeng He*
- 79 NTO3D: Neural Target Object 3D Reconstruction with Segment Anything, *Xiaobao Wei, Renrui Zhang, Jiarui Wu, Jiaming Liu, Ming Lu, Yandong Guo, Shanghang Zhang*
- 80 Loopy-SLAM: Dense Neural SLAM with Loop Closures, *Lorenzo Liso, Erik Sandström, Vladimir Yugay, Luc Van Gool, Martin R. Oswald*
- 81 BSNet: Box-Supervised Simulation-assisted Mean Teacher for 3D Instance Segmentation, *Jiahao Lu, Jiacheng Deng, Tianzhu Zhang*
- 82 ExtraNeRF: Visibility-Aware View Extrapolation of Neural Radiance Fields with Diffusion Models, *Meng-Li Shih, Wei-Chiu Ma, Lorenzo Boyice, Aleksander Holynski, Forrester Cole, Brian Curless, Janne Kontkanen*
- 83 Alpha Invariance: On Inverse Scaling Between Distance and Volume Density in Neural Radiance Fields, *Joshua Ahn, Haochen Wang, Raymond A. Yeh, Greg Shakhnarovich*
- 84 SpatialTracker: Tracking Any 2D Pixels in 3D Space, *Yuxi Xiao, Qianqian Wang, Shangzhan Zhang, Nan Xue, Sida Peng, Yujun Shen, Xiaowei Zhou*
- 85 GauHuman: Articulated Gaussian Splatting from Monocular Human Videos, *Shoukang Hu, Tao Hu, Ziwei Liu*
- 86 IPoD: Implicit Field Learning with Point Diffusion for Generalizable 3D Object Reconstruction from Single RGB-D Images, *Yushuang Wu, Luyue Shi, Junhao Cai, Weihao Yuan, Lingteng Qiu, Zilong Dong, Liefeng Bo, Shuguang Cui, Xiaoguang Han*
- 87 GOV-NeSF: Generalizable Open-Vocabulary Neural Semantic Fields, *Yunsong Wang, Hanlin Chen, Gim Hee Lee*
- 88 LASA: Instance Reconstruction from Real Scans using A Large-scale Aligned Shape Annotation Dataset, *Haolin Liu, Chongjie Ye, Yinyu Nie, Yingfan He, Xiaoguang Han*
- 89 GenZI: Zero-Shot 3D Human-Scene Interaction Generation, *Lei Li, Angela Dai*
- 90 MVCPS-NeuS: Multi-view Constrained Photometric Stereo for Neural Surface Reconstruction, *Hiroaki Santo, Fumio Okura, Yasuyuki Matsushita*
- 91 DVMNet: Computing Relative Pose for Unseen Objects Beyond Hypotheses, *Chen Zhao, Tong Zhang, Zheng Dang, Mathieu Salzmann*
- 92 Motion2VecSets: 4D Latent Vector Set Diffusion for Non-rigid Shape Reconstruction and Tracking, *Wei Cao, Chang Luo, Biao Zhang, Matthias Nießner, Jiapeng Tang*
- 93 DiffuScene: Denoising Diffusion Models for Generative Indoor Scene Synthesis, *Jiapeng Tang, Yinyu Nie, Lev Markhasin, Angela Dai, Justus Thies, Matthias Nießner*
- 94 Test-Time Adaptation for Depth Completion, *Hyoungseob Park, Anjali Gupta, Alex Wong*
- 95 Global and Hierarchical Geometry Consistency Priors for Few-shot NeRFs in Indoor Scenes, *Xiaotian Sun, Qingshan Xu, Xinjie Yang, Yu Zang, Cheng Wang*
- 96 KP-RED: Exploiting Semantic Keypoints for Joint 3D Shape Retrieval and Deformation, *Ruida Zhang, Chenyangguang Zhang, Yan Di, Fabian Manhardt, Xingyu Liu, Federico Tombari, Xiangyang Ji*
- 97 Unsigned Orthogonal Distance Fields: An Accurate Neural Implicit Representation for Diverse 3D Shapes, *Yujie Lu, Long Wan, Nayu Ding, Yulong Wang, Shuhan Shen, Shen Cai, Lin Gao*
- 98 DiSR-NeRF: Diffusion-Guided View-Consistent Super-Resolution NeRF, *Jie Long Lee, Chen Li, Gim Hee Lee*
- 99 BANF: Band-Limited Neural Fields for Levels of Detail Reconstruction, *Akhmedkhan Shabanov, Shrisudhan Govindarajan, Cody Reading, Lily Goli, Daniel Rebain, Kwang Moo Yi, Andrea Tagliasacchi*
- 100 SuperNormal: Neural Surface Reconstruction via Multi-View Normal Integration, *Xu Cao, Takafumi Taketomi*
- 101 ADFactory: An Effective Framework for Generalizing Optical Flow with NeRF, *Han Ling, Quansen Sun, Yinghui Sun, Xian Xu, Xinfeng Li*
- 102 Dr.Hair: Reconstructing Scalp-Connected Hair Strands without Pre-Training via Differentiable Rendering of Line Segments, *Yusuke Takimoto, Hikari Takehara, Hiroyuki Sato, Zihao Zhu, Bo Zheng*
- 103 OmniSeg3D: Omniversal 3D Segmentation via Hierarchical Contrastive Learning, *Haiyang Ying, Yixuan Yin, Jinzhi Zhang, Fan Wang, Tao Yu, Ruqi Huang, Lu Fang*
- 104 Visual Programming for Zero-shot Open-Vocabulary 3D Visual Grounding, *Zihao Yuan, Jinke Ren, Chun-Mei Feng, Hengshuang Zhao, Shuguang Cui, Zhen Li*
- 105 GEARS: Local Geometry-aware Hand-object Interaction Synthesis, *Keyang Zhou, Bharat Lal Bhatnagar, Jan Eric Lenssen, Gerard Pons-Moll*
- 106 Edge-Aware 3D Instance Segmentation Network with Intelligent Semantic Prior, *Wonseok Roh, Hwanhee Jung, Giljoo Nam, Jinseop Yeom, Hyunje Park, Sang Ho Yoon, Sangpil Kim*
- 107 Scaffold-GS: Structured 3D Gaussians for View-Adaptive Rendering, *Tao Lu, Mulin Yu, Linning Xu, Yuanbo Xiangli, Limin Wang, Dahua Lin, Bo Dai*
- 108 Map-Relative Pose Regression for Visual Re-Localization, *Shuai Chen, Tommaso Cavallari, Victor Adrian Prisacariu, Eric Brachmann*
- 109 3DGStream: On-the-Fly Training of 3D Gaussians for Efficient Streaming of Photo-Realistic Free-Viewpoint Videos, *Jiakai Sun, Han Jiao, Guangyuan Li, Zhanjie Zhang, Lei Zhao, Wei Xing*
- 110 Revisiting Global Translation Estimation with Feature Tracks, *Peilin Tao, Hainan Cui, Mengqi Rong, Shuhan Shen*
- 111 DUST3R: Geometric 3D Vision Made Easy, *Shuzhe Wang, Vincent Leroy, Yann Cabon, Boris Chidlovskii, Jerome Revaud*
- 112 Robust Depth Enhancement via Polarization Prompt Fusion Tuning, *Kei Ikemura, Yiming Huang, Felix Heide, Zhaoxiang Zhang, Qifeng Chen, Chenyang Lei*
- 113 StraightPCF: Straight Point Cloud Filtering, *Dasith de Silva Edirimuni, Xuequan Lu, Gang Li, Lei Wei, Antonio Robles-Kelly, Hongdong Li*
- 114 NeRFiller: Completing Scenes via Generative 3D inpainting, *Ethan Weber, Aleksander Holynski, Varun Jampani, Saurabh Saxena, Noah Snaveley, Abhishek Kar, Angjoo Kanazawa*
- 115 NeRF Director: Revisiting View Selection in Neural Volume Rendering, *Wenhui Xiao, Rodrigo Santa Cruz, David Ahmmedt-Aristizabal, Olivier Salvado, Clinton Fookes, Leo Lebrat*
- 116 Learning Intra-view and Cross-view Geometric Knowledge for Stereo Matching, *Rui Gong, Weide Liu, Zaiwang Gu, Xulei Yang, Jun Cheng*
- 117 Sherpa3D: Boosting High-Fidelity Text-to-3D Generation via Coarse 3D Prior, *Fangfu Liu, Diankun Wu, Yi Wei, Yongming Rao, Yueqi Duan*
- 118 DNGaussian: Optimizing Sparse-View 3D Gaussian Radiance Fields with Global-Local Depth Normalization, *Jiahe Li, Jiawei Zhang, Xiao Bai, Jin Zheng, Xin Ning, Jun Zhou, Lin Gu*
- 119 A Conditional Denoising Diffusion Probabilistic Model for Point Cloud Upsampling, *Wentao Qu, Yuantian Shao, Lingwu Meng, Xiaoshui Huang, Liang Xiao*

- 120 COLMAP-Free 3D Gaussian Splatting,
* Yang Fu, Sifei Liu, Amey Kulkarni, Jan Kautz, Alexei A. Efros, Xiaolong Wang
- 121 GSNeRF: Generalizable Semantic Neural Radiance Fields with Enhanced 3D Scene Understanding, Zi-Ting Chou, Sheng-Yu Huang, I-Jieh Liu, Yu-Chiang Frank Wang
- 122 Extend Your Own Correspondences: Unsupervised Distant Point Cloud Registration by Progressive Distance Extension, Quan Liu, Hongzi Zhu, Zhenxi Wang, Yunsong Zhou, Shan Chang, Minyi Guo
- 123 Fully Geometric Panoramic Localization, Junho Kim, Jiwon Jeong, Young Min Kim
- 124 Multiway Point Cloud Mosaicking with Diffusion and Global Optimization, Shengze Jin, Iro Armeni, Marc Pollefeys, Daniel Barath
- 125 Mip-Splatting: Alias-free 3D Gaussian Splatting,
* Zehao Yu, Anpei Chen, Binbin Huang, Torsten Sattler, Andreas Geiger
- 126 Generative 3D Part Assembly via Part-Whole-Hierarchy Message Passing, Bi'an Du, Xiang Gao, Wei Hu, Renjie Liao
- 127 Total-Decom: Decomposed 3D Scene Reconstruction with Minimal Interaction, Xiaoyang Lyu, Chirui Chang, Peng Dai, Yang-Tian Sun, Xiaojuan Qi
- 128 Absolute Pose from One or Two Scaled and Oriented Features,
* Jonathan Ventura, Zuzana Kukelova, Torsten Sattler, Daniel Barath
- 129 DGC-GNN: Leveraging Geometry and Color Cues for Visual Descriptor-Free 2D-3D Matching, Shuzhe Wang, Juho Kannala, Daniel Barath
- 130 Entity-NeRF: Detecting and Removing Moving Entities in Urban Scenes, Takashi Otonari, Satoshi Ikehata, Kiyoharu Aizawa
- 131 GaussianEditor: Editing 3D Gaussians Delicately with Text Instructions, Junjie Wang, Jiemin Fang, Xiaopeng Zhang, Lingxi Xie, Qi Tian
- 132 The More You See in 2D the More You Perceive in 3D, Xinyang Han,
* Zelin Gao, Angjoo Kanazawa, Shubham Goel, Yossi Gandelsman
- 133 Multi-Scale 3D Gaussian Splatting for Anti-Aliased Rendering, Zhiwen Yan, Weng Fei Low, Yu Chen, Gim Hee Lee
- 134 Practical Measurements of Translucent Materials with Inter-Pixel Translucency Prior, Zhenyu Chen, Jie Guo, Shuichang Lai, Ruoyu Fu, Mengxun Kong, Chen Wang, Hongyu Sun, Zhebin Zhang, Chen Li, Yanwen Guo
- 135 OneFormer3D: One Transformer for Unified Point Cloud Segmentation, Maxim Kolodiazhnyi, Anna Vorontsova, Anton Konushin, Danila Rukhovich
- 136 General Point Model Pretraining with Autoencoding and Autoregressive, Zhe Li, Zhangyang Gao, Cheng Tan, Bocheng Ren, Laurence T. Yang, Stan Z. Li
- 137 MorpheuS: Neural Dynamic 360° Surface Reconstruction from Monocular RGB-D Video, Hengyi Wang, Jingwen Wang, Lourdes Agapito
- 138 pixelSplat: 3D Gaussian Splats from Image Pairs for Scalable
* Generalizable 3D Reconstruction, David Charatan, Sizhe Lester Li, Andrea Tagliasacchi, Vincent Sitzmann
- 139 Object Dynamics Modeling with Hierarchical Point Cloud-based Representations, Chanho Kim, Li Fuxin
- 140 Neural Refinement for Absolute Pose Regression with Feature Synthesis, Shuai Chen, Yash Bhalgat, Xinghui Li, Jia-Wang Bian, Kejie Li, Zirui Wang, Victor Adrian Prisacariu
- 141 Gaussian Shadow Casting for Neural Characters, Luis Bolanos, Shih-Yang Su, Helge Rhodin
- 142 PAPR in Motion: Seamless Point-level 3D Scene Interpolation,
* Shichong Peng, Yanshu Zhang, Ke Li
- 143 ShapeMatcher: Self-Supervised Joint Shape Canonicalization Segmentation Retrieval and Deformation, Yan Di, Chenyangguang Zhang, Chaowei Wang, Ruida Zhang, Guangyao Zhai, Yanyan Li, Bowen Fu, Xiangyang Ji, Shan Gao
- 144 XScale-NVS: Cross-Scale Novel View Synthesis with Hash Featurized Manifold, Guangyu Wang, Jinzhi Zhang, Fan Wang, Ruqi Huang, Lu Fang
- 145 Instance-Adaptive and Geometric-Aware Keypoint Learning for Category-Level 6D Object Pose Estimation, Xiao Lin, Wenfei Yang, Yuan Gao, Tianzhu Zhang
- 146 RepKPU: Point Cloud Upsampling with Kernel Point Representation and Deformation, Yi Rong, Haoran Zhou, Kang Xia, Cheng Mei, Jiahao Wang, Tong Lu
- 147 ColorPCR: Color Point Cloud Registration with Multi-Stage Geometric-Color Fusion, Juncheng Mu, Lin Bie, Shaoyi Du, Yue Gao
- 148 ConsistDreamer: 3D-Consistent 2D Diffusion for High-Fidelity Scene Editing, Jun-Kun Chen, Samuel Rota Bulò, Norman Müller, Lorenzo Porzi, Peter Kotschieder, Yu-Xiong Wang
- 149 SceneTex: High-Quality Texture Synthesis for Indoor Scenes via * Diffusion Priors, Dave Zhenyu Chen, Haoxuan Li, Hsin-Ying Lee, Sergey Tulyakov, Matthias Nießner
- 150 Aerial Lifting: Neural Urban Semantic and Building Instance Lifting from Aerial Imagery, Yuqi Zhang, Guanying Chen, Jiaying Chen, Shuguang Cui
- 151 Improving Depth Completion via Depth Feature Upsampling, Yufei Wang, Ge Zhang, Shaoqian Wang, Bo Li, Qi Liu, Le Hui, Yuchao Dai
- 152 ZeroRF: Fast Sparse View 360° Reconstruction with Zero Pretraining, Ruoxi Shi, Xinyue Wei, Cheng Wang, Hao Su
- 153 Multi-Level Neural Scene Graphs for Dynamic Urban Environments, Tobias Fischer, Lorenzo Porzi, Samuel Rota Bulò, Marc Pollefeys, Peter Kotschieder
- 154 Gaussian-Flow: 4D Reconstruction with Dynamic 3D Gaussian * Particle, Youtian Lin, Zuo Zhou Dai, Siyu Zhu, Yao Yao
- 155 L4D-Track: Language-to-4D Modeling Towards 6-DoF Tracking and Shape Reconstruction in 3D Point Cloud Stream, Jingtao Sun, Yaonan Wang, Mingtao Feng, Yulan Guo, Ajmal Mian, Mike Zheng Shou
- 156 Neural Directional Encoding for Efficient and Accurate View- * Dependent Appearance Modeling, Liwen Wu, Sai Bi, Zexiang Xu, Fujun Luan, Kai Zhang, Iliyan Georgiev, Kalyan Sunkavalli, Ravi Ramamoorthi
- 157 SNI-SLAM: Semantic Neural Implicit SLAM, Siting Zhu, Guangming Wang, Hermann Blum, Jiuming Liu, Liang Song, Marc Pollefeys, Hesheng Wang
- 158 Enhancing 3D Object Detection with 2D Detection-Guided Query Anchors, Haoxuanye Ji, Pengpeng Liang, Erkang Cheng
- 159 SpecNeRF: Gaussian Directional Encoding for Specular Reflections,
* Li Ma, Vasu Agrawal, Haithem Turki, Changil Kim, Chen Gao, Pedro Sander, Michael Zollhöfer, Christian Richardt
- 160 Correspondence-Free Non-Rigid Point Set Registration Using * Unsupervised Clustering Analysis, Mingyang Zhao, Jingen Jiang, Lei Ma, Shiqing Xin, Gaofeng Meng, Dong-Ming Yan
- 161 GAFusion: Adaptive Fusing LiDAR and Camera with Multiple Guidance for 3D Object Detection, Xiaotian Li, Baojie Fan, Jiandong Tian, Huijie Fan
- 162 3D Neural Edge Reconstruction, Lei Li, Songyou Peng, Zehao Yu, Shaohui Liu, Rémi Pautrat, Xiaochuan Yin, Marc Pollefeys
- 163 AlignMiF: Geometry-Aligned Multimodal Implicit Field for LiDAR- * Camera Joint Synthesis, Tang Tao, Guangrun Wang, Yixing Lao, Peng Chen, Jie Liu, Liang Lin, Kaicheng Yu, Xiaodan Liang
- 164 Polarization Wavefront Lidar: Learning Large Scene Reconstruction from Polarized Wavefronts, Dominik Scheuble, Chenyang Lei, Seung-Hwan Baek, Mario Bijelic, Felix Heide
- 165 A Unified Diffusion Framework for Scene-aware Human Motion Estimation from Sparse Signals, Jiangnan Tang, Jingya Wang, Kaiyang Ji, Lan Xu, Jingyi Yu, Ye Shi
- 166 FaceTalk: Audio-Driven Motion Diffusion for Neural Parametric Head Models, Shivangi Aneja, Justus Thies, Angela Dai, Matthias Nießner
- 167 NeRFCoDec: Neural Feature Compression Meets Neural Radiance Fields for Memory-Efficient Scene Representation, Sicheng Li, Hao Li, Yiyi Liao, Lu Yu
- 168 Open-Vocabulary 3D Semantic Segmentation with Foundation * Models, Li Jiang, Shaoshuai Shi, Bernt Schiele
- 169 GraphDreamer: Compositional 3D Scene Synthesis from Scene Graphs, Gege Gao, Weiyang Liu, Anpei Chen, Andreas Geiger, Bernhard Schölkopf
- 170 OA-CNNs: Omni-Adaptive Sparse CNNs for 3D Semantic Segmentation, Bohao Peng, Xiaoyang Wu, Li Jiang, Yukang Chen, Hengshuang Zhao, Zhuotao Tian, Jiaya Jia
- 171 Efficient Solution of Point-Line Absolute Pose,
* Petr Hruby, Timothy Duff, Marc Pollefeys

- 172 CN-RMA: Combined Network with Ray Marching Aggregation for 3D Indoor Object Detection from Multi-view Images, *Guanlin Shen, Jingwei Huang, Zhihua Hu, Bin Wang*
- 173 HUGS: Holistic Urban 3D Scene Understanding via Gaussian Splatting, *Hongyu Zhou, Jiahao Shao, Lu Xu, Dongfeng Bai, Weichao Qiu, Bingbing Liu, Yue Wang, Andreas Geiger, Yiyi Liao*
- 174 Benchmarking Implicit Neural Representation and Geometric Rendering in Real-Time RGB-D SLAM, *Tongyan Hua, Lin Wang*
- 175 SplatAM: Splat Track & Map 3D Gaussians for Dense RGB-D SLAM, *Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, Jonathon Luiten*
- 176 Lift3D: Zero-Shot Lifting of Any 2D Vision Model to 3D, *Mukund Varma T, Peihao Wang, Zhiwen Fan, Zhangyang Wang, Hao Su, Ravi Ramamoorthi*
- 177 TutteNet: Injective 3D Deformations by Composition of 2D Mesh
* Deformations, *Bo Sun, Thibault Groueix, Chen Song, Qixing Huang, Noam Aigerman*
- 178 L0-Sampler: An L0 Model Guided Volume Sampling for NeRF, *Liangchen Li, Juyong Zhang*
- 179 Text-to-3D using Gaussian Splatting, *Zilong Chen, Feng Wang, Yikai Wang, Huaping Liu*
- 180 TMM: TriAdapter Multi-Modal Learning for 3D Shape Understanding, *Zhihao Zhang, Shengcao Cao, Yu-Xiong Wang*
- 181 FreGS: 3D Gaussian Splatting with Progressive Frequency Regularization, *Jiahui Zhang, Fangneng Zhan, Muyu Xu, Shijian Lu, Eric Xing*
- 182 NelSF: Neural Incident Stokes Field for Geometry and Material
* Estimation, *Chenhao Li, Taishi Ono, Takeshi Uemori, Hajime Mihara, Alexander Gatto, Hajime Nagahara, Yusuke Moriuchi*
- 183 Non-Rigid Structure-from-Motion: Temporally-Smooth Procrustean Alignment and Spatially-Variant Deformation Modeling, *Jiawei Shi, Hui Deng, Yuchao Dai*
- 184 Small Steps and Level Sets: Fitting Neural Surface Models with Point Guidance, *Chamin Hewa Koneputugodage, Yizhak Ben-Shabat, Dylan Campbell, Stephen Gould*
- 185 CVT-xRF: Contrastive In-Voxel Transformer for 3D Consistent Radiance Fields from Sparse Inputs, *Yingji Zhong, Lanqing Hong, Zhenguo Li, Dan Xu*
- 186 GaussianEditor: Swift and Controllable 3D Editing with Gaussian Splatting, *Yiwen Chen, Zilong Chen, Chi Zhang, Feng Wang, Xiaofeng Yang, Yikai Wang, Zhongang Cai, Lei Yang, Huaping Liu, Guosheng Lin*
- 187 Cam4DOcc: Benchmark for Camera-Only 4D Occupancy Forecasting in Autonomous Driving Applications, *Junyi Ma, Xieyuanli Chen, Jiawei Huang, Jingyi Xu, Zhen Luo, Jintao Xu, Weihao Gu, Rui Ai, Hesheng Wang*
- 188 UDiFF: Generating Conditional Unsigned Distance Fields with Optimal Wavelet Diffusion, *Junsheng Zhou, Weiqi Zhang, Baorui Ma, Kanle Shi, Yu-Shen Liu, Zhizhong Han*
- 189 PanoRecon: Real-Time Panoptic 3D Reconstruction from Monocular Video, *Dong Wu, Zike Yan, Hongbin Zha*
- 190 Three Pillars Improving Vision Foundation Model Distillation for Lidar, *Gilles Puy, Spyros Gidaris, Alexandre Boulch, Oriane Siméoni, Corentin Sautier, Patrick Pérez, Andrei Bursuc, Renaud Marlet*
- 191 GARField: Group Anything with Radiance Fields, *Chung Min Kim, Mingxuan Wu, Justin Kerr, Ken Goldberg, Matthew Tancik, Angjoo Kanazawa*
- 192 Flexible Depth Completion for Sparse and Varying Point Densities, *Jinhyung Park, Yu-Jhe Li, Kris Kitani*
- 193 ReconFusion: 3D Reconstruction with Diffusion Priors, *Rundi Wu, Ben Mildenhall, Philipp Henzler, Keunhong Park, Ruiqi Gao, Daniel Watson, Pratul P. Srinivasan, Dor Verbin, Jonathan T. Barron, Ben Poole, Aleksander Holynski*
- 194 GLACE: Global Local Accelerated Coordinate Encoding, *Fangjinhua Wang, Xudong Jiang, Silvano Galliani, Christoph Vogel, Marc Pollefeys*
- 195 NARUTO: Neural Active Reconstruction from Uncertain Target Observations, *Ziyue Feng, Huangying Zhan, Zheng Chen, Qingan Yan, Xiangyu Xu, Changjiang Cai, Bing Li, Qilun Zhu, Yi Xu*
- 196 Photo-SLAM: Real-time Simultaneous Localization and Photorealistic Mapping for Monocular Stereo and RGB-D Cameras, *Huajian Huang, Longwei Li, Hui Cheng, Sai-Kit Yeung*
- 197 Detector-Free Structure from Motion, *Xingyi He, Jiaming Sun, Yifan Wang, Sida Peng, Qixing Huang, Hujun Bao, Xiaowei Zhou*
- 198 Memory-based Adapters for Online 3D Scene Perception, *Xiuwei Xu, Chong Xia, Ziwei Wang, Lingqing Zhao, Yueqi Duan, Jie Zhou, Jiwen Lu*
- 199 SurroundSDF: Implicit 3D Scene Understanding Based on Signed
* Distance Field, *Lizhe Liu, Bohua Wang, Hongwei Xie, Daqi Liu, Li Liu, Zhiqiang Tian, Kuiyuan Yang, Bing Wang*
- 200 CoGS: Controllable Gaussian Splatting, *Heng Yu, Joel Julin, Zoltán A. Milacski, Koichiro Niinuma, László A. Jeni*
- 201 DrivingGaussian: Composite Gaussian Splatting for Surrounding Dynamic Autonomous Driving Scenes, *Xiaoyu Zhou, Zhiwei Lin, Xiaojun Shan, Yongtao Wang, Deqing Sun, Ming-Hsuan Yang*
- 202 GS-IR: 3D Gaussian Splatting for Inverse Rendering, *Zhihao Liang, Qi Zhang, Ying Feng, Ying Shan, Kui Jia*
- 203 Cross-spectral Gated-RGB Stereo Depth Estimation,
* *Samuel Brucker, Stefanie Walz, Mario Bijelic, Felix Heide*
- 204 Efficient LoFTR: Semi-Dense Local Feature Matching with Sparse-Like
* Speed, *Yifan Wang, Xingyi He, Sida Peng, Dongli Tan, Xiaowei Zhou*
- 205 Feature 3DGS: Supercharging 3D Gaussian Splatting to Enable
* Distilled Feature Fields, *Shijie Zhou, Haoran Chang, Sicheng Jiang, Zhiwen Fan, Zehao Zhu, Dejia Xu, Pradyumna Chari, Suyu You, Zhangyang Wang, Achuta Kadambi*
- 206 VGGsFM: Visual Geometry Grounded Deep Structure From Motion,
* *Jianyuan Wang, Nikita Karaev, Christian Rupprecht, David Novotny*
- 207 Dynamic Cues-Assisted Transformer for Robust Point Cloud
* Registration, *Hong Chen, Pei Yan, Sihe Xiang, Yihua Tan*
- 208 Learning to Produce Semi-dense Correspondences for Visual
* Localization, *Khang Truong Giang, Soohwan Song, Sungho Jo*
- 209 GP-NeRF: Generalized Perception NeRF for Context-Aware 3D
* Scene Understanding, *Hao Li, Dingwen Zhang, Yalun Dai, Nian Liu, Lechao Cheng, Jingfeng Li, Jingdong Wang, Junwei Han*
- 210 Compact 3D Gaussian Representation for Radiance Field,
* *Joo Chan Lee, Daniel Rho, Xiangyu Sun, Jong Hwan Ko, Eunbyung Park*
- 211 Unsupervised Occupancy Learning from Sparse Point Cloud,
* *Amine Ouasfi, Adnane Boukhayma*
- 212 Grounding and Enhancing Grid-based Models for Neural Fields,
* *Zelin Zhao, Fenglei Fan, Wenlong Liao, Junchi Yan*
- 213 TACO: Benchmarking Generalizable Bimanual Tool-Action-Object
* Understanding, *Yun Liu, Haolin Yang, Xu Si, Ling Liu, Zipeng Li, Yuxiang Zhang, Yebin Liu, Li Yi*
- 214 ImageNet-D: Benchmarking Neural Network Robustness on
* Diffusion Synthetic Object, *Chenshuang Zhang, Fei Pan, Junmo Kim, In So Kweon, Chengzhi Mao*
- 215 SynFog: A Photo-realistic Synthetic Fog Dataset based on End-to-end Imaging Simulation for Advancing Real-World Defogging in Autonomous Driving, *Yiming Xie, Henglu Wei, Zhenyi Liu, Xiaoyu Wang, Xiangyang Ji*
- 216 FineSports: A Multi-person Hierarchical Sports Video Dataset for Fine-grained Action Understanding, *Jinglin Xu, Guohao Zhao, Sibao Yin, Wenhao Zhou, Yuxin Peng*
- 217 Infinigen Indoors: Photorealistic Indoor Scenes using Procedural Generation, *Alexander Raistrick, Lingjie Mei, Karhan Kayan, David Yan, Yiming Zuo, Beining Han, Hongyu Wen, Meenal Parakh, Stamatis Alexandropoulos, Lahav Lipson, Zeyu Ma, Jia Deng*
- 218 Probing the 3D Awareness of Visual Foundation Models, *Mohamed El Banani, Amit Raj, Kevis-Kokitsi Maninis, Abhishek Kar, Yuanzhen Li, Michael Rubinstein, Deqing Sun, Leonidas Guibas, Justin Johnson, Varun Jampani*
- 219 VBench: Comprehensive Benchmark Suite for Video Generative
* Models, *Ziqi Huang, Yinan He, Jiashuo Yu, Fan Zhang, Chenyang Si, Yuming Jiang, Yuanhan Zhang, Tianxing Wu, Qingyang Jin, Nattapol Chanpaisit, Yaohui Wang, Xinyuan Chen, Limin Wang, Dahua Lin, Yu Qiao, Ziwei Liu*
- 220 MAPLM: A Real-World Large-Scale Vision-Language Benchmark for Map and Traffic Scene Understanding, *Xu Cao, Tong Zhou, Yunsheng Ma, Wenqian Ye, Can Cui, Kun Tang, Zhipeng Cao, Kaizhao Liang, Ziran Wang, James M. Rehg, Chao Zheng*
- 221 Video Recognition in Portrait Mode, *Mingfei Han, Linjie Yang, Xiaojie Jin, Jiashi Feng, Xiaojun Chang, Heng Wang*

- 222 MMVP: A Multimodal MoCap Dataset with Vision and Pressure Sensors, *He Zhang, Shenghao Ren, Haolei Yuan, Jianhui Zhao, Fan Li, Shuangpeng Sun, Zhenghao Liang, Tao Yu, Qiu Shen, Xun Cao*
- 223 What If the TV Was Off? Examining Counterfactual Reasoning Abilities of Multi-modal Language Models, *Letian Zhang, Xiaotong Zhai, Zhongkai Zhao, Yongshuo Zong, Xin Wen, Bingchen Zhao*
- 224 COCONut: Modernizing COCO Segmentation, *Xueqing Deng, Qihang Yu, Peng Wang, Xiaohui Shen, Liang-Chieh Chen*
- 225 Traffic Scene Parsing through the TSP6K Dataset, *Peng-Tao Jiang, Yuqi Yang, Yang Cao, Qibin Hou, Ming-Ming Cheng, Chunhua Shen*
- 226 Real Acoustic Fields: An Audio-Visual Room Acoustics Dataset and Benchmark, *Ziyang Chen, Israel D. Gebru, Christian Richardt, Anurag Kumar, William Laney, Andrew Owens, Alexander Richard*
- 227 Rethinking the Evaluation Protocol of Domain Generalization, *Han Yu, Xingxuan Zhang, Renzhe Xu, Jiashuo Liu, Yue He, Peng Cui*
- 228 MMSum: A Dataset for Multimodal Summarization and Thumbnail Generation of Videos, *Jielin Qiu, Jiacheng Zhu, William Han, Aditesh Kumar, Karthik Mittal, Claire Jin, Zhengyuan Yang, Linjie Li, Jianfeng Wang, Ding Zhao, Bo Li, Lijuan Wang*
- 229 Learning from Synthetic Human Group Activities, *Che-Jui Chang, Danrui Li, Deep Patel, Parth Goel, Honglu Zhou, Seonghyeon Moon, Samuel S. Sohn, Sejong Yoon, Vladimir Pavlovic, Mubbasir Kapadia*
- 230 Instance Tracking in 3D Scenes from Egocentric Videos, *Yunhan Zhao, Haoyu Ma, Shu Kong, Charles Fowlkes*
- 231 Insect-Foundation: A Foundation Model and Large-scale 1M Dataset for Visual Insect Understanding, *Hoang-Quan Nguyen, Thanh-Dat Truong, Xuan Bac Nguyen, Ashley Dowling, Xin Li, Khoa Luu*
- 232 Low-Resource Vision Challenges for Foundation Models, *Yunhua Zhang, Hazel Doughty, Cees G. M. Snoek*
- 233 OpenStreetView-5M: The Many Roads to Global Visual Geolocation, *Guillaume Astruc, Nicolas Dufour, Ioannis Siglidis, Constantin Aronsohn, Nacim Bouia, Stephanie Fu, Romain Loiseau, Van Nguyen Nguyen, Charles Raude, Elliot Vincent, Lintao Xu, Hongyu Zhou, Loic Landrieu*
- 234 FreeMan: Towards Benchmarking 3D Human Pose Estimation under Real-World Conditions, *Jiong Wang, Fengyu Yang, Bingliang Li, Wenbo Gou, Danqi Yan, Ailing Zeng, Yijun Gao, Junle Wang, Yanqing Jing, Ruimao Zhang*
- 235 LiDAR-Net: A Real-scanned 3D Point Cloud Dataset for Indoor Scenes, *Yanwen Guo, Yuanqi Li, Dayong Ren, Xiaohong Zhang, Jiawei Li, Liang Pu, Changfeng Ma, Xiaoyu Zhan, Jie Guo, Mingqiang Wei, Yan Zhang, Piaopiao Yu, Shuangyu Yang, Donghao Ji, Huisheng Ye, Hao Sun, Yansong Liu, YINUO Chen, Jiaqi Zhu, Hongyu Liu*
- 236 View-decoupled Transformer for Person Re-identification under Aerial-ground Camera Network, *Quan Zhang, Lei Wang, Vishal M. Patel, Xiaohua Xie, Jianhaung Lai*
- 237 UFineBench: Towards Text-based Person Retrieval with Ultra-fine Granularity, *Jialong Zuo, Hanyu Zhou, Ying Nie, Feng Zhang, Tianyu Guo, Nong Sang, Yunhe Wang, Changxin Gao*
- 238 Towards Automatic Power Battery Detection: New Challenge Benchmark Dataset and Baseline, *Xiaoqi Zhao, Youwei Pang, Zhenyu Chen, Qian Yu, Lihe Zhang, Hanqi Liu, Jiaming Zuo, Huchuan Lu*
- 239 Abductive Ego-View Accident Video Understanding for Safe Driving Perception, *Jianwu Fang, Lei-lei Li, Junfei Zhou, Junbin Xiao, Hongkai Yu, Chen Lv, Jianru Xue, Tat-Seng Chua*
- 240 Multiagent Multitraversal Multimodal Self-Driving: Open MARS Dataset, *Yiming Li, Zhiheng Li, Nuo Chen, Moonjun Gong, Zonglin Lyu, Zehong Wang, Peili Jiang, Chen Feng*
- 241 Towards Surveillance Video-and-Language Understanding: New Dataset Baselines and Challenges, *Tongtong Yuan, Xuange Zhang, Kun Liu, Bo Liu, Chen Chen, Jian Jin, Zhenzhen Jiao*
- 242 Pre-training Vision Models with Mandelbulb Variations, *Benjamin Naoto Chiche, Yuto Horikawa, Ryo Fujita*
- 243 EgoExoLearn: A Dataset for Bridging Asynchronous Ego- and Exo-centric View of Procedural Activities in Real World, *Yifei Huang, Guo Chen, Jilan Xu, Mingfang Zhang, Lijin Yang, Baoqi Pei, Hongjie Zhang, Lu Dong, Yali Wang, Limin Wang, Yu Qiao*
- 244 JRDB-Social: A Multifaceted Robotic Dataset for Understanding of Context and Dynamics of Human Interactions Within Social Groups, *Simindokht Jahangard, Zhixi Cai, Shiki Wen, Hamid Rezatofighi*
- 245 Spectral and Polarization Vision: Spectro-polarimetric Real-world Dataset, *Yujin Jeon, Eunsue Choi, Youngchan Kim, Yunseong Moon, Khalid Omer, Felix Heide, Seung-Hwan Baek*
- 246 MatSynth: A Modern PBR Materials Dataset, *Giuseppe Vecchio, Valentin Deschaintre*
- 247 When Visual Grounding Meets Gigapixel-level Large-scale Scenes: Benchmark and Approach, *Tao Ma, Bing Bai, Haozhe Lin, Heyuan Wang, Yu Wang, Lin Luo, Lu Fang*
- 248 HoloVIC: Large-scale Dataset and Benchmark for Multi-Sensor Holographic Intersection and Vehicle-Infrastructure Cooperative, *Cong Ma, Lei Qiao, Chengkai Zhu, Kai Liu, Zelong Kong, Qing Li, Xueqi Zhou, Yuheng Kan, Wei Wu*
- 249 EvalCrafter: Benchmarking and Evaluating Large Video Generation Models, *Yaofang Liu, Xiaodong Cun, Xuebo Liu, Xintao Wang, Yong Zhang, Haoxin Chen, Yang Liu, Tiejong Zeng, Raymond Chan, Ying Shan*
- 250 Localization Is All You Evaluate: Data Leakage in Online Mapping Datasets and How to Fix It, *Adam Lilja, Junsheng Fu, Erik Stenborg, Lars Hammarstrand*
- 251 DL3DV-10K: A Large-Scale Scene Dataset for Deep Learning-based 3D Vision, *Lu Ling, Yichen Sheng, Zhi Tu, Wentian Zhao, Cheng Xin, Kun Wan, Lantao Yu, Qianyu Guo, Zixun Yu, Yawen Lu, Xuanmao Li, Xingpeng Sun, Rohan Ashok, Aniruddha Mukherjee, Hao Kang, Xiangrui Kong, Gang Hua, Tianyi Zhang, Bedrich Benes, Aniket Bera*
- 252 OmniMedVQA: A New Large-Scale Comprehensive Evaluation Benchmark for Medical LLM, *Yutao Hu, Tianbin Li, Quanfeng Lu, Wenqi Shao, Junjun He, Yu Qiao, Ping Luo*
- 253 Can Biases in ImageNet Models Explain Generalization?, *Paul Gavrikov, Janis Keuper*
- 254 MVBench: A Comprehensive Multi-modal Video Understanding Benchmark, *Kunchang Li, Yali Wang, Yinan He, Yizhuo Li, Yi Wang, Yi Liu, Zun Wang, Jilan Xu, Guo Chen, Ping Luo, Limin Wang, Yu Qiao*
- 255 Towards Scalable 3D Anomaly Detection and Localization: A Benchmark via 3D Anomaly Synthesis and A Self-Supervised Learning Network, *Wenqiao Li, Xiaohao Xu, Yao Gu, Bozhong Zheng, Shenghua Gao, Yingna Wu*
- 256 Point-VOS: Pointing Up Video Object Segmentation, *Sabarinath Mahadevan, Idil Esen Zulfikar, Paul Voigtlaender, Bastian Leibe*
- 257 GPT-4V(ision) is a Human-Aligned Evaluator for Text-to-3D Generation, *Tong Wu, Guandaog Yang, Zhibing Li, Kai Zhang, Ziwei Liu, Leonidas Guibas, Dahua Lin, Gordon Wetzstein*
- 258 ConCon-Chi: Concept-Context Chimera Benchmark for Personalized Vision-Language Tasks, *Andrea Rosasco, Stefano Berti, Giulia Pasquale, Damiano Malafronte, Shogo Sato, Hiroyuki Segawa, Tetsugo Inada, Lorenzo Natale*
- 259 FISBe: A Real-World Benchmark Dataset for Instance Segmentation of Long-Range Thin Filamentous Structures, *Lisa Mais, Peter Hirsch, Claire Managan, Ramya Kandarpa, Josef Lorenz Rumberger, Annika Reinke, Lena Maier-Hein, Gudrun Ihrke, Dagmar Kainmueller*
- 260 Inter-X: Towards Versatile Human-Human Interaction Analysis, *Liang Xu, Xintao Lv, Yichao Yan, Xin Jin, Shuwen Wu, Congsheng Xu, Yifan Liu, Yizhou Zhou, Fengyun Rao, Xingdong Sheng, Yunhui Liu, Wenjun Zeng, Xiaokang Yang*
- 261 TextNeRF: A Novel Scene-Text Image Synthesis Method based on Neural Radiance Fields, *Jialei Cui, Jianwei Du, Wenzhuo Liu, Zhouhui Lian*
- 262 Systematic Comparison of Semi-supervised and Self-supervised Learning for Medical Image Classification, *Zhe Huang, Ruijie Jiang, Shuchin Aeron, Michael C. Hughes*
- 263 Unexplored Faces of Robustness and Out-of-Distribution: Covariate Shifts in Environment and Sensor Domains, *Eunsu Baek, Keondo Park, Jiyeon Kim, Hyung-Sin Kim*

- 264 MCD: Diverse Large-Scale Multi-Campus Dataset for Robot Perception, *Thien-Minh Nguyen, Shenghai Yuan, Thien Hoang Nguyen, Pengyu Yin, Haozhi Cao, Lihua Xie, Maciej Wozniak, Patric Jensfelt, Marko Thiel, Justin Ziegenbein, Noel Blunder*
- 265 360Loc: A Dataset and Benchmark for Omnidirectional Visual Localization with Cross-device Queries, *Huajian Huang, Changkun Liu, Yipeng Zhu, Hui Cheng, Tristan Braud, Sai-Kit Yeung*
- 266 Deep Generative Model based Rate-Distortion for Image Downscaling Assessment, *Yuanbang Liang, Bhavesh Garg, Paul Rosin, Yipeng Qin*
- 267 JRDB-PanoTrack: An Open-world Panoptic Segmentation and Tracking Robotic Dataset in Crowded Human Environments, *Duy Tho Le, Chenhui Gou, Stavva Datta, Hengcan Shi, Ian Reid, Jianfei Cai, Hamid Rezaatofghi*
- 268 MTMMC: A Large-Scale Real-World Multi-Modal Camera Tracking Benchmark, *Sanghyun Woo, Kwanyong Park, Inkyu Shin, Myungchul Kim, In So Kweon*
- 269 RCooper: A Real-world Large-scale Dataset for Roadside Cooperative Perception, *Ruiyang Hao, Siqi Fan, Yingru Dai, Zhenlin Zhang, Chenxi Li, Yuntian Wang, Haibao Yu, Wenxian Yang, Jirui Yuan, Zaiqing Nie*
- 270 UVEB: A Large-scale Benchmark and Baseline Towards Real-World Underwater Video Enhancement, *Yaofeng Xie, Lingwei Kong, Kai Chen, Ziqiang Zheng, Xiao Yu, Zhibin Yu, Bing Zheng*
- 271 Real-World Mobile Image Denoising Dataset with Efficient Baselines, *Roman Flepp, Andrey Ignatov, Radu Timofte, Luc Van Gool*
- 272 RGBD Objects in the Wild: Scaling Real-World 3D Object Learning from RGB-D Videos, *Hongchi Xia, Yang Fu, Sifei Liu, Xiaolong Wang*
- 273 Evaluating Transferability in Retrieval Tasks: An Approach Using MMD and Kernel Methods, *Mengyu Dai, Amir Hossein Raffiee, Aashish Jain, Joshua Correa*
- 274 BEHAVIOR Vision Suite: Customizable Dataset Generation via Simulation, *Yunhao Ge, Yihe Tang, Jiashu Xu, Cem Gokmen, Chengshu Li, Wensi Ai, Benjamin Jose Martinez, Arman Aydin, Mona Anvari, Ayush K Chakravarthy, Hong-Xing Yu, Josiah Wong, Sanjana Srivastava, Sharon Lee, Shengxin Zha, Laurent Itti, Yunzhu Li, Roberto Martín-Martín, Miao Liu, Pengchuan Zhang, Ruohan Zhang, Li Fei-Fei, Jiayun Wu*
- 275 MULAN: A Multi Layer Annotated Dataset for Controllable Text-to-Image Generation, *Petru-Daniel Tudosiu, Yongxin Yang, Shifeng Zhang, Fei Chen, Steven McDonagh, Gerasimos Lampouras, Ignacio Iacobacci, Sarah Parisot*
- 276 Sieve: Multimodal Dataset Pruning using Image Captioning Models, *Anas Mahmoud, Mostafa Elhoushi, Amro Abbas, Yu Yang, Newsha Ardalani, Hugh Leather, Ari S. Morcos*
- 277 Perceptual Assessment and Optimization of HDR Image Rendering, *Peibei Cao, Rafal K. Mantiuk, Kede Ma*
- 278 GlitchBench: Can Large Multimodal Models Detect Video Game Glitches?, *Mohammad Reza Taesiri, Tianjun Feng, Cor-Paul Bezemer, Anh Nguyen*
- 279 WinSyn: A High Resolution Testbed for Synthetic Data, *Tom Kelly, John Femiani, Peter Wonka*
- 280 DiVa-360: The Dynamic Visual Dataset for Immersive Neural Fields, *Cheng-You Lu, Peisen Zhou, Angela Xing, Chandradeep Pokharia, Arnab Dey, Ishaan Nikhil Shah, Rugved Mavidipalli, Dylan Hu, Andrew I. Comport, Kefan Chen, Srinath Sridhar*
- 281 Learning Discriminative Dynamics with Label Corruption for Noisy Label Detection, *Suyeon Kim, Dongha Lee, SeongKu Kang, Sukang Chae, Sangwan Jang, Hwanjo Yu*
- 282 DriveTrack: A Benchmark for Long-Range Point Tracking in Real-World Videos, *Arjun Balasingam, Joseph Chandler, Chenning Li, Zhoutong Zhang, Hari Balakrishnan*
- 283 HouseCat6D - A Large-Scale Multi-Modal Category Level 6D Object Perception Dataset with Household Objects in Realistic Scenarios, *HyunJun Jung, Shun-Cheng Wu, Patrick Ruhkamp, Guangyao Zhai, Hannah Schieber, Giulia Rizzoli, Pengyuan Wang, Hongcheng Zhao, Lorenzo Garattoni, Sven Meier, Daniel Roth, Nassir Navab, Benjamin Busam*
- 284 Benchmarking Segmentation Models with Mask-Preserved Attribute Editing, *Zijin Yin, Kongming Liang, Bing Li, Zhanyu Ma, Jun Guo*
- 285 The Devil is in the Fine-Grained Details: Evaluating Open-Vocabulary Object Detectors for Fine-Grained Understanding, *Lorenzo Bianchi, Fabio Carrara, Nicola Messina, Claudio Gennaro, Fabrizio Falchi*
- 286 PKU-DyMVHumans: A Multi-View Video Benchmark for High-Fidelity Dynamic Human Modeling, *Xiaoyun Zheng, Liwei Liao, Xufeng Li, Jianbo Jiao, Rongjie Wang, Feng Gao, Shiqi Wang, Ronggang Wang*
- 287 Insights from the Use of Previously Unseen Neural Architecture Search Datasets, *Rob Geada, David Towers, Matthew Forshaw, Amir Atapour-Abarghouei, A. Stephen McGough*
- 288 TULIP: Multi-camera 3D Precision Assessment of Parkinson's Disease, *Kyungdo Kim, Sihan Lyu, Sneha Mantri, Timothy W. Dunn*
- 289 LUWA Dataset: Learning Lithic Use-Wear Analysis on Microscopic Images, *Jing Zhang, Irving Fang, Hao Wu, Akshat Kaushik, Alice Rodriguez, Hanwen Zhao, Juexiao Zhang, Zhuo Zheng, Radu Iovita, Chen Feng*
- 290 ShapeWalk: Compositional Shape Editing Through Language-Guided Chains, *Habib Slim, Mohamed Elhoseiny*
- 291 360+x: A Panoptic Multi-modal Scene Understanding Dataset, *Hao Chen, Yuqi Hou, Chenyuan Qu, Irene Testini, Xiaohan Hong, Jianbo Jiao*
- 292 Ego-Exo4D: Understanding Skilled Human Activity from First- and Third-Person Perspectives, *Kristen Grauman, Andrew Westbury, Lorenzo Torresani, Kris Kitani, Jitendra Malik, Triantafyllos Afouras, Kumar Ashutosh, Vijay Baiyya, Siddhant Bansal, Bikram Boote, Eugene Byrne, Zach Chavis, Joya Chen, Feng Cheng, Fu-Jen Chu, Sean Crane, Avijit Dasgupta, Jing Dong, Maria Escobar, Cristhian Forigua, Abrahm Gebreselasie, Sanjay Haresh, Jing Huang, Md Mohaiminul Islam, Suyog Jain, Rawal Khirodkar, Devansh Kukreja, Kevin J Liang, Jia-Wei Liu, Sagnik Majumder, Yongsun Mao, Miguel Martin, Effrosyni Mavroudi, Tushar Nagarajan, Francesco Ragusa, Santhosh Kumar Ramakrishnan, Luigi Seminara, Arjun Somayazulu, Yale Song, Shan Su, Zihui Xue, Edward Zhang, Jinxu Zhang, Angela Castillo, Changan Chen, Xinzhu Fu, Ryosuke Furuta, Cristina Gonzalez, Prince Gupta, Jiabo Hu, Yifei Huang, Yiming Huang, Weslie Khoo, Anush Kumar, Robert Kuo, Sach Lakhavani, Miao Liu, Mi Luo, Zhengyi Luo, Brighid Meredith, Austin Miller, Oluwatuminu Oguntola, Xiaqing Pan, Penny Peng, Shraman Pramanick, Merey Ramazanov, Fiona Ryan, Wei Shan, Kiran Somasundaram, Chenan Song, Audrey Southerland, Masatoshi Tateno, Huiyu Wang, Yuchen Wang, Takuma Yagi, Mingfei Yan, Xitong Yang, Zecheng Yu, Shengxin Cindy Zha, Chen Zhao, Ziwei Zhao, Zhifan Zhu, Jeff Zhuo, Pablo Arbelaez, Gedas Bertasius, Dima Damen, Jakob Engel, Giovanni Maria Farinella, Antonino Furnari, Bernard Ghanem, Judy Hoffman, C.V. Jawahar, Richard Newcombe, Hyun Soo Park, James M. Rehg, Yoichi Sato, Manolis Savva, Jianbo Shi, Mike Zheng Shou, Michael Wray*
- 293 Rich Human Feedback for Text-to-Image Generation, *Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, Junjie Ke, Krishnamurthy Dj Dvijotham, Katherine M. Collins, Yiwen Luo, Yang Li, Kai J Kohlhoff, Deepak Ramachandran, Vidhya Navalpakkam*
- 294 TRINS: Towards Multimodal Language Models that Can Read, *Ruiyi Zhang, Yanzhe Zhang, Jian Chen, Yufan Zhou, Jiuxiang Gu, Changyou Chen, Tong Sun*
- 295 MAGIC: A Large-scale Captioned Dataset from Matting Generated Images using Chroma Keying, *Ryan D. Burgert, Brian L. Price, Jason Kuen, Yijun Li, Michael S. Ryoo*
- 296 EFHQ: Multi-purpose ExtremePose-Face-HQ dataset, *Trung Tuan Dao, Duc Hong Vu, Cuong Pham, Anh Tran*
- 297 How to Train Neural Field Representations: A Comprehensive Study and Benchmark, *Samuele Papa, Riccardo Valperga, David Knigge, Miltiadis Kofinas, Phillip Lippe, Jan-Jakob Sonke, Efstratios Gavves*
- 298 BioCLIP: A Vision Foundation Model for the Tree of Life, *Samuel Stevens, Jiaman Wu, Matthew J Thompson, Elizabeth G Campolongo, Chan Hee Song, David Edward Carlyn, Li Dong, Wasila M Dahdul, Charles Stewart, Tanya Berger-Wolf, Wei-Lun*

- Chao, Yu Su
- 299 A Noisy Elephant in the Room: Is Your Out-of-Distribution Detector Robust to Label Noise?, *Galadrielle Humblot-Renaux, Sergio Escalera, Thomas B. Moeslund*
- 300 eTraM: Event-based Traffic Monitoring Dataset, *Aayush Atul Verma, Bharatesh Chakravarthi, Arpitsinh Vaghela, Hua Wei, Yezhou Yang*
- 301 SubT-MRS Dataset: Pushing SLAM Towards All-weather Environments, *Shibo Zhao, Yuanjun Gao, Tianhao Wu, Damanpreet Singh, Rushan Jiang, Haoxiang Sun, Mansi Sarawata, Yuheng Qiu, Warren Whittaker, Ian Higgins, Yi Du, Shaoshu Su, Can Xu, John Keller, Jay Karhade, Lucas Nogueira, Sourjit Saha, Ji Zhang, Wenshan Wang, Chen Wang, Sebastian Scherer*
- 302 MSU-4S - The Michigan State University Four Seasons Dataset, *Daniel Kent, Mohammed Alyaqoub, Xiaohu Lu, Hamed Khatounabadi, Kookjin Sung, Cole Scheller, Alexander Dalat, Asma bin Thabit, Roberto Whitley, Hayder Radha*
- 303 TUMTraF V2X Cooperative Perception Dataset, *Walter Zimmer, Gerhard Arya Wardana, Suren Sritharan, Xingcheng Zhou, Rui Song, Alois C. Knoll*
- 304 Multiview Aerial Visual REcognition (MAVREC): Can Multi-view Improve Aerial Visual Perception?, *Aritra Dutta, Srijan Das, Jacob Nielsen, Rajatsubhra Chakraborty, Mubarak Shah*
- 305 Towards Co-Evaluation of Cameras HDR and Algorithms for Industrial-Grade 6DoF Pose Estimation, *Agastya Kalra, Guy Stoppi, Dmitrii Marin, Vage Taamazyan, Aarrushi Shandilya, Rishav Agarwal, Anton Boykov, Tze Hao Chong, Michael Stark*
- 306 Scaling Laws for Data Filtering— Data Curation cannot be Compute Agnostic, *Sachin Goyal, Pratyush Maini, Zachary C. Lipton, Aditi Raghunathan, J. Zico Kolter*
- 307 Benchmarking Audio Visual Segmentation for Long-Untrimmed Videos, *Chen Liu, Peike Patrick Li, Qingtao Yu, Hongwei Sheng, Dadong Wang, Lincheng Li, Xin Yu*
- 308 MLP Can Be A Good Transformer Learner, *Si Hao Lin, Pumeng Lyu, Dongrui Liu, Tao Tang, Xiaodan Liang, Andy Song, Xiaojun Chang*
- 309 From SAM to CAMs: Exploring Segment Anything Model for Weakly Supervised Semantic Segmentation, *Hyeokjun Kweon, Kuk-Jin Yoon*
- 310 Domain-Specific Block Selection and Paired-View Pseudo-Labeling for Online Test-Time Adaptation, *Yeonguk Yu, Sungho Shin, Seunghyeok Back, Mihwan Ko, Sangjun Noh, Kyoobin Lee*
- 311 VideoMAC: Video Masked Autoencoders Meet ConvNets, *Gensheng Pei, Tao Chen, Xiruo Jiang, Huafeng Liu, Zeren Sun, Yazhou Yao*
- 312 Unsupervised Universal Image Segmentation, *Dantong Niu, Xudong Wang, Xinyang Han, Long Lian, Roei Herzig, Trevor Darrell*
- 313 VideoCutLER: Surprisingly Simple Unsupervised Video Instance Segmentation, *Xudong Wang, Ishan Misra, Ziyun Zeng, Rohit Girdhar, Trevor Darrell*
- 314 What You See is What You GAN: Rendering Every Pixel for High-Fidelity Geometry in 3D GANs, *Alex Trevithick, Matthew Chan, Towaki Takikawa, Umar Iqbal, Shalini De Mello, Manmohan Chandraker, Ravi Ramamoorthi, Koki Nagano*
- 315 SPOT: Self-Training with Patch-Order Permutation for Object-Centric Learning with Autoregressive Transformers, *Ioannis Kakogeorgiou, Spyros Gidaris, Konstantinos Karantzas, Nikos Komodakis*
- 316 Unsupervised Learning of Category-Level 3D Pose from Object-Centric Videos, *Leonhard Sommer, Artur Jesslen, Eddy Ilg, Adam Kortylewski*
- 317 Distributionally Generative Augmentation for Fair Facial Attribute Classification, *Fengda Zhang, Qianpei He, Kun Kuang, Jiashuo Liu, Long Chen, Chao Wu, Jun Xiao, Hanwang Zhang*
- 318 Estimating Noisy Class Posterior with Part-level Labels for Noisy Label Learning, *Rui Zhao, Bin Shi, Jianfei Ruan, Tianze Pan, Bo Dong*
- 319 Unsupervised Keypoints from Pretrained Diffusion Models, *Eric Hedlin, Gopal Sharma, Shweta Mahajan, Xingzhe He, Hossam Isack, Abhishek Kar, Helge Rhodin, Andrea Tagliasacchi, Kwang Moo Yi*
- 320 Learning to Rank Patches for Unbiased Image Redundancy Reduction, *Yang Luo, Zhineng Chen, Peng Zhou, Zuxuan Wu, Xieping Gao, Yu-Gang Jiang*
- 321 Rethinking the Representation in Federated Unsupervised Learning with Non-IID Data, *Xinting Liao, Weiming Liu, Chaochao Chen, Pengyang Zhou, Fengyuan Yu, Huabin Zhu, Binhui Yao, Tao Wang, Xiaolin Zheng, Yanchao Tan*
- 322 GLID: Pre-training a Generalist Encoder-Decoder Vision Model, *Jihao Liu, Jinliang Zheng, Yu Liu, Hongsheng Li*
- 323 Sequential Modeling Enables Scalable Learning for Large Vision Models, *Yutong Bai, Xinyang Geng, Karttikeya Mangalam, Amir Bar, Alan L. Yuille, Trevor Darrell, Jitendra Malik, Alexei A. Efros*
- 324 VoCo: A Simple-yet-Effective Volume Contrastive Learning Framework for 3D Medical Image Analysis, *Linshan Wu, Jiaxin Zhuang, Hao Chen*
- 325 Real-IAD: A Real-World Multi-View Dataset for Benchmarking Versatile Industrial Anomaly Detection, *Chengjie Wang, Wenbing Zhu, Bin-Bin Gao, Zhenye Gan, Jiangning Zhang, Zhihao Gu, Shuguang Qian, Mingang Chen, Lizhuang Ma*
- 326 CroSel: Cross Selection of Confident Pseudo Labels for Partial-Label Learning, *Shiyu Tian, Hongxin Wei, Yiqun Wang, Lei Feng*
- 327 BEM: Balanced and Entropy-based Mix for Long-Tailed Semi-Supervised Learning, *Hongwei Zheng, Linyuan Zhou, Han Li, Jinming Su, Xiaoming Wei, Xiaoming Xu*
- 328 ReCoRe: Regularized Contrastive Representation Learning of World Model, *Rudra P.K. Poudel, Harit Pandya, Stephan Liwicki, Roberto Cipolla*
- 329 Universal Novelty Detection Through Adaptive Contrastive Learning, *Hossein Mirzaei, Mojtaba Nafez, Mohammad Jafari, Mohammad Bagher Soltani, Mohammad Azizmalayeri, Jafar Habibi, Mohammad Sabokrou, Mohammad Hossein Rohban*
- 330 Learning to Count without Annotations, *Lukas Knobel, Tengda Han, Yuki M. Asano*
- 331 Point Cloud Pre-training with Diffusion Models, *Xiao Zheng, Xiaoshui Huang, Guofeng Mei, Yuenan Hou, Zhaoyang Lyu, Bo Dai, Wanli Ouyang, Yongshun Gong*
- 332 Improving Unsupervised Hierarchical Representation with Reinforcement Learning, *Ruyi An, Yewen Li, Xu He, Pengjie Gu, Mengchen Zhao, Dong Li, Jianye Hao, Chaojie Wang, Bo An, Mingyuan Zhou*
- 333 Investigating and Mitigating the Side Effects of Noisy Views for Self-Supervised Clustering Algorithms in Practical Multi-View Scenarios, *Jie Xu, Yazhou Ren, Xiaolong Wang, Lei Feng, Zheng Zhang, Gang Niu, Xiaofeng Zhu*
- 334 Self-Supervised Representation Learning from Arbitrary Scenarios, *Zhaowen Li, Yousong Zhu, Zhiyang Chen, Zongxin Gao, Rui Zhao, Chaoyang Zhao, Ming Tang, Jinqiao Wang*
- 335 Learning SO(3)-Invariant Semantic Correspondence via Local Shape Transform, *Chunghyun Park, Seungwook Kim, Jaesik Park, Minsu Cho*
- 336 A Bayesian Approach to OOD Robustness in Image Classification, *Prakhar Kaushik, Adam Kortylewski, Alan Yuille*
- 337 Sculpting Holistic 3D Representation in Contrastive Language-Image-3D Pre-training, *Yipeng Gao, Zeyu Wang, Wei-Shi Zheng, Cihang Xie, Yuyin Zhou*
- 338 Solving Masked Jigsaw Puzzles with Diffusion Vision Transformers, *Jinyang Liu, Wondmgezahu Teshome, Sandesh Ghimire, Mario Szaier, Octavia Camps*
- 339 DS-NeRV: Implicit Neural Video Representation with Decomposed Static and Dynamic Codes, *Hao Yan, Zhihui Ke, Xiaobo Zhou, Tie Qiu, Xidong Shi, Dadong Jiang*
- 340 Brain Decodes Deep Nets, *Huzheng Yang, James Gee, Jianbo Shi*
- 341 Pose-Guided Self-Training with Two-Stage Clustering for Unsupervised Landmark Discovery, *Siddharth Tourani, Ahmed Alwehbi, Arif Mahmood, Muhammad Haris Khan*
- 342 Mitigating Object Dependencies: Improving Point Cloud Self-Supervised Learning through Object Exchange, *Yanhao Wu, Tong Zhang, Wei Ke, Congpei Qiu, Sabine Süssstrunk, Mathieu Salzmann*

- 343 Adaptive Slot Attention: Object Discovery with Dynamic Slot Number, *Ke Fan, Zechen Bai, Tianjun Xiao, Tong He, Max Horn, Yanwei Fu, Francesco Locatello, Zheng Zhang*
- 344 Targeted Representation Alignment for Open-World Semi-Supervised Learning, *Ruixuan Xiao, Lei Feng, Kai Tang, Junbo Zhao, Yixuan Li, Gang Chen, Haobo Wang*
- 345 Hierarchical Correlation Clustering and Tree Preserving Embedding, *Morteza Haghir Chehreghani, Mostafa Haghir Chehreghani*
- 346 Contrastive Mean-Shift Learning for Generalized Category Discovery, *Sua Choi, Dahyun Kang, Minsu Cho*
- 347 CuVLER: Enhanced Unsupervised Object Discoveries through Exhaustive Self-Supervised Transformers, *Shahaf Arica, Or Rubin, Sapir Gershov, Shlomi Laufer*
- 348 SODA: Bottleneck Diffusion Models for Representation Learning, *Drew A. Hudson, Daniel Zoran, Mateusz Malinowski, Andrew K. Lampinen, Andrew Jaegle, James L. McClelland, Loic Matthey, Felix Hill, Alexander Lerchner*
- 349 HPL-ESS: Hybrid Pseudo-Labeling for Unsupervised Event-based Semantic Segmentation, *Linglin Jing, Yiming Ding, Yunpeng Gao, Zhigang Wang, Xu Yan, Dong Wang, Gerald Schaefer, Hui Fang, Bin Zhao, Xuelong Li*
- 350 Positive-Unlabeled Learning by Latent Group-Aware Meta Disambiguation, *Lin Long, Haobo Wang, Zhijie Jiang, Lei Feng, Chang Yao, Gang Chen, Junbo Zhao*
- 351 Aligning Logits Generatively for Principled Black-Box Knowledge Distillation, *Jing Ma, Xiang Xiang, Ke Wang, Yuchuan Wu, Yongbin Li*
- 352 Improving Semantic Correspondence with Viewpoint-Guided
- ✱ Spherical Maps, *Octave Mariotti, Oisín Mac Aodha, Hakan Bilen*
- 353 Neural Modes: Self-supervised Learning of Nonlinear Modal Subspaces, *Jiahong Wang, Yinwei Du, Stelian Coros, Bernhard Thomaszewski*
- 354 Decentralized Directed Collaboration for Personalized Federated Learning, *Yingqi Liu, Yifan Shi, Qinglun Li, Baoyuan Wu, Xueqian Wang, Li Shen*
- 355 Improving Graph Contrastive Learning via Adaptive Positive Sampling, *Jiaming Zhuo, Feiyang Qin, Can Cui, Kun Fu, Bingxin Niu, Mengzhu Wang, Yuanfang Guo, Chuan Wang, Zhen Wang, Xiaochun Cao, Liang Yang*
- 356 Integrating Efficient Optimal Transport and Functional Maps For Unsupervised Shape Correspondence Learning, *Tung Le, Khai Nguyen, Shanlin Sun, Nhat Ho, Xiaohui Xie*
- 357 Unsupervised Feature Learning with Emergent Data-Driven Prototypicality, *Yunhui Guo, Youren Zhang, Yubei Chen, Stella X. Yu*
- 358 Label Propagation for Zero-shot Classification with Vision-Language Models, *Vladan Stojnić, Yannis Kalantidis, Giorgos Tolias*
- 359 Boosting Continual Learning of Vision-Language Models via Mixture-of-Experts Adapters, *Jiazuo Yu, Yunzhi Zhuge, Lu Zhang, Ping Hu, Dong Wang, Huchuan Lu, You He*
- 360 Backpropagation-free Network for 3D Test-time Adaptation, *Yanshuo Wang, Ali Cheraghian, Zeeshan Hayder, Jie Hong, Sameera Ramasinghe, Shafin Rahman, David Ahmedt-Aristizabal, Xuesong Li, Lars Petersson, Mehrtash Harandi*
- 361 GDA: Generalized Diffusion for Robust Test-time Adaptation, *Yun-Yun Tsai, Fu-Chen Chen, Albert Y. C. Chen, Junfeng Yang, Che-Chun Su, Min Sun, Cheng-Hao Kuo*
- 362 Semantically-Shifted Incremental Adapter-Tuning is a Continual ViTransformer, *Yuwen Tan, Qin hao Zhou, Xiang Xiang, Ke Wang, Yuchuan Wu, Yongbin Li*
- 363 Few-shot Learner Parameterization by Diffusion Time-steps, *Zhongqi Yue, Pan Zhou, Richang Hong, Hanwang Zhang, Qianru Sun*
- 364 FREE: Faster and Better Data-Free Meta-Learning, *Yongxian Wei, Zixuan Hu, Zhenyi Wang, Li Shen, Chun Yuan, Dacheng Tao*
- 365 Classes Are Not Equal: An Empirical Study on Image Recognition Fairness, *Jiequan Cui, Beier Zhu, Xin Wen, Xiaojuan Qi, Bei Yu, Hanwang Zhang*
- 366 DAVE - A Detect-and-Verify Paradigm for Low-Shot Counting, *Jer Pelhan, Alan Lukežič, Vitjan Zavrtnik, Matej Kristan*
- 367 Density-guided Translator Boosts Synthetic-to-Real Unsupervised Domain Adaptive Segmentation of 3D Point Clouds, *Zhimin Yuan, Wankang Zeng, Yanfei Su, Weiquan Liu, Ming Cheng, Yulan Guo, Cheng Wang*
- 368 D3T: Distinctive Dual-Domain Teacher Zigzagging Across RGB-Thermal Gap for Domain-Adaptive Object Detection, *Dinh Phat Do, Taehoon Kim, Jaemin Na, Jiwon Kim, Keonho Lee, Kyunghwan Cho, Wonjun Hwang*
- 369 AMU-Tuning: Effective Logit Bias for CLIP-based Few-shot Learning, *Yuwei Tang, Zhenyi Lin, Qilong Wang, Pengfei Zhu, Qinghua Hu*
- 370 LEAD: Learning Decomposition for Source-free Universal Domain Adaptation, *Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, Changjun Jiang*
- 371 Improving Generalized Zero-Shot Learning by Exploring the Diverse Semantics from External Class Names, *Yapeng Li, Yong Luo, Zengmao Wang, Bo Du*
- 372 What How and When Should Object Detectors Update in Continually Changing Test Domains?, *Jayeon Yoo, Dongkwan Lee, Inseop Chung, Donghyun Kim, Nojun Kwak*
- 373 Split to Merge: Unifying Separated Modalities for Unsupervised Domain Adaptation, *Xinyao Li, Yuke Li, Zhekai Du, Fengling Li, Ke Lu, Jingjing Li*
- 374 Domain-Agnostic Mutual Prompting for Unsupervised Domain Adaptation, *Zhekai Du, Xinyao Li, Fengling Li, Ke Lu, Lei Zhu, Jingjing Li*
- 375 Improving the Generalization of Segmentation Foundation Model under Distribution Shift via Weakly Supervised Adaptation, *Haojie Zhang, Yongyi Su, Xun Xu, Kui Jia*
- 376 DeiT-LT: Distillation Strikes Back for Vision Transformer Training on Long-Tailed Datasets, *Harsh Rangwani, Pradipto Mondal, Mayank Mishra, Ashish Ramayee Asokan, R. Venkatesh Babu*
- 377 Unified Language-driven Zero-shot Domain Adaptation, *Senqiao Yang, Zhuotao Tian, Li Jiang, Jiaya Jia*
- 378 Stable Neighbor Denoising for Source-free Domain Adaptive Segmentation, *Dong Zhao, Shuang Wang, Qi Zang, Licheng Jiao, Nicu Sebe, Zhun Zhong*
- 379 A Simple Recipe for Language-guided Domain Generalized Segmentation, *Mohammad Fahes, Tuan-Hung Vu, Andrei Bursuc, Patrick Pérez, Raoul de Charette*
- 380 TCP: Textual-based Class-aware Prompt tuning for Visual-Language Model, *Hantao Yao, Rui Zhang, Changsheng Xu*
- 381 Adapters Strike Back, *Jan-Martin O. Steitz, Stefan Roth*
- 382 Improving Plasticity in Online Continual Learning via Collaborative Learning, *Maorong Wang, Nicolas Michel, Ling Xiao, Toshihiko Yamasaki*
- 383 Visual Prompting for Generalized Few-shot Segmentation: A Multi-scale Approach, *Mir Rayat Imtiaz Hossain, Mennatullah Siam, Leonid Sigal, James J. Little*
- 384 Adaptive Random Feature Regularization on Fine-tuning Deep Neural Networks, *Shin'ya Yamaguchi, Sekitoshi Kanai, Kazuki Adachi, Daiki Chijiwa*
- 385 ESCAPE: Encoding Super-keypoints for Category-Agnostic Pose Estimation, *Khoi Duc Nguyen, Chen Li, Gim Hee Lee*
- 386 PracticalDG: Perturbation Distillation on Vision-Language Models for Hybrid Domain Generalization, *Zining Chen, Weiqiu Wang, Zhicheng Zhao, Fei Su, Aidong Men, Hongying Meng*
- 387 Rethinking Multi-domain Generalization with a General Learning Objective, *Zhaorui Tan, Xi Yang, Kaizhu Huang*
- 388 L2B: Learning to Bootstrap Robust Models for Combating Label Noise, *Yuyin Zhou, Xianhang Li, Fengze Liu, Qingyue Wei, Xuxi Chen, Lequan Yu, Cihang Xie, Matthew P. Lungren, Lei Xing*
- 389 Meta-Point Learning and Refining for Category-Agnostic Pose Estimation, *Junjie Chen, Jiebin Yan, Yuming Fang, Li Niu*
- 390 A2XP: Towards Private Domain Generalization, *Geunhyeok Yu, Hyoseok Hwang*
- 391 Expandable Subspace Ensemble for Pre-Trained Model-Based Class-Incremental Learning, *Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, De-Chuan Zhan*
- 392 VRP-SAM: SAM with Visual Reference Prompt, *Yanpeng Sun, Jiahui Chen, Shan Zhang, Xinyu Zhang, Qiang Chen, Gang Zhang, Errui Ding, Jingdong Wang, Zechao Li*
- 393 Flatten Long-Range Loss Landscapes for Cross-Domain Few-Shot Learning, *Yixiong Zou, Yicong Liu, Yiman Hu, Yuhua Li, Ruixuan Li*

- 394 MAP: MAsk-Pruning for Source-Free Model Intellectual Property Protection, *Boyang Peng, Sanqing Qu, Yong Wu, Tianpei Zou, Lianghua He, Alois Knoll, Guang Chen, Changjun Jiang*
- 395 Disentangled Prompt Representation for Domain Generalization, *De Cheng, Zhipeng Xu, Xinyang Jiang, Nannan Wang, Dongsheng Li, Xinbo Gao*
- 396 Adapt Before Comparison: A New Perspective on Cross-Domain Few-Shot Segmentation, *Jonas Herzog*
- 397 Convolutional Prompting meets Language Models for Continual Learning, *Anurag Roy, Riddhiman Moulick, Vinay K. Verma, Saptarshi Ghosh, Abir Das*
- 398 Visual-Augmented Dynamic Semantic Prototype for Generative Zero-Shot Learning, *Wenjin Hou, Shiming Chen, Shuhuang Chen, Ziming Hong, Yan Wang, Xuetao Feng, Salman Khan, Fahad Shahbaz Khan, Xinge You*
- 399 InfLoRA: Interference-Free Low-Rank Adaptation for Continual Learning, *Yan-Shuo Liang, Wu-Jun Li*
- 400 Discriminative Pattern Calibration Mechanism for Source-Free Domain Adaptation, *Haifeng Xia, Siyu Xia, Zhengming Ding*
- 401 NICE: Neurogenesis Inspired Contextual Encoding for Replay-free Class Incremental Learning, *Mustafa Burak Gurbuz, Jean Michael Moorman, Constantine Dovrolis*
- 402 Orchestrate Latent Expertise: Advancing Online Continual Learning with Multi-Level Supervision and Reverse Self-Distillation, *Hongwei Yan, Liyuan Wang, Kaisheng Ma, Yi Zhong*
- 403 A Closer Look at the Few-Shot Adaptation of Large Vision-Language Models, *Julio Silva-Rodríguez, Sina Hajimiri, Ismail Ben Ayed, Jose Dolz*
- 404 Towards Generalizing to Unseen Domains with Few Labels, *Chamuditha Jayanga Galappaththige, Sanooran Baliah, Malitha Gunawardhana, Muhammad Haris Khan*
- 405 Improved Self-Training for Test-Time Adaptation, *Jing Ma*
- 406 Source-Free Domain Adaptation with Frozen Multimodal Foundation Model, *Song Tang, Wenxin Su, Mao Ye, Xi Tian Zhu*
- 407 Deep Imbalanced Regression via Hierarchical Classification Adjustment, *Haipeng Xiong, Angela Yao*
- 408 A Versatile Framework for Continual Test-Time Domain Adaptation: Balancing Discriminability and Generalizability, *Xu Yang, Xuan Chen, Moqi Li, Kun Wei, Cheng Deng*
- 409 DYSON: Dynamic Feature Space Self-Organization for Online Task-Free Class Incremental Learning, *Yuhang He, Yingjie Chen, Yuhan Jin, Songlin Dong, Xing Wei, Yihong Gong*
- 410 Test-Time Linear Out-of-Distribution Detection, *Ke Fan, Tong Liu, Xingyu Qiu, Yikai Wang, Lian Huai, Zeyu Shangguan, Shuang Gou, Fengjian Liu, Yuqian Fu, Yanwei Fu, Xingqun Jiang*
- 411 LTGC: Long-tail Recognition via Leveraging LLMs-driven
* Generated Content, *Qihao Zhao, Yalun Dai, Hao Li, Wei Hu, Fan Zhang, Jun Liu*
- 412 APSeg: Auto-Prompt Network for Cross-Domain Few-Shot Semantic Segmentation, *Weizhao He, Yang Zhang, Wei Zhuo, Linlin Shen, Jiaqi Yang, Songhe Deng, Liang Sun*
- 413 LP++: A Surprisingly Strong Linear Probe for Few-Shot CLIP, *Yunshi Huang, Fereshteh Shakeri, Jose Dolz, Malik Boudiaf, Houa Bahig, Ismail Ben Ayed*
- 414 On the Test-Time Zero-Shot Generalization of Vision-Language Models: Do We Really Need Prompt Learning?, *Maxime Zanella, Ismail Ben Ayed*
- 415 Discriminative Sample-Guided and Parameter-Efficient Feature Space Adaptation for Cross-Domain Few-Shot Learning, *Rashindrie Perera, Saman Halgamuge*
- 416 Regularized Parameter Uncertainty for Improving Generalization in Reinforcement Learning, *Pehuen Moure, Longbiao Cheng, Joachim Ott, Zuowen Wang, Shih-Chii Liu*
- 417 An Empirical Study of the Generalization Ability of Lidar 3D Object Detectors to Unseen Domains, *George Eskandar*
- 418 MMA: Multi-Modal Adapter for Vision-Language Models, *Lingxiao Yang, Ru-Yuan Zhang, Yanchen Wang, Xiaohua Xie*
- 419 PerAda: Parameter-Efficient Federated Learning Personalization with Generalization Guarantees, *Chulin Xie, De-An Huang, Wenda Chu, Daguang Xu, Chaowei Xiao, Bo Li, Anima Anandkumar*
- 420 Bayesian Exploration of Pre-trained Models for Low-shot Image Classification, *Yibo Miao, Yu Lei, Feng Zhou, Zhijie Deng*
- 421 NAYER: Noisy Layer Data Generation for Efficient and Effective Data-free Knowledge Distillation, *Minh-Tuan Tran, Trung Le, Xuan-May Le, Mehrtash Harandi, Quan Hung Tran, Dinh Phung*
- 422 Text-Enhanced Data-free Approach for Federated Class-Incremental Learning, *Minh-Tuan Tran, Trung Le, Xuan-May Le, Mehrtash Harandi, Dinh Phung*
- 423 Pre-trained Vision and Language Transformers Are Few-Shot Incremental Learners, *Keon-Hee Park, Kyungwoo Song, Gyeong-Moon Park*
- 424 CDMAD: Class-Distribution-Mismatch-Aware Debiasing for Class-Imbalanced Semi-Supervised Learning, *Hyuck Lee, Heeyoung Kim*
- 425 TEA: Test-time Energy Adaptation, *Yige Yuan, Bingbing Xu, Liang Hou, Fei Sun, Huawei Shen, Xueqi Cheng*
- 426 Universal Semi-Supervised Domain Adaptation by Mitigating Common-Class Bias, *Wenyu Zhang, Qingmu Liu, Felix Ong Wei Cong, Mohamed Ragab, Chuan-Sheng Foo*
- 427 Leveraging Vision-Language Models for Improving Domain Generalization in Image Classification, *Sravanti Addepalli, Ashish Ramayee Asokan, Lakshay Sharma, R. Venkatesh Babu*
- 428 Learning Equi-angular Representations for Online Continual Learning, *Minhyuk Seo, Hyunseo Koh, Wonje Jeung, Minjae Lee, San Kim, Hankook Lee, Sungjun Cho, Sungik Choi, Hyunwoo Kim, Jonghyun Choi*
- 429 Open-Set Domain Adaptation for Semantic Segmentation, *Seun-An Choe, Ah-Hyung Shin, Keon-Hee Park, Jinwoo Choi, Gyeong-Moon Park*
- 430 Task-Adaptive Saliency Guidance for Exemplar-free Class Incremental Learning, *Xialei Liu, Jiang-Tian Zhai, Andrew D. Bagdanov, Ke Li, Ming-Ming Cheng*
- 431 Progressive Semantic-Guided Vision Transformer for Zero-Shot Learning, *Shiming Chen, Wenjin Hou, Salman Khan, Fahad Shahbaz Khan*
- 432 Unified Entropy Optimization for Open-Set Test-Time Adaptation, *Zhengqing Gao, Xu-Yao Zhang, Cheng-Lin Liu*
- 433 FedSelect: Personalized Federated Learning with Customized Selection of Parameters for Fine-Tuning, *Rishub Tamirisa, Chulin Xie, Wenxuan Bao, Andy Zhou, Ron Arel, Aviv Shamsian*
- 434 Dual-Enhanced Coreset Selection with Class-wise Collaboration for Online Blurry Class Incremental Learning, *Yutian Luo, Shiqi Zhao, Haoran Wu, Zhiwu Lu*
- 435 Troika: Multi-Path Cross-Modal Traction for Compositional Zero-Shot Learning, *Siteng Huang, Biao Gong, Yutong Feng, Min Zhang, Yiliang Lv, Donglin Wang*
- 436 Unveiling the Unknown: Unleashing the Power of Unknown to Known in Open-Set Source-Free Domain Adaptation, *Fuli Wan, Han Zhao, Xu Yang, Cheng Deng*
- 437 Dual-Consistency Model Inversion for Non-Exemplar Class Incremental Learning, *Zihuan Qiu, Yi Xu, Fanman Meng, Hongliang Li, Linfeng Xu, Qingbo Wu*
- 438 Domain-Rectifying Adapter for Cross-Domain Few-Shot Segmentation, *Jiapeng Su, Qi Fan, Wenjie Pei, Guangming Lu, Fanlin Chen*
- 439 Overcoming Generic Knowledge Loss with Selective Parameter Update, *Wenxuan Zhang, Paul Janson, Rahaf Aljundi, Mohamed Elhoseiny*
- 440 BrainWash: A Poisoning Attack to Forget in Continual Learning, *Ali Abbasi, Parsa Nooralinejad, Hamed Pirsiavash, Soheil Kolouri*
- 441 Enhancing Visual Continual Learning with Language-Guided Supervision, *Bolin Ni, Hongbo Zhao, Chenghao Zhang, Ke Hu, Gaofeng Meng, Zhaoxiang Zhang, Shiming Xiang*
- 10:30 - 18:45 Art Program (Arch 4CDE)**
- 10:30 - 18:45 DEMOS (Arch 4CDE)**
- 1 The Visual Remix: Swap Objects with Ease, *Bhushan Garware*
- 2 Better Call SAL: Towards Learning to Segment Anything in Lidar, *Aljosa Osep, Tim Meinhardt, Francesco Ferroni, Neehar Peri, Deva Ramanan, Laura Leal-Taixé*

- 3 ScribblePrompt: Fast and Flexible Interactive Segmentation for Any Biomedical Image, *Hallee Wong, Marianne Rakic, John Guttag, Adrian Dalca*
- 4 DART: Implicit Doppler Tomography for Radar Novel View Synthesis, *Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, Anthony Rowe*
- 5 Visual Place Recognition using 3D City Models, *Gabriele Berton, Lorenz Junglas, Tom Pollock, Carlo Masone, Barbara Caputo*
- 6 A Computer Vision Testbed for New York City Street Intersections, *Mehmet Kerem Turkcan, Mahshid Ghasemi Dehkordi, Sofia Kleisarchaki, Thomas Calmant, Levent Gürgen, Javad Ghaderi, Gil Zussman, Zoran Kostic*
- 7 L-MAGIC: Language Model Assisted Generation of Images with Coherence, *Zhipeng Cai, Tien Pei Chou*
- 8 Building UBC in Minecraft, *Ashtan Mistal*
- 9 SuperPrimitive: Scene Reconstruction at a Primitive Level, *Kirill Mazur, Gwangbin Bae, Andrew J. Davison*
- 10 H-Unique: 3D Hand Reconstruction and Automated Mapping of Anatomical Detail for Forensic Identification, *Bryan M. Williams, Hossein Rahmani, Sue Black, Xinyu Yang, Zheheng Jiang, Andrei Banica*
- 11 Universal 3D Reconstruction: Interactive Demonstration of the Scalable 3D Lifting Foundation Model (3D-LFM), *Mosam Dabhi, László A. Jeni, Simon Lucey*
- 12 Neuro-Symbolic Olympics Diving Judge, *Lauren Okamoto, Paritosh Parmar*
- 13 Grounding Everything: Emerging Localization Properties in Vision-Language Transformers, *Walid Boussefham*
- 14 CoGS: Controllable Gaussian Splatting, *Heng Yu, Joel Julin, Zoltan Á Milacski, Koichiro Niinuma, László A. Jeni*
- 15 Cutting-edge Text-Image Comprehension and Composition in Vision-Language Large Model, *Jiaqi Wang, Xiaoyi Dong, Pan Zhang, Yuhang Zang*
- 16 Collaborative Score Distillation for Consistent Visual Editing of My Own Visual Assets, *Subin Kim, Sooyeon Park*
- 17 Semantic Class-Adaptive Diffusion Model (SCA-DM), *Alex Ergasti, Claudio Ferrari, Tomaso Fontanini, Massimo Bertozzi, Andrea Prati*
- 18 A Real-Time Speech-Driven Vocal Tract Avatar, *Tejas Prabhune, Peter Wu, Cheol Jun Cho, Bohan Yu, Gopala Anumanchipalli*


12:00 - 14:00 LUNCH (Summit ExHall 1-2)

13:00 - 14:30 Orals 6A: Low-level Vision and Remote Sensing (Summit Ballroom)


- 1 LDP: Language-driven Dual-Pixel Image Defocus Deblurring Network, *Hao Yang, Liyuan Pan, Yan Yang, Richard Hartley, Miaomiao Liu*
- 2 S2MAE: A Spatial-Spectral Pretraining Foundation Model for Spectral Remote Sensing Data, *Xuyang Li, Danfeng Hong, Jocelyn Chanussot*
- 3 Task-Driven Wavelets using Constrained Empirical Risk Minimization,  *Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke, Jonas Teuwen*
- 4 Image Processing GNN: Breaking Rigidity in Super-Resolution,  *Yuchuan Tian, Hanting Chen, Chao Xu, Yunhe Wang*
- 5 DART: Implicit Doppler Tomography for Radar Novel View Synthesis, *Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana Laroia, Zico Kolter, Anthony Rowe*

13:00 - 14:30 Orals 6B: Image & Video Synthesis (Summit Flex Hall AB)

- 1 Alchemist: Parametric Control of Material Properties with Diffusion Models, *Prafull Sharma, Varun Jampani, Yuanzhen Li, Xuhui Jia, Dmitry Lagun, Fredo Durand, Bill Freeman, Mark Matthews*
- 2 Generative Image Dynamics,  *Zhengqi Li, Richard Tucker, Noah Snavely, Aleksander Holynski*
- 3 Visual Anagrams: Generating Multi-View Optical Illusions with Diffusion Models, *Daniel Geng, Inbum Park, Andrew Owens*
- 4 MonoHair: High-Fidelity Hair Modeling from a Monocular Video, *Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, Youyi Zheng*

- 5 Analyzing and Improving the Training Dynamics of Diffusion Models,  *Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, Samuli Laine*

13:00 - 14:30 Orals 6C: Multi-Modal Learning (Summit Flex Hall C)

- 1 InternVL: Scaling up Vision Foundation Models and Aligning for Generic Visual-Linguistic Tasks, *Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, Bin Li, Ping Luo, Tong Lu, Yu Qiao, Jifeng Dai*
- 2 Describing Differences in Image Sets with Natural Language, *Lisa Dunlap, Yuhui Zhang, Xiaohan Wang, Ruiqi Zhong, Trevor Darrell, Jacob Steinhardt, Joseph E. Gonzalez, Serena Yeung-Levy*
- 3 NoiseCLR: A Contrastive Learning Approach for Unsupervised Discovery of Interpretable Directions in Diffusion Models, *Yusuf Dalva, Pinar Yanardag*
- 4 MetaCloak: Preventing Unauthorized Subject-driven Text-to-image Diffusion-based Synthesis via Meta-learning, *Yixin Liu, Chenrui Fan, Yutong Dai, Xun Chen, Pan Zhou, Lichao Sun*
- 5 EGTR: Extracting Graph from Transformer for Scene Graph Generation,  *Jinbae Im, Jeongyeon Nam, Nokyung Park, Hyungmin Lee, Seunghyun Park*

14:30 - 14:45 Courtesy Break

14:45 - 15:45 KEYNOTE 3 - Sofia Crespo, Artist (Summit Flex Hall ABC)

15:45 - 16:00 Courtesy Break

16:00 - 17:00 PANEL 2 (Summit Flex Hall ABC)

16:45 - 17:15 Poster Setup (Arch 4E)

17:15 - 18:45 Poster Session 6 & Exhibit Hall (Arch 4A-E)

- 1 MonoHair: High-Fidelity Hair Modeling from a Monocular Video,  *Keyu Wu, Lingchen Yang, Zhiyi Kuang, Yao Feng, Xutao Han, Yuefan Shen, Hongbo Fu, Kun Zhou, Youyi Zheng*
- 2 BadCLIP: Trigger-Aware Prompt Learning for Backdoor Attacks on CLIP, *Jiawang Bai, Kuofeng Gao, Shaobo Min, Shu-Tao Xia, Zhifeng Li, Wei Liu*
- 3 Semantic-Aware Multi-Label Adversarial Attacks, *Hassan Mahmood, Ehsan Elhamifar*
- 4 Defense without Forgetting: Continual Adversarial Defense with Anisotropic & Isotropic Pseudo Replay, *Yuhang Zhou, Zhongyuan Hua*
- 5 Learning to Transform Dynamically for Better Adversarial Transferability, *Rongyi Zhu, Zeliang Zhang, Susan Liang, Zhuo Liu, Chenliang Xu*
- 6 Infrared Adversarial Car Stickers, *Xiaopei Zhu, Yuqiu Liu, Zhanhao Hu, Jianmin Li, Xiaolin Hu*
- 7 Unsegment Anything by Simulating Deformation, *Jiahao Lu, Xingyi Yang, Xinchao Wang*
- 8 Efficient Model Stealing Defense with Noise Transition Matrix, *Dong-Dong Wu, Chilin Fu, Weichang Wu, Wenwen Xia, Xiaolu Zhang, Jun Zhou, Min-Ling Zhang*
- 9 Fully Exploiting Every Real Sample: SuperPixel Sample Gradient Model Stealing, *Yunlong Zhao, Xiaoheng Deng, Yijing Liu, Xinjun Pei, Jiazhi Xia, Wei Chen*
- 10 Hide in Thicket: Generating Imperceptible and Rational Adversarial Perturbations on 3D Point Clouds, *Tianrui Lou, Xiaojun Jia, Jindong Gu, Li Liu, Siyuan Liang, Bangyan He, Xiaochun Cao*
- 11 Boosting Adversarial Transferability by Block Shuffle and Rotation, *Kunyu Wang, Xuanran He, Wenxuan Wang, Xiaosen Wang*
- 12 Robust Overfitting Does Matter: Test-Time Adversarial Purification With FGSM, *Linyu Tang, Lei Zhang*
- 13 Data Poisoning based Backdoor Attacks to Contrastive Learning, *Jinghuai Zhang, Hongbin Liu, Jinyuan Jia, Neil Zhenqiang Gong*
- 14 NAPGuard: Towards Detecting Naturalistic Adversarial Patches, *Siyang Wu, Jiakai Wang, Jiejie Zhao, Yazhe Wang, Xianglong Liu*
- 15 Ensemble Diversity Facilitates Adversarial Transferability, *Bowen Tang, Zheng Wang, Yi Bin, Qi Dou, Yang Yang, Heng Tao Shen*

- 16 Revamping Federated Learning Security from a Defender's Perspective: A Unified Defense with Homomorphic Encrypted Data Space, *K Naveen Kumar, Reshmi Mitra, C Krishna Mohan*
- 17 Can Protective Perturbation Safeguard Personal Data from Being Exploited by Stable Diffusion?, *Zhengyue Zhao, Jinhao Duan, Kaidi Xu, Chenan Wang, Rui Zhang, Zidong Du, Qi Guo, Xing Hu*
- 18 One Prompt Word is Enough to Boost Adversarial Robustness for Pre-trained Vision-Language Models, *Lin Li, Haoyan Guan, Jianing Qiu, Michael Spratling*
- 19 Watermark-embedded Adversarial Examples for Copyright Protection against Diffusion Models, *Peifei Zhu, Tsubasa Takahashi, Hirokatsu Kataoka*
- 20 Not All Prompts Are Secure: A Switchable Backdoor Attack Against Pre-trained Vision Transformers, *Sheng Yang, Jiawang Bai, Kuofeng Gao, Yong Yang, Yiming Li, Shu-Tao Xia*
- 21 Focus on Hiders: Exploring Hidden Threats for Enhancing Adversarial Training, *Qian Li, Yuxiao Hu, Yinpeng Dong, Dongxiao Zhang, Yuntian Chen*
- 22 Physical 3D Adversarial Attacks against Monocular Depth Estimation in Autonomous Driving, *Junhao Zheng, Chenhao Lin, Jiahao Sun, Zhengyu Zhao, Qian Li, Chao Shen*
- 23 Distraction is All You Need: Memory-Efficient Image Immunization against Diffusion-Based Image Editing Highlight: Top 10% of the accepted papers No Ling Lo, Cheng Yu Yeo, Hong-Han Shuai, Wen-Huang Cheng
- 24 PAD: Patch-Agnostic Defense against Adversarial Patch Attacks, *Lihua Jing, Rui Wang, Wenqi Ren, Xin Dong, Cong Zou*
- 25 PeerAiD: Improving Adversarial Distillation from a Specialized Peer Tutor, *Jaewon Jung, Hongsun Jang, Jaeyong Song, Jinho Lee*
- 26 Revisiting Adversarial Training Under Long-Tailed Distributions, *Xinli Yue, Ningping Mou, Qian Wang, Lingchen Zhao*
- 27 Pre-trained Model Guided Fine-Tuning for Zero-Shot Adversarial Robustness, *Sibo Wang, Jie Zhang, Zheng Yuan, Shiguang Shan*
- 28 Towards Transferable Targeted 3D Adversarial Attack in the Physical World, *Yao Huang, Yinpeng Dong, Shouwei Ruan, Xiao Yang, Hang Su, Xingxing Wei*
- 29 Nearest is Not Dearest: Towards Practical Defense against Quantization-conditioned Backdoor Attacks, *Boheng Li, Yishuo Cai, Haowei Li, Feng Xue, Zhifeng Li, Yiming Li*
- 30 Perturbing Attention Gives You More Bang for the Buck: Subtle Imaging Perturbations That Efficiently Fool Customized Diffusion Models, *Jingyao Xu, Yuetong Lu, Yandong Li, Siyang Lu, Dongdong Wang, Xiang Wei*
- 31 Boosting Adversarial Training via Fisher-Rao Norm-based Regularization, *Xiangyu Yin, Wenjie Ruan*
- 32 Random Entangled Tokens for Adversarially Robust Vision Transformer, *Huihui Gong, Minjing Dong, Siqi Ma, Seyit Camtepe, Surya Nepal, Chang Xu*
- 33 Backdoor Defense via Test-Time Detecting and Repairing, *Jiyang Guan, Jian Liang, Ran He*
- 34 1-Lipschitz Layers Compared: Memory Speed and Certifiable Robustness, *Bernd Prach, Fabio Brau, Giorgio Buttazzo, Christoph H. Lampert*
- 35 DiffAM: Diffusion-based Adversarial Makeup Transfer for Facial Privacy Protection, *Yuhao Sun, Lingyun Yu, Hongtao Xie, Jiaming Li, Yongdong Zhang*
- 36 DAP: A Dynamic Adversarial Patch for Evading Person Detectors, *Amira Guesmi, Ruitian Ding, Muhammad Abdullah Hanif, Ihsen Alouani, Muhammad Shafique*
- 37 Adversarial Distillation Based on Slack Matching and Attribution Region Alignment, *Shenglin Yin, Zhen Xiao, Mingxuan Song, Jieyi Long*
- 38 Improving Transferable Targeted Adversarial Attacks with Model Self-Enhancement, *Han Wu, Guanyan Ou, Weibin Wu, Zibin Zheng*
- 39 On the Robustness of Large Multimodal Models Against Image Adversarial Attacks, *Xuanming Cui, Alejandro Aparcedo, Young Kyun Jang, Ser-Nam Lim*
- 40 Intriguing Properties of Diffusion Models: An Empirical Study of the Natural Attack Capability in Text-to-Image Generative Models, *Takami Sato, Justin Yue, Nanze Chen, Ningfei Wang, Qi Alfred Chen*
- 41 BadCLIP: Dual-Embedding Guided Backdoor Attack on
- ✳ iMultimodal Contrastive Learning, *Yuan Liang, Mingli Zhu, Aishan Liu, Baoyuan Wu, Xiaochun Cao, Ee-Chien Chang*
- 42 MMCert: Provable Defense against Adversarial Attacks to Multimodal Models, *Yanting Wang, Hongye Fu, Wei Zou, Jinyuan Jia*
- 43 MimicDiffusion: Purifying Adversarial Perturbation via Mimicking Clean Diffusion Model, *Kaiyu Song, Hanjiang Lai, Yan Pan, Jian Yin*
- 44 Revisiting Adversarial Training at Scale, *Zeyu Wang, Xianhang Li, Hongru Zhu, Cihang Xie*
- 45 Language-Driven Anchors for Zero-Shot Adversarial Robustness, *Xiao Li, Wei Zhang, Yining Liu, Zhanhao Hu, Bo Zhang, Xiaolin Hu*
- 46 Transferable Structural Sparse Adversarial Attack Via Exact Group Sparsity Training, *Di Ming, Peng Ren, Yunlong Wang, Xin Feng*
- 47 Fooling Polarization-Based Vision using Locally Controllable Polarizing Projection, *Zhuoxiao Li, Zhihang Zhong, Shohei Nobuhara, Ko Nishino, Yinqiang Zheng*
- 48 Overload: Latency Attacks on Object Detection for Edge Devices, *Erh-Chung Chen, Pin-Yu Chen, I-Hsin Chung, Che-Rung Lee*
- 49 Attack To Defend: Exploiting Adversarial Attacks for Detecting Poisoned Models, *Samar Fares, Karthik Nandakumar*
- 50 Towards Understanding and Improving Adversarial Robustness of Vision Transformers, *Samyukta Jain, Tanima Dutta*
- 51 Towards Fairness-Aware Adversarial Learning, *Yanghao Zhang, Tianle Zhang, Ronghui Mu, Xiaowei Huang, Wenjie Ruan*
- 52 Byzantine-robust Decentralized Federated Learning via Dual-domain Clustering and Trust Bootstrapping, *Peng Sun, Xinyang Liu, Zhibo Wang, Bo Liu*
- 53 Towards General Robustness Verification of MaxPool-based Convolutional Neural Networks via Tightening Linear Approximation, *Yuan Xiao, Shiqing Ma, Juan Zhai, Chunrong Fang, Jinyuan Jia, Zhenyu Chen*
- 54 Soften to Defend: Towards Adversarial Robustness via Self-Guided Label Refinement, *Zhuorong Li, Daiwei Yu, Lina Wei, Canghong Jin, Yun Zhang, Sixian Chan*
- 55 SlowFormer: Adversarial Attack on Compute and Energy Consumption of Efficient Vision Transformers, *K L Navaneet, Soroush Abbasi Koohpayegani, Essam Sleiman, Hamed Pirsiavash*
- 56 LOTUS: Evasive and Resilient Backdoor Attacks through Sub-Partitioning, *Siyan Cheng, Guan hong Tao, Yingqi Liu, Guangyu Shen, Shengwei An, Shiwei Feng, Xiangzhe Xu, Kaiyuan Zhang, Shiqing Ma, Xiangyu Zhang*
- 57 Deep-TROJ: An Inference Stage Trojan Insertion Algorithm through Efficient Weight Replacement Attack, *Sabbir Ahmed, Ranyang Zhou, Shaahin Angizi, Adnan Siraj Rakin*
- 58 Semantic Shield: Defending Vision-Language Models Against Backdooring and Poisoning via Fine-grained Knowledge Alignment, *Alvi Md Ishmam, Christopher Thomas*
- 59 Initialization Matters for Adversarial Transfer Learning, *Andong Hua, Jindong Gu, Zhiyu Xue, Nicholas Carlini, Eric Wong, Yao Qin*
- 60 Strong Transferable Adversarial Attacks via Ensembled
- ✳ Asymptotically Normal Distribution Learning, *Zhengwei Fang, Rui Wang, Tao Huang, Liping Jing*
- 61 HDRFlow: Real-Time HDR Video Reconstruction with Large Motions, *Gangwei Xu, Yujin Wang, Jinwei Gu, Tianfan Xue, Xin Yang*
- 62 A Physics-informed Low-rank Deep Neural Network for Blind and Universal Lens Aberration Correction, *Jin Gong, Runzhaoyang, Weihang Zhang, Jinli Suo, Qionghai Dai*
- 63 Super-Resolution Reconstruction from Bayer-Pattern Spike Streams, *Yanchen Dong, Ruiqin Xiong, Jian Zhang, Zhaofei Yu, Xiaopeng Fan, Shuyuan Zhu, Tiejun Huang*
- 64 In2SET: Intra-Inter Similarity Exploiting Transformer for Dual-Camera Compressive Hyperspectral Imaging, *Xin Wang, Lizhi Wang, Xiangtian Ma, Maoqing Zhang, Lin Zhu, Hua Huang*
- 65 SuperSVG: Superpixel-based Scalable Vector Graphics Synthesis, *Teng Hu, Ran Yi, Baihong Qian, Jiangning Zhang, Paul L. Rosin, Yu-Kun Lai*
- 66 Language-driven All-in-one Adverse Weather Removal, *Hao Yang, Liyuan Pan, Yan Yang, Wei Liang*

- 67 LDP: Language-driven Dual-Pixel Image Defocus Deblurring Network,
* Hao Yang, Liyuan Pan, Yan Yang, Richard Hartley, Miaomiao Liu
- 68 Language-guided Image Reflection Separation, Haofeng Zhong,
Yuchen Hong, Shuchen Weng, Jinxiu Liang, Boxin Shi
- 69 Time-Efficient Light-Field Acquisition Using Coded Aperture and
Events, Shuji Habuchi, Keita Takahashi, Chihiro Tsutake, Toshiaki
Fujii, Hajime Nagahara
- 70 NB-GTR: Narrow-Band Guided Turbulence Removal, Yifei Xia,
Chu Zhou, Chengxuan Zhu, Mingguai Teng, Chao Xu, Boxin Shi
- 71 Complementing Event Streams and RGB Frames for Hand Mesh
Reconstruction, Jianping Jiang, Xinyu Zhou, Bingxuan Wang,
Xiaoming Deng, Chao Xu, Boxin Shi
- 72 Boosting Spike Camera Image Reconstruction from a Perspective
of Dealing with Spike Fluctuations, Rui Zhao, Ruiqin Xiong, Jing
Zhao, Jian Zhang, Xiaopeng Fan, Zhaofei Yu, Tiejun Huang
- 73 Frequency-aware Event-based Video Deblurring for Real-World
Motion Blur, Taewoo Kim, Hoonhee Cho, Kuk-Jin Yoon
- 74 Latency Correction for Event-guided Deblurring and Frame
Interpolation, Yixin Yang, Jinxiu Liang, Bohan Yu, Yan Chen, Jimmy
S. Ren, Boxin Shi
- 75 Learning to Remove Wrinkled Transparent Film with Polarized Prior,
Jiaqi Tang, Ruizheng Wu, Xiaogang Xu, Sixing Hu, Ying-Cong Chen
- 76 Dispersed Structured Light for Hyperspectral 3D Imaging,
Suhyun Shin, Seokjun Choi, Felix Heide, Seung-Hwan Baek
- 77 Generalized Event Cameras, Varun Sundar, Matthew Dutton, Andrei
Ardelean, Claudio Bruschini, Edoardo Charbon, Mohit Gupta
- 78 Intensity-Robust Autofocus for Spike Camera, Changqing Su,
Zhiyuan Ye, Yongsheng Xiao, You Zhou, Zhen Cheng, Bo Xiong,
Zhaofei Yu, Tiejun Huang
- 79 Selective Nonlinearities Removal from Digital Signals,
Krzysztof A. Maliszewski, Magdalena A. Urbańska, Varvara Vetrova,
Sylvia M. Kolenderska
- 80 Close Imitation of Expert Retouching for Black-and-White
Photography, Seunghyun Shin, Jisu Shin, Jihwan Bae, Inwook
Shim, Hae-Gon Jeon
- 81 Spike-guided Motion Deblurring with Unknown Modal
Spatiotemporal Alignment, Jiyuan Zhang, Shiyan Chen, Yajing
Zheng, Zhaofei Yu, Tiejun Huang
- 82 Coherence As Texture – Passive Textureless 3D Reconstruction
* by Self-interference, Wei-Yu Chen, Aswin C. Sankaranarayanan,
Anat Levin, Matthew O’Toole
- 83 TurboSL: Dense Accurate and Fast 3D by Neural Inverse
Structured Light, Parsa Mirdehghan, Maxx Wu, Wenzheng Chen,
David B. Lindell, Kiriakos N. Kutulakos
- 84 SPIDeRS: Structured Polarization for Invisible Depth and Reflectance
Sensing, Tomoki Ichikawa, Shohei Nobuhara, Ko Nishino
- 85 CPP-Net: Embracing Multi-Scale Feature Fusion into Deep
Unfolding CP-PPA Network for Compressive Sensing, Zhen Guo,
Hongping Gan
- 86 SwitchLight: Co-design of Physics-driven Architecture and Pre-training
* Framework for Human Portrait Relighting, Hoon Kim, Minje Jang,
Wonjun Yoon, Jisoo Lee, Donghyun Na, Sanghyun Woo
- 87 Diffeomorphic Template Registration for Atmospheric Turbulence
* Mitigation, Dong Lao, Congli Wang, Alex Wong, Stefano Soatto
- 88 Towards HDR and HFR Video from Rolling-Mixed-Bit
Spikings, Yakun Chang, Yeliduosi Xiaokaiti, Yujia Liu, Bin Fan,
Zhaojun Huang, Tiejun Huang, Boxin Shi
- 89 Progressive Divide-and-Conquer via Subsampling
* Decomposition for Accelerated MRI, Chong Wang, Lanqing Guo,
Yufei Wang, Hao Cheng, Yi Yu, Bihan Wen
- 90 Generative Quanta Color Imaging, Vishal Purohit, Junjie Luo,
Yiheng Chi, Qi Guo, Stanley H. Chan, Qiang Qiu
- 91 UFC-Net: Unrolling Fixed-point Continuous Network for Deep
Compressive Sensing, Xiaoyang Wang, Hongping Gan
- 92 Batch Normalization Alleviates the Spectral Bias in Coordinate
Networks, Zhicheng Cai, Hao Zhu, Qiu Shen, Xinran Wang, Xun Cao
- 93 EVS-assisted Joint Deblurring Rolling-Shutter Correction and
Video Frame Interpolation through Sensor Inverse Modeling, Rui
Jiang, Fangwen Tu, Yixuan Long, Aabhaas Vaish, Bowen Zhou,
Qinyi Wang, Wei Zhang, Yuntan Fang, Luis Eduardo Garcia Capel,
Bo Mu, Tiejun Dai, Andreas Suess
- 94 Unsupervised Deep Unrolling Networks for Phase
Unwrapping, Zhile Chen, Yuhui Quan, Hui Ji
- 95 LAN: Learning to Adapt Noise for Image Denoising, Changjin
Kim, Tae Hyun Kim, Sungyong Baik
- 96 Snapshot Lidar: Fourier Embedding of Amplitude and Phase
for Single-Image Depth Reconstruction, Sarah Friday, Yunzi Shi,
Yaswanth Cherivirala, Vishwanath Saragadam, Adithya Pediredla
- 97 FC-GNN: Recovering Reliable and Accurate Correspondences
from Interferences, Haobo Xu, Jun Zhou, Hua Yang, Renjie Pan,
Cunyan Li
- 98 Projecting Trackable Thermal Patterns for Dynamic Computer
Vision, Mark Sheinin, Aswin C. Sankaranarayanan, Srinivasa G.
Narasimhan
- 99 PixelRNN: In-pixel Recurrent Neural Networks for End-to-end-
optimized Perception with Neural Sensors, Haley M. So, Laurie
Bose, Piotr Dudek, Gordon Wetzstein
- 100 Image Restoration by Denoising Diffusion Models with Iteratively
Preconditioned Guidance, Tomer Garber, Tom Tirer
- 101 DART: Implicit Doppler Tomography for Radar Novel View Synthesis,
* Tianshu Huang, John Miller, Akarsh Prabhakara, Tao Jin, Tarana
Laroia, Zico Kolter, Anthony Rowe
- 102 Equivariant Plug-and-Play Image Reconstruction, Matthieu Terris,
Thomas Moreau, Nelly Pustelnik, Julian Tachella
- 103 CodedEvents: Optimal Point-Spread-Function Engineering for
3D-Tracking with Event Cameras, Sachin Shah, Matthew A. Chan,
Haoming Cai, Jingxi Chen, Sakshum Kulshrestha, Chahat Deep
Singh, Yiannis Aloimonos, Christopher A. Metzler
- 104 WaveMo: Learning Wavefront Modulations to See Through
Scattering, Mingyang Xie, Haiyun Guo, Brandon Y. Feng, Lingbo
Jin, Ashok Veeraraghavan, Christopher A. Metzler
- 105 Turb-Seg-Res: A Segment-then-Restore Pipeline for Dynamic
Videos with Atmospheric Turbulence, Ripon Kumar Saha, Dehao
Qin, Nianyi Li, Jinwei Ye, Suren Jayasuriya
- 106 DiffSCI: Zero-Shot Snapshot Compressive Imaging via Iterative
Spectral Diffusion Model, Zhenghao Pan, Haijin Zeng, Jiezhong
Cao, Kai Zhang, Yongyong Chen
- 107 Resolution Limit of Single-Photon LiDAR, Stanley H. Chan,
Hashan K. Weerasooriya, Weijian Zhang, Pamela Abshire, Istvan
Gyongy, Robert K. Henderson
- 108 QN-Mixer: A Quasi-Newton MLP-Mixer Model for Sparse-View CT
Reconstruction, Ishak Ayad, Nicolas Larue, Mai K. Nguyen
- 109 Dual-Scale Transformer for Large-Scale Single-Pixel
Imaging, Gang Qu, Ping Wang, Xin Yuan
- 110 Rolling Shutter Correction with Intermediate Distortion Flow
Estimation, Mingdeng Cao, Sidi Yang, Yujiu Yang, Yinqiang Zheng
- 111 Passive Snapshot Coded Aperture Dual-Pixel RGB-D
Imaging, Bhargav Ghanekar, Salman Siddique Khan, Pranav
Sharma, Shreyas Singh, Vivek Boominathan, Kaushik Mitra, Ashok
Veeraraghavan
- 112 Single View Refractive Index Tomography with Neural
Fields, Brandon Zhao, Aviad Levis, Liam Connor, Pratul P.
Srinivasan, Katherine L. Bouman
- 113 SPECAT: SPatial-spEctral Cumulative-Attention Transformer for
High-Resolution Hyperspectral Image Reconstruction, Zhiyang
Yao, Shuyang Liu, Xiaoyun Yuan, Lu Fang
- 114 Task-Driven Wavelets using Constrained Empirical Risk
* Minimization, Eric Marcus, Ray Sheombarsing, Jan-Jakob Sonke,
Jonas Teuwen
- 115 Describing Differences in Image Sets with Natural Language,
* Lisa Dunlap, Yuhui Zhang, Xiaohan Wang, Ruiqi Zhong, Trevor
Darrell, Jacob Steinhardt, Joseph E. Gonzalez, Serena Yeung-Levy
- 116 Alchemist: Parametric Control of Material Properties with
* Diffusion Models, Prafull Sharma, Varun Jampani, Yuanzhen Li, Xuhui
Jia, Dmitry Lagun, Fredo Durand, Bill Freeman, Mark Matthews
- 117 Generative Image Dynamics, Zhengqi Li, Richard Tucker,
* Noah Snavely, Aleksander Holynski
- 118 Visual Anagrams: Generating Multi-View Optical Illusions with
* Diffusion Models, Daniel Geng, Inbum Park, Andrew Owens
- 119 NoiseCLR: A Contrastive Learning Approach for Unsupervised

- ❖ Discovery of Interpretable Directions in Diffusion Models, *Yusuf Dalva, Pinar Yanardag*
- 120 Analyzing and Improving the Training Dynamics of Diffusion Models, *Tero Karras, Miika Aittala, Jaakko Lehtinen, Janne Hellsten, Timo Aila, Samuli Laine*
- 🏆 121 Fourier Priors-Guided Diffusion for Zero-Shot Joint Low-Light Enhancement and Deblurring, *Xiaoqian Lv, Shengping Zhang, Chenyang Wang, Yichen Zheng, Bineng Zhong, Chongyi Li, Liqiang Nie*
- 122 Color Shift Estimation-and-Correction for Image Enhancement, *Yiyu Li, Ke Xu, Gerhard Petrus Hancke, Rynson W.H. Lau*
- 123 Video Super-Resolution Transformer with Masked Inter&Intra-Frame Attention, *Xingyu Zhou, Leheng Zhang, Xiaorui Zhao, Keze Wang, Leida Li, Shuhang Gu*
- 124 Distilling Semantic Priors from SAM to Efficient Image Restoration Models, *Quan Zhang, Xiaoyu Liu, Wei Li, Hanting Chen, Junchao Liu, Jie Hu, Zhiwei Xiong, Chun Yuan, Yunhe Wang*
- 125 Beyond Average: Individualized Visual Scanpath Prediction, *Xianyu Chen, Ming Jiang, Qi Zhao*
- 126 Multimodal Prompt Perceiver: Empower Adaptiveness Generalizability and Fidelity for All-in-One Image Restoration, *Yang Ai, Huaibo Huang, Xiaoqiang Zhou, Jiexiang Wang, Ran He*
- 127 Selective Hourglass Mapping for Universal Image Restoration Based on Diffusion Model, *Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, Wei-Shi Zheng*
- 128 SeeSR: Towards Semantics-Aware Real-World Image Super-Resolution, *Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, Lei Zhang*
- 129 Revisiting Single Image Reflection Removal In the Wild, *Yurui Zhu, Xueyang Fu, Peng-Tao Jiang, Hao Zhang, Qibin Sun, Jinwei Chen, Zheng-Jun Zha, Bo Li*
- 130 ODCR: Orthogonal Decoupling Contrastive Regularization for Unpaired Image Dehazing, *Zhongze Wang, Haitao Zhao, Jingchao Peng, Lujian Yao, Kaijie Zhao*
- 131 Q-Instruct: Improving Low-level Visual Abilities for Multi-modality Foundation Models, *Haoning Wu, Zicheng Zhang, Erli Zhang, Chaofeng Chen, Liang Liao, Annan Wang, Kaixin Xu, Chunyi Li, Jingwen Hou, Guangtao Zhai, Geng Xue, Wenxiu Sun, Qiong Yan, Weisi Lin*
- 132 Enhancing Quality of Compressed Images by Mitigating Enhancement Bias Towards Compression Domain, *Qunliang Xing, Mai Xu, Shengxi Li, Xin Deng, Meisong Zheng, Huaida Liu, Ying Chen*
- 133 Attentive Illumination Decomposition Model for Multi-Illuminant White Balancing, *Dongyoung Kim, Jinwoo Kim, Junsang Yu, Seon Joo Kim*
- 134 NightCC: Nighttime Color Constancy via Adaptive Channel Masking, *Shuwei Li, Robby T. Tan*
- 135 Navigating Beyond Dropout: An Intriguing Solution towards Generalizable Image Super Resolution, *Hongjun Wang, Jiyuan Chen, Yinqiang Zheng, Tiejong Zeng*
- 136 Learning Inclusion Matching for Animation Paint Bucket Colorization, *Yuekun Dai, Shangchen Zhou, Qinyue Li, Chongyi Li, Chen Change Loy*
- 137 Defense Against Adversarial Attacks on No-Reference Image Quality Models with Gradient Norm Regularization, *Yujia Liu, Chenxi Yang, Dingquan Li, Jianhao Ding, Tingting Jiang*
- 138 Towards Backward-Compatible Continual Learning of Image Compression, *Zhihao Duan, Ming Lu, Justin Yang, Jiangpeng He, Zhan Ma, Fengqing Zhu*
- 139 APISR: Anime Production Inspired Real-World Anime Super-Resolution, *Boyang Wang, Fengyu Yang, Xihang Yu, Chao Zhang, Hanbin Zhao*
- 140 Unifying Automatic and Interactive Matting with Pretrained ViTs, *Zixuan Ye, Wenze Liu, He Guo, Yujia Liang, Chaoyi Hong, Hao Lu, Zhiguo Cao*
- 141 Motion-adaptive Separable Collaborative Filters for Blind Motion Deblurring, *Chengxu Liu, Xuan Wang, Xiangyu Xu, Ruhao Tian, Shuai Li, Xueming Qian, Ming-Hsuan Yang*
- 142 Genuine Knowledge from Practice: Diffusion Test-Time Adaptation for Video Adverse Weather Removal, *Yijun Yang, Hongtao Wu, Angelica I. Aviles-Rivero, Yulun Zhang, Jing Qin, Lei Zhu*
- 143 HomoFormer: Homogenized Transformer for Image Shadow Removal, *Jie Xiao, Xueyang Fu, Yurui Zhu, Dong Li, Jie Huang, Kai Zhu, Zheng-Jun Zha*
- 144 Bidirectional Multi-Scale Implicit Neural Representations for Image Deraining, *Xiang Chen, Jinshan Pan, Jiangxin Dong*
- 145 LED: A Large-scale Real-world Paired Dataset for Event Camera Denoising, *Yuxing Duan*
- 146 Seeing Motion at Nighttime with an Event Camera, *Haoyue Liu, Shihan Peng, Lin Zhu, Yi Chang, Hanyu Zhou, Luxin Yan*
- 147 Leveraging Frame Affinity for sRGB-to-RAW Video De-rendering, *Chen Zhang, Wencheng Han, Yang Zhou, Jianbing Shen, Cheng-zhong Xu, Wentao Liu*
- 148 Scaling Up to Excellence: Practicing Model Scaling for Photo-Realistic Image Restoration In the Wild, *Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, Chao Dong*
- 149 AdaRevD: Adaptive Patch Exiting Reversible Decoder Pushes the Limit of Image Deblurring, *Xintian Mao, Qingli Li, Yan Wang*
- 150 Unsupervised Blind Image Deblurring Based on Self-Enhancement, *Lufeifei Chen, Xiangpeng Tian, Shuhua Xiong, Yinjie Lei, Chao Ren*
- 151 TTA-EVF: Test-Time Adaptation for Event-based Video Frame Interpolation via Reliable Pixel and Sample Estimation, *Hoonhee Cho, Taewoo Kim, Yuhwan Jeong, Kuk-Jin Yoon*
- 152 Learning Coupled Dictionaries from Unpaired Data for Image Super-Resolution, *Longguang Wang, Juncheng Li, Yingqian Wang, Qingyong Hu, Yulan Guo*
- 153 Empowering Resampling Operation for Ultra-High-Definition Image Enhancement with Model-Aware Guidance, *Wei Yu, Jie Huang, Bing Li, Kaiwen Zheng, Qi Zhu, Man Zhou, Feng Zhao*
- 154 Generating Content for HDR Deghosting from Frequency View, *Tao Hu, Qingsen Yan, Yuankai Qi, Yanning Zhang*
- 155 Dual Prior Unfolding for Snapshot Compressive Imaging, *Jiancheng Zhang, Haijin Zeng, Jiezhong Cao, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao*
- 156 Binarized Low-light Raw Video Enhancement, *Gengchen Zhang, Yulun Zhang, Xin Yuan, Ying Fu*
- 157 Neural Spline Fields for Burst Image Fusion and Layer Separation, *Ilya Chugunov, David Shustin, Ruyi Yan, Chenyang Lei, Felix Heide*
- 158 Learning Degradation-Independent Representations for Camera ISP Pipelines, *Yanhui Guo, Fangzhou Luo, Xiaolin Wu*
- 159 SeD: Semantic-Aware Discriminator for Image Super-Resolution, *Bingchen Li, Xin Li, Hanxin Zhu, Yeying Jin, Ruoyu Feng, Zhizheng Zhang, Zhibo Chen*
- 160 SinSR: Diffusion-Based Image Super-Resolution in a Single Step, *Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C. Kot, Bihan Wen*
- 161 Self-Adaptive Reality-Guided Diffusion for Artifact-Free Super-Resolution, *Qingping Zheng, Ling Zheng, Yuanfan Guo, Ying Li, Songcen Xu, Jiankang Deng, Hang Xu*
- 162 Improving Spectral Snapshot Reconstruction with Spectral-Spatial Rectification, *Jiancheng Zhang, Haijin Zeng, Yongyong Chen, Dengxiu Yu, Yin-Ping Zhao*
- 163 Diffusion-based Blind Text Image Super-Resolution, *Yuzhe Zhang, Jiawei Zhang, Hao Li, Zhouxia Wang, Luwei Hou, Dongqing Zou, Liheng Bian*
- 164 CAMixerSR: Only Details Need More "Attention", *Yan Wang, Yi Liu, Shijie Zhao, Junlin Li, Li Zhang*
- 165 ID-Blau: Image Deblurring by Implicit Diffusion-based reBLurring AUgmentation, *Jia-Hao Wu, Fu-Jen Tsai, Yan-Tsung Peng, Chung-Chi Tsai, Chia-Wen Lin, Yen-Yu Lin*
- 166 Low-Res Leads the Way: Improving Generalization for Super-Resolution by Self-Supervised Learning, *Haoyu Chen, Wenbo Li, Jinjin Gu, Jingjing Ren, Haoze Sun, Xueyi Zou, Zhensong Zhang, Youliang Yan, Lei Zhu*
- 167 CoSeR: Bridging Image and Language for Cognitive Super-Resolution, *Haoze Sun, Wenbo Li, Jianzhuang Liu, Haoyu Chen, Renjing Pei, Xueyi Zou, Youliang Yan, Yujiu Yang*

- 168 Real-World Efficient Blind Motion Deblurring via Blur Pixel Discretization, *Insoo Kim, Jae Seok Choi, Geonseok Seo, Kinam Kwon, Jinwoo Shin, Hyong-Euk Lee*
- 169 SeNM-VAE: Semi-Supervised Noise Modeling with Hierarchical Variational Autoencoder, *Dihan Zheng, Yihang Zou, Xiaowen Zhang, Chenglong Bao*
- 170 Text-guided Explorable Image Super-resolution, *Kanchana Vaishnavi Gandikota, Paramanand Chandramouli*
- 171 Equivariant Multi-Modality Image Fusion, *Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Kai Zhang, Shuang Xu, Dongdong Chen, Radu Timofte, Luc Van Gool*
- 172 Revisiting Spatial-Frequency Information Integration from a Hierarchical Perspective for Panchromatic and Multi-Spectral Image Fusion, *Jiangtong Tan, Jie Huang, Naishan Zheng, Man Zhou, Keyu Yan, Danfeng Hong, Feng Zhao*
- 173 MCNet: Rethinking the Core Ingredients for Accurate and Efficient Homography Estimation, *Haokai Zhu, Si-Yuan Cao, Jianxin Hu, Sitong Zuo, Beinan Yu, Jiacheng Ying, Junwei Li, Hui-Liang Shen*
- 174 Contrastive Pre-Training with Multi-View Fusion for No-Reference Point Cloud Quality Assessment, *Ziyu Shan, Yujie Zhang, Qi Yang, Haichen Yang, Yiling Xu, Jenq-Neng Hwang, Xiaozhong Xu, Shan Liu*
- 175 MuGE: Multiple Granularity Edge Detection, *Caixia Zhou, Yaping Huang, Mengyang Pu, Qingji Guan, Ruoxi Deng, Haibin Ling*
- 176 KVQ: Kwai Video Quality Assessment for Short-form Videos, *Yiting Lu, Xin Li, Yajing Pei, Kun Yuan, Qizhi Xie, Yunpeng Qu, Ming Sun, Chao Zhou, Zhibo Chen*
- 177 Transfer CLIP for Generalizable Image Denoising, *Jun Cheng, Dong Liang, Shan Tan*
- 178 Improved Implicit Neural Representation with Fourier Reparameterized Training, *Kexuan Shi, Xingyu Zhou, Shuhang Gu*
- 179 Deep Video Inverse Tone Mapping Based on Temporal Clues, *Yuyao Ye, Ning Zhang, Yang Zhao, Hongbin Cao, Ronggang Wang*
- 180 Boosting Flow-based Generative Super-Resolution Models via Learned Prior, *Li-Yuan Tsao, Yi-Chen Lo, Chia-Che Chang, Hao-Wei Chen, Roy Tseng, Chien Feng, Chun-Yi Lee*
- 181 Look-Up Table Compression for Efficient Image Restoration,
* *Yinglong Li, Jiacheng Li, Zhiwei Xiong*
- 182 Latent Modulated Function for Computational Optimal Continuous Image Representation, *Zongyao He, Zhi Jin*
- 183 Task-Aware Encoder Control for Deep Video Compression, *Xingtong Ge, Jixiang Luo, Xinjie Zhang, Tongda Xu, Guo Lu, Dailan He, Jing Geng, Yan Wang, Jun Zhang, Hongwei Qin*
- 184 A Dynamic Kernel Prior Model for Unsupervised Blind Image Super-Resolution, *Zhixiong Yang, Jingyuan Xia, Shengxi Li, Xinghua Huang, Shuanghui Zhang, Zhen Liu, Yaowen Fu, Yongxiang Liu*
- 185 Zero-Reference Low-Light Enhancement via Physical Quadruple Priors, *Wenjing Wang, Huan Yang, Jianlong Fu, Jiaying Liu*
- 186 ParamISP: Learned Forward and Inverse ISPs using Camera Parameters, *Woohyeok Kim, Geonu Kim, Junyong Lee, Seungyong Lee, Seung-Hwan Baek, Sunghyun Cho*
- 187 FSC: Few-point Shape Completion, *Xianzu Wu, Xianfeng Wu, Tianyu Luan, Yajing Bai, Zhongyuan Lai, Junsong Yuan*
- 188 Generative Latent Coding for Ultra-Low Bitrate Image Compression, *Zhaoyang Jia, Jiahao Li, Bin Li, Houqiang Li, Yan Lu*
- 189 Neural Video Compression with Feature Modulation, *Jiahao Li, Bin Li, Yan Lu*
- 190 Driving-Video Dehazing with Non-Aligned Regularization for Safety Assistance, *Junkai Fan, Jiangwei Weng, Kun Wang, Yijun Yang, Jianjun Qian, Jun Li, Jian Yang*
- 191 Image Processing GNN: Breaking Rigidity in Super-Resolution,
* *Yuchuan Tian, Hanting Chen, Chao Xu, Yunhe Wang*
- 192 CFAT: Unleashing Triangular Windows for Image Super-resolution,
* *Abhisek Ray, Gaurav Kumar, Maheshkumar H. Kolekar*
- 193 Zero-Shot Structure-Preserving Diffusion Model for High Dynamic Range Tone Mapping, *Ruoxi Zhu, Shusong Xu, Peiye Liu, Sicheng Li, Yanheng Lu, Dimin Niu, Zihao Liu, Zihao Meng, Zhiyong Li, Xinhua Chen, Yibo Fan*
- 194 Calibrating Multi-modal Representations: A Pursuit of Group Robustness without Annotations, *Chenyu You, Yifei Min, Weicheng Dai, Jasjeet S. Sekhon, Lawrence Staib, James S. Duncan*
- 195 Learn from View Correlation: An Anchor Enhancement Strategy for Multi-view Clustering, *Suyuan Liu, Ke Liang, Zhibin Dong, Siwei Wang, Xihong Yang, Sihang Zhou, En Zhu, Xinwang Liu*
- 196 Circuit Design and Efficient Simulation of Quantum Inner Product and Empirical Studies of Its Effect on Near-Term Hybrid Quantum-Classic Machine Learning, *Hao Xiong, Yehui Tang, Xinyu Ye, Junchi Yan*
- 197 Discriminability-Driven Channel Selection for Out-of-Distribution Detection, *Yue Yuan, Rundong He, Yicong Dong, Zhongyi Han, Yilong Yin*
- 198 Efficient Hyperparameter Optimization with Adaptive Fidelity Identification, *Jiantong Jiang, Zeyi Wen, Atif Mansoor, Ajmal Mian*
- 199 Probabilistic Sampling of Balanced K-Means using Adiabatic Quantum Computing, *Jan-Nico Zaech, Martin Danelljan, Tolga Birdal, Luc Van Gool*
- 200 Online Task-Free Continual Generative and Discriminative Learning via Dynamic Cluster Memory, *Fei Ye, Adrian G. Bors*
- 201 S²MVTC: a Simple yet Efficient Scalable Multi-View Tensor Clustering, *Zhen Long, Qiyuan Wang, Yazhou Ren, Yipeng Liu, Ce Zhu*
- 202 Spanning Training Progress: Temporal Dual-Depth Scoring (TDDS) for Enhanced Dataset Pruning, *Xin Zhang, Jiawei Du, Yunsong Li, Weiyang Xie, Joey Tianyi Zhou*
- 203 An Aggregation-Free Federated Learning for Tackling Data Heterogeneity, *Yuan Wang, Huazhu Fu, Renuga Kanagavelu, Qingsong Wei, Yong Liu, Rick Siow Mong Goh*
- 204 POCE: Primal Policy Optimization with Conservative Estimation for Multi-constraint Offline Reinforcement Learning, *Jiayi Guan, Li Shen, Ao Zhou, Lusong Li, Han Hu, Xiaodong He, Guang Chen, Changjun Jiang*
- 205 SVDinsTN: A Tensor Network Paradigm for Efficient Structure Search from Regularized Modeling Perspective, *Yu-Bang Zheng, Xi-Le Zhao, Junhua Zeng, Chao Li, Qibin Zhao, Heng-Chao Li, Ting-Zhu Huang*
- 206 Fine-Grained Bipartite Concept Factorization for Clustering, *Chong Peng, Pengfei Zhang, Yongyong Chen, Zhao Kang, Chenglizhao Chen, Qiang Cheng*
- 207 Embodied Multi-Modal Agent trained by an LLM from a Parallel TextWorld, *Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, Yuhui Shi*
- 208 The Mirrored Influence Hypothesis: Efficient Data Influence Estimation by Harnessing Forward Passes, *Myeongseob Ko, Feiyang Kang, Weiyan Shi, Ming Jin, Zhou Yu, Ruoxi Jia*
- 209 Improved Baselines with Visual Instruction Tuning,
* *Haotian Liu, Chunyuan Li, Yuheng Li, Yong Jae Lee*
- 210 Linguistic-Aware Patch Slimming Framework for Fine-grained Cross-Modal Alignment, *Zheren Fu, Lei Zhang, Hou Xia, Zhendong Mao*
- 211 FlowVQTalker: High-Quality Emotional Talking Face Generation through Normalizing Flow and Quantization, *Shuai Tan, Bin Ji, Ye Pan*
- 212 Audio-Visual Segmentation via Unlabeled Frame Exploitation, *Jinxiang Liu, Yikun Liu, Fei Zhang, Chen Ju, Ya Zhang, Yanfeng Wang*
- 213 Binding Touch to Everything: Learning Unified Multimodal Tactile Representations, *Fengyu Yang, Chao Feng, Ziyang Chen, Hyoungseob Park, Daniel Wang, Yiming Dou, Ziyao Zeng, Xien Chen, Rit Gangopadhyay, Andrew Owens, Alex Wong*
- 214 MoDE: CLIP Data Experts via Clustering, *Jiawei Ma, Po-Yao Huang, Saining Xie, Shang-Wen Li, Luke Zettlemoyer, Shih-Fu Chang, Wen-Tau Yih, Hu Xu*
- 215 X-MIC: Cross-Modal Instance Conditioning for Egocentric Action Generalization, *Anna Kukleva, Fadime Sener, Edoardo Remelli, Bugra Tekin, Eric Sauser, Bernt Schiele, Shugao Ma*
- 216 PixelLM: Pixel Reasoning with Large Multimodal Model, *Zhongwei Ren, Zhicheng Huang, Yunchao Wei, Yao Zhao, Dongmei Fu, Jiashi Feng, Xiaojie Jin*
- 217 Probing Synergistic High-Order Interaction in Infrared and Visible Image Fusion, *Naishan Zheng, Man Zhou, Jie Huang, Junming Hou, Haoying Li, Yuan Xu, Feng Zhao*

- 218 The Audio-Visual Conversational Graph: From an Egocentric-Exocentric Perspective, *Wenqi Jia, Miao Liu, Hao Jiang, Ishwarya Ananthabhotla, James M. Rehg, Vamsi Krishna Ithapu, Ruohan Gao*
- 219 MultiPLY: A Multisensory Object-Centric Embodied Large Language Model in 3D World, *Yining Hong, Zishuo Zheng, Peihao Chen, Yian Wang, Junyan Li, Chuang Gan*
- 220 GPT4Point: A Unified Framework for Point-Language
* Understanding and Generation, *Zhangyang Qi, Ye Fang, Zeyi Sun, Xiaoyang Wu, Tong Wu, Jiaqi Wang, Dahua Lin, Hengshuang Zhao*
- 221 LL3DA: Visual Interactive Instruction Tuning for Omni-3D Understanding Reasoning and Planning, *Sijin Chen, Xin Chen, Chi Zhang, Mingsheng Li, Gang Yu, Hao Fei, Hongyuan Zhu, Jiayuan Fan, Tao Chen*
- 222 Unified-IO 2: Scaling Autoregressive Multimodal Models with
* Vision Language Audio and Action, *Jiasen Lu, Christopher Clark, Sangho Lee, Zichen Zhang, Savya Khosla, Ryan Marten, Derek Hoiem, Aniruddha Kembhavi*
- 223 SHAP-EDITOR: Instruction-Guided Latent 3D Editing in Seconds, *Minghao Chen, Junyu Xie, Iro Laina, Andrea Vedaldi*
- 224 Learning to Visually Localize Sound Sources from Mixtures without Prior Source Knowledge, *Dongjin Kim, Sung Jin Um, Sangmin Lee, Jung Uk Kim*
- 225 Bring Event into RGB and LiDAR: Hierarchical Visual-Motion Fusion for Scene Flow, *Hanyu Zhou, Yi Chang, Zhiwei Shi*
- 226 Dispel Darkness for Better Fusion: A Controllable Visual Enhancer based on Cross-modal Conditional Adversarial Learning, *Hao Zhang, Linfeng Tang, Xinyu Xiang, Xuhui Zuo, Jiayi Ma*
- 227 Unraveling Instance Associations: A Closer Look for Audio-Visual Segmentation, *Yuanhong Chen, Yuyuan Liu, Hu Wang, Fengbei Liu, Chong Wang, Helen Frazer, Gustavo Carneiro*
- 228 DMR: Decomposed Multi-Modality Representations for Frames and Events Fusion in Visual Reinforcement Learning, *Haoran Xu, Peixi Peng, Guang Tan, Yuan Li, Xinhai Xu, Yonghong Tian*
- 229 Text-Guided Variational Image Generation for Industrial Anomaly Detection and Segmentation, *Mingyu Lee, Jongwon Choi*
- 230 Tactile-Augmented Radiance Fields, *Yiming Dou, Fengyu Yang, Yi Liu, Antonio Loquercio, Andrew Owens*
- 231 LION: Empowering Multimodal Large Language Model with Dual-Level Visual Knowledge, *Gongwei Chen, Leyang Shen, Rui Shao, Xiang Deng, Liqiang Nie*
- 232 SDSTrack: Self-Distillation Symmetric Adapter Learning for Multi-Modal Visual Object Tracking, *Xiaojuan Hou, Jiazheng Xing, Yijie Qian, Yaowei Guo, Shuo Xin, Junhao Chen, Kai Tang, Mengmeng Wang, Zhengkai Jiang, Liang Liu, Yong Liu*
- 233 Exploring the Transferability of Visual Prompting for Multimodal
* Large Language Models, *Yichi Zhang, Yinpeng Dong, Siyuan Zhang, Tianzan Min, Hang Su, Jun Zhu*
- 234 Mask Grounding for Referring Image Segmentation, *Yong Xien Chng, Henry Zheng, Yizeng Han, Xuchong Qiu, Gao Huang*
- 235 OneLLM: One Framework to Align All Modalities with Language, *Jiaming Han, Kaixiong Gong, Yiyuan Zhang, Jiaqi Wang, Kaipeng Zhang, Dahua Lin, Yu Qiao, Peng Gao, Xiangyu Yue*
- 236 EmoVIT: Revolutionizing Emotion Insights with Visual Instruction Tuning, *Hongxia Xie, Chu-Jun Peng, Yu-Wen Tseng, Hung-Jen Chen, Chan-Feng Hsu, Hong-Han Shuai, Wen-Huang Cheng*
- 237 ModaVerse: Efficiently Transforming Modalities with LLMs, *Xinyu Wang, Bohan Zhuang, Qi Wu*
- 238 PromptKD: Unsupervised Prompt Distillation for Vision-Language Models, *Zheng Li, Xiang Li, Xinyi Fu, Xin Zhang, Weiqiang Wang, Shuo Chen, Jian Yang*
- 239 Dynamic Prompt Optimizing for Text-to-Image Generation, *Wenyi Mo, Tianyu Zhang, Yalong Bai, Bing Su, Ji-Rong Wen, Qing Yang*
- 240 Domain Prompt Learning with Quaternion Networks,
* *Qinglong Cao, Zhengqin Xu, Yuntian Chen, Chao Ma, Xiaokang Yang*
- 241 ViT-Lens: Towards Omni-modal Representations, *Weixian Lei, Yixiao Ge, Kun Yi, Jianfeng Zhang, Difei Gao, Dylan Sun, Yuying Ge, Ying Shan, Mike Zheng Shou*
- 242 Rotated Multi-Scale Interaction Network for Referring Remote Sensing Image Segmentation, *Sihan Liu, Yiwei Ma, Xiaoqing Zhang, Haowei Wang, Jiayi Ji, Xiaoshuai Sun, Rongrong Ji*
- 243 Cyclic Learning for Binaural Audio Generation and Localization, *Zhaojian Li, Bin Zhao, Yuan Yuan*
- 244 Learning to Rematch Mismatched Pairs for Robust Cross-Modal Retrieval, *Haochen Han, Qinghua Zheng, Guang Dai, Minnan Luo, Jingdong Wang*
- 245 VILA: On Pre-training for Visual Language Models, *Ji Lin, Hongxu Yin, Wei Ping, Pavlo Molchanov, Mohammad Shoeybi, Song Han*
- 246 A Picture is Worth More Than 77 Text Tokens: Evaluating CLIP-Style Models on Dense Captions, *Jack Urbanek, Florian Bordes, Pietro Astolfi, Mary Williamson, Vasu Sharma, Adriana Romero-Soriano*
- 247 How to Configure Good In-Context Sequence for Visual Question Answering, *Li Li, Jiawei Peng, Huiyi Chen, Chongyang Gao, Xu Yang*
- 248 CrossMAE: Cross-Modality Masked Autoencoders for Region-Aware Audio-Visual Pre-Training, *Yuxin Guo, Siyang Sun, Shuailei Ma, Kecheng Zheng, Xiaoyi Bao, Shijie Ma, Wei Zou, Yun Zheng*
- 249 Modality-Collaborative Test-Time Adaptation for Action Recognition, *Baochen Xiong, Xiaoshan Yang, Yaguang Song, Yaowei Wang, Changsheng Xu*
- 250 T-VSL: Text-Guided Visual Sound Source Localization in Mixtures, *Tanvir Mahmud, Yapeng Tian, Diana Marculescu*
- 251 UniBind: LLM-Augmented Unified and Balanced Representation Space to Bind Them All, *Yuanhuiyi Lyu, Xu Zheng, Jiazhou Zhou, Lin Wang*
- 252 Monkey: Image Resolution and Text Label Are Important Things
* for Large Multi-modal Models, *Zhang Li, Biao Yang, Qiang Liu, Zhiyin Ma, Shuo Zhang, Jingxu Yang, Yabo Sun, Yuliang Liu, Xiang Bai*
- 253 Rethinking Multi-view Representation Learning via Distilled Disentangling, *Guanzhou Ke, Bo Wang, Xiaoli Wang, Shengfeng He*
- 254 Causal Mode Multiplexer: A Novel Framework for Unbiased Multispectral Pedestrian Detection, *Taeheon Kim, Sebin Shin, Youngjoon Yu, Hak Gu Kim, Yong Man Ro*
- 255 Image-Text Co-Decomposition for Text-Supervised Semantic Segmentation, *Ji-Jia Wu, Andy Chia-Hao Chang, Chieh-Yu Chuang, Chun-Pei Chen, Yu-Lun Liu, Min-Hung Chen, Hou-Ning Hu, Yung-Yu Chuang, Yen-Yu Lin*
- 256 Mirasol3B: A Multimodal Autoregressive Model for Time-Aligned and Contextual Modalities, *AJ Piergiovanni, Isaac Noble, Dahun Kim, Michael S. Ryoo, Victor Gomes, Anelia Angelova*
- 257 Efficient Vision-Language Pre-training by Cluster Masking, *Zihao Wei, Zixuan Pan, Andrew Owens*
- 258 MeLFusion: Synthesizing Music from Image and Language Cues using Diffusion Models, *anjay Chowdhury, Sayan Nag, K J Joseph, Balaji Vasan Srinivasan, Dinesh Manocha*
- 259 Weakly Misalignment-free Adaptive Feature Alignment for UAVs-based Multimodal Object Detection, *Chen Chen, Jiahao Qi, Xingyue Liu, Kangcheng Bin, Ruigang Fu, Xikun Hu, Ping Zhong*
- 260 DiVAS: Video and Audio Synchronization with Dynamic Frame Rates, *Clara Fernandez-Labrador, Mertcan Akçay, Eitan Abecassis, Joan Massich, Christopher Schroers*
- 261 Querying as Prompt: Parameter-Efficient Learning for Multimodal Language Model, *Tian Liang, Jing Huang, Ming Kong, Luyuan Chen, Qiang Zhu*
- 262 SonicVisionLM: Playing Sound with Vision Language Models, *Zhifeng Xie, Shengye Yu, Qile He, Mengtian Li*
- 263 Embracing Unimodal Aleatoric Uncertainty for Robust Multimodal Fusion, *Zixian Gao, Xun Jiang, Xing Xu, Fumin Shen, Yujie Li, Heng Tao Shen*
- 264 C3Net: Compound Conditioned ControlNet for Multimodal Content Generation, *Juntao Zhang, Yuehuai Liu, Yu-Wing Tai, Chi-Keung Tang*
- 265 Composed Video Retrieval via Enriched Context and Discriminative Embeddings, *Omkar Thawakar, Muzammal Naseer, Rao Muhammad Anwer, Salman Khan, Michael Felsberg, Mubarak Shah, Fahad Shahbaz Khan*
- 266 Looking Similar Sounding Different: Leveraging Counterfactual Cross-Modal Pairs for Audiovisual Representation Learning, *Nikhil Singh, Chih-Wei Wu, Iro Orife, Mahdi Kalayeh*
- 267 Anchor-based Robust Finetuning of Vision-Language Models, *Jinwei Han, Zhiwen Lin, Zhongyisun Sun, Yingguo Gao, Ke Yan, Shouhong Ding, Yuan Gao, Gui-Song Xia*

- 268 Event-based Visible and Infrared Fusion via Multi-task Collaboration, *Mengyue Geng, Lin Zhu, Lizhi Wang, Wei Zhang, Ruiqin Xiong, Yonghong Tian*
- 269 Prompt Learning via Meta-Regularization, *Jinyoung Park, Juyeon Ko, Hyunwoo J. Kim*
- 270 Knowledge-Enhanced Dual-stream Zero-shot Composed Image Retrieval, *Yucheng Suo, Fan Ma, Linchao Zhu, Yi Yang*
- 271 Contextual Augmented Global Contrast for Multimodal Intent Recognition, *Kaili Sun, Zhiwen Xie, Mang Ye, Huyin Zhang*
- 272 MRFS: Mutually Reinforcing Image Fusion and Segmentation, *Hao Zhang, Xuhui Zuo, Jie Jiang, Chunchao Guo, Jiayi Ma*
- 273 POPDG: Popular 3D Dance Generation with PopDanceSet, *Zhenye Luo, Min Ren, Xuecai Hu, Yongzhen Huang, Li Yao*
- 274 How to Make Cross Encoder a Good Teacher for Efficient Image-Text Retrieval?, *Yuxin Chen, Zongyang Ma, Ziqi Zhang, Zhongang Qi, Chunfeng Yuan, Bing Li, Junfu Pu, Ying Shan, Xiaojuan Qi, Weiming Hu*
- 275 Active Prompt Learning in Vision Language Models, *Jihwan Bang, Sumyeong Ahn, Jae-Gil Lee*
- 276 Descriptor and Word Soups: Overcoming the Parameter Efficiency Accuracy Tradeoff for Out-of-Distribution Few-shot Learning, *Christopher Liao, Theodoros Tsiligkaridis, Brian Kulis*
- 277 Text-IF: Leveraging Semantic Text Guidance for Degradation-Aware and Interactive Image Fusion, *Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, Jiayi Ma*
- 278 Hallucination Augmented Contrastive Learning for Multimodal Large Language Model, *Chaoya Jiang, Haiyang Xu, Mengfan Dong, Jiaxing Chen, Wei Ye, Ming Yan, Qinghao Ye, Ji Zhang, Fei Huang, Shikun Zhang*
- 279 Beyond Text: Frozen Large Language Models in Visual Signal Comprehension, *Lei Zhu, Fangyun Wei, Yanye Lu*
- 280 Learning Spatial Features from Audio-Visual Correspondence in Egocentric Videos, *Sagnik Majumder, Ziad Al-Halah, Kristen Grauman*
- 281 ES³: Evolving Self-Supervised Learning of Robust Audio-Visual Speech Representations, *Yuanhang Zhang, Shuang Yang, Shiguang Shan, Xilin Chen*
- 282 PortraitBooth: A Versatile Portrait Model for Fast Identity-preserved Personalization, *Xu Peng, Junwei Zhu, Boyuan Jiang, Ying Tai, Donghao Luo, Jiangning Zhang, Wei Lin, Taisong Jin, Chengjie Wang, Rongrong Ji*
- 283 ULIP-2: Towards Scalable Multimodal Pre-training for 3D Understanding, *Le Xue, Ning Yu, Shu Zhang, Artemis Panagopoulou, Junnan Li, Roberto Martín-Martín, Jiajun Wu, Caiming Xiong, Ran Xu, Juan Carlos Nieves, Silvio Savarese*
- 284 AVFF: Audio-Visual Feature Fusion for Video Deepfake Detection, *Trevine Oorloff, Surya Koppiseti, Nicolò Bonettini, Divyaraj Solanki, Ben Colman, Yaser Yacoob, Ali Shahriyari, Gaurav Bharaj*
- 285 Language-aware Visual Semantic Distillation for Video Question Answering, *Bo Zou, Chao Yang, Yu Qiao, Chengbin Quan, Youjian Zhao*
- 286 PerceptionGPT: Effectively Fusing Visual Perception into LLM, *Renjie Pi, Lewei Yao, Jiahui Gao, Jipeng Zhang, Tong Zhang*
- 287 Cooperation Does Matter: Exploring Multi-Order Bilateral Relations for Audio-Visual Segmentation, *Qi Yang, Xing Nie, Tong Li, Pengfei Gao, Ying Guo, Cheng Zhen, Pengfei Yan, Shiming Xiang*
- 288 MV-Adapter: Multimodal Video Transfer Learning for Video Text Retrieval, *Xiaojie Jin, Bowen Zhang, Weibo Gong, Kai Xu, Xueqing Deng, Peng Wang, Zhao Zhang, Xiaohui Shen, Jiashi Feng*
- 289 Draw Step by Step: Reconstructing CAD Construction Sequences from Point Clouds via Multimodal Diffusion., *Weijian Ma, Shuaiqi Chen, Yunzhong Lou, Xueyang Li, Xiangdong Zhou*
- 290 AV-RIR: Audio-Visual Room Impulse Response Estimation, *Anton Ratnarajah, Sreyan Ghosh, Sonal Kumar, Purva Chiniya, Dinesh Manocha*
- 291 Link-Context Learning for Multimodal LLMs, *Yan Tai, Weichen Fan, Zhao Zhang, Ziwei Liu*
- 292 Unveiling the Power of Audio-Visual Early Fusion Transformers with Dense Interactions through Masked Modeling, *Shentong Mo, Pedro Morgado*
- 293 Noisy-Correspondence Learning for Text-to-Image Person Re-identification, *Yang Qin, Yingke Chen, Dezhong Peng, Xi Peng, Joey Tianyi Zhou, Peng Hu*
- 294 Mind Artist: Creating Artistic Snapshots with Human Thought, *Jiaxuan Chen, Yu Qi, Yueming Wang, Gang Pan*
- 295 VTQA: Visual Text Question Answering via Entity Alignment and Cross-Media Reasoning, *Kang Chen, Xiangqian Wu*
- 296 THRONE: An Object-based Hallucination Benchmark for the Free-form Generations of Large Vision-Language Models, *Prannay Kaul, Zhizhong Li, Hao Yang, Yonatan Dukler, Ashwin Swaminathan, C. J. Taylor, Stefano Soatto*
- 297 Data-Efficient Multimodal Fusion on a Single GPU, *✱ Noël Vouitsis, Zhaoyan Liu, Satya Krishna Gorti, Valentin Vilcecroze, Jesse C. Cresswell, Guangwei Yu, Gabriel Loaiza-Ganem, Maksims Volkovs*
- 298 SoundingActions: Learning How Actions Sound from Narrated Egocentric Videos, *Changan Chen, Kumar Ashutosh, Rohit Girdhar, David Harwath, Kristen Grauman*
- 299 Accept the Modality Gap: An Exploration in the Hyperbolic Space, *✱ ameera Ramasinghe, Violetta Shevchenko, Gil Avraham, Ajanthan Thalaisyasingam*
- 300 DiffSal: Joint Audio and Video Learning for Diffusion Saliency Prediction, *Junwen Xiong, Peng Zhang, Tao You, Chuanyue Li, Wei Huang, Yufei Zha*
- 301 DiPrompt: Disentangled Prompt Tuning for Multiple Latent Domain Generalization in Federated Learning, *Sikai Bai, Jie Zhang, Song Guo, Shuaicheng Li, Jingcai Guo, Jun Hou, Tao Han, Xiao Cheng Lu*
- 302 Probabilistic Speech-Driven 3D Facial Motion Synthesis: New Benchmarks Methods and Applications, *Karren D. Yang, Anurag Ranjan, Jen-Hao Rick Chang, Raviteja Vemulapalli, Oncel Tuzel*
- 303 DIEM: Decomposition-Integration Enhancing Multimodal Insights, *Xinyi Jiang, Guoming Wang, Junhao Guo, Juncheng Li, Wenqiao Zhang, Rongxing Lu, Siliang Tang*
- 304 MAFA: Managing False Negatives for Vision-Language Pre-training, *Jaeseok Byun, Dohoon Kim, Taesup Moon*
- 305 AV2AV: Direct Audio-Visual Speech to Audio-Visual Speech *✱ Translation with Unified Audio-Visual Speech Representation, Jeongsoo Choi, Se Jin Park, Minsu Kim, Yong Man Ro*
- 306 Enhancing Multimodal Cooperation via Sample-level Modality Valuation, *Yake Wei, Ruoxuan Feng, Ziheng Wang, Di Hu*
- 307 Diff-BGM: A Diffusion Model for Video Background Music Generation, *Sizhe Li, Yiming Qin, Minghang Zheng, Xin Jin, Yang Liu*
- 308 SaCo Loss: Sample-wise Affinity Consistency for Vision-Language Pre-training, *Sitong Wu, Haoru Tan, Zhuotao Tian, Yukang Chen, Xiaojuan Qi, Jiaya Jia*
- 309 MoPE-CLIP: Structured Pruning for Efficient Vision-Language Models with Module-wise Pruning Error Metric, *Haokun Lin, Haoli Bai, Zhili Liu, Lu Hou, Muye Sun, Linqi Song, Ying Wei, Zhenan Sun*
- 310 Mitigating Noisy Correspondence by Geometrical Structure Consistency Learning, *Zihua Zhao, Mengxi Chen, Tianjie Dai, Jiangchao Yao, Bo Han, Ya Zhang, Yanfeng Wang*
- 311 DetCLIPv3: Towards Versatile Generative Open-vocabulary Object Detection, *Lewei Yao, Renjie Pi, Jianhua Han, Xiaodan Liang, Hang Xu, Wei Zhang, Zhenguo Li, Dan Xu*
- 312 Leveraging Cross-Modal Neighbor Representation for Improved CLIP Classification, *Chao Yi, Lu Ren, De-Chuan Zhan, Han-Jia Ye*
- 313 OmniVec2 - A Novel Transformer based Network for Large Scale Multimodal and Multitask Learning, *Siddharth Srivastava, Gaurav Sharma*
- 314 CoDi-2: In-Context Interleaved and Interactive Any-to-Any Generation, *✱ Zineng Tang, Ziyi Yang, Mahmoud Khademi, Yang Liu, Chenguang Zhu, Mohit Bansal*
- 315 Differentiable Information Bottleneck for Deterministic Multi-view Clustering, *Xiaoqiang Yan, Zhixiang Jin, Fengshou Han, Yangdong Ye*
- 316 A Study of Dropout-Induced Modality Bias on Robustness

- to Missing Video Frames for Audio-Visual Speech Recognition, *Yusheng Dai, Hang Chen, Jun Du, Ruoyu Wang, Shihao Chen, Haotian Wang, Chin-Hui Lee*
- 317 Multimodal Representation Learning by Alternating Unimodal Adaptation, *Xiaohui Zhang, Jaehong Yoon, Mohit Bansal, Huaxiu Yao*
- 318 View-Category Interactive Sharing Transformer for Incomplete
* Multi-View Multi-Label Learning, *hilong Ou, Zhe Xue, Yawen Li, Meiyu Liang, Yuanqiang Cai, Junjiang Wu*
- 319 Scalable 3D Registration via Truncated Entry-wise Absolute Residuals, *Tianyu Huang, Liangzu Peng, Rene Vidal, Yun-Hui Liu*
- 320 Partial-to-Partial Shape Matching with Geometric Consistency, *Viktoria Ehm, Maolin Gao, Paul Roetzer, Marvin Eisenberger, Daniel Cremers, Florian Bernard*
- 321 Towards Robust Learning to Optimize with Theoretical Guarantees, *Qingyu Song, Wei Lin, Juncheng Wang, Hong Xu*
- 322 From Variance to Veracity: Unbundling and Mitigating Gradient Variance in Differentiable Bundle Adjustment Layers, *Swaminathan Gurumurthy, Karnik Ram, Bingqing Chen, Zachary Manchester, Zico Kolter*
- 323 DIMAT: Decentralized Iterative Merging-And-Training for Deep Learning Models, *Nastaran Saadati, Minh Pham, Nasla Saleem, Joshua R. Waite, Aditya Balu, Zhanong Jiang, Chinmay Hegde, Soumik Sarkar*
- 324 Ink Dot-Oriented Differentiable Optimization for Neural Image Halftoning, *Hao Jiang, Bingfeng Zhou, Yadong Mu*
- 325 Are Conventional SNNs Really Efficient? A Perspective from
* Network Quantization, *Guobin Shen, Dongcheng Zhao, Tenglong Li, Jindong Li, Yi Zeng*
- 326 FedMef: Towards Memory-efficient Federated Dynamic Pruning, *Hong Huang, Weiming Zhuang, Chen Chen, Lingjuan Lyu*
- 327 SD4Match: Learning to Prompt Stable Diffusion Model for Semantic Matching, *Xinghui Li, Jingyi Lu, Kai Han, Victor Adrian Prisacariu*
- 328 Purified and Unified Steganographic Network, *Guobiao Li, Sheng Li, Zicong Luo, Zhenxing Qian, Xinpeng Zhang*
- 329 Learned Lossless Image Compression based on Bit Plane Slicing, *Zhe Zhang, Huairui Wang, Zhenzhong Chen, Shan Liu*
- 330 Towards Calibrated Multi-label Deep Neural Networks, *Jiacheng Cheng, Nuno Vasconcelos*
- 331 Improving Generalization via Meta-Learning on Hard Samples, *Nishant Jain, Arun S. Suggala, Pradeep Shenoy*
- 332 Learning with Structural Labels for Learning with Noisy Labels, *Noo-ri Kim, Jin-Seop Lee, Jee-Hyong Lee*
- 333 DiffuseMix: Label-Preserving Data Augmentation with Diffusion Models, *Khawar Islam, Muhammad Zaigham Zaheer, Arif Mahmood, Karthik Nandakumar*
- 334 Improving Out-of-Distribution Generalization in Graphs via Hierarchical Semantic Environments, *Yinhua Piao, Sangseon Lee, Yijingxiu Lu, Sun Kim*
- 335 Patch2Self2: Self-supervised Denoising on Coresets via Matrix Sketching, *Shreyas Fadnavis, Agniva Chowdhury, Joshua Batson, Petros Drineas, Eleftherios Garyfallidis*
- 336 G-FARS: Gradient-Field-based Auto-Regressive Sampling for 3D Part Grouping, *Junfeng Cheng, Tania Stathaki*
- 337 Decompose-and-Compose: A Compositional Approach to Mitigating Spurious Correlation, *Fahimeh Hosseini Noohdani, Parsa Hosseini, Aryan Yazdan Parast, Hamidreza Yaghoubi Araghi, Mahdieh Soleymani Baghshah*
- 338 SkySense: A Multi-Modal Remote Sensing Foundation Model Towards Universal Interpretation for Earth Observation Imagery, *Xin Guo, Jiangwei Lao, Bo Dang, Yingying Zhang, Lei Yu, Lixiang Ru, Liheng Zhong, Ziyuan Huang, Kang Wu, Dingxiang Hu, Huimei He, Jian Wang, Jingdong Chen, Ming Yang, Yongjun Zhang, Yansheng Li*
- 339 Building Bridges across Spatial and Temporal Resolutions: Reference-Based Super-Resolution via Change Priors and Conditional Diffusion Model, *Runmin Dong, Shuai Yuan, Bin Luo, Mengxuan Chen, Jinxiao Zhang, Lixian Zhang, Weijia Li, Juepeng Zheng, Haohuan Fu*
- 340 SatSynth: Augmenting Image-Mask Pairs through Diffusion Models for Aerial Semantic Segmentation, *Aysim Toker, Marvin Eisenberger, Daniel Cremers, Laura Leal-Taixé*
- 341 S2MAE: A Spatial-Spectral Pretraining Foundation Model for
* Spectral Remote Sensing Data, *Xuyang Li, Danfeng Hong, Jocelyn Chanussot*
- 342 Poly Kernel Inception Network for Remote Sensing Detection, *Xinhao Cai, Qiuxia Lai, Yuwei Wang, Wenguan Wang, Zeren Sun, Yazhou Yao*
- 343 Learning without Exact Guidance: Updating Large-scale High-resolution Land Cover Maps from Low-resolution Historical Labels, *Zhuohong Li, Wei He, Jiepan Li, Fangxiao Lu, Hongyan Zhang*
- 344 3D Building Reconstruction from Monocular Remote Sensing Images with Multi-level Supervisions, *Weijia Li, Haote Yang, Zhenghao Hu, Juepeng Zheng, Gui-Song Xia, Conghui He*
- 345 Content-Adaptive Non-Local Convolution for Remote Sensing Pansharpening, *Yule Duan, Xiao Wu, Haoyu Deng, Liang-Jian Deng*
- 346 SG-BEV: Satellite-Guided BEV Fusion for Cross-View Semantic
* Segmentation, *Junyan Ye, Qiyang Luo, Jinhua Yu, Huaping Zhong, Zhimeng Zheng, Conghui He, Weijia Li*
- 347 DiffCast: A Unified Framework via Residual Diffusion for Precipitation Nowcasting, *Demin Yu, Xutao Li, Yunming Ye, Baoquan Zhang, Chuyao Luo, Kuai Dai, Rui Wang, Xunlai Chen*
- 348 MoCha-Stereo: Motif Channel Attention Network for Stereo Matching, *Ziyang Chen, Wei Long, He Yao, Yongjun Zhang, Bingshu Wang, Yongbin Qin, Jia Wu*
- 349 PBWR: Parametric-Building-Wireframe Reconstruction from Aerial LiDAR Point Clouds, *Shangfeng Huang, Ruisheng Wang, Bo Guo, Hongxin Yang*
- 350 Multi-modal Learning for Geospatial Vegetation Forecasting, *Vitus Benson, Claire Robin, Christian Requena-Mesa, Lazaro Alonso, Nuno Carvalhais, José Cortés, Zhihan Gao, Nora Linscheid, Mélanie Weynants, Markus Reichstein*
- 351 Relational Matching for Weakly Semi-Supervised Oriented Object Detection, *Wenhao Wu, Hau-San Wong, Si Wu, Tianyou Zhang*
- 352 Rethinking Transformers Pre-training for Multi-Spectral Satellite Imagery, *Mubashir Noman, Muzammal Naseer, Hisham Cholakkal, Rao Muhammad Anwer, Salman Khan, Fahad Shahbaz Khan*
- 353 Unmixing Diffusion for Self-Supervised Hyperspectral Image Denoising, *Haijin Zeng, Jiezhong Cao, Kai Zhang, Yongyong Chen, Hiep Luong, Wilfried Philips*
- 354 GeoChat: Grounded Large Vision-Language Model for Remote Sensing, *Kartik Kuckreja, Muhammad Sohail Danish, Muzammal Naseer, Abhijit Das, Salman Khan, Fahad Shahbaz Khan*
- 355 Parameter Efficient Self-Supervised Geospatial Domain Adaptation, *Linus Scheibenreif, Michael Mommert, Damian Borth*
- 356 Bridging Remote Sensors with Multisensor Geospatial Foundation Models, *Boran Han, Shuai Zhang, Xingjian Shi, Markus Reichstein*
- 357 CLIP-Driven Open-Vocabulary 3D Scene Graph Generation via
* Cross-Modality Contrastive Learning, *Lianggangxun Chen, Xuejiao Wang, Jiale Lu, Shaohui Lin, Changbo Wang, Gaoqi He*
- 358 Learnable Earth Parser: Discovering 3D Prototypes in Aerial Scans, *Romain Loiseau, Elliot Vincent, Mathieu Aubry, Loic Landrieu*
- 359 Semantics Distortion and Style Matter: Towards Source-free UDA for Panoramic Segmentation, *Xu Zheng, Pengyuan Zhou, Athanasios V. Vasilakos, Lin Wang*
- 360 Geometrically-driven Aggregation for Zero-shot 3D Point Cloud
* Understanding, *Guofeng Mei, Luigi Riz, Yiming Wang, Fabio Poiesi*
- 361 SAM-6D: Segment Anything Model Meets Zero-Shot 6D Object Pose Estimation, *Jiehong Lin, Lihua Liu, Dekun Lu, Kui Jia*
- 362 Construct to Associate: Cooperative Context Learning for Domain Adaptive Point Cloud Segmentation, *Guangrui Li*
- 363 Multi-Task Dense Prediction via Mixture of Low-Rank Experts, *Yuqi Yang, Peng-Tao Jiang, Qibin Hou, Hao Zhang, Jinwei Chen, Bo Li*
- 364 OED: Towards One-stage End-to-End Dynamic Scene Graph Generation, *Guan Wang, Zhimin Li, Qingchao Chen, Yang Liu*
- 365 OMG-Seg: Is One Model Good Enough For All Segmentation?, *Xiangtai Li, Haobo Yuan, Wei Li, Henghui Ding, Size Wu, Wenwei Zhang, Yining Li, Kai Chen, Chen Change Loy*
- 366 DiffusionMTL: Learning Multi-Task Denoising Diffusion Model from Partially Annotated Data, *Hanrong Ye, Dan Xu*

- 367 Bilateral Adaptation for Human-Object Interaction Detection with Occlusion-Robustness, *Guangzhi Wang, Yangyang Guo, Ziwei Xu, Mohan Kankanhalli*
- 368 CurveCloudNet: Processing Point Clouds with 1D Structure, *Colton Stearns, Alex Fu, Jiateng Liu, Jeong Joon Park, Davis Remppe, Despoina Paschalidou, Leonidas J. Guibas*
- 369 VCoder: Versatile Vision Encoders for Multimodal Large Language Models, *Jitesh Jain, Jianwei Yang, Humphrey Shi*
- 370 Amodal Ground Truth and Completion in the Wild, *Guanqi Zhan, Chuanxia Zheng, Weidi Xie, Andrew Zisserman*
- 371 Living Scenes: Multi-object Relocalization and Reconstruction in
* Changing 3D Environments, *Liyuan Zhu, Shengyu Huang, Konrad Schindler, Iro Armeni*
- 372 Single Domain Generalization for Crowd Counting, *Zhuoxuan Peng, S.-H. Gary Chan*
- 373 LTA-PCS: Learnable Task-Agnostic Point Cloud Sampling, *Jiaheng Liu, Jianhao Li, Kaisiyuan Wang, Hongcheng Guo, Jian Yang, Junran Peng, Ke Xu, Xianglong Liu, Jinyang Guo*
- 374 Prompt3D: Random Prompt Assisted Weakly-Supervised 3D Object Detection, *Xiaohong Zhang, Huisheng Ye, Jingwen Li, Qinyu Tang, Yuanqi Li, Yanwen Guo, Jie Guo*
- 375 No More Ambiguity in 360° Room Layout via Bi-Layout Estimation, *Yu-Ju Tsai, Jin-Cheng Jhang, Jingjing Zheng, Wei Wang, Albert Y. C. Chen, Min Sun, Cheng-Hao Kuo, Ming-Hsuan Yang*
- 376 Semantic Line Combination Detector, *Jinwon Ko, Dongkwon Jin, Chang-Su Kim*
- 377 From Pixels to Graphs: Open-Vocabulary Scene Graph Generation with Vision-Language Models, *Rongjie Li, Songyang Zhang, Dahua Lin, Kai Chen, Xuming He*
- 378 PanoContext-Former: Panoramic Total Scene Understanding with a Transformer, *Yuan Dong, Chuan Fang, Liefeng Bo, Zilong Dong, Ping Tan*
- 379 DiffAssemble: A Unified Graph-Diffusion Model for 2D and 3D Reassembly, *Gianluca Scarpellini, Stefano Fiorini, Francesco Giuliari, Pietro Moreiro, Alessio Del Bue*
- 380 ProMotion: Prototypes As Motion Learners, *Yawen Lu, Dongfang Liu, Qifan Wang, Cheng Han, Yiming Cui, Zhiwen Cao, Xueling Zhang, Yingjie Victor Chen, Heng Fan*
- 381 HUNTER: Unsupervised Human-centric 3D Detection via Transferring Knowledge from Synthetic Instances to Real Scenes, *Yichen Yao, Zimo Jiang, Yujing Sun, Zhencai Zhu, Xinge Zhu, Runnan Chen, Yuexin Ma*
- 382 Rethinking the Up-Sampling Operations in CNN-based Generative Network for Generalizable Deepfake Detection, *Chuangchuang Tan, Yao Zhao, Shikui Wei, Guanghua Gu, Ping Liu, Yunchao Wei*
- 383 Shadows Don't Lie and Lines Can't Bend! Generative Models don't know Projective Geometry...for now, *Ayush Sarkar, Hanlin Mai, Amitabh Mahapatra, Svetlana Lazebnik, D.A. Forsyth, Anand Bhattad*
- 384 Text Grouping Adapter: Adapting Pre-trained Text Detector for Layout Analysis, *Tianci Bi, Xiaoyi Zhang, Zhizheng Zhang, Wenxuan Xie, Cuiling Lan, Yan Lu, Nanning Zheng*
- 385 Groupwise Query Specialization and Quality-Aware Multi-Assignment for Transformer-based Visual Relationship Detection, *Jongha Kim, Jihwan Park, Jinyoung Park, Jinyoung Kim, Sehyung Kim, Hyunwoo J. Kim*
- 386 CoralSCOP: Segment any COral Image on this Planet,
* *Ziqiang Zheng, Haixin Liang, Binh-Son Hua, Yue Him Wong, Put Ang Jr, Apple Pui Yi Chui, Sai-Kit Yeung*
- 387 Going Beyond Multi-Task Dense Prediction with Synergy Embedding Models, *Huimin Huang, Yawen Huang, Lanfen Lin, Ruofeng Tong, Yen-Wei Chen, Hao Zheng, Yuexiang Li, Yefeng Zheng*
- 388 Disentangled Pre-training for Human-Object Interaction Detection, *Zhuolong Li, Xingao Li, Changxing Ding, Xiangmin Xu*
- 389 Osprey: Pixel Understanding with Visual Instruction Tuning, *Yuqian Yuan, Wentong Li, Jian Liu, Dongqi Tang, Xinjie Luo, Chi Qin, Lei Zhang, Jianke Zhu*
- 390 Discovering Syntactic Interaction Clues for Human-Object Interaction Detection, *Jinguo Luo, Weihong Ren, Weibo Jiang, Xi'ai Chen, Qiang Wang, Zhi Han, Honghai Liu*
- 391 Flattening the Parent Bias: Hierarchical Semantic Segmentation in the Poincaré Ball, *Simon Weber, Barış Zöngür, Nikita Araslanov, Daniel Cremers*
- 392 HiKER-SGG: Hierarchical Knowledge Enhanced Robust Scene Graph Generation, *Ce Zhang, Simon Stepputtis, Joseph Campbell, Katia Sycara, Yaqi Xie*
- 393 Hierarchical Intra-modal Correlation Learning for Label-free 3D Semantic Segmentation, *Xin Kang, Lei Chu, Jiahao Li, Xuejin Chen, Yan Lu*
- 394 FreePoint: Unsupervised Point Cloud Instance Segmentation, *Zhikai Zhang, Jian Ding, Li Jiang, Dengxin Dai, Guisong Xia*
- 395 GoodSAM: Bridging Domain and Capacity Gaps via Segment Anything Model for Distortion-aware Panoramic Semantic Segmentation, *Weiming Zhang, Yexin Liu, Xu Zheng, Lin Wang*
- 396 MaskClustering: View Consensus based Mask Graph Clustering for Open-Vocabulary 3D Instance Segmentation, *Mi Yan, Jiazhao Zhang, Yan Zhu, He Wang*
- 397 ECoDepth: Effective Conditioning of Diffusion Models for Monocular Depth Estimation, *Suraj Patni, Aradhya Agarwal, Chetan Arora*
- 398 Physical Property Understanding from Language-Embedded Feature Fields, *Albert J. Zhai, Yuan Shen, Emily Y. Chen, Gloria X. Wang, Xinlei Wang, Sheng Wang, Kaiyu Guan, Shenlong Wang*
- 399 LLM4SGG: Large Language Models for Weakly Supervised Scene Graph Generation, *Kibum Kim, Kanghoon Yoon, Jaehyeong Jeon, Yeonjun In, Jinyoung Moon, Donghyun Kim, Chanyoung Park*
- 400 DSGG: Dense Relation Transformer for an End-to-end Scene Graph Generation, *Zeeshan Hayder, Xuming He*
- 401 OTE: Exploring Accurate Scene Text Recognition Using One Token, *Jianjun Xu, Yuxin Wang, Hongtao Xie, Yongdong Zhang*
- 402 SemCity: Semantic Scene Generation with Triplane Diffusion, *Jumin Lee, Sebin Lee, Changho Jo, Woobin Im, Juhyeong Seon, Sung-Eui Yoon*
- 403 Advancing Saliency Ranking with Human Fixations: Dataset Models and Benchmarks, *Bowen Deng, Siyang Song, Andrew P. French, Denis Schluppeck, Michael P. Pound*
- 404 Choose What You Need: Disentangled Representation Learning for Scene Text Recognition Removal and Editing, *Boqiang Zhang, Hongtao Xie, Zuan Gao, Yuxin Wang*
- 405 Leveraging Predicate and Triplet Learning for Scene Graph Generation, *Jiankai Li, Yunhong Wang, Xiefan Guo, Ruijie Yang, Weixin Li*
- 406 Regressor-Segmenter Mutual Prompt Learning for Crowd Counting, *Mingyue Guo, Li Yuan, Zhaoyi Yan, Binghui Chen, Yaowei Wang, Qixiang Ye*
- 407 Learning from Observer Gaze: Zero-Shot Attention Prediction Oriented by Human-Object Interaction Recognition, *Yuchen Zhou, Linkai Liu, Chao Gou*
- 408 EGTR: Extracting Graph from Transformer for Scene Graph Generation,
* *Jinbae Im, JeongYeon Nam, Nokyung Park, Hyungmin Lee, Seunghyun Park*
- 409 SG-PGM: Partial Graph Matching Network with Semantic Geometric Fusion for 3D Scene Graph Alignment and Its Downstream Tasks, *Yaxu Xie, Alain Pagani, Didier Stricker*
- 410 Open-Vocabulary Semantic Segmentation with Image Embedding Balancing, *Xiangheng Shan, Dongyue Wu, Guilin Zhu, Yuanjie Shao, Nong Sang, Changxin Gao*
- 411 Bridging the Synthetic-to-Authentic Gap: Distortion-Guided Unsupervised Domain Adaptation for Blind Image Quality Assessment, *Aobo Li, Jinjian Wu, Yongxu Liu, Leida Li*
- 412 InternVL: Scaling up Vision Foundation Models and Aligning for
* Generic Visual-Linguistic Tasks, *Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, Bin Li, Ping Luo, Tong Lu, Yu Qiao, Jifeng Dai*
- 413 Robust Distillation via Untargeted and Targeted Intermediate Adversarial Samples, *Junhao Dong, Piotr Koniusz, Junxi Chen, Z. Jane Wang, Yew-Soon Ong*

- 414 Class Incremental Learning with Multi-Teacher Distillation, *Haitao Wen, Lili Pan, Yu Dai, Heqian Qiu, Lanxiao Wang, Qingbo Wu, Hongliang Li*
- 415 Large Language Models are Good Prompt Learners for Low-Shot Image Classification, *Zhaoheng Zheng, Jingmin Wei, Xuefeng Hu, Haidong Zhu, Ram Nevatia*
- 416 Consistent Prompting for Rehearsal-Free Continual Learning, *Zhanxin Gao, Jun Cen, Xiaobin Chang*
- 417 Tuning Stable Rank Shrinkage: Aiming at the Overlooked Structural Risk in Fine-tuning, *Sicong Shen, Yang Zhou, Bingzheng Wei, Eric I-Chao Chang, Yan Xu*
- 418 Coherent Temporal Synthesis for Incremental Action Segmentation, *Guodong Ding, Hans Golong, Angela Yao*
- 419 FCS: Feature Calibration and Separation for Non-Exemplar Class Incremental Learning, *Qiwei Li, Yuxin Peng, Jiahuan Zhou*
- 420 DELL: Direct-and-Inverse CLIP for Open-World Few-Shot Learning, *Shuai Shao, Yu Bai, Yan Wang, Baodi Liu, Yicong Zhou*
- 421 Understanding and Improving Source-free Domain Adaptation from a Theoretical Perspective, *Yu Mitsuzumi, Akisato Kimura, Hisashi Kashima*
- 422 Resurrecting Old Classes with New Data for Exemplar-Free Continual Learning, *Dipam Goswami, Albin Soutif-Cormerais, Yuyang Liu, Sandesh Kamath, Bartłomiej Twardowski, Joost van de Weijer*
- 423 Adversarially Robust Few-shot Learning via Parameter Co-distillation of Similarity and Class Concept Learners, *Junhao Dong, Piotr Koniusz, Junxi Chen, Xiaohua Xie, Yew-Soon Ong*
- 424 Learning CNN on ViT: A Hybrid Model to Explicitly Class-specific Boundaries for Domain Adaptation, *Ba Hung Ngo, Nhat-Tuong Do-Tran, Tuan-Ngoc Nguyen, Hae-Gon Jeon, Tae Jong Choi*
- 425 Efficient Stitchable Task Adaptation, *Haoyu He, Zizheng Pan, Jing Liu, Jianfei Cai, Bohan Zhuang*
- 426 Gradient-based Parameter Selection for Efficient Fine-Tuning, *Zhi Zhang, Qizhe Zhang, Zijun Gao, Renrui Zhang, Ekaterina Shutova, Shiji Zhou, Shanghang Zhang*
- 427 ArGue: Attribute-Guided Prompt Tuning for Vision-Language Models, *Xinyu Tian, Shu Zou, Zhaoyuan Yang, Jing Zhang*
- 428 Simple Semantic-Aided Few-Shot Learning, *Hai Zhang, Junzhe Xu, Shanlin Jiang, Zhenan He*
- 429 Long-Tail Class Incremental Learning via Independent Sub-prototype Construction, *Xi Wang, Xu Yang, Jie Yin, Kun Wei, Cheng Deng*
- 430 Few-Shot Object Detection with Foundation Models, *Guangxing Han, Ser-Nam Lim*
- 431 Stronger Fewer & Superior: Harnessing Vision Foundation Models for Domain Generalized Semantic Segmentation, *Zhixiang Wei, Lin Chen, Yi Jin, Xiaoxiao Ma, Tianle Liu, Pengyang Ling, Ben Wang, Huaian Chen, Jinjin Zheng*
- 432 Continual Forgetting for Pre-trained Vision Models, *Hongbo Zhao, Bolin Ni, Junsong Fan, Yuxi Wang, Yuntao Chen, Gaofeng Meng, Zhaoxiang Zhang*
- 433 AETTA: Label-Free Accuracy Estimation for Test-Time Adaptation, *Taeckyoung Lee, Sorn Chottananurak, Taesik Gong, Sung-Ju Lee*
- 434 Continual-MAE: Adaptive Distribution Masked Autoencoders for Continual Test-Time Adaptation, *Jiaming Liu, Ran Xu, Senqiao Yang, Renrui Zhang, Qizhe Zhang, Zehui Chen, Yandong Guo, Shanghang Zhang*
- 435 LEAD: Exploring Logit Space Evolution for Model Selection, *Zixuan Hu, Xiaotong Li, Shixiang Tang, Jun Liu, Yichun Hu, Ling-Yu Duan*
- 436 Instance-based Max-margin for Practical Few-shot Recognition, *Minghao Fu, Ke Zhu*
- 437 Domain Gap Embeddings for Generative Dataset Augmentation, *Yinong Oliver Wang, Younjoon Chung, Chen Henry Wu, Fernando De la Torre*
- 438 JoAPR: Cleaning the Lens of Prompt Learning for Vision-Language Models, *Yuncheng Guo, Xiaodong Gu*
- 439 Generative Multi-modal Models are Good Class Incremental Learners, *Xusheng Cao, Haori Lu, Linlan Huang, Xialei Liu, Ming-Ming Cheng*
- 440 Dual Memory Networks: A Versatile Adaptation Approach for Vision-Language Models, *Yabin Zhang, Wenjie Zhu, Hui Tang, Zhiyuan Ma, Kaiyang Zhou, Lei Zhang*
- 441 UniPT: Universal Parallel Tuning for Transfer Learning with Efficient Parameter and Memory, *Haiwen Diao, Bo Wan, Ying Zhang, Xu Jia, Huchuan Lu, Long Chen*
- 442 Federated Generalized Category Discovery, *Nan Pu, Wenjing Li, Xingyuan Ji, Yalan Qin, Nicu Sebe, Zhun Zhong*
- 443 Learning from One Continuous Video Stream, *João Carreira, Michael King, Viorica Patraucean, Dilara Gokay, Catalin Ionescu, Yi Yang, Daniel Zoran, Joseph Heyward, Carl Doersch, Yusuf Aytar, Dima Damen, Andrew Zisserman*
- 444 OrCo: Towards Better Generalization via Orthogonality and
* Contrast for Few-Shot Class-Incremental Learning, *Noor Ahmed, Anna Kukleva, Bernt Schiele*
- 445 SDDGR: Stable Diffusion-based Deep Generative Replay for Class
* Incremental Object Detection, *Junsu Kim, Hoseong Cho, Jihyeon Kim, Yihalem Yimolal Tiruneh, Seungryul Baek*
- 446 Active Domain Adaptation with False Negative Prediction for
* Object Detection, *Yuzuru Nakamura, Yasunori Ishii, Takayoshi Yamashita*
- 447 Stationary Representations: Optimally Approximating
* Compatibility and Implications for Improved Model Replacements, *Niccolò Biondi, Federico Pernici, Simone Ricci, Alberto Del Bimbo*
- 448 Your Transferability Barrier is Fragile: Free-Lunch for Transferring
* the Non-Transferable Learning, *Ziming Hong, Li Shen, Tongliang Liu*
- 449 Transductive Zero-Shot and Few-Shot CLIP,
* égolène Martin, Yunshi Huang, Fereshteh Shakeri, Jean-Christophe Pesquet, Ismail Ben Ayed
- 450 Task2Box: Box Embeddings for Modeling Asymmetric Task
* Relationships, *Rangel Daroya, Aaron Sun, Subhransu Maji*
- 451 Unbiased Faster R-CNN for Single-source Domain Generalized
* Object Detection, *Yajing Liu, Shijun Zhou, Xiyao Liu, Chunhui Hao, Baojie Fan, Jiandong Tian*
- 452 MetaCloak: Preventing Unauthorized Subject-driven Text-to-
* image Diffusion-based Synthesis via Meta-learning, *Yixin Liu, Chenrui Fan, Yutong Dai, Xun Chen, Pan Zhou, Lichao Sun*

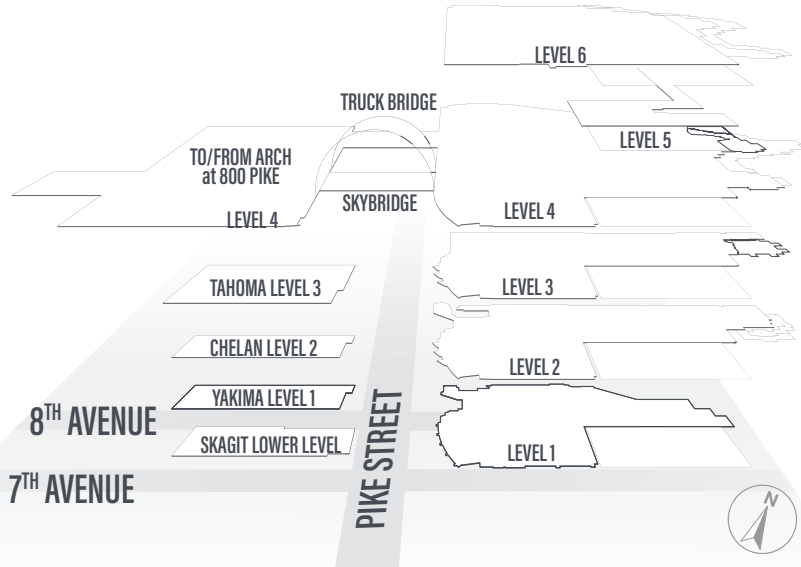
Notes:

Notes:



The Arch building is at 705 Pike Street.

The main pedestrian entrance to Arch is on the corner of 7th Avenue and Pike Street, and the Arch drop-off points are 725 Pike (private/rideshare) and 800 Convention Place (bus).



The Summit building is at 900 Pine Street, just over a block away from Arch.

The main pedestrian entrance is on the corner of 9th Avenue and Pine Street, and the drop-off location is on 9th Avenue between Pine Street and Olive Way (private/rideshare and bus).

