

ICCV **HONOLULU**
OCT 19-23, 2025 **HAWAII**

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

PROGRAM – MAIN CONFERENCE

ULTIMATE SPONSORS



PLATINUM SPONSORS



GOLD SPONSORS



SILVER SPONSORS



Welcome to the 2025 IEEE/CVF International Conference on Computer Vision (ICCV) in Honolulu, Hawai'i! ICCV is the premier biennial conference in computer vision, gathering researchers and practitioners from academia, industry, and government to present and discuss the latest advances in computer vision. Our program features invited keynotes, oral and poster presentations, workshops, tutorials, demonstrations, exhibitions, and numerous networking opportunities. ICCV 2025 is primarily an in-person conference, but a virtual component will also be available, providing access to conference papers, posters, videos, and talks for those unable to attend physically.

ICCV 2025 received 11,239 valid paper submissions, reflecting the community's continued growth and engagement. The review process was overseen by six program chairs, 510 area chairs, and 11,859 reviewers. Each submission underwent a rigorous evaluation, including at least three independent reviews and an author rebuttal period, culminating in final recommendations by the area chairs. A total of 2,701 papers were accepted, resulting in an acceptance rate of 24%, a figure that was not pre-determined and is consistent with historical ICCV rates. All accepted papers will be presented in one of six poster sessions, with 64 selected for oral presentations and 263 designated as highlights in recognition of their exceptional quality and potential impact. A subset of accepted papers has been named award candidates; the best papers, as selected by a committee of distinguished community members, will be announced during the conference's opening session.

For ICCV 2025, we introduced several major changes, some of which have been adopted at other computer vision and AI conferences. To partially address the explosive growth of submissions in the field and the impact on the reviewing process, paper submissions per author were capped at 25. In addition, every qualified author (unless serving in another organizational capacity) was required to serve as a reviewer. Reviewers who failed to submit reviews by the deadline, or whose reviews were flagged as highly irresponsible, faced desk rejection of all papers on which they were authors. These changes were meant to improve accountability, reduce review delays, and preserve the high review standards of the computer vision community.

We extend our deepest appreciation to everyone who has contributed to ICCV 2025. Special thanks go to our area chairs, reviewers, authors, demo

and tutorial participants, and our generous sponsors and exhibitors. We also gratefully acknowledge Nicole Finn and the C to C Events team for conference logistics, as well as all the committee members whose tireless efforts have made this event possible.

Thank you for joining us for ICCV 2025 in beautiful Honolulu. We encourage you to meaningfully engage with our vibrant community and also explore the scenic beauty and rich culture of Hawai'i.

Enjoy ICCV 2025!

ICCV 2025 General Chairs

Hilde Kuehne, *University of Tübingen*
Gerard Medioni, *Amazon*
Dimitris Samaras, *Stony Brook University*
Jingyi Yu, *ShanghaiTech University*
Ramin Zabih, *Cornell University*

ICCV 2025 Program Chairs

Bohyung Han, *Seoul National University*
Aniruddha Kembhavi, *Wayve*
Richard Souvenir, *Temple University*
Deqing Sun, *Google DeepMind*
Ayellet Tal, *Technion*
Jingdong Wang, *Baidu*

ICCV 2025 QR CODES

ICCV 2025 Virtual Platform



Access schedules,
papers, workshops,
tutorials, etc

ICCV 2025 Slido Site



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

Exhibit Hall Floorplan



Access location of
exhibitors on the
main floor

Scholar Inbox Paper Recommender



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

General Chairs

Hilde Kuehne, *University of Tuebingen*
 Gerard Medioni, *Amazon*
 Dimitris Samaras, *Stony Brook University*
 Jingyi Yu, *ShanghaiTech University*
 Ramin Zabih, *Cornell*

Program Chairs

Bohyung Han, *Seoul National University*
 Aniruddha Kembhavi, *Wayve*
 Richard Souvenir, *Temple University*
 Deqing Sun, *Google DeepMind*
 Ayellet Tal, *Technion*
 Jingdong Wang, *Baidu*

Advisor to the Program Committee

David Forsyth, *University of Illinois at Urbana-Champaign*

Tutorial Chairs

Lijie Fan, *Massachusetts Institute of Technology*
 Tejas Gokhale, *University of Maryland, Baltimore County*
 Jun-Yan Zhu, *Carnegie Mellon University*

Workshop Chairs

Vitor Albiero, *Meta*
 Boqing Gong, *Boston University*
 Seunghoon Hong, *KAIST*
 James Tompkin, *Brown University*
 Maria Vakalopoulou, *CentraleSupélec*

Demonstration Chairs

Huaizu Jiang, *Northeastern University*
 Luca Weihs, *Allen Institute for Artificial Intelligence*

Broadening Participation Chairs

Paola Cascante-Bonilla, *Stony Brook University*
 Mahdi Hosseini, *Concordia University*
 Vicky Kalogeiton, *Ecole polytechnique, IP Paris*
 Lingjie Liu, *Saarland Informatics Campus, Max-Planck Institute*
 Si Liu, *Beihang University*
 Yael Moses, *Interdisciplinary Center*
 Tamar Rott Shaham, *MIT*
 Saining Xie, *Courant Institute of Mathematical Sciences*

Technical Chair

Yoshitomo Matsubara, *Yahoo!*

Publications Coordinator

Patrick Kellenberger, *IEEE Computer Society*

Ombud

David Forsyth, *University of Illinois at Urbana-Champaign*

Finance Chair

Walter Scheirer, *University of Notre Dame*

Doctoral Consortium Chairs

Tanmay Gupta, *Allen Institute for Artificial Intelligence*
 Anna Kukleva, *MPII*

Local Chair

Huaijin Chen, *University of Hawaii*

Publicity Chairs

Kosta Derpanis, *York University*
 Antonio Furnari, *University of Catania*
 Nataniel Ruiz, *Google*
 Siyu Zhu, *Fudan University*

Web Developers

Lee Campbell, *ICCV Staff*
 Tony Manzo, *ICCV Staff*

Conference Producer

Nicole Finn, *c to c events*

* - Indicates lead area chairs

Abhinav Shrivastava*	Cheng Deng	Hajime Nagahara	Jonathan T. Barron*	Marcus A Brubaker
Abhishek Gupta	Chi-Keung Tang	Hamid Rezatofighi	Jong Chul Ye*	Marcus Rohrbach
Adam Czajka	Chih-Yuan Yang	Han Hu	Jonghyun Choi*	Margret Keuper
Adrian Hilton	Chong Luo	Hanbyul Joo	Jongwoo Lim*	Maria Vakalopoulou
Adriana Kovashka*	Christoph Feichtenhofer*	Hao Zhang	Joon-Young Lee*	Marius Leordeanu
Ahmet Iscen	Christopher Funk	Hazel Doughty	Juan Carlos Niebles	Mathieu Salzmann*
Aishwarya Agrawal*	Christopher Pal	He Wang	Juergen Gall	Matthew A. Turk
Akihiro Sugimoto*	Chuang Gan	Heng Fan	Jun-Cheng Chen	Mengmeng Wang
Alex Schwing*	Chunhua Shen*	Hengshuang Zhao	Junfeng He	Michael Maire*
Alex Wong	Chunyu Wang	Hiroshi Kawasaki	Jungseock Joo	Michael Niemeyer
Alexander C. Berg*	Cihang Xie	Hongbin Zha*	Junhwa Hur	Michael Rubinstein
Alexandre Alahi	Cordelia Schmid*	Hongdong Li*	Junmo Kim	Michael S. Brown*
Alina Kuznetsova	Cornelia Fermuller*	Hongsheng Li*	Junseok Kwon	Mike Zheng Shou
Alireza Fathi*	Cristian Canton Ferrer	Hongyang Li	Junsong Yuan*	Min H. Kim*
Ameesh Makadia	Daniel Zoran	Huaizu Jiang*	Kai Han	Min Sun
Amit Raj	David Fouhey*	Humphrey Shi	Kaleem Siddiqi	Ming-Hsuan Yang*
Amit Roy-Chowdhury*	David J. Crandall*	Hyun Soo Park*	Kalyan Sunkavalli*	Ming-Ming Cheng
Anand Bhattad	Davide Modolo	Hyung Jin Chang	Karteek Alahari	Mingkui Tan
Andre Araujo*	Diane Larlus	Hyunjung Shim	Kfir Aberman	Mingyu Ding
Andrea Burns	Dima Damen*	Hyunwoo J. Kim	Ko Nishino*	Minhyuk Sung
Andrea Fusiello	Dimitrios Tzionas	Ian Reid	Konrad Schindler	Minsu Cho*
Andrew Owens	Dingwen Zhang	Ilan Shimshoni*	Konstantinos G. Derpanis*	Mohit Gupta*
Andrew Zisserman*	Dong Gong	In Kyu Park*	Kostas Daniilidis	Nalini K. Ratha
Angel X Chang*	Donglai Wei	Ioannis Gkioulekas	Kris Kitani*	Nathan Jacobs*
Angela Yao*	Du Tran	Ioannis Patras	Krishna Kumar Singh	Neal Wadhwa
Angjoo Kanazawa*	Elahe Arani	Ivan Laptev*	Kuk-Jin Yoon*	Neel Joshi*
Anna Rohrbach	Eli Shechtman	Jaesik Park	Kuo-Hao Zeng	Nianyi Li
Anthony Hoogs*	Enze Xie	Jakob Verbeek*	Kushal Kafle*	Nicoletta Noceti
Anton Milan	Evan Shelhamer	Jan van Gemert	Kwang In Kim	Nicu Sebe*
Antoni B. Chan*	Evangelos Kalogerakis*	Jana Kosecka	Kwang Moo Yi	Niloy Mitra*
Antonis Argyros	Federica Bogo	Jean Ponce*	Kyle Genova	Noah Snavely*
Anurag Arnab	Federico Tombari	Jean-Francois Lalonde	Laura Leal-Taixé*	Norimichi Ukita
Anurag Mittal*	Feng Lu	Jean-Marc Odobez	Laura Sevilla-Lara	Nuno Vasconcelos*
Ariel Shamir*	Feng Xu	Jeany Son	Le Lu*	Octavia Camps*
Arsha Nagrani	Feng Yang*	Jennifer J. Sun	Le Wang	Ohad Ben-Shahar
Ashok Veeraraghavan	Frank Dellaert*	Jeong Joon Park	Lei Zhang*	Ohad Fried
Bastian Leibe	Frederic Jurie*	Jia Deng	Lei Zhang*	Oisin Mac Aodha*
Benjamin Busam	Fredrik Kahl	Jia-Bin Huang*	Leonid Sigal*	Olga Veksler
Bernt Schiele*	Friedrich Fraundorfer	Jiajun Wu*	Li Yi	Ondrej Chum*
Bin Xiao	Fumio Okura	Jian Liang	Liang Zheng*	Or Litany
Bing Su	Gabriel Brostow*	Jian Sun	Liang-Chieh Chen*	Orazio Gallo*
Bingbing Ni	Gang Hua*	Jianbo Shi*	Lianli Gao	Oren Freifeld
Bingyi Cao	Gao Huang	Jianfei Cai*	Lijie Fan	Oriane Siméoni
Bingyi Kang	Gengshan Yang	Jiangxin Dong	Limin Wang	P J Narayanan*
Björn Ommer	Georgios Pavlakos	Jiankang Deng	Lin Gu	Pablo Arbelaez
Boqing Gong*	Georgios Tzimiropoulos*	Jianmin Bao	Lingjie Liu	Paolo Favaro
Boxin Shi	Gerard Pons-Moll	Jianwei Yang	Lingxi Xie	Pascal Fua*
Boyi Li	Gim Hee Lee	Jianxin Wu*	Long Zhao	Paul Hongsuck Seo
Brian L. Price	Giorgos Tolias*	Jiaolong Yang*	Lorenzo Baraldi	Peter Vincent Gehler
Bryan A. Plummer	Giovanni Maria Farinella*	Jiasen Lu*	Lorenzo Torresani*	Philipp Kraehenbuehl
Bumsub Ham	Gordon Wetzstein	Jiashi Feng	Lourdes Agapito	Philippos Mordohai*
C.V. Jawahar*	Greg Mori*	Jiawen Chen	Lu Jiang	Ping Luo
Camillo Jose Taylor*	Greg Shakhnarovich*	Jiaya Jia*	Lu Qi	Ping Tan*
Carlos Esteves	Guanbin Li	Jiaying Liu	Lu Yuan*	Piotr Koniusz*
Carsten Rother	Gunhee Kim*	Jifeng Dai	Luisa Verdoliva*	Pratul P. Srinivasan
Chang Xu	Guofeng Zhang	Jingwan Lu	M. Usman Rafique	Qi Dai
Chao Ma	Guorong Li	Jingya Wang	Maja Pantic	Qi Mao
Charles Herrmann	Guy Gilboa*	Jinwei Ye	Manmohan Chandraker*	Qi Tian*
Charless Fowlkes	Gyeongsik Moon	Jiri Matas*	Manolis Savva*	Qibin Hou
Chen Change Loy*	Hadar Averbuch-Elor	Jiwen Lu	Marc Pollefeys*	Qifeng Chen
Chen Chen	Hae-Gon Jeon	Joao P Barreto	Marcella Cornia	Qing Guo
Chen Sun	Haibin Ling*	John Collomosse*	Marcello Pelillo	Qixiang Ye

Qixing Huang*	Shubham Tulsiani	Tobias Ritschel	Xiaolong Wang*	Yong Jae Lee*
Radu Timofte	Sibei Yang	Tomas Pajdla	Xiaoqing Ye	Yongtao Wang*
Raja Giryes*	Silvia Zuffi	Tomer Michaeli	Xiaowei Zhou	Young Min Kim
Ram Nevatia*	Simon Niklaus	Torsten Sattler*	Xiatian Zhu	Yu Kong
Rama Chellappa	Siyu Zhu	Trevor Darrell*	Xilin Chen*	Yu-Chiang Frank Wang
Ran He	Slobodan Ilic	Tushar Nagarajan*	Xin Li	Yu-Wing Tai
Rana Hanocka	Song Bai	Umar Iqbal	Xinggang Wang	Yu-Xiong Wang
Ranjay Krishna*	Songyou Peng	Unnat Jain*	Xinlei Chen*	Yuanjun Xiong
Ravi Ramamoorthi*	Sourav Garg	Varun Jampani*	Xinlong Wang	Yuankai Qi
Rene Vidal	Srinath Sridhar	Vasileios Belagiannis	Xinyu Zhang	Yubei Chen
Robby T. Tan*	Srinivasa Narasimhan*	Venkatesh Babu	Xuejin Chen	Yuchao Dai
Ron Kimmel	Stefan Roth*	Radhakrishnan	Xuhui Jia	Yueqi Duan
Roозbeh Mottaghi*	Stefanos Zafeiriou	Vicente Ordonez*	Xuming He*	Yuexin Ma
Ruigang Yang*	Stella X. Yu*	Vicky Kalogeiton	Yadong MU	Yuhui Yuan
Ruiping Wang	Stephen Gould	Vinay P. Nambodiri*	Yair Weiss*	Yujun Cai
Ryan Farrell*	Stephen Lin*	Vincent Sitzmann	Yale Song	Yulan Guo
S Kevin Zhou*	Stratis Gavves	Vineeth N.	Yan Huang	Yuliang Xiu
Saeed Anwar	Subhransu Maji*	Balasubramanian	Yan Xu	Yulun Zhang
Saining Xie	Suha Kwak*	Vitomir Struc*	Yang Liu	Yumin Suh
Saket Anand*	Sunghyun Cho	Vittorio Ferrari	Yang Wang	Yung-Yu Chuang
Sanghyun Woo	Supasorn Suwajanakorn	Vlad I Morariu*	Yannis Kalantidis	Yuri Boykov
Sara Beery*	Tae Hyun Kim	Vladlen Koltun*	Yao Qin	Yusuke Matsui
Scott McCloskey*	Tae-Hyun Oh*	Wangmeng Zuo	Yasutaka Furukawa*	Yusuke Sugano
Scott Workman	Tae-Kyun Kim	Wankou Yang	Yasuyuki Matsushita*	Yuting Zhang
Seon Joo Kim*	Taesup Moon	Wanli Ouyang*	Yebin Liu	Zhaoxiang Zhang*
Seong Joon Oh	Takayuki Okatani	Wei Liu	Yen-Yu Lin	Zhaozheng Yin
Ser-Nam Lim	Tal Hassner*	Wei Shen	Yezhou Yang	Zhe Gan*
Sergey Tulyakov*	Tat-Jen Cham*	Wei Tang	Yi Fang	Zhe Lin*
Seunghoon Hong	Tat-Jun Chin*	Wei Yang	Yi-Hsuan Tsai	Zhengqi Li
Seungryong Kim	Tatiana Tommasi	Wei-Chiu Ma	Yibing Song	Zhengzhong Tu
Seungyong Lee*	Tatsuya Harada*	Wei-Shi Zheng*	Yijun Li	Zhiding Yu
Shalini De Mello*	Tejas Gokhale	Wenqi Ren	Yilun Du	Zhirong Wu
Shang-Hong Lai	Theo Gevers	Wu Liu	Yin Cui*	Zhouchen Lin*
Shanshan Zhang	Thomas Kollar	Xi Peng	Yin Li	Zhouhui Lian
Shaodi You	Tian Han	Xiang Bai*	Ying Fu	Zhuang Liu
Shengcai Liao	Tianfan Xue	Xiangtai Li	Ying Wu	Zicheng Liu
Shenghua Gao*	Tianhong Li	Xiangyu Xu	Yingli Tian*	Zuxuan Wu
Shenlong Wang	Tianzhu Zhang	Xiangyu Zhang	Yinxiao Li	Zuzana Kukelova
Shiguang Shan*	Ting Liu	Xiaodan Liang	Yizhou Yu	
Shiliang Zhang	Ting Yao	Xiaoguang Han	Yoav Schechner	
Shin'ichi Satoh	Ting Zhang	XIAOJUAN QI	Yoichi Sato*	

We are grateful to all of the 11,859 reviewers who helped make ICCV 2025 possible. We are especially pleased to recognize the following Outstanding Reviewers, whose high-quality reviews (as judged by their Area Chairs) placed them among the top 3% of reviewers.

Ajian Liu	Eduardo Pérez-Pellitero	Junyong Lee	Anh-Quan Cao	Tu Bui
Abhijith Punnappurath	Eloi Zablocki	Juwon Kang	Quentin Bouniot	Urs Waldmann
Adin Ramírez Rivera	Eric Slyman	Kai Fong Ernest Chong	R. Teal Witter	Utkarsh Ojha
Aditya Arora	Fabio Tosi	Kai Han	Raoul de Charette	Viraj Shah
Ahmad Sajedi	Fei Gao	Kanchana Ranasinghe	Ratheesh Kalarot	Vishaal Udandara
Akshay Krishnan	Fei Xue	Kim Yu-Ji	Raza Yunus	Vishwanath Saragadam
Alan Lukezic	Fenggen Yu	Kyros Kutulakos	Renaud Marlet	Vittorio Murino
Albert J. Zhai	Filippo Aleotti	Konstantinos Kamnitsas	Renjun Xu	Vivek Boominathan
Alessandro D'Amelio	Florin-Alexandru Vasluiianu	Kowshik Thopalli	Riccardo de Lutio	Wei Chen
Alessio Del Bue	Franziska Boenisch	Kumara Kahatapitiya	Rinon Gal	Wei-Lun Chao
Alessio Tonioni	Gabriel Schwartz	Lars Hammarstrand	Rishubh Parihar	Weichen Fan
Alexander Becker	Gabriela Csurka	Lei Wang	Rita Cucchiara	Weiwei Sun
Alexander Vilesov	Gabriele Berton	Lena Maier-hein	Robin Hesse	Wolfgang Foerstner
Alina Marcu	Garvita Tiwari	Liangchen Song	Rohit Gandikota	Woong Oh Cho
Ancong Wu	Giuseppe Cartella	Liliane Momeni	Rohit Lal	Xiang An
Andrei Bursuc	Guanghai Wang	Luc Van Gool	Roman Bachmann	Xiang Chen
Andres Bruhn	Guillaume Jeanneret	Luigi Di Stefano	Ross Girshick	Xiang Ji
Aniruddh Sikdar	Guofeng Mei	Mahmoud Afifi	Ruixuan Wang	Xiangjie Sui
Antoine Guedon	Gustau Camps-Valls	Manuel Brack	Rynson W. H. Lau	Xiangyu Chang
Arda Senocak	Han-Hung Lee	Marc Comino-Trinidad	Ryo Yonetani	Xiaoqian Wang
Arijit Ray	Hang Su	Marcel Klemmt	Salman Asif	Xiaoyu Zhou
Aron Monszpart	Hanxi Yin	Maria Taktasheva	Samet Akcay	Xin Huang
Atharva Sehgal	Haochen Zhang	Mariliza Tzes	Samuele Poppi	Xingyu Chen
Auke Wiggers	Hermann Blum	Martin Urschler	Sang Wan Lee	Xinyu Yi
Ayush Jain	Hila Chefer	Massimiliano Mancini	Sejong Yoon	Xinyu Zhou
Azin Jahedi	Hiroshi Kera	Matej Kristan	Seunggyu Chang	Xu Cao
Benjamin Planche	Hongjie Wang	Matias Mattamala	Seyedmorteza Sadat	Xuan Luo
Bhavna Gopal	Hooman Shayani	Matteo Poggi	Shakiba Kheradmand	Yan-Pei Cao
Robert T. Collins	Hoonhee Cho	Matthew Kowal	Shan Zhang	Yana Hasson
Boseung Jeong	Hsuan-I Ho	Matthew Stamm	Shangzhe Wu	Yang Li
Brian K. S. Isaac-Medina	Huajin Tang	Mengyuan Liu	Shaokui Wei	Yangyi Huang
Byoung-Tak Zhang	Hyeongjin Nam	Merey Ramazanov	Shenghai Yuan	Yanpeng Sun
Byung Cheol Song	Hyun Kim	Michael J. Black	Shenghao Fu	Yeqing Li
Byungjun Kim	Idil Esen Zulfikar	Michael Milford	Shizhe Chen	Yicong Hong
Carl Lindström	Inhwan Bae	Ming Liu	Shuai Chen	Yijie Lin
Carlo Masone	Ioannis Papoutsis	Mohammad Fahes	Shuhan Tan	Yilu Wu
Chandrakanth Gudavalli	Ismi Lourentzou	Mohammad Mostafavi	Shuo Wang	Ying Shen
Chengzhou Tang	Jacob Chalk	Moritz Wolter	Siddhant Bansal	Yingyan Celine Lin
Chonghao Sima	Jae-Young Sim	Mutian Xu	Simon Donné	Yiyi Liao
Christian Ledig	Jaejun Yoo	Naoto Inoue	Simon Giebenhain	Yotam Nitzan
Christian Richardt	Jan Dirk Wegner	Ngoc-Son Vu	Sinisa Stekovic	Youwei Lyu
Christoforos N. Spartalis	Jason J. Yu	Niccolò Bisagno	Siyu Lin	Yu Liu
Christopher Metzler	Javier Civera	Nicholas Konz	Sjoerd van Steenkiste	Yue Chen
Chris Thomas	Javier Vazquez-Corral	Nico Lang	Somak Aditya	Yue Gao
Chu Zhou	Jer Pelhan	Nicola Strisciunglio	SouYoung Jin	Yuezun Li
Chuanguang Yang	Jiacheng Chen	Nicolas Audebert	Stan Birchfield	Yuhang He
Chuanxia Zheng	Jiacheng Li	Nikhil Varma Keetha	Stanislaw Antol	Yuke Zhu
Chun-Peng Chang	Jiaming Liu	Niki Amini-Naieni	Stefan Gumhold	Yunzhu Li
Clinton Fookes	Jian Zheng	Nikolaos Sarafianos	Stefan Lee	Yuxuan Xue
Cyrill Stachniss	Jianhao Zheng	Nikos Athanasiou	Stephan Günnemann	Yuzhu Ji
Cyrus Rashtchian	Jianlong Wu	Nyle Siddiqui	Subhankar Roy	Zaid Tasneem
Dan Casas	Jianping Gou	Oren Barkan	Sudipta N. Sinha	Ze Wang
Daniel Glasner	Jichang Li	Ouxiang Li	Sungyeon Kim	Zhaoyang Liu
Darshan Thaker	Jihyong Oh	Ozgur Kara	Takuhiro Kaneko	Zhiqin Chen
David Bau	Jin-Hwi Park	Peiqi Duan	Tao Hu	Zhixiang Wang
David Bourgin	Jinkyu Kim	Pengfei Xu	Tao Wang	Zhonghan Zhao
Daxuan Ren	Johan Edstedt	Peter Kulits	Theodore P. Pavlic	Zike Yan
Deep Chakraborty	Johan W. Verjans	Prithvijit Chattopadhyay	Teddy Furon	Zitian Tang
Delin Qu	Jonas Henry Grebe	Prune Truong	Terrance E. Boulton	Zixiang Zhao
Devis Tuia	Jonathan Sauder	Qihua Dong	Thomas Wimmer	Zixuan Zhu
Donghyeon Kwon	Jongseong Jang	Qimin Chen	Tobias Kirschstein	Zongwei Zhou
Dor Verbin	Juan C. Caicedo	Qing Liu	Tommaso Cavallari	
Duolikhun Danier	Julian F. P. Kooij	Qing Yu	Torben Peters	
Dylan Campbell	Julian Wiederer	Qingping Zheng	Triantafyllos Afouras	
Eduard Zamfir	Junxuan Li	Qize Yang	Truong Nguyen	

Tuesday, October 21

7:00 - 17:00 **Registration/Badge Pickup** (HCC Main Lobby)7:00 - 17:00 **Nursing Pod** 7:00 - 17:00 **Prayer / Quiet Room / Mother's Room** (326 A)7:00 - 9:00 **Breakfast** (Exhibit Hall II)8:00 - 8:45 **Welcome & Awards** (Exhibit Hall III)8:45 - 10:00 **Oral Session 1A: Multi-Modal Learning**
(Exhibit Hall III)8:45 - 10:00 **Oral Session 1B: Structure and Motion**
(Kalakaua Ballroom)10:00 - 11:00 **Coffee Break** (Exhall II)10:00 - 10:30 **Poster Setup** (Exhibit Hall I)10:15 - 11:15 **Keynote** (Exhall III)11:30 - 13:30 **Lunch** (Exhibit Hall II /
Overflow Seating Rooftop and Third Floor)11:30 - 13:30 **Poster Session 1 and Exhibit Hall** (Exhibit Hall I)11:30 - 13:30 **Demos 1** (Exhibit Hall I)13:30 - 14:45 **Oral Session 2A: View Synthesis and Scene**
(Exhibit Hall III)13:30 - 14:45 **Oral Session 2B: Efficient Learning**
(Kalakaua Ballroom)14:30 - 15:00 **Poster Setup** (Exhibit Hall I)15:00 - 17:00 **Poster Session 2 & Exhibit Hall with**
Coffee Break (Exhibit Hall I)15:00 - 17:00 **Demos 2** (Exhibit Hall I)8:45 - 10:00 **Oral Session 1A: Multi-Modal Learning**
(Exhibit Hall III)

Session Chairs: Hengshuang Zhao, Andrea Burns

 - Award candidate paper  - Outstanding Reviewer1 GT-Loc: Unifying When and Where in Images Through a Joint
Embedding Space, *David G. Shatwell, Ishan Rajendrakumar Dave,*
*Sirnam Swetha, Mubarak Shah*2 Scaling Laws for Native Multimodal Models, *Mustafa Shukor,*
Enrico Fini, Victor Guilherme Turrissi da Costa, Matthieu Cord,
*Joshua Susskind, Alaaeldin El-Nouby*3 FixTalk: Taming Identity Leakage for High-Quality Talking Head
Generation in Extreme Cases, *Shuai Tan, Bill Gong, Bin Ji, Ye Pan*4 Differentiable Room Acoustic Rendering with Multi-View Vision
Priors, *Derong Jin, Ruohan Gao*5 Token Activation Map to Visually Explain Multimodal LLMs,
*Yi Li, Hualiang Wang, Xinpeng Ding, Haonan Wang, Xiaomeng Li*8:45 - 10:00 **Oral Session 1B: Structure and Motion**
(Kalakaua Ballroom)

Session Chairs: Junhwa Hur, Ping Tan

1 Multi-View 3D Point Tracking, *Frano Rajić, Haofei Xu,*
Marko Mihajlovic, Siyuan Li, Irem Demir, Emircan Gündoğdu, Lei Ke,
*Sergey Prokudin, Marc Pollefeys, Siyu Tang*2 Uncalibrated Structure from Motion on a Sphere, *Jonathan Ventura,*
*Viktor Larsson, Fredrik Kahl*3 Removing Cost Volumes from Optical Flow Estimators,
*Simon Kiefhaber, Stefan Roth, Simone Schaub-Meyer*4 Image as an IMU: Estimating Camera Motion from a Single Motion-
Blurred Image, *Jerred Chen, Ronald Clark*5 TrajectoryCrafter: Redirecting Camera Trajectory for Monocular
Videos via Diffusion Models, *Mark Yu, Wenbo Hu, Jinbo Xing,*
*Ying Shan*10:00 - 11:00 **Coffee Break** (Exhall II)10:00 - 10:30 **Poster Setup** (Exhibit Hall I)10:15 - 11:15 **Keynote** (Exhibit Hall III)Sheperd Doleman, Astrophysicist at the Center for
Astrophysics | Harvard & Smithsonian and the Director of
the Event Horizon Telescope (EHT)11:30 - 13:30 **Lunch** (Exhibit Hall II / Overflow Seating Rooftop and
Third Floor)11:30 - 13:30 **Poster Session 1 and Exhibit Hall** (Exhibit Hall I)* - Highlight paper  - Award candidate paper

○ - Oral Paper ☆ - Outstanding Reviewer

1 Secure On-Device Video OOD Detection Without Backpropagation,
Shawn Li, Peilin Cai, Yuxiao Zhou, Zhiyu Ni, Renjie Liang, You Qin,
*Yi Nian, Zhengzhong Tu, Xiyang Hu, Yue Zhao*2 Learning Counterfactually Decoupled Attention for Open-World Model
Attribution, *Yu Zheng, Boyang Gong, Fanye Kong, Yueqi Duan,*
*Bingyao Yu, Wenzhao Zheng, Lei Chen, Jiwen Lu, Jie Zhou*3 Latte: Collaborative Test-Time Adaptation of Vision-Language
Models in Federated Learning, *Wenxuan Bao, Ruxi Deng,*
*Ruizhong Qiu, Tianxin Wei, Hanghang Tong, Jingrui He*4 Is Less More? Exploring Token Condensation as Training-free
Test-time Adaptation, *Zixin Wang, Dong Gong, Sen Wang, Zi Huang,*
*Yadan Luo*5 Mastering Collaborative Multi-modal Data Selection: A Focus on
Informativeness, Uniqueness, and Representativeness, *Qifan Yu,*
Zhebei Shen, Zhongqi Yue, Yang Wu, Bosheng Qin, Wenqiao Zhang,
*Yunfei Li, Juncheng Li, Siliang Tang, Yueting Zhuang*6 IRGPT: Understanding Real-world Infrared Image with Bi-cross-modal
Curriculum on Large-scale Benchmark, *Zhe Cao, Jin Zhang,*
*Ruiheng Zhang*7 SMOLoRA: Exploring and Defying Dual Catastrophic Forgetting in
Continual Visual Instruction Tuning, *Ziqi Wang, Chang Che, Qi Wang,*
*Yangyang Li, Zenglin Shi, Meng Wang*8 One Object, Multiple Lies: A Benchmark for Cross-task Adversarial Attack
on Unified Vision-Language Models, *Jiale Zhao, Xinyang Jiang,*
*Junyao Gao, Yuhao Xue, Cairong Zhao*9 Generalized Tensor-based Parameter-Efficient Fine-Tuning via
Lie Group Transformations, *Chongjie Si, Zhiyi Shi, Xuehui Wang,*
*Yichen Xiao, Xiaokang Yang, Wei Shen*10 Bootstrapping Grounded Chain-of-Thought in Multimodal LLMs for
Data-Efficient Model Adaptation, *Jiaer Xia, Bingkui Tong,*
*Yuhang Zang, Rui Shao, Kaiyang Zhou*11 Deciphering Cross-Modal Alignment in Large Vision-Language
Models via Modality Integration Rate, *Qidong Huang, Xiaoyi Dong,*
Pan Zhang, Yuhang Zang, Yuhang Cao, Jiaqi Wang, Weiming Zhang,
*Nenghai Yu*12 X-Fusion: Introducing New Modality to Frozen Large Language
Models, *Sicheng Mo, Thao Nguyen, Xun Huang, Siddharth Srinivasan Iyer,*
Yijun Li, Yuchen Liu, Abhishek Tandon, Eli Shechtman, Krishna Kumar Singh,
*Yong Jae Lee, Bolei Zhou, Yuheng Li*13 LLaVA-KD: A Framework of Distilling Multimodal Large Language
Models, *Yuxuan Cai, Jiangning Zhang, Haoyang He, Xinwei He,*
*Ao Tong, Zhenye Gan, Chengjie Wang, Zhucun Xue, Yong Liu, Xiang Bai*14 TITAN: Query-Token based Domain Adaptive Adversarial Learning,
*Tajamul Ashraf, Janibul Bashir*15 StolenLoRA: Exploring LoRA Extraction Attacks via Synthetic
Data, *Yixu Wang, Yan Teng, Yingchun Wang, Xingjun Ma*16 MMReason: An Open-Ended Multi-Modal Multi-Step Reasoning
Benchmark for MLLMs Toward AGI, *Huanjin Yao, Jiaying Huang,*
Yawen Qiu, Michael K. Chen, Wenzheng Liu, Wei Zhang, Wenjie Zeng,
*Xikun Zhang, Jingyi Zhang, YuXin Song, Wenhao Wu, Dacheng Tao*17 Doodle Your Keypoints: Sketch-Based Few-Shot Keypoint
Detection, *Subhajit Maity, Ayan Kumar Bhunia, Subhadeep Koley,*
*Pinaki Nath Chowdhury, Aneeshan Sain, Yi-Zhe Song*18 Dissecting Generalized Category Discovery: Multiplex Consensus
under Self-Deconstruction, *Luyao Tang, Kunze Huang, Chaoqi Chen,*
Yuxuan Yuan, Chenxin Li, Xiaotong Tu, Xinghao Ding, Yue Huang

- 19 LMM-Det: Make Large Multimodal Models Excel in Object Detection, *Jincheng Li, Chunyu Xie, Ji Ao, Dawei Leng, Yuhui Yin*
- 20 Partial Forward Blocking: A Novel Data Pruning Paradigm for Lossless Training Acceleration, *Dongyue Wu, Zilin Guo, Jialong Zuo, Nong Sang, Changxin Gao*
- 21 ShortV: Efficient Multimodal Large Language Models by Freezing Visual Tokens in Ineffective Layers, *Qianhao Yuan, Qingyu Zhang, Yanjiang Liu, Jiawei Chen, Yaojie Lu, Hongyu Lin, Jia Zheng, Xianpei Han, Le Sun*
- 22 Achieving More with Less: Additive Prompt Tuning for Rehearsal-Free Class-Incremental Learning, *Haoran Chen, Ping Wang, Zihan Zhou, Xu Zhang, Zuxuan Wu, Yu-Gang Jiang*
- 23 CIARD: Cyclic Iterative Adversarial Robustness Distillation, *Liming Lu, Shuchao Pang, Xu Zheng, Xiang Gu, Anan Du, Yunhuai Liu, Yongbin Zhou*
- 24 Moderating the Generalization of Score-based Generative Model, *Wan Jiang, He Wang, Xin Zhang, Dan Guo, Zhaoxin Fan, Yunfeng Diao, Richang Hong*
- 25 Scaling Language-Free Visual Representation Learning,
* *David Fan, Shengbang Tong, Jiachen Zhu, Koustuv Sinha, Zhuang Liu, Xinlei Chen, Michael Rabbat, Nicolas Ballas, Yann LeCun, Amir Bar, Saining Xie*
- 26 LLM-assisted Entropy-based Adaptive Distillation for Unsupervised Fine-grained Visual Representation Learning, *Jianfeng Dong, Danfeng Luo, Daizong Liu, Jie Sun, Xiaoye Qu, Xun Yang, Dongsheng Liu, Xun Wang*
- 27 InfoBridge: Balanced Multimodal Integration through Conditional Dependency Modeling, *Chenxin Li, Yifan Liu, Panwang Pan, Hengyu Liu, Xinyu Liu, Wuyang Li, Cheng Wang, Weihao Yu, Yiyang Lin, Yixuan Yuan*
- 28 A Hidden Stumbling Block in Generalized Category Discovery: Distracted Attention, *Qiyu Xu, Zhanxuan Hu, Yu Duan, Ercheng Pei, Yonghang Tai*
- 29 Meta-Learning Dynamic Center Distance: Hard Sample Mining for Learning with Noisy Labels, *Chenyu Mu, Yijun Qu, Jiexi Yan, Erkun Yang, Cheng Deng*
- 30 ChartPoint: Guiding MLLMs with Grounding Reflection for Chart Reasoning, *Zhengzhuo Xu, SiNan Du, Yiyang Qi, Siwen Lu, Chengjin Xu, Chun Yuan, Jian Guo*
- 31 Multimodal Large Language Model-Guided ISP Hyperparameter Optimization with Dynamic Preference Learning, *Xinyu Sun, Zhikun Zhao, Congyan Lang, Bing Li, Juan Wang*
- 32 Creation-MMBench: Assessing Context-Aware Creative Intelligence in MLLMs, *Xinyu Fang, Zhijian Chen, Kai Lan, Lixin Ma, Shengyuan Ding, Yingji Liang, Xiangyu Zhao, Farong Wen, Zicheng Zhang, Guofeng Zhang, Haodong Duan, Kai Chen, Dahua Lin*
- 33 Gradient Short-Circuit: Efficient Out-of-Distribution Detection via Feature Intervention, *Jiawei Gu, Ziyue Qiao, Zechao Li*
- 34 Boundary Probing for Input Privacy Protection When Using LMM Services, *Xiaofei Hui, Haoxuan Qu, Ping Hu, Hossein Rahmani, Jun Liu*
- 35 Interpretable Zero-Shot Learning with Locally-Aligned Vision-Language Model, *Shiming Chen, Bowen Duan, Salman Khan, Fahad Shahbaz Khan*
- 36 MMCR: Benchmarking Cross-Source Reasoning in Scientific Papers, *Yang Tian, Zheng Lu, Mingqi Gao, Zheng Liu, Bo Zhao*
- 37 ConsNoTrainLoRA: Data-driven Weight Initialization of Low-rank Adapters using Constraints, *Debasmit Das, Hyoungwoo Park, Munawar Hayat, Seokeon Choi, Sungrack Yun, Fatih Porikli*
- 38 UPRE: Zero-Shot Domain Adaptation for Object Detection via Unified Prompt and Representation Enhancement, *Xiao Zhang, Fei Wei, Yong Wang, Wenda Zhao, Feiyi Li, Xiangxiang Chu*
- 39 Dataset Distillation as Data Compression: A Rate-Utility Perspective, *Youneng Bao, Yiping Liu, Zhuo Chen, Yongsheng Liang, Mu Li, Kede Ma*
- 40 Boosting Generative Adversarial Transferability with Self-supervised Vision Transformer Features, *Shangbo Wu, Yu-an Tan, Ruinan Ma, Wencong Ma, Dehua Zhu, Yuanzhang Li*
- 41 Open-set Cross Modal Generalization via Multimodal Unified Representation, *Hai Huang, Yan Xia, Shulei Wang, Hanting Wang, Minghui Fang, Shengpeng Ji, Sashuai Zhou, Tao Jin, Zhou Zhao*
- 42 Adversarial Data Augmentation for Single Domain Generalization via Lyapunov Exponent-Guided Optimization, *Zuyu Zhang, Ning Chen, Yongshan Liu, Qinghua Zhang, Xu Zhang*
- 43 Adversarial Robust Memory-Based Continual Learner, *Xiaoyue Mi, Fan Tang, Zonghan Yang, Danding Wang, Juan Cao, Peng Li, Yang Liu*
- 44 NegRefine: Refining Negative Label-Based Zero-Shot OOD Detection, *Amirhossein Ansari, Ke Wang, Pulei Xiong*
- 45 Divide-and-Conquer for Enhancing Unlabeled Learning, Stability, and Plasticity in Semi-supervised Continual Learning, *Yue Duan, Taicai Chen, Lei Qi, Yinghuan Shi*
- 46 A Unified Framework to BRIDGE Complete and Incomplete Deep Multi-View Clustering under Non-IID Missing Patterns, *Xiaorui Jiang, Buyun He, Peng Yuan Zhou, Xinyue Chen, Jingcai Guo, Jie Xu, Yong Liao*
- 47 HumorDB: Can AI understand graphical humor?, *Vedaant V Jain, Gabriel Kreiman, Felipe dos Santos Alves Feitosa*
- 48 GCAV: A Global Concept Activation Vector Framework for Cross-Layer Consistency in Interpretability, *Zhenghao He, Sanchit Sinha, Guangzhi Xiong, Aidong Zhang*
- 49 Confound from All Sides, Distill with Resilience: Multi-Objective
* *Adversarial Paths to Zero-Shot Robustness, Junhao Dong, Jiao Liu, Xinghua Qu, Yew-Soon Ong*
- 50 Mitigating Object Hallucinations via Sentence-Level Early Intervention, *Shangpin Peng, Senqiao Yang, Li Jiang, Zhuotao Tian*
- 51 Active Membership Inference Test (aMINT): Enhancing Model Auditability with Multi-Task Learning., *Daniel DeAlcala, Aythami Morales, Julian Fierrez, Gonzalo Mancera, Ruben Tolosana, Javier Ortega-Garcia*
- 52 Unknown Text Learning for CLIP-based Few-Shot Open-set Recognition, *Rui Ma, Qilong Wang, Bing Cao, Qinghua Hu, Yahong Han*
- 53 One-Shot Knowledge Transfer for Scalable Person Re-identification, *Longhua Li, Lei Qi, Xin Geng*
- 54 ShortFT: Diffusion Model Alignment via Shortcut-based Fine-Tuning, *Xiefan Guo, Miaomiao Cui, Liefeng Bo, Di Huang*
- 55 PRISM: Reducing Spurious Implicit Biases in Vision-Language Models with LLM-Guided Embedding Projection, *Mahdiyar Molahassani, Azadeh Motamedi, Michael Greenspan, Il-Min Kim, Ali Etemad*
- 56 Open-Unfairness Adversarial Mitigation for Generalized Deepfake Detection, *Zhaoyang Li, Zhu Teng, Baopeng Zhang, Jianping Fan*
- 57 MultiVerse: A Multi-Turn Conversation Benchmark for Evaluating Large Vision and Language Models, *Young-Jun Lee, Byung-Kwan Lee, Jianshu Zhang, Yechan Hwang, Byungsoo Ko, Han-Gyu Kim, Dongyu Yao, Xuankun Rong, Eojin Joo, Seung-Ho Han, Bowon Ko, Ho-Jin Choi*
- 58 Spatial Preference Rewarding for MLLMs Spatial Understanding, *Han Qiu, Peng Gao, Lewei Lu, Xiaoqin Zhang, Ling Shao, Shijian Lu*
- 59 EA-KD: Entropy-based Adaptive Knowledge Distillation, *Chi-Ping Su, Ching-Hsun Tseng, Bin Pu, Lei Zhao, Jiewen Yang, Zhuangzhuang Chen, Shin-Jye Lee*
- 60 Structured Policy Optimization: Enhance Large Vision-Language Model via Self-referenced Dialogue, *Guohao Sun, Can Qin, Yihao Feng, Zeyuan Chen, Ran Xu, Sohail Dianat, Majid Rabbani, Raghuveer Rao, Zhiqiang Tao*
- 61 Seal Your Backdoor with Variational Defense, *Ivan Sabolić, Matej Grcić, Siniša Šegvić*
- 62 Semi-ViM: Bidirectional State Space Model for Mitigating Label Imbalance in Semi-Supervised Learning, *Hongyang He, Hongyang Xie, Haochen You, Victor Sanchez*
- 63 CODE-CL: Conceptor-Based Gradient Projection for Deep Continual Learning, *Marco P. E. Apolinario, Sakshi Choudhary, Kaushik Roy*
- 64 SAMO: A Lightweight Sharpness-Aware Approach for Multi-Task Optimization with Joint Global-Local Perturbation, *Hao Ban, Gokul Ram Subramani, Kaiyi Ji*
- 65 Beyond the Limits: Overcoming Negative Correlation of Activation-Based Training-Free NAS, *Haidong Kang, Lianbo Ma, Pengjun Chen, Guo Yu, Xingwei Wang, Min Huang*
- 66 Integrating Task-Specific and Universal Adapters for Pre-Trained Model-based Class-Incremental Learning, *Yan Wang, Da-Wei Zhou, Han-Jia Ye*

- 67 I Am Big, You Are Little; I Am Right, You Are Wrong, *David A. Kelly, Akchunya Chanchal, Nathan Blake*
- 68 Semi-supervised Deep Transfer for Regression without Domain Alignment, *Mainak Biswas, Ambedkar Dukkipati, Devarajan Sridharan*
- 69 DocThinker: Explainable Multimodal Large Language Models with Rule-based Reinforcement Learning for Document Understanding, *Wenwen Yu, Zhibo Yang, Yuliang Liu, Xiang Bai*
- 70 CVPT: Cross Visual Prompt Tuning, *Lingyun Huang, Jianxu Mao, Junfei Yi, Ziming Tao, Yaonan Wang*
- 71 From Easy to Hard: The MIR Benchmark for Progressive Interleaved Multi-Image Reasoning, *Hang Du, Jiayang Zhang, Guoshun Nan, Wendi Deng, Zhenyan Chen, Chenyang Zhang, Wang Xiao, Shan Huang, Yuqi Pan, Tao Qi, Sicong Leng*
- 72 Learning Separable Fine-Grained Representation via Dendrogram Construction from Coarse Labels for Fine-grained Visual Recognition, *Guanghui Shi, Xuefeng Liang, Wenjie Li, Xiaoyu Lin*
- 73 Diffusion Guided Adaptive Augmentation for Generalization in Visual Reinforcement Learning, *Jeong Woon Lee, Hyoseok Hwang*
- 74 Causal Disentanglement and Cross-Modal Alignment for Enhanced Few-Shot Learning, *Tianjiao Jiang, Zhen Zhang, Yuhang Liu, Javen Qinfeng Shi*
- 75 Multi-View 3D Point Tracking, *Frano Rajič, Haofer Xu, Marko Mihajlovic, Siyuan Li, Irem Demir, Emircan Gündoğdu, Lei Ke, Sergey Prokudin, Marc Pollefeys, Siyu Tang*
- 76 Removing Cost Volumes from Optical Flow Estimators, *Simon Kiefhaber, Stefan Roth, Simone Schaub-Meyer*
- 77 CA2C: A Prior-Knowledge-Free Approach for Robust Label Noise Learning via Asymmetric Co-learning and Co-training, *Mengmeng Sheng, Zeren Sun, Tianfei Zhou, Xiangbo Shu, Jinshan Pan, Yazhou Yao*
- 78 Fast Globally Optimal and Geometrically Consistent 3D Shape Matching, *Paul Roetzer, Florian Bernard*
- 79 A Framework for Double-Blind Federated Adaptation of Foundation Models, *Nurbek Tastan, Karthik Nandakumar*
- 80 Customizing Domain Adapters for Domain Generalization, *Yuyang Ji, Zeyi Huang, Haohan Wang, Yong Jae Lee*
- 81 Towards Scalable Spatial Intelligence via 2D-to-3D Data Lifting, *Xingyu Miao, Haoran Duan, Qianhao Qian, Jiuniu Wang, Yang Long, Ling Shao, Deli Zhao, Ran Xu, Gongjie Zhang*
- 82 SRefiner: Soft-Braid Attention for Multi-Agent Trajectory Refinement, *Liwen Xiao, Zhiyu Pan, Zhicheng Wang, Zhiguo Cao, Wei Li*
- 83 Underwater Visual SLAM with Depth Uncertainty and Medium Modeling, *Rui Liu, Sheng Fan, Wenguan Wang, Yi Yang*
- 84 Aligning Vision to Language: Annotation-Free Multimodal Knowledge Graph Construction for Enhanced LLMs Reasoning, *Junming Liu, Siyuan Meng, Yanting Gao, Song Mao, Pinlong Cai, Guohang Yan, Yirong Chen, Zilin Bian, Ding Wang, Botian Shi*
- 85 AIM: Amending Inherent Interpretability via Self-Supervised Masking, *Eyad Alshami, Shashank Agnihotri, Bernt Schiele, Margret Keuper*
- 86 Reinforcement Learning-Guided Data Selection via Redundancy Assessment, *Suorong Yang, Peijia Li, Furao Shen, Jian Zhao*
- 87 Deep Incomplete Multi-view Clustering with Distribution Dual-Consistency Recovery Guidance, *Jiaqi Jin, Siwei Wang, Zhibin Dong, Xihong Yang, Xinwang Liu, En Zhu, Kunlun He*
- 88 VGGSounder: Audio-Visual Evaluations for Foundation Models, *Daniil Zverev, Thaddäus Wiedemer, Ameysa Prabhu, Matthias Bethge, Wieland Brendel, A. Sophia Koepke*
- 89 EA-Vit: Efficient Adaptation for Elastic Vision Transformer, *Chen Zhu, Wangbo Zhao, Huiwen Zhang, Yuhao Zhou, Weidong Tang, Shuo Wang, Zhihang Yuan, Yuzhang Shang, Xiaojiang Peng, Kai Wang, Dawei Yang*
- 90 Web Artifact Attacks Disrupt Vision Language Models, *Maan Qraitem, Piotr Teterwak, Kate Saenko, Bryan A. Plummer*
- 91 Tensor-aggregated LoRA in Federated Fine-tuning, *Zhixuan Li, Bingqian Xu, Xiangbo Shu, Jiachao Zhang, Yazhou Yao, Guo-Sen Xie, Jinhui Tang*
- 92 Feature Coding in the Era of Large Models: Dataset, Test Conditions, and Benchmark, *Changsheng Gao, Yifan Ma, Qiaoxi Chen, Yenan Xu, Dong Liu, Weisi Lin*
- 93 Generate, Refine, and Encode: Leveraging Synthesized Novel Samples for On-the-Fly Fine-Grained Category Discovery, *Xiao Liu, Nan Pu, Haiyang Zheng, Wenjing Li, Nicu Sebe, Zhun Zhong*
- 94 MMOne: Representing Multiple Modalities in One Scene, *Zhifeng Gu, Bing Wang*
- 95 MM-IFEngine: Towards Multimodal Instruction Following, *Shengyuan Ding, Shenxi Wu, Xiangyu Zhao, Yuhang Zang, Haodong Duan, Xiaoyi Dong, Pan Zhang, Yuhang Cao, Dahua Lin, Jiaqi Wang*
- 96 Knowledge Distillation with Refined Logits, *Wujie Sun, Defang Chen, Siwei Lyu, Genlang Chen, Chun Chen, Can Wang*
- 97 CoST: Efficient Collaborative Perception From Unified Spatiotemporal Perspective, *Zongheng Tang, Yi Liu, Yifan Sun, Yulu Gao, Jinyu Chen, Runsheng Xu, Si Liu*
- 98 RainbowPrompt: Diversity-Enhanced Prompt-Evolving for Continual Learning, *Kiseong Hong, Gyeong-hyeon Kim, Eunwoo Kim*
- 99 Causality-guided Prompt Learning for Vision-language Models via Visual Granulation, *Mengyu Gao, QiuLei Dong*
- 100 FA: Forced Prompt Learning of Vision-Language Models for Out-of-Distribution Detection, *Xinhua Lu, Runhe Lai, Yanqi Wu, Kanghao Chen, Wei-Shi Zheng, Ruixuan Wang*
- 101 VisionMath: Vision-Form Mathematical Problem-Solving, *Zongyang Ma, Yuxin Chen, Ziqi Zhang, Zhongang Qi, Chunfeng Yuan, Shaojie Zhu, Chengxiang Zhuo, Bing Li, Ye Liu, Zang Li, Ying Shan, Weiming Hu*
- 102 Scaling Inference-Time Search with Vision Value Model for Improved Visual Comprehension, *Xiyao Wang, Zhengyuan Yang, Linjie Li, Hongjin Lu, Yuancheng Xu, Chung-Ching Lin, Kevin Lin, Furong Huang, Lijuan Wang*
- 103 Contrastive Flow Matching, *George Stoica, Vivek Ramanujan, Xiang Fan, Ali Farhadi, Ranjay Krishna, Judy Hoffman*
- 104 SemiVisBooster: Boosting Semi-Supervised Learning for Fine-Grained Classification through Pseudo-Label Semantic Guidance, *Wenjin Zhang, Xinyu Li, Chenyang Gao, Ivan Marsic*
- 105 Dataset Distillation via the Wasserstein Metric, *Haoyang Liu, Yijiang Li, Tiancheng Xing, Peiran Wang, Vibhu Dalal, Luwei Li, Jingrui He, Haohan Wang*
- 106 Membership Inference Attacks with False Discovery Rate Control, *Chenxu Zhao, Wei Qian, Aobo Chen, Mengdi Huai*
- 107 Acknowledging Focus Ambiguity in Visual Questions, *Chongyan Chen, Yu-Yun Tseng, Zhuoheng Li, Anush Venkatesh, Danna Gurari*
- 108 A Good Teacher Adapts Their Knowledge for Distillation, *Chengyao Qian, Trung Le, Mehrtaash Harandi*
- 109 Evading Data Provenance in Deep Neural Networks, *Hongyu Zhu, Sichu Liang, Wenwen Wang, Zhuomeng Zhang, Fangqi Li, Shi-Lin Wang*
- 110 Boosting Adversarial Transferability via Residual Perturbation Attack, *Jinjia Peng, Zeze Tao, Huibing Wang, Meng Wang, Yang Wang*
- 111 MAVias: Mitigate any Visual Bias, *Ioannis Sarridis, Christos Koutlis, Symeon Papadopoulos, Christos Diou*
- 112 MPBR: Multimodal Progressive Bidirectional Reasoning for Open-Set Fine-Grained Recognition, *Junfu Tan, Peiguang Jing, Yu Zhu, Yu Liu*
- 113 Towards Higher Effective Rank in Parameter-Efficient Fine-tuning using Khatri-Rao Product, *Paul Albert, Frederic Z. Zhang, Hemanth Saratchandran, Anton van den Hengel, Ehsan Abbasnejad*
- 114 Revisiting Pool-based Prompt Learning for Few-shot Class-incremental Learning, *Yongwei Jiang, Yixiong Zou, Yuhua Li, Ruixuan Li*
- 115 Federated Representation Angle Learning, *Liping Yi, Han Yu, Gang Wang, Xiaoguang Liu, Xiaoxiao Li*
- 116 Federated Continual Instruction Tuning, *Haiyang Guo, Fanhu Zeng, Fei Zhu, Wenzhuo Liu, Da-Han Wang, Jian Xu, Xu-Yao Zhang, Cheng-Lin Liu*
- 117 Scaling Omni-modal Pretraining with Multimodal Context: Advancing Universal Representation Learning Across Modalities, *Yiyuan Zhang, Handong Li, Jing Liu, Xiangyu Yue*
- 118 More Reliable Pseudo-labels, Better Performance: A Generalized Approach to Single Positive Multi-label Learning, *Luong Tran, Thieu Vo, Anh Nguyen, Sang Dinh, Van Nguyen*

- 119 TWIST & SCOUT: Grounding Multimodal LLM-Experts by Forget-Free Tuning, *Aritra Bhowmik, Mohammad Mahdi Derakhshani, Dennis Koelma, Yuki M. Asano, Martin R. Oswald, Cees G. M. Snoek*
- 120 Generate, Transduct, Adapt: Iterative Transduction with VLMs, *Oindrila Saha, Logan Lawrence, Grant Van Horn, Subhransu Maji*
- 121 BabyVLM: Data-Efficient Pretraining of VLMs Inspired by Infant Learning, *Shengao Wang, Arjun Chandra, Aoming Liu, Venkatesh Saligrama, Boqing Gong*
- 122 Controlling Multimodal LLMs via Reward-guided Decoding, *Oscar Mañas, Pierluca D'Oro, Koustuv Sinha, Adriana Romero-Soriano, Michal Drozdal, Aishwarya Agrawal*
- 123 Improving Large Vision and Language Models by Learning from a Panel of Peers, *Jefferson Hernandez, Jing Shi, Simon Jenni, Vicente Ordóñez, Kushal Kafle*
- 124 CE-FAM: Concept-Based Explanation via Fusion of Activation Maps, *Michihiro Kuroki, Toshihiko Yamasaki*
- 125 Leveraging Spatial Invariance to Boost Adversarial Transferability, *Zihan Zhou, Li Li, Yanli Ren, Chuan Qin, Guorui Feng*
- 126 Client2Vec: Improving Federated Learning by Distribution Shifts Aware Client Indexing, *Yongxin Guo, Lin Wang, Xiaoying Tang, Tao Lin*
- 127 A Tiny Change, A Giant Leap: Long-Tailed Class-Incremental Learning via Geometric Prototype Alignment, *Xinyi Lai, LuoJun Lin, Weijie Chen, Yuanlong Yu*
- 128 PEFTDiff: Diffusion-Guided Transferability Estimation for Parameter-Efficient Fine-Tuning, *Pravful Kumar Khoba, Zijian Wang, Chetan Arora, Mahsa Baktashmotlagh*
- 129 One Last Attention for Your Vision-Language Model, *Liang Chen, Ghazi Shazan Ahmad, Tianjun Yao, Lingqiao Liu, Zhiqiang Shen*
- 130 Forgetting Through Transforming: Enabling Federated Unlearning via Class-Aware Representation Transformation, *Qi Guo, Zhen Tian, Minghao Yao, Saiyu Qi, Yong Qi, Bingyi Liu*
- 131 MMAT-1M: A Large Reasoning Dataset for Multimodal Agent Tuning, *Tianhong Gao, Yannian Fu, Weiqun Wu, Haixiao Yue, Shanshan Liu, Gang Zhang*
- 132 Geometry Distributions, *Biao Zhang, Jing Ren,*
* *Peter Wonka*
- 133 FastJSMA: Accelerating Jacobian-based Saliency Map Attacks through Gradient Decoupling, *Zhenghao Gao, Shengjie Xu, Zijing Li, Meixi Chen, Chaojian Yu, Yuanjie Shao, Changxin Gao*
- 134 Boosting Class Representation via Semantically Related Instances for Robust Long-Tailed Learning with Noisy Labels, *Yuhang Li, ZhuYing Li, Yuheng Jia*
- 135 Pi-GPS: Enhancing Geometry Problem Solving by Unleashing the Power of Diagrammatic Information, *Junbo Zhao, Ting Zhang, Jiayu Sun, Mi Tian, Hua Huang*
- 136 VAGUE: Visual Contexts Clarify Ambiguous Expressions, *Heejeong Nam, Jinwoo Ahn, Keummin Ka, Jiwan Chung, Youngjae Yu*
- 137 Diffusion-based Source-biased Model for Single Domain Generalized Object Detection, *Han Jiang, Wenfei Yang, Tianzhu Zhang, Yongdong Zhang*
- 138 PRO-VPT: Distribution-Adaptive Visual Prompt Tuning via Prompt Relocation, *Chikai Shang, Mengke Li, Yiqun Zhang, Zhen Chen, Jinlin Wu, Fangqing Gu, Yang Lu, Yiu-Ming Cheung*
- 139 BATCLIP: Bimodal Online Test-Time Adaptation for CLIP, *Sarthak Maharana, Baoming Zhang, Leonid Karlinsky, Rogerio Feris, Yunhui Guo*
- 140 Mind the Cost of Scaffold! Benign Clients May Even Become Accomplices of Backdoor Attack, *Xingshuo Han, Xuanye Zhang, Xiang Lan, Haozhao Wang, Shengmin Xu, Shen Ren, Jason Zeng, Ming Wu, Michael Heinrich, Tianwei Zhang*
- 141 AVTrustBench: Assessing and Enhancing Reliability and Robustness in Audio-Visual LLMs, *Sanjoy Chowdhury, Sayan Nag, Subhrajyoti Dasgupta, Yaoting Wang, Mohamed Elhoseiny, Ruohan Gao, Dinesh Manocha*
- 142 Verbalized Representation Learning for Interpretable Few-Shot Generalization, *Cheng-Fu Yang, Da Yin, Wenbo Hu, Heng Ji, Nanyun Peng, Bolei Zhou, Kai-Wei Chang*
- 143 UIPro: Unleashing Superior Interaction Capability For GUI Agents, *Hongxin Li, Jingran Su, Jingfan Chen, Zheng Ju, Yuntao Chen, Qing Li, Zhaoxiang Zhang*
- 144 Loss Functions for Predictor-based Neural Architecture Search, *Han Ji, Yuqi Feng, Jiahao Fan, Yanan Sun*
- 145 DiMPLe - Disentangled Multi-Modal Prompt Learning: Enhancing Out-Of-Distribution Alignment with Invariant and Spurious Feature Separation, *Umaira Rahman, Mohammad Yaqub, Dwarikanath Mahapatra*
- 146 GRAB: A Challenging GRaph Analysis Benchmark for Large Multimodal Models, *Jonathan Roberts, Kai Han, Samuel Albanie*
- 147 MosaicDiff: Training-free Structural Pruning for Diffusion Model Acceleration Reflecting Pretraining Dynamics, *Bowei Guo, Shengkun Tang, Cong Zeng, Zhiqiang Shen*
- 148 GLEAM: Enhanced Transferable Adversarial Attacks for Vision-Language Pre-training Models via Global-Local Transformations, *Yunqi Liu, Xue Ouyang, Xiaohui Cui*
- 149 Adversarial Training for Probabilistic Robustness, *Yi Zhang, Yuhang Chen, Zhen Chen, Wenjie Ruan, Xiaowei Huang, Siddhartha Khastgir, Xingyu Zhao*
- 150 RMultiplex200K: Toward Reliable Multimodal Process Supervision for Visual Language Models on Telecommunications, *Sijia Chen, Bin Song*
- 151 Diffusion Curriculum: Synthetic-to-Real Data Curriculum via Image-Guided Diffusion, *Yijun Liang, Shweta Bhardwaj, Tianyi Zhou*
- 152 Backdoor Defense via Enhanced Splitting and Trap Isolation, *Hongrui Yu, Lu Qi, Wanyu Lin, Jian Chen, Hailong Sun, Chengbin Sun*
- 153 GT-Loc: Unifying When and Where in Images Through a Joint Embedding Space, *David G. Shatwell, Ishan Rajendrakumar Dave, Sirnam Swetha, Mubarak Shah*
- 154 TrajectoryCrafter: Redirecting Camera Trajectory for Monocular Videos via Diffusion Models, *Mark Yu, Wenbo Hu, Jinbo Xing, Ying Shan*
- 155 Benchmarking Multimodal CoT Reward Model Stepwise by Visual Program, *Minghe Gao, Xuqi Liu, Zhongqi Yue, Yang Wu, Shuang Chen, Juncheng Li, Siliang Tang, Fei Wu, Tat-Seng Chua, Yueting Zhuang*
- 156 AIRA: Activation-Informed Low-Rank Adaptation for Large Models, *Lujun Li, Dezhi Li, Cheng Lin, Wei Li, Wei Xue, Sirui Han, Yike Guo*
- 157 Social Debiasing for Fair Multi-modal LLMs, *Harry Cheng, Yangyang Guo, Qingpei Guo, Ming Yang, Tian Gan, Weili Guan, Liqiang Nie*
- 158 Equipping Vision Foundation Model with Mixture of Experts for Out-of-Distribution Detection, *Shizhen Zhao, Jiahui Liu, Xin Wen, Haoru Tan, Xiaojuan Qi*
- 159 LIRA: Reasoning Reconstruction via Multimodal Large Language Models, *Zhen Zhou, Tong Wang, Yunkai Ma, Xiao Tan, Fengshui Jing*
- 160 Class-Wise Federated Averaging for Efficient Personalization, *Gyuejeong Lee, Daeyoung Choi*
- 161 Towards Cross-modal Backward-compatible Representation Learning for Vision-Language Models, *Young Kyun Jang, Ser-nam Lim*
- 162 Region-based Cluster Discrimination for Visual Representation Learning, *Yin Xie, Kaicheng Yang, Xiang An, Kun Wu, Yongle Zhao,*
* *Weimo Deng, Zimin Ran, Yumeng Wang, Ziyong Feng, Roy Miles, Ismail Elezi, Jiankang Deng*
- 163 Towards Privacy-preserved Pre-training of Remote Sensing Foundation Models with Federated Mutual-guidance Learning, *Jieyi Tan, Chengwei Zhang, Bo Dang, Yansheng Li*
- 164 EFTViT: Efficient Federated Training of Vision Transformers with Masked Images on Resource-Constrained Clients, *Meihan Wu, Tao Chang, Cui Miao, Jie Zhou, Chun Li, Xiangyu Xu, Ming Li, Xiaodong Wang*
- 165 HOLa: Zero-Shot HOI Detection with Low-Rank Decomposed VLM Feature Adaptation, *Qinqian Lei, Bo Wang, Robby T. Tan*
- 166 Diagnosing Pretrained Models for Out-of-distribution Detection, *Haipeng Xiong, Kai Xu, Angela Yao*
- 167 ODP-Bench: Benchmarking Out-of-Distribution Performance Prediction, *Han Yu, Kehan Li, Dongbai Li, Yue He, Xingxuan Zhang, Peng Cui*
- 168 R1-VL: Learning to Reason with Multimodal Large Language Models via Step-wise Group Relative Policy Optimization, *Jingyi Zhang, Jiaxing Huang, Huanjin Yao, Shunyu Liu, Xikun Zhang, Shijian Lu, Dacheng Tao*

- 169 Supervised Exploratory Learning for Long-Tailed Visual Recognition, *Zhongquan Jian, Yanhao Chen, Yancheng Wang, Junfeng Yao, Meihong Wang, Qingqiang Wu*
- 170 Synergistic Prompting for Robust Visual Recognition with Missing Modalities, *Zhihui Zhang, Luanyuan Dai, Qika Lin, Yunfeng Diao, Guangyin Jin, Yufei Guo, Jing Zhang, Xiaoshuai Hao*
- 171 Hierarchical Cross-modal Prompt Learning for Vision-Language Models, *Hao Zheng, Shunzhi Yang, Zhuoxin He, Jinfeng Yang, Zhenhua Huang*
- 172 Rethinking Few Shot CLIP Benchmarks: A Critical Analysis in the Inductive Setting, *Alexey Kravets, Da Chen, Vinay P. Namboodiri*
- 173 Boosting Domain Generalized and Adaptive Detection with
* Diffusion Models: Fitness, Generalization, and Transferability, *Boyong He, Yuxiang Ji, Zhuoyue Tan, Liaoni Wu*
- 174 Entropy-Adaptive Diffusion Policy Optimization with Dynamic Step Alignment, *RenYe Yan, Jikang Cheng, Yaozhong Gan, Shikun Sun, You Wu, Yunfan Yang, Liang Ling, Jinlong Lin, Yeshuang Zhu, Jie Zhou, Jinchao Zhang, Junliang Xing, Yimao Cai, Ru Huang*
- 175 Target Bias Is All You Need: Zero-Shot Debiasing of Vision-Language
☆ Models with Bias Corpus, *Taeuk Jang, Hoin Jung, Xiaoqian Wang*
- 176 Joint Asymmetric Loss for Learning with Noisy Labels, *Jialiang Wang, Xianming Liu, Xiong Zhou, Gangfeng Hu, Deming Zhai, Junjun Jiang, Xiangyang Ji*
- 177 FG-OrLU: Towards Better Forgetting via Feature-Gradient Orthogonality for Incremental Unlearning, *Qian Feng, JiaHang Tu, Mintong Kang, Hanbin Zhao, Chao Zhang, Hui Qian*
- 178 Language-Driven Multi-Label Zero-Shot Learning with Semantic Granularity, *Shouwen Wang, Qian Wan, Junbin Gao, Zhigang Zeng*
- 179 ViT-Split: Unleashing the Power of Vision Foundation Models via Efficient Splitting Heads, *Yifan Li, Xin Li, Tianqin Li, Wenbin He, Yu Kong, Liu Ren*
- 180 Generalized Deep Multi-view Clustering via Causal Learning with Partially Aligned Cross-view Correspondence, *Xihong Yang, Siwei Wang, Jiaqi Jin, Fangdi Wang, Tianrui Liu, Yueming Jin, Xinwang Liu, En Zhu, Kunlun He*
- 181 ViT-EnsembleAttack: Augmenting Ensemble Models for Stronger Adversarial Transferability in Vision Transformers, *Hanwen Cao, Haobo Lu, Xiaosen Wang, Kun He*
- 182 Instruction-Oriented Preference Alignment for Enhancing Multi-Modal Comprehension Capability of MLLMs, *Zitian Wang, Yue Liao, Kang Rong, Fengyun Rao, Yibo Yang, Si Liu*
- 183 Progressive Homeostatic and Plastic Prompt Tuning for Audio-Visual Multi-Task Incremental Learning, *Jiong Yin, Liang Li, Jiehua Zhang, Yuhao Gao, Chenggang Yan, Xichun Sheng*
- 184 Visual-RFT: Visual Reinforcement Fine-Tuning, *Ziyu Liu, Zeyi Sun, Yuhang Zang, Xiaoyi Dong, Yuhang Cao, Haodong Duan, Dahua Lin, Jiaqi Wang*
- 185 Jailbreaking Multimodal Large Language Models via Shuffle Inconsistency, *Shiji Zhao, Ranjie Duan, Fengxiang Wang, Chi Chen, Caixin Kang, Shouwei Ruan, Jialing Tao, YueFeng Chen, Hui Xue, Xingxing Wei*
- 186 Large Learning Rates Simultaneously Achieve Robustness to Spurious Correlations and Compressibility, *Melih Barsbey, Lucas Prieto, Stefanos Zafeiriou, Tolga Birdal*
- 187 Dual-Rate Dynamic Teacher for Source-Free Domain Adaptive Object Detection, *Qi He, Xiao Wu, Jun-Yan He, Shuai Li*
- 188 Dynamic Multi-Layer Null Space Projection for Vision-Language
☆ Continual Learning, *Borui Kang, Lei Wang, Zhiping Wu, Tao Feng, Yawen Li, Yang Gao, Wenbin Li*
- 189 LLaVA-CoT: Let Vision Language Models Reason Step-by-Step, *Guowei Xu, Peng Jin, Ziang Wu, Hao Li, Yibing Song, Lichao Sun, Li Yuan*
- 190 DALIP: Distribution Alignment-based Language-Image Pre-Training for Domain-Specific Data, *Junjie Wu, Jiangtao Xie, Zhaolin Zhang, Qilong Wang, Qinghua Hu, Peihua Li, Sen Xu*
- 191 Semi-supervised Concept Bottleneck Models, *Lijie Hu, Tianhao Huang, Huanyi Xie, Xilin Gong, Chenyang Ren, Zhengyu Hu, Lu Yu, Ping Ma, Di Wang*
- 192 FRET: Feature Redundancy Elimination for Test Time Adaptation, *Linjing You, Jiabao Lu, Xiayuan Huang, Xiangli Nie*
- 193 Meta-Unlearning on Diffusion Models: Preventing Relearning Unlearned Concepts, *Hongcheng Gao, Tianyu Pang, Chao Du, Taihang Hu, Zhijie Deng, Min Lin*
- 194 A Unified Interpretation of Training-Time Out-of-Distribution
* Detection, *Xu Cheng, Xin Jiang, Zechao Li*
- 195 Coupling the Generator with Teacher for Effective Data-Free Knowledge Distillation, *Xu Chen, Yang Li, Yahong Han, Guangquan Xu, Jialie Shen*
- 196 FedMeNF: Privacy-Preserving Federated Meta-Learning for Neural Fields, *Junhyeog Yun, Minui Hong, Gunhee Kim*
- 197 Visual Modality Prompt for Adapting Vision-Language Object Detectors, *Heitor R. Medeiros, Atif Belal, Srikanth Muralidharan, Eric Granger, Marco Pedersoli*
- 198 What's in a Latent? Leveraging Diffusion Latent Space for Domain Generalization, *Xavier Thomas, Deepti Ghadiyaram*
- 199 Prototype Guided Backdoor Defense via Activation Space Manipulation, *Venkat Adithya Amula, Sunayana Samavedam, Saurabh Saini, Avani Gupta, P J Narayanan*
- 200 Analyzing Finetuning Representation Shift for Multimodal LLMs Steering, *Pegah Khayatan, Mustafa Shukor, Jayneel Parekh, Arnaud Dapogny, Matthieu Cord*
- 201 Efficient Unsupervised Shortcut Learning Detection and Mitigation in Transformers, *Lukas Kuhn, Sari Sadiya, Jörg Schlötterer, Florian Buettner, Christin Seifert, Gemma Roig*
- 202 Understanding Museum Exhibits using Vision-Language
☆ Reasoning, *Ada-Astrid Balauca, Sanjana Garai, Stefan Balauca, Rasesh Udayakumar Shetty, Naitik Agrawal, Dhwanil Subhashbhai Shah, Yuqian Fu, Xi Wang, Kristina Toutanova, Danda Pani Paudel, Luc Van Gool*
- 203 Looking in the Mirror: A Faithful Counterfactual Explanation Method
☆ for Interpreting Deep Image Classification Models, *Townim Chowdhury, Vu Minh Hieu Phan, Kewen Liao, Nanyu Dong, Minh-Son To, Anton van den Hengel, Johan W. Verjans, Zhibin Liao*
- 204 Improving Multimodal Learning via Imbalanced Learning, *Shicai Wei, Chunbo Luo, Yang Luo*
- 205 NAPPure: Adversarial Purification for Robust Image Classification under Non-Additive Perturbations, *Junjie Nan, Jianing Li, Wei Chen, Mingkun Zhang, Xueqi Cheng*
- 206 VSP: Diagnosing the Dual Challenges of Perception and Reasoning in Spatial Planning Tasks for MLLMs, *Qiucheng Wu, Handong Zhao, Michael Saxon, Trung Bui, William Yang Wang, Yang Zhang, Shiyu Chang*
- 207 Multi-Cache Enhanced Prototype Learning for Test-Time Generalization of Vision-Language Models, *Xinyu Chen, Haotian Zhai, Can Zhang, Xiupeng Shi, Ruirui Li*
- 208 AVAM: a Universal Training-free Adaptive Visual Anchoring Embedded into Multimodal Large Language Model for Multi-image Question Answering, *Kang Zeng, Guojin Zhong, Jintao Cheng, Jin Yuan, Zhiyong Li*
- 209 Unsupervised Visual Chain-of-Thought Reasoning via Preference Optimization, *Kesen Zhao, Beier Zhu, Qianru Sun, Hanwang Zhang*
- 210 Adversarial Robustness of Discriminative Self-Supervised Learning in Vision, *Ömer Veysel Çağatan, Ömer Faruk Tal, M. Emre Gursoy*
- 211 Hierarchical Variational Test-Time Prompt Generation for Zero-Shot Generalization, *Zhaoyang Wu, Fang Liu, Licheng Jiao, Shuo Li, Lingling Li, Xu Liu, Puhua Chen, Wenping Ma*
- 212 TRNAS: A Training-Free Robust Neural Architecture Search, *Yeming Yang, Qingling Zhu, Jianping Luo, Ka-Chun Wong, Qiuzhen Lin, Jianqiang Li*
- 213 Staining and Locking Computer Vision Models Without Retraining, *Oliver J. Sutton, Qinghua Zhou, George Leete, Alexander N. Gorban, Ivan Y. Tyukin*
- 214 Granular Concept Circuits: Toward a Fine-Grained Circuit Discovery for Concept Representations, *Dahee Kwon, Sehyun Lee, Jaesik Choi*
- 215 Federated Domain Generalization with Domain-specific Soft Prompts Generation, *Jianhan Wu, Xiaoyang Qu, Zhangcheng Huang, Jianzong Wang*
- 216 R1-Onevision: Advancing Generalized Multimodal Reasoning through Cross-Modal Formalization, *Yi Yang, Xiaoxuan He, Hongkun Pan,*

- Xiyan Jiang, Yan Deng, Xingtao Yang, Haoyu Lu, Dacheng Yin, Fengyun Rao, Minfeng Zhu, Bo Zhang, Wei Chen*
 217 Boosting Adversarial Transferability via Negative Hessian Trace Regularization, *Yunfei Long, Zilin Tian, Liguozhang, Huosheng Xu*
 218 The Inter-Intra Modal Measure: A Predictive Lens on Fine-Tuning Outcomes in Vision-Language Models, *Laura Niss, Kevin Vogt-Lowell, Theodoros Tsiligkaridis*
 219 What to Distill? Fast Knowledge Distillation with Adaptive Sampling, * *Byungchul Chae, Seonyeong Heo*
 220 Invisible Watermarks, Visible Gains: Steering Machine Unlearning with Bi-Level Watermarking Design, *Yuhao Sun, Yihua Zhang, Gaowen Liu, Hongtao Xie, Sijia Liu*
 221 Federated Continuous Category Discovery and Learning, *Lixu Wang, Chenxi Liu, Junfeng Guo, Qingqing Ye, Heng Huang, Haibo Hu, Wei Dong*
 222 Beyond Losses Reweighting: Empowering Multi-Task Learning via the * Generalization Perspective, *Hoang Phan, Lam Tran, Quyen Tran, Ngoc Tran, Tuan Truong, Qi Lei, Nhat Ho, Dinh Phung, Trung Le*
 223 Flexi-FSCIL: Adaptive Knowledge Retention for Breaking the Stability-Plasticity Dilemma in Few-Shot Class-Incremental Learning, *Wufei Xie, Yalin Wang, Chenliang Liu, Zhaohui Jiang, Xue Yang*
 224 FDPT: Federated Discrete Prompt Tuning for Black-Box Visual-Language Models, *Jiaqi Wu, Simin Chen, Jing Tang, Yuzhe Yang, Yiming Chen, Lixu Wang, Song Lin, Zehua Wang, Wei Chen, Zijian Tian*
 225 PROL: Rehearsal Free Continual Learning in Streaming Data via Prompt Online Learning, *M. Anwar Ma'sum, Mahardhika Pratama, Savitha Ramasamy, Lin Liu, Habibullah Habibullah, Ryszard Kowalczyk*
 226 Can Knowledge be Transferred from Unimodal to Multimodal? Investigating the Transitivity of Multimodal Knowledge Editing, *Lingyong Fang, Xinzhang Wang, Depeng Wang, Zongru Wu, Ya Guo, Huijia Zhu, Zhuosheng Zhang, Gongshen Liu*
 227 Scaling Laws for Native Multimodal Models, *Mustafa Shukor, Enrico Fini, Victor Guilherme Turrissi da Costa, Matthieu Cord, Joshua Susskind, Alaaeldin El-Nouby*
 228 Image as an IMU: Estimating Camera Motion from a Single * Motion-Blurred Image, *Jerred Chen, Ronald Clark*
 229 Visual-Oriented Fine-Grained Knowledge Editing for MultiModal Large Language Models, *Zhen Zeng, Leijiang Gu, Xun Yang, Zhangling Duan, Zenglin Shi, Meng Wang*
 230 Dynamic Multimodal Prototype Learning in Vision-Language ☆ Models, *Xingyu Zhu, Shuo Wang, Beier Zhu, Miaoge Li, Yunfan Li, Junfeng Fang, Zhicai Wang, Dongsheng Wang, Hanwang Zhang*
 231 Visual Intention Grounding for Egocentric Assistants, *Pengzhan Sun, Junbin Xiao, Tze Ho Elden Tse, Yicong Li, Arjun Akula, Angela Yao*
 232 Unleashing Vecset Diffusion Model for Fast Shape Generation, * *Zeqiang Lai, Yunfei Zhao, Zibo Zhao, Haolin Liu, Fuyun Wang, Huiwen Shi, Xianghui Yang, Qingxiang Lin, Jingwei Huang, Yuhong Liu, Jie Jiang, Chunhao Guo, Xiangyu Yue*
 233 INTER: Mitigating Hallucination in Large Vision-Language Models by Interaction Guidance Sampling, *Xin Dong, Shichao Dong, Jin Wang, Jing Huang, Li Zhou, Zenghui Sun, Lihua Jing, Jinsong Lan, Xiaoyong Zhu, Bo Zheng*
 234 VisRL: Intention-Driven Visual Perception via Reinforced Reasoning, *Zhangquan Chen, Xufang Luo, Dongsheng Li*
 235 Auto-Regressionally Generating Multi-View Consistent Images, *JiaKui Hu, Yuxiao Yang, Jialun Liu, Jinbo Wu, Chen Zhao, Yanye Lu*
 236 Towards Performance Consistency in Multi-Level Model Collaboration, *Qi Li, Runpeng Yu, Xinchao Wang*
 237 Debiased Teacher for Day-to-Night Domain Adaptive Object Detection, *Yiming Cui, Liang Li, Haibing Yin, Yuhao Gao, Yaoqi Sun, Chenggang Yan*
 238 From Easy to Hard: Progressive Active Learning Framework for Infrared Small Target Detection with Single Point Supervision, *Chuang Yu, Jinmiao Zhao, Yunpeng Liu, Sicheng Zhao, Yimian Dai, Xiangyu Yue*
 239 Disentangled World Models: Learning to Transfer Semantic Knowledge from Distracting Videos for Reinforcement Learning, *Qi Wang, Zhipeng Zhang, Baao Xie, Xin Jin, Yunbo Wang, Shiyu Wang, Liaomo Zheng, Xiaokang Yang, Wenjun Zeng*
 240 Stable-Sim2Real: Exploring Simulation of Real-Captured 3D Data with Two-Stage Depth Diffusion, *Mutian Xu, Chongjie Ye, Haolin Liu, Yushuang Wu, Jiahao Chang, Xiaoguang Han* ☆
 241 ULTHO: Ultra-Lightweight yet Efficient Hyperparameter Optimization in Deep Reinforcement Learning, *Mingqi Yuan, Bo Li, Xin Jin, Wenjun Zeng*
 242 Task-Aware Prompt Gradient Projection for Parameter-Efficient Tuning Federated Class-Incremental Learning, *Hualong Ke, Jiangming Shi, Yachao Zhang, Fangyong Wang, Yuan Xie, Yanyun Qu*
 243 Is Visual in-Context Learning for Compositional Medical Tasks within Reach?, *Simon Reiß, Zdravko Marinov, Alexander Jaus, Constantin Seibold, M. Saquib Sarfraz, Erik Rodner, Rainer Stiefelhagen*
 244 Effective Training Data Synthesis for Improving MLLM Chart Understanding, *Yuwei Yang, Zeyu Zhang, Yunzhong Hou, Zhuowan Li, Gaowen Liu, Ali Payani, Yuan-Sen Ting, Liang Zheng*
 245 Learnable Logit Adjustment for Imbalanced Semi-Supervised Learning under Class Distribution Mismatch, *Hyuck Lee, Taemin Park, Heeyoung Kim*
 246 Where, What, Why: Towards Explainable Driver Attention Prediction, * *Yuchen Zhou, Jiayu Tang, Xiaoyan Xiao, Yueyao Lin, Linkai Liu, Zipeng Guo, Hao Fei, Xiaobo Xia, Chao Gou*
 247 Heuristic-Induced Multimodal Risk Distribution Jailbreak Attack for Multimodal Large Language Models, *Teng Ma, Xiaojun Jia, Ranjie Duan, Xinfeng Li, Yihao Huang, Xiaoshuang Jia, Zhixuan Chu, Wenqi Ren*
 248 Hypergraph Clustering Network with Partial Attribute Imputation, *Qianqian Wang, Bowen Zhao, Zhengming Ding, Wei Feng, Quanxue Gao*
 249 VRM: Knowledge Distillation via Virtual Relation Matching, * *Weijia Zhang, Fei Xie, Weidong Cai, Chao Ma*
 250 Geminio: Language-Guided Gradient Inversion Attacks in Federated Learning, *Junjie Shan, Ziqi Zhao, Jialin Lu, Rui Zhang, Siu Ming Yiu, Ka-Ho Chow*
 251 From Holistic to Localized: Local Enhanced Adapters for Efficient Visual Instruction Fine-Tuning, *Pengkun Jiao, Bin Zhu, Jingjing Chen, Chong-Wah Ngo, Yu-Gang Jiang*
 252 Gaze-Language Alignment for Zero-Shot Prediction of Visual Search Targets from Human Gaze Scanpaths, *Sounak Mondal, Naveen Sendhilnathan, Ting Zhang, Yue Liu, Michael Proulx, Michael Louis Iuzzolino, Chuan Qin, Tanya R. Jonker*
 253 You Are Your Own Best Teacher: Achieving Centralized-level Performance in Federated Learning under Heterogeneous and Long-tailed Data, *Shanshan Yan, Zexi Li, Chao Wu, Meng Pang, Yang Lu, Yan Yan, Hanzhi Wang*
 254 Enhancing Few-Shot Vision-Language Classification with Large Multimodal Model Features, *Chancharik Mitra, Brandon Huang, Tianning Chai, Zhiqiu Lin, Assaf Arbelle, Rogerio Feris, Leonid Karlinsky, Trevor Darrell, Deva Ramanan, Roi Herzig*
 255 Safeguarding Vision-Language Models: Mitigating Vulnerabilities to Gaussian Noise in Perturbation-based Attacks, *Jiawei Wang, Yushen Zuo, Yuanjun Chai, Zhendong Liu, Yicheng Fu, Yichun Feng, Kin-Man Lam*
 256 Towards Adversarial Robustness via Debiased High-Confidence Logit Alignment, *Kejia Zhang, Juanjuan Weng, Shaozi Li, Zhiming Luo*
 257 Learning Visual Proxy for Compositional Zero-Shot Learning, *Shiyu Zhang, Cheng Yan, Yang Liu, Chenchen Jing, Lei Zhou, Wenjun Wang*
 258 Taming Flow Matching with Unbalanced Optimal Transport into Fast Pansharpening, *Zihan Cao, Yu Zhong, Liang-Jian Deng*
 259 Evidential Knowledge Distillation, *Liangyu Xiang, Junyu Gao, Changsheng Xu*
 260 Know "No" Better: A Data-Driven Approach for Enhancing Negation Awareness in CLIP, *Junsung Park, Jungbeom Lee, Jongyoon Song, Sangwon Yu, Dahuin Jung, Sungroh Yoon*
 261 Intervening in Black Box: Concept Bottleneck Model for Enhancing Human Neural Network Mutual Understanding, *Nuoye Xiong, Anqi Dong, Ning Wang, Cong Hua, Guangming Zhu, Lin Mei, Peiyi Shen, Liang Zhang*
 262 VALLR: Visual ASR Language Model for Lip Reading, *Marshall Thomas, Edward Fish, Richard Bowden*
 263 Robust Dataset Condensation using Supervised Contrastive Learning, *Nicole Hee-Yeon Kim, Hwanjun Song*

- 264 CaliMatch: Adaptive Calibration for Improving Safe Semi-supervised Learning, *Jinsoo Bae, Seoung Bum Kim, Hyungrok Do*
- 265 Hybrid-TTA: Continual Test-time Adaptation via Dynamic Domain Shift Detection, *Hyewon Park, Hyejin Park, Jueun Ko, Dongbo Min*
- 266 Resolving Token-Space Gradient Conflicts: Token Space Manipulation for Transformer-Based Multi-Task Learning, *Wooseong Jeong, Kuk-Jin Yoon*
- 267 DisCoPatch: Taming Adversarially-driven Batch Statistics for Improved Out-of-Distribution Detection, *Francisco Caetano, Christiaan Viviers, Luis A. Zavala-Mondragón, Peter H.N. De With, Fons van der Sommen*
- 268 PLAN: Proactive Low-Rank Allocation for Continual Learning, *Xiequn Wang, Zhan Zhuang, Yu Zhang*
- 269 Think Twice: Test-Time Reasoning for Robust CLIP Zero-Shot ☆ Classification, *Shenyu Lu, Zhaoying Pan, Xiaoqian Wang*
- 270 Differential-informed Sample Selection Accelerates Multimodal Contrastive Learning, *Zihua Zhao, Feng Hong, Mengxi Chen, Pengyi Chen, Benyuan Liu, Jiangchao Yao, Ya Zhang, Yanfeng Wang*
- 271 Dataset Distillation via Vision-Language Category Prototype, * *Yawen Zou, Guang Li, Duo Su, Zi Wang, Jun Yu, Chao Zhang*
- 272 Scaling and Taming Adversarial Training with Synthetic Data, *Juntao Wu, Xianting Huang, Yu Chen, Shuai Pang, Ke Wang*
- 273 A Constrained Optimization Approach for Gaussian Splatting * *From Coarsely-posed Images and Noisy Lidar Point Clouds, Jizong Peng, Tze Ho Elden Tse, Kai Xu, Wenchao Gao, Angela Yao*
- 274 Soft Separation and Distillation: Toward Global Uniformity in Federated Unsupervised Learning, *Hung-Chieh Fang, Hsuan-Tien Lin, Irwin King, Yifei Zhang*
- 275 Competitive Distillation: A Simple Learning Strategy for Improving Visual Classification, *Daqian Shi, Xiaolei Diao, Xu Chen, Cédric M John*
- 276 Mitigating Catastrophic Overfitting in Fast Adversarial Training via Label Information Elimination, *Chao Pan, Ke Tang, Qing Li, Xin Yao*
- 277 Fewer Denoising Steps or Cheaper Per-Step Inference: Towards ☆ Compute-Optimal Diffusion Model Deployment, *Zhenbang Du, Yonggan Fu, Lifu Wang, Jiayi Qian, Xiao Luo, Yingyan Celine Lin*
- 278 Chimera: Improving Generalist Model with Domain-Specific Experts, *Tianshuo Peng, Mingsheng Li, Jiakang Yuan, Hongbin Zhou, Renqiu Xia, Renrui Zhang, Lei Bai, Song Mao, Bin Wang, Aojun Zhou, Botian Shi, Tao Chen, Bo Zhang, Xiangyu Yue*
- 279 SAUCE: Selective Concept Unlearning in Vision-Language Models with Sparse Autoencoders, *Jiahui Geng, Qing Li*
- 280 Corvid: Improving Multimodal Large Language Models Towards * Chain-of-Thought Reasoning, *Jingjing Jiang, Chao Ma, Xurui Song, Hanwang Zhang, Jun Luo*
- 281 Any-SSR: How Recursive Least Squares Works in Continual Learning of Large Language Model, *Kai Tong, Kang Pan, Xiao Zhang, Erli Meng, Run He, Yawen Cui, Nuoyan Guo, Huiping Zhuang*
- 282 Reminiscence Attack on Residuals: Exploiting Approximate Machine Unlearning for Privacy, *Yaxin Xiao, Qingqing Ye, Li Hu, Huadi Zheng, Haibo Hu, Zi Liang, Haoyang Li, Yijie Jiao*
- 283 STEP-DETR: Advancing DETR-based Semi-Supervised Object Detection with Super Teacher and Pseudo-Label Guided Text Queries, *Tahira Shehzadi, Khurram Azeem Hashmi, Shalini Sarode, Didier Stricker, Muhammad Zeshan Afzal*
- 284 Adversarial Reconstruction Feedback for Robust Fine-grained Generalization, *Shijie Wang, Jian Shi, Haojie Li*
- 285 Not Only Vision: Evolve Visual Speech Recognition via Peripheral Information, *Zhaoxin Yuan, Shuang Yang, Shiguang Shan, Xilin Chen*
- 286 KOEnsAttack: Towards Efficient Data-Free Black-Box Adversarial Attacks via Knowledge-Orthogonalized Substitute Ensembles, *Chaoyong Yang, Jia-Li Yin, Bin Chen, Zhaozhe Hu, Xiaolei Liu, Wei Lin*
- 287 FedPall: Prototype-based Adversarial and Collaborative Learning for Federated Learning with Feature Drift, *Yong Zhang, Feng Liang, Guanghu Yuan, Min Yang, Chengming Li, Xiping Hu*
- 288 DuET: Dual Incremental Object Detection via Exemplar-Free Task Arithmetic, *Munish Monga, Vishal Chudasama, Pankaj Wasnik, Biplob Banerjee*
- 289 Dataset Ownership Verification for Pre-trained Masked Models, *Yuechen Xie, Jie Song, Yicheng Shan, Xiaoyan Zhang, Yuanyu Wan, Shengxuming Zhang, Jiarui Duan, Mingli Song*
- 290 Sparsity Outperforms Low-Rank Projections in Few-Shot Adaptation, *Nairouz Mrabah, Nicolas Richet, Ismail Ben Ayed, Eric Granger*
- 291 Feature Decomposition-Recomposition in Large Vision-Language Model for Few-Shot Class-Incremental Learning, *Zongyao Xue, Meina Kan, Shiguang Shan, Xilin Chen*
- 292 VLRLMBench: A Comprehensive and Challenging Benchmark for Vision-Language Reward Models, *Jiacheng Ruan, Wenzhen Yuan, Xian Gao, Ye Guo, Daoxin Zhang, Zhe Xu, Yao Hu, Ting Liu, Yuzhuo Fu*
- 293 ConstStyle: Robust Domain Generalization with Unified Style Transformation, *Nam Duong Tran, Nam Nguyen Phuong, Hieu H. Pham, Phi Le Nguyen, My T. Thai*
- 294 Adversarial Attention Perturbations for Large Object Detection Transformers, *Zachary Yahn, Selim Furkan Tekin, Fatih Ilhan, Sihao Hu, Tiansheng Huang, Yichang Xu, Margaret Loper, Ling Liu*
- 295 DISTIL: Data-Free Inversion of Suspicious Trojan Inputs via Latent Diffusion, *Hossein Mirzaei, Zeinab Taghavi, Sepehr Rezaee, Masoud Hadi, Moein Madadi, Mackenzie W. Mathis*
- 296 Guiding Diffusion-Based Articulated Object Generation by Partial * Point Cloud Alignment and Physical Plausibility Constraints, *Jens U. Kreber, Joerg Stueckler*
- 297 MissRAG: Addressing the Missing Modality Challenge in Multimodal ☆ Large Language Models, *Vittorio Pipoli, Alessia Saporita, Federico Bolelli, Marcella Cornia, Lorenzo Baraldi, Costantino Grana, Rita Cucchiara, Elisa Ficarra*
- 298 ONLY: One-Layer Intervention Sufficiently Mitigates Hallucinations in Large Vision-Language Models, *Zifu Wan, Ce Zhang, Silong Yong, Martin Q. Ma, Simon Stepputtis, Louis-Philippe Morency, Deva Ramanan, Katia Sycara, Yaqi Xie*
- 299 Adaptive Prompt Learning via Gaussian Outlier Synthesis for Out-of-distribution Detection, *Yongkang Zhang, Dongyu She, Zhong Zhou*
- 300 FinMMR: Make Financial Numerical Reasoning More Multimodal, Comprehensive, and Challenging, *Zichen Tang, Haihong E, Jiacheng Liu, Zhongjun Yang, Rongjin Li, Zihua Rong, Haoyang He, Zhuodi Hao, Xinyang Hu, Kun Ji, Ziyang Ma, Mengyuan Ji, Jun Zhang, Chenghao Ma, Qianhe Zheng, Yang Liu, Yiling Huang, Xinyi Hu, Qing Huang, Zijian Xie, Shiyao Peng*
- 301 Trust but Verify: Programmatic VLM Evaluation in the Wild, *Viraj Prabhu, Senthil Purushwalkam, An Yan, Caiming Xiong, Ran Xu*
- 302 FixTalk: Taming Identity Leakage for High-Quality Talking Head Generation in Extreme Cases, *Shuai Tan, Bill Gong, Bin Ji, Ye Pan*
- 303 Uncalibrated Structure from Motion on a Sphere, * *Jonathan Ventura, Viktor Larsson, Fredrik Kahl*
- 304 FreeDNA: Endowing Domain Adaptation of Diffusion-Based Dense Prediction with Training-Free Domain Noise Alignment, *Hang Xu, Jie Huang, Linjiang Huang, Dong Li, Yidi Liu, Feng Zhao*
- 305 Progressive Distribution Bridging: Unsupervised Adaptation for Large-scale Pre-trained Models via Adaptive Auxiliary Data, *Weinan He, Yixin Zhang, Zilei Wang*
- 306 ALLGCD: Leveraging All Unlabeled Data for Generalized Category Discovery, *Xinzi Cao, Ke Chen, Feidiao Yang, Xianwu Zheng, Yonghong Tian, Yutong Lu*
- 307 RALoc: Enhancing Outdoor LiDAR Localization via Rotation Awareness, *Yuyang Yang, Wen Li, Sheng Ao, Qingshan Xu, Shangshu Yu, Yu Guo, Yin Zhou, Siqi Shen, Cheng Wang*
- 308 External Knowledge Injection for CLIP-Based Class-Incremental Learning, *Da-Wei Zhou, Kai-Wen Li, Jingyi Ning, Han-Jia Ye, Lijun Zhang, De-Chuan Zhan*
- 309 Cooperative Pseudo Labeling for Unsupervised Federated Classification, *Kuangpu Guo, Lijun Sheng, Yongcan Yu, Jian Liang, Zilei Wang, Ran He*
- 310 Beyond Low-Rank Tuning: Model Prior-Guided Rank Allocation for Effective Transfer in Low-Data and Large-Gap Regimes, *Chuyan Zhang, Kefan Wang, Yun Gu*
- 311 SVIP: Semantically Contextualized Visual Patches for Zero-Shot Learning, *Zhi Chen, Zecheng Zhao, Jingcai Guo, Jingjing Li, Zi Huang*
- 312 DreamLayer: Simultaneous Multi-Layer Generation via Diffusion * Model, *Junjia Huang, Pengxiang Yan, Jinhang Cai, Jiyang Liu, Zhao Wang, Yitong Wang, Xinglong Wu, Guanbin Li*
- 313 AutoOcc: Automatic Open-Ended Semantic Occupancy Annotation * via Vision-Language Guided Gaussian Splatting, *Xiaoyu Zhou,*

- ☆ *Jingqi Wang, Yongtao Wang, Yufei Wei, Nan Dong, Ming-Hsuan Yang*
- 314 DWIM: Towards Tool-aware Visual Reasoning via Discrepancy-aware Workflow Generation & Instruct-Masking Tuning, *Fucaí Ke, Vijay Kumar B G, Xingjian Leng, Zhixi Cai, Zaid Khan, Weiqing Wang, Pari Delir Haghighi, Hamid Rezaatofghi, Manmohan Chandraker*
- 315 FW-Merging: Scaling Model Merging with Frank-Wolfe Optimization, *Hao Mark Chen, Shell Xu Hu, Wayne Luk, Timothy Hospedales, Hongxiang Fan*
- 316 Augmenting Moment Retrieval: Zero-Dependency Two-Stage Learning, *Zhengxuan Wei, Jiajin Tang, Sibeí Yang*
- 317 Rep-MTL: Unleashing the Power of Representation-level Task Saliency for Multi-Task Learning, *Zedong Wang, Siyuan Li, Dan Xu*
- 318 Enhancing Numerical Prediction of MLLMs with Soft Labeling, *Pei Wang, Zhaowei Cai, Hao Yang, Davide Modolo, Ashwin Swaminathan*
- 319 Transparent Vision: A Theory of Hierarchical Invariant Representations, *Shuren Qi, Yushu Zhang, Chao Wang, Zhihua Xia, Xiaochun Cao, Fenglei Fan*
- 320 Fuse Before Transfer: Knowledge Fusion for Heterogeneous Distillation, *Guopeng Li, Qiang Wang, Ke Yan, Shouhong Ding, Yuan Gao, Gui-Song Xia*
- 321 What You Have is What You Track: Adaptive and Robust Multimodal Tracking, *Yuedong Tan, Jiawei Shao, Eduard Zamfir, Ruanjun Li, Zhaochong An, Chao Ma, Danda Paudel, Luc Van Gool, Radu Timofte, Zongwei Wu*
- 322 Instruction-Grounded Visual Projectors for Continual Learning of Generative Vision-Language Models, *Hyundong Jin, Hyung Jin Chang, Eunwoo Kim*
- 323 OrderChain: Towards General Instruct-Tuning for Stimulating the Ordinal Understanding Ability of MLLM, *Jinhong Wang, Shuo Tong, Jian Liu, Dongqi Tang, Weiqiang Wang, Wentong Li, Hongxia Xu, Danny Z. Chen, Jintai Chen, Jian Wu*
- 324 DAP-MAE: Domain-Adaptive Point Cloud Masked Autoencoder for Effective Cross-Domain Learning, *Ziqi Gao, Qiufu Li, Linlin Shen*
- 325 SFUOD: Source-Free Unknown Object Detection, *Keon-Hee Park, Seun-An Choe, Gyeong-Moon Park*
- 326 Activation Subspaces for Out-of-Distribution Detection, *Bariş Zöngür, Robin Hesse, Stefan Roth*
- 327 A Recipe for Generating 3D Worlds from a Single Image, *Katja Schwarz, Denis Rozumny, Samuel Rota Bulò, Lorenzo Porzi, Peter Kotschieder*
- 328 On the Complexity-Faithfulness Trade-off of Gradient-Based Explanations, *Amir Mehrpanah, Matteo Gamba, Kevin Smith, Hossein Azizpour*
- 329 Oasis: One Image is All You Need for Multimodal Instruction Data Synthesis, *Letian Zhang, Quan Cui, Bingchen Zhao, Cheng Yang*
- 330 Learning to Inference Adaptively for Multimodal Large Language Models, *Zhuoyan Xu, Khoi Duc Nguyen, Preeti Mukherjee, Saurabh Bagchi, Somali Chatterji, Yingyu Liang, Yin Li*
- 331 Self-Reinforcing Prototype Evolution with Dual-Knowledge Cooperation for Semi-Supervised Lifelong Person Re-Identification, *Kunlun Xu, Fan Zhuo, Jiangmeng Li, Xu Zou, Jiahuan Zhou*
- 332 Hierarchical Divide-and-Conquer Grouping for Classification Adaptation of Pre-Trained Models, *Ziqian Lu, Yunlong Yu, Qinyue Tong, Jun Liu*
- 333 Knowledge Transfer from Interaction Learning, *Yilin Gao, Kangyi Chen, Zhongxing Peng, Hengjie Lu, Shugong Xu*
- 334 DOGR: Towards Versatile Visual Document Grounding and Referring, *Yinan Zhou, Yuxin Chen, Haokun Lin, Yichen Wu, Shuyu Yang, Zhongang Qi, Chen Ma, Li Zhu*
- 335 Lark: Low-Rank Updates After Knowledge Localization for Few-shot Class-Incremental Learning, *Jinxin Shi, Jiabao Zhao, Yifan Yang, Xingjiao Wu, Jiawen Li, Liang He*
- 336 Advancing Textual Prompt Learning with Anchored Attributes, *Zheng Li, Yibing Song, Ming-Ming Cheng, Xiang Li, Jian Yang*
- 337 Decoupled Multi-Predictor Optimization for Inference-Efficient Model Tuning, *Liwei Luo, Shuaitengyuan Li, Dongwei Ren, Qilong Wang, Pengfei Zhu, Qinghua Hu*
- 338 SHIFT: Smoothing Hallucinations by Information Flow Tuning for Multimodal Large Language Models, *Sudong Wang, Yunjian Zhang, Yao Zhu, Enci Liu, Jianing Li, Yanwei Liu, Xiangyang Ji*
- 339 Large Multi-modal Models Can Interpret Features in Large Multi-modal Models, *Kaichen Zhang, Yifei Shen, Bo Li, Ziwei Liu*
- 340 SCAN: Bootstrapping Contrastive Pre-training for Data Efficiency, *Yangyang Guo, Mohan Kankanhalli*
- 341 A Conditional Probability Framework for Compositional Zero-shot Learning, *Peng Wu, Qiuxia Lai, Hao Fang, Guo-Sen Xie, Yilong Yin, Xiankai Lu, Wenguan Wang*
- 342 Backdoor Self-Supervised Contrastive Learning by Noisy Alignment, *Tuo Chen, Jie Gui, Minjing Dong, Ju Jia, Lanting Fang, Jian Liu*
- 343 Lyra: An Efficient and Speech-Centric Framework for Omni-Cognition, *Zhisheng Zhong, Chengyao Wang, Yuqi Liu, Senqiao Yang, Longxiang Tang, Yuechen Zhang, Jingyao Li, Tianyuan Qu, Yanwei Li, Yukang Chen, Shaozuo Yu, Sitong Wu, Eric Lo, Shu Liu, Jiaya Jia*
- 344 p-MoD: Building Mixture-of-Depths MLLMs via Progressive Ratio Decay, *Jun Zhang, Desen Meng, Zhengming Zhang, Zhenpeng Huang, Tao Wu, Limin Wang*
- 345 Enhancing Transferability of Targeted Adversarial Examples via Inverse Target Gradient Competition and Spatial Distance Stretching, *Zhankai Li, Weiping Wang, Jie Li, Shigeng Zhang, Yunan Hu, Song Guo*
- 346 FedDiffRC: Unlocking the Potential of Text-to-Image Diffusion Models in Heterogeneous Federated Learning, *Huan Wang, Haoran Li, Huaming Chen, Jun Yan, Jiahua Shi, Jun Shen*
- 347 LoRA-FAIR: Federated LoRA Fine-Tuning with Aggregation and Initialization Refinement, *Jieming Bian, Lei Wang, Letian Zhang, Jie Xu*
- 348 Diversity-Enhanced Distribution Alignment for Dataset Distillation, *Hongcheng Li, Yucan Zhou, Xiaoyan Gu, Bo Li, Weiping Wang*
- 349 Category-Specific Selective Feature Enhancement for Long-Tailed Multi-Label Image Classification, *Ruiqi Du, Xu Tang, Xiangrong Zhang, Jingjing Ma*
- 350 Registration beyond Points: General Affine Subspace Alignment via Geodesic Distance on Grassmann Manifold, *Jaeho Shin, Hyeonjae Gil, Junwoo Jang, Maani Ghaffari, Ayoung Kim*
- 351 Mind the Gap: Preserving and Compensating for the Modality Gap in CLIP-Based Continual Learning, *Linlan Huang, Xusheng Cao, Haori Lu, Yifan Meng, Fei Yang, Xialei Liu*
- 352 Sibai: A Few-Shot Meta-Classifer for Poisoning Detection in Federated Learning, *Melanie Götz, Torsten Krauß, Alexandra Dmitrienko*
- 353 Generalization-Preserved Learning: Closing the Backdoor to Catastrophic Forgetting in Continual Deepfake Detection, *Xueyi Zhang, Peiyin Zhu, Chengwei Zhang, Zhiyuan Yan, Jikang Cheng, Mingrui Lao, Siqi Cai, Yanming Guo*
- 354 Dual Domain Control via Active Learning for Remote Sensing Domain Incremental Object Detection, *Jiachen Sun, De Cheng, Xi Yang, Nannan Wang*
- 355 Gradient Extrapolation for Debaised Representation Learning, *Ihab Asaad, Maha Shadaydeh, Joachim Denzler*
- 356 VisNumBench: Evaluating Number Sense of Multimodal Large Language Models, *Tengjin Weng, Jingyi Wang, Wenhao Jiang, Zhong Ming*
- 357 FedAGC: Federated Continual Learning with Asymmetric Gradient Correction, *Chengchao Zhang, Fanhua Shang, Hongying Liu, Liang Wan, Wei Feng*
- 358 BUFFER-X: Towards Zero-Shot Point Cloud Registration in Diverse Scenes, *Minkyun Seo, Hyungtae Lim, Kanghee Lee, Luca Carlone, Jaesik Park*
- 359 FREE-Merging: Fourier Transform for Efficient Model Merging, *Shenghe Zheng, Hongzhi Wang*
- 360 RANKCLIP: Ranking-Consistent Language-Image Pretraining, *Yiming Zhang, Zhuokai Zhao, Zhaorun Chen, Zhili Feng, Zenghui Ding, Yining Sun*
- 361 Enhancing Adversarial Transferability by Balancing Exploration and Exploitation with Gradient-Guided Sampling, *Zenghao Niu, Weicheng Xie, Siyang Song, Zitong Yu, Feng Liu, Linlin Shen*
- 362 An Efficient Post-hoc Framework for Reducing Task Discrepancy of Text Encoders for Composed Image Retrieval, *Jaeseok Byun, Seokhyeon Jeong, Wonjae Kim, Sanghyuk Chun, Taesup Moon*
- 363 Robin3D: Improving 3D Large Language Model via Robust Instruction Tuning, *Weitai Kang, Haifeng Huang, Yuzhang Shang, Mubarak Shah, Yan Yan*

- 364 FLSeg: Enhancing Privacy and Robustness in Federated Learning under Heterogeneous Data via Model Segmentation, *Zichun Su, Zhi Lu, Yutong Wu, Renfei Shen, Songfeng Lu*
- 365 ZIUM: Zero-Shot Intent-Aware Adversarial Attack on Unlearned Models, *Hyun Jun Yook, Ga San Jhun, Jae Hyun Cho, Min Jeon, Donghyun Kim, Tae Hyung Kim, Youn Kyu Lee*
- 366 Federated Prompt-Tuning with Heterogeneous and Incomplete Multimodal Client Data, *Thu Hang Phung, Duong M. Nguyen, Thanh Trung Huynh, Quoc Viet Hung Nguyen, Trong Nghia Hoang, Phi Le Nguyen*
- 367 Learning Interpretable Queries for Explainable Image Classification with Information Pursuit, *Stefan Kolek, Aditya Chattopadhyay, Kwan Ho Ryan Chan, Hector Andrade-Loarca, Gitta Kutyniok, René Vidal*
- 368 ACAM-KD: Adaptive and Cooperative Attention Masking for Knowledge Distillation, *Qizhen Lan, Qing Tian*
- 369 Find a Scapegoat: Poisoning Membership Inference Attack and Defense to Federated Learning, *Wenjin Mo, Zhiyuan Li, Minghong Fang, Mingwei Fang*
- 370 OpenVision: A Fully-Open, Cost-Effective Family of Advanced Vision Encoders for Multimodal Learning, *Xianhang Li, Yanqing Liu, Haoqin Tu, Cihang Xie*
- 371 Integrating Visual Interpretation and Linguistic Reasoning for Geometric Problem Solving, *Zixian Guo, Ming Liu, Qilong Wang, Zhilong Ji, Jinfeng Bai, Lei Zhang, Wangmeng Zuo*
- 372 SAFER: Sharpness Aware layer-selective Finetuning for Enhanced Robustness in vision transformers, *Bhavna Gopal, Huanrui Yang, Mark Horton, Yiran Chen*
- 373 Adding Additional Control to One-Step Diffusion with Joint Distribution Matching, *Yihong Luo, Tianyang Hu, Yifan Song, Jiacheng Sun, Zhenguo Li, Jing Tang*
- 374 Semantic Equitable Clustering: A Simple and Effective Strategy for Clustering Vision Tokens, *Qihang Fan, Huaibo Huang, Mingrui Chen, Ran He*
- 375 SPD: Shallow Backdoor Protecting Deep Backdoor Against Backdoor Detection, *Shunjie Yuan, Xinghua Li, Xuelin Cao, Haiyan Zhang, Mengyao Zhu, Robert H. Deng*
- 376 To Label or Not to Label: PALM – A Predictive Model for Evaluating Sample Efficiency in Active Learning Models, *Julia Machnio, Mads Nielsen, Mostafa Mehdipour Ghazi*
- 377 Inference-Time Diffusion Model Distillation, *Geon Yeong Park, Sang Wan Lee, Jong Chul Ye*
- 378 G2D: Boosting Multimodal Learning with Gradient-Guided Distillation, *Mohammed Rakib, Arunkumar Bagavathi*
- 379 Personalized Federated Learning under Local Supervision, *Qiqi Liu, Jiaqiang Li, Yuchen Liu, Yaochu Jin, Lingjuan Lyu, Xiaohu Wu, Han Yu*
- 380 Token Activation Map to Visually Explain Multimodal LLMs, *Yi Li, Hualiang Wang, Xinpeng Ding, Haonan Wang, Xiaomeng Li*
- 381 Differentiable Room Acoustic Rendering with Multi-View Vision Priors, *Derong Jin, Ruohan Gao*
- 382 Noise-Modeled Diffusion Models for Low-Light Spike Image Restoration, *Ruonan Liu, Lin Zhu, Xijie Xiang, Lizhi Wang, Hua Huang*
- 383 One Perturbation is Enough: On Generating Universal Adversarial Perturbations against Vision-Language Pre-training Models, *Hao Fang, Jiawei Kong, Wenbo Yu, Bin Chen, Jiawei Li, Hao Wu, Shu-Tao Xia, Ke Xu*
- 384 Why LVLs Are More Prone to Hallucinations in Longer Responses: The Role of Context, *Ge Zheng, Jiaye Qian, Jiajin Tang, Sibe Yang*
- 385 Intra-view and Inter-view Correlation Guided Multi-view Novel Class Discovery, *Xinhang Wan, Jiyuan Liu, Qian Qu, Suyuan Liu, Chuyu Zhang, Fangdi Wang, Xinwang Liu, En Zhu, Kunlun He*
- 386 Prototype-based Contrastive Learning with Stage-wise Progressive Augmentation for Self-Supervised Fine-Grained Learning, *Baofeng Tan, Xiu-Shen Wei, Lin Zhao*
- 387 Radiant Foam: Real-Time Differentiable Ray Tracing, *Shrisudhan Govindarajan, Daniel Rebain, Kwang Moo Yi, Andrea Tagliasacchi*
- 388 COSTARR: Consolidated Open Set Technique with Attenuation for Robust Recognition, *Ryan Rabinowitz, Steve Cruz, Walter Scheirer, Terrance E. Boulton*
- 389 ALOcc: Adaptive Lifting-Based 3D Semantic Occupancy and Cost Volume-Based Flow Predictions, *Dubing Chen, Jin Fang, Wencheng Han, Xinjing Cheng, Junbo Yin, Chengzhong Xu, Fahad Shahbaz Khan, Jianbing Shen*
- 390 Information Density Principle for MLLM Benchmarks, *Chunyi Li, Xiaozhe Li, Zicheng Zhang, Yuan Tian, Ziheng Jia, Xiaohong Liu, Xiongkuo Min, Jia Wang, Haodong Duan, Kai Chen, Guangtao Zhai*
- 391 Perspective-Aware Teaching: Adapting Knowledge for Heterogeneous Distillation, *Jhe-Hao Lin, Yi Yao, Chan-Feng Hsu, Hong-Xia Xie, Hong-Han Shuai, Wen-Huang Cheng*
- 392 Is Meta-Learning Out? Rethinking Unsupervised Few-Shot Classification with Limited Entropy, *Yunchuan Guan, Yu Liu, Ke Zhou, Zhiqi Shen, Jenq-Neng Hwang, Serge Belongie, Lei Li*
- 393 GenieBlue: Integrating both Linguistic and Multimodal Capabilities for Large Language Models on Mobile Devices, *Xudong Lu, Yinghao Chen, Renshou Wu, Haozhao Gao, Xi Chen, Xue Yang, Xiangyu Zhao, Aojun Zhou, Fangyuan Li, Yafei Wen, Xiaoxin Chen, Shuai Ren, Hongsheng Li*
- 394 Learning to Unlearn while Retaining: Combating Gradient Conflicts in Machine Unlearning, *Gaurav Patel, Qiang Qiu*
- 395 Neural Architecture Search Driven by Locally Guided Diffusion for Personalized Federated Learning, *Peng Liao, Xilu Wang, Yaochu Jin, Wenli Du, Han Hu*
- 396 Robust Multi-View Learning via Representation Fusion of Sample-Level Attention and Alignment of Simulated Perturbation, *Jie Xu, Na Zhao, Gang Niu, Masashi Sugiyama, Xiaofeng Zhu*
- 397 PAN-Crafter: Learning Modality-Consistent Alignment for PAN-Sharpener, *Jeonghyeok Do, Sungpyo Kim, Geunhyuk Youk, Jaehyup Lee, Munchul Kim*
- 398 Attention to the Burstiness in Visual Prompt Tuning!, *Yuzhu Wang, Manni Duan, Shu Kong*
- 399 DIP: Unsupervised Dense In-Context Post-training of Visual Representations, *Sophia Sirko-Galouchenko, Spyros Gidaris, Antonin Vobecky, Andrei Bursuc, Nicolas Thome*
- 400 Zero-Shot Vision Encoder Grafting via LLM Surrogates, *Kaiyu Yue, Vasu Singla, Menglin Jia, John Kirchenbauer, Rifaa Qadri, Zikui Cai, Abhinav Bhatele, Furong Huang, Tom Goldstein*
- 401 Long-Tailed Classification with Multi-Granularity Semantics, *Yuting Liu, Liu Yang, Yu Wang*
- 402 LLaVA-3D: A Simple yet Effective Pathway to Empowering LMMs with 3D Capabilities, *Chenming Zhu, Tai Wang, Wenwei Zhang, Jiangmiao Pang, Xihui Liu*
- 404 ReTracker: Exploring Image Matching for Robust Online Any Point Tracking, *Dongli Tan, Xingyi He, Sida Peng, Yiqing Gong, Xing Zhu, Jiaming Sun, Ruizhen Hu, Yujun Shen, Hujun Bao, Xiaowei Zhou*
- 405 HIS-GPT: Towards 3D Human-In-Scene Multimodal Understanding, *Jiahe Zhao, Ruibing Hou, Zefei Tian, Hong Chang, Shiguang Shan*
- 406 Backdoor Attacks on Neural Networks via One-Bit Flip, *Xiang Li, Lannan Luo, Qiang Zeng*
- 407 A Linear N-Point Solver for Structure and Motion from Asynchronous Tracks, *Hang Su, Yunlong Feng, Daniel Gehrig, Panfeng Jiang, Ling Gao, Xavier Lagorce, Laurent Kneip*
- 408 Long-LRM: Long-sequence Large Reconstruction Model for Wide-coverage Gaussian Splats, *Chen Ziwen, Hao Tan, Kai Zhang, Sai Bi, Fujun Luan, Yicong Hong, Li Fuxin, Zexiang Xu*
- 409 TR-PTS: Task-Relevant Parameter and Token Selection for Efficient Tuning, *Siqi Luo, Haoran Yang, Yi Xin, Mingyang Yi, Guangyang Wu, Guangtao Zhai, Xiaohong Liu*
- 410 Exploring The Visual Feature Space for Multimodal Neural Decoding, *Weihao Xia, Cengiz Oztireli*
- 411 Joint Diffusion Models in Continual Learning, *Paweł Skiers, Kamil Deja*
- 412 TurboTrain: Towards Efficient and Balanced Multi-Task Learning for Multi-Agent Perception and Prediction, *Zewei Zhou, Seth Z. Zhao, Tianhui Cai, Zhiyu Huang, Bolei Zhou, Jiaqi Ma*
- 413 VITAL: More Understandable Feature Visualization through Distribution Alignment and Relevant Information Flow, *Ada Görgün, Bernt Schiele, Jonas Fischer*
- 414 Overcoming Dual Drift for Continual Long-Tailed Visual Question

- Answering, *Feifei Zhang, Zhihao Wang, Xi Zhang, Changsheng Xu*
- 415 ToolVQA: A Dataset for Multi-step Reasoning VQA with External Tools, *Shaofeng Yin, Ting Lei, Yang Liu*
- 416 Continual Adaptation: Environment-Conditional Parameter Generation for Object Detection in Dynamic Scenarios, *Deng Li, Aming Wu, Yang Li, Yaowei Wang, Yahong Han*
- 417 SMP-Attack: Boosting the Transferability of Feature Importance-based Adversarial Attack with Semantics-aware Multi-granularity Patchout, *Wen Yang, Guodong Liu, Di Ming*
- 418 Towards Comprehensive Lecture Slides Understanding: Large-scale Dataset and Effective Method, *Enming Zhang, Yuzhe Li, Yuliang Liu, Yingying Zhu, Xiang Bai*
- 419 Backdoor Mitigation by Distance-Driven Detoxification, ☆☆ *Shaokui Wei, Jiayin Liu, Hongyuan Zha*
- 420 Zeroth-Order Fine-Tuning of LLMs in Random Subspaces, *Ziming Yu, Pan Zhou, Sike Wang, Jia Li, Mi Tian, Hua Huang*
- 421 Gradient Decomposition and Alignment for Incremental Object Detection, *Wenlong Luo, Shizhou Zhang, De Cheng, Yinghui Xing, Guoqiang Liang, Peng Wang, Yanning Zhang*
- 422 Synthesizing Near-Boundary OOD Samples for Out-of-Distribution Detection, ☆ *Jinglun Li, Kaixun Jiang, Zhaoyu Chen, Bo Lin, Yao Tang, Weifeng Ge, Wenqiang Zhang*
- 423 CAVIS: Context-Aware Video Instance Segmentation, *Seunghun Lee, Jiwan Seo, Kiljoon Han, Minwoo Choi, Sunghoon Im*
- 424 Unlearning the Noisy Correspondence Makes CLIP More Robust, *Haochen Han, Alex Jinpeng Wang, Peijun Ye, Fangming Liu*
- 425 FEVER-ODD: Free Energy Vulnerability Elimination for Robust Out-of-Distribution Detection, ☆ *Brian K.S. Isaac-Medina, Mauricio Che, Yona Falinie A. Gaus, Samet Akcay, Toby P. Breckon*
- 426 Local Dense Logit Relations for Enhanced Knowledge Distillation, *Liuchi Xu, Kang Liu, Jinshuai Liu, Lu Wang, Lisheng Xu, Jun Cheng*
- 427 Controllable Feature Whitening for Hyperparameter-Free Bias Mitigation, *Yoosin Cho, Hanbyel Cho, Janghyeon Lee, HyeongGwon Hong, Jaesung Ahn, Junmo Kim*
- 428 Differentially Private Fine-Tuning of Diffusion Models, *Yu-Lin Tsai, Yizhe Li, Chia-Mu Yu, Xuebin Ren, Po-Yu Chen, Zekai Chen, Francois Buet-Golfouse*
- 429 FedXDS: Leveraging Model Attribution Methods to counteract Data Heterogeneity in Federated Learning, *Maximilian Andreas Hoefer, Karsten Mueller, Wojciech Samek*
- 430 Towards Effective Foundation Model Adaptation for Extreme Cross-Domain Few-Shot Learning, *Fei Zhou, Peng Wang, Lei Zhang, Wei Wei, Chen Ding, Guosheng Lin, Yanning Zhang*
- 431 Consensus-Driven Active Model Selection, ☆ *Justin Kay, Grant Van Horn, Subhransu Maji, Daniel Sheldon, Sara Beery*
- 432 Adversarial Purification via Super-Resolution and Diffusion, *Mincheol Park, Cheonjun Park, Seungseop Lim, Mijin Koo, Hyunwuk Lee, Won Woo Ro, Suhyun Kim*
- 433 FedWSQ: Efficient Federated Learning with Weight Standardization and Distribution-Aware Non-Uniform Quantization, *Seung-Wook Kim, Seongyeol Kim, Jiah Kim, Seowon Ji, Se-Ho Lee*
- 434 CMAD: Correlation-Aware and Modalities-Aware Distillation for Multimodal Sentiment Analysis with Missing Modalities, *Yan Zhuang, Minhao Liu, Wei Bai, Yanru Zhang, Xiaoyue Zhang, Jiawen Deng, Fuji Ren*
- 435 SimpleVQA: Multimodal Factuality Evaluation for Multimodal Large Language Models, *Xianfu Cheng, Wei Zhang, Shiwei Zhang, Jian Yang, Xiangyuan Guan, Xianjie Wu, Xiang Li, Ge Zhang, Jiaheng Liu, Yuying Mai, Yutao Zeng, Zhoufutu Wen, Ke Jin, Baorui Wang, Weixiao Zhou, Yunhong Lu, Hangyuan Ji, Tongliang Li, Wenhao Huang, Zhoujun Li*
- 436 2.5 Years in Class: A Multimodal Textbook for Vision-Language Pretraining, ☆ *Wenqi Zhang, Hang Zhang, Xin Li, Jiahuo Sun, Yongliang Shen, Weiming Lu, Deli Zhao, Yueting Zhuang, Lidong Bing*
- 437 Revelio: Interpreting and leveraging semantic information in diffusion models, *Dahye Kim, Xavier Thomas, Deepti Ghadiyaram*
- 438 CLIP-GS: Unifying Vision-Language Representation with 3D Gaussian Splatting, *Siyu Jiao, Haoye Dong, Yuyang Yin, Zequn Jie, Yinlong Qian, Yao Zhao, Humphrey Shi, Yunchao Wei*
- 439 ProJUDGE: A Multi-Modal Multi-Discipline Benchmark and Instruction-Tuning Dataset for MLLM-based Process Judges, *Jiaxin Ai, Pengfei Zhou, Zhaopan Xu, Ming Li, Fanrui Zhang, Zizhen Li, Jianwen Sun, Yukang Feng, Baojin Huang, Zhongyuan Wang, Kaipeng Zhang*
- 440 Failure Cases Are Better Learned But Boundary Says Sorry: Facilitating Smooth Perception Change for Accuracy-Robustness Trade-Off in Adversarial Training, *Yanyun Wang, Li Liu*
- 441 VehicleMAE: View-asymmetry Mutual Learning for Vehicle Re-identification Pre-training via Masked AutoEncoders, *Qi Wang, Zeyu Zhang, Dong Wang, Di Gai, Xin Xiong, Jiyang Xu, Ruihua Zhou*
- 442 SplatTalk: 3D VQA with Gaussian Splatting, *Anh Thai, Songyou Peng, Kyle Genova, Leonidas Guibas, Thomas Funkhouser*
- 443 CaO2: Rectifying Inconsistencies in Diffusion-Based Dataset Distillation, *Haoxuan Wang, Zhenghao Zhao, Junyi Wu, Yuzhang Shang, Gaowen Liu, Yan Yan*
- 444 Taming the Untamed: Graph-Based Knowledge Retrieval and Reasoning for MLLMs to Conquer the Unknown, *Bowen Wang, Zhouqiang Jiang, Yasuaki Susumu, Shotaro Miwa, Tianwei Chen, Yuta Nakashima*
- 445 MambaML: Exploring State Space Models for Multi-Label Image Classification, *Xuelin Zhu, Jian Liu, Jiuxin Cao, Bing Wang*
- 446 MUNBA: Machine Unlearning via Nash Bargaining, *Jing Wu, Mehrtaash Harandi*
- 447 SMSTracker: Tri-path Score Mask Sigma Fusion for Multi-Modal Tracking, *Sixian Chan, Zedong Li, Wenhao Li, Shijian Lu, Chunhua Shen, Xiaoqin Zhang*
- 448 Auxiliary Prompt Tuning of Vision-Language Models for Few-Shot Out-of-Distribution Detection, *Wenjun Miao, Guansong Pang, Zihan Wang, Jin Zheng, Xiao Bai*
- 449 Enhancing Transformers Through Conditioned Embedded Tokens, *Hemanth Saratchandran, Simon Lucey*
- 450 Improved Noise Schedule for Diffusion Training, *Tiankai Hang, Shuyang Gu, Jianmin Bao, Fangyun Wei, Dong Chen, Xin Geng, Baining Guo*
- 451 How Do Multimodal Large Language Models Handle Complex Multimodal Reasoning? Placing Them in An Extensible Escape Game, *Ziyue Wang, Yurui Dong, Fuwen Luo, Minyuan Ruan, Zhili Cheng, Chi Chen, Peng Li, Yang Liu*
- 452 One Encoder to Rule them All: Representation Learning for Model-free Visual Reinforcement Learning using Fourier Neural Operators, *Parag Dutta, Mohd Ayyoob, Shalabh Bhatnagar, Ambedkar Dukkipati*
- 453 LIFT: Latent Implicit Functions for Task- and Data-Agnostic Encoding, *Amirhossein Kazerouni, Soroush Mehraban, Michael Brudno, Babak Taati*
- 454 Improving Noise Efficiency in Privacy-preserving Dataset Distillation, *Runkai Zheng, Vishnu Asutosh Dasu, Yinong Oliver Wang, Haohan Wang, Fernando De La Torre*
- 455 Boosting MLLM Reasoning with Text-Debiased Hint-GRPO, *Qihan Huang, Weilong Dai, Jinlong Liu, Wanggui He, Hao Jiang, Mingli Song, Jingyuan Chen, Chang Yao, Jie Song*
- 456 Towards Real Unsupervised Anomaly Detection Via Confident Meta-Learning, *Muhammad Aqeel, Shakiba Sharifi, Marco Cristani, Francesco Setti*
- 457 RIPE: Reinforcement Learning on Unlabeled Image Pairs for Robust Keypoint Extraction, *Johannes Künzel, Anna Hilsmann, Peter Eisert*
- 458 Efficient Adaptation of Pre-trained Vision Transformer underpinned by Approximately Orthogonal Fine-Tuning Strategy, *Yiting Yang, Hao Luo, Yuan Sun, Qingsen Yan, Haokui Zhang, Wei Dong, Guoqing Wang, Peng Wang, Yang Yang, Hengtao Shen*
- 459 DiffRefine: Diffusion-based Proposal Specific Point Cloud Densification for Cross-Domain Object Detection, ☆ *Sangyun Shin, Yuhang He, Xinyu Hou, Samuel Hodgson, Andrew Markham, Niki Trigoni*
- 460 On the Robustness Tradeoff in Fine-Tuning, *Kunyang Li, Jean-Charles Noiret Ferrand, Ryan Sheatsley, Blaine Hoak, Yohan Beugin, Eric Pauley, Patrick McDaniel*
- 461 Understanding Flatness in Generative Models: Its Role and Benefits, ☆ *Taehwan Lee, Kyeongkook Seo, Jaejun Yoo, Sung Whan Yoon*

11:30 - 13:30 Demos 1 (Exhibit Hall I)

- Spatially-Varying Autofocus, *Yingsi Qin, Aswin C Sankaranarayanan, Matthew O'Toole*
- Measuring the Performance of 6 DoF Object Pose Estimation for Robotic Bin Picking, *Kamel Saidi*
- Multi-turn Consistent Image Editing, *Zijun Zhou*
- AstroLoc: Robust Space to Ground Image Localizer, *Gabriele Berton, Alex Stoken, Carlo Masone*

13:30 - 14:45 Oral Session 2A: View Synthesis and Scene

(Exhibit Hall III)

Session Chairs: Hongdong Li, Federico Tombari

🏆 - Award candidate paper ☆ - Outstanding Reviewer

- 1 RayZer: A Self-supervised Large View Synthesis Model, *Hanwen Jiang, Hao Tan, Peng Wang, Haian Jin, Yue Zhao, Sai Bi, Kai Zhang, Fujun Luan, Kalyan Sunkavalli, Qixing Huang, Georgios Pavlakos*
- 2 EVER: Exact Volumetric Ellipsoid Rendering for Real-time View Synthesis, *Alexander Mai, Peter Hedman, George Kopanas, Dor Verbin, David Futschik, Qiangeng Xu, Falko Kuester, Jonathan T. Barron, Yinda Zhang*
- 3 Self-Ensembling Gaussian Splatting for Few-Shot Novel View Synthesis, *Chen Zhao, Xuan Wang, Tong Zhang, Saqib Javed, Mathieu Salzmann*
- 4 Back on Track: Bundle Adjustment for Dynamic Scene Reconstruction, *Weirong Chen, Ganlin Zhang, Felix Wimbauer, Rui Wang, Nikita Araslanov, Andrea Vedaldi, Daniel Cremers*
- 5 SceneSplat: Gaussian Splatting-based Scene Understanding with Vision-Language Pretraining, *Yue Li, Qi Ma, Runyi Yang, Huapeng Li, Mengjiao Ma, Bin Ren, Nikola Popovic, Nicu Sebe, Ender Konukoglu, Theo Gevers, Luc Van Gool, Martin R. Oswald, Danda Pani Paudel*

13:30 - 14:45 Oral Session 2B: Efficient Learning

(Kalakaua Ballroom)

Session Chairs: Jianfei Cai, Sara Beery

- 1 Variance-Based Pruning for Accelerating and Compressing Trained Networks, *Uranik Berisha, Jens Mehnert, Alexandru Paul Condurache*
- 2 Importance-Based Token Merging for Efficient Image and Video Generation, *Haoyu Wu, Jingyi Xu, Hieu Le, Dimitris Samaras*
- 3 Knowledge Distillation for Learned Image Compression, *Yunuo Chen, Zezheng Lyu, Bing He, Ning Cao, Gang Chen, Guo Lu, Wenjun Zhang*
- 4 Distilling Diffusion Models to Efficient 3D LiDAR Scene Completion, *Shengyuan Zhang, An Zhao, Ling Yang, Zejian Li, Chenye Meng, Haoran Xu, Tianrun Chen, AnYang Wei, Perry Pengyun Gu, Lingyun Sun*
- 5 Heavy Labels Out! Dataset Distillation with Label Space Lightening, *Ruonan Yu, Songhua Liu, Zigeng Chen, Jingwen Ye, Xinchao Wang*

14:30 - 15:00 Poster Setup (Exhibit Hall I)**15:00 - 17:00 Poster Session 2 & Exhibit Hall with Coffee Break**

(Exhibit Hall I)

* - Highlight paper 🏆 - Award candidate paper

● - Oral Paper ☆ - Outstanding Reviewer

- 1 HAMSt3R: Human-Aware Multi-view Stereo 3D Reconstruction, *Sara Rojas, Matthieu Armando, Bernard Ghanem, Philippe Weinzaepfel, Vincent Leroy, Grégory Rogez*
- 2 Time-Aware Auto White Balance in Mobile Photography, *Mahmoud Afifi, Luxi Zhao, Abhijith Punnappurath, Mohamed A. Abdelsalam, Ran Zhang, Michael S. Brown*
- 3 Detect Anything 3D in the Wild, *Hanxue Zhang, Haoran Jiang, Qingsong Yao, Yanan Sun, Renrui Zhang, Hao Zhao, Hongyang Li, Hongzi Zhu, Zetong Yang*
- 4 VolumetricSMPL: A Neural Volumetric Body Model for Efficient Interactions, Contacts, and Collisions, *Marko Mihajlovic, Siwei Zhang, Gen Li, Kaifeng Zhao, Lea Muller, Siyu Tang*
- 5 Contact-Aware Refinement of Human Pose Pseudo-Ground Truth via Bioimpedance Sensing, *Maria-Paola Forte, Nikos Athanasiou, Giulia Ballardini, Jan Ulrich Bartels, Katherine J. Kuchenbecker, Michael J. Black*
- 6 Hyper-Depth: Hypergraph-based Multi-Scale Representation Fusion for Monocular Depth Estimation, *Lin Bie, Siqi Li, Yifan Feng, Yue Gao*

- 7 Quanta Neural Networks: From Photons to Perception, *Varun Sundar, Tianyi Zhang, Sacha Jungerman, Mohit Gupta*
- 8 EDFFDNet: Towards Accurate and Efficient Unsupervised Multi-Grid Image Registration, *Haokai Zhu, Bo Qu, Si-Yuan Cao, Runmin Zhang, Shujie Chen, Bailin Yang, Hui-Liang Shen*
- 9 AdaDrive: Self-Adaptive Slow-Fast System for Language-Grounded Autonomous Driving, *Ruifei Zhang, Junlin Xie, Wei Zhang, Weikai Chen, Xiao Tan, Xiang Wan, Guanbin Li*
- 10 SIGMAN: Scaling 3D Human Gaussian Generation with Millions of Assets, *Yuhang Yang, Fengqi Liu, Yixing Lu, Qin Zhao, Pingyu Wu, Wei Zhai, Ran Yi, Yang Cao, Lizhuang Ma, Zheng-Jun Zha, Junting Dong*
- 11 TAR3D: Creating High-Quality 3D Assets via Next-Part Prediction, *Xuying Zhang, Yutong Liu, Yangguang Li, Renrui Zhang, Yufei Liu, Kai Wang, Wanli Ouyang, Zhiwei Xiong, Peng Gao, Qibin Hou, Ming-Ming Cheng*
- 12 Depth Any Event Stream: Enhancing Event-based Monocular Depth Estimation via Dense-to-Sparse Distillation, *Jinjing Zhu, Tianbo Pan, Zidong Cao, Yexin Liu, James T. Kwok, Hui Xiong*
- 13 PRVQL: Progressive Knowledge-guided Refinement for Robust Egocentric Visual Query Localization, *Bing Fan, Yunhe Feng, Yapeng Tian, James Chenhao Liang, Yuewei Lin, Yan Huang, Heng Fan*
- 14 WildSeg3D: Segment Any 3D Objects in the Wild from 2D Images, *Yansong Guo, Jie Hu, Yansong Qu, Liujuan Cao*
- 15 AerialVG: A Challenging Benchmark for Aerial Visual Grounding by Exploring Positional Relations, *Junli Liu, Qizhi Chen, Zhigang Wang, Yiwen Tang, Yiting Zhang, Chi Yan, Dong Wang, Xuelong Li, Bin Zhao*
- 16 Consistent Time-of-Flight Depth Denoising via Graph-Informed Geometric Attention, *Weida Wang, Changyong He, Jin Zeng, Di Qiu*
- 17 Extending Foundational Monocular Depth Estimators to Fisheye Cameras with Calibration Tokens, *Suchisrit Gangopadhyay, Jung-Hee Kim, Xien Chen, Patrick Rim, Hyoungseob Park, Alex Wong*
- 18 Selective Contrastive Learning for Weakly Supervised Affordance Grounding, *WonJun Moon, Hyun Seok Seong, Jae-Pil Heo*
- 19 OpenSubstance: A High-quality Measured Dataset of Multi-View and -Lighting Images and Shapes, *Fan Pei, Jinchen Bai, Xiang Feng, Zoubin Bi, Kun Zhou, Hongzhi Wu*
- 20 MGSfM: Multi-Camera Geometry Driven Global Structure-from-Motion, *Peilin Tao, Hainan Cui, Diantao Tu, Shuhan Shen*
- 21 PhysSplat: Efficient Physics Simulation for 3D Scenes via MLLM-Guided Gaussian Splatting, *Haoyu Zhao, Hao Wang, Xingyue Zhao, Hao Fei, Hongqiu Wang, Chengjiang Long, Hua Zou*
- 22 AllTracker: Efficient Dense Point Tracking at High Resolution, *Adam W. Harley, Yang You, Xinglong Sun, Yang Zheng, Nikhil Raghuraman, Yunqi Gu, Sheldon Liang, Wen-Hsuan Chu, Achal Dave, Suyu You, Rares Ambrus, Katerina Fragkiadaki, Leonidas Guibas*
- 23 PseudoMapTrainer: Learning Online Mapping without HD Maps, *Christian Löwens, Thorben Funke, Jingchao Xie, Alexandru Paul Condurache*
- 24 LONG3R: Long Sequence Streaming 3D Reconstruction, *Zhuoguang Chen, Minghui Qin, Tianyuan Yuan, Zhe Liu, Hang Zhao*
- 25 DAMap: Distance-aware MapNet for High Quality HD Map Construction, *Jinpeng Dong, Chen Li, Yutong Lin, Jingwen Fu, Sanping Zhou, Nanning Zheng*
- 26 VGMamba: Attribute-to-Location Clue Reasoning for Quantity-Agnostic 3D Visual Grounding, *Yihang Zhu, Jinhao Zhang, Yuxuan Wang, Aming Wu, Cheng Deng*
- 27 OV3D-CG: Open-vocabulary 3D Instance Segmentation with Contextual Guidance, *Mingquan Zhou, Chen He, Ruiping Wang, Xilin Chen*
- 28 AnnotFreeOD: Detecting All Classes at Low Frame Rates Without Human Annotations, *Boyi Sun, Yuhang Liu, Houxin He, Yonglin Tian, Fei-Yue Wang*
- 29 PriOr-Flow: Enhancing Primitive Panoramic Optical Flow with Orthogonal View, *Longliang Liu, Miaojie Feng, Junda Cheng, Jijun Xiang, Xuan Zhu, Xin Yang*
- 30 Learning 4D Embodied World Models, *Haoyu Zhen, Qiao Sun, Hongxin Zhang, Junyan Li, Siyuan Zhou, Yilun Du, Chuang Gan*
- 31 DAVID: Data-efficient and Accurate Vision Models from Synthetic Data, *Fatemeh Saleh, Sadegh Aliakbarian, Charlie Hewitt*

- Lohit Petikam, Xian Xiao, Antonio Criminisi, Thomas J. Cashman, Tadas Baltrusaitis
- 32 Marigold-DC: Zero-Shot Monocular Depth Completion with Guided
☆ Diffusion, Massimiliano Viola, Kevin Qu, Nando Metzger, Bingxin Ke, Alexander Becker, Konrad Schindler, Anton Obukhov
- 33 Multi-view Gaze Target Estimation, Qiaomu Miao, Vivek Raju Golani, Jingyi Xu, Progga Paromita Dutta, Minh Hoai, Dimitris Samaras
- 34 Bayesian-Inspired Space-Time Superpixels, Kent Gauen, Stanley Chan
- 35 Multimodal LLM Guided Exploration and Active Mapping using Fisher Information, Wen Jiang, Boshu Lei, Katrina Ashton, Kostas Daniilidis
- 36 Multispectral Demosaicing via Dual Cameras, SaiKiran Tedla,
*☆ Junyong Lee, Beixuan Yang, Mahmoud Afifi, Michael S. Brown
- 37 DepthSync: Diffusion Guidance-Based Depth Synchronization for Scale- and Geometry-Consistent Video Depth Estimation, Yue-Jiang Dong, Wang Zhao, Jiale Xu, Ying Shan, Song-Hai Zhang
- 38 Generative Modeling of Shape-Dependent Self-Contact Human Poses, Takehiko Ohkawa, Jihyun Lee, Shunsuke Saito, Jason Saragih, Fabian Prada, Yichen Xu, Shouo-I Yu, Ryosuke Furuta, Yoichi Sato, Takaaki Shiratori
- 39 Enhanced Event-based Dense Stereo via Cross-Sensor Knowledge Distillation, Haihao Zhang, Yunjian Zhang, Jianing Li, Lin Zhu, Meng Lv, Yao Zhu, Yanwei Liu, Xiangyang Ji
- 40 PBFG: A New Physically-Based Dataset and Removal of Lens Flares and Glares, Jie Zhu, Sungkil Lee
- 41 Met2Net: A Decoupled Two-Stage Spatio-Temporal Forecasting Model for Complex Meteorological Systems, Shaohan Li, Hao Yang, Min Chen, Xiaolin Qin
- 42 GSOT3D: Towards Generic 3D Single Object Tracking in the Wild, Yifan Jiao, Yunhao Li, Junhua Ding, Qing Yang, Song Fu, Heng Fan, Libo Zhang
- 43 MCAM: Multimodal Causal Analysis Model for Ego-Vehicle-Level Driving Video Understanding, Tongtong Cheng, Rongzhen Li, Yixin Xiong, Tao Zhang, Jing Wang, Kai Liu
- 44 Knowledge-Guided Part Segmentation, Xuejian Gou, Fang Liu, Licheng Jiao, Shuo Li, Lingling Li, Hao Wang, Xu Liu, Puhua Chen, Wenping Ma
- 45 Robust 3D-Masked Part-level Editing in 3D Gaussian Splatting with Regularized Score Distillation Sampling, Hayeon Kim, Ji Ha Jang, Se Young Chun
- 46 COSMO: Combination of Selective Memorization for Low-cost
☆ Vision-and-Language Navigation, Siqi Zhang, Yanyuan Qiao, Qunbo Wang, Zike Yan, Qi Wu, Zhihua Wei, Jing Liu
- 47 TriDi: Trilateral Diffusion of 3D Humans, Objects, and Interactions, Ilya A. Petrov, Riccardo Marin, Julian Chibane, Gerard Pons-Moll
- 48 NavMorph: A Self-Evolving World Model for Vision-and-Language Navigation in Continuous Environments, Xuan Yao, Junyu Gao, Changsheng Xu
- 49 Beyond RGB: Adaptive Parallel Processing for RAW Object Detection, Shani Gamrian, Hila Barel, Feiran Li, Masakazu Yoshimura, Daisuke Iso
- 50 GLEAM: Learning Generalizable Exploration Policy for Active Mapping in Complex 3D Indoor Scene, Xiao Chen, Tai Wang, Quanyi Li, Tao Huang, Jiangmiao Pang, Tianfan Xue
- 51 SMARTIES: Spectrum-Aware Multi-Sensor Auto-Encoder for Remote Sensing Images, Gencer Sumbul, Chang Xu, Emanuele Dalsasso, Devis Tuia
- 52 egoPPG: Heart Rate Estimation from Eye-Tracking Cameras in Egocentric Systems to Benefit Downstream Vision Tasks, Björn Braun, Rayan Armani, Manuel Meier, Max Moebus, Christian Holz
- 53 PHATNet: A Physics-guided Haze Transfer Network for Domain-adaptive Real-world Image Dehazing, Fu-Jen Tsai, Yan-Tsung Peng, Yen-Yu Lin, Chia-Wen Lin
- 54 Efficient Visual Place Recognition Through Multimodal Semantic Knowledge Integration, Sitao Zhang, Hongda Mao, Qingshuang Chen, Yelin Kim
- 55 PoseSyn: Synthesizing Diverse 3D Pose Data from In-the-Wild 2D Data, ChangHee Yang, Hyeonseop Song, Seokhun Choi, Seungwoo Lee, Jaechul Kim, Hoseok Do
- 56 STI-Bench: Are MLLMs Ready for Precise Spatial-Temporal World Understanding?, Yun Li, Yiming Zhang, Tao Lin, Xiangrui Liu, Wenxiao Cai, Zheng Liu, Bo Zhao
- 57 Articulate3D: Holistic Understanding of 3D Scenes as Universal
☆ Scene Description, Anna-Maria Halacheva, Yang Miao, Jan-Nico Zaech, Xi Wang, Luc Van Gool, Danda Pani Paudel
- 58 TorchAdapt: Towards Light-Agnostic Real-Time Visual Perception, Khuram Azeem Hashmi, Karthik Palyakere Suresh, Didier Stricker, Muhammad Zeshan Afzal
- 59 Human-in-the-Loop Local Corrections of 3D Scene Layouts via
* Infilling, Christopher Xie, Armen Avetisyan, Henry Howard-Jenkins, Yawar Siddiqui, Julian Straub, Richard Newcombe, Vasileios Balntas, Jakob Engel
- 60 FlowSeek: Optical Flow Made Easier with Depth Foundation Models
☆ and Motion Bases, Matteo Poggi, Fabio Tosi
- 61 POMATO: Marrying Pointmap Matching with Temporal Motions for Dynamic 3D Reconstruction, Songyan Zhang, Yongtao Ge, Jinyuan Tian, Guangkai Xu, Hao Chen, Chen Lv, Chunhua Shen
- 62 CAT: A Unified Click-and-Track Framework for Realistic Tracking, Yongsheng Yuan, Jie Zhao, Dong Wang, Huchuan Lu
- 63 Diffusion-Based Extreme High-speed Scenes Reconstruction with the Complementary Vision Sensor, Yapeng Meng, Yihan Lin, Taoyi Wang, Yuguo Chen, Lijian Wang, Rong Zhao
- 64 GARF: Learning Generalizable 3D Reassembly for Real-World Fractures, Sihang Li, Zeyu Jiang, Grace Chen, Chenyang Xu, Siqi Tan, Xue Wang, Irving Fang, Kristof Zyskowski, Shannon P. McPherron, Radu Iovita, Chen Feng, Jing Zhang
- 65 DepR: Depth Guided Single-view Scene Reconstruction with Instance-level Diffusion, Qingcheng Zhao, Xiang Zhang, Haiyang Xu, Zeyuan Chen, Jianwen Xie, Yuan Gao, Zhuowen Tu
- 66 XTrack: Multimodal Training Boosts RGB-X Video Object
☆ Trackers, Yuedong Tan, Zongwei Wu, Yuqian Fu, Zhuoyun Zhou, Guolei Sun, Eduard Zamfir, Chao Ma, Danda Paudel, Luc Van Gool, Radu Timofte
- 67 DiffuMatch: Category-Agnostic Spectral Diffusion Priors for Robust Non-rigid Shape Matching, Emery Pierson, Lei Li, Angela Dai, Maks Ovsjanikov
- 68 Joint Learning of Pose Regression and Denoising Diffusion with Score Scaling Sampling for Category-level 6D Pose Estimation, Seunghyun Lee, Tae-Kyun Kim
- 69 UnrealZoo: Enriching Photo-realistic Virtual Worlds for Embodied
* AI, Fangwei Zhong, Kui Wu, Churan Wang, Hao Chen, Hai Ci, Zhoujun Li, Yizhou Wang
- 70 SAC-GNC: Sample Consensus for adaptive Graduated
* Non-Convexity, Valter Piedade, Chitturi Sidhartha, José Gaspar, Venu Madhav Govindu, Pedro Miraldo
- 71 Performing Defocus Deblurring by Modeling its Formation Process, Zhengbo Zhang, Lin Geng Foo, Hossein Rahmani, Jun Liu, De Wen Soh
- 72 Ultra-Precision 6DoF Pose Estimation Using 2-D Interpolated Discrete Fourier Transform, Guowei Shi, Zian Mao, Peisen Huang
- 73 AstroLoc: Robust Space to Ground Image Localizer, Gabriele Berton,
☆ Alex Stokes, Carlo Masone
- 74 RayZer: A Self-supervised Large View Synthesis Model,
○ Hanwen Jiang, Hao Tan, Peng Wang, Haian Jin, Yue Zhao,
🔗 Sai Bi, Kai Zhang, Fujun Luan, Kalyan Sunkavalli, Qixing Huang, Georgios Pavlakos
- 75 Back on Track: Bundle Adjustment for Dynamic Scene Reconstruction,
○ Weirong Chen, Ganlin Zhang, Felix Wimbauer, Rui Wang,
🔗 Nikita Araslanov, Andrea Vedaldi, Daniel Cremers
- 76 Real3D: Towards Scaling Large Reconstruction Models with Real Images, Hanwen Jiang, Qixing Huang, Georgios Pavlakos
- 77 Do It Yourself: Learning Semantic Correspondence from
☆ Pseudo-Labels, Olaf Dünkel, Thomas Wimmer, Christian Theobalt, Christian Rupprecht, Adam Kortylewski
- 78 EMatch: A Unified Framework for Event-based Optical Flow and Stereo Matching, Pengjie Zhang, Lin Zhu, Xiao Wang, Lizhi Wang, Hua Huang
- 79 PanSt3R: Multi-view Consistent Panoptic Segmentation,
☆ Lojze Zust, Yohann Cabon, Juliette Marrie, Leonid Antsfeld, Boris Chidlovskii, Jerome Revaud, Gabriela Csurka
- 80 Stochastic Interpolants for Revealing Stylistic Flows across the History of Art, Pingchuan Ma, Ming Gui, Johannes Schusterbauer,

- Xiaopei Yang, Olga Grebenkova, Vincent Tao Hu, Björn Ommer
 81 Is Tracking Really More Challenging in First Person Egocentric
 * Vision?, Matteo Dunnhofer, Zaira Manigrasso, Christian Micheloni
 82 Where am I? Cross-View Geo-localization with Natural Language
 Descriptions, Junyan Ye, Honglin Lin, Leyan Ou, Dairong Chen,
 Zihao Wang, Qi Zhu, Conghui He, Weijia Li
 83 Less is More: Empowering GUI Agent with Context-Aware
 * Simplification, Gongwei Chen, Xurui Zhou, Rui Shao, Yibo Lyu,
 Kaiwen Zhou, Shuai Wang, Wentao Li, Yinchuan Li, Zhongang Qi,
 Liqiang Nie
 84 CityNav: A Large-Scale Dataset for Real-World Aerial
 Navigation, Jungdae Lee, Taiki Miyamishi, Shuhei Kurita,
 Koya Sakamoto, Daichi Azuma, Yutaka Matsuo, Nakamasa Inoue
 85 VLDrive: Vision-Augmented Lightweight LLMs for Efficient
 Language-grounded Autonomous Driving, Ruifei Zhang, Wei Zhang,
 Xiao Tan, Sibe Yang, Xiang Wan, Xiaonan Luo, Guanbin Li
 86 Rethinking Cross-Modal Interaction in Multimodal Diffusion
 Transformers, Zhengyao Lv, Tianlin Pan, Chenyang Si, Zhaoxi Chen,
 Wangmeng Zuo, Ziwei Liu, Kwan-Yee K. Wong
 87 Toward Material-Agnostic System Identification from Videos,
 Yizhou Zhao, Haoyu Chen, Chunjiang Liu, Zhenyang Li,
 Charles Herrmann, Junhwa Hur, Yinxiao Li, Ming-Hsuan Yang,
 Bhiksha Raj, Min Xu
 88 MagicHOL: Leveraging 3D Priors for Accurate Hand-object
 Reconstruction from Short Monocular Video Clips, Shibo Wang,
 Haonan He, Maria Parelli, Christoph Gebhardt, Zicong Fan, Jie Song
 89 Exploiting Frequency Dynamics for Enhanced Multimodal Event-
 based Action Recognition, Meiqi Cao, Xiangbo Shu, Xin Jiang,
 Rui Yan, Yazhou Yao, Jinhui Tang
 90 Boosting Multi-View Indoor 3D Object Detection via Adaptive
 3D Volume Construction, Runmin Zhang, Zhu Yu, Si-Yuan Cao,
 Lingyu Zhu, Guangyi Zhang, Xiaokai Bai, Hui-Liang Shen
 91 CleanPose: Category-Level Object Pose Estimation via Causal
 Learning and Knowledge Distillation, Xiao Lin, Yun Peng, Liuyi Wang,
 Xianyou Zhong, Minghao Zhu, Yi Feng, Jingwei Yang, Chengju Liu,
 Qijun Chen
 92 ETA: Energy-based Test-time Adaptation for Depth Completion,
 Younjoon Chung, Hyungseob Park, Patrick Rim, Xiaoran Zhang,
 Jihe He, Ziyao Zeng, Safa Cicek, Byung-Woo Hong, James S. Duncan,
 Alex Wong
 93 CoTracker3: Simpler and Better Point Tracking by Pseudo-
 * Labelling Real Videos, Nikita Karaev, Yuri Makarov,
 Janyuan Wang, Natalia Neverova, Andrea Vedaldi, Christian Rupprecht
 94 FiffDepth: Feed-forward Transformation of Diffusion-Based
 Generators for Detailed Depth Estimation, Yunpeng Bai, Qixing Huang
 95 SceneMI: Motion In-betweening for Modeling Human-Scene
 * Interaction, Inwoo Hwang, Bing Zhou, Young Min Kim, Jian Wang,
 Chuan Guo
 96 HORT: Monocular Hand-held Objects Reconstruction with
 * Transformers, Zerui Chen, Rolandos Alexandros Potamias,
 Shizhe Chen, Cordelia Schmid
 97 Active Learning Meets Foundation Models: Fast Remote
 Sensing Data Annotation for Object Detection, Marvin Burges,
 Philippe Ambrozio Dias, Carson Woody, Sarah Walters, Dalton Lunga
 98 Humans as Checkboards: Calibrating Camera Motion Scale for
 World-Coordinate Human Mesh Recovery, Fengyuan Yang, Kerui Gu,
 Ha Linh Nguyen, Tze Ho Elden Tse, Angela Yao
 99 Learnable Fractional Reaction-Diffusion Dynamics for Under-
 * Display ToF Imaging and Beyond, Xin Qiao, Matteo Poggi, Xing Wei,
 Pengchao Deng, Yanhui Zhou, Stefano Mattoccia
 100 ArgoTweak: Towards Self-Updating HD Maps through Structured
 Priors, Lena Wild, Rafael Valencia, Patric Jensfelt
 101 DEPTHOR: Depth Enhancement from a Practical Light-Weight dToF
 Sensor and RGB Image, Jijun Xiang, Xuan Zhu, Xianqi Wang, Yu Wang,
 Hong Zhang, Fei Guo, Xin Yang
 102 GT-Mean Loss: A Simple Yet Effective Solution for Brightness
 Mismatch in Low-Light Image Enhancement, Jingxi Liao, Shijie Hao,
 Richang Hong, Meng Wang
 103 GeoExplorer: Active Geo-localization with Curiosity-Driven
 * Exploration, Li Mi, Manon Béchaz, Zeming Chen, Antoine Bosselut,
 Devis Tuia
 104 ROADWork: A Dataset and Benchmark for Learning to Recognize,
 Observe, Analyze and Drive Through Work Zones, Anurag Ghosh,
 Shen Zheng, Robert Tamburo, Khiem Vuong, Juan Alvarez-Padilla,
 Hailiang Zhu, Michael Cardei, Nicholas Dunn, Christoph Mertz,
 Srinivasa G. Narasimhan
 105 WildSAT: Learning Satellite Image Representations from Wildlife
 Observations, Rangel Daroya, Elijah Cole, Oisín Mac Aodha, Grant Van
 Horn, Subhansu Maji
 106 RoMo: Robust Motion Segmentation Improves Structure from
 Motion, Lily Goli, Sara Sabour, Mark Matthews, Marcus A. Brubaker,
 Dmitry Lagun, Alec Jacobson, David J. Fleet, Saurabh Saxena,
 Andrea Tagliasacchi
 107 Hints of Prompt: Enhancing Visual Representation for Multimodal
 LLMs in Autonomous Driving, Hao Zhou, Zhanning Gao, Zhili Chen,
 Maosheng Ye, Qifeng Chen, Tongyi Cao, Honggang Qi
 108 Learning Large Motion Estimation from Intermediate Representations
 * with a High-Resolution Optical Flow Dataset Featuring Long-Range
 ☆ Dynamic Motion, Hoonhee Cho, Yuhwan Jeong, Kuk-Jin Yoon
 109 Robust Low-light Scene Restoration via Illumination Transition,
 Ze Li, Feng Zhang, Xiatian Zhu, Meng Zhang, Yanghong Zhou, P. Y. Mok
 110 CCMNet: Leveraging Calibrated Color Correction Matrices for Cross-
 ☆ Camera Color Constancy, Dongyoung Kim, Mahmoud Afifi,
 Dongyun Kim, Michael S. Brown, Seon Joo Kim
 111 UrbanLLaVA: A Multi-modal Large Language Model for Urban
 Intelligence, Jie Feng, Shengyuan Wang, Tianhui Liu, Yanxin Xi,
 Yong Li
 112 CO2-Net: A Physics-Informed Spatio-Temporal Model for Global
 Surface CO2 Reconstruction, Hao Zheng, Yuting Zheng,
 Hanbo Huang, Chaofan Sun, Enhui Liao, Lin Liu, Yi Han, Hao Zhou,
 Shiyu Liang
 113 Zero-shot Inexact CAD Model Alignment from a Single Image,
 Pattaramanee Arsomngern, Sasikarn Khwanmuang, Matthias Nießner,
 Supasorn Suwajanakorn
 114 Balanced Sharpness-Aware Minimization for Imbalanced
 Regression, Yahao Liu, Qin Wang, Lixin Duan, Wen Li
 115 Dual Reciprocal Learning of Language-based Human Motion
 Understanding and Generation, Chen Liang, Zhicheng Shi,
 Wenguan Wang, Yi Yang
 116 HazeFlow: Revisit Haze Physical Model as ODE and Non-Homogeneous
 Haze Generation for Real-World Dehazing, Junseong Shin,
 Seungwoo Chung, Yunjeong Yang, Tae Hyun Kim
 117 Mamba-3VL: Taming State Space Model for 3D Vision Language
 Learning, Yuan Wang, Yuxin Chen, Zhongang Qi, Lijun Liu, Jile Jiao,
 Xuetao Feng, Yujia Liang, Ying Shan, Zhipeng Zhang
 118 Motat: Unsupervised 3D Object Detection by Modality and Task-
 specific Knowledge Transfer, Hai Wu, Hongwei Lin, Xusheng Guo,
 Xin Li, Mingming Wang, Cheng Wang, Chenglu Wen
 119 DeGauss: Dynamic-Static Decomposition with Gaussian Splatting
 for Distractor-free 3D Reconstruction, Rui Wang, Quentin Lohmeyer,
 Mirko Meboldt, Siyu Tang
 120 Manual-PA: Learning 3D Part Assembly from Instruction Diagrams,
 Jiahao Zhang, Anoop Cherian, Cristian Rodriguez, Weijian Deng,
 Stephen Gould
 121 MoMa-Kitchen: A 100K+ Benchmark for Affordance-Grounded
 Last-Mile Navigation in Mobile Manipulation, Pingrui Zhang,
 Xianqiang Gao, Yuhao Wu, Kehui Liu, Dong Wang, Zhigang Wang,
 Bin Zhao, Yan Ding, Xuelong Li
 122 NavQ: Learning a Q-Model for Foresighted Vision-and-Language
 Navigation, Peiran Xu, Xicheng Gong, Yadong Mu
 123 Embodied VideoAgent: Persistent Memory from Egocentric Videos
 * and Embodied Sensors Enables Dynamic Scene Understanding,
 Yue Fan, Xiaojian Ma, Rongpeng Su, Jun Guo, Ruijie Wu, Xi Chen, Qing Li
 124 LightsOut: Diffusion-based Outpainting for Enhanced Lens Flare
 Removal, Shr-Ruei Tsai, Wei-Cheng Chang, Jie-Ying Lee, Chih-Hai Su,
 Yu-Lun Liu
 125 Rethinking Multi-modal Object Detection from the Perspective
 of Mono-Modality Feature Learning, Tianyi Zhao, Boyang Liu,
 Yanglei Gao, Yiming Sun, Maoxun Yuan, Xingxing Wei
 126 GeoDiffusion: A Training-Free Framework for Accurate 3D Geometric

- ☆ Conditioning in Image Generation, *Phillip Mueller, Talip Uenlue, Sebastian Schmidt, Marcel Kollovieh, Jiajie Fan, Stephan Günnemann, Lars Mikelsons*
- 127 OVA-Fields: Weakly Supervised Open-Vocabulary Affordance Fields for Robot Operational Part Detection, *Heng Su, Mengying Xie, Nieqing Cao, Yan Ding, Beichen Shao, Xianlei Long, Fuqiang Gu, Chao Chen*
- 128 Arti-PG: A Toolbox for Procedurally Synthesizing Large-Scale and Diverse Articulated Objects with Rich Annotations, *Jianhua Sun, Yuxuan Li, Jiude Wei, Longfei Xu, Nange Wang, Yining Zhang, Cewu Lu*
- 129 Scaling 3D Compositional Models for Robust Classification and Pose Estimation, *Xiaoding Yuan, Guofeng Zhang, Prakhar Kaushik, Artur Jesslen, Adam Kortylewski, Alan Yuille*
- 130 RoboTrom-Nav: A Unified Framework for Embodied Navigation Integrating Perception, Planning, and Prediction, *Yufeng Zhong, Chengjian Feng, Feng Yan, Fanfan Liu, Liming Zheng, Lin Ma*
- 131 Optical Model-Driven Sharpness Mapping for Autofocus in Small Depth-of-Field and Severe Defocus Scenarios, *Chen-Liang Fan, Mingpei Cao, Chih Chien Hung, Yuesheng Zhu*
- 132 X-Capture: An Open-Source Portable Device for Multi-Sensory Learning, *Samuel Clarke, Suzannah Wistreich, Yanjie Ze, Jiajun Wu*
- 133 DRaM-LHM: A Quaternion Framework for Iterative Camera Pose Estimation, *Chen Lin, Weizhi Du, Zhixiang Min, Baochen She, Enrique Dunn, Sonya M. Hanson*
- 134 VoteSplat: Hough Voting Gaussian Splatting for 3D Scene Understanding, *Minchao Jiang, Shunyu Jia, Jiaming Gu, Xiaoyuan Lu, Guangming Zhu, Anqi Dong, Liang Zhang*
- 135 UMDATrack: Unified Multi-Domain Adaptive Tracking Under Adverse Weather Conditions, *Siyuan Yao, Rui Zhu, Ziqi Wang, Wenqi Ren, Yanyang Yan, Xiaochun Cao*
- 136 Prior-aware Dynamic Temporal Modeling Framework for Sequential 3D Hand Pose Estimation, *Pengfei Ren, Jingyu Wang, Haifeng Sun, Qi Qi, Xingyu Liu, Menghao Zhang, Lei Zhang, Jing Wang, Jianxin Liao*
- 137 Epipolar Consistent Attention Aggregation Network for Unsupervised Light Field Disparity Estimation, *Chen Gao, Shuo Zhang, Youfang Lin*
- 138 Fish2Mesh Transformer: 3D Human Mesh Recovery from Egocentric Vision, *Tianma Shen, Aditya Puranik, James Vong, Vrushabh Deogirikar, Ryan Fell, Julianna Dietrich, Maria Kyrarini, Christopher Kitts, David C. Jeong*
- 139 ATLAS: Decoupling Skeletal and Shape Parameters for Expressive Parametric Human Modeling, *Jinhyung Park, Javier Romero, Shunsuke Saito, Fabian Prada, Takaaki Shiratori, Yichen Xu, Federica Bogo, Shou-I Yu, Kris Kitani, Rawal Khrodar*
- 140 GloPER: Unsupervised Animal Pattern Extraction from Local Reconstruction, *Bowen Chen, Yun Sing Koh, Gillian Dobbie*
- 141 ObjectRelator: Enabling Cross-View Object Relation Understanding
- ☆ Across Ego-Centric and Exo-Centric Perspectives, *Yuqian Fu, Runze Wang, Bin Ren, Guolei Sun, Biao Gong, Yanwei Fu, Danda Pani Paudel, Xuanjing Huang, Luc Van Gool*
- 142 MoSiC: Optimal-Transport Motion Trajectory for Dense Self-Supervised Learning, *Mohammadreza Salehi, Shashanka Venkataramanan, Ioana Simion, Efstratios Gavves, Cees G. M. Snoek, Yuki M Asano*
- 143 On the Generalization of Representation Uncertainty in Earth Observation, *Spyros Kondylatos, Nikolaos Ioannis Bountos, Dimitrios Michail, Xiao Xiang Zhu, Gustau Camps-Valls, Ioannis Papoutsis*
- 144 MeshMamba: State Space Models for Articulated 3D Mesh Generation and Reconstruction, *Yusuke Yoshiyasu, Leyuan Sun, Ryusuke Sagawa*
- 145 Predict-Optimize-Distill: A Self-Improving Cycle for 4D Object Understanding, *Mingxuan Wu, Huang Huang, Justin Kerr, Chung Min Kim, Anthony Zhang, Brent Yi, Angjoo Kanazawa*
- 146 SceneSplat: Gaussian Splatting-based Scene Understanding with
- Vision-Language Pretraining, *Yue Li, Qi Ma, Runyi Yang, Huapeng Li, Mengjiao Ma, Bin Ren, Nikola Papanic, Nicu Sebe, Ender Konukoglu, Theo Gevers, Luc Van Gool, Martin R. Oswald, Danda Pani Paudel*
- 147 Self-Ensembling Gaussian Splatting for Few-Shot Novel View
- Synthesis, *Chen Zhao, Xuan Wang, Tong Zhang, Saqib Javed, Mathieu Salzmann*
- 148 Are VLMs Ready for Autonomous Driving? An Empirical Study
- ☆ from the Reliability, Data and Metric Perspectives, *Shaoyuan Xie, Lingdong Kong, Yuhao Dong, Chonghao Sima, Wenwei Zhang, Qi Alfred Chen, Ziwei Liu, Liang Pan*
- 149 Humans as a Calibration Pattern: Dynamic 3D Scene Reconstruction from Unsynchronized and Uncalibrated Videos, *Changwoon Choi, Jeongjun Kim, Geonho Cha, Minkwan Kim, Dongyoon Wee, Young Min Kim*
- 150 PhysRig: Differentiable Physics-Based Skinning and Rigging Framework for Realistic Articulated Object Modeling, *Hao Zhang, Haolan Xu, Chun Feng, Varun Jampani, Narendra Ahuja*
- 152 Learning Dense Feature Matching via Lifting Single 2D Image to 3D Space, *Yingping Liang, Yutao Hu, Wenqi Shao, Ying Fu*
- 153 GeometryCrafter: Consistent Geometry Estimation for Open-world Videos with Diffusion Priors, *Tian-Xing Xu, Xiangjun Gao, Wenbo Hu, Xiaoyu Li, Song-Hai Zhang, Ying Shan*
- 154 Placelt3D: Language-Guided Object Placement in Real 3D
- ☆ Scenes, *Ahmed Abdelreheem, Filippo Aleotti, Jamie Watson, Zawar Qureshi, Abdelrahman Eldesokey, Peter Wonka, Gabriel Brostow, Sara Vicente, Guillermo Garcia-Hernando*
- 155 Trace3D: Consistent Segmentation Lifting via Gaussian Instance Tracing, *Hongyu Shen, Junfeng Ni, Yixin Chen, Weishuo Li, Mingtao Pei, Siyuan Huang*
- 156 Low-Light Image Enhancement Using Event-Based Illumination
- ☆ Estimation, *Lei Sun, Yuhao Bao, Jiajun Zhai, Jingyun Liang, Yulun Zhang, Kaiwei Wang, Danda Pani Paudel, Luc Van Gool*
- 157 Hybrid-grained Feature Aggregation with Coarse-to-fine Language Guidance for Self-supervised Monocular Depth Estimation, *Wenyao Zhang, Hongsi Liu, Bohan Li, Jiawei He, Zekun Qi, Yunnan Wang, Shengyang Zhao, Xinqiang Yu, Wenjun Zeng, Xin Jin*
- 158 Zero-AVSR: Zero-Shot Audio-Visual Speech Recognition with LLMs by Learning Language-Agnostic Speech Representations, *Jeong Hun Yeo, Minsu Kim, Chae Won Kim, Stavros Petridis, Yong Man Ro*
- 159 Jigsaw++: Imagining Complete Shape Priors for Object Reassembly, *Jiaxin Lu, Gang Hua, Qixing Huang*
- 160 Seeing and Seeing Through the Glass: Real and Synthetic Data for Multi-Layer Depth Estimation, *Hongyu Wen, Yiming Zuo, Venkat Subramanian, Patrick Chen, Jia Deng*
- 161 SpatialTrackerV2: Advancing 3D Point Tracking with Explicit Camera Motion, *Yuxi Xiao, Jianyuan Wang, Nan Xue, Nikita Karaev, Yuri Makarov, Bingyi Kang, Xing Zhu, Hujun Bao, Yujun Shen, Xiaowei Zhou*
- 162 UST-SSM: Unified Spatio-Temporal State Space Models for Point
- ☆ Cloud Video Modeling, *Peiming Li, Ziyi Wang, Yulin Yuan, Hong Liu, Xiangming Meng, Junsong Yuan, Mengyuan Liu*
- 163 A Simple yet Mighty Hartley Diffusion Versatilist for Generalizable Dense Vision Tasks, *Qi Bi, Jingjun Yi, Huimin Huang, Hao Zheng, Haolan Zhan, Wei Ji, Yawen Huang, Yuexiang Li, Yefeng Zheng*
- 164 Seeing 3D Through 2D Lenses: 3D Few-Shot Class-Incremental Learning via Cross-Modal Geometric Rectification, *Tuo Xiang, Xuemiao Xu, Bangzhen Liu, Jinyi Li, Yong Li, Shengfeng He*
- 165 DSO: Aligning 3D Generators with Simulation Feedback for
- ☆ Physical Soundness, *Ruining Li, Chuanxia Zheng, Christian Rupprecht, Andrea Vedaldi*
- 166 Feed-Forward SceneDINO for Unsupervised Semantic Scene Completion, *Aleksandar Jevtić, Christoph Reich, Felix Wimbauer, Oliver Hahn, Christian Rupprecht, Stefan Roth, Daniel Cremers*
- 167 Cycle-Consistent Learning for Joint Layout-to-Image Generation and Object Detection, *Xinhao Cai, Qixia Lai, Gensheng Pei, Xiangbo Shu, Yazhou Yao, Wenguan Wang*
- 168 IGL-Nav: Incremental 3D Gaussian Localization for Image-goal Navigation, *Wenxuan Guo, Xiuwei Xu, Hang Yin, Ziwei Wang, Jianjiang Feng, Jie Zhou, Jiwen Lu*
- 169 AR-VRM: Imitating Human Motions for Visual Robot Manipulation with Analogical Reasoning, *Dejie Yang, Zijiang Zhao, Yang Liu*
- 170 MemDistill: Distilling LiDAR Knowledge into Memory for Camera-Only
- ☆ 3D Object Detection, *Donghyeon Kwon, Youngseok Yoon, Hyeongseok Son, Suha Kwak*
- 171 Towards Long-Horizon Vision-Language-Action System: Reasoning, Acting and Memory, *Daixun Li, Yusi Zhang, Mingxiang Cao,*

- Donglai Liu, Weiying Xie, Tianlin Hui, Lunkai Lin, Zhiqiang Xie, Yunsong Li
- 172 Head2Body: Body Pose Generation from Multi-sensory Head-mounted Inputs, Minh Tran, Hongda Mao, Qingshuang Chen, Yelin Kim
- 173 Task-Decoupled Bézier Surface Constraint for Uneven Low-Light Image Enhancement, Xingxiang Zhou, Xiangdong Su, Haoran Zhang, Wei Chen, Guanglai Gao
- 174 Unleashing the Temporal Potential of Stereo Event Cameras for Continuous-Time 3D Object Detection, Jae-Young Kang, Hoonhee Cho, Kuk-Jin Yoon
- 175 PlaneRAS: Learning Planar Primitives for 3D Plane Recovery, Fang Zhang, Wenzhao Zheng, Linqing Zhao, Zelan Zhu, Jiwen Lu, Xiuzhuang Zhou
- 176 O-MaMa: Learning Object Mask Matching between Egocentric and Exocentric Views, Lorenzo Mur-Labadia, Maria Santos-Villafranca, Jesus Bermudez-Cameo, Alejandro Perez-Yus, Ruben Martinez-Cantin, Jose J. Guerrero
- 177 FedVLA: Federated Vision-Language-Action Learning with Dual Gating Mixture-of-Experts for Robotic Manipulation, Cui Miao, Tao Chang, Meihan Wu, Hongbin Xu, Chun Li, Ming Li, Xiaodong Wang
- 178 How To Make Your Cell Tracker Say "I dunno!", Richard D. Paul, Johannes Seiffarth, David Rügamer, Katharina Nöh, Hanno Scharr
- 179 3DSRBench: A Comprehensive 3D Spatial Reasoning Benchmark, Wufei Ma, Haoyu Chen, Guofeng Zhang, Yu-Cheng Chou, Jieneng Chen, Celso M de Melo, Alan Yuille
- 180 Harnessing Massive Satellite Imagery with Efficient Masked Image Modeling, Fengxiang Wang, Hongzhen Wang, Di Wang, Zonghao Guo, Zhenyu Zhong, Long Lan, Wenjing Yang, Jing Zhang
- 181 Exploring the Adversarial Vulnerabilities of Vision-Language-Action Models in Robotics, Taowen Wang, Cheng Han, James Liang, Wenhao Yang, Dongfang Liu, Luna Xinyu Zhang, Qifan Wang, Jiebo Luo, Ruixiang Tang
- 182 Simultaneous Motion And Noise Estimation with Event Cameras, Shintaro Shiba, Yoshimitsu Aoki, Guillermo Gallego
- 183 EgoAgent: A Joint Predictive Agent Model in Egocentric Worlds, Lu Chen, Yizhou Wang, Shixiang Tang, Qianhong Ma, Tong He, Wanli Ouyang, Xiaowei Zhou, Hujun Bao, Sida Peng
- 184 Weakly-Supervised Learning of Dense Functional Correspondences, Stefan Stojanov, Linan Zhao, Yunzhi Zhang, Daniel L. K. Yamins, Jiajun Wu
- 185 Power of Cooperative Supervision: Multiple Teachers Framework for Advanced 3D Semi-Supervised Object Detection, Jin-Hee Lee, Jae-Keun Lee, Jeseok Kim, Kwon Soon
- 186 Layer-wise Vision Injection with Disentangled Attention for Efficient LVLMs, Xuange Zhang, Dengjie Li, Bo Liu, Zenghao Bao, Yao Zhou, Baisong Yang, Zhongying Liu, Yujie Zhong, Tongtong Yuan
- 187 CMT: A Cascade MAR with Topology Predictor for Multimodal Conditional CAD Generation, Jianyu Wu, Yizhou Wang, Xiangyu Yue, Xinzhu Ma, Jinyang Guo, Dongzhan Zhou, Wanli Ouyang, Shixiang Tang
- 188 Embodied Navigation with Auxiliary Task of Action Description Prediction, Haru Kondoh, Asako Kanezaki
- 189 Open-Vocabulary Octree-Graph for 3D Scene Understanding, Zhigang Wang, Yifei Su, Chenhui Li, Dong Wang, Yan Huang, Xuelong Li, Bin Zhao
- 190 Learning an Implicit Physics Model for Image-based Fluid Simulation, Emily Yue-Ting Jia, Jiageng Mao, Zhiyuan Gao, Yajie Zhao, Yue Wang
- 191 PropVG: End-to-End Proposal-Driven Visual Grounding with Multi-Granularity Discrimination, Ming Dai, Wenxuan Cheng, Jiedong Zhuang, Jiang-jiang Liu, Hongshen Zhao, Zhenhua Feng, Wankou Yang
- 192 StableDepth: Scene-Consistent and Scale-Invariant Monocular Depth, Zheng Zhang, Lihe Yang, Tianyu Yang, Chaohui Yu, Xiaoyang Guo, Yixing Lao, Hengshuang Zhao
- 193 PoseAnchor: Robust Root Position Estimation for 3D Human Pose Estimation, Jun-Hee Kim, Jumin Han, Seong-Whan Lee
- 194 4DSegStreamer: Streaming 4D Panoptic Segmentation via Dual Threads, Ling Liu, Jun Tian, Li Yi
- 195 Color Matching Using Hypernetwork-Based Kolmogorov-Arnold Networks, Artem Nikonov, Georgy Perevozchikov, Andrei Korepanov, Nancy Mehta, Mahmoud Afifi, Egor Ershov, Radu Timofte
- 196 Boost 3D Reconstruction using Diffusion-based Monocular Camera Calibration, Junyuan Deng, Wei Yin, Xiaoyang Guo, Qian Zhang, Xiaotao Hu, Weiqiang Ren, Xiao-Xiao Long, Ping Tan
- 197 Spatial-Temporal Aware Visuomotor Diffusion Policy Learning, Zhenyang Liu, Yikai Wang, Kuanning Wang, Longfei Liang, Xiangyang Xue, Yanwei Fu
- 198 GEOBench-VLM: Benchmarking Vision-Language Models for Geospatial Tasks, Muhammad Danish, Muhammad Akhtar Munir, Syed Roshan Ali Shah, Kartik Kuckreja, Fahad Shahbaz Khan, Paolo Fraccaro, Alexandre Lacoste, Salman Khan
- 199 Single-Scanline Relative Pose Estimation for Rolling Shutter Cameras, Petr Hruby, Marc Pollefeys
- 200 3D Mesh Editing using Masked LRMs, Will Gao, Dilin Wang, Yuchen Fan, Aljaz Bozic, Tuur Stuyck, Zhengqin Li, Zhao Dong, Rakesh Ranjan, Nikolaos Sarafianos
- 201 HcPose(BF): Predicting Front & Back Surfaces to Construct Ultra-Dense 2D-3D Correspondences for Pose Estimation, Yulin Wang, Mengting Hu, Hongli Li, Chen Luo
- 202 Information-Bottleneck Driven Binary Neural Network for Change Detection, Kaijie Yin, Zhiyuan Zhang, Shu Kong, Tian Gao, Cheng-Zhong Xu, Hui Kong
- 203 GaussianVideo: Efficient Video Representation via Hierarchical Gaussian Splatting, Andrew Bond, Jui-Hsien Wang, Long Mai, Erkut Erdem, Aykut Erdem
- 204 VPR-Cloak: A First Look at Privacy Cloak Against Visual Place Recognition, Shuting Dong, Mingzhi Chen, Feng Lu, Hao Yu, Guanghao Li, Zhe Wu, Ming Tang, Chun Yuan
- 205 Event-based Tiny Object Detection: A Benchmark Dataset and Baseline, Nuo Chen, Chao Xiao, Yimian Dai, Shiman He, Miao Li, Wei An
- 206 PhysTwin: Physics-Informed Reconstruction and Simulation of Deformable Objects from Videos, Hanxiao Jiang, Hao-Yu Hsu, Kaifeng Zhang, Hsin-Ni Yu, Shenlong Wang, Yunzhu Li
- 207 GaussianProperty: Integrating Physical Properties to 3D Gaussians with LMMs, Xinli Xu, Wenhong Ge, Dicong Qiu, ZhiFei Chen, Dongyu Yan, Zhuoyun Liu, Haoyu Zhao, Hanfeng Zhao, Shunsi Zhang, Junwei Liang, Ying-Cong Chen
- 208 Neural Solver of Dichromatic Reflection Model for Specular Highlight Removal, Gang Fu
- 209 Retinex-MEF: Retinex-based Glare Effects Aware Unsupervised Multi-Exposure Image Fusion, Haowen Bai, Jiangshe Zhang, Zixiang Zhao, Lilun Deng, Yukun Cui, Shuang Xu
- 210 Frequency-Aligned Knowledge Distillation for Lightweight Spatiotemporal Forecasting, Yuqi Li, Chuanguang Yang, Hansheng Zeng, Zeyu Dong, Zhulin An, Yongjun Xu, Yingli Tian, Hao Wu
- 211 Adaptive Dual Uncertainty Optimization: Boosting Monocular 3D Object Detection under Test-Time Shifts, Zixuan Hu, Dongxiao Li, Xinzhu Ma, Shixiang Tang, Xiaotong Li, Wenhan Yang, Ling-Yu Duan
- 212 CAD-Assistant: Tool-Augmented VLLMs as Generic CAD Task Solvers, Dimitrios Mallis, Ahmet Serda Karadeniz, Sebastian Cavada, Danila Rukhovich, Niki Foteinopoulou, Kseniya Cherenkova, Anis Kacem, Djamila Aouada
- 213 Dynamic Point Maps: A Versatile Representation for Dynamic 3D Reconstruction, Edgar Sucar, Zihang Lai, Eldar Insafutdinov, Andrea Vedaldi
- 214 Diffusion-based 3D Hand Motion Recovery with Intuitive Physics, Yufei Zhang, Zijun Cui, Jeffrey O. Kephart, Qiang Ji
- 215 Physics Context Builders: A Modular Framework for Physical Reasoning in Vision-Language Models, Vahid Balazadeh, Mohammadmehdi Ataei, Hyunmin Cheong, Amir Hosein Khasahmadi, Rahul G. Krishnan
- 216 Measuring the Impact of Rotation Equivariance on Aerial Object Detection, Xiuyu Wu, Xinhao Wang, Xiubin Zhu, Lan Yang, Jiyuan Liu, Xingchen Hu
- 217 Exploring View Consistency for Scene-Adaptive Low-Light Light Field Image Enhancement, Shuo Zhang, Chen Gao, Youfang Lin
- 218 VOccl3D: A Video Benchmark Dataset for 3D Human Pose and Shape Estimation under real Occlusions, Yash Garg, Saketh Bachu, Arindam Dutta, Rohit Lal, Sarosij Bose, Calvin-Khang Ta, M. Salman Asif, Amit Roy-Chowdhury

- 219 Tracking Tiny Drones against Clutter: Large-Scale Infrared Benchmark with Motion-Centric Adaptive Algorithm, *Jiahao Zhang, Zongli Jiang, Jinli Zhang, Yixin Wei, Liang Li, Yizheng Wang, Gang Wang*
- 220 TruthPrint: Mitigating Large Vision-Language Models Object Hallucination Via Latent Truthful-Guided Pre-Intervention, *Jinhao Duan, Fei Kong, Hao Cheng, James Diffenderfer, Bhavya Kailkhura, Lichao Sun, Xiaofeng Zhu, Xiaoshuang Shi, Kaidi Xu*
- 221 TerraMind: Large-Scale Generative Multimodality for Earth Observation, *Johannes Jakubik, Felix Yang, Benedikt Blumenstiel, Erik Scheurer, Rocco Sedona, Stefano Maurogiovanni, Jente Bosmans, Nikolaos Dionelis, Valerio Marsocci, Niklas Kopp, Rahul Ramachandran, Paolo Fraccaro, Thomas Brunschweiler, Gabriele Cavallaro, Juan Bernabe-Moreno, Nicolas Longépé*
- 222 MM-Spatial: Exploring 3D Spatial Understanding in Multimodal LLMs, *Erik Daxberger, Nina Wenzel, David Griffiths, Haiming Gang, Justin Lazarow, Gefen Kohavi, Kai Kang, Marcin Eichner, Yinfei Yang, Afshin Dehghan, Peter Grasch*
- 223 AutoComPose: Automatic Generation of Pose Transition Descriptions for Composed Pose Retrieval Using Multimodal LLMs, *Yi-Ting Shen, Sungmin Eum, Doheon Lee, Rohit Shete, Chiao-Yi Wang, Heesung Kwon, Shuvra S. Bhattacharyya*
- 224 Image-Guided Shape-from-Template Using Mesh Inextensibility Constraints, *Thuy Tran, Ruochen Chen, Shaifali Parashar*
- 225 Distilling Diffusion Models to Efficient 3D LiDAR Scene Completion, *Shengyuan Zhang, An Zhao, Ling Yang, Zejian Li, Chenye Meng, Haoran Xu, Tianrun Chen, AnYang Wei, Perry Pengyun Gu, Lingyun Sun*
- 226 Heavy Labels Out! Dataset Distillation with Label Space Lightening, *Ruonan Yu, Songhua Liu, Zigeng Chen, Jingwen Ye, Xinchao Wang*
- 227 3D-MOOD: Lifting 2D to 3D for Monocular Open-Set Object Detection, *Yung-Hsu Yang, Luigi Piccinelli, Mattia Segu, Siyuan Li, Rui Huang, Yuqian Fu, Marc Pollefeys, Hermann Blum, Zuria Bauer*
- 228 LUDVIG: Learning-Free Uplifting of 2D Visual Features to Gaussian Splatting Scenes, *Juliette Marrie, Romain Menegaux, Michael Arbel, Diane Larlus, Julien Mairal*
- 229 GeoMan: Temporally Consistent Human Geometry Estimation using Image-to-Video Diffusion, *Gwanghyun Kim, Xueting Li, Ye Yuan, Koki Nagano, Tianye Li, Jan Kautz, Se Young Chun, Umar Iqbal*
- 230 ClearSight: Human Vision-Inspired Solutions for Event-Based Motion Deblurring, *Xiaopeng Lin, Yulong Huang, Hongwei Ren, Zunchang Liu, Hongxiang Huang, Yue Zhou, Haotian Fu, Bojun Cheng*
- 231 VOVTrack: Exploring the Potentiality in Raw Videos for Open-Vocabulary Multi-Object Tracking, *Zekun Qian, Ruize Han, Junhui Hou, Lingqi Song, Wei Feng*
- 232 Unsupervised Identification of Protein Compositions and Conformations via Implicit Content-Transformation Disentanglement, *Mostofa Rafid Uddin, Jana Armouti, Min Xu*
- 233 Not All Frame Features Are Equal: Video-to-4D Generation via Decoupling Dynamic-Static Features, *Liyang Yang, Chen Liu, Zhenwei Zhu, Ajian Liu, Hui Ma, Jian Nong, Yanyan Liang*
- 234 Future-Aware Interaction Network For Motion Forecasting, *Shijie Li, Chunyu Liu, Xun Xu, Si Yong Yeo, Xulei Yang*
- 235 EventUPS: Uncalibrated Photometric Stereo Using an Event Camera, *Jinxiu Liang, Bohan Yu, Siqi Yang, Haotian Zhuang, Jieji Ren, Peiqi Duan, Boxin Shi*
- 236 PHD: Personalized 3D Human Body Fitting with Point Diffusion, *Hsuan-I Ho, Chen Guo, Po-Chen Wu, Ivan Shugurov, Chengcheng Tang, Abhay Mittal, Sizhe An, Manuel Kaufmann, Linguang Zhang*
- 237 Frequency Domain-Based Diffusion Model for Unpaired Image Dehazing, *Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, Ming-Hsuan Yang*
- 238 Language Driven Occupancy Prediction, *Zhu Yu, Bowen Pang, Lizhe Liu, Runmin Zhang, Qiang Li, Si-Yuan Cao, Maochun Luo, Mingxia Chen, Sheng Yang, Hui-Liang Shen*
- 239 Rethinking the Upsampling Process in Light Field Super-Resolution with Spatial-Epipolar Implicit Image Function, *Ruixuan Cong, Yu Wang, Mingyuan Zhao, Da Yang, Rongshan Chen, Hao Sheng*
- 240 C4D: 4D Made from 3D through Dual Correspondences, *Shizun Wang, Zhenxiang Jiang, Xingyi Yang, Xinchao Wang*
- 241 Instance-Level Video Depth in Groups Beyond Occlusions, *Yuan Liang, Yang Zhou, Ziming Sun, Tianyi Xiang, Guiqing Li, Shengfeng He*
- 242 ScoreHOI: Physically Plausible Reconstruction of Human-Object Interaction via Score-Guided Diffusion, *Ao Li, Jinpeng Liu, Yixuan Zhu, Yansong Tang*
- 243 Active Perception Meets Rule-Guided RL: A Two-Phase Approach for Precise Object Navigation in Complex Environments, *Liang Qin, Min Wang, Peiwei Li, Wengang Zhou, Houqiang Li*
- 244 MonoSOWA: Scalable Monocular 3D Object Detector Without Human Annotations, *Jan Skvrna, Lukas Neumann*
- 245 Estimating 2D Camera Motion with Hybrid Motion Basis, *Haipeng Li, Tianhao Zhou, Zhanglei Yang, Yi Wu, Yan Chen, Zijing Mao, Shen Cheng, Bing Zeng, Shuaicheng Liu*
- 246 AgroBench: Vision-Language Model Benchmark in Agriculture, *Risa Shinoda, Nakamasa Inoue, Hirokatsu Kataoka, Masaki Onishi, Yoshitaka Ushiku*
- 247 Princeton365: A Diverse Dataset with Accurate Camera Pose, *Karhan Kayan, Stamatis Alexandropoulos, Rishabh Jain, Yiming Zuo, Erich Liang, Jia Deng*
- 248 H3R: Hybrid Multi-view Correspondence for Generalizable 3D Reconstruction, *Heng Jia, Linchao Zhu, Na Zhao*
- 249 From Abyssal Darkness to Blinding Glare: A Benchmark on Extreme Exposure Correction in Real World, *Bo Wang, Huiyuan Fu, Zhiye Huang, Siru Zhang, Xin Wang, Huadong Ma*
- 250 Learning to See in the Extremely Dark, *Hai Jiang, Binhao Guan, Zhen Liu, Xiaohong Liu, Jian Yu, Zheng Liu, Songchen Han, Shuaicheng Liu*
- 251 Dita: Scaling Diffusion Transformer for Generalist Vision-Language-Action Policy, *Zhi Hou, Tianyi Zhang, Yuwen Xiong, Haonan Duan, Hengjun Pu, Ronglei Tong, Chengyang Zhao, Xizhou Zhu, Yu Qiao, Jifeng Dai, Yuntao Chen*
- 252 Voyaging into Perpetual Dynamic Scenes from a Single View, *Fengrui Tian, Tianjiao Ding, Jinqi Luo, Hancheng Min, Rene Vidal*
- 253 PacGDC: Label-Efficient Generalizable Depth Completion with Projection Ambiguity and Consistency, *Haotian Wang, Aoran Xiao, Xiaoqin Zhang, Meng Yang, Shijian Lu*
- 254 ARMO: Autoregressive Rigging for Multi-Category Objects, *Mingze Sun, Shiwei Mao, Keyi Chen, Yurun Chen, Shunlin Lu, Jingbo Wang, Junting Dong, Ruqi Huang*
- 255 DCHM: Depth-Consistent Human Modeling for Multiview Detection, *Jiahao Ma, Tianyu Wang, Miaomiao Liu, David Ahmed-Aristizabal, Chuong Nguyen*
- 256 Completing 3D Partial Assemblies with View-Consistent 2D-3D Correspondence, *Weihao Wang, Yu Lan, Mingyu You, Bin He*
- 257 GSV3D: Gaussian Splatting-based Geometric Distillation with Stable Video Diffusion for Single-Image 3D Object Generation, *Ye Tao, Jiawei Zhang, Yahao Shi, Dongqing Zou, Bin Zhou*
- 258 Learnable Feature Patches and Vectors for Boosting Low-light Image Enhancement without External Knowledge, *Xiaogang Xu, Jiafei Wu, Qingsen Yan, Jiequan Cui, Richang Hong, Bei Yu*
- 259 InstaScene: Towards Complete 3D Instance Decomposition and Reconstruction from Cluttered Scenes, *Zesong Yang, Bangbang Yang, Wenqi Dong, Chenxuan Cao, Liyuan Cui, Yuewen Ma, Zhaopeng Cui, Hujun Bao*
- 260 TESPEC: Temporally-Enhanced Self-Supervised Pretraining for Event Cameras, *Mohammad Mohammadi, Ziyi Wu, Igor Gilitschenski*
- 261 SAME: Learning Generic Language-Guided Visual Navigation with State-Adaptive Mixture of Experts, *Gengze Zhou, Yicong Hong, Zun Wang, Chongyang Zhao, Mohit Bansal, Qi Wu*
- 262 CL-Splats: Continual Learning of Gaussian Splatting with Local Optimization, *Jan Ackermann, Jonas Kulhanek, Shengqu Cai, Haoqi Xu, Marc Pollefeys, Gordon Wetzstein, Leonidas J. Guibas, Songyou Peng*
- 263 Find Any Part in 3D, *Ziqi Ma, Yisong Yue, Georgia Gkioxari*
- 264 Learning 3D Scene Analogies with Neural Contextual Scene Maps, *Junho Kim, Gwangtak Bae, Eun Sun Lee, Young Min Kim*
- 265 GausSim: Foreseeing Reality by Gaussian Simulator for Elastic Objects, *Yidi Shao, Mu Huang, Chen Change Loy, Bo Dai*

- 266 Enhancing Spatial Reasoning in Multimodal Large Language Models through Reasoning-based Segmentation, *Zhenhua Ning, Zhuotao Tian, Shaoshuai Shi, Guangming Lu, Daojing He, Wenjie Pei, Li Jiang*
- 267 M-SpecGene: Generalized Foundation Model for RGBT Multispectral Vision, *Kailai Zhou, Fuqiang Yang, Shixian Wang, Bihan Wen, Chongde Zi, Linsen Chen, Qiu Shen, Xun Cao*
- 268 Cross-modal Ship Re-Identification via Optical and SAR Imagery: A Novel Dataset and Method, *Han Wang, Shengyang Li, Jian Yang, Yuxuan Liu, Yixuan Lv, Zhuang Zhou*
- 269 Global Motion Corresponder for 3D Point-Based Scene Interpolation under Large Motion, *Junru Lin, Chirag Vashist, Mikaela Angelina Uy, Colton Stearns, Xuan Luo, Leonidas Guibas, Ke Li*
- 270 AdvDreamer Unveils: Are Vision-Language Models Truly Ready for Real-World 3D Variations?, *Shouwei Ruan, Hanqing Liu, Yao Huang, Xiaoqi Wang, Caixin Kang, Hang Su, Yinpeng Dong, Xingxing Wei*
- 271 SpikeDiff: Zero-shot High-Quality Video Reconstruction from Chromatic Spike Camera and Sub-millisecond Spike Streams, *Siqi Yang, Jinxiu Liang, Zhaojun Huang, Yeliduosi Xiaokaiti, Yakun Chang, Zhaofei Yu, Boxin Shi*
- 272 VA-MoE: Variables-Adaptive Mixture of Experts for Incremental Weather Forecasting, *Hao Chen, Han Tao, Guo Song, Jie Zhang, Yonghan Dong, Yunlong Yu, Lei Bai*
- 273 AJAHR: Amputated Joint Aware 3D Human Mesh Recovery, *Hyunjin Cho, Giyun Choi, Jongwon Choi*
- 274 Event-aided Dense and Continuous Point Tracking: Everywhere and Anytime, *Zhexiong Wan, Jianqin Luo, Yuchao Dai, Gim Hee Lee*
- 275 EquiCaps: Predictor-Free Pose-Aware Pre-Trained Capsule Networks, *Athinoula Konstantinou, Georgios Leontidis, Mamatha Thota, Aiden Durrant*
- 276 A Structure-aware and Motion-adaptive Framework for 3D Human Pose Estimation with Mamba, *Ye Lu, Jie Wang, Jianjun Gao, Rui Gong, Chen Cai, Kim-Hui Yap*
- 277 Learning Normal Flow Directly From Events, *Dehao Yuan, Levi Burner, Jiayi Wu, Minghui Liu, Jingxi Chen, Yiannis Aloimonos, Cornelia Fermüller*
- 278 Unsupervised Joint Learning of Optical Flow and Intensity with Event Cameras, *Shuang Guo, Friedhelm Hamann, Guillermo Gallego*
- 279 OV-SCAN: Semantically Consistent Alignment for Novel Object Discovery in Open-Vocabulary 3D Object Detection, *Adrian Chow, Evelien Riddell, Yimu Wang, Sean Sedwards, Krzysztof Czarnecki*
- 280 CAPTURE: Evaluating Spatial Reasoning in Vision Language Models via Occluded Object Counting, *Atin Pothiraj, Elias Stengel-Eskin, Jaemin Cho, Mohit Bansal*
- 281 RoboTron-Drive: All-in-One Large Multimodal Model for Autonomous Driving, *Zhijian Huang, Chengjian Feng, Feng Yan, Baihui Xiao, Zequn Jie, Yujie Zhong, Xiaodan Liang, Lin Ma*
- 282 Trial-Oriented Visual Rearrangement, *Yuyi Liu, Xinhong Song, Tianliang Qi, Shuqiang Jiang*
- 283 6DOPE-GS: Online 6D Object Pose Estimation using Gaussian Splatting, *Yufeng Jin, Vignesh Prasad, Snehal Jauhri, Mathias Franzius, Georgia Chalvatzaki*
- 284 AnyCalib: On-Manifold Learning for Model-Agnostic Single-View Camera Calibration, *Javier Tirado-Garín, Javier Civera*
- 285 Background Invariance Testing According to Semantic Proximity, *Zukang Liao, Min Chen*
- 286 NuPlanQA: A Large-Scale Dataset and Benchmark for Multi-View Driving Scene Understanding in Multi-Modal Large Language Models, *Sung-Yeon Park, Can Cui, Yunsheng Ma, Ahmadreza Moradipari, Rohit Gupta, Kyungtae Han, Ziran Wang*
- 287 One Look is Enough: Seamless Patchwise Refinement for Zero-Shot Monocular Depth Estimation on High-Resolution Images, *Byeongjun Kwon, Munchul Kim*
- 288 Adapting Vehicle Detectors for Aerial Imagery to Unseen Domains with Weak Supervision, *Xiao Fang, Minhyek Jeon, Zheyang Qin, Stanislav Panev, Celso De Melo, Shuowen Hu, Shayok Chakraborty, Fernando De La Torre*
- 289 RegGS: Unposed Sparse Views Gaussian Splatting with 3DGS Registration, *Chong Cheng, Yu Hu, Sicheng Yu, Beizhen Zhao, Zijian Wang, Hao Wang*
- 290 PersPose: 3D Human Pose Estimation with Perspective Encoding and Perspective Rotation, *Xiaoyang Hao, Han Li*
- 291 Move to Understand a 3D Scene: Bridging Visual Grounding and Exploration for Efficient and Versatile Embodied Navigation, *Ziyu Zhu, Xilin Wang, Yixuan Li, Zhuofan Zhang, Xiaojian Ma, Yixin Chen, Baoxiong Jia, Wei Liang, Qian Yu, Zhidong Deng, Siyuan Huang, Qing Li*
- 292 Training-free Generation of Temporally Consistent Rewards from VLMs, *Yinuo Zhao, Jiale Yuan, Zhiyuan Xu, Xiaoshuai Hao, Xinyi Zhang, Kun Wu, Zhengping Che, Chi Harold Liu, Jian Tang*
- 293 LGA-Net: Learning Local and Global Affinities for Sparse Scribble based Image Colorization, *Hongjin Lyu, Bo Li, Paul L. Rosin, Yu-Kun Lai*
- 294 COBL: Toward Zero-Shot Ordinal Layering without User Prompting, *Aneel Damaraju, Dean Hazineh, Todd Zickler*
- 295 Hierarchical Material Recognition from Local Appearance, *Matthew Beveridge, Shree K. Nayar*
- 296 Correspondence as Video: Test-Time Adaption on SAM2 for Reference Segmentation in the Wild, *Haoran Wang, Zekun Li, Jian Zhang, Lei Qi, Yinghuan Shi*
- 297 MEMFOT: High-Resolution Training for Memory-Efficient Multi-Frame Optical Flow Estimation, *Vladislav Bargatin, Egor Chistov, Alexander Yakovenko, Dmitriy Vatolin*
- 298 Breaking Rectangular Shackles: Cross-View Object Segmentation for Fine-Grained Object Geo-Localization, *Qingwang Zhang, Yingying Zhu*
- 299 MVGBench: a Comprehensive Benchmark for Multi-view Generation Models, *Xianghui Xie, Jan Eric Lessen, Gerard Pons-Moll*
- 300 Harnessing Input-Adaptive Inference for Efficient VLN, *Dongwoo Kang, Akhil Perincherry, Zachary Coalson, Aiden Gabriel, Stefan Lee, Sanghyun Hong*
- 301 From One to More: Contextual Part Latents for 3D Generation, *Shaocong Dong, Lihe Ding, Xiao Chen, Yaokun Li, Yuxin Wang, Yucheng Wang, Qi Wang, Jaehyeok Kim, Chenjian Gao, Zhanpeng Huang, Zibin Wang, Tianfan Xue, Dan Xu*
- 302 TopicGeo: An Efficient Unified Framework for Geolocation, *Xin Wang, Xinlin Wang, Shuiping Gou*
- 303 Importance-Based Token Merging for Efficient Image and Video Generation, *Haoyu Wu, Jingyi Xu, Hieu Le, Dimitris Samaras*
- 304 Knowledge Distillation for Learned Image Compression, *Yunuo Chen, Zezheng Lyu, Bing He, Ning Cao, Gang Chen, Guo Lu, Wenjun Zhang*
- 305 MonoFusion: Sparse-View 4D Reconstruction via Monocular Fusion, *Zihan Wang, Jeff Tan, Tarasha Khurana, Neehar Peri, Deva Ramanan*
- 306 ETCH: Generalizing Body Fitting to Clothed Humans via Equivariant Tightness, *Boqian Li, Haiwen Feng, Zeyu Cai, Michael J. Black, Yuliang Xiu*
- 307 Revisiting Image Fusion for Multi-Illuminant White-Balance Correction, *David Serrano-Lozano, Aditya Arora, Luis Herranz, Konstantinos G. Derpanis, Michael S. Brown, Javier Vazquez-Corral*
- 308 Stronger, Steadier & Superior: Geometric Consistency in Depth VFM Forges Domain Generalized Semantic Segmentation, *Siyu Chen, Ting Han, Changshe Zhang, Xin Luo, Meiliu Wu, Guorong Cai, Jinhe Su*
- 309 Partially Matching Submap Helps: Uncertainty Modeling and Propagation for Text to Point Cloud Localization, *Mingtao Feng, Longlong Mei, Zijie Wu, Jianqiao Luo, Fenghao Tian, Jie Feng, Weisheng Dong, Yaonan Wang*
- 310 SAS: Segment Any 3D Scene with Integrated 2D Priors, *Zhuoyuan Li, Jiahao Lu, Jiacheng Deng, Hanzhi Chang, Lifan Wu, Yanzhe Liang, Tianzhu Zhang*
- 311 Medical World Model, *Yijun Yang, Zhao-Yang Wang, Qiuping Liu, Shuwen Sun, Kang Wang, Rama Chellappa, Zongwei Zhou, Alan Yuille, Lei Zhu, Yu-Dong Zhang, Jieneng Chen*
- 312 NormalCrafter: Learning Temporally Consistent Normals from Video Diffusion Priors, *Yanrui Bin, Wenbo Hu, Haoyuan Wang, Xinya Chen, Bing Wang*
- 313 MATE: Motion-Augmented Temporal Consistency for Event-based Point Tracking, *Han Han, Wei Zhai, Yang Cao, Bin Li, Zheng-jun Zha*

- 314 ObjectGS: Object-aware Scene Reconstruction and Scene Understanding via Gaussian Splatting, *Ruijie Zhu, Mulin Yu, Linning Xu, Lihan Jiang, Yixuan Li, Tianzhu Zhang, Jiangmiao Pang, Bo Dai*
- 315 DuCos: Duality Constrained Depth Super-Resolution via Foundation Model, *Zhiqiang Yan, Zhengxue Wang, Haoye Dong, Jun Li, Jian Yang, Gim Hee Lee*
- 316 MaskHand: Generative Masked Modeling for Robust Hand Mesh Reconstruction in the Wild, *Muhammad Usama Saleem, Ekkasit Pinyoanuntapong, Mayur Jagdishbhai Patel, Hongfei Xue, Ahmed Helmy, Srijan Das, Pu Wang*
- 317 OpenRSD: Towards Open-prompts for Object Detection in Remote Sensing Images, *Ziyue Huang, Yongchao Feng, Ziqi Liu, Shuai Yang, Qingjie Liu, Yunhong Wang*
- 318 Passing the Driving Knowledge Test, *Maolin Wei, Wanzhou Liu, Eshed Ohn-Bar*
- 319 Uncertainty-Aware Gradient Stabilization for Small Object Detection, *Huixin Sun, Yanjing Li, Linlin Yang, Xianbin Cao, Baochang Zhang*
- 320 Learning 3D Object Spatial Relationships from Pre-trained 2D Diffusion Models, *Sangwon Baik, Hyeonwoo Kim, Hanbyul Joo*
- 321 Can Generative Geospatial Diffusion Models Excel as Discriminative Geospatial Foundation Models?, *Yuru Jia, Valerio Marsocci, Ziyang Gong, Xue Yang, Maarten Vergauwen, Andrea Nascetti*
- 322 Towards Annotation-Free Evaluation: KPAScore for Human Keypoint Detection, *Xiaoxiao Wang, Chunxiao Li, Peng Sun, Boming Miao, Yunjian Zhang, Yao Zhu*
- 323 4D Visual Pre-training for Robot Learning, *Chengkai Hou, Yanjie Ze, Yankai Fu, Zeyu Gao, Songbo Hu, Yue Yu, Shanghang Zhang, Huazhe Xu*
- 324 CryoFastAR: Fast Cryo-EM Ab initio Reconstruction Made Easy, *Jiakai Zhang, Shouchen Zhou, Haizhao Dai, Xinhang Liu, Peihao Wang, Zhiwen Fan, Yuan Pei, Jingyi Yu*
- 325 Beyond Pixel Uncertainty: Bounding the OoD Objects in Road Scenes, *Huachao Zhu, Zelong Liu, Zhichao Sun, Yuda Zou, Gui-Song Xia, Yongchao Xu*
- 326 HoliTracer: Holistic Vectorization of Geographic Objects from Large-Size Remote Sensing Imagery, *Yu Wang, Bo Dang, Wanchun Li, Wei Chen, Yansheng Li*
- 327 Generative Zoo, *Tomasz Niewiadomski, Anastasios Yiannakidis, ☆ Hanz Cuevas-Velasquez, Soubhik Sanyal, Michael J. Black, Silvia Zuffi, Peter Kulits*
- 328 St4RTrack: Simultaneous 4D Reconstruction and Tracking in the World, *Haiwen Feng, Junyi Zhang, Qianqian Wang, Yufei Ye, Pengcheng Yu, Michael J. Black, Trevor Darrell, Angjoo Kanazawa*
- 329 DialNav: Multi-turn Dialog Navigation with a Remote Guide, *Leekyeung Han, Hyunji Min, Gyeom Hwangbo, Jonghyun Choi, Paul Hongsuck Seo*
- 330 Event-guided Unified Framework for Low-light Video Enhancement, Frame Interpolation, and Deblurring, *Taewoo Kim, Kuk-Jin Yoon*
- 331 Aether: Geometric-Aware Unified World Modeling, *Haoyi Zhu, Yifan Wang, Jianjun Zhou, Wenzheng Chang, Yang Zhou, Zizun Li, Junyi Chen, Chunhua Shen, Jiangmiao Pang, Tong He*
- 332 PARTE: Part-Guided Texturing for 3D Human Reconstruction from a Single Image, *Hyeongjin Nam, Donghwan Kim, Gyeongsik Moon, Kyoung Mu Lee*
- 333 PLMP - Point-Line Minimal Problems for Projective SfM, *☆ Kim Kiehn, Albin Ahlback, Kathlén Kohn*
- 334 PS-Mamba: Spatial-Temporal Graph Mamba for Pose Sequence Refinement, *Haoye Dong, Gim Hee Lee*
- 335 TaxaDiffusion: Progressively Trained Diffusion Model for Fine-Grained Species Generation, *Amin Karimi Monsefi, Mridul Khurana, Rajiv Ramnath, Anuj Karpatne, Wei-Lun Chao, Cheng Zhang*
- 336 Dynamic Reconstruction of Hand-Object Interaction with Distributed Force-aware Contact Representation, *Zhenjun Yu, Wenqiang Xu, Pengfei Xie, Yutong Li, Brian W. Anthony, Zhuorui Zhang, Cewu Lu*
- 337 VLM4D: Towards Spatiotemporal Awareness in Vision Language Models, *☆ Shijie Zhou, Alexander Vilesov, Xuehai He, Ziyu Wan, Shuwang Zhang, Aditya Nagachandra, Di Chang, Dongdong Chen, Xin Eric Wang, Achuta Kadambi*
- 338 Aligning Constraint Generation with Design Intent in Parametric CAD, *Evan Casey, Tianyu Zhang, Shu Ishida, John Roger Thompson, Amir Khasahmadi, Joseph George Lambourne, Pradeep Kumar Jayaraman, Karl D.D. Willis*
- 339 Spatial Alignment and Temporal Matching Adapter for Video-Radar Remote Physiological Measurement, *Qian Liang, Ruixu Geng, Jinbo Chen, Haoyu Wang, Yan Chen, Yang Hu*
- 340 Bias in Gender Bias Benchmarks: How Spurious Features Distort Evaluation, *Yusuke Hirota, Ryo Hachiuma, Boyi Li, Ximing Lu, Michael Ross Boone, Boris Ivanovic, Yejin Choi, Marco Pavone, Yu-Chiang Frank Wang, Noa Garcia, Yuta Nakashima, Chao-Han Huck Yang*
- 341 AGO: Adaptive Grounding for Open World 3D Occupancy Prediction, *Peizheng Li, Shuxiao Ding, You Zhou, Qingwen Zhang, Onat Inak, Larissa Triess, Niklas Hanselmann, Marius Cordts, Andreas Zell*
- 342 Efficient Event Camera Data Pretraining with Adaptive Prompt Fusion, *Quanmin Liang, Qiang Li, Shuai Liu, Xinzi Cao, Jinyi Lu, Feidiao Yang, Wei Zhang, Kai Huang, Yonghong Tian*
- 343 Unsupervised Part Discovery via Descriptor-Based Masked Image Restoration with Optimized Constraints, *Jiahao Xia, Yike Wu, Wenjian Huang, Jianguo Zhang, Jian Zhang*
- 344 Environment-Agnostic Pose: Generating Environment-independent Object Representations for 6D Pose Estimation, *Shaobo Zhang, Yuhang Huang, Wangqing Zhao, Wei Zhao, Ziyu Guan, Jinye Peng*
- 345 OpenM3D: Open Vocabulary Multi-view Indoor 3D Object Detection without Human Annotations, *Peng-Hao Hsu, Ke Zhang, Fu-En Wang, Tao Tu, Ming-Feng Li, Yu-Lun Liu, Albert Y. C. Chen, Min Sun, Cheng-Hao Kuo*
- 346 CATP-LLM: Empowering Large Language Models for Cost-Aware Tool Planning, *Duo Wu, Jinghe Wang, Yuan Meng, Yanning Zhang, Le Sun, Zhi Wang*
- 347 Online Dense Point Tracking with Streaming Memory, *Qiaole Dong, Yanwei Fu*
- 348 TimeFormer: Capturing Temporal Relationships of Deformable 3D Gaussians for Robust Reconstruction, *Dadong Jiang, Zhi Hou, Zhihui Ke, Xianghui Yang, Xiaobo Zhou, Tie Qiu*
- 349 Long-Context State-Space Video World Models, *Ryan Po, ☆ Yotam Nitzan, Richard Zhang, Berlin Chen, Tri Dao, Eli Shechtman, Gordon Wetzstein, Xun Huang*
- 350 MaGS: Reconstructing and Simulating Dynamic 3D Objects with Mesh-adsorbed Gaussian Splatting, *☆ Shaojie Ma, Yawei Luo, Wei Yang, Yi Yang*
- 351 When Schrödinger Bridge Meets Real-World Image Dehazing with Unpaired Training, *Yunwei Lan, Zhigao Cui, Xin Luo, Chang Liu, Nian Wang, Menglin Zhang, Yanzhao Su, Dong Liu*
- 352 GM-MoE: Low-Light Enhancement with Gated-Mechanism Mixture-of-Experts, *Minwen Liao, Haobo Dong, Xinyi Wang, Kurban Ubul, Yihua Shao, Ziyang Yan*
- 353 CHARM3R: Towards Unseen Camera Height Robust Monocular 3D Detector, *Abhinav Kumar, Yuliang Guo, Zhihao Zhang, Xinyu Huang, Liu Ren, Xiaoming Liu*
- 354 CWNet: Causal Wavelet Network for Low-Light Image Enhancement, *Tongshun Zhang, Pingping Liu, Yubing Lu, Mengen Cai, Zijian Zhang, Zhe Zhang, Qiuzhan Zhou*
- 355 MonoMobility: Zero-Shot 3D Mobility Analysis from Monocular Videos, *Hongyi Zhou, Yulan Guo, Xiaogang Wang, Kai Xu*
- 356 Test-Time Retrieval-Augmented Adaptation for Vision-Language Models, *Xinqi Fan, Xueli Chen, Luoxiao Yang, Chuin Hong Yap, Rizwan Qureshi, Qi Dou, Moi Hoon Yap, Mubarak Shah*
- 357 TRACE: Learning 3D Gaussian Physical Dynamics from Multi-view Videos, *Jinxi Li, Ziyang Song, Bo Yang*
- 358 RnGCam: High-speed video from rolling & global shutter measurements, *Kevin Tandi, Xiang Dai, Chinmay Talegaonkar, Gal Mishne, Nick Antipa*
- 359 SplArt: Articulation Estimation and Part-Level Reconstruction with 3D Gaussian Splatting, *Shengjie Lin, Jiading Fang, Muhammad Zubair Irshad, Vitor Campagnolo Guizilini, Rares Andrei Ambrus, Greg Shakhnarovich, Matthew R. Walter*
- 360 FuXi-RTM: A Physics-Guided Prediction Framework with Radiative Transfer Modeling, *☆ Qiusheng Huang, Xiaohui Zhong, Xu Fan, Hao Li*

- 361 Self-Supervised Monocular 4D Scene Reconstruction for Egocentric Videos, *Chengbo Yuan, Geng Chen, Li Yi, Yang Gao*
- 362 IDEATOR: Jailbreaking and Benchmarking Large Vision-Language Models Using Themselves, *Ruofan Wang, Juncheng Li, Yixu Wang, Bo Wang, Xiaosen Wang, Yan Teng, Yingchun Wang, Xingjun Ma, Yu-Gang Jiang*
- 363 3DGraphLLM: Combining Semantic Graphs and Large Language Models for 3D Scene Understanding, *Tatiana Zemskova, Dmitry Yudin*
- 364 Diorama: Unleashing Zero-shot Single-view 3D Indoor Scene
* Modeling, *Qirui Wu, Denys Iliash, Daniel Ritchie, Manolis Savva, Angel X. Chang*
- 365 Bokehlicious: Photorealistic Bokeh Rendering with Controllable
* Apertures, *Tim Seizinger, Florin-Alexandru Vasluianu,*
☆ *Marcos V. Conde, Zongwei Wu, Radu Timofte*
- 366 SuperEvent: Cross-Modal Learning of Event-based Keypoint
* Detection for SLAM, *Yannick Burkhardt, Simon Schaefer, Stefan Leutenegger*
- 367 High-Resolution Spatiotemporal Modeling with Global-Local State Space Models for Video-Based Human Pose Estimation, *Runyang Feng, Hyung Jin Chang, Tze Ho Elden Tse, Boeun Kim, Yi Chang, Yixing Gao*
- 368 Learning on the Go: A Meta-learning Object Navigation Model, *Xiaorong Qin, Xinhang Song, Sixian Zhang, Xinyao Yu, Xinmiao Zhang, Shuqiang Jiang*
- 369 PASG: A Closed-Loop Framework for Automated Geometric Primitive Extraction and Semantic Anchoring in Robotic Manipulation, *Zhihao Zhu, Yifan Zheng, Siyu Pan, Yaohui Jin, Yao Mu*
- 370 Unlocking Constraints: Source-Free Occlusion-Aware Seamless Segmentation, *Yihong Cao, Jiaming Zhang, Xu Zheng, Hao Shi, Kunyu Peng, Hang Liu, Kailun Yang, Hui Zhang*
- 371 Kestrel: 3D Multimodal LLM for Part-Aware Grounded Description, *Mahmoud Ahmed, Junjie Fei, Jian Ding, Eslam Mohamed Bakr, Mohamed Elhoseiny*
- 372 ProGait: A Multi-Purpose Video Dataset and Benchmark for
* Transfemoral Prosthesis Users, *Xiangyu Yin, Boyuan Yang, Weichen Liu, Qiyao Xue, Abrar Alamri, Goeran Fiedler, Wei Gao*
- 373 HFD-Teacher: High-Frequency Depth Distillation from Depth Foundation Models for Enhanced Depth Completion, *Zhiyuan Yang, Anqi Cheng, Haiyue Zhu, Tianjiao Li, Pey Yuen Tao, Kezhi Mao*
- 374 Detection, Pose Estimation and Segmentation for Multiple Bodies: Closing the Virtuous Circle, *Miroslav Purkrabek, Jiri Matas*
- 375 Benchmarking Multimodal Large Language Models Against Image Corruptions, *Xinkuan Qiu, Meina Kan, Yongbin Zhou, Shiguang Shan*
- 376 MixRI: Mixing Features of Reference Images for Novel Object Pose Estimation, *Xinhang Liu, Jiawei Shi, Zheng Dang, Yuchao Dai*
- 377 INS-MMBench: A Comprehensive Benchmark for Evaluating LVLMS' Performance in Insurance, *Chenwei Lin, Hanjia Lyu, Xian Xu, Jiebo Luo*
- 378 ReassembleNet: Learnable Keypoints and Diffusion for 2D
☆ Fresco Reconstruction, *Adeela Islam, Stefano Fiorini, Stuart James, Pietro Morerio, Alessio Del Bue*
- 379 SITE: towards Spatial Intelligence Thorough Evaluation, *Wenqi Wang, Reuben Tan, Pengyue Zhu, Jianwei Yang, Zhengyuan Yang, Lijuan Wang, Andrey Kolobov, Jianfeng Gao, Boqing Gong*
- 380 PASD: A Pixel-Adaptive Swarm Dynamics Approach for Unsupervised Low-Light Image Enhancement, *Shuai Jin, Yuhua Qian, Feijiang Li, Guoqing Liu, Xinyan Liang*
- 381 EVER: Exact Volumetric Ellipsoid Rendering for Real-time View
○ Synthesis, *Alexander Mai, Peter Hedman, George Kopanas, Dor Verbin, David Futschik, Qiangeng Xu, Falko Kuester, Jonathan T. Barron, Yinda Zhang*
- 382 Variance-Based Pruning for Accelerating and Compressing
○ Trained Networks, *Uranik Berisha, Jens Mehnert, Alexandru Paul Condurache*
- 383 WonderPlay: Dynamic 3D Scene Generation from a Single Image and
* Actions, *Zizhang Li, Hong-Xing Yu, Wei Liu, Yin Yang, Charles Herrmann, Gordon Wetzstein, Jiajun Wu*
- 384 Beyond the Destination: A Novel Benchmark for Exploration-Aware Embodied Question Answering, *Kaixuan Jiang, Yang Liu, Weixing Chen, Jingzhou Luo, Ziliang Chen, Ling Pan, Guanbin Li,*
Liang Lin
- 385 RayPose: Ray Bundling Diffusion for Template Views in Unseen 6D Object Pose Estimation, *Junwen Huang, Shishir Reddy Vutukur, Peter KT Yu, Nassir Navab, Slobodan Ilic, Benjamin Busam*
- 386 Not all Views are Created Equal: Analyzing Viewpoint Instabilities in Vision Foundation Models, *Mateusz Michalkiewicz, Sheena Bai, Mahsa Baktashmotlagh, Varun Jampani, Guha Balakrishnan*
- 387 CHROME: Clothed Human Reconstruction with Occlusion-
* Resilience and Multiview-Consistency from a Single Image,
☆ *Arindam Dutta, Meng Zheng, Zhongpai Gao, Benjamin Planche, Anwesa Choudhuri, Terrence Chen, Amit K. Roy-Chowdhury, Ziyang Wu*
- 388 SkySense V2: A Unified Foundation Model for Multi-modal Remote Sensing, *Yingying Zhang, Lixiang Ru, Kang Wu, Lei Yu, Lei Liang, Yansheng Li, Jingdong Chen*
- 389 ReCoT: Reflective Self-Correction Training for Mitigating Confirmation Bias in Large Vision-Language Models, *Mengxue Qu, Yibo Hu, Kunyang Han, Yunchao Wei, Yao Zhao*
- 390 Easi3R: Estimating Disentangled Motion from DUST3R Without
☆ Training, *Xingyu Chen, Yue Chen, Yuliang Xiu, Andreas Geiger, Anpei Chen*
- 391 PRE-Mamba: A 4D State Space Model for Ultra-High-Frequent Event Camera Deraining, *Ciyu Ruan, Ruishan Guo, Zihang Gong, Jingao Xu, Wenhan Yang, Xinlei Chen*
- 392 Amodal3R: Amodal 3D Reconstruction from Occluded 2D Images,
☆ *Tianhao Wu, Chuanxia Zheng, Frank Guan, Andrea Vedaldi, Tat-Jen Cham*
- 393 GenHaze: Pioneering Controllable One-Step Realistic Haze Generation for Real-World Dehazing, *Sixiang Chen, Tian Ye, Yunlong Lin, Yeying Jin, Yijun Yang, Haoyu Chen, Jianyu Lai, Song Fei, Zhaohu Xing, Fugee Tsung, Lei Zhu*
- 394 When Large Vision-Language Model Meets Large Remote Sensing Imagery: Coarse-to-Fine Text-Guided Token Pruning, *Junwei Luo, Yingying Zhang, Xue Yang, Kang Wu, Qi Zhu, Lei Liang, Jingdong Chen, Yansheng Li*
- 395 After the Party: Navigating the Mapping From Color to Ambient
☆ Lighting, *Florin-Alexandru Vasluianu, Tim Seizinger, Zongwei Wu, Radu Timofte*
- 396 Harnessing Uncertainty-aware Bounding Boxes for Unsupervised 3D Object Detection, *Ruiyang Zhang, Hu Zhang, Zhedong Zheng*
- 397 Perspective-Aware Reasoning in Vision-Language Models via Mental Imagery Simulation, *Phillip Y. Lee, Jihyeon Je, Chanh Park, Mikaela Angelina Uy, Leonidas Guibas, Minhyuk Sung*
- 398 3D Gaussian Map with Open-Set Semantic Grouping for Vision-Language Navigation, *Jianzhe Gao, Rui Liu, Wenguan Wang*
- 399 GWM: Towards Scalable Gaussian World Models for Robotic Manipulation, *Guanxing Lu, Baoxiong Jia, Puhao Li, Yixin Chen, Ziwei Wang, Yansong Tang, Siyuan Huang*
- 400 Ross3D: Reconstructive Visual Instruction Tuning with 3D-Awareness, *Haochen Wang, Yucheng Zhao, Tiancai Wang, Haoqiang Fan, Xiangyu Zhang, Zhaoxiang Zhang*
- 401 OMNI-DC: Highly Robust Depth Completion with Multiresolution Depth Integration, *Yiming Zuo, Willow Yang, Zeyu Ma, Jia Deng*
- 402 BlinkTrack: Feature Tracking over 80 FPS via Events and Images, *Yichen Shen, Yijin Li, Shuo Chen, Guanglin Li, Zhaoyang Huang, Hujun Bao, Zhaopeng Cui, Guofeng Zhang*
- 403 GEO: Geometrically Consistent Embedding with Lightspeed Inference, *Regine Hartwig, Dominik Muhle, Riccardo Marin, Daniel Cremers*
- 404 Can3Tok: Canonical 3D Tokenization and Latent Modeling of Scene-Level 3D Gaussians, *Quankai Gao, Iliyan Georgiev, Tuanfeng Y. Wang, Krishna Kumar Singh, Ulrich Neumann, Jae Shin Yoon*
- 405 GEOPARD: Geometric Pretraining for Articulation Prediction in 3D Shapes, *Pradyumn Goyal, Dmitry Petrov, Sheldon Andrews, Yizhak Ben-Shabat, Hsueh-Ti Derek Liu, Evangelos Kalogerakis*
- 406 EMoTive: Event-guided Trajectory Modeling for 3D Motion Estimation, *Zengyu Wan, Wei Zhai, Yang Cao, Zhengjun Zha*
- 407 Dream-to-Recon: Monocular 3D Reconstruction with Diffusion-Depth Distillation from Single Images, *Philipp Wulff, Felix Wimbauer, Dominik Muhle, Daniel Cremers*
- 408 FusionPhys: A Flexible Framework for Fusing Complementary Sensing Modalities in Remote Physiological Measurement,

- Chenhang Ying, Huiyu Yang, Jieyi Ge, Zhaodong Sun, Xu Cheng, Kui Ren, Xiaobai Li
- 409 BoxDreamer: Dreaming Box Corners for Generalizable Object Pose Estimation, Yuanhong Yu, Xingyi He, Chen Zhao, Junhao Yu, Jiaqi Yang, Ruizhen Hu, Yujun Shen, Xing Zhu, Xiaowei Zhou, Sida Peng
- 410 Fine-grained Spatiotemporal Grounding on Egocentric Videos, Shuo Liang, Yiwu Zhong, Zi-Yuan Hu, Yeyao Tao, Liwei Wang
- 411 Deep Space Weather Model: Long-Range Solar Flare Prediction from Multi-Wavelength Images, Shunya Nagashima, Komei Sugiura
- 412 From Sharp to Blur: Unsupervised Domain Adaptation for 2D
- ☆ Human Pose Estimation Under Extreme Motion Blur Using Event Cameras, Youngho Kim, Hoonhee Cho, Kuk-Jin Yoon
- 413 CMB-ML: A Cosmic Microwave Background Dataset for the Oldest Possible Computer Vision Task, James Amato, Yunan Xie, Leonel Medina-Varela, Ammar Aljerwi, Adam McCutcheon, T. Seth Rippentrop, Kristian Gonzalez, Jacques Delabrouille, Mustapha Ishak, Nicholas Ruozzi
- 414 Fine-Grained Evaluation of Large Vision-Language Models in Autonomous Driving, Yue Li, Meng Tian, Zhenyu Lin, Jiangtong Zhu, Dechang Zhu, Haiqiang Liu, Yueyi Zhang, Zhiwei Xiong, Xinhai Zhao
- 415 Test-Time Prompt Tuning for Zero-Shot Depth Completion, *☆ Chanhwi Jeong, Inhwan Bae, Jin-Hwi Park, Hae-Gon Jeon
- 416 Rethinking the Embodied Gap in Vision-and-Language Navigation: A Holistic Study of Physical and Visual Disparities, Liuyi Wang, Xinyuan Xia, Hui Zhao, Hanqing Wang, Tai Wang, Yilun Chen, Chengju Liu, Qijun Chen, Jiangmiao Pang
- 417 Learning to See Inside Opaque Liquid Containers using Speckle Vibrometry, Matan Kichler, Shai Bagon, Mark Sheinin
- 418 monoVLN: Bridging the Observation Gap between Monocular and Panoramic Vision and Language Navigation, Renjie Lu, Yu Zhou, Hao Cheng, Jingke Meng, Wei-Shi Zheng
- 419 Contact-Aware Amodal Completion for Human-Object Interaction * via Multi-Regional Inpainting, Seunggeun Chi, Enna Sachdeva, Pin-Hao Huang, Kwonjoon Lee
- 420 Bring Your Rear Cameras for Egocentric 3D Human Pose Estimation, Hiroyasu Akada, Jian Wang, Vladislav Golyanik, Christian Theobalt
- 421 BokehDiff: Neural Lens Blur with One-Step Diffusion, Chengxuan Zhu, Qingnan Fan, Qi Zhang, Jinwei Chen, Huaqi Zhang, Chao Xu, Boxin Shi
- 422 LocalDyGS: Multi-view Global Dynamic Scene Modeling via Adaptive Local Implicit Feature Decoupling, Jiahao Wu, Rui Peng, Jianbo Jiao, Jiayu Yang, Luyang Tang, Kaiqiang Xiong, Jie Liang, Jinbo Yan, Runling Liu, Ronggang Wang
- 423 Closed-Loop Transfer for Weakly-supervised Affordance Grounding, Jiajin Tang, Zhengxuan Wei, Ge Zheng, Sibe Yang
- 424 Combinative Matching for Geometric Shape Assembly, * Nahyuk Lee, Juhong Min, Junhong Lee, Chunghyun Park, Minsu Cho
- 425 CogNav: Cognitive Process Modeling for Object Goal Navigation with LLMs, Yihan Cao, Jiazhao Zhang, Zhinan Yu, Shuzhen Liu, Zheng Qin, Qin Zou, Bo Du, Kai Xu
- 426 DyGS-SLAM: Real-Time Accurate Localization and Gaussian Reconstruction for Dynamic Scenes, Xinggang Hu, Chenyangguang Zhang, Mingyuan Zhao, Yuanze Gui, Xiangkui Zhang, Xiangyang Ji
- 427 Teaching VLMs to Localize Specific Objects from In-context Examples, Sivan Doveh, Nimrod Shabtay, Eli Schwartz, Hilde Kuehne, Raja Giryes, Rogerio Feris, Leonid Karlinsky, James Glass, Assaf Arbelle, Shimon Ullman, M. Jehanzeb Mirza
- 428 Dark-ISP: Enhancing RAW Image Processing for Low-Light Object Detection, Jiasheng Guo, Xin Gao, Yuxiang Yan, Guanghao Li, Jian Pu
- 429 Parameter-Efficient Adaptation of Geospatial Foundation Models through Embedding Deflection, Romain Thoreau, Valerio Marsocci, Dawa Derksen
- 430 MeasureXpert: Automatic Anthropometric Measurement Extraction from Two Unregistered, Partial, Posed, and Dressed Body Scans, Ran Zhao, Xinxin Dai, Pengpeng Hu, Vasile Palade, Adrian Munteanu
- 431 SDFit: 3D Object Pose and Shape by Fitting a Morphable SDF to a Single Image, Dimitrije Antić, Georgios Paschalidis, Shashank Tripathi, Theo Gevers, Sai Kumar Dwivedi, Dimitrios Tzionas
- 432 Details Matter for Indoor Open-vocabulary 3D Instance Segmentation, Sanghun Jung, Jingjing Zheng, Ke Zhang, Nan Qiao, Albert Y. C. Chen, Lu Xia, Chi Liu, Yuyin Sun, Xiao Zeng, Hsiang-Wei Huang, Byron Boots, Min Sun, Cheng-Hao Kuo
- 433 FlashDepth: Real-time Streaming Video Depth Estimation at 2K Resolution, * Gene Chou, Wenqi Xian, Guandao Yang, Mohamed Abdelfattah, Bharath Hariharan, Noah Snively, Ning Yu, Paul Debevec
- 434 Self-supervised Learning of Hybrid Part-aware 3D Representations of 2D Gaussians and Superquadrics, Zhirui Gao, Renjiao Yi, Yuhang Huang, Wei Chen, Chenyang Zhu, Kai Xu
- 435 Shape of Motion: 4D Reconstruction from a Single Video, * Qianqian Wang, Vickie Ye, Hang Gao, Weijia Zeng, Jake Austin, Zhengqi Li, Angjoo Kanazawa
- 436 Amodal Depth Anything: Amodal Depth Estimation in the Wild, Zhenyu Li, Mykola Lavreniuk, Jian Shi, Shariq Farooq Bhat, Peter Wonka
- 437 Training-Free Personalization via Retrieval and Reasoning on Fingerprints, ☆ Deepayan Das, Davide Talon, Yiming Wang, Massimiliano Mancini, Elisa Ricci
- 438 TAPNext: Tracking Any Point (TAP) as Next Token Prediction, Artem Zholus, Carl Doersch, Yi Yang, Skanda Koppula, Viorica Patraucean, Xu Owen He, Ignacio Rocco, Mehdi S. M. Sajjadi, Sarath Chandra, Ross Goroshin
- 439 PartField: Learning 3D Feature Fields for Part Segmentation and Beyond, Minghua Liu, Mikaela Angelina Uy, Donglai Xiang, Hao Su, Sanja Fidler, Nicholas Sharp, Jun Gao
- 440 RetinexMCNet: A Memory Controller Dominated Network for Low-Light Video Enhancement Based on Retinex, Meiao Wang, Xuejing Kang, Yaxi Lu, Jie Xu
- 441 ViewSRD: 3D Visual Grounding via Structured Multi-View Decomposition, Ronggang Huang, Haoxin Yang, Yan Cai, Xuemiao Xu, Huaidong Zhang, Shengfeng He
- 442 GeoProg3D: Compositional Visual Reasoning for City-Scale 3D Language Fields, Shunsuke Yasuki, Taiki Miyanishi, Nakamasa Inoue, Shuhei Kurita, Koya Sakamoto, Daichi Azuma, Masato Taki, Yutaka Matsuo
- 443 Bridging the Sky and Ground: Towards View-Invariant Feature Learning for Aerial-Ground Person Re-Identification, Wajahat Khalid, Bin Liu, Xulin Li, Muhammad Waqas, Muhammad Sher Afgan
- 444 CoA-VLA: Improving Vision-Language-Action Models via Visual-Text Chain-of-Affordance, Jinming Li, Yichen Zhu, Zhibin Tang, Junjie Wen, Minjie Zhu, Xiaoyu Liu, Chengmeng Li, Ran Cheng, Yaxin Peng, Yan Peng, Feifei Feng
- 445 Removing Out-of-Focus Reflective Flares via Color Alignment, Fengbo Lan, Chang Wen Chen
- 446 Proactive Scene Decomposition and Reconstruction, Baicheng Li, ☆ Zike Yan, Dong Wu, Hongbin Zha
- 447 Unified Category-Level Object Detection and Pose Estimation from RGB Images using 3D Prototypes, Tom Fischer, Xiaojie Zhang, Eddy Ilg
- 448 CAD-Recode: Reverse Engineering CAD Code from Point Clouds, Danila Rukhovich, Elona Dupont, Dimitrios Mallis, Kseniya Cherenkova, Anis Kacem, Djamilia Aouada
- 449 EvRT-DETR: Latent Space Adaptation of Image Detectors for Event-based Vision, Dmitrii Torbunov, Yihui Ren, Animesh Ghose, Odera Dim, Yonggang Cui
- 450 A Hyperdimensional One Place Signature to Represent Them All: ☆ Stackable Descriptors For Visual Place Recognition, Connor Malone, Somayeh Hussaini, Tobias Fischer, Michael Milford
- 451 IRASim: A Fine-Grained World Model for Robot Manipulation, Fangqi Zhu, Hongtao Wu, Song Guo, Yuxiao Liu, Chilam Cheang, Tao Kong
- 452 WalkVLM: Aid Visually Impaired People Walking by Vision Language Model, Zhiqiang Yuan, Ting Zhang, Yeshuang Zhu, Jiawei Zhang, Ying Deng, Zexi Jia, Peixiang Luo, Xiaoyue Duan, Jie Zhou, Jinchao Zhang
- 453 CCL-LGS: Contrastive Codebook Learning for 3D Language Gaussian Splatting, Lei Tian, Xiaomin Li, Liqian Ma, Hao Yin, Zirui Zheng,

454 LDPose: Towards Inclusive Human Pose Estimation for Limb-Deficient Individuals in the Wild, *Jiaying Ying, Heming Du, Kaihao Zhang, Lincheng Li, Xin Yu*

15:00 - 17:00 Demos 2 (Exhibit Hall I)

- Notes:**

Wednesday, October 22

7:30 - 17:00	Registration / Badge Pickup (HCC Main Lobby)
7:00 - 17:00	Nursing Pod 
7:00 - 17:00	Prayer / Quiet Room / Mother's Room (326 A)
7:30 - 9:00	Breakfast (Exhibit Hall II)
8:00 - 9:15	Oral Session 3A: Foundation Models and Representation (Exhibit Hall III)
8:00 - 9:15	Oral Session 3B: Human Modeling (Kalakaua Ballroom)
9:15 - 10:15	Coffee Break (Exhibit Hall II)
9:00 - 9:30	Poster Setup (Exhibit Hall I)
9:30 - 10:30	Keynote (Exhibit Hall III)
10:45 - 12:45	Poster Session 3 & Exhibit Hall (Exhibit Hall I)
10:45 - 12:45	Demos 3 (Exhibit Hall I)
11:00 - 13:00	Lunch (Exhibit Hall II / Overflow Seating Rooftop and Third Floor)
11:00 - 13:00	Doctoral Consortium (301 A+ B)
13:00 - 14:15	Oral Session 4A: Vision + Graphics (Exhibit Hall III)
13:00 - 14:15	Oral Session 4B: 3D Pose Understanding (Kalakaua Ballroom)
14:00 - 14:30	Poster Setup (Exhibit Hall I)
14:30 - 16:30	Poster Session 4 & Exhibit Hall with Coffee Break (Exhibit Hall I)
14:30 - 16:30	Demos 4 (Exhibit Hall I)
16:45 - 17:45	PAMI TC Meeting (Exhibit Hall III)
19:00 - 20:00	Reception (Hilton Lawn)

8:00-9:15 Oral Session 3A: Foundation Models and Representation (Exhibit Hall III)

Session Chairs: Bryan A. Plummer, Yujun Cai

 - Award candidate paper  - Outstanding Reviewer

- 1 RS-vHeat: Heat Conduction Guided Efficient Remote Sensing Foundation Model, *Huiyang Hu, Peijin Wang, Hanbo Bi, Boyuan Tong, Zhaozhi Wang, Wenhui Diao, Hao Chang, Yingchao Feng, Ziqi Zhang, Yaowei Wang, Qixiang Ye, Kun Fu, Xian Sun*
- 2  Towards a Unified Copernicus Foundation Model for Earth Vision, *Yi Wang, Zhitong Xiong, Chenying Liu, Adam J. Stewart, Thomas Dujardin, Nikolaos Ioannis Bountos, Angelos Zavras, Franziska Gerken, Ioannis Papoutsis, Laura Leal-Taixé, Xiao Xiang Zhu*
- 3 Learning Streaming Video Representation via Multitask Training, *Yibin Yan, Jilan Xu, Shangzhe Di, Yikun Liu, Yudi Shi, Qirui Chen, Zeqian Li, Yifei Huang, Weidi Xie*
- 4 LoftUp: Learning a Coordinate-Based Feature Upsampler for Vision Foundation Models, *Haiwen Huang, Anpei Chen, Volodymyr Havrylov, Andreas Geiger, Dan Zhang*
- 5 Learning Visual Hierarchies in Hyperbolic Space for Image Retrieval, *Ziwei Wang, Sameera Ramasinghe, Chenchen Xu, Julien Monteil, Loris Bazzani, Thalayasingam Ajanthan*
- 6 GMMamba: Group Masking Mamba for Whole Slide Image Classification, *Tingting Zheng, Hongxun Yao, Kui Jiang, Yi Xiao, Sicheng Zhao*

8:00-9:15 Oral Session 3B: Human Modeling (Kalakaua Ballroom)

Session Chairs: Jiri Matas, Yuexin Ma

- 1 NullSwap: Proactive Identity Cloaking Against Deepfake Face Swapping, *Tianyi Wang, Shuaicheng Niu, Harry Cheng, Xiao Zhang, Yinglong Wang*

- 2  MaskControl: Spatio-Temporal Control for Masked Motion Synthesis, *Ekkasit Pinyoanuntapong, Muhammad Saleem, Korrawe Karunratanakul, Pu Wang, Hongfei Xue, Chen Chen, Chuan Guo, Junli Cao, Jian Ren, Sergey Tulyakov*
- 3  HairCUP: Hair Compositional Universal Prior for 3D Gaussian Avatars, *Byungjun Kim, Shunsuke Saito, Giljoo Nam, Tomas Simon, Jason Saragih, Hanbyul Joo, Junxuan Li*
- 4 Understanding Co-speech Gestures in-the-wild, *Sindhu B Hegde, K R Prajwal, Taein Kwon, Andrew Zisserman*
- 5 DPoser-X: Diffusion Model as Robust 3D Whole-body Human Pose Prior, *Junzhe Lu, Jing Lin, Hongkun Dou, Ailing Zeng, Yue Deng, Xian Liu, Zhongang Cai, Lei Yang, Yulun Zhang, Haoqian Wang, Ziwei Liu*
- 6 Teeth Reconstruction and Performance Capture Using a Phone Camera, *Weixi Zheng, Jingwang Ling, Zhibo Wang, Quan Wang, Feng Xu*

9:15 - 10:15 Coffee Break (Exhibit Hall II)

9:00 - 9:30 Poster Setup (Exhibit Hall I)

9:30 - 10:30 Keynote (Exhibit Hall III)

Brent Seales, Stanley and Karen Pigman Chair of Heritage Science and Professor of Computer Science at the University of Kentucky

10:45 - 12:45 Poster Session 3 & Exhibit Hall (Exhibit Hall I)

* - Highlight paper  - Award candidate paper

○ - Oral Paper  - Outstanding Reviewer

- 1 Error Recognition in Procedural Videos using Generalized Task Graph, *Shih-Po Lee, Ehsan Elhamifar*
- 2 MoMaps: Semantics-Aware Scene Motion Generation with Motion Maps, *Jiahui Lei, Kyle Genova, George Kopanas, Noah Snively, Leonidas Guibas*
- 3 LongAnimation: Long Animation Generation with Dynamic Global-Local Memory, *Nan Chen, Mengqi Huang, Yihao Meng, Zhendong Mao*
- 4 VIGFace: Virtual Identity Generation for Privacy-Free Face Recognition Dataset, *Minsoo Kim, Min-Cheol Sagong, Gi Pyo Nam, Junghyun Cho, Ig-Jae Kim*
- 5 COVTrack: Continuous Open-Vocabulary Tracking via Adaptive Multi-Cue Fusion, *Zekun Qian, Ruize Han, Zhixiang Wang, Junhui Hou, Wei Feng*
- 6 CanonSwap: High-Fidelity and Consistent Video Face Swapping via Canonical Space Modulation, *Xiangyang Luo, Ye Zhu, Yunfei Liu, Lijian Lin, Cong Wan, Zijian Cai, Yu Li, Shao-Lun Huang*
- 7 RoboFactory: Exploring Embodied Agent Collaboration with Compositional Constraints, *Yiran Qin, Li Kang, Xiufeng Song, Zhenfei Yin, Xiaohong Liu, Xihui Liu, Ruimao Zhang, Lei Bai*
- 8 MotionStreamer: Streaming Motion Generation via Diffusion-based Autoregressive Model in Causal Latent Space, *Lixing Xiao, Shunlin Lu, Huaijin Pi, Ke Fan, Liang Pan, Yueer Zhou, Ziyong Feng, Xiaowei Zhou, Sida Peng, Jingbo Wang*
- 9 RapVerse: Coherent Vocals and Whole-Body Motion Generation from Text, *Jiabao Chen, Xin Yan, Yihang Chen, Siyuan Cen, Zixin Wang, Qinwei Ma, Haoyu Zhen, Kaizhi Qian, Lie Lu, Chuang Gan*
- 10  GGTalker: Talking Head Synthesis with Generalizable Gaussian Priors and Identity-Specific Adaptation, *Wentao Hu, Shunkai Li, Ziqiao Peng, Haoxian Zhang, Fan Shi, Xiaoqiang Liu, Pengfei Wan, Di Zhang, Hui Tian*
- 11 RoboPearls: Editable Video Simulation for Robot Manipulation, *Tang Tao, Likui Zhang, Youpeng Wen, Kaidong Zhang, Jia-Wang Bian, Xia Zhou, Tianyi Yan, Kun Zhan, Peng Jia, Hefeng Wu, Liang Lin, Xiaodan Liang*
- 12 GraspCoT: Integrating Physical Property Reasoning for 6-DoF Grasping under Flexible Language Instructions, *Xiaomeng Chu, Jiajun Deng, Guoliang You, Wei Liu, Xingchen Li, Jianmin Ji, Yanyong Zhang*
- 13 Asynchronous Event Error-Minimizing Noise for Safeguarding
- 14  Event Dataset, *Ruofei Wang, Peiqi Duan, Boxin Shi, Renjie Wan*
- 14 Exploiting Diffusion Prior for Task-driven Image Restoration, *Jaeha Kim, Junghun Oh, Kyoung Mu Lee*

- 15 Bridging Class Imbalance and Partial Labeling via Spectral-Balanced Energy Propagation for Skeleton-based Action Recognition, *Yandan Wang, Chenqi Guo, Yinglong Ma, Jiangyan Chen, Yuan Gao, Weiming Dong*
- 16 Towards Immersive Human-X Interaction: A Real-Time Framework for Physically Plausible Motion Synthesis, *Kaiyang Ji, Ye Shi, Zichen Jin, Kangyi Chen, Lan Xu, Yuexin Ma, Jingyi Yu, Jingya Wang*
- 17 SynFER: Towards Boosting Facial Expression Recognition with Synthetic Data, *Xilin He, Cheng Luo, Xiaole Xian, Bing Li, Muhammad Haris Khan, Zongyuan Ge, Weicheng Xie, Siyang Song, Linlin Shen, Bernard Ghanem, Xiangyu Yue*
- 18 ImHead: A Large-scale Implicit Morphable Model for Localized Head Modeling, *Rolandos Alexandros Potamias, Stathis Galanakis, Jiankang Deng, Athanasios Papaioannou, Stefanos Zafeiriou*
- 19 Animate Anyone 2: High-Fidelity Character Image Animation with Environment Affordance, *Li Hu, Guangyuan Wang, Zhen Shen, Xin Gao, Dechao Meng, Lian Zhuo, Peng Zhang, Bang Zhang, Liefeng Bo*
- 20 PVMamba: Parallelizing Vision Mamba via Dynamic State Aggregation, *Fei Xie, Zhongdao Wang, Weijia Zhang, Chao Ma*
- 21 FlowStyler: Artistic Video Stylization via Transformation Fields Transports, *Yuning Gong, Jiaming Chen, Xiaohua Ren, Yuanjun Liao, Yanci Zhang*
- 22 Controllable and Expressive One-Shot Video Head Swapping, *Chaonan Ji, Jinwei Qi, Peng Zhang, Bang Zhang, Liefeng Bo*
- 23 Multi-modal Multi-platform Person Re-identification: Benchmark and Method, *Ruiyang Ha, Songyi Jiang, Bin Li, Bikang Pan, Yihang Zhu, Junjie Zhang, Xiatian Zhu, Shaogang Gong, Jingya Wang*
- 24 HERO: Human Reaction Generation from Videos, *Chengjun Yu, Wei Zhai, Yuhang Yang, Yang Cao, Zheng-Jun Zha*
- 25 Mixture of Experts Guided by Gaussian Splatters Matters: A new Approach to Weakly-Supervised Video Anomaly Detection, *Giacomo D'Amicantonio, Snehashis Majhi, Quan Kong, Lorenzo Garattoni, Gianpiero Francesca, Francois Bremond, Egor Bondarev*
- 26 What If: Understanding Motion Through Sparse Interactions, *Stefan Andreas Baumann, Nick Stracke, Timy Phan, Björn Ommer*
- 27 PROGRESSOR: A Perceptually Guided Reward Estimator with Self-Supervised Online Refinement, *Tewodros W. Ayalew, Xiao Zhang, Kevin Yuanbo Wu, Tianchong Jiang, Michael Maire, Matthew R. Walter*
- 28 How Far are AI-generated Videos from Simulating the 3D Visual World: A Learned 3D Evaluation Approach, *Chirui Chang, Jiahui Liu, Zhengzhe Liu, Xiaoyang Lyu, Yi-Hua Huang, Xin Tao, Pengfei Wan, Di Zhang, Xiaojuan Qi*
- 29 UniEgoMotion: A Unified Model for Egocentric Motion Reconstruction, Forecasting, and Generation, *Chaitanya Patel, Hiroki Nakamura, Yuta Kyuragi, Kazuki Kozuka, Juan Carlos Niebles, Ehsan Adeli*
- 30 DAViD: Modeling Dynamic Affordance of 3D Objects Using Pre-trained Video Diffusion Models, *Hyeonwoo Kim, Sangwon Baik, Hanbyul Joo*
- 31 DNF-Intrinsic: Deterministic Noise-Free Diffusion for Indoor Inverse Rendering, *Rongjia Zheng, Qing Zhang, Chengjiang Long, Wei-Shi Zheng*
- 32 RoboAnnotatorX: A Comprehensive and Universal Annotation Framework for Accurate Understanding of Long-horizon Robot Demonstration, *Longxin Kou, Fei Ni, Yan Zheng, Peilong Han, Jinyi Liu, Haiqin Cui, Rui Liu, Jianye Hao*
- 33 FaceShield: Defending Facial Image against Deepfake Threats, *Jaehwan Jeong, Sumin In, Sieun Kim, Hannie Shin, Jongheon Jeong, Sang Ho Yoon, Jaewook Chung, Sangpil Kim*
- 34 Task-Oriented Human Grasp Synthesis via Context- and Task-Aware Diffusers, *An-Lun Liu, Yu-Wei Chao, Yi-Ting Chen*
- 35 Ouroboros: Single-step Diffusion Models for Cycle-consistent Forward and Inverse Rendering, *Shanlin Sun, Yifan Wang, Hanwen Zhang, Yifeng Xiong, Qin Ren, Ruogu Fang, Xiaohui Xie, Chenyu You*
- 36 Expressive Talking Human from Single-Image with Imperfect Priors, *Jun Xiang, Yudong Guo, Leipeng Hu, Boyang Guo, Yancheng Yuan, Juyong Zhang*
- 37 InteractAvatar: Modeling Hand-Face Interaction in Photorealistic Avatars with Deformable Gaussians, *Kefan Chen, Sreyas Mohan, Justin Theiss, Sergiu Oprea, Srinath Sridhar, Aayush Prakash*
- 38 Beyond Label Semantics: Language-Guided Action Anatomy for Few-shot Action Recognition, *Zefeng Qian, Xincheng Yao, Yifei Huang, Chongyang Zhang, Jiangyong Ying, Hong Sun*
- 39 DynFaceRestore: Balancing Fidelity and Quality in Diffusion-Guided Blind Face Restoration with Dynamic Blur-Level Mapping and Guidance, *Huu-Phu Do, Yu-Wei Chen, Yi-Cheng Liao, Chi-Wei Hsiao, Han-Yang Wang, Wei-Chen Chiu, Ching-Chun Huang*
- 40 Few-Shot Image Quality Assessment via Adaptation of Vision-Language Models, *Xudong Li, Zihao Huang, Yan Zhang, Yunhang Shen, Ke Li, Xiawu Zheng, Liujuan Cao, Rongrong Ji*
- 41 Unleashing High-Quality Image Generation in Diffusion Sampling Using Second-Order Levenberg-Marquardt-Langevin, *Fangyikang Wang, Hubery Yin, Lei Qian, Yinan Li, Shaobin Zhuang, Huminhao Zhu, Yilin Zhang, Yanlong Tang, Chao Zhang, Hanbin Zhao, Hui Qian, Chen Li*
- 42 VISION-XL: High Definition Video Inverse Problem Solver using Latent Image Diffusion Models, *Taesung Kwon, Jong Chul Ye*
- 43 Dynamic Group Detection using VLM-augmented Temporal Groupness Graph, *Kaname Yokoyama, Chihiro Nakatani, Norimichi Ukita*
- 44 When Lighting Deceives: Exposing Vision-Language Models' Illumination Vulnerability Through Illumination Transformation Attack, *Hanqing Liu, Shouwei Ruan, Yao Huang, Shiji Zhao, Xingxing Wei*
- 45 Self-Calibrated Variance-Stabilizing Transformations for Real-World Image Denoising, *Sébastien Herbreteau, Michael Unser*
- 46 Reverse Convolution and Its Applications to Image Restoration, *Xuhong Huang, Shiqi Liu, Kai Zhang, Ying Tai, Jian Yang, Hui Zeng, Lei Zhang*
- 47 MamTiff-CAD: Multi-Scale Latent Diffusion with Mamba+ for Complex Parametric Sequence, *Liyuan Deng, Yunpeng Bai, Yongkang Dai, Xiaoshui Huang, Hongping Gan, Dongshuo Huang, Hao Jiacheng, Yilei Shi*
- 48 Local Scale Equivariance with Latent Deep Equilibrium Canonicalizer, *Md Ashiqur Rahman, Chiao-An Yang, Michael N. Cheng, Lim Jun Hao, Jeremiah Jiang, Teck-Yian Lim, Raymond A. Yeh*
- 49 HOMO-Feature: Cross-Arbitrary-Modal Image Matching with Homomorphism of Organized Major Orientation, *Chenzhong Gao, Wei Li, Desheng Weng*
- 50 DynamicID: Zero-Shot Multi-ID Image Personalization with Flexible Facial Editability, *Xirui Hu, Jiahao Wang, Hao Chen, Weizhan Zhang, Benqi Wang, Yikun Li, Haishun Nan*
- 51 FreeDance: Towards Harmonic Free-Number Group Dance Generation via a Unified Framework, *Yiwen Zhao, Yang Wang, Liting Wen, Hengyuan Zhang, Xingqun Qi*
- 52 SILO: Solving Inverse Problems with Latent Operators, *Ron Raphaelli, Sean Man, Michael Elad*
- 53 Learning Precise Affordances from Egocentric Videos for Robotic Manipulation, *Gen Li, Nikolaos Tsagkas, Jifei Song, Ruairidh Mon-Williams, Sethu Vijayakumar, Kun Shao, Laura Sevilla-Lara*
- 54 EfficientMT: Efficient Temporal Adaptation for Motion Transfer in Text-to-Video Diffusion Models, *Yufei Cai, Hu Han, Yuxiang Wei, Shiguang Shan, Xilin Chen*
- 55 X-Dancer: Expressive Music to Human Dance Video Generation, *Zeyuan Chen, Hongyi Xu, Guoxian Song, You Xie, Chenxu Zhang, Xin Chen, Chao Wang, Di Chang, Linjie Luo*
- 56 DeepMesh: Auto-Regressive Artist-mesh Creation with Reinforcement Learning, *Ruowen Zhao, Junliang Ye, Zhengyi Wang, Guangce Liu, Yiwen Chen, Yikai Wang, Jun Zhu*
- 57 SAM Encoder Breach by Adversarial Simplicial Complex Triggers Downstream Model Failures, *Yi Qin, Rui Wang, Tao Huang, Tong Xiao, Liping Jing*
- 58 Text-IRSTD: Leveraging Semantic Text to Promote Infrared Small Target Detection in Complex Scenes, *Feng Huang, Shuyuan Zheng, Zhaobing Qiu, Huanxian Liu, Huanxin Bai, Liqiong Chen*
- 59 AMDANet: Attention-Driven Multi-Perspective Discrepancy Alignment for RGB-Infrared Image Fusion and Segmentation, *Haifeng Zhong, Fan Tang, Zhuo Chen, Hyung Jin Chang, Yixing Gao*
- 60 Im2Haircut: Single-view Strand-based Hair Reconstruction for Human Avatars, *Vanessa Sklyarova, Egor Zakharov, Malte Prinzler,*

- Giorgio Becherini, Michael J. Black, Justus Thies
- 61 AFUNet: Cross-Iterative Alignment-Fusion Synergy for HDR Reconstruction via Deep Unfolding Paradigm, Xinyue Li, Zhangkai Ni, Wenhan Yang
- 62 PINO: Person-Interaction Noise Optimization for Long-Duration and
- ☆ Customizable Motion Generation of Arbitrary-Sized Groups, Sakuya Ota, Qing Yu, Kent Fujiwara, Satoshi Ikehata, Ikuro Sato
- 63 TeRA: Rethinking Text-guided Realistic 3D Avatar Generation, Yanwen Wang, Yiyu Zhuang, Jiawei Zhang, Li Wang, Yifei Zeng, Xun Cao, Xinxin Zuo, Hao Zhu
- 64 A Unified Framework for Motion Reasoning and Generation in Human Interaction, Jeongeun Park, Sungjoon Choi, Sangdoo Yun
- 65 Open-World Skill Discovery from Unsegmented Demonstration Videos, Jingwen Deng, Zihao Wang, Shaofei Cai, Anji Liu, Yitao Liang
- 66 CLOT: Closed Loop Optimal Transport for Unsupervised Action Segmentation, Elena Bueno-Benito, Mariella Dimiccoli
- 67 Deep Adaptive Unfolded Network via Spatial Morphology Stripping and Spectral Filtration for Pan-sharpening, Hebaixu Wang, Jiayi Ma
- 68 EgoAdapt: Adaptive Multisensory Distillation and Policy Learning for Efficient Egocentric Perception, Sanjoy Chowdhury, Subrata Biswas, Sayan Nag, Tushar Nagarajan, Calvin Murdock, Ishwarya Ananthabhotla, Yijun Qian, Vamsi Krishna Ithapu, Dinesh Manocha, Ruohan Gao
- 69 Towards Human-like Virtual Beings: Simulating Human Behavior in 3D Scenes, Chen Liang, Wenguan Wang, Yi Yang
- 70 Reference-based Super-Resolution via Image-based Retrieval-Augmented Generation Diffusion, Byeonghun Lee, Hyunmin Cho, Hong Gyu Choi, Soo Min Kang, Iljun Ahn, Kyong Hwan Jin
- 71 Dual-Temporal Exemplar Representation Network for Video Semantic Segmentation, Xiaolong Xu, Lei Zhang, Jiayi Li, Lituan Wang, Yifan Guan, Yu Yan, Leyi Zhang, Hao Song
- 72 Vulnerability-Aware Spatio-Temporal Learning for Generalizable Deepfake Video Detection, Dat Nguyen, Marcella Astrid, Anis Kacem, Enjie Ghorbel, Djamilia Aouada
- 73 Multi-modal Identity Extraction, Ryan Webster,
- ☆ Teddy Furon
- 74 NullSwap: Proactive Identity Cloaking Against Deepfake Face Swapping, Tianyi Wang, Shuaicheng Niu, Harry Cheng, Xiao Zhang, Yinglong Wang
- 75 LoftUp: Learning a Coordinate-Based Feature Upsampler for
- Vision Foundation Models, Haiwen Huang, Anpei Chen, Volodymyr Havrylov, Andreas Geiger, Dan Zhang
- 76 Joint Self-Supervised Video Alignment and Action Segmentation, Ali Shah Ali, Syed Ahmed Mahmood, Mubin Saeed, Andrey Konin, M. Zeeshan Zia, Quoc-Huy Tran
- 77 VSSD: Vision Mamba with Non-Causal State Space Duality, Yuheng Shi, Mingjia Li, Minjing Dong, Chang Xu
- 78 EgoM2P: Egocentric Multimodal Multitask Pretraining, Gen Li, Yutong Chen, Yiqian Wu, Kaifeng Zhao, Marc Pollefeys, Siyu Tang
- 79 Efficient Multi-Person Motion Prediction by Lightweight Spatial and Temporal Interactions, Yuanhong Zheng, Ruixuan Yu, Jian Sun
- 80 E-NeMF: Event-based Neural Motion Field for Novel Space-time View
- ☆ Synthesis of Dynamic Scenes, Yan Liu, Zehao Chen, Haojie Yan, De Ma, Huajin Tang, Qian Zheng, Gang Pan
- 81 LayerAnimate: Layer-level Control for Animation, Yuxue Yang, Lue Fan, Zuzeng Lin, Feng Wang, Zhaoxiang Zhang
- 82 AnimeGamer: Infinite Anime Life Simulation with Next Game State Prediction, Junhao Cheng, Yuying Ge, Yixiao Ge, Jing Liao, Ying Shan
- 83 HUMOTO: A 4D Dataset of Mocap Human Object Interactions, Jiabin Lu, Chun-Hao Paul Huang, Uttaran Bhattacharya, Qixing Huang, Yi Zhou
- 84 InfiniteYou: Flexible Photo Recrafting While Preserving Your Identity,
- * Liming Jiang, Qing Yan, Yumin Jia, Zichuan Liu, Hao Kang, Xin Lu
- 85 Enhanced Pan-sharpening via Quaternion Spatial-Spectral Interactions, Dong Li, Chunhui Luo, Yuanfei Bao, Gang Yang, Jie Xiao, Xueyang Fu, Zheng-Jun Zha
- 86 CombatVLA: An Efficient Vision-Language-Action Model for Combat Tasks in 3D Action Role-Playing Games, Peng Chen, Pi Bu, Yingyao Wang, Xinyi Wang, Ziming Wang, Jie Guo, Yingxiu Zhao, Qi Zhu, Jun Song, Siran Yang, Jiamang Wang, Bo Zheng
- 87 GestureLSM: Latent Shortcut based Co-Speech Gesture Generation with Spatial-Temporal Modeling, Pinxin Liu, Luchuan Song, Junhua Huang, Haiyang Liu, Chenliang Xu
- 88 MonStER: a Unified Model for Motion, Scene, Text Retrieval, Luca Collorone, Matteo Gioia, Massimiliano Pappa, Paolo Leoni, Giovanni Ficarra, Or Litany, Indro Spinelli, Fabio Galasso
- 89 UDC-VIT: A Real-World Video Dataset for Under-Display Cameras,
- * Kyusu Ahn, JiSoo Kim, Sangik Lee, HyunGyu Lee, Byeonghyun Ko, Chanwoo Park, Jaejin Lee
- 90 Nautilus: Locality-aware Autoencoder for Scalable Mesh Generation, Yuxuan Wang, Xuanyu Yi, Haohan Weng, Qingshan Xu, Xiaokang Wei, Xianghui Yang, Chunchao Guo, Long Chen, Hanwang Zhang
- 91 Blind Video Super-Resolution based on Implicit Kernels, Qiang Zhu, Yuxuan Jiang, Shuyuan Zhu, Fan Zhang, David Bull, Bing Zeng
- 92 F-Bench: Rethinking Human Preference Evaluation Metrics
- * for Benchmarking Face Generation, Customization, and Restoration, Lu Liu, Huiyu Duan, Qiang Hu, Liu Yang, Chunlei Cai, Tianxiao Ye, Huayu Liu, Xiaoyun Zhang, Guangtao Zhai
- 93 Decouple and Track: Benchmarking and Improving Video Diffusion Transformers For Motion Transfer, Qingyu Shi, Jianzong Wu, Jinbin Bai, Jiangning Zhang, Lu Qi, Yunhai Tong, Xiangtai Li
- 94 Latent Swap Joint Diffusion for 2D Long-Form Latent Generation, Yusheng Dai, Chenxi Wang, Chang Li, Chen Wang, Kewei Li, Jun Du, Lei Sun, Jianqing Gao, Ruoyu Wang, Jiefeng Ma
- 95 Blind Noisy Image Deblurring Using Residual Guidance Strategy, Heyan Liu, Jianing Sun, Jun Liu, Xi-Le Zhao, Tingting Wu, Tiejong Zeng
- 96 Drawing Developmental Trajectory from Cortical Surface Reconstruction, Wenxuan Wu, Ruowen Qu, Zhongliang Liu, Zhuoyan Dai, Dongzi Shi, Sijin Yu, Tong Xiong, Shiping Liu, Xiangmin Xu, Xiaofen Xing, Xin Zhang
- 97 DreamActor-M1: Holistic, Expressive and Robust Human Image Animation with Hybrid Guidance, Yuxuan Luo, Zhengkun Rong, Lizhen Wang, Longhao Zhang, Tianshu Hu
- 98 Diffuman4D: 4D Consistent Human View Synthesis from Sparse-View Videos with Spatio-Temporal Diffusion Models, Yudong Jin, Sida Peng, Xuan Wang, Tao Xie, Zhen Xu, Yifan Yang, Yujun Shen, Hujun Bao, Xiaowei Zhou
- 99 DyWA: Dynamics-adaptive World Action Model for Generalizable Non-prehensile Manipulation, Jiangran Lyu, Ziming Li, Xuesong Shi, Chaoyi Xu, Yizhou Wang, He Wang
- 100 Less is More: Improving Motion Diffusion Models with Sparse Keyframes, Jinseok Bae, Inwoo Hwang, Young-Yoon Lee, Ziyu Guo, Joseph Liu, Yizhak Ben-Shabat, Young Min Kim, Mubbasir Kapadia
- 101 DGTalker: Disentangled Generative Latent Space Learning for Audio-Driven Gaussian Talking Heads, Xiaoxi Liang, Yanbo Fan, Qiya Yang, Xuan Wang, Wei Gao, Ge Li
- 102 VQ-VLA: Improving Vision-Language-Action Models via Scaling Vector-Quantized Action Tokenizers, Yating Wang, Haoyi Zhu, Mingyu Liu, Jiange Yang, Hao-Shu Fang, Tong He
- 103 Augmented and Softened Matching for Unsupervised Visible-Infrared Person Re-Identification, Zhiqi Pang, Chunyu Wang, Lingling Zhao, Junjie Wang
- 104 Temporal Unlearnable Examples: Preventing Personal Video Data from Unauthorized Exploitation by Object Tracking, Qiangqiang Wu, Yi Yu, Chenqi Kong, Ziquan Liu, Jia Wan, Haoliang Li, Alex C. Kot, Antoni B. Chan
- 105 Frequency-Semantic Enhanced Variational Autoencoder for Zero-Shot Skeleton-based Action Recognition, Wenhan Wu, Zhishuai Guo, Chen Chen, Hongfei Xue, Aidong Lu
- 106 Learning Hierarchical Line Buffer for Image Processing, Jiacheng Li,
- ☆ Feiran Li, Daisuke Iso
- 107 VLABench: A Large-Scale Benchmark for Language-Conditioned Robotics Manipulation with Long-Horizon Reasoning Tasks, Shiduo Zhang, Zhe Xu, Peiju Liu, Xiaopeng Yu, Yuan Li, Qinghui Gao, Zhaoye Fei, Zhangyue Yin, Zuxuan Wu, Yu-Gang Jiang, Xipeng Qiu
- 108 TrackVerse: A Large-Scale Object-Centric Video Dataset for Image-Level Representation Learning, Yibing Wei, Samuel Church, Victor Suciu, Jinhong Lin, Cheng-En Wu, Pedro Morgado

- 109 Reangle-A-Video: 4D Video Generation as Video-to-Video Translation, *Hyeonho Jeong, Suhyeon Lee, Jong Chul Ye*
- 110 Human-Object Interaction from Human-Level Instructions, *Zhen Wu, Jiaman Li, Pei Xu, C. Karen Liu*
- 111 KinMo: Kinematic-aware Human Motion Understanding and Generation, *Pengfei Zhang, Pinxin Liu, Pablo Garrido, Hyeonwoo Kim, Bindita Chaudhuri*
- 112 Beyond Spatial Frequency: Pixel-wise Temporal Frequency-based Deepfake Video Detection, *Taehoon Kim, Jongwook Choi,*
* *Yonghyun Jeong, Haeun Noh, Jaemin Yoo, Seungryl Baek, Jongwon Choi*
- 113 Causal-Entity Reflected Egocentric Traffic Accident Video Synthesis, *Lei-Lei Li, Jianwu Fang, Junbin Xiao, Shanmin Pang, Hongkai Yu, Chen Lv, Jianru Xue, Tat-Seng Chua*
- 114 WaveMamba: Wavelet-Driven Mamba Fusion for RGB-Infrared Object Detection, *Haodong Zhu, Wenhao Dong, Linlin Yang, Hong Li, Yuguang Yang, Yangyang Ren, Qingcheng Zhu, Zichao Feng, Changbai Li, Shaohui Lin, Runqi Wang, Xiaoyan Luo, Baochang Zhang*
- 115 Robust Test-Time Adaptation for Single Image Denoising Using Deep Gaussian Prior, *Qing Ma, Pengwei Liang, Xiong Zhou, Jiayi Ma, Junjun Jiang, Zhe Peng*
- 116 MemoryTalker: Personalized Speech-Driven 3D Facial Animation via Audio-Guided Stylization, *Hyung Kyu Kim, Sangmin Lee, Hak Gu Kim*
- 117 Hierarchical-aware Orthogonal Disentanglement Framework for Fine-grained Skeleton-based Action Recognition, *Haochen Chang, Pengfei Ren, Haoyang Zhang, Liang Xie, Hongbo Chen, Erwei Yin*
- 118 Balancing Task-invariant Interaction and Task-specific Adaptation for Unified Image Fusion, *Xingyu Hu, Junjun Jiang, Chenyang Wang, Kui Jiang, Xianming Liu, Jiayi Ma*
- 119 EVOLVE: Event-Guided Deformable Feature Transfer and Dual-Memory Refinement for Low-Light Video Object Segmentation, *Jong-Hyeon Baek, Jiwon Oh, Yeong Jun Koh*
- 120 PatchScaler: An Efficient Patch-Independent Diffusion Model for Image Super-Resolution, *Yong Liu, Hang Dong, Jinshan Pan, Qingji Dong, Kai Chen, Rongxiang Zhang, Lean Fu, Fei Wang*
- 121 OneGT: One-Shot Geometry-Texture Neural Rendering for Head Avatars, *Jinshu Chen, Bingchuan Li, Fan Zhang, Songtao Zhao, Qian He*
- 122 MotionAgent: Fine-grained Controllable Video Generation via Motion Field Agent, *Xinyao Liao, Xianfang Zeng, Liao Wang, Gang Yu, Guosheng Lin, Chi Zhang*
- 123 D2ST-Adapter: Disentangled-and-Deformable Spatio-Temporal Adapter for Few-shot Action Recognition, *Wenjie Pei, Qizhong Tan, Guangming Lu, Jiandong Tian, Jun Yu*
- 124 Disentangled Clothed Avatar Generation with Layered Representation, *Weitian Zhang, Yichao Yan, Sijing Wu, Manwen Liao, Xiaokang Yang*
* *Xiaokang Yang*
- 125 Augmented Mass-Spring Model for Real-Time Dense Hair Simulation, *J. Alejandro Amador H., Yi Zhou, Xin Sun, Zhixin Shu, Chengan He, Soren Pirk, Dominik L. Michels*
- 126 Punching Bag vs. Punching Person: Motion Transferability in Videos, *Raiyaan Abdullah, Jared Claypoole, Michael Cogswell, Ajay Divakaran, Yogesh Rawat*
- 127 OCK: Unsupervised Dynamic Video Prediction with Object-Centric Kinematics, *Yeon-Ji Song, Jaein Kim, Suhyung Choi, Jin-Hwa Kim, Byoung-Tak Zhang*
* *Yeon-Ji Song, Jaein Kim, Suhyung Choi, Jin-Hwa Kim, Byoung-Tak Zhang*
- 128 FaceXFormer: A Unified Transformer for Facial Analysis, *Kartik Narayan, Vibashan VS, Rama Chellappa, Vishal M. Patel*
- 129 ContextFace: Generating Facial Expressions from Emotional Contexts, *Min-jung Kim, Minsang Kim, Seung Jun Baek*
- 130 Laboring on less labors: RPCA Paradigm for Pan-sharpening, *Honghui Xu, Chuangjie Fang, Yibin Wang, Jie Wu, Jianwei Zheng*
- 131 ShadowHack: Hacking Shadows via Luminance-Color Divide and Conquer, *Jin Hu, Mingjia Li, Xiaojie Guo*
- 132 What we need is explicit controllability: Training 3D gaze estimator using only facial images, *Tingwei Li, Jun Bao, Zhenzhong Kuang, Buyu Liu*
- 133 Riemannian-Geometric Fingerprints of Generative Models, *Hae Jin Song, Laurent Itti*
* *Hae Jin Song, Laurent Itti*
- 134 Unraveling the Smoothness Properties of Diffusion Models: A Gaussian Mixture Perspective, *Yingyu Liang, Zhizhou Sha, Zhenmei Shi, Zhao Song, Mingda Wan, Yufa Zhou*
- 135 G-DexGrasp: Generalizable Dexterous Grasping Synthesis Via Part-Aware Prior Retrieval and Prior-Assisted Generation, *Juntao Jian, Xiuping Liu, Zixuan Chen, Manyi Li, Jian Liu, Ruizhen Hu*
- 136 FPEM: Face Prior Enhanced Facial Attractiveness Prediction for Live Videos with Face Retouching, *Hui Li, Xiaoyu Ren, Hongjiu Yu, Ying Chen, Kai Li, L. Wang, Xiongkuo Min, Huiyu Duan, Guangtao Zhai, Xu Liu*
* *Hui Li, Xiaoyu Ren, Hongjiu Yu, Ying Chen, Kai Li, L. Wang, Xiongkuo Min, Huiyu Duan, Guangtao Zhai, Xu Liu*
- 137 Diffusion-Based Imaginative Coordination for Bimanual Manipulation, *Huailin Xu, Jian Ding, Jiakun Xu, Ruixiang Wang, Jun Chen, Jinjie Mai, Yanwei Fu, Bernard Ghanem, Feng Xu, Mohamed Elhoseiny*
- 138 WarpHE4D: Dense 4D Head Map toward Full Head Reconstruction, *Jongseob Yun, Yong-Hoon Kwon, Min-Gyu Park, Ju-Mi Kang, Min-Ho Lee, Inho Chang, Ju Hong Yoon, Kuk-Jin Yoon*
- 139 PrimHOL: Compositional Human-Object Interaction via Reusable Primitives, *Kai Jia, Tengyu Liu, Mingtao Pei, Yixin Zhu, Siyuan Huang*
- 140 Continuous-Time Human Motion Field from Event Cameras, *Ziyun Wang, Ruijun Zhang, Zi-Yan Liu, Yufu Wang, Kostas Daniilidis*
- 141 Efficient Track Anything, *Yunyang Xiong, Chong Zhou, Xiaoyu Xiang, Lemeng Wu, Chenchen Zhu, Zechun Liu, Saksham Suri, Balakrishnan Varadarajan, Ramya Akula, Forrest Iandola, Raghuraman Krishnamoorthi, Bilge Soran, Vikas Chandra*
- 142 HAMoBE: Hierarchical and Adaptive Mixture of Biometric Experts for Video-based Person ReID, *Yiyang Su, Yunping Shi, Feng Liu, Xiaoming Liu*
- 143 Multi-Object Sketch Animation by Scene Decomposition and Motion Planning, *Jingyu Liu, Zijie Xin, Yuhan Fu, Ruixiang Zhao, Bangxiang Lan, Xirong Li*
- 144 ISP2HRNet: Learning to Reconstruct High Resolution Image from Irregularly Sampled Pixels via Hierarchical Gradient Learning, *Yuanlin Wang, Ruiqin Xiong, Rui Zhao, Jin Wang, Xiaopeng Fan, Tiejun Huang*
* *Yuanlin Wang, Ruiqin Xiong, Rui Zhao, Jin Wang, Xiaopeng Fan, Tiejun Huang*
- 145 LDIP: Long Distance Information Propagation for Video Super-Resolution, *Michael Bernasconi, Abdelaziz Djelouah, Yang Zhang, Markus Gross, Christopher Schroers*
- 146 MBTI: Masked Blending Transformers with Implicit Positional Encoding for Frame-rate Agnostic Motion Estimation, *Jungwoo Huh, Yeseung Park, Seongjean Kim, Jungsu Kim, Sanghoon Lee*
- 147 Sequential keypoint density estimator: an overlooked baseline of skeleton-based video anomaly detection, *Anja Delić, Matej Grcic, Siniša Šegvić*
* *Anja Delić, Matej Grcic, Siniša Šegvić*
- 148 MaskControl: Spatio-Temporal Control for Masked Motion Synthesis, *Ekkasit Pinyoanuntapong, Muhammad Saleem, Korrawe Karunratanakul, Pu Wang, Hongfei Xue, Chen Chen, Chuan Guo, Junli Cao, Jian Ren, Sergey Tulyakov*
* *Ekkasit Pinyoanuntapong, Muhammad Saleem, Korrawe Karunratanakul, Pu Wang, Hongfei Xue, Chen Chen, Chuan Guo, Junli Cao, Jian Ren, Sergey Tulyakov*
- 149 RS-vHeat: Heat Conduction Guided Efficient Remote Sensing Foundation Model, *Huiyang Hu, Peijin Wang, Hanbo Bi, Boyuan Tong, Zhaozhi Wang, Wenhui Diao, Hao Chang, Yingchao Feng, Ziqi Zhang, Yaowei Wang, Qixiang Ye, Kun Fu, Xian Sun*
- 150 GameFactory: Creating New Games with Generative Interactive Videos, *Jiwen Yu, Yiran Qin, Xintao Wang, Pengfei Wan, Di Zhang, Xihui Liu*
* *Jiwen Yu, Yiran Qin, Xintao Wang, Pengfei Wan, Di Zhang, Xihui Liu*
- 151 MOSCATO: Predicting Multiple Object State Change Through Actions, *Parnian Zameni, Yuhan Shen, Ehsan Elhamifar*
- 152 FaceCraft4D: Animated 3D Facial Avatar Generation from a Single Image, *Fei Yin, Mallikarjun B R, Chun-Han Yao, Rafal K. Mantiuk, Varun Jampani*
- 153 Decouple to Reconstruct: High Quality UHD Restoration via Active Feature Disentanglement and Reversible Fusion, *Yidi Liu, Dong Li, Yuxin Ma, Jie Huang, Wenlong Zhang, Xueyang Fu, Zheng-Jun Zha*
- 154 MOVE: Motion-Guided Few-Shot Video Object Segmentation, *Kaining Ying, Hengrui Hu, Henghui Ding*
- 155 V2M4: 4D Mesh Animation Reconstruction from a Single Monocular Video, *Jianqi Chen, Biao Zhang, Xiangjun Tang, Peter Wonka*
- 156 Event-Driven Storytelling with Multiple Lifelike Humans in a 3D Scene, *Donggeun Lim, Jinseok Bae, Inwoo Hwang, Seungmin Lee, Hwanhee Lee, Young Min Kim*

- 157 EvolvingGrasp: Evolutionary Grasp Generation via Efficient Preference Alignment, *Yufei Zhu, Yiming Zhong, Zemin Yang, Peishan Cong, Jingyi Yu, Xinge Zhu, Yuexin Ma*
- 158 Structure-Guided Diffusion Models for High-Fidelity Portrait Shadow Removal, *Wanchang Yu, Qing Zhang, Rongjia Zheng, Wei-Shi Zheng*
- 159 Efficient Autoregressive Shape Generation via Octree-Based Adaptive Tokenization, *Kangle Deng, Hsueh-Ti Derek Liu, Yiheng Zhu, Xiaoxia Sun, Chong Shang, Kiran S. Bhat, Deva Ramanan, Jun-Yan Zhu, Maneesh Agrawala, Tinghui Zhou*
- 160 MedVSR: Medical Video Super-Resolution with Cross State-Space Propagation, *Xinyu Liu, Guolei Sun, Cheng Wang, Yixuan Yuan, Ender Konukoglu*
- 161 EAMamba: Efficient All-Around Vision State Space Model for Image Restoration, *Yu-Cheng Lin, Yu-Syuan Xu, Hao-Wei Chen, Hsien-Kai Kuo, Chun-Yi Lee*
- 162 Beyond Walking: A Large-Scale Image-Text Benchmark for Text-based Person Anomaly Search, *Shuyu Yang, Yaxiong Wang, Li Zhu, Zhedong Zheng*
- 163 SyncDiff: Synchronized Motion Diffusion for Multi-Body Human-Object Interaction Synthesis, *Wenkun He, Yun Liu, Ruitao Liu, Li Yi*
- 164 MMAIF: Multi-task and Multi-degradation All-in-One for Image Fusion with Language Guidance, *Zihan Cao, Yu Zhong, Ziqi Wang, Liang-Jian Deng*
- 165 Fast Image Super-Resolution via Consistency Rectified Flow, *Jiaqi Xu, Wenbo Li, Haoze Sun, Fan Li, Zhixin Wang, Long Peng, Jingjing Ren, Haoran Yang, Xiaowei Hu, Renjing Pei, Pheng-Ann Heng*
- 166 GENMO: A GENeralist Model for Human MOTion, *Jiefeng Li, Jinkun Cao, Haotian Zhang, Davis Rempe, Jan Kautz, Umar Iqbal, Ye Yuan*
- 167 VAFFlow: Video-to-Audio Generation with Cross-Modality Flow Matching, *Xihua Wang, Xin Cheng, Yuyue Wang, Ruihua Song, Yunfeng Wang*
- 168 Event-guided HDR Reconstruction with Diffusion Priors, *Yixin Yang, Jiawei Zhang, Yang Zhang, Yunxuan Wei, Dongqing Zou, Jimmy S. Ren, Boxin Shi*
- 169 Learning Efficient and Generalizable Human Representation with Human Gaussian Model, *Yifan Liu, Shengjun Zhang, Chensheng Dai, Yang Chen, Hao Liu, Chen Li, Yueqi Duan*
- 170 SMGDiff: Soccer Motion Generation using Diffusion Probabilistic Models, *Hongdi Yang, Chengyang Li, Zhenxuan Wu, Gaozheng Li, Jingya Wang, Jingyi Yu, Zhuo Su, Lan Xu*
- 171 AffordDexGrasp: Open-set Language-guided Dexterous Grasp with Generalizable-Instructive Affordance, *Yi-Lin Wei, Mu Lin, Yuhao Lin, Jian-Jian Jiang, Xiao-Ming Wu, Ling-An Zeng, Wei-Shi Zheng*
- 172 Hierarchical Visual Prompt Learning for Continual Video Instance Segmentation, *Jiahua Dong, Hui Yin, Wenqi Liang, Hanbin Zhao, Henghui Ding, Nicu Sebe, Salman Khan, Fahad Shahbaz Khan*
- 173 LVFace: Progressive Cluster Optimization for Large Vision Models in Face Recognition, *Jinghan You, Shanglin Li, Yuanrui Sun, Jiangchuan Wei, Mingyu Guo, Chao Feng, Jiao Ran*
- 174 Empowering Your Pansharpening Models with Generalizability: Unified Distribution is All You Need, *Yongchuan Cui, Peng Liu, Hui Zhang*
- 175 MotionShot: Adaptive Motion Transfer across Arbitrary Objects for Text-to-Video Generation, *Yanchen Liu, Yanan Sun, Zhening Xing, Junyao Gao, Kai Chen, Wenjie Pei*
- 176 Robust Adverse Weather Removal via Spectral-based Spatial Grouping, *Yuhwan Jeong, Yunseo Yang, Youngho Yoon, Kuk-Jin Yoon*
- 177 CarGait: Cross-Attention based Re-ranking for Gait recognition, *Gavriel Habib, Noa Barzilay, Or Shimshi, Rami Ben-Ari, Nir Darshan*
- 178 Lightweight and Fast Real-time Image Enhancement via Decomposition of the Spatial-aware Lookup Tables, *Wontae Kim, Keuntek Lee, Nam Ik Cho*
- 179 WAVE: Warp-Based View Guidance for Consistent Novel View Synthesis Using a Single Image, *Jiwoo Park, Tae Eun Choi, Youngjun Jun, Seong Jae Hwang*
- 180 Unsupervised Visible-Infrared Person Re-identification under Unpaired Settings, *Haoyu Yao, Bin Yang, Wenke Huang, Bo Du, Mang Ye*
- 181 LightBSR: Towards Lightweight Blind Super-Resolution via Discriminative Implicit Degradation Representation Learning, *Jiang Yuan, Ji Ma, Bo Wang, Guanzhou Ke, Weiming Hu*
- 182 Multi-identity Human Image Animation with Structural Video Diffusion, *Zhenzhi Wang, Yixuan Li, Yanhong Zeng, Yuwei Guo, Dahua Lin, Tianfan Xue, Bo Dai*
- 183 Embodied Representation Alignment with Mirror Neurons, *Wentao Zhu, Zhining Zhang, Yuwei Ren, Yin Huang, Hao Xu, Yizhou Wang*
- 184 EC-Flow: Enabling Versatile Robotic Manipulation from Action-Unlabeled Videos via Embodiment-Centric Flow, *Yixiang Chen, Peiyan Li, Yan Huang, Jiabing Yang, Kehan Chen, Liang Wang*
- 185 Switch-a-View: View Selection Learned from Unlabeled In-the-wild Videos, *Sagnik Majumder, Tushar Nagarajan, Ziad Al-Halah, Kristen Grauman*
- 186 RAGNet: Large-scale Reasoning-based Affordance Segmentation Benchmark towards General Grasping, *Dongming Wu, Yanping Fu, Saikhe Huang, Yingfei Liu, Fan Jia, Nian Liu, Feng Dai, Tiancai Wang, Rao Muhammad Anwer, Fahad Shahbaz Khan, Jianbing Shen*
- 187 DiTailListener: Controllable High Fidelity Listener Video Generation with Diffusion, *Maksim Siniukov, Di Chang, Minh Tran, Hongkun Gong, Ashutosh Chaubey, Mohammad Soleymani*
- 188 Hipandas: Hyperspectral Image Joint Denoising and Super-Resolution by Image Fusion with the Panchromatic Image, *Shuang Xu, Zixiang Zhao, Haowen Bai, Chang Yu, Jiangjun Peng, Xiangyong Cao, Deyu Meng*
- 189 You Think, You ACT: The New Task of Arbitrary Text to Motion Generation, *Runqi Wang, Caoyuan Ma, Guopeng Li, Hanrui Xu, Yuke Li, Zheng Wang*
- 190 EgoMusic-driven Human Dance Motion Estimation with Skeleton Mamba, *Quang Nguyen, Nhat Le, Baoru Huang, Minh Nhat Vu, Chengcheng Tang, Van Nguyen, Ngan Le, Thieu Vo, Anh Nguyen*
- 191 PersonaCraft: Personalized and Controllable Full-Body Multi-Human Scene Generation Using Occlusion-Aware 3D-Conditioned Diffusion, *Gwanghyun Kim, Suh Yoon Jeon, Seunggyu Lee, Se Young Chun*
- 192 DADM: Dual Alignment of Domain and Modality for Face Anti-spoofing, *Jingyi Yang, Xun Lin, Zitong Yu, Liepiao Zhang, Xin Liu, Hui Li, Xiaochen Yuan, Xiaochun Cao*
- 193 Autoregressive Denoising Score Matching is a Good Video Anomaly Detector, *Hanwen Zhang, Congqi Cao, Qinyi Lv, Lingtong Min, Yanning Zhang*
- 194 Ponimatter: Unfolding Interactive Pose for Versatile Human-human Interaction Animation, *Shaowei Liu, Chuan Guo, Bing Zhou, Jian Wang*
- 195 Monocular Facial Appearance Capture in the Wild, *Yingyan Xu, Kate Gadola, Prashanth Chandran, Sebastian Weiss, Markus Gross, Gaspard Zoss, Derek Bradley*
- 196 Avat3r: Large Animatable Gaussian Reconstruction Model for High-fidelity 3D Head Avatars, *Tobias Kirschstein, Javier Romero, Artem Sevastopolsky, Matthias Nießner, Shunsuke Saito*
- 197 Skeleton Motion Words for Unsupervised Skeleton-Based Temporal Action Segmentation, *Uzay Gökal, Federico Spurio, Dominik R. Bach, Juergen Gall*
- 198 MagicMotion: Controllable Video Generation with Dense-to-Sparse Trajectory Guidance, *Quanhao Li, Zhen Xing, Rui Wang, Hui Zhang, Qi Dai, Zuxuan Wu*
- 199 DH-FaceVid-1K: A Large-Scale High-Quality Dataset for Face Video Generation, *Donglin Di, He Feng, Wenzhang Sun, Yongjia Ma, Hao Li, Wei Chen, Lei Fan, Tonghua Su, Xun Yang*
- 200 Synthetic Video Enhances Physical Fidelity in Video Synthesis, *Qi Zhao, Xingyu Ni, Ziyu Wang, Feng Cheng, Ziyang Yang, Lu Jiang, Bohan Wang*
- 201 TimeBooth: Disentangled Facial Invariant Representation for Diverse and Personalized Face Aging, *Zepeng Su, Zhulin Liu, Zongyan Zhang, Tong Zhang, C.L. Philip Chen*
- 202 DanceEditor: Towards Iterative Editable Music-driven Dance Generation with Open-Vocabulary Descriptions, *Hengyuan Zhang, Zhe Li, Xingqun Qi, Mengze Li, Muiy Sun, Siye Wang, Man Zhang, Sirui Han*
- 203 Identity Preserving 3D Head Stylization with Multiview Score Distillation, *Bahri Batuhan Bilecen, Ahmet Berke Gökmen, Furkan Guzelant, Aysegül Dundar*

- 204 IDF: Iterative Dynamic Filtering Networks for Generalizable Image Denoising, *Dongjin Kim, Jaekyun Ko, Muhammad Kashif Ali, Tae Hyun Kim*
- 205 GDKVM: Echocardiography Video Segmentation via Spatiotemporal Key-Value Memory with Gated Delta Rule, *Rui Wang, Yimu Sun, Jingxing Guo, Huisi Wu, Jing Qin*
- 206 Who is a Better Talker: Subjective and Objective Quality Assessment for AI-Generated Talking Heads, *Yingjie Zhou, Jiezhong Cao, Zicheng Zhang, Farong Wen, Yanwei Jiang, Jun Jia, Xiaohong Liu, Xiongkuo Min, Guangtao Zhai*
- 207 Towards Efficient General Feature Prediction in Masked Skeleton Modeling, *Shengkai Sun, Zefan Zhang, Jianfeng Dong, Zhiyong Cheng, Xiaojun Chang, Meng Wang*
- 208 Scaling Action Detection: AdaTAD++ with Transformer-Enhanced Temporal-Spatial Adaptation, *Tanay Agrawal, Abid Ali, Antitza Dantcheva, Francois Bremond*
- 209 How Would It Sound? Material-Controlled Multimodal Acoustic Profile Generation for Indoor Scenes, *Mahnoor Fatima Saad, Ziad Al-Halah*
- 210 VideoSetDiff: Identifying and Reasoning Similarities and Differences in Similar Videos, *Yue Qiu, Yanjun Sun, Takuma Yagi, Shusaku Egami, Natsuki Miyata, Ken Fukuda, Kensho Hara, Ryusuke Sagawa*
- 211 MotionCtrl: A Real-time Controllable Vision-Language-Motion Model, *Bin Cao, Sipeng Zheng, Ye Wang, Lujie Xia, Qianshan Wei, Qin Jin, Jing Liu, Zongqing Lu*
- 212 Occlusion-robust Stylization for Drawing-based 3D Animation, *Sunjae Yoon, Gwanhyeong Koo, Younghwan Lee, Ji Woo Hong, Chang D. Yoo*
- 213 Rethinking Key-frame-based Micro-expression Recognition:
* A Robust and Accurate Framework Against Key-frame Errors, *Zheyuan Zhang, Weihao Tang, Hong Chen*
- 214 Video Individual Counting for Moving Drones, *Yaowu Fan, **
Jia Wan, Tao Han, Antoni B. Chan, Andy J. Ma
- 215 What Changed and What Could Have Changed? State-Change Counterfactuals for Procedure-Aware Video Representation Learning, *Chi-Hsi Kung, Frangil Ramirez, Juhyung Ha, Yi-Ting Chen, David Crandall, Yi-Hsuan Tsai*
- 216 Devil is in the Uniformity: Exploring Diverse Learners within Transformer for Image Restoration, *Shihao Zhou, Dayu Li, Jinshan Pan, Juncheng Zhou, Jinglei Shi, Jufeng Yang*
- 217 HADES: Human Avatar with Dynamic Explicit Hair Strands, *Zhanfeng Liao, Hanzhang Tu, Cheng Peng, Hongwen Zhang, Boyao Zhou, Yebin Liu*
- 218 FlowDPS : Flow-Driven Posterior Sampling for Inverse Problems, *Jeongsol Kim, Bryan Sangwoo Kim, Jong Chul Ye*
- 219 ZFusion: Efficient Deep Compositional Zero-shot Learning for Blind Image Super-Resolution with Generative Diffusion Prior, *Alireza Esmaeilzahi, Hossein Zareadar, Yapeng Tian, Laleh Seyyed-Kalantari*
- 220 Wavelet Policy: Lifting Scheme for Policy Learning in Long-Horizon Tasks, *Hao Huang, Shuaihang Yuan, Geeta Chandra Raju Bethala, Congcong Wen, Anthony Tzes, Yi Fang*
- 221 VLIPP: Towards Physically Plausible Video Generation with Vision and Language Informed Physical Prior, *Xindi Yang, Baolu Li, Yiming Zhang, Zhenfei Yin, Lei Bai, Liqian Ma, Zhiyong Wang, Jianfei Cai, Tien-Tsin Wong, Huchuan Lu, Xu Jia*
- 222 StreamDiffusion: A Pipeline-level Solution for Real-Time Interactive Generation, *Akio Kodaira, Chenfeng Xu, Toshiaki Hazama, Takanori Yoshimoto, Kohei Ohno, Shogo Mitsuho, Soichi Sugano, Hanying Cho, Zhijian Liu, Masayoshi Tomizuka, Kurt Keutzer*
- 223 HairCUP: Hair Compositional Universal Prior for 3D Gaussian
○ Avatars, *Byungjun Kim, Shunsuke Saito, Giljoo Nam, Tomas Simon, ☆*
Jason Saragih, Hanbyul Joo, Junxuan Li
- 224 Learning Streaming Video Representation via Multitask Training,
○ *Yibin Yan, Jilan Xu, Shangzhe Di, Yikun Liu, Yudi Shi, Qirui Chen, Zeqian Li, Yifei Huang, Weidi Xie*
- 225 DreamRelation: Relation-Centric Video Customization, *Yujie Wei, Shiwei Zhang, Hangjie Yuan, Biao Gong, Longxiang Tang, Xiang Wang, Haonan Qiu, Hengjia Li, Shuai Tan, Yingya Zhang, Hongming Shan*
- 226 ModSkill: Physical Character Skill Modularization, *Yiming Huang, Zhiyang Dou, Lingjie Liu*
- 227 Stable Virtual Camera: Generative View Synthesis with Diffusion Models, *Jensen Zhou, Hang Gao, Vikram Voleti, Aaryaman Vasishta, Chun-Han Yao, Mark Boss, Philip Torr, Christian Rupprecht, Varun Jampani*
- 228 * Fine-structure Preserved Real-world Image Super-resolution
via Transfer VAE Training, *Qiaosi Yi, Shuai Li, Rongyuan Wu, Lingchen Sun, Yuhui Wu, Lei Zhang*
- 229 Rethinking Bimanual Robotic Manipulation: Learning with Decoupled Interaction Framework, *Jian-Jian Jiang, Xiao-Ming Wu, Yi-Xiang He, Ling-An Zeng, Yi-Lin Wei, Dandan Zhang, Wei-Shi Zheng*
- 230 LA-MOTR: End-to-End Multi-Object Tracking by Learnable Association, *Peng Wang, Yongcai Wang, Hualong Cao, Wang Chen, Deying Li*
- 231 Free-Form Motion Control: Controlling the 6D Poses of Camera and Objects in Video Generation, *Xincheng Shuai, Henghui Ding, Zhenyuan Qin, Hao Luo, Xingjun Ma, Dacheng Tao*
- 232 Learning A Unified Template for Gait Recognition, *Panjuan Huang, Saihui Hou, Junzhou Huang, Yongzhen Huang*
- 233 Sliced Wasserstein Bridge for Open-Vocabulary Video Instance
* Segmentation, *Zheyun Qin, Deng Yu, Chuanchen Luo, Zhumin Chen*
- 234 Semantic Alignment and Reinforcement for Data-Free Quantization of Vision Transformers, *Yunshan Zhong, Yuyao Zhou, Yuxin Zhang, Wanchen Sui, Shen Li, Yong Li, Fei Chao, Rongrong Ji*
- 235 NeurOp-Diff: Continuous Remote Sensing Image Super-Resolution via Neural Operator Diffusion, *Zihao Xu, Yuzhi Tang, Bowen Xu, Qingquan Li*
- 236 Gaussian Variation Field Diffusion for High-fidelity Video-to-4D Synthesis, *Bowen Zhang, Sicheng Xu, Chuxin Wang, Jiaolong Yang, Feng Zhao, Dong Chen, Baining Guo*
- 237 Synchronization of Multiple Videos, *Avihai Naaman, Ron Shapira Weber, Oren Freifeld*
- 238 DeepShield: Fortifying Deepfake Video Detection with Local and
☆ Global Forgery Analysis, *Yinqi Cai, Jichang Li, Zhaolun Li, Weikai Chen, Rushi Lan, Xi Xie, Xiaonan Luo, Guanbin Li*
- 239 Perceiving and Acting in First-Person: A Dataset and Benchmark for Egocentric Human-Object-Human Interactions, *Liang Xu, Chengqun Yang, Zili Lin, Fei Xu, Yifan Liu, Congsheng Xu, Yiyi Zhang, Jie Qin, Xingdong Sheng, Yunhui Liu, Xin Jin, Yichao Yan, Wenjun Zeng, Xiaokang Yang*
- 240 Audio-visual Controlled Video Diffusion with Masked Selective State Spaces Modeling for Natural Talking Head Generation, *Fa-Ting Hong, Zunnan Xu, Zixiang Zhou, Jun Zhou, Xiu Li, Qin Lin, Qinglin Lu, Dan Xu*
- 241 SCORE: Scene Context Matters in Open-Vocabulary Remote
* Sensing Instance Segmentation, *Shiqi Huang, Shuting He, Huaiyuan Qin, Bihan Wen*
- 242 VertexRegen: Mesh Generation with Continuous Level of Detail, *Xiang Zhang, Yawar Siddiqui, Armen Avetisyan, Chris Xie, Jakob Engel, Henry Howard-Jenkins*
- 243 Multimodal Prompt Alignment for Facial Expression Recognition, *Fuyan Ma, Yiran He, Bin Sun, Shutao Li*
- 244 Vivid4D: Improving 4D Reconstruction from Monocular Video by
☆ Video Inpainting, *Jiaxin Huang, Sheng Miao, Bangbang Yang, Yewen Ma, Yiyi Liao*
- 245 AU-Blendshape for Fine-grained Stylized 3D Facial Expression Manipulation, *Hao Li, Ju Dai, Feng Zhou, Kaida Ning, Lei Li, Junjun Pan*
- 246 GestureHYDRA: Semantic Co-speech Gesture Synthesis via Hybrid Modality Diffusion Transformer and Cascaded-Synchronized Retrieval-Augmented Generation, *Quanwei Yang, Luying Huang, Kaisiyuan Wang, Jiazhi Guan, Shengyi He, Fengguo Li, Hang Zhou, Lingyun Yu, Yingying Li, Haocheng Feng, Hongtao Xie*
- 247 FoundIR: Unleashing Million-scale Training Data to Advance
☆ Foundation Models for Image Restoration, *Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, Jinshan Pan*
- 248 Highlight What You Want: Weakly-Supervised Instance-Level Controllable Infrared-Visible Image Fusion, *Zeyu Wang, Jizheng Zhang, Haiyu Song, Mingyu Ge, Jiayu Wang, Haoran Duan*
- 249 Adaptive Hyper-Graph Convolution Network for Skeleton-based Human Action Recognition with Virtual Connections, *Youwei Zhou, Tianyang Xu, Cong Wu, Xiaojun Wu, Josef Kittler*

- 250 Weakly Supervised Visible-Infrared Person Re-Identification via Heterogeneous Expert Collaborative Consistency Learning, *Yafei Zhang, Lingqi Kong, Huafeng Li, Jie Wen*
- 251 PERSONA: Personalized Whole-Body 3D Avatar with Pose-Driven Deformations from a Single Image, *Geonhee Sim, Gyeongsik Moon*
- 252 Derm1M: A Million-scale Vision-Language Dataset Aligned
* with Clinical Ontology Knowledge for Dermatology, *Siyuan Yan, Ming Hu, Yiwen Jiang, Xieji Li, Hao Fei, Philipp Tschandl, Harald Kittler, Zongyuan Ge*
- 253 FaceLift: Learning Generalizable Single Image 3D Face Reconstruction from Synthetic Heads, *Weijie Lyu, Yi Zhou, Ming-Hsuan Yang, Zhixin Shu*
- 254 DexH2R: A Benchmark for Dynamic Dexterous Grasping in Human-to-Robot Handover, *Youzhuo Wang, Jiayi Ye, Chuyang Xiao, Yiming Zhong, Heng Tao, Hang Yu, Yumeng Liu, Jingyi Yu, Yuexin Ma*
- 255 Precise Action-to-Video Generation Through Visual Action Prompts, *Yuang Wang, Chao Wen, Haoyu Guo, Sida Peng, Minghan Qin, Hujun Bao, Xiaowei Zhou, Ruizhen Hu*
- 256 PRIMAL: Physically Reactive and Interactive Motor Model for Avatar
☆ Learning, *Yan Zhang, Yao Feng, Alpár Cseke, Nitin Saini, Nathan Bajandas, Nicolas Heron, Michael J. Black*
- 257 MagicID: Hybrid Preference Optimization for ID-Consistent and Dynamic-Preserved Video Customization, *Hengjia Li, Lifan Jiang, Xi Xiao, Tianyang Wang, Hongwei Yi, Boxi Wu, Deng Cai*
- 258 Consistency Trajectory Matching for One-Step Generative Super-Resolution, *Weiyi You, Mingyang Zhang, Leheng Zhang, Xingyu Zhou, Kexuan Shi, Shuhang Gu*
- 259 Bridging the Skeleton-Text Modality Gap: Diffusion-Powered Modality Alignment for Zero-shot Skeleton-based Action Recognition, *Jeonghyeok Do, Munchul Kim*
- 260 Visual Chronicles: Using Multimodal LLMs to Analyze
* Massive Collections of Images, *Boyang Deng, Songyou Peng, Kyle Genova, Gordon Wetzstein, Noah Snaveley, Leonidas Guibas, Thomas Funkhouser*
- 261 Latent-Reframe: Enabling Camera Control for Video Diffusion Models without Training, *Zhenghong Zhou, Jie An, Jiebo Luo*
- 262 Neuromanifold-Regularized KANs for Shape-fair Feature Representations, *Mazlum Ferhat Arslan, Weihong Guo, Shuo Li*
- 263 Learning to Generalize without Bias for Open-Vocabulary Action
* Recognition, *Yating Yu, Congqi Cao, Yifan Zhang, Yanning Zhang*
- 264 GeoAvatar: Adaptive Geometrical Gaussian Splatting for 3D Head Avatar, *SeungJun Moon, Hah Min Lew, Seungeun Lee, Ji-Su Kang, Gyeong-Moon Park*
- 265 MotionFollower: Editing Video Motion via Score-Guided Diffusion, *Shuyuan Tu, Qi Dai, Zihao Zhang, Sicheng Xie, Zhi-Qi Cheng, Chong Luo, Xintong Han, Zuxuan Wu, Yu-Gang Jiang*
- 266 InterSyn: Interleaved Learning for Dynamic Motion Synthesis in the Wild, *Yiyi Ma, Yuanzhi Liang, Xiu Li, Chi Zhang, Xuelong Li*
- 267 MoFRR: Mixture of Diffusion Models for Face Retouching Restoration, *Jiaxin Liu, Qichao Ying, Zhenxing Qian, Sheng Li, Runqi Zhang, Jian Liu, Xinpeng Zhang*
- 268 D3: Training-Free AI-Generated Video Detection Using Second-Order Features, *Chende Zheng, Ruiqi Suo, Chenhao Lin, Zhengyu Zhao, Le Yang, Shuai Liu, Minghui Yang, Cong Wang, Chao Shen*
- 269 Image Intrinsic Scale Assessment: Bridging the Gap Between Quality and Resolution, *Vlad Hosu, Lorenzo Agnolucci, Daisuke Iso, Dietmar Saupe*
- 270 Frequency-Guided Posterior Sampling for Diffusion-Based Image
☆ Restoration, *Darshan Thaker, Abhishek Goyal, Rene Vidal*
- 271 GAS: Generative Avatar Synthesis from a Single Image, *Yixing Lu, Juntong Dong, Youngjoong Kwon, Qin Zhao, Bo Dai, Fernando De la Torre*
- 272 Less Static, More Private: Towards Transferable Privacy-Preserving Action Recognition by Generative Decoupled Learning, *Zhi-Wei Xia, Kun-Yu Lin, Yuan-Ming Li, Wei-Jin Huang, Xian-Tuo Tan, Wei-Shi Zheng*
- 273 Bitrate-Controlled Diffusion for Disentangling Motion and Content in Video, *Xiao Li, Qi Chen, Xiulian Peng, Kai Yu, Xie Chen, Yan Lu*
- 274 MR-FIQA: Face Image Quality Assessment with Multi-Reference Representations from Synthetic Data Generation, *Fu-Zhao Ou, Chongyi Li, Shiqi Wang, Sam Kwong*
- 275 Text-to-Any-Skeleton Motion Generation Without Retargeting, *Qingyuan Liu, Ke Lv, Kun Dong, Jian Xue, Zehai Niu, Jinbao Wang*
- 276 Blind2Sound: Self-Supervised Image Denoising without Residual Noise, *Jiazhen Liu, Zejin Wang, Bohao Chen, Hua Han*
- 277 STaR: Seamless Spatial-Temporal Aware Motion Retargeting with Penetration and Consistency Constraints, *Xiaohang Yang, Qing Wang, Jiahao Yang, Gregory Slabaugh, Shanxin Yuan*
- 278 ARIG: Autoregressive Interactive Head Generation for Real-time Conversations, *Ying Guo, Xi Liu, Cheng Zhen, Pengfei Yan, Xiaoming Wei*
- 279 Towards a Universal Image Degradation Model via Content-Degradation Disentanglement, *Wenbo Yang, Zhongling Wang, Zhou Wang*
- 280 Unified Multimodal Understanding via Byte-Pair Visual Encoding,
* *Wanpeng Zhang, Yicheng Feng, Hao Luo, Yijiang Li, Zihao Yue, Sipeng Zheng, Zongqing Lu*
- 281 IMoRe: Implicit Program-Guided Reasoning for Human Motion Q&A, *Chen Li, Chinthani Sugandhika, Yeo Keat Ee, Eric Peh, Hao Zhang, Hong Yang, Deepu Rajan, Basura Fernando*
- 282 AdaDCP: Learning an Adapter with Discrete Cosine Prior for Clear-to-Adverse Domain Generalization, *Qi Bi, Yixian Shen, Jingjun Yi, Gui-Song Xia*
- 283 MP-HSIR: A Multi-Prompt Framework for Universal Hyperspectral Image Restoration, *Zhehui Wu, Yong Chen, Naoto Yokoya, Wei He*
- 284 MorphoGen: Efficient Unconditional Generation of Long-Range Projection Neuronal Morphology via a Global-to-Local Framework, *Tianfang Zhu, Hongyang Zhou, Anan Li*
- 285 Adaptive Articulated Object Manipulation On The Fly with Foundation Model Reasoning and Part Grounding, *Xiaojie Zhang, Yuanfei Wang, Ruihai Wu, Kunqi Xu, Yu Li, Liuyu Xiang, Hao Dong, Zhao Feng He*
- 286 Scoring, Remember, and Reference: Catching Camouflaged Objects in Videos, *Yu'ang Feng, Shuyong Gao, Fuzhen Yan, Yicheng Song, Lingyi Hong, Junjie Hu, Wenqiang Zhang*
- 287 Stylized-Face: A Million-level Stylized Face Dataset for Face Recognition, *Zhengyuan Peng, Jianqing Xu, Yuge Huang, Jinkun Hao, Shouhong Ding, Zhizhong Zhang, Xin Tan, Lizhuang Ma*
- 288 GaussianSpeech: Audio-Driven Personalized 3D Gaussian
☆ Avatars, *Shivangi Aneja, Artem Sevastopolsky, Tobias Kirschstein, Justus Thies, Angela Dai, Matthias Nießner*
- 289 A Quality-Guided Mixture of Score-Fusion Experts Framework for Human Recognition, *Jie Zhu, Yiyang Su, Minchul Kim, Anil Jain, Xiaoming Liu*
- 290 VMBench: A Benchmark for Perception-Aligned Video Motion Generation, *Xinran Ling, Chen Zhu, Meiqi Wu, Hangyu Li, Xiaokun Feng, Cundian Yang, Aiming Hao, Jiashu Zhu, Jiahong Wu, Xiangxiang Chu*
- 291 Capturing head avatar with hand contacts from a monocular video, *Haonan He, Yufeng Zheng, Jie Song*
- 292 Tiling artifacts and trade-offs of feature normalization in
* the segmentation of large biological images, *Elena Buglakova, Anwai Archit, Edoardo D'Imprima, Julia Mahamid, Constantin Pape, Anna Kreshuk*
- 293 BlueNeg: A 35mm Negative Film Dataset for Restoring Channel-Heterogeneous Deterioration, *Hanyuan Liu, Chengze Li, Minshan Xie, Zhenni Wang, Jiawen Liang, Chi-Sing Leung, Tien-Tsin Wong*
- 294 GenM3: Generative Pretrained Multi-path Motion Model for Text Conditional Human Motion Generation, *Junyu Shi, Lijiang Liu, Yong Sun, Zhiyuan Zhang, Jinni Zhou, Qiang Nie*
- 295 Saliency-Aware Quantized Imitation Learning for Efficient Robotic Control, *Seongmin Park, Hyungmin Kim, Sangwoo Kim, Wonseok Jeon, Juyoung Yang, Byeongwook Jeon, Yoonseon Oh, Jungwook Choi*
- 296 MAVFlow: Preserving Paralinguistic Elements with Conditional Flow Matching for Zero-Shot AV2AV Multilingual Translation, *Sungwoo Cho, Jeongsoo Choi, Sungnyun Kim, Se-Young Yun*
- 297 Privacy-centric Deep Motion Retargeting for Anonymization of Skeleton-Based Motion Visualization, *Thomas Carr, Depeng Xu, Shuhan Yuan, Aidong Lu*
- 298 ChartCap: Mitigating Hallucination of Dense Chart Captioning,
* *Junyoung Lim, Jaewoo Ahn, Gunhee Kim*

- 299 GenFlowRL: Shaping Rewards with Generative Object-Centric Flow in Visual Reinforcement Learning, *Kelin Yu, Sheng Zhang, Harshit Soora, Furong Huang, Heng Huang, Pratap Tokekar, Ruohan Gao*
- 300 Align Your Rhythm: Generating Highly Aligned Dance Poses with Gating-Enhanced Rhythm-Aware Feature Representation, *Congyi Fan, Jian Guan, Xuanjia Zhao, Dongli Xu, Youtian Lin, Tong Ye, Pengming Feng, Haiwei Pan*
- 301 GMMamba: Group Masking Mamba for Whole Slide Image Classification, *Tingting Zheng, Hongxun Yao, Kui Jiang, Yi Xiao, Sicheng Zhao*
- 302 Understanding Co-speech Gestures in-the-wild, *Sindhu B Hegde, K R Prajwal, Taein Kwon, Andrew Zisserman*
- 303 Motion Synthesis with Sparse and Flexible Keyjoint Control, *Inwoo Hwang, Jinseok Bae, Donggeun Lim, Young Min Kim*
- 304 UniPhys: Unified Planner and Controller with Diffusion for Flexible Physics-Based Character Control, *Yan Wu, Korrawe Karunratanakul, Zhengyi Luo, Siyu Tang*
- 305 MMAD: Multi-label Micro-Action Detection in Videos, *Kun Li, Pengyu Liu, Dan Guo, Fei Wang, Zhiliang Wu, Hehe Fan, Meng Wang*
- 306 UniRes: Universal Image Restoration for Complex Degradations, *Mo Zhou, Keren Ye, Mauricio Delbracio, Peyman Milanfar, Vishal M. Patel, Hossein Talebi*
- 307 SV4D 2.0: Enhancing Spatio-Temporal Consistency in Multi-View Video Diffusion for High-Quality 4D Generation, *Chun-Han Yao, Yiming Xie, Vikram Voleti, Huaizu Jiang, Varun Jampani*
- 308 Gait-X: Exploring X modality for Generalized Gait Recognition, *Zengbin Wang, Saihui Hou, Junjie Li, Xu Liu, Chunshui Cao, Yongzhen Huang, Siye Wang, Man Zhang*
- 309 MVTrajecter: Multi-View Pedestrian Tracking with Trajectory Motion Cost and Trajectory Appearance Cost, *Taiga Yamane, Ryo Masumura, Satoshi Suzuki, Shota Orihashi*
- 310 A Plug-and-Play Physical Motion Restoration Approach for In-the-Wild High-Difficulty Motions, *Youliang Zhang, Ronghui Li, Yachao Zhang, Liang Pan, Jingbo Wang, Yebin Liu, Xiu Li*
- 311 GazeGaussian: High-Fidelity Gaze Redirection with 3D Gaussian Splatting, *Xiaobao Wei, Peng Chen, Guangyu Li, Ming Lu, Hui Chen, Feng Tian*
- 312 MoGA: 3D Generative Avatar Prior for Monocular Gaussian Avatar Reconstruction, *Zijian Dong, Longteng Duan, Jie Song, Michael J. Black, Andreas Geiger*
- 313 Light-A-Video: Training-free Video Relighting via Progressive Light Fusion, *Yujie Zhou, Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, Qidong Huang, Jinsong Li, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Anyi Rao, Jiaqi Wang, Li Niu*
- 314 SVG-Head: Hybrid Surface-Volumetric Gaussians for High-Fidelity Head Reconstruction and Real-Time Editing, *Heyi Sun, Cong Wang, Tian-Xing Xu, Jingwei Huang, Di Kang, Chunchao Guo, Song-Hai Zhang*
- 315 Go to Zero: Towards Zero-shot Motion Generation with Million-scale Data, *Ke Fan, Shunlin Lu, Minyue Dai, Runyi Yu, Lixing Xiao, Zhiyang Dou, Junting Dong, Lizhuang Ma, Jingbo Wang*
- 316 StyleMotif: Multi-Modal Motion Stylization using Style-Content Cross Fusion, *Ziyu Guo, Young Yoon Lee, Joseph Liu, Yizhak Ben-Shabat, Victor Zordan, Mubbasir Kapadia*
- 317 I2V3D: Controllable Image-to-video Generation with 3D Guidance, *Zhiyuan Zhang, Dongdong Chen, Jing Liao*
- 318 Group-wise Scaling and Orthogonal Decomposition for Domain-Invariant Feature Extraction in Face Anti-Spoofing, *Seungjin Jung, Kanghee Lee, Yonghyun Jeong, Haeun Noh, Jungmin Lee, Jongwon Choi*
- 319 FakeRadar: Probing Forgery Outliers to Detect Unknown Deepfake Videos, *Zhaolun Li, Jichang Li, Yinqi Cai, Junye Chen, Xiaonan Luo, Guanbin Li, Rushi Lan*
- 320 StrandHead: Text to Hair-Disentangled 3D Head Avatars Using Human-Centric Priors, *Xiaokun Sun, Zeyu Cai, Ying Tai, Jian Yang, Zhenyu Zhang*
- 321 Puppet-Master: Scaling Interactive Video Generation as a Motion Prior for Part-Level Dynamics, *Ruining Li, Chuanxia Zheng, Christian Rupprecht, Andrea Vedaldi*
- 322 CameraCtrl II: Dynamic Scene Exploration via Camera-controlled Video Diffusion Models, *Hao He, Ceyuan Yang, Shanchuan Lin, Yinghao Xu, Meng Wei, Liangke Gui, Qi Zhao, Gordon Wetzstein, Lu Jiang, Hongsheng Li*
- 323 General Compression Framework for Efficient Transformer Object Tracking, *Lingyi Hong, Jinglun Li, Xinyu Zhou, Shilin Yan, Pinxue Guo, Kaixun Jiang, Zhaoyu Chen, Shuyong Gao, Runze Li, Xingdong Sheng, Wei Zhang, Hong Lu, Wenqiang Zhang*
- 324 DynamicFace: High-Quality and Consistent Face Swapping for Image and Video using Composable 3D Facial Priors, *Runqi Wang, Yang Chen, Sijie Xu, Tianyao He, Wei Zhu, Dejia Song, Nemo Chen, Xu Tang, Yao Hu*
- 325 StreamMind: Unlocking Full Frame Rate Streaming Video Dialogue through Event-Gated Cognition, *Xin Ding, Hao Wu, Yifan Yang, Shiqi Jiang, Qianxi Zhang, Donglin Bai, Zhibo Chen, Ting Cao*
- 326 CARP: Visuomotor Policy Learning via Coarse-to-Fine Autoregressive Prediction, *Zhefei Gong, Pengxiang Ding, Shangke Lyu, Siteng Huang, Mingyang Sun, Wei Zhao, Zhaoxin Fan, Donglin Wang*
- 327 Unlocking the Potential of Diffusion Priors in Blind Face Restoration, *Yunqi Miao, Zhiyu Qu, Mingqi Gao, Changrui Chen, Jifei Song, Jungong Han, Jiankang Deng*
- 328 Bridging Diffusion Models and 3D Representations: A 3D Consistent Super-Resolution Framework, *Yi-Ting Chen, Ting-Hsuan Liao, Pengsheng Guo, Alexander Schwing, Jia-Bin Huang*
- 329 A₂: An Affordance-Aware Hierarchical Model for General Robotic Manipulation, *Rongtao Xu, Jian Zhang, Minghao Guo, Youpeng Wen, Haoting Yang, Min Lin, Jianzheng Huang, Zhe Li, Kaidong Zhang, Liqiong Wang, Yuxuan Kuang, Meng Cao, Feng Zheng, Xiaodan Liang*
- 330 DisenQ: Disentangling Q-Former for Activity-Biometrics, *Shehreen Azad, Yogesh Singh Rawat*
- 331 The Source Image is the Best Attention for Infrared and Visible Image Fusion, *Song Wang, Xie Han, Liqun Kuang, Boying Wang, Zhongyu Chen, Zherui Qiao, Fan Yang, Xiaoxia Liu, Bingyu Zhang, Zhixun Wang*
- 332 HUST: High-Fidelity Unbiased Skin Tone Estimation via Texture Quantization, *Zimin Ran, Xingyu Ren, Xiang An, Kaicheng Yang, Ziyong Feng, Jing Yang, Rolandos Alexandros Potamias, Linchao Zhu, Jiankang Deng*
- 333 AdaHuman: Animatable Detailed 3D Human Generation with Compositional Multiview Diffusion, *Yangyi Huang, Ye Yuan, Xueting Li, Jan Kautz, Umar Iqbal*
- 334 Trokens: Semantic-Aware Relational Trajectory Tokens for Few-Shot Action Recognition, *Pulkit Kumar, Shuaiyi Huang, Matthew Walmer, Sai Saketh Rambhatla, Abhinav Shrivastava*
- 335 AnimateAnyMesh: A Feed-Forward 4D Foundation Model for Text-Driven Universal Mesh Animation, *Zijie Wu, Chaohui Yu, Fan Wang, Xiang Bai*
- 336 Auto-Regressive Transformation for Image Alignment, *Kanggeon Lee, Soochahn Lee, Kyoung Mu Lee*
- 337 Controllable Weather Synthesis and Removal with Video Diffusion Models, *Chih-Hao Lin, Zian Wang, Ruofan Liang, Yuxuan Zhang, Sanja Fidler, Shenlong Wang, Zan Gojcic*
- 338 Sequential Gaussian Avatars with Hierarchical Motion Context, *Wangze Xu, Yifan Zhan, Zhihang Zhong, Xiao Sun*
- 339 TokenUnify: Scaling Up Autoregressive Pretraining for Neuron Segmentation, *Yinda Chen, Haoyuan Shi, Xiaoyu Liu, Te Shi, Ruobing Zhang, Dong Liu, Zhiwei Xiong, Feng Wu*
- 340 SAM2Long: Enhancing SAM 2 for Long Video Segmentation with a Training-Free Memory Tree, *Shuangrui Ding, Rui Qian, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Yuwei Guo, Dahua Lin, Jiaqi Wang*
- 341 T2Bs: Text-to-Character Blendshapes via Video Generation, *Jiahao Luo, Chaoyang Wang, Michael Vasilkovsky, Vladislav Shakhrai, Di Liu, Peiye Zhuang, Sergey Tulyakov, Peter Wonka, Hsin-Ying Lee, James Davis, Jian Wang*
- 342 TACO: Taming Diffusion for in-the-wild Video Amodal Completion, *Ruijie Lu, Yixin Chen, Yu Liu, Jiaxiang Tang, Junfeng Ni, Diwen Wan, Gang Zeng, Siyuan Huang*
- 343 Unfolding-Associative Encoder-Decoder Network with Progressive Alignment for Pansharpening, *Shijie Fang, Hongping Gan*
- 344 AnyBimanual: Transferring Unimanual Policy for General Bimanual Manipulation, *Guanxing Lu, Tengbo Yu, Haoyuan Deng, Season Si Chen, Yansong Tang, Ziwei Wang*
- 345 MOERL: When Mixture-of-Experts Meet Reinforcement Learning for

- Adverse Weather Image Restoration, *Tao Wang, Peiwen Xia, Bo Li, Peng-Tao Jiang, Zhe Kong, Kaihao Zhang, Tong Lu, Wenhan Luo*
- 346 LD-RPS: Zero-Shot Unified Image Restoration via Latent Diffusion Recurrent Posterior Sampling, *Huaqiu Li, Yong Wang, Tongwen Huang, Hailang Huang, Haoqian Wang, Xiangxiang Chu*
- 347 DimensionX: Create Any 3D and 4D Scenes from a Single Image with Decoupled Video Diffusion, *Wenqiang Sun, Shuo Chen, Fangfu Liu, Zilong Chen, Yueqi Duan, Jun Zhu, Jun Zhang, Yikai Wang*
- 348 RoboTron-Mani: All-in-One Multimodal Large Model for Robotic Manipulation, *Feng Yan, Fanfan Liu, Yiyang Huang, Zechao Guan, Liming Zheng, Yufeng Zhong, Chengjian Feng, Lin Ma*
- 349 LOMM: Latest Object Memory Management for Temporally Consistent Video Instance Segmentation, *Seunghun Lee, Jiwan Seo, Minwoo Choi, Kiljoon Han, Jahoon Jeong, Zane Durante, Ehsan Adeli, Sang Hyun Park, Sunghoon Im*
- 350 Video Motion Graphs, *Haiyang Liu, Zhan Xu, Fa-Ting Hong, * Hsin-Ping Huang, Yi Zhou, Yang Zhou*
- 351 Online Generic Event Boundary Detection, *Hyungrok Jung, Daneul Kim, Seunggyun Lim, Jeany Son, Jonghyun Choi*
- 352 SD2Actor: Continuous State Decomposition via Diffusion Embeddings for Robotic Manipulation, *Jiayi Li*
- 353 SemTalk: Holistic Co-speech Motion Generation with Frame-level Semantic Emphasis, *Xiangyue Zhang, Jianfang Li, Jiaxu Zhang, Ziqiang Dang, Jianqiang Ren, Liefeng Bo, Zhigang Tu*
- 354 DuoCLR: Dual-Surrogate Contrastive Learning for Skeleton-based Human Action Segmentation, *Haitao Tian*
- 355 VoluMe – Authentic 3D Video Calls from Live Gaussian Splat Prediction, *Martin de La Gorce, Charlie Hewitt, Tibor Takács, Robert Gerdisch, Zafirah Hosenie, Givi Meishvili, Marek Kowalski, Thomas J. Cashman, Antonio Criminisi*
- 356 EVDM: Event-based Real-world Video Deblurring with Mamba, *Zhijing Sun, Senyan Xu, Kean Liu, Runze Tian, Xueyang Fu, Zheng-Jun Zha*
- 357 From Enhancement to Understanding: Build a Generalized Bridge for Low-light Vision via Semantically Consistent Unsupervised Fine-tuning, *Sen Wang, Shao Zeng, Tianjun Gu, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, Lizhuang Ma*
- 358 Physical Degradation Model-Guided Interferometric Hyperspectral Reconstruction with Unfolding Transformer, *Yuansheng Li, Yunhao Zou, Linwei Chen, Ying Fu*
- 359 Intra-modal and Cross-modal Synchronization for Audio-visual Deepfake Detection and Temporal Localization, *Ashutosh Anshul, Shreyas Gopal, Deepu Rajan, Eng Siong Chng*
- 360 FineMotion: A Dataset and Benchmark with both Spatial and Temporal Annotation for Fine-grained Motion Generation and Editing, *Bizhu Wu, Jinheng Xie, Meidan Ding, Zhe Kong, Jianfeng Ren, Ruibin Bai, Rong Qu, Linlin Shen*
- 361 OmniHuman-1: Rethinking the Scaling-Up of One-Stage * Conditioned Human Animation Models, *Gaojie Lin, Jianwen Jiang, Jiaqi Yang, Zerong Zheng, Chao Liang, Yuan Zhang, Jingtuo Liu*
- 362 Context-Aware Academic Emotion Dataset and Benchmark, *Luming Zhao, Jingwen Xuan, Jiamin Lou, Yonghui Yu, Wenwu Yang*
- 363 MotionLab: Unified Human Motion Generation and Editing via the Motion-Condition-Motion Paradigm, *Ziyan Guo, Zeyu Hu, De Wen Soh, Na Zhao*
- 364 Gradient-Reweighted Adversarial Camouflage for Physical Object Detection Evasion, *Jiawei Liang, Siyuan Liang, Tianrui Lou, Ming Zhang, Wenjin Li, Dunqiu Fan, Xiaochun Cao*
- 365 iManip: Skill-Incremental Learning for Robotic Manipulation, ☆ *Zexin Zheng, Jia-Feng Cai, Xiao-Ming Wu, Yi-Lin Wei, Yu-Ming Tang, Ancong Wu, Wei-Shi Zheng*
- 366 Q-Norm: Robust Representation Learning via Quality-Adaptive ☆ Normalization, *Lanning Zhang, Ying Zhou, Fei Gao, Ziyun Li, Maoying Qiao, Jinlan Xu, Nannan Wang*
- 367 Proxy-Bridged Game Transformer for Interactive Extreme Motion Prediction, *Yanwen Fang, Wenqi Jia, Xu Cao, Peng-Tao Jiang, Guodong Li, Jintai Chen*
- 368 MeshAnything V2: Artist-Created Mesh Generation with Adjacent Mesh Tokenization, *Yiwen Chen, Yikai Wang, Yihao Luo, Zhengyi Wang, Zilong Chen, Jun Zhu, Chi Zhang, Guosheng Lin*
- 369 MDD: A Dataset for Text-and-Music Conditioned Duet Dance Generation, *Prerit Gupta, Jason Alexander Fotso-Puepi, Zhengyuan Li, Jay Mehta, Aniket Bera*
- 370 π -AVAS: Can Physics-Integrated Audio-Visual Modeling Boost Neural Acoustic Synthesis?, *Susan Liang, Chao Huang, Yunlong Tang, Zeliang Zhang, Chenliang Xu*
- 371 Exploring Weather-aware Aggregation and Adaptation for Semantic Segmentation under Adverse Conditions, *Yuwen Pan, Rui Sun, Wangkai Li, Tianzhu Zhang*
- 372 SemGes: Semantics-aware Co-Speech Gesture Generation using Semantic Coherence and Relevance Learning, *Linmiao Liu, Esam Ghaleb, Asli Ozyurek, Zerrin Yumak*
- 373 Metric Convolutions: A Unifying Theory to Adaptive Image Convolutions, *Thomas Dagès, Michael Lindenbaum, Alfred M. Bruckstein*
- 374 RobAVA: A Large-scale Dataset and Baseline Towards Video based Robotic Arm Action Understanding, *Baoli Sun, Ning Wang, Xinzhu Ma, Anqi Zou, Yihang Lu, Chuixuan Fan, Zhihui Wang, Kun Lu, Zhiyong Wang*
- 375 IDFace: Face Template Protection for Efficient and Secure Identification, *Sunpill Kim, Seunghun Paik, Chanwoo Hwang, Dongsoo Kim, Junbum Shin, Jae Hong Seo*
- 376 Learning Visual Hierarchies in Hyperbolic Space for Image Retrieval, *Ziwei Wang, Sameera Ramasinghe, Chenchen Xu, Julien Monteil, Loris Bazzani, Thalaisyasingam Ajanthan*
- 377 DPoser-X: Diffusion Model as Robust 3D Whole-body Human Pose Prior, *Junzhe Lu, Jing Lin, Hongkun Dou, Ailing Zeng, Yue Deng, Xian Liu, Zhongang Cai, Lei Yang, Yulun Zhang, Haoqian Wang, Ziwei Liu*
- 378 Dual-level Prototype Learning for Composite Degraded Image Restoration, *Zhongze Wang, Haitao Zhao, Lujian Yao, Jingchao Peng, Kaijie Zhao*
- 379 BVINet: Unlocking Blind Video Inpainting with Zero Annotations, *Zhiliang Wu, Kerui Chen, Kun Li, Hehe Fan, Yi Yang*
- 380 HumanSAM: Classifying Human-centric Forgery Videos in Human Spatial, Appearance, and Motion Anomaly, *Chang Liu, Yunfan Ye, Fan Zhang, Qingyang Zhou, Yuchuan Luo, Zhiping Cai*
- 381 DreamDance: Animating Human Images by Enriching 3D Geometry Cues from 2D Poses, *Yatian Pang, Bin Zhu, Bin Lin, Mingzhe Zheng, Francis E. H. Tay, Ser-Nam Lim, Harry Yang, Li Yuan*
- 382 I2VControl: Disentangled and Unified Video Motion Synthesis Control, *Wanquan Feng, Tianhao Qi, Jiawei Liu, Mingzhen Sun, Pengqi Tu, Tianxiang Ma, Fei Dai, Songtao Zhao, Siyu Zhou, Qian He*
- 383 MeshLLM: Empowering Large Language Models to Progressively Understand and Generate 3D Mesh, *Shuangkang Fang, I-Chao Shen, Yufeng Wang, Yi-Hsuan Tsai, Yi Yang, Shuchang Zhou, Wenrui Ding, Takeo Igarashi, Ming-Hsuan Yang*
- 384 On-Device Diffusion Transformer Policy for Efficient Robot Manipulation, *Yiming Wu, Huan Wang, Zhenghao Chen, Jianxin Pang, Dong Xu*
- 385 Generic Event Boundary Detection via Denoising Diffusion, *Jaeeun Hwang, Dayoung Gong, Manjin Kim, Minsu Cho*
- 386 Learning Pixel-adaptive Multi-layer Perceptrons for Real-time Image Enhancement, *Junyu Lou, Xiaorui Zhao, Kexuan Shi, Shuhang Gu*
- 387 Multi-Modal Few-Shot Temporal Action Segmentation, *Zijia Lu, Ehsan Elhamifar*
- 388 SIMS: Simulating Stylized Human-Scene Interactions with Retrieval-Augmented Script Generation, *Wenjia Wang, Liang Pan, Zhiyang Dou, Jidong Mei, Zhouyingcheng Liao, Yuke Lou, Yifan Wu, Lei Yang, Jingbo Wang, Taku Komura*
- 389 ProbRes: Probabilistic Jump Diffusion for Open-World Egocentric Activity Recognition, *Sanjoy Kundu, Shanmukha Vellamcheti, Sathyanarayanan N. Aakur*
- 390 Unified Adversarial Augmentation for Improving Palmprint Recognition, *Jianlong Jin, Chenglong Zhao, Ruixin Zhang, Sheng Shang, Yang Zhao, Jun Wang, Jingyun Zhang, Shouhong Ding, Wei Jia, Yunsheng Wu*
- 391 Not All Degradations Are Equal: A Targeted Feature Denoising Framework for Generalizable Image Super-Resolution, *Hongjun Wang, Jiyeun Chen, Zhengwei Yin, Xuan Song, Yinqiang Zheng*

- 392 SHeaP: Self-Supervised Head Geometry Predictor Learned via 2D Gaussians, *Liam Schoneveld, Zhe Chen, Davide Davoli, Jiapeng Tang, Saimon Terazawa, Ko Nishino, Matthias Nießner*
- 393 MultiModal Action Conditioned Video Simulation, *Yichen Li, Antonio Torralba*
- 394 LHM: Large Animatable Human Reconstruction Model for Single Image to 3D in Seconds, *Lingteng Qiu, Xiaodong Gu, Peihao Li, Qi Zuo, Weichao Shen, Junfei Zhang, Kejie Qiu, Weihao Yuan, Guanying Chen, Zilong Dong, Liefeng Bo*
- 395 Learning Deblurring Texture Prior from Unpaired Data with Diffusion Model, *Chengxu Liu, Lu Qi, Jinshan Pan, Xueming Qian, Ming-Hsuan Yang*
- 396 GUAVA: Generalizable Upper Body 3D Gaussian Avatar, *Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Yang Li, Minghan Qin, Yu Li, Haoqian Wang*
- 397 Recognizing Actions from Robotic View for Natural Human-Robot Interaction, *Ziyi Wang, Peiming Li, Hong Liu, Zhichao Deng, Can Wang, Jun Liu, Junsong Yuan, Mengyuan Liu*
- 398 TARO: Timestep-Adaptive Representation Alignment with Onset-Aware Conditioning for Synchronized Video-to-Audio Synthesis, *Tri Ton, Ji Woo Hong, Chang D. Yoo*
- 399 UniFuse: A Unified All-in-One Framework for Multi-Modal Medical Image Fusion Under Diverse Degradations and Misalignments, *Dayong Su, Yafei Zhang, Huafeng Li, Jinxing Li, Yu Liu*
- 400 DexVLG: Dexterous Vision-Language-Grasp Model at Scale, *Jiawei He, Danshi Li, Xinqiang Yu, Zekun Qi, Wenyao Zhang, Jiayi Chen, Zhaoxiang Zhang, Zhizheng Zhang, Li Yi, He Wang*
- 401 Towards Explicit Exoskeleton for the Reconstruction of Complicated 3D Human Avatars, *Yifan Zhan, Qingtian Zhu, Muyao Niu, Mingze Ma, Jiancheng Zhao, Zhihang Zhong, Xiao Sun, Yu Qiao, Yinqiang Zheng*
- 402 AV-Flow: Transforming Text to Audio-Visual Human-like Interactions, *Aggelina Chatziagapi, Louis-Philippe Morency, Hongyu Gong, Michael Zollhöfer, Dimitris Samaras, Alexander Richard*
- 403 Democratizing High-Fidelity Co-Speech Gesture Video Generation, *Xu Yang, Shaoli Huang, Shenbo Xie, Xuelin Chen, Yifei Liu, Changxing Ding*
- 404 Fine-Grained 3D Gaussian Head Avatars Modeling from Static Captures via Joint Reconstruction and Registration, *Yuan Sun, Xuan Wang, Cong Wang, WeiLi Zhang, Yanbo Fan, Yu Guo, Fei Wang*
- 405 Motion-2-to-3: Leveraging 2D Motion Data for 3D Motion Generations, *Ruoxi Guo, Huaijin Pi, Zehong Shen, Qing Shuai, Zechen Hu, Zhumei Wang, Yajiao Dong, Ruizhen Hu, Taku Komura, Sida Peng, Xiaowei Zhou*
- 406 IM-LUT: Interpolation Mixing Look-Up Tables for Image Super-Resolution, *Sejin Park, Sangmin Lee, Kyong Hwan Jin, Seung-Won Jung*
- 407 TemCoCo: Temporally Consistent Multi-modal Video Fusion with Visual-Semantic Collaboration, *Meiqi Gong, Hao Zhang, Xunpeng Yi, Linfeng Tang, Jiayi Ma*
- 408 Beyond the Frame: Generating 360° Panoramic Videos from Perspective Videos, *Rundong Luo, Matthew Wallingford, Ali Fahardi, Noah Snaveley, Wei-Chiu Ma*
- 409 SpinMeRound: Consistent Multi-View Identity Generation Using Diffusion Models, *Stathis Galanakis, Alexandros Lattas, Stylianos Moschoglou, Bernhard Kainz, Stefanos Zafeiriou*
- 410 DIMO: Diverse 3D Motion Generation for Arbitrary Objects, *Linzhan Mou, Jiahui Lei, Chen Wang, Lingjie Liu, Kostas Daniilidis*
- 411 OpenAnimals: Revisiting Person Re-Identification for Animals Towards Better Generalization, *Saihui Hou, Panjian Huang, Zengbin Wang, Yuan Liu, Zeyu Li, Man Zhang, Yongzhen Huang*
- 412 Perception-as-Control: Fine-grained Controllable Image Animation with 3D-aware Motion Representation, *Yingjie Chen, Yifang Men, Yuan Yao, Miaomiao Cui, Liefeng Bo*
- 413 Attention to Trajectory: Trajectory-Aware Open-Vocabulary Tracking, *Yunhao Li, Yifan Jiao, Dan Meng, Heng Fan, Libo Zhang*
- 414 UniPortrait: A Unified Framework for Identity-Preserving Single- and Multi-Human Image Personalization, *Junjie He, Yifeng Geng, Liefeng Bo*
- 415 SAMPLE: Semantic Alignment through Temporal-Adaptive Multimodal Prompt Learning for Event-Based Open-Vocabulary Action Recognition, *Jing Wang, Rui Zhao, Ruiqin Xiong, Xingtao Wang, Xiaopeng Fan, Tiejun Huang*
- 416 Music-Aligned Holistic 3D Dance Generation via Hierarchical Motion Modeling, *Xiaojie Li, Ronghui Li, Shukai Fang, Shuzhao Xie, Xiaoyang Guo, Jiaqing Zhou, Junkun Peng, Zhi Wang*
- 417 Structure Matters: Revisiting Boundary Refinement in Video Object Segmentation, *Guanyi Qin, Ziyue Wang, Daiyun Shen, Haofeng Liu, Hantao Zhou, Junde Wu, Runze Hu, Yueming Jin*
- 418 NoiseController: Towards Consistent Multi-view Video Generation via Noise Decomposition and Collaboration, *Haotian Dong, Xin Wang, Di Lin, Yipeng Wu, Qin Chen, Ruonan Liu, Kairui Yang, Ping Li, Qing Guo*
- 419 FED-PsyAU: Privacy-Preserving Micro-Expression Recognition via Psychological AU Coordination and Dynamic Facial Motion Modeling, *Jingting Li, Yu Qian, Lin Zhao, Su-Jing Wang*
- 420 MagicMirror: ID-Preserved Video Generation in Video Diffusion Transformers, *Yuechen Zhang, Yaoyang Liu, Bin Xia, Bohao Peng, Zexin Yan, Eric Lo, Jiaya Jia*
- 421 MotionDiff: Training-free Zero-shot Interactive Motion Editing via Flow-assisted Multi-view Diffusion, *Yikun Ma, Yiqing Li, Jiawei Wu, Xing Luo, Zhi Jin*
- 422 Dense Policy: Bidirectional Autoregressive Learning of Actions, *Yue Su, Xinyu Zhan, Hongjie Fang, Han Xue, Hao-Shu Fang, Yong-Lu Li, Cewu Lu, Lixin Yang*
- 423 PUMPS: Skeleton-Agnostic Point-based Universal Motion Pre-Training for Synthesis in Human Motion Tasks, *Clinton Ansun Mo, Kun Hu, Chengjiang Long, Dong Yuan, Wan-Chi Siu, Zhiyong Wang*
- 424 Dirichlet-Constrained Variational Codebook Learning for Temporally Coherent Video Face Restoration, *Baoyou Chen, Ce Liu, Weihao Yuan, Zilong Dong, Siyu Zhu*
- 425 Seeing Through Deepfakes: A Human-Inspired Framework for Multi-Face Detection, *Juan Hu, Shaojing Fan, Terence Sim*
- 426 MistSense: Versatile Online Detection of Procedural and Execution Mistakes, *Constantin Patsch, Yuankai Wu, Marsil Zakour, Driton Salihu, Eckehard Steinbach*
- 427 SEREP: Semantic Facial Expression Representation for Robust In-the-Wild Capture and Retargeting, *Arthur Josi, Luiz Gustavo Hafemann, Abdallah Dib, Emeline Got, Rafael M. O. Cruz, Marc-André Carboneau*
- 428 Clink! Chop! Thud! - Learning Object Sounds from Real-World Interactions, *Mengyu Yang, Yiming Chen, Haozheng Pei, Siddhant Agarwal, Arun Balajee Vasudevan, James Hays*
- 429 LUT-Fuse: Towards Extremely Fast Infrared and Visible Image Fusion via Distillation to Learnable Look-Up Tables, *Xunpeng Yi, Yibing Zhang, Xinyu Xiang, Qinglong Yan, Han Xu, Jiayi Ma*
- 430 LiON-LoRA: Rethinking LoRA Fusion to Unify Controllable Spatial and Temporal Generation for Video Diffusion, *Yisu Zhang, Chenjie Cao, Chaohui Yu, Jianke Zhu*
- 431 Morph: A Motion-free Physics Optimization Framework for Human Motion Generation, *Zhuo Li, Mingshuang Luo, Ruibing Hou, Xin Zhao, Hao Liu, Hong Chang, Zimo Liu, Chen Li*
- 432 RoBridge: A Hierarchical Architecture Bridging Cognition and Execution for General Robotic Manipulation, *Kaidong Zhang, Rongtao Xu, Pengzhen Ren, Junfan Lin, Hefeng Wu, Liang Lin, Xiaodan Liang*
- 433 DisCoRD: Discrete Tokens to Continuous Motion via Rectified Flow Decoding, *Junbin Cho, Junwan Kim, Jisoo Kim, Minseo Kim, Mingu Kang, Sungeun Hong, Tae-Hyun Oh, Youngjae Yu*
- 434 MixANT: Observation-dependent Memory Propagation for Stochastic Dense Action Anticipation, *Syed Talal Wasim, Hamid Suleman, Olga Zatsarynna, Muzammal Naseer, Juergen Gall*
- 435 VoiceCraft-Dub: Automated Video Dubbing with Neural Codec Language Models, *Kim Sung-Bin, Jeongsoo Choi, Puyuan Peng, Joon Son Chung, Tae-Hyun Oh, David Harwath*
- 436 DeSPITE: Exploring Contrastive Deep Skeleton-Pointcloud-IMU-Text Embeddings for Advanced Point Cloud Human Activity Understanding, *Thomas Kreutz, Max Mühlhäuser, Alejandro Sanchez Guinea*
- 437 Temporal Rate Reduction Clustering for Human Motion Segmentation, *Xianghan Meng, Zhengyu Tong, Zhiyuan Huang, Chun-Guang Li*
- 438 DISTA-Net: Dynamic Closely-Spaced Infrared Small Target Unmixing, *Shengdong Han, Shangdong Yang, Yuxuan Li, Xin Zhang, Xiang Li, Jian Yang, Ming-Ming Cheng, Yimian Dai*
- 439 Efficient Concertormer for Image Deblurring and Beyond, *Pin-Hung Kuo, Jinshan Pan, Shao-Yi Chien, Ming-Hsuan Yang*

- 440 ASCENT: Annotation-free Self-supervised Contrastive Embeddings for 3D Neuron Tracking in Fluorescence Microscopy, *Haejun Han, Hang Lu*
- 441 InfiniDreamer: Arbitrarily Long Human Motion Generation via Segment Score Distillation, *Wenjie Zhuo, Fan Ma, Hehe Fan*
- 442 FLOAT: Generative Motion Latent Flow Matching for Audio-driven Talking Portrait, *Taekyung Ki, Dongchan Min, Gyeongsu Chae*
- 443 VSRM: A Robust Mamba-Based Framework for Video Super-Resolution, *Dinh Phu Tran, Dao Duy Hung, Daeyoung Kim*
- 444 Face Retouching with Diffusion Data Generation and Spectral Restoration, *Zhidan Xu, Xiaoqin Zhang, Shijian Lu*
- 445 Separation for Better Integration: Disentangling Edge and Motion in Event-based Deblurring, *Yufei Zhu, Hao Chen, Yongjian Deng, Wei You*
- 446 2HandedAfforder: Learning Precise Actionable Bimanual Affordances from Human Videos, *Marvin Heidinger, Snehal Jauhri, Vignesh Prasad, Georgia Chalvatzaki*
- 447 Flow Stochastic Segmentation Networks, *Fabio De Sousa Ribeiro, Omar Todd, Charles Jones, Avinash Kori, Raghav Mehta, Ben Glocker*
- 448 IAP: Invisible Adversarial Patch Attack through Perceptibility-Aware Localization and Perturbation Optimization, *Subrat Kishore Dutta, Xiao Zhang*
- 449 Towards a Unified Copernicus Foundation Model for Earth Vision, *Yi Wang, Zhitong Xiong, Chenying Liu, Adam J. Stewart, Thomas Dujardin, Nikolaos Ioannis Bountos, Angelos Zavras, Franziska Gerken, Ioannis Papoutsis, Laura Leal-Taixé, Xiao Xiang Zhu*
- 450 Teeth Reconstruction and Performance Capture Using a Phone Camera, *Weixi Zheng, Jingwang Ling, Zhibo Wang, Quan Wang, Feng Xu*
- 451 AnimalClue: Recognizing Animals by their Traces, *Risa Shinoda, Nakamasa Inoue, Iro Laina, Christian Rupprecht, Hirokatsu Kataoka*

10:45 - 12:45 Demos 3 (Exhibit Hall I)

- Layered Diffusion Brushes, *Peyman Gholami, Robert Xiao*
- Semantic navigation, *Yue Hu*
- Demo for One-Step Specular Highlight Removal with Adapted Diffusion Models, *Mahir Atmis, Levent Karacan, Mehmet Sarigul*
- Image as an IMU, *Jerred Chen, Ronald Clark*

11:00 - 13:00 Lunch (Exhibit Hall II / Overflow Seating Rooftop and Third Floor)

11:00 - 13:00 Doctoral Consortium (301 A+ B)

13:00 - 14:15 Oral Session 4A: Vision + Graphics (Exhibit Hall III)

Session Chairs: Boxin Shi, Neal Wadhwa

- 🏆 - Award candidate paper ☆ - Outstanding Reviewer
- 1 Dynamic Typography: Bringing Text to Life via Video Diffusion Prior, *Zichen Liu, Yihao Meng, Hao Ouyang, Yue Yu, Bolin Zhao, Daniel Cohen-Or, Huamin Qu*
- 2 Generating Physically Stable and Buildable Brick Structures from Text, *Ava Pun, Kangle Deng, Ruixuan Liu, Deva Ramanan, Changliu Liu, Jun-Yan Zhu*
- 3 WIR3D: Visually-Informed and Geometry-Aware 3D Shape Abstraction, *Richard Liu, Daniel Fu, Noah Tan, Itai Lang, Rana Hanocka*
- 4 SparseFlex: High-Resolution and Arbitrary-Topology 3D Shape Modeling, *Xianglong He, Zi-Xin Zou, Chia-Hao Chen, Yuan-Chen Guo, Ding Liang, Chun Yuan, Wanli Ouyang, Yan-Pei Cao, Yangguang Li*
- 5 ReCamMaster: Camera-Controlled Generative Rendering from A Single Video, *Jianhong Bai, Menghan Xia, Xiao Fu, Xintao Wang, Lianrui Mu, Jinwen Cao, Zuozhu Liu, Haoji Hu, Xiang Bai, Pengfei Wan, Di Zhang*

13:00 - 14:15 Oral Session 4B: 3D Pose Understanding (Kalakaua Ballroom)

Session Chairs: P J Narayanan, Hyung Jin Chang

- 1 Forecasting Continuous Non-Conservative Dynamical Systems in $SO(3)$, *Lennart Bastian, Mohammad Rashed, Nassir Navab, Tolga Birdal*
- 2 Certifiably Optimal Anisotropic Rotation Averaging, *Carl Olsson, Yaroslava Lochman, Johan Malmport, Christopher Zach*

- 3 Deterministic Object Pose Confidence Region Estimation, *Jinghao Wang, Zhang Li, Zi Wang, Banglei Guan, Yang Shang, Qifeng Yu*
- 4 RePosED: Efficient Relative Pose Estimation With Known Depth Information, *Yaqing Ding, Viktor Kocur, Václav Vávra, Zuzana Berger Haladová, Jian Yang, Torsten Sattler, Zuzana Kukelova*
- 5 Diving into the Fusion of Monocular Priors for Generalized Stereo Matching, *Chengtang Yao, Lidong Yu, Zhidan Liu, Jiaxi Zeng, Yuwei Wu, Yunde Jia*

14:00 - 14:30 Poster Setup (Exhibit Hall I)

14:30 - 16:30 Poster Session 4 & Exhibit Hall with Coffee Break (Exhibit Hall I)

- * - Highlight paper 🏆 - Award candidate paper
 ○ - Oral Paper ☆ - Outstanding Reviewer
- 1 TIP-I2V: A Million-Scale Real Text and Image Prompt Dataset for Image-to-Video Generation, *Wenhao Wang, Yi Yang*
- 2 IntroStyle: Training-Free Introspective Style Attribution using Diffusion Features, *Anand Kumar, Jiteng Mu, Nuno Vasconcelos*
- 3 SCFlow: Implicitly Learning Style and Content Disentanglement with Flow Models, *Pingchuan Ma, Xiaopei Yang, Yusong Li, Ming Gui, Felix Krause, Johannes Schusterbauer, Björn Ommer*
- 4 Compression-Aware One-Step Diffusion Model for JPEG Artifact Removal, *Jinpei Guo, Zheng Chen, Wenbo Li, Yong Guo, Yulun Zhang*
- 5 OminiControl: Minimal and Universal Control for Diffusion Transformer, *Zhenxiong Tan, Songhua Liu, Xingyi Yang, Qiaochu Xue, Xinchao Wang*
- 6 Phantom: Subject-Consistent Video Generation via Cross-Modal Alignment, *Lijie Liu, Tianxiang Ma, Bingchuan Li, Zhuowei Chen, Jiawei Liu, Gen Li, Siyu Zhou, Qian He, Xinglong Wu*
- 7 Penalizing Boundary Activation for Object Completeness in Diffusion Models, *Haoyang Xu, Tianhao Zhao, Sibe Yang, Yutian Lin*
- 8 MatchDiffusion: Training-free Generation of Match-Cuts, *Alejandro Pardo, Fabio Pizzati, Tong Zhang, Alexander Pondaven, Philip Torr, Juan Camilo Perez, Bernard Ghanem*
- 9 Dual-Expert Consistency Model for Efficient and High-Quality Video Generation, *Zhengyao Lv, Chenyang Si, Tianlin Pan, Zhaoxi Chen, Kwan-Yee K. Wong, Yu Qiao, Ziwei Liu*
- 10 RealGeneral: Unifying Visual Generation via Temporal In-Context Learning with Video Models, *Yijing Lin, Mengqi Huang, Shuhan Zhuang, Zhendong Mao*
- 11 Straighten Viscous Rectified Flow via Noise Optimization, *Jimin Dai, Jiexi Yan, Jian Yang, Lei Luo*
- 12 Trans-Adapter: A Plug-and-Play Framework for Transparent Image Inpainting, *Yuekun Dai, Haitian Li, Shangchen Zhou, Chen Change Loy*
- 13 Scalable Dual Fingerprinting for Hierarchical Attribution of Text-to-Image Models, *Jianwei Fei, Yunshu Dai, Peipeng Yu, Zhe Kong, Jiantao Zhou, Zhihua Xia*
- 14 QuantCache: Adaptive Importance-Guided Quantization with Hierarchical Latent and Layer Caching for Video Generation, *Junyi Wu, Zhiteng Li, Zheng Hui, Yulun Zhang, Linghe Kong, Xiaokang Yang*
- 15 CRAM: Large Scale Video Continual Learning with Bootstrapped Compression, *Shivani Mall, Joao F. Henriques*
- 16 Learning Robust Image Watermarking with Lossless Cover Recovery, *Jiale Chen, Wei Wang, Chongyang Shi, Li Dong, Xiping Hu*
- 17 Cross-Subject Mind Decoding from Inaccurate Representations, *Yangyang Xu, Bangzhen Liu, Wenqi Shao, Yong Du, Shengfeng He, Tingting Zhu*
- 18 Tree-NeRV: Efficient Non-Uniform Sampling for Neural Video Representation via Tree-Structured Feature Grids, *Jiancheng Zhao, Yifan Zhan, Qingtian Zhu, Mingze Ma, Muyao Niu, Zunian Wan, Xiang Ji, Yinqiang Zheng*
- 19 HPSv3: Towards Wide-Spectrum Human Preference Score, *Yuhang Ma, Xiaoshi Wu, Keqiang Sun, Hongsheng Li*
- 20 SummDiff: Generative Modeling of Video Summarization with Diffusion, *Kwanseok Kim, Jaehoon Hahm, Sumin Kim, Jinhwan Sul, Byungchak Kim, Joonseok Lee*
- 21 Leveraging Panoptic Scene Graph for Evaluating Fine-Grained Text-to-Image Generation, *Xueqing Deng, Linjie Yang, Qihang Yu, Chenglin Yang, Liang-Chieh Chen*

- 22 MaTe: Images Are All You Need for Material Transfer via Diffusion Transformer, *Nisha Huang, Henglin Liu, Yizhou Lin, Kaer Huang, Chubin Chen, Jie Guo, Tong-ye Lee, Xiu Li*
- 23 LEGO-Maker: A Semantic-Driven Algorithm for Text-to-3D Generation, *Yifei Zhang, Lei Chen*
- 24 ForCenNet: Foreground-Centric Network for Document Image Rectification, *Peng Cai, Qiang Li, Kaicheng Yang, Dong Guo, Jia Li, Nan Zhou, Xiang An, Ninghua Yang, Jiankang Deng*
- 25 VEGGIE: Instructional Editing and Reasoning Video Concepts with Grounded Generation, *Shoubin Yu, Difan Liu, Ziqiao Ma, Yicong Hong, Yang Zhou, Hao Tan, Joyce Chai, Mohit Bansal*
- 26 Textured 3D Regenerative Morphing with 3D Diffusion Prior, *Songlin Yang, Yushi Lan, Honghua Chen, Xingang Pan*
- 27 Scale Your Instructions: Enhance the Instruction-Following Fidelity of Unified Image Generation Model by Self-Adaptive Attention Scaling, *Chao Zhou, Tianyi Wei, Nenghai Yu*
- 28 Efficient Input-level Backdoor Defense on Text-to-Image
- * Synthesis via Neuron Activation Variation, *Shengfang Zhai, Jiajun Li, Yue Liu, Huanran Chen, Zhihua Tian, Wenjie Qu, Qingni Shen, Ruoxi Jia, Yinpeng Dong, Jiaheng Zhang*
- 29 CycleVAR: Repurposing Autoregressive Model for Unsupervised One-Step Image Translation, *Yi Liu, Shengqian Li, Zuzeng Lin, Feng Wang, Si Liu*
- 30 MagicColor: Multi-Instance Sketch Colorization, *Yinhan Zhang, Yue Ma, Bingyuan Wang, Qifeng Chen, Zeyu Wang*
- 31 EmotiCrafter: Text-to-Emotional-Image Generation based on Valence-Arousal Model, *Shengqi Dang, Yi He, Long Ling, Ziqing Qian, Nanxuan Zhao, Nan Cao*
- 32 SDMatte: Grafting Diffusion Models for Interactive Matting, *Longfei Huang, Yu Liang, Hao Zhang, Jinwei Chen, Wei Dong, Lunde Chen, Wanyu Liu, Bo Li, Peng-Tao Jiang*
- 33 Adaptive Caching for Faster Video Generation with Diffusion Transformers, *Kumara Kahatapitiya, Haozhe Liu, Sen He, Ding Liu, Menglin Jia, Chenyang Zhang, Michael S. Ryoo, Tian Xie*
- 34 CoMPaSS: Enhancing Spatial Understanding in Text-to-Image Diffusion Models, *Gaoyang Zhang, Bingtao Fu, Qingnan Fan, Qi Zhang, Runxing Liu, Hong Gu, Huaqi Zhang, Xinguo Liu*
- 35 Pinco: Position-induced Consistent Adapter for Diffusion Transformer in Foreground-conditioned Inpainting, *Guangben Lu, Yuzhen Du, Yizhe Tang, Zhimin Sun, Ran Yi, Yifan Qi, Tianyi Wang, Lizhuang Ma, Fangyuan Zou*
- 36 Edicho: Consistent Image Editing in the Wild, *Qingyan Bai, Hao Ouyang, Yinghao Xu, Qiuyu Wang, Ceyuan Yang, Ka Leong Cheng, Yujun Shen, Qifeng Chen*
- 37 Photolithography Overlay Map Generation with Implicit Knowledge Distillation Diffusion Transformer, *Yuan-Fu Yang, Hsiu-Hui Hsiao*
- 38 LUSD: Localized Update Score Distillation for Text-Guided Image Editing, *Worameth Chinchuthakun, Tossaporn Saengja, Nontawat Tritrong, Pitchaporn Rewatbowornwong, Pramook Khungurn, Supasorn Suwajanakorn*
- 39 FlowChef: Steering of Rectified Flow Models for Controlled Generations, *Maitreya Patel, Song Wen, Dimitris N. Metaxas, Yezhou Yang*
- 40 DiffVSR: Revealing an Effective Recipe for Taming Robust Video Super-Resolution Against Complex Degradations, *Xiaohui Li, Yihao Liu, Shuo Cao, Ziyang Chen, Shaobin Zhuang, Xiangyu Chen, Yinan He, Yi Wang, Yu Qiao*
- 41 From Reflection to Perfection: Scaling Inference-Time Optimization for Text-to-Image Diffusion Models via Reflection Tuning, *Le Zhuo, Liangbing Zhao, Sayak Paul, Yue Liao, Renrui Zhang, Yi Xin, Peng Gao, Mohamed Elhoseiny, Hongsheng Li*
- 42 PanoLlama: Generating Endless and Coherent Panoramas with Next-Token-Prediction LLMs, *Teng Zhou, Xiaoyu Zhang, Yongchuan Tang*
- 43 Visual Interestingness Decoded: How GPT-4o Mirrors Human Interests, *Fitim Abdullahu, Helmut Grabner*
- 44 Translation of Text Embedding via Delta Vector to Suppress Strongly Entangled Content in Text-to-Image Diffusion Models, *Eunseo Koh, Seunghoo Hong, Tae-Young Kim, Simon S. Woo, Jae-Pil Heo*
- 45 Grouped Speculative Decoding for Autoregressive Image Generation, *Junhyuk So, Juncheol Shin, Hyunho Kook, Eunhyeok Park*
- 46 Generating, Fast and Slow: Scalable Parallel Video Generation with Video Interface Networks, *Bhishma Dedhia, David Bourgin, Krishna Kumar Singh, Yuheng Li, Yan Kang, Zhan Xu, Niraj K. Jha, Yuchen Liu*
- 47 DuoLoRA : Cycle-consistent and Rank-disentangled Content-Style Personalization, *Aniket Roy, Shubhankar Borse, Shreya Kadambi, Debasmit Das, Shweta Mahajan, Risheek Garrepalli, Hyojin Park, Ankita Nayak, Rama Chellappa, Munawar Hayat, Fatih Porikli*
- 48 Progressive Artwork Outpainting via Latent Diffusion Models, *Dae-Young Song, Jung-Jae Yu, Donghyeon Cho*
- 49 SynTag: Enhancing the Geometric Robustness of Inversion-based Generative Image Watermarking, *Han Fang, Kejiang Chen, Zehua Ma, Jiajun Deng, Yicong Li, Weiming Zhang, Ee-Chien Chang*
- 50 Text Embedding Knows How to Quantize Text-Guided Diffusion Models, *Hongjae Lee, Myungjun Son, Dongjae Kang, Seung-Won Jung*
- 51 Scene Graph Guided Generation: Enable Accurate Relations Generation in Text-to-Image Models via Textual Rectification, *Guibao Shen, Luozhou Wang, Jiantao Lin, Wenhang Ge, Chaozhe Zhang, Xin Tao, Di Zhang, Pengfei Wan, Guangyong Chen, Yijun Li, Ying-cong Chen*
- 52 PUMA: Empowering Unified MLLM with Multi-granular Visual Generation, *Rongyao Fang, Chengqi Duan, Kun Wang, Hao Li, Linjiang Huang, Hao Tian, Xingyu Zeng, Rui Zhao, Jifeng Dai, Hongsheng Li, Xihui Liu*
- 53 NeuralSVG: An Implicit Representation for Text-to-Vector Generation, *Sagi Polaczek, Yuval Alaluf, Elad Richardson, Yael Vinker, Daniel Cohen-Or*
- 54 IQA-Adapter: Exploring Knowledge Transfer from Image Quality Assessment to Diffusion-based Generative Models, *Khaled Abud, Sergey Lavrushkin, Alexey Kirillov, Dmitriy Vatolin*
- 55 DICE: Staleness-Centric Optimizations for Parallel Diffusion MoE Inference, *Jiajun Luo, Lizhuo Luo, Jianru Xu, Jiajun Song, Rongwei Lu, Chen Tang, Zhi Wang*
- 56 Dual Recursive Feedback on Generation and Appearance Latents for Pose-Robust Text-to-Image Diffusion, *Jiwon Kim, Pureum Kim, SeonHwa Kim, Soobin Park, Eunju Cha, Kyong Hwan Jin*
- 57 Anti-Tamper Protection for Unauthorized Individual Image Generation, *Zelin Li, Ruohan Zong, Yifan Liu, Ruichen Yao, Yaokun Liu, Yang Zhang, Dong Wang*
- 58 Continual Personalization for Diffusion Models, *Yu-Chien Liao, Jr-Jen Chen, Chi-Pin Huang, Ci-Siang Lin, Meng-Lin Wu, Yu-Chiang Frank Wang*
- 59 FreeCus: Free Lunch Subject-driven Customization in Diffusion Transformers, *Yanbing Zhang, Zhe Wang, Qin Zhou, Mengping Yang*
- 60 WikiAutoGen: Towards Multi-Modal Wikipedia-Style Article Generation, *Zhongyu Yang, Jun Chen, Dannong Xu, Junjie Fei, Xiaoqian Shen, Liangbing Zhao, Chun-Mei Feng, Mohamed Elhoseiny*
- 61 QuEST: Low-bit Diffusion Model Quantization via Efficient Selective Finetuning, *Haoyuan Wang, Yuzhang Shang, Zhihang Yuan, Junyi Wu, Junchi Yan, Yan Yan*
- 62 LazyMAR: Accelerating Masked Autoregressive Models via Feature Caching, *Feihong Yan, Qingyan Wei, Jiayi Tang, Jiajun Li, Yulin Wang, Xuming Hu, Huiqi Li, Linfeng Zhang*
- 63 SPADE: Spatial-Aware Denoising Network for Open-vocabulary Panoptic Scene Graph Generation with Long- and Local-range Context Reasoning, *Xin Hu, Ke Qin, Guiduo Duan, Ming Li, Yuan-Fang Li, Tao He*
- 64 An Information-Theoretic Regularizer for Lossy Neural Image Compression, *Yingwen Zhang, Meng Wang, Xihua Sheng, Peilin Chen, Junru Li, Li Zhang, Shiqi Wang*
- 65 DropletVideo: A Dataset and Approach to Explore Integral Spatio-Temporal Consistent Video Generation, *Runze Zhang, Guoguang Du, Xiaochuan Li, Qi Jia, Liang Jin, Lu Liu, Jingjing Wang, Cong Xu, Zhenhua Guo, Yaqian Zhao, Xiaoli Gong, Rengang Li, Baoyu Fan*
- 66 Timestep-Aware Diffusion Model for Extreme Image Rescaling, *Ce Wang, Zhenyu Hu, Wanjie Sun, Zhenzhong Chen*
- 67 The Best of Both Worlds: Integrating Language Models and Diffusion Models for Video Generation, *Aoxiong Yin, Xu Tan, Kai Shen, Yichong Leng, Xinyu Zhou, Juncheng Li, Siliang Tang*

- 68 Split-and-Combine: Enhancing Style Augmentation for Single Domain Generalization, *Zhen Zhang, Shuai Yang, Qianlong Dang, Zhize Wu, Lichuan Gu*
- 69 Att-Adapter: A Robust and Precise Domain-Specific Multi-Attributes T2I Diffusion Adapter via Conditional Variational Autoencoder, *Wonwoong Cho, Yan-Ying Chen, Matthew Klenk, David I. Inouye, Yanxia Zhang*
- 70 VPO: Aligning Text-to-Video Generation Models with Prompt Optimization, *Jiale Cheng, Ruiliang Lyu, Xiaotao Gu, Xiao Liu, Jiazheng Xu, Yida Lu, Jiayan Teng, Zhuoyi Yang, Yuxiao Dong, Jie Tang, Hongning Wang, Minlie Huang*
- 71 ReEdit: A Benchmark and Method for Improving Instruction-based Image Editing Model on Referring Expressions, *Bimsara Pathiraja, Maitreya Patel, Shivam Singh, Yezhou Yang, Chitta Baral*
- 72 Reflect-DiT: Inference-Time Scaling for Text-to-Image Diffusion Transformers via In-Context Reflection, *Shufan Li, Konstantinos Kallidromitis, Akash Gokul, Arsh Koneru, Yusuke Kato, Kazuki Kozuka, Aditya Grover*
- 73 TryOn-Refiner: Conditional Rectified-flow-based TryOn Refiner for More Accurate Detail Reconstruction, *Wen Qian*
- 74 ReCamMaster: Camera-Controlled Generative Rendering from A Single Video, *Jianhong Bai, Menghan Xia, Xiao Fu, Xintao Wang, Lianrui Mu, Jinwen Cao, Zuozhu Liu, Haoji Hu, Xiang Bai, Pengfei Wan, Di Zhang*
- 75 SparseFlex: High-Resolution and Arbitrary-Topology 3D Shape Modeling, *Xianglong He, Zi-Xin Zou, Chia-Hao Chen, Yuan-Chen Guo, Ding Liang, Chun Yuan, Wanli Ouyang, Yan-Pei Cao, Yangguang Li*
- 76 CuRe: Cultural Gaps in the Long Tail of Text-to-Image Systems, *Aniket Rege, Zinnia Nie, Mahesh Ramesh, Unmesh Raskar, Zhuoran Yu, Aditya Kusupati, Yong Jae Lee, Ramya Korlakai Vinayak*
- 77 LeanVAE: An Ultra-Efficient Reconstruction VAE for Video Diffusion Models, *Yu Cheng, Fajie Yuan*
- 78 TREAD: Token Routing for Efficient Architecture-agnostic Diffusion Training, *Felix Krause, Timy Phan, Ming Gui, Stefan Andreas Baumann, Vincent Tao Hu, Björn Ommer*
- 79 Bootstrap3D: Improving Multi-view Diffusion Model with Synthetic Data, *Zeyi Sun, Tong Wu, Pan Zhang, Yuhang Zang, Xiaoyi Dong, Yuanjun Xiong, Dahua Lin, Jiaqi Wang*
- 80 Cassic: Towards Content-Adaptive State-Space Models for Learned Image Compression, *Shiyu Qin, Jinpeng Wang, Yimin Zhou, Bin Chen, Tianci Luo, Baoyi An, Tao Dai, Shu-Tao Xia, Yaowei Wang*
- 81 FullDiT: Video Generative Foundation Models with Multimodal Control via Full Attention, *Xuan Ju, Weicai Ye, Qande Liu, Qiulin Wang, Xintao Wang, Pengfei Wan, Di Zhang, Kun Gai, Qiang Xu*
- 82 Zero-Shot Depth Aware Image Editing with Diffusion Models, *Rishabh Parihar, Sachidanand VS, R. Venkatesh Babu*
- 83 StyleKeeper: Prevent Content Leakage using Negative Visual Query Guidance, *Jaeseok Jeong, Junho Kim, Gayoung Lee, Yunjey Choi, Youngjung Uh*
- 84 Global and Local Entailment Learning for Natural World Imagery, *Srikumar Sastry, Aayush Dhakal, Eric Xing, Subash Khanal, Nathan Jacobs*
- 85 Beyond Next-Token: Next-X Prediction for Autoregressive Visual Generation, *Sucheng Ren, Qihang Yu, Ju He, Xiaohui Shen, Alan Yuille, Liang-Chieh Chen*
- 86 Multi-turn Consistent Image Editing, *Zijun Zhou, Yingying Deng, Xiangyu He, Weiming Dong, Fan Tang*
- 87 Domain Generalizable Portrait Style Transfer, *Xinbo Wang, Wenju Xu, Qing Zhang, Wei-Shi Zheng*
- 88 TRKT: Weakly Supervised Dynamic Scene Graph Generation with Temporal-enhanced Relation-aware Knowledge Transferring, *Zhu Xu, Ting Lei, Zhimin Li, Guan Wang, Qingchao Chen, Yuxin Peng, Yang Liu*
- 89 Pose-Star: Anatomy-Aware Editing for Open-World Fashion Images, *Yuran Dong, Mang Ye*
- 90 Who Controls the Authorization? Invertible Networks for Copyright Protection in Text-to-Image Synthesis, *Baoyue Hu, Yang Wei, Junhao Xiao, Wendong Huang, Xiuli Bi, Bin Xiao*
- 91 SpectralAR: Spectral Autoregressive Visual Generation, *Yuanhui Huang, Weiliang Chen, Wenzhao Zheng, Yueqi Duan, Jie Zhou, Jiwen Lu*
- 92 From Reusing to Forecasting: Accelerating Diffusion Models with TaylorSeers, *Jiacheng Liu, Chang Zou, Yuanhuiyi Lyu, Junjie Chen, Linfeng Zhang*
- 93 SegmentDreamer: Towards High-fidelity Text-to-3D Synthesis with Segmented Consistency Trajectory Distillation, *Jiahao Zhu, Zixuan Chen, Guangcong Wang, Xiaohua Xie, Yi Zhou*
- 94 The Curse of Conditions: Analyzing and Improving Optimal Transport for Conditional Flow-Based Generation, *Ho Kei Cheng, Alexander Schwing*
- 95 MUSE: Multi-Subject Unified Synthesis via Explicit Layout Semantic Expansion, *Fei Peng, Junqiang Wu, Yan Li, Tingting Gao, Di Zhang, Huiyuan Fu*
- 96 Rethink Sparse Signals for Pose-guided Text-to-image Generation, *Wenjie Xuan, Jing Zhang, Juhua Liu, Bo Du, Dacheng Tao*
- 97 Steering Guidance for Personalized Text-to-Image Diffusion Models, *Sunghyun Park, Seokeon Choi, Hyoungwoo Park, Sunguck Yun*
- 98 CAFA: a Controllable Automatic Foley Artist, *Roi Benita, Michael Finkelson, Tavi Halperin, Gleb Sterkin, Yossi Adi*
- 99 M2SFormer: Multi-Spectral and Multi-Scale Attention with Edge-Aware Difficulty Guidance for Image Forgery Localization, *Ju-Hyeon Nam, Dong-Hyun Moon, Sang-Chul Lee*
- 100 ReFlex: Text-Guided Editing of Real Images in Rectified Flow via Mid-Step Feature Extraction and Attention Adaptation, *Jimyeong Kim, Jungwon Park, Yeji Song, Nojun Kwak, Wonjong Rhee*
- 101 IFAdapter: Instance Feature Control for Grounded Text-to-Image Generation, *Yinwei Wu, Xianpan Zhou, Bing Ma, Xuefeng Su, Kai Ma, Xinchao Wang*
- 102 EditCLIP: Representation Learning for Image Editing, *Qian Wang, Aleksandar Cvejić, Abdelrahman EldeSokey, Peter Wonka*
- 103 Magic Insert: Style-Aware Drag-and-Drop, *Nataniel Ruiz, Yuanzhen Li, Neal Wadhwa, Yael Pritch, Michael Rubinstein, David E. Jacobs, Shlomi Fruchter*
- 104 Training-free and Adaptive Sparse Attention for Efficient Long Video Generation, *Yifei Xia, Suhan Ling, Fangcheng Fu, Yujie Wang, Huixia Li, Xuefeng Xiao, Bin Cui*
- 105 SliderSpace: Decomposing the Visual Capabilities of Diffusion Models, *Rohit Gandikota, Zongze Wu, Richard Zhang, David Bau, Eli Shechtman, Nick Kolkin*
- 106 DIVE: Taming DINO for Subject-Driven Video Editing, *Yi Huang, Wei Xiong, He Zhang, Chaoqi Chen, Jianzhuang Liu, Mingfu Yan, Shifeng Chen*
- 107 FontAnimate: High Quality Few-shot Font Generation via Animating Font Transfer Process, *Bin Fu, Zixuan Wang, Kainan Yan, Shitian Zhao, Qi Qin, Jie Wen, Junjun He, Peng Gao*
- 108 PromptDresser: Improving the Quality and Controllability of Virtual Try-On via Generative Textual Prompt and Prompt-aware Mask, *Jeongho Kim, Hoieyoung Jin, Sunghyun Park, Jaegul Choo*
- 109 Scalable Image Tokenization with Index Backpropagation Quantization, *Fengyuan Shi, Zhuoyan Luo, Yixiao Ge, Yujiu Yang, Ying Shan, Limin Wang*
- 110 Enhancing Image Restoration Transformer via Adaptive Translation Equivariance, *JiaKui Hu, Zhengjian Yao, Lujia Jin, Hangzhou He, Yanye Lu*
- 111 CharaConsist: Fine-Grained Consistent Character Generation, *Mengyu Wang, Henghui Ding, Jianing Peng, Yao Zhao, Yunpeng Chen, Yunchao Wei*
- 112 LiT: Delving into a Simple Linear Diffusion Transformer for Image Generation, *Jiahao Wang, Ning Kang, Lewei Yao, Mengzhao Chen, Chengyue Wu, Songyang Zhang, Shuchen Xue, Yong Liu, Taiqiang Wu, Xihui Liu, Kaipeng Zhang, Shifeng Zhang, Wenqi Shao, Zhenguo Li, Ping Luo*
- 113 IMG: Calibrating Diffusion Models via Implicit Multimodal Guidance, *Jiayi Guo, Chuanhao Yan, Xingqian Xu, Yulin Wang, Kai Wang, Gao Huang, Humphrey Shi*
- 114 SAGI: Semantically Aligned and Uncertainty Guided AI Image Inpainting, *Paschalis Giakoumoglou, Dimitrios Karageorgiou, Symeon Papadopoulos, Panagiotis C. Petrantonis*
- 115 Lay2Story: Extending Diffusion Transformers for Layout-Toggable Story Generation, *Ao Ma, Jiasong Feng, Ke Cao, Jing Wang, Yun Wang, Quanwei Zhang, Zhanjie Zhang*

- 116 TextMaster: A Unified Framework for Realistic Text Editing via Glyph-Style Dual-Control, *Zhenyu Yan, Jian Wang, Aoqiang Wang, Yuhan Li, Wenxiang Shang, Ran Lin, Zhu Hangcheng*
- 117 Guiding Diffusion Models with Adaptive Negative Sampling Without External Resources, *Alakh Desai, Nuno Vasconcelos*
- 118 LoRA.rar: Learning to Merge LoRAs via Hypernetworks for Subject-Style Conditioned Image Generation, *Donald Shenaj, Ondrej Bohdal, Mete Ozay, Pietro Zanuttigh, Umberto Michieli*
- 119 Beyond Perspective: Neural 360-Degree Video Compression, *Andy Regensky, Marc Windsheimer, Fabian Brand, Andre Kaup*
- 120 MCID: Multi-aspect Copyright Infringement Detection for Generated Images, *Chuanwei Huang, Zexi Jia, Hongyan Fei, Yeshuang Zhu, Zhiqiang Yuan, Ying Deng, Jiawei Zhang, Xiaoyue Duan, Jinchao Zhang, Jie Zhou*
- 121 Text2Outfit: Controllable Outfit Generation with Multimodal Language Models, *Yuanhao Zhai, Yen-Liang Lin, Minxu Peng, Larry S. Davis, Ashwin Chandramouli, Junsong Yuan, David Doermann*
- 122 Outlier-Aware Post-Training Quantization for Image Super-Resolution, *Hailing Wang, Jianglin Lu, Yitian Zhang, Yun Fu*
- 123 SANA-Sprint: One-Step Diffusion with Continuous-Time Consistency Distillation, *Junsong Chen, Shuchen Xue, Yuyang Zhao, Jincheng Yu, Sayak Paul, Junyu Chen, Han Cai, Song Han, Enze Xie*
- 124 SEAL: Semantic Aware Image Watermarking, *Kasra Arabi, R. Teal Witter, Chinmay Hegde, Niv Cohen*
- 125 Modeling Human Gaze Behavior with Diffusion Models for Unified Scanpath Prediction, *Giuseppe Cartella, Vittorio Cuculo, Alessandro D'Amelio, Marcella Cornia, Giuseppe Boccignone, Rita Cucchiara*
- 126 What Changed? Detecting and Evaluating Instruction-Guided Image Edits with Multimodal Large Language Models, *Lorenzo Baraldi, Davide Bucciarelli, Federico Betti, Marcella Cornia, Lorenzo Baraldi, Nicu Sebe, Rita Cucchiara*
- 127 MeshPad: Interactive Sketch-Conditioned Artist-Reminiscent Mesh Generation and Editing, *Haoxuan Li, Ziya Erkoç, Lei Li, Daniele Sirigatti, Vladislav Rosov, Angela Dai, Matthias Nießner*
- 128 PLADIS: Pushing the Limits of Attention in Diffusion Models at Inference Time by Leveraging Sparsity, *Kwanyoung Kim, Byeongsu Sim*
- 129 STIV: Scalable Text and Image Conditioned Video Generation, *Zongyu Lin, Wei Liu, Chen Chen, Jiasen Lu, Wenzhe Hu, Tsu-Jui Fu, Jesse Allardice, Zhengfeng Lai, Liangchen Song, Bowen Zhang, Cha Chen, Yiran Fei, Lezhi Li, Yinfei Yang, Yizhou Sun, Kai-Wei Chang*
- 130 TLB-VFI: Temporal-Aware Latent Brownian Bridge Diffusion for Video Frame Interpolation, *Zonglin Lyu, Chen Chen*
- 131 ForgeLens: Data-Efficient Forgery Focus for Generalizable Forgery Image Detection, *Yingjian Chen, Lei Zhang, Yakun Niu*
- 132 ObjectMate: A Recurrence Prior for Object Insertion and Subject-Driven Generation, *Daniel Winter, Asaf Shul, Matan Cohen, Dana Berman, Yael Pritch, Alex Rav-Acha, Yedid Hoshen*
- 133 D3QE: Learning Discrete Distribution Discrepancy-aware Quantization Error for Autoregressive-Generated Image Detection, *Yanran Zhang, Bingyao Yu, Yu Zheng, Wenzhao Zheng, Yueqi Duan, Lei Chen, Jie Zhou, Jiwen Lu*
- 134 OmniCache: A Trajectory-Oriented Global Perspective on Training-Free Cache Reuse for Diffusion Transformer Models, *Huanpeng Chu, Wei Wu, Guanyu Feng, Yutao Zhang*
- 135 One-Step Specular Highlight Removal with Adapted Diffusion Models, *Mahir Atmis, Levent Karacan, Mehmet Sarıgöl*
- 136 ART: Adaptive Relation Tuning for Generalized Relation Prediction, *Gopika Sudhakaran, Hikaru Shindo, Patrick Schramowski, Simone Schaub-Meyer, Kristian Kersting, Stefan Roth*
- 137 Transformed Low-rank Adaptation via Tensor Decomposition and Its Applications to Text-to-image Models, *Zerui Tao, Yuhta Takida, Naoki Murata, Qibin Zhao, Yuki Mitsufuji*
- 138 DiGA3D: Coarse-to-Fine Diffusional Propagation of Geometry and Appearance for Versatile 3D Inpainting, *Jingyi Pan, Dan Xu, Qiong Luo*
- 139 Holistic Unlearning Benchmark: A Multi-Faceted Evaluation for Text-to-Image Diffusion Model Unlearning, *Saemi Moon, Minjong Lee, Sangdon Park, Dongwoo Kim*
- 140 Make Me Happier: Evoking Emotions Through Image Diffusion Models, *Qing Lin, Jingfeng Zhang, Yew-Soon Ong, Mengmi Zhang*
- 141 MV-Adapter: Multi-View Consistent Image Generation Made Easy, *Zehuan Huang, Yuan-Chen Guo, Haoran Wang, Ran Yi, Lizhuang Ma, Yan-Pei Cao, Lu Sheng*
- 142 On Large Multimodal Models as Open-World Image Classifiers, *Alessandro Conti, Massimiliano Mancini, Enrico Fini, Yiming Wang, Paolo Rota, Elisa Ricci*
- 143 DiTFastAttnV2: Head-wise Attention Compression for Multi-Modality Diffusion Transformers, *Hanling Zhang, Rundong Su, Zhihang Yuan, Pengtao Chen, Mingzhu Shen, Yibo Fan, Shengen Yan, Guohao Dai, Yu Wang*
- 144 DLFR-Gen: Diffusion-based Video Generation with Dynamic Latent Frame Rate, *Zhihang Yuan, Rui Xie, Yuzhang Shang, Hanling Zhang, Siyuan Wang, Shengen Yan, Guohao Dai, Yu Wang*
- 145 Erasing More Than Intended? How Concept Erasure Degrades the Generation of Non-Target Concepts, *Ibtihel Amara, Ahmed Imtiaz Humayun, Ivana Kajić, Zarana Parekh, Natalie Harris, Sarah Young, Chirag Nagpal, Najoung Kim, Junfeng He, Cristina Nader Vasconcelos, Deepak Ramachandran, Golnoosh Farnadi, Katherine Heller, Mohammad Havaei, Negar Rostamzadeh*
- 146 DIMCIM: A Quantitative Evaluation Framework for Default-mode Diversity and Generalization in Text-to-Image Generative Models, *Revant Teotia, Candace Ross, Karen Ullrich, Sumit Chopra, Adriana Romero-Soriano, Melissa Hall, Matthew Muckley*
- 147 From Linearity to Non-Linearity: How Masked Autoencoders Capture Spatial Correlations, *Anthony Bisulco, Rahul Ramesh, Randall Balestriero, Pratik Chaudhari*
- 148 Addressing Text Embedding Leakage in Diffusion-based Image Editing, *Sunung Mun, Jinhwan Nam, Sunghyun Cho, Jungseul Ok*
- 149 JailbreakDiffBench: A Comprehensive Benchmark for Jailbreaking Diffusion Models, *Xiaolong Jin, Zixuan Weng, Hanxi Guo, Chenlong Yin, Siyuan Cheng, Guangyu Shen, Xiangyu Zhang*
- 150 GReg: Geometry-Aware Region Refinement for Sign Language Video Generation, *Tongkai Shi, Lianyu Hu, Fanhua Shang, Liqing Gao, Wei Feng*
- 151 RePoseD: Efficient Relative Pose Estimation With Known Depth Information, *Yaqing Ding, Viktor Kocur, Václav Vávra, Zuzana Berger Haladová, Jian Yang, Torsten Sattler, Zuzana Kukelova*
- 152 Diving into the Fusion of Monocular Priors for Generalized Stereo Matching, *Chengtang Yao, Lidong Yu, Zhidan Liu, Jiaxi Zeng, Yuwei Wu, Yunde Jia*
- 153 Reusing Computation in Text-to-Image Diffusion for Efficient Generation of Image Sets, *Dale Decatur, Thibault Groueix, Wang Yifan, Rana Hanocka, Vladimir Kim, Matheus Gadelha*
- 154 Repurposing 2D Diffusion Models with Gaussian Atlas for 3D Generation, *Tiange Xiang, Kai Li, Chengjiang Long, Christian Häne, Peihong Guo, Scott Delp, Ehsan Adeli, Li Fei-Fei*
- 155 Shot-by-Shot: Film-Grammar-Aware Training-Free Audio Description Generation, *Junyu Xie, Tengda Han, Max Bain, Arsha Nagrani, Eshika Khandelwal, Gül Varol, Weidi Xie, Andrew Zisserman*
- 156 Cross-Granularity Online Optimization with Masked Compensated Information for Learned Image Compression, *Haowei Kuang, Wenhan Yang, Zongming Guo, Jiaying Liu*
- 157 Generating Multi-Image Synthetic Data for Text-to-Image Customization, *Nupur Kumari, Xi Yin, Jun-Yan Zhu, Ishan Misra, Samaneh Azadi*
- 158 Deeply Supervised Flow-Based Generative Models, *Inkyu Shin, Chenglin Yang, Liang-Chieh Chen*
- 159 Stroke2Sketch: Harnessing Stroke Attributes for Training-Free Sketch Generation, *Rui Yang, Huining Li, Yiyi Long, Xiaojun Wu, Shengfeng He*
- 160 What Makes for Text to 360-degree Panorama Generation with Stable Diffusion?, *Jinhong Ni, Chang-Bin Zhang, Qiang Zhang, Jing Zhang*
- 161 DualReal: Adaptive Joint Training for Lossless Identity-Motion Fusion in Video Customization, *Wenchuan Wang, Mengqi Huang, Yijing Tu, Zhendong Mao*
- 162 Your Text Encoder Can Be An Object-Level Watermarking Controller, *Naresh Kumar Devulapally, Mingzhen Huang, Vishal Asnani, Shruti Agarwal, Siwei Lyu, Vishnu Suresh Lokhande*
- 163 ICE-Bench: A Unified and Comprehensive Benchmark for Image

- Creating and Editing, *Yulin Pan, Xiangteng He, Chaojie Mao, Zhen Han, Zeyinzi Jiang, Jingfeng Zhang, Yu Liu*
- 164 Stable Score Distillation, *Haiming Zhu, Yangyang Xu, Chenshu Xu, Tingrui Shen, Wenxi Liu, Yong Du, Jun Yu, Shengfeng He*
- 165 KV-Edit: Training-Free Image Editing for Precise Background Preservation, *Tianrui Zhu, Shiyi Zhang, Jiawei Shao, Yansong Tang*
- 166 Edit360: 2D Image Edits to 3D Assets from Any Angle, * *Junchao Huang, Xinting Hu, Shaoshuai Shi, Zhuotao Tian, Li Jiang*
- 167 FlowTok: Flowing Seamlessly Across Text and Image Tokens, *Ju He, Qihang Yu, Qihao Liu, Liang-Chieh Chen*
- 168 Principal Components Enable A New Language of Images, *Xin Wen, Bingchen Zhao, Ismail Elezi, Jiankang Deng, Xiaojuan Qi*
- 169 TeEFusion: Blending Text Embeddings to Distill Classifier-Free Guidance, *Minghao Fu, Guo-Hua Wang, Xiaohao Chen, Qing-Guo Chen, Zhao Xu, Weihua Luo, Kaifu Zhang*
- 170 TITAN-Guide: Taming Inference-Time Alignment for Guided Text-to-Video Diffusion Models, *Christian Simon, Masato Ishii, Akio Hayakawa, Zhi Zhong, Shusuke Takahashi, Takashi Shibuya, Yuki Mitsufuji*
- 171 FiVE-Bench: A Fine-grained Video Editing Benchmark for Evaluating Emerging Diffusion and Rectified Flow Models, *Minghan Li, Chenxi Xie, Yichen Wu, Lei Zhang, Mengyu Wang*
- 172 CompSlider: Compositional Slider for Disentangled Multiple-Attribute Image Generation, *Zixin Zhu, Kevin Duarte, Mamshad Nayeem Rizve, Chengyuan Xu, Ratheesh Kalarot, Junsong Yuan*
- 173 InsViE-1M: Effective Instruction-based Video Editing with Elaborate Dataset Construction, *Yuhui Wu, Liyi Chen, Ruibin Li, Shihao Wang, Chenxi Xie, Lei Zhang*
- 174 OmniVTON: Training-Free Universal Virtual Try-On, *Zhaotong Yang, Yuhui Li, Shengfeng He, Xinzhe Li, Yangyang Xu, Junyu Dong, Yong Du*
- 175 DreamRenderer: Taming Multi-Instance Attribute Control in Large-Scale Text-to-Image Models, *Dewei Zhou, Mingwei Li, Zongxin Yang, Yi Yang*
- 176 TAG-WM: Tamper-Aware Generative Image Watermarking via Diffusion Inversion Sensitivity, *Yuzhuo Chen, Zehua Ma, Han Fang, Weiming Zhang, Nenghai Yu*
- 177 X2I: Seamless Integration of Multimodal Understanding into Diffusion Transformer via Attention Distillation, *Jian Ma, Qirong Peng, Xu Guo, Chen Chen, Haonan Lu, Zhenyu Yang*
- 178 FreeFlux: Understanding and Exploiting Layer-Specific Roles in RoPE-Based MMDiT for Versatile Image Editing, *Tianyi Wei, Yifan Zhou, Dongdong Chen, Xingang Pan*
- 179 Vision-Language Interactive Relation Mining for Open-Vocabulary Scene Graph Generation, *Yukuan Min, Muli Yang, Jinhao Zhang, Yuxuan Wang, Aming Wu, Cheng Deng*
- 180 YOLO-Count: Differentiable Object Counting for Text-to-Image Generation, *Guanning Zeng, Xiang Zhang, Zirui Wang, Haiyang Xu, Zeyuan Chen, Bingnan Li, Zhuowen Tu*
- 181 UnZipLoRA: Separating Content and Style from a Single Image, *☆ *Chang Liu, Viraj Shah, Aiyu Cui, Svetlana Lazebnik*
- 182 Generative Adversarial Diffusion, *U-Chae Jun, Jaeun Ko, Jiwoo Kang*
- 183 Colors See Colors Ignore: Clothes Changing ReID with Color Disentanglement, *Priyank Pathak, Yogesh S. Rawat*
- 184 InstantEdit: Text-Guided Few-Step Image Editing with Piecewise Rectified Flow, *Yiming Gong, Zhen Zhu, Minjia Zhang*
- 185 Adversarial Distribution Matching for Diffusion Distillation * *Towards Efficient Image and Video Synthesis, Yanzuo Lu, Yuxi Ren, Xin Xia, Shanchuan Lin, Xing Wang, Xuefeng Xiao, Andy J. Ma, Xiaohua Xie, Jian-Huang Lai*
- 186 Co-Painter: Fine-Grained Controllable Image Stylization via Implicit Decoupling and Adaptive Injection, *Bowen Fu, Wei Wei, Jiaqi Tang, Jiangtao Nie, Yanyu Ye, Xiaogang Xu, Ying-Cong Chen, Lei Zhang*
- 187 AIComposer: Any Style and Content Image Composition via Feature Integration, *Haowen Li, Zhenfeng Fan, Zhang Wen, Zhengzhou Zhu, Yunjin Li*
- 188 PLA: Prompt Learning Attack against Text-to-Image Generative Models, *Xinqi Lyu, Yihao Liu, Yanjie Li, Bin Xiao*
- 189 Rethinking Layered Graphic Design Generation with a Top-Down Approach, *Jingye Chen, Zhaowen Wang, Nanxuan Zhao, Li Zhang, Difan Liu, Jimei Yang, Qifeng Chen*
- 190 Memory-Efficient Generative Models via Product Quantization, *Jie Shao, Hanxiao Zhang, Hao Yu, Jianxin Wu*
- 191 Text2VDM: Text to Vector Displacement Maps for Expressive and Interactive 3D Sculpting, *Hengyu Meng, Duotun Wang, Zhijing Shao, Ligang Liu, Zeyu Wang*
- 192 FreeScale: Unleashing the Resolution of Diffusion Models via Tuning-Free Scale Fusion, *Haonan Qiu, Shiwei Zhang, Yujie Wei, Ruihang Chu, Hangjie Yuan, Xiang Wang, Yingya Zhang, Ziwei Liu*
- 193 DiffSim: Taming Diffusion Models for Evaluating Visual Similarity, *Yiren Song, Xiaokang Liu, Mike Zheng Shou*
- 194 Holistic Tokenizer for Autoregressive Image Generation, *Anlin Zheng, Haochen Wang, Yucheng Zhao, Weipeng Deng, Tiancai Wang, Xiangyu Zhang, Xiaojuan Qi*
- 195 Scendi Score: Prompt-Aware Diversity Evaluation via Schur * *Complement of CLIP Embeddings, Azim Osmanov, Mohammad Jalali, Farzan Farnia*
- 196 Toward Better Out-painting: Improving the Image Composition with Initialization Policy Model, *Xuan Han, Yihao Zhao, Yanhao Ge, Mingyu You*
- 197 From Image to Video: An Empirical Study of Diffusion * *Representations, Pedro Vélaz, Luisa F. Polanía, Yi Yang, Chuhan Zhang, Rishabh Kabra, Anurag Arnab, Mehdi S. M. Sajjadi*
- 198 An Inversion-based Measure of Memorization for Diffusion Models, *Zhe Ma, Qingming Li, Xuhong Zhang, Tianyu Du, Ruixiao Lin, Zonghui Wang, Shouling Ji, Wenzhi Chen*
- 199 CAP: Evaluation of Persuasive and Creative Image Generation, *Aysan Aghazadeh, Adriana Kovashka*
- 200 Versatile Transition Generation with Image-to-Video Diffusion, *Zuhao Yang, Jiahui Zhang, Yingchen Yu, Shijian Lu, Song Bai*
- 201 Ultra High-Resolution Image Inpainting with Patch-Based Content Consistency Adapter, *Jianhui Zhang, Shen Cheng, Qirui Sun, Jia Liu, Wang Luyang, Chaoyu Feng, Chen Fang, Lei Lei, Jue Wang, Shuaicheng Liu*
- 202 MetaMorph: Multimodal Understanding and Generation via Instruction Tuning, *Shengbang Tong, David Fan, Jiachen Li, Yunyang Xiong, Xinlei Chen, Koustuv Sinha, Michael Rabbat, Yann LeCun, Saining Xie, Zhuang Liu*
- 203 CSD-VAR: Content-Style Decomposition in Visual Autoregressive Models, *Quang-Binh Nguyen, Minh Luu, Quang Nguyen, Anh Tran, Khoi Nguyen*
- 204 AlignGuard: Scalable Safety Alignment for Text-to-Image Generation, *Runtao Liu, I Chieh Chen, Jindong Gu, Jipeng Zhang, Renjie Pi, Qifeng Chen, Philip Torr, Ashkan Khakzar, Fabio Pizzati*
- 205 DiffIP: Representation Fingerprints for Robust IP Protection of Diffusion Models, *Zhuoling Li, Haoxuan Qu, Jason Kuen, Jiuxiang Gu, Qihong Ke, Jun Liu, Hossein Rahmani*
- 206 FairHuman: Boosting Hand and Face Quality in Human Image Generation with Minimum Potential Delay Fairness in Diffusion Models, *Yuxuan Wang, Tianwei Cao, Huayu Zhang, Zhongjiang He, Kongming Liang, Zhanyu Ma*
- 207 Processing and acquisition traces in visual encoders: What * *does CLIP know about your camera?, Ryan Ramos, Vladan Stojnić, Giorgos Kordopatis-Zilos, Yuta Nakashima, Giorgos Tolias, Noa Garcia*
- 208 Spatial-Temporal Forgery Trace based Forgery Image Identification, *Yilin Wang, Zunlei Feng, Jiachi Wang, Hengrui Lou, Binjia Zhou, Jie Lei, Mingli Song, Yijun Bei*
- 209 AM-Adapter: Appearance Matching Adapter for Exemplar-based Semantic Image Synthesis in-the-Wild, *Siyoon Jin, Jisu Nam, Jiyoung Kim, Dahyun Chung, Yeong-Seok Kim, Joonhyung Park, Heonjeong Chu, Seungryong Kim*
- 210 Aligning Global Semantics and Local Textures in Generative Video Enhancement, *Zhikai Chen, Fuchen Long, Zhaofan Qiu, Ting Yao, Wengang Zhou, Jiebo Luo, Tao Mei*
- 211 Diffusion Epistemic Uncertainty with Asymmetric Learning for Diffusion-Generated Image Detection, *Yingsong Huang, Hui Guo, Jing Huang, Bing Bai, Qi Xiong*
- 212 HypDAE: Hyperbolic Diffusion Autoencoders for Hierarchical Few-shot Image Generation, *Lingxiao Li, Kaixuan Fan, Boqing Gong, Xiangyu Yue*

- 213 STAR: Spatial-Temporal Augmentation with Text-to-Video Models for Real-World Video Super-Resolution, *Rui Xie, Yinhong Liu, Penghao Zhou, Chen Zhao, Jun Zhou, Kai Zhang, Zhenyu Zhang, Jian Yang, Zhenheng Yang, Ying Tai*
- 214 Preacher: Paper-to-Video Agentic System, *Jingwei Liu, Ling Yang, Hao Luo, Fan Wang, Hongyan Li, Mengdi Wang*
- 215 Frequency-Aware Autoregressive Modeling for Efficient High-Resolution Image Synthesis, *Zhuokun Chen, Jugang Fan, Zhuowei Yu, Bohan Zhuang, Minghui Tan*
- 216 Model Reveals What to Cache: Profiling-Based Feature Reuse for Video Diffusion Models, *Xuran Ma, Yexin Liu, Yaofu Liu, Xianfeng Wu, Mingzhe Zheng, Zihao Wang, Ser-Nam Lim, Harry Yang*
- 217 UniVG: A Generalist Diffusion Model for Unified Image Generation and Editing, *Tsu-Jui Fu, Yusu Qian, Chen Chen, Wenze Hu, Zhe Gan, Yinfei Yang*
- 218 Draw Your Mind: Personalized Generation via Condition-Level Modeling in Text-to-Image Diffusion Models, *Hyungjin Kim, Seokho Ahn, Young-Duk Seo*
- 219 Spectral Image Tokenizer, *Carlos Esteves, Mohammed Suhail, Ameesh Makadia*
- 220 VACE: All-in-One Video Creation and Editing, *Zeyinzi Jiang, Zhen Han, Chaojie Mao, Jingfeng Zhang, Yulin Pan, Yu Liu*
- 221 RomanTex: Decoupling 3D-aware Rotary Positional Embedded Multi-Attention Network for Texture Synthesis, *Yifei Feng, Mingxin Yang, Shuhui Yang, Sheng Zhang, Jiaao Yu, Zibo Zhao, Yuhong Liu, Jie Jiang, Chunchao Guo*
- 222 ImageGen-CoT: Enhancing Text-to-Image In-context Learning with Chain-of-Thought Reasoning, *Jiaqi Liao, Zhengyuan Yang, Linjie Li, Dianqi Li, Kevin Lin, Yu Cheng, Lijuan Wang*
- 223 Calibrating MLLM-as-a-judge via Multimodal Bayesian Prompt Ensembles, *Eric Slyman, Mehrab Tanjim, Kushal Kalle, Stefan Lee*
- 224 V.I.P.: Iterative Online Preference Distillation for Efficient Video Diffusion Models, *Jisoo Kim, Wooseok Seo, Junwan Kim, Seungho Park, Sooyeon Park, Youngjae Yu*
- 225 LOTA: Bit-Planes Guided AI-Generated Image Detection, *Hongsong Wang, Renxi Cheng, Yang Zhang, Chaolei Han, Jie Gui*
- 226 Forecasting Continuous Non-Conservative Dynamical Systems in $SO(3)$, *Lennart Bastian, Mohammad Rashed, Nassir Navab, Tolga Birdal*
- 227 Dynamic Typography: Bringing Text to Life via Video Diffusion Prior, *Zichen Liu, Yihao Meng, Hao Ouyang, Yue Yu, Bolin Zhao, Daniel Cohen-Or, Huamin Qu*
- 228 Trade-offs in Image Generation: How Do Different Dimensions Interact?, *Sicheng Zhang, Binzhu Xie, Zhonghao Yan, Yuli Zhang, Donghao Zhou, Xiaofei Chen, Shi Qiu, Jiaqi Liu, Guoyang Xie, Zhichao Lu*
- 229 X-Prompt: Generalizable Auto-Regressive Visual Learning with In-Context Prompting, *Zeyi Sun, Ziyang Chu, Pan Zhang, Tong Wu, Yuhang Zang, Xiaoyi Dong, Yuanjun Xiong, Dahua Lin, Jiaqi Wang*
- 230 Long Context Tuning for Video Generation, *Yuwei Guo, Ceyuan Yang, Ziyang Yang, Zhibei Ma, Zhijie Lin, Zhenheng Yang, Dahua Lin, Lu Jiang*
- 231 DreamFuse: Adaptive Image Fusion with Diffusion Transformer, *Junjia Huang, Pengxiang Yan, Jiyang Liu, Jie Wu, Zhao Wang, Yitong Wang, Liang Lin, Guanbin Li*
- 232 AnyI2V: Animating Any Conditional Image with Motion Control, *Ziye Li, Hao Luo, Xincheng Shuai, Henghui Ding*
- 233 LMM4LMM: Benchmarking and Evaluating Large-multimodal Image Generation with LMMs, *Jiarui Wang, Huiyu Duan, Yu Zhao, Juntong Wang, Guangtao Zhai, Xiongkuo Min*
- 234 PlugMark: A Plug-in Zero-Watermarking Framework for Diffusion Models, *Pengzhen Chen, Yanwei Liu, Xiaoyan Gu, Enci Liu, Zhuoyi Shang, Xiangyang Ji, Wu Liu*
- 235 OmniPaint: Mastering Object-Oriented Editing via Disentangled Insertion-Removal Inpainting, *Yongsheng Yu, Ziyun Zeng, Haitian Zheng, Jiebo Luo*
- 236 Balanced Image Stylization with Style Matching Score, *Yuxin Jiang, Liming Jiang, Shuai Yang, Jia-Wei Liu, Ivor W. Tsang, Mike Zheng Shou*
- 237 GLViC: Generative Implicit Video Compression, *Ge Gao, Siyue Teng, Tianhao Peng, Fan Zhang, David Bull*
- 238 Streamlining Image Editing with Layered Diffusion Brushes, *Peyman Gholami, Robert Xiao*
- 239 StableCodec: Taming One-Step Diffusion for Extreme Image Compression, *Tianyu Zhang, Xin Luo, Li Li, Dong Liu*
- 240 Rethinking Discrete Tokens: Treating Them as Conditions for Continuous Autoregressive Image Synthesis, *Peng Zheng, Junke Wang, Yi Chang, Yizhou Yu, Rui Ma, Zuxuan Wu*
- 241 Can We Achieve Efficient Diffusion Without Self-Attention? Distilling Self-Attention into Convolutions, *Ziyi Dong, Chengxing Zhou, Weijian Deng, Pengxu Wei, Xiangyang Ji, Liang Lin*
- 242 DADet: Safeguarding Image Conditional Diffusion Models against Adversarial and Backdoor Attacks via Diffusion Anomaly Detection, *Hongwei Yu, Xinlong Ding, Jiawei Li, Jinlong Wang, Yudong Zhang, Rongquan Wang, Huimin Ma, Jiansheng Chen*
- 243 Latent Diffusion Models with Masked AutoEncoders, *Junho Lee, Jeongwoo Shin, Hyungwook Choi, Joonseok Lee*
- 244 Imbalance in Balance: Online Concept Balancing in Generation Models, *Yukai Shi, Jiarong Ou, Rui Chen, Haotian Yang, Jiahao Wang, Xin Tao, Pengfei Wan, Di Zhang, Kun Gai*
- 245 OCR Hinders RAG: Evaluating the Cascading Impact of OCR on Retrieval-Augmented Generation, *Junyuan Zhang, Qintong Zhang, Bin Wang, Linke Ouyang, Zichen Wen, Ying Li, Ka-Ho Chow, Conghui He, Wentao Zhang*
- 246 Multi-scenario Overlapping Text Segmentation with Depth Awareness, *Yang Liu, Xudong Xie, Yuliang Liu, Xiang Bai*
- 247 TextSSR: Diffusion-based Data Synthesis for Scene Text Recognition, *Xingsong Ye, Yongkun Du, Yunbo Tao, Zhineng Chen*
- 248 EEdit: Rethinking the Spatial and Temporal Redundancy for Efficient Image Editing, *Zexuan Yan, Yue Ma, Chang Zou, Wenteng Chen, Qifeng Chen, Linfeng Zhang*
- 249 RAGDiffusion: Faithful Cloth Generation via External Knowledge Assimilation, *Yuhan Li, Xianfeng Tan, Wenxiang Shang, Yubo Wu, Jian Wang, Xuanhong Chen, Yi Zhang, Hangcheng Zhu, Bingbing Ni*
- 250 Wasserstein Style Distribution Analysis and Transform for Stylized Image Generation, *Xi Yu, Xiang Gu, Zhihao Shi, Jian Sun*
- 251 Instruction-based Image Editing with Planning, Reasoning, and Generation, *Liya Ji, Chenyang Qi, Qifeng Chen*
- 252 Discovering Divergent Representations between Text-to-Image Models, *Lisa Dunlap, Joseph E. Gonzalez, Trevor Darrell, Fabian Caba Heilbron, Josef Sivic, Bryan Russell*
- 253 DDB: Diffusion Driven Balancing to Address Spurious Correlations, *Aryan Yazdan Parast, Basim Azam, Naveed Akhtar*
- 254 HDR Image Generation via Gain Map Decomposed Diffusion, *Yuanshen Guan, Ruikang Xu, Yinyao Liao, Mingde Yao, Lizhi Wang, Zhiwei Xiong*
- 255 ESSENTIAL: Episodic and Semantic Memory Integration for Video Class-Incremental Learning, *Jongseo Lee, Kyungho Bae, Kyle Min, Gyeong-Moon Park, Jinwoo Choi*
- 256 AutoPrompt: Automated Red-Teaming of Text-to-Image Models via LLM-Driven Adversarial Prompts, *Yufan Liu, Wangqian Zhang, Huashan Chen, Lin Wang, Xiaojun Jia, Zheng Lin, Weiping Wang*
- 257 Fair Generation without Unfair Distortions: Debiasing Text-to-Image Generation with Entanglement-Free Attention, *Jeonghoon Park, Juyoung Lee, Chaeyeon Chung, Jaeseong Lee, Jaegul Choo, Jindong Gu*
- 258 Training-Free Text-Guided Image Editing with Visual Autoregressive Model, *Yufei Wang, Lanqing Guo, Zhihao Li, Jiaying Huang, Pichao Wang, Bihan Wen, Jian Wang*
- 259 QR-LoRA: Efficient and Disentangled Fine-tuning via QR Decomposition for Customized Generation, *Jiahui Yang, Yongjia Ma, Donglin Di, Jianxun Cui, Hao Li, Wei Chen, Yan Xie, Xun Yang, Wangmeng Zuo*
- 260 Uncover Treasures in DCT: Advancing JPEG Quality Enhancement by Exploiting Latent Correlations, *Jing Yang, Qunliang Xing, Mai Xu, Minglang Qiao*
- 261 Accelerating Diffusion Transformer via Gradient-Optimized Cache, *Xunxiang Qiu, Lin Liu, Shuo Wang, Jinda Lu, Kezhou Chen, Yanbin Hao*
- 262 The Silent Assistant: NoiseQuery as Implicit Guidance for Goal-Driven Image Generation, *Ruoyu Wang, Huayang Huang, Ye Zhu, Olga Russakovsky, Yu Wu*

- 263 Progressive Growing of Video Tokenizers for Temporally Compact
☆ Latent Spaces, *Aniruddha Mahapatra, Long Mai, David Bourgin, Yitian Zhang, Feng Liu*
- 264 Multimodal Latent Diffusion Model for Complex Sewing Pattern Generation, *Shengqi Liu, Yuhao Cheng, Zhuo Chen, Xingyu Ren, Wenhan Zhu, Lincheng Li, Mengxiao Bi, Xiaokang Yang, Yichao Yan*
- 265 ArtEditor: Learning Customized Instructional Image Editor from Few-Shot Examples, *Shijie Huang, Yiren Song, Yuxuan Zhang, Hailong Guo, Xueyin Wang, Jiaming Liu*
- 266 Are They the Same? Exploring Visual Correspondence Shortcomings of Multimodal LLMs, *Yikang Zhou, Tao Zhang, Shilin Xu, Shihao Chen, Qianyu Zhou, Yunhai Tong, Shunping Ji, Jiangning Zhang, Lu Qi, Xiangtai Li*
- 267 MC-Bench: A Benchmark for Multi-Context Visual Grounding in the Era of MLLMs, *Yunqiu Xu, Linchao Zhu, Yi Yang*
- 268 Golden Noise for Diffusion Models: A Learning Framework, *Zikai Zhou, Shitong Shao, Lichen Bai, Shufei Zhang, Zhiqiang Xu, Bo Han, Zeke Xie*
- 269 Disrupting Model Merging: A Parameter-Level Defense Without Sacrificing Accuracy, *Wei Junhao, Yu Zhe, Jun Sakuma*
- 270 Towards Stabilized and Efficient Diffusion Transformers through Long-Skip-Connections with Spectral Constraints, *Guanjie Chen, Xinyu Zhao, Yucheng Zhou, Xiaoye Qu, Tianlong Chen, Yu Cheng*
- 271 Learning Few-Step Diffusion Models by Trajectory Distribution Matching, *Yihong Luo, Tianyang Hu, Jiacheng Sun, Yujun Cai, Jing Tang*
- 272 End-to-End Entity-Predicate Association Reasoning for Dynamic Scene Graph Generation, *Liwei Wang, Yanduo Zhang, Tao Lu, Fang Liu, Huiqin Zhang, Jiayi Ma, Huabing Zhou*
- 273 Harmonizing Visual Representations for Unified Multimodal Understanding and Generation, *Size Wu, Wenwei Zhang, Lumin Xu, Sheng Jin, Zhonghua Wu, Qingyi Tao, Wentao Liu, Wei Li, Chen Change Loy*
- 274 A3GS: Arbitrary Artistic Style into Arbitrary 3D Gaussian Splatting, *Zhiyuan Fang, Rengan Xie, Xuancheng Jin, Qi Ye, Wei Chen, Wenting Zheng, Rui Wang, Yuchi Huo*
- 275 HouseTour: A Virtual Real Estate A(I)gent, *Ata Çelen, Marc Pollefeys, Daniel Barath, Iro Armeni*
- 276 Forensic-MoE: Exploring Comprehensive Synthetic Image Detection Traces with Mixture of Experts, *Mingqi Fang, Ziguang Li, Lingyun Yu, Quanwei Yang, Hongtao Xie, Yongdong Zhang*
- 277 LayerD: Decomposing Raster Graphic Designs into Layers,
☆ Tomoyuki Suzuki, Kang-Jun Liu, Naoto Inoue, Kota Yamaguchi
- 278 TikZero: Zero-Shot Text-Guided Graphics Program Synthesis,
* Jonas Belouadi, Eddy Ilg, Margret Keuper, Hideki Tanaka, Masao Utiyama, Raj Dabre, Steffen Eger, Simone Ponzetto
- 279 ViLU: Learning Vision-Language Uncertainties for Failure Prediction, *Marc Lafon, Yannis Karmim, Julio Silva-Rodríguez, Paul Couairon, Clément Rambour, Raphael Fournier-Sniehotta, Ismail Ben Ayed, Jose Dolz, Nicolas Thome*
- 280 Beyond Blur: A Fluid Perspective on Generative Diffusion Models, *Grzegorz Gruszczynski, Jakub Meixner, Michal Wlodarczyk, Przemyslaw Musialski*
- 281 Conditional Visual Autoregressive Modeling for Pathological Image
☆ Restoration, *Ziyi Liu, Zhe Xu, Jiabo Ma, Wenqiang Li, Ruixuan Wang, Bo Du, Hao Chen*
- 282 Subjective Camera 1.0: Bridging Human Cognition and Visual
☆ Reconstruction through Sequence-Aware Sketch-Guided Diffusion, *Haoyang Chen, Dongfang Sun, Caoyuan Ma, Shiqin Wang, Kewei Zhang, Zheng Wang, Zhixiang Wang*
- 283 GlassWizard: Harvesting Diffusion Priors for Glass Surface Detection, *Wenxue Li, Tian Ye, Xinyu Xiong, Jinbin Bai, Feilong Tang, Wenxuan Song, Zhaohu Xing, Lie Ju, Guanbin Li, Lei Zhu*
- 284 SKALD: Learning-Based Shot Assembly for Coherent Multi-Shot Video Creation, *Chen-Yi Lu, Md Mehrab Tanjim, Ishita Dasgupta, Somdeb Sarkhel, Gang Wu, Saayan Mitra, Somali Chaterji*
- 285 FedMVP: Federated Multimodal Visual Prompt Tuning for Vision-Language Models, *Mainak Singha, Subhankar Roy, Sarthak Mehrotra, Ankit Jha, Moloud Abdar, Biplab Banerjee, Elisa Ricci*
- 286 LoRAverse: A Submodular Framework to Retrieve Diverse Adapters for Diffusion Models, *Mert Sonmezer, Matthew Zheng, Pinar Yanardag*
- 287 HyTIP: Hybrid Temporal Information Propagation for Masked Conditional Residual Video Coding, *Yi-Hsin Chen, Yi-Chen Yao, Kuan-Wei Ho, Chun-Hung Wu, Huu-Tai Phung, Martin Benjak, Jörn Ostermann, Wen-Hsiao Peng*
- 288 DACoN: DINO for Anime Paint Bucket Colorization with Any Number of Reference Images, *Kazuma Nagata, Naoshi Kaneko*
- 289 Lay-Your-Scene: Natural Scene Layout Generation with Diffusion Transformers, *Divyansh Srivastava, Xiang Zhang, He Wen, Chenru Wen, Zhuowen Tu*
- 290 Free2Guide: Training-Free Text-to-Video Alignment using Image LVLm, *Jaemin Kim, Bryan Sangwoo Kim, Jong Chul Ye*
- 291 Hybrid Layout Control for Diffusion Transformer: Fewer Annotations, Superior Aesthetics, *Keming Wu, Junwen Chen, Zhanhao Liang, Yinuo Wang, Ji Li, Chao Zhang, Bin Wang, Yuhui Yuan*
- 292 InfGen: A Resolution-Agnostic Paradigm for Scalable Image Synthesis, *Tao Han, Wanghan Xu, Junchao Gong, Xiaoyu Yue, Song Guo, Luping Zhou, Lei Bai*
- 293 VideoVAE+: Large Motion Video Autoencoding with Cross-modal Video VAE, *Yazhou Xing, Yang Fei, Yingqing He, Jingye Chen, Jiaxin Xie, Xiaowei Chi, Qifeng Chen*
- 294 DOLLAR: Few-Step Video Generation via Distillation and Latent Reward Optimization, *Zihan Ding, Chi Jin, Difan Liu, Haitian Zheng, Krishna Kumar Singh, Qiang Zhang, Yan Kang, Zhe Lin, Yuchen Liu*
- 295 Dual-Process Image Generation, *Grace Luo, Jonathan Granskog, Aleksander Holynski, Trevor Darrell*
- 296 SpecGuard: Spectral Projection-based Advanced Invisible Watermarking, *Inzamamul Alam, Md Tanvir Islam, Simon S. Woo, Khan Muhammad*
- 297 DIA: The Adversarial Exposure of Deterministic Inversion in Diffusion Models, *Seunghoo Hong, Geonho Son, Juhun Lee, Simon S. Woo*
- 298 Supercharged One-step Text-to-Image Diffusion Models with Negative Prompts, *Viet Nguyen, Anh Nguyen, Trung Dao, Khoi Nguyen, Cuong Pham, Toan Tran, Anh Tran*
- 299 GFPack++: Attention-Driven Gradient Fields for Optimizing 2D
* Irregular Packing, *Tianyang Xue, Lin Lu, Yang Liu, Mingdong Wu, Hao Dong, Yanbin Zhang, Renmin Han, Baoquan Chen*
- 300 Denoising Token Prediction in Masked Autoregressive Models, *Ting Yao, Yehao Li, Yingwei Pan, Zhaofan Qiu, Tao Mei*
- 301 DC-AR: Efficient Masked Autoregressive Image Generation with Deep Compression Hybrid Tokenizer, *Yecheng Wu, Han Cai, Junyu Chen, Zhuoyang Zhang, Enze Xie, Jincheng Yu, Junsong Chen, Jinyi Hu, Yao Lu, Song Han*
- 302 LACONIC: A 3D Layout Adapter for Controllable Image Creation, *Léopold Maillard, Tom Durand, Adrien Ramanana Rahary, Maks Ovsjanikov*
- 303 Preserve Anything: Controllable Image Synthesis with Object Preservation, *Prasen Kumar Sharma, Neeraj Matiyali, Siddharth Srivastava, Gaurav Sharma*
- 304 Rethinking DPO-style Diffusion Aligning Frameworks, *Xun Wu, Shaohan Huang, Lingjie Jiang, Furu Wei*
- 305 Certifiably Optimal Anisotropic Rotation Averaging,
○ Carl Olsson, Yaroslava Lochman, Johan Malmport,
☞ Christopher Zach
- 306 Generating Physically Stable and Buildable Brick Structures from
○ Text, *Ava Pun, Kangle Deng, Ruixuan Liu, Deva Ramanan, Changliu Liu, Jun-Yan Zhu*
- 307 Ask and Remember: A Questions-Only Replay Strategy for Continual Visual Question Answering, *Imad Eddine Marouf, Enzo Tartaglione, Stéphane Lathuilière, Joost Van De Weijer*
- 308 Contrastive Test-Time Composition of Multiple LoRA Models
* for Image Generation, *Tuna Han Salih Meral, Enis Simsar, Federico Tombari, Pinar Yanardag*
- 309 LLM Thought Divergence and Convergence for Dialogue-Based Image Generation Control, *Hui Li*
- 310 FreeMorph: Tuning-Free Generalized Image Morphing with Diffusion Model, *Yukang Cao, Chenyang Si, Jinghao Wang, Ziwei Liu*
- 311 FramePainter: Endowing Interactive Image Editing with Video Diffusion Priors, *Yabo Zhang, Xinpeng Zhou, Yihan Zeng, Hang Xu, Hui Li, Wangmeng Zuo*
- 312 TurboVSR: Fantastic Video Upscalers and Where to Find Them,
* Zhongdao Wang, Guodongfang Zhao, Jingjing Ren, Bailan Feng, Shifeng Zhang, Wenbo Li

- 313 PlanGen: Towards Unified Layout Planning and Image Generation in Auto-Regressive Vision Language Models, *Runze He, Bo Cheng, Yuhang Ma, Qingxiang Jia, Shanyuan Liu, Ao Ma, Xiaoyu Wu, Liebucha Wu, Dawei Leng, Yuhui Yin*
- 314 Turbo2K: Towards Ultra-Efficient and High-Quality 2K Video Synthesis, *Jingjing Ren, Wenbo Li, Zhongdao Wang, Haoze Sun, Bangzhen Liu, Haoyu Chen, Jiaqi Xu, Aoxue Li, Shifeng Zhang, Bin Shao, Yong Guo, Lei Zhu*
- 315 Anchor Token Matching: Implicit Structure Locking for Training-free AR Image Editing, *Taihang Hu, Linxuan Li, Kai Wang, Yaxing Wang, Jian Yang, Ming-Ming Cheng*
- 316 Improving Rectified Flow with Boundary Conditions, ☆ *Xixi Hu, Runlong Liao, Keyang Xu, Bo Liu, Yeqing Li, Eugene Ie, Hongliang Fei, Qiang Liu*
- 317 Decoding Correlation-Induced Misalignment in the Stable Diffusion Workflow for Text-to-Image Generation, *Yunze Tong, Fengda Zhang, Didi Zhu, Jun Xiao, Kun Kuang*
- 318 TokensGen: Harnessing Condensed Tokens for Long Video Generation, *Wenqi Ouyang, Zeqi Xiao, Danni Yang, Yifan Zhou, Shuai Yang, Lei Yang, Jianlou Si, Xingang Pan*
- 319 Parametric Shadow Control for Portrait Generation in Text-to-Image Diffusion Models, ☆ *Haoming Cai, Tsung-Wei Huang, Shiv Gehlot, Brandon Y. Feng, Sachin Shah, Guan-Ming Su, Christopher Metzler*
- 320 IGD: Instructional Graphic Design with Multimodal Layer Generation, *Yadong Qu, Shancheng Fang, Yuxin Wang, Xiaorui Wang, Zhineng Chen, Hongtao Xie, Yongdong Zhang*
- 321 GenDoP: Auto-regressive Camera Trajectory Generation as a Director of Photography, *Mengchen Zhang, Tong Wu, Jing Tan, Ziwei Liu, Gordon Wetzstein, Dahua Lin*
- 322 OURO: A Self-Bootstrapped Framework for Enhancing Multimodal Scene Understanding, *Tianrun Xu, Guanyu Chen, Ye Li, Yuxin Xi, Zeyu Mu, Ruichen Wang, Tianren Zhang, Haichuan Gao, Feng Chen*
- 323 CompleteMe: Reference-based Human Image Completion, ☆ *Yu-Ju Tsai, Brian Price, Qing Liu, Luis Figueroa, Daniil Pakhomov, Zhihong Ding, Scott Cohen, Ming-Hsuan Yang*
- 324 REPA-E: Unlocking VAE for End-to-End Tuning of Latent Diffusion Transformers, *Xingjian Leng, Jaskirat Singh, Yunzhong Hou, Zhenchang Xing, Saining Xie, Liang Zheng*
- 325 EEGMirror: Leveraging EEG Data in the Wild via Montage-Agnostic Self-Supervision for EEG to Video Decoding, *Xuan-Hao Liu, Bao-Liang Lu, Wei-Long Zheng*
- 326 Accelerating Diffusion Sampling via Exploiting Local Transition Coherence, *Shangwen Zhu, Han Zhang, Zhantao Yang, Qianyu Peng, Zhao Pu, Huangji Wang, Fan Cheng*
- 327 SA-LUT: Spatial Adaptive 4D Look-Up Table for Photorealistic Style Transfer, *Zerui Gong, Zhonghua Wu, Qingyi Tao, Qinyue Li, Chen Change Loy*
- 328 Inpaint4Drag: Repurposing Inpainting Models for Drag-Based Image Editing via Bidirectional Warping, ☆ *Jingyi Lu, Kai Han*
- 329 UniversalBooth: Model-Agnostic Personalized Text-to-Image Generation, *Songhua Liu, Ruonan Yu, Xinchao Wang*
- 330 UniCombine: Unified Multi-Conditional Combination with Diffusion Transformer, *Haoxuan Wang, Jinlong Peng, Qingdong He, Hao Yang, Ying Jin, Jiafu Wu, Xiaobin Hu, Yanjie Pan, Zhenye Gan, Mingmin Chi, Bo Peng, Yabiao Wang*
- 331 UniGlyph: Unified Segmentation-Conditioned Diffusion for Precise Visual Text Synthesis, *Yuanrui Wang, Cong Han, Yafei Li, Zhipeng Jin, Xiawei Li, SiNan Du, Wen Tao, Shuanglong Li, Yi Yang, Chun Yuan, Liu Lin*
- 332 ADIEE: Automatic Dataset Creation and Scorer for Instruction-Guided Image Editing Evaluation, *Sherry X. Chen, Yi Wei, Luowei Zhou, Suren Kumar*
- 333 SUV: Suppressing Undesired Video Content via Semantic Modulation Based on Text Embeddings, *Xiang Lv, Mingwen Shao, Lingzhuang Meng, Chang Liu, Yecong Wan, Xinyuan Chen*
- 334 Neurons: Emulating the Human Visual Cortex Improves Fidelity and Interpretability in fMRI-to-Video Reconstruction, *Haonan Wang, Qixiang Zhang, Lehan Wang, Xuanqi Huang, Xiaomeng Li*
- 335 TF-TI2I: Training-Free Text-and-Image-to-Image Generation via Multi-Modal Implicit-Context Learning In Text-to-Image Models, *Teng-Fang Hsiao, Bo-Kai Ruan, Yi-Lun Wu, Tzu-Ling Lin, Hong-Han Shuai*
- 336 Semantic Discrepancy-aware Detector for Image Forgery Identification, *Ziye Wang, Minghang Yu, Chunyan Xu, Zhen Cui*
- 337 Scalable Ranked Preference Optimization for Text-to-Image Generation, *Shyamgopal Karthik, Huseyin Coskun, Zeynep Akata, Sergey Tulyakov, Jian Ren, Anil Kag*
- 338 FairGen: Enhancing Fairness in Text-to-Image Diffusion Models via Self-Discovering Latent Directions, *Yilei Jiang, Wei-Hong Li, Yiyuan Zhang, Minghong Cai, Xiangyu Yue*
- 339 ConceptSplit: Decoupled Multi-Concept Personalization of Diffusion Models via Token-wise Adaptation and Attention Disentanglement, *Habin Lim, Yeongseob Won, Juwon Seo, Gyeong-Moon Park*
- 340 Randomized Autoregressive Visual Generation, *Qihang Yu, Ju He, Xueqing Deng, Xiaohui Shen, Liang-Chieh Chen*
- 341 Democratizing Text-to-Image Masked Generative Models with Compact Text-Aware One-Dimensional Tokens, *Dongwon Kim, Ju He, Qihang Yu, Chenglin Yang, Xiaohui Shen, Suha Kwak, Liang-Chieh Chen*
- 342 REGEN: Learning Compact Video Embedding with (Re-)Generative ☆ *Decoder, Yitian Zhang, Long Mai, Aniruddha Mahapatra, David Bourgin, Yicong Hong, Jonah Casebeer, Feng Liu, Yun Fu*
- 343 FontS: Text Rendering With Typography and Style Controls, *Wenda Shi, Yiren Song, Dengming Zhang, Jiaming Liu, Xingxing Zou*
- 344 USP: Unified Self-Supervised Pretraining for Image Generation and Understanding, *Xiangxiang Chu, Renda Li, Yong Wang*
- 345 CreatiLayout: Siamese Multimodal Diffusion Transformer for Creative Layout-to-Image Generation, *Hui Zhang, Dexiang Hong, Yitong Wang, Jie Shao, Xinglong Wu, Zuxuan Wu, Yu-Gang Jiang*
- 346 MVQA: Mamba with Unified Sampling for Efficient Video Quality ☆ *Assessment, Yachun Mi, Yu Li, Weicheng Meng, Chaofeng Chen, Chen Hui, Shaohui Liu*
- 347 DMQ: Dissecting Outliers of Diffusion Models for Post-Training Quantization, *Dongyeun Lee, Jiwan Hur, Hyounguk Shon, Jae Young Lee, Junmo Kim*
- 348 CoMatch: Dynamic Covisibility-Aware Transformer for Bilateral ☆ *Subpixel-Level Semi-Dense Image Matching, Zizhuo Li, Yifan Lu, Linfeng Tang, Shihua Zhang, Jiayi Ma*
- 349 Guiding Noisy Label Conditional Diffusion Models with Score-based Discriminator Correction, *Dat Nguyen Cong, Hieu Tran Bao, Tung Hoang-Thanh*
- 350 HiGarment: Cross-modal Harmony Based Diffusion Model for Flat Sketch to Realistic Garment Image, *Junyi Guo, Jingxuan Zhang, Fangyu Wu, Huanda Lu, Qiufeng Wang, Wenmian Yang, Eng Gee Lim, Dongming Lu*
- 351 TCFG: Truncated Classifier-Free Guidance for Efficient and Scalable Text-to-Image Acceleration, *Xiaomeng Fu, Jia Li*
- 352 Benchmarking and Learning Multi-Dimensional Quality Evaluator for Text-to-3D Generation, *Yujie Zhang, Bingyang Cui, Qi Yang, Zhu Li, Yiting Xu*
- 353 Automated Red Teaming for Text-to-Image Models through Feedback-Guided Prompt Iteration with Vision-Language Models, *Wei Xu, Kangjie Chen, Jiawei Qiu, Yuyang Zhang, Run Wang, Jin Mao, Tianwei Zhang, Lina Wang*
- 354 PASTA: Part-Aware Sketch-to-3D Shape Generation with Text-Aligned Prior, *Seunggywan Lee, Hwanhee Jung, Byoungsoo Koh, Qixing Huang, Sang Ho Yoon, Sangpil Kim*
- 355 Bridging Continuous and Discrete Tokens for Autoregressive Visual Generation, *Yuqing Wang, Zhijie Lin, Yao Teng, Yuanzhi Zhu, Shuhuai Ren, Jiashi Feng, Xihui Liu*
- 356 Di[M]O: Distilling Masked Diffusion Models into One-step Generator, *Yuanzhi Zhu, Xi Wang, Stéphane Lathuilière, Vicky Kalogeiton*
- 357 Gain-MLP: Improving HDR Gain Map Encoding via a Lightweight MLP, *Trevor D. Canham, SaiKiran Tedla, Michael J. Murdoch, Michael S. Brown*
- 358 TrustMark: Robust Watermarking and Watermark Removal for ☆ *Arbitrary Resolution Images, Tu Bui, Shruti Agarwal, John Collomosse*
- 359 Perceive, Understand and Restore: Real-World Image Super-Resolution with Autoregressive Multimodal Generative Models, *Hongyang Wei, Shuaizheng Liu, Chun Yuan, Lei Zhang*
- 360 From Prompt to Progression: Taming Video Diffusion Models for Seamless Attribute Transition, *Ling Lo, Kelvin C.K. Chan,*

- Wen-Huang Cheng, Ming-Hsuan Yang
- 361 Dense2MoE: Restructuring Diffusion Transformer to MoE for Efficient Text-to-Image Generation, Youwei Zheng, Yuxi Ren, Xin Xia, Xuefeng Xiao, Xiaohua Xie
- 362 Video-T1: Test-time Scaling for Video Generation, Fangfu Liu, Hanyang Wang, Yimo Cai, Kaiyan Zhang, Xiaohang Zhan, Yueqi Duan
- 363 Less-to-More Generalization: Unlocking More Controllability by In-Context Generation, Shaojin Wu, Mengqi Huang, Wenxu Wu, Yufeng Cheng, Fei Ding, Qian He
- 364 StyleSRN: Scene Text Image Super-Resolution with Text Style Embedding, Shengrong Yuan, Runmin Wang, Ke Hao, Xuqi Ma, Changxin Gao, Li Liu, Nong Sang
- 365 Sparse Fine-Tuning of Transformers for Generative Tasks, ☆ Wei Chen, Jingxi Yu, Zichen Miao, Qiang Qiu
- 366 FlexGen: Flexible Multi-View Generation from Text and Image Inputs, Xinli Xu, Wenhang Ge, Jiantao Lin, Jiawei Feng, Lie Xu, Hanfeng Zhao, Shunsi Zhang, Ying-Cong Chen
- 367 Prompt-A-Video: Prompt Your Video Diffusion Model via Preference-Aligned LLM, Yatai Ji, Jiacheng Zhang, Jie Wu, Shilong Zhang, Shoufa Chen, Chongjian Ge, Peize Sun, Weifeng Chen, Wenqi Shao, Xuefeng Xiao, Weilin Huang, Ping Luo
- 368 Learning Implicit Features with Flow-Infused Transformations for Realistic Virtual Try-On, Delong Zhang, Qiwei Huang, Yang Sun, Yuanliu Liu, Wei-Shi Zheng, Pengfei Xiong, Wei Zhang
- 369 ALGI-Holmes: Towards Explainable and Generalizable AI-Generated Image Detection via Multimodal Large Language Models, Ziyin Zhou, Yunpeng Luo, Yuanchen Wu, Ke Sun, Jiayi Ji, Ke Yan, Shouhong Ding, Xiaoshuai Sun, Yunsheng Wu, Rongrong Ji
- 370 Semantic Watermarking Reinvented: Enhancing Robustness and Generation Quality with Fourier Integrity, Sung Ju Lee, Nam Ik Cho
- 371 GigaTok: Scaling Visual Tokenizers to 3 Billion Parameters for Autoregressive Image Generation, Tianwei Xiong, Jun Hao Liew, Zilong Huang, Jiashi Feng, Xihui Liu
- 372 Collaborative Instance Object Navigation: Leveraging Uncertainty-Awareness to Minimize Human-Agent Dialogues, Francesco Taioli, Edoardo Zorzi, Gianni Franchi, Alberto Castellini, Alessandro Farinelli, Marco Cristani, Yiming Wang
- 373 ReasonVQA: A Multi-hop Reasoning Benchmark with Structural Knowledge for Visual Question Answering, Duong T. Tran, Trung-Kien Tran, Manfred Hauswirth, Danh Le Phuoc
- 374 LOCATEdit: Graph Laplacian Optimized Cross Attention for Localized Text-Guided Image Editing, Achint Soni, Meet Soni, Sirisha Rambhatla
- 375 Enhancing Mamba Decoder with Bidirectional Interaction in Multi-Task Dense Prediction, Mang Cao, Sanping Zhou, Yizhe Li, Ye Deng, Wenli Huang, Le Wang
- 376 Region-Level Data Attribution for Text-to-Image Generative Models, Trong Bang Nguyen, Phi Le Nguyen, Simon Lucey, Minh Hoai
- 377 Learned Image Compression with Hierarchical Progressive Context Modeling, Yuqi Li, Haotian Zhang, Li Li, Dong Liu
- 378 Early Timestep Zero-Shot Candidate Selection for Instruction-Guided Image Editing, Joowon Kim, Ziseok Lee, Donghyeon Cho, Sanghyun Jo, Yeonsung Jung, Kyungsu Kim, Eunho Yang
- 379 GTR: Guided Thought Reinforcement Prevents Thought Collapse in RL-based VLM Agent Training, Tong Wei, Yijun Yang, Junliang Xing, Yuanchun Shi, Zongqing Lu, Deheng Ye
- 380 Teleportraits: Training-Free People Insertion into Any Scene, Jialu Gao, K J Joseph, Fernando De La Torre
- 381 DCT-Shield: A Robust Frequency Domain Defense against ☆ Malicious Image Editing, Aniruddha Bala, Rohit Chowdhury, Rohan Jaiswal, Siddharth Roheda
- 382 Context Guided Transformer Entropy Modeling for Video Compression, Junlong Tong, Wei Zhang, Yaohui Jin, Xiaoyu Shen
- 383 Deterministic Object Pose Confidence Region Estimation, ○ Jinghao Wang, Zhang Li, Zi Wang, Banglei Guan, Yang Shang, Qifeng Yu
- 384 WIR3D: Visually-Informed and Geometry-Aware 3D Shape ○ Abstraction, Richard Liu, Daniel Fu, Noah Tan, Itai Lang, Rana Hanocka
- 385 UIP2P: Unsupervised Instruction-based Image Editing via Edit ☆ Reversibility Constraint, Enis Simsar, Alessio Tonioni, Yongqin Xian, Thomas Hofmann, Federico Tombari
- 386 CODA: Repurposing Continuous VAEs for Discrete Tokenization, Zeyu Liu, Zanlin Ni, Yeguo Hua, Xin Deng, Xiao Ma, Cheng Zhong, Gao Huang
- 387 DiffDoctor: Diagnosing Image Diffusion Models Before Treating, Yiyang Wang, Xi Chen, Xiaogang Xu, Sihui Ji, Yu Liu, Yujun Shen, Hengshuang Zhao
- 388 TRCE: Towards Reliable Malicious Concept Erasure in Text-to-Image Diffusion Models, Ruidong Chen, Honglin Guo, Lanjun Wang, Chenyu Zhang, Weizhi Nie, An-An Liu
- 389 LEGION: Learning to Ground and Explain for Synthetic Image * Detection, Hengrui Kang, Siwei Wen, Zichen Wen, Junyan Ye, Weijia Li, Peilin Feng, Baichuan Zhou, Bin Wang, Dahua Lin, Linfeng Zhang, Conghui He
- 390 DiT4SR: Taming Diffusion Transformer for Real-World Image Super-Resolution, Zheng-Peng Duan, Jiawei Zhang, Xin Jin, Ziheng Zhang, Zheng Xiong, Dongqing Zou, Jimmy S. Ren, Chunle Guo, Chongyi Li
- 391 Bi-Level Optimization for Self-Supervised AI-Generated Face * Detection, Mian Zou, Nan Zhong, Baosheng Yu, Yibing Zhan, Kede Ma
- 392 VisualCloze: A Universal Image Generation Framework via Visual In-Context Learning, Zhong-Yu Li, Ruoyi Du, Juncheng Yan, Le Zhuo, Zhen Li, Peng Gao, Zhanyu Ma, Ming-Ming Cheng
- 393 From Imitation to Innovation: The Emergence of AI's Unique Artistic Styles and the Challenge of Copyright Protection, Zexi Jia, Chuanwei Huang, Yeshuang Zhu, Hongyan Fei, Ying Deng, Zhiqiang Yuan, Jiawei Zhang, Jinchao Zhang, Jie Zhou
- 394 AnyPortal: Zero-Shot Consistent Video Background Replacement, Wenshuo Gao, Xicheng Lan, Shuai Yang
- 395 Neighboring Autoregressive Modeling for Efficient Visual Generation, Yefei He, Yuanyu He, Shaoyuan He, Feng Chen, Hong Zhou, Kaipeng Zhang, Bohan Zhuang
- 396 FastVAR: Linear Visual Autoregressive Modeling via Cached Token Pruning, Hang Guo, Yawei Li, Taolin Zhang, Jiangshan Wang, Tao Dai, Shu-Tao Xia, Luca Benini
- 397 Enhancing Reward Models for High-quality Image Generation: Beyond Text-Image Alignment, Ying Ba, Tianyu Zhang, Yalong Bai, Wenyi Mo, Tao Liang, Bing Su, Ji-Rong Wen
- 398 Tune-Your-Style: Intensity-tunable 3D Style Transfer with Gaussian Splatting, Yian Zhao, Rushi Ye, Ruochong Zheng, Zesen Cheng, Chaoran Feng, Jiashu Yang, Pengchong Qiao, Chang Liu, Jie Chen
- 399 QK-Edit: Revisiting Attention-based Injection in MM-DiT for Image and Video Editing, Tiancheng Shen, Zilong Huang, Xiangtai Li, Zhijie Lin, Jiyang Liu, Yitong Wang, Jiashi Feng, Ming-Hsuan Yang, Jun Hao Liew
- 400 Beyond Isolated Words: Diffusion Brush for Handwritten Text-Line Generation, Gang Dai, Yifan Zhang, Yutao Qin, Qiangya Guo, Shuangping Huang, Shuicheng Yan
- 401 DC-ControlNet: Decoupling Inter- and Intra-Element Conditions in Image Generation with Diffusion Models, Hongji Yang, Wencheng Han, Yucheng Zhou, Jianbing Shen
- 402 BadVideo: Stealthy Backdoor Attack against Text-to-Video Generation, Ruotong Wang, Mingli Zhu, Jiarong Ou, Rui Chen, Xin Tao, Pengfei Wan, Baoyuan Wu
- 403 Any2AnyTryon: Leveraging Adaptive Position Embeddings for Versatile Virtual Clothing Tasks, Hailong Guo, Bohan Zeng, Yiren Song, Wentao Zhang, Jiaming Liu, Chuang Zhang
- 404 FICGen: Frequency-Inspired Contextual Disentanglement for Layout-driven Degraded Image Generation, Wenzhuang Wang, Yifan Zhao, Mingcan Ma, Ming Liu, Zhonglin Jiang, Yong Chen, Jia Li
- 405 An Empirical Study of Autoregressive Pre-training from Videos, Jathushan Rajasegaran, Ilija Radosavovic, Rahul Ravishanker, Yossi Gandelsman, Christoph Feichtenhofer, Jitendra Malik
- 406 Blended Point Cloud Diffusion for Localized Text-guided Shape * Editing, Etai Sella, Noam Atia, Ron Mokady, Hadar Averbuch-Elor
- 407 Training-free Geometric Image Editing on Diffusion Models, Hanshen Zhu, Zhen Zhu, Kaile Zhang, Yiming Gong, Yuliang Liu, Xiang Bai
- 408 Video Color Grading via Look-Up Table Generation, Seunghyun Shin, Dongmin Shin, Jisu Shin, Hae-Gon Jeon, Joon-Young Lee
- 409 VSC: Visual Search Compositional Text-to-Image Diffusion Model, Do Huu Dat, Nam Hyeon-Woo, Po-Yuan Mao, Tae-Hyun Oh

- 410 VideoAuteur: Towards Long Narrative Video Generation, *Junfei Xiao, Feng Cheng, Lu Qi, Liangke Gui, Yang Zhao, Shanchuan Lin, Jiepeng Cen, Zhibei Ma, Alan Yuille, Lu Jiang*
- 411 Fine-Tuning Visual Autogressive Models for Subject-Driven Generation, *Jiwoo Chung, Sangeek Hyun, Hyunjun Kim, Eunseo Koh, MinKyu Lee, Jae-Pil Heo*
- 412 Describe, Don't Dictate: Semantic Image Editing with Natural Language Intent, *En Ci, Shanyan Guan, Yanhao Ge, Yilin Zhang, Wei Li, Zhenyu Zhang, Jian Yang, Ying Tai*
- 413 Frequency-Guided Diffusion for Training-Free Text-Driven Image Translation, *Zheng Gao, Jifei Song, Zhensong Zhang, Jiankang Deng, Ioannis Patras*
- 414 SuperEdit: Rectifying and Facilitating Supervision for Instruction-Based Image Editing, *Ming Li, Xin Gu, Fan Chen, Xiaoying Xing, Longyin Wen, Chen Chen, Sijie Zhu*
- 415 Pretrained Reversible Generation as Unsupervised Visual Representation Learning, *Rongkun Xue, Jinouwen Zhang, Yazhe Niu, Dazhong Shen, Bingqi Ma, Yu Liu, Jing Yang*
- 416 DLF: Extreme Image Compression with Dual-generative Latent Fusion, *Naifu Xue, Zhaoyang Jia, Jiahao Li, Bin Li, Yuan Zhang, Yan Lu*
- 417 REDUCIO! Generating 1K Video within 16 Seconds using Extremely Compressed Motion Latents, *Rui Tian, Qi Dai, Jianmin Bao, Kai Qiu, Yifan Yang, Chong Luo, Zuxuan Wu, Yu-Gang Jiang*
- 418 Tracing Copied Pixels and Regularizing Patch Affinity in Copy Detection, *Yichen Lu, Siwei Nie, Minlong Lu, Xudong Yang, Xiaobo Zhang, Peng Zhang*
- 419 Beyond Brain Decoding: Visual-Semantic Reconstructions to Mental Creation Extension Based on fMRI, *Haodong Jing, Dongyao Jiang, Yongqiang Ma, Haibo Hua, Bo Huang, Nanning Zheng*
- 420 PixTalk: Controlling Photorealistic Image Processing and Editing with Language, *Marcos V. Conde, Zihao Lu, Radu Timofte*
- 421 ADCD-Net: Robust Document Image Forgery Localization via Adaptive DCT Feature and Hierarchical Content Disentanglement, *Kahim Wong, Jicheng Zhou, Haiwei Wu, Yain-Whar Si, Jiantao Zhou*
- 422 Towards Robust Defense against Customization via Protective Perturbation Resistant to Diffusion-based Purification, *Wenkui Yang, Jie Cao, Junxian Duan, Ran He*
- 423 A Unified Framework for Industrial Cel-Animation Colorization with Temporal-Structural Awareness, *Xiaoyi Feng, Tao Huang, Peng Wang, Zizhou Huang, Zhang Haihang, Yuntao Zou, Dagang Li, Kaifeng Zou*
- 424 VQ-SGen: A Vector Quantized Stroke Representation for Creative Sketch Generation, *Jiawei Wang, Zhiming Cui, Changjian Li*
- 425 SEGA: A Stepwise Evolution Paradigm for Content-Aware Layout Generation with Design Prior, *Haoran Wang, Bo Zhao, Jinghui Wang, Hanzhang Wang, Huan Yang, Wei Ji, Hao Liu, Xinyan Xiao*
- 426 RAGD: Regional-Aware Diffusion Model for Text-to-Image Generation, *Zhennan Chen, Yajie Li, Haofan Wang, Zhibo Chen, Zhengkai Jiang, Jun Li, Qian Wang, Jian Yang, Ying Tai*
- 427 Engage for All: Making Ordinary Image Descriptions Appealing Again!, *Yuyan Chen, Yifan Jiang, Li Zhou, Jinghan Cao, Yu Guan, Ming Yang, Qingpei Guo*
- 428 Omegance: A Single Parameter for Various Granularities in Diffusion-Based Synthesis, *Xinyu Hou, Zongsheng Yue, Xiaoming Li, Chen Change Loy*
- 429 Generative Video Bi-flow, *Chen Liu, Tobias Ritschel*
- 430 AV-Link: Temporally-Aligned Diffusion Features for Cross-Modal Audio-Video Generation, *Moayed Haji-Ali, Willi Menapace, Aliaksandr Siarohin, Ivan Skorokhodov, Alper Canberk, Kwot Sin Lee, Vicente Ordóñez, Sergey Tulyakov*
- 431 CHORDS: Diffusion Sampling Accelerator with Multi-core Hierarchical ODE Solvers, *Jiaqi Han, Haotian Ye, Puheng Li, Minkai Xu, James Zou, Stefano Ermon*
- 432 T2I-Copilot: A Training-Free Multi-Agent Text-to-Image System for Enhanced Prompt Interpretation and Interactive Generation, *Chieh-Yun Chen, Min Shi, Gong Zhang, Humphrey Shi*
- 433 PersonalVideo: High ID-Fidelity Video Customization without Dynamic and Semantic Degradation, *Hengjia Li, Haonan Qiu, Shiwei Zhang, Xiang Wang, Yujie Wei, Zekun Li, Yingya Zhang, Boxi Wu, Deng Cai*
- 434 CopyrightShield: Enhancing Diffusion Model Security Against Copyright Infringement Attacks, *Zhixiang Guo, Siyuan Liang, Aishan Liu, Dacheng Tao*
- 435 Unified Video Generation via Next-Set Prediction in Continuous Domain, *Zhanzhou Feng, Qingpei Guo, Xinyu Xiao, Ruihan Xu, Ming Yang, Shiliang Zhang*
- 436 When and Where do Data Poisons Attack Textual Inversion?, *Jeremy Styborski, Mingzhi Lyu, Jiayou Lu, Nupur Kapur, Adams Wai-Kin Kong*
- 437 Mobile Video Diffusion, *Haitam Ben Yahia, Denis Korzhenkov, Ioannis Lelekas, Amir Ghodrati, Amirhossein Habibi*
- 438 LayerLock: Non-collapsing Representation Learning with Progressive Freezing, *Goker Erdogan, Nikhil Parthasarathy, Catalin Ionescu, Drew A. Hudson, Alexander Lerchner, Andrew Zisserman, Mehdi S. M. Sajjadi, Joao Carreira*
- 439 Flow to the Mode: Mode-Seeking Diffusion Autoencoders for State-of-the-Art Image Tokenization, *Kyle Sargent, Kyle Hsu, Justin Johnson, Li Fei-Fei, Jiajun Wu*
- 440 Adaptive Routing of Text-to-Image Generation Requests Between Large Cloud Model and Light-Weight Edge Model, *Zewei Xin, Qinya Li, Chaoyue Niu, Fan Wu, Guihai Chen*
- 441 Exploring Multimodal Diffusion Transformers for Enhanced Prompt-based Image Editing, *Joonghyuk Shin, Alchan Hwang, Yujin Kim, Daneul Kim, Jaesik Park*
- 442 JPEG Processing Neural Operator for Backward-Compatible Coding, *Woo Kyoung Han, Yongjun Lee, Byeonghun Lee, Sang Hyun Park, Sunghoon Im, Kyong Hwan Jin*
- 443 EasyControl: Adding Efficient and Flexible Control for Diffusion Transformer, *Yuxuan Zhang, Yirui Yuan, Yiren Song, Haofan Wang, Jiaming Liu*
- 444 All Parts Matter: A Unified Mask-Free Virtual Try-On Framework, *Chenghu Du, Shengwu Xiong, Yi Rong*
- 445 Function-centric Bayesian Network for Zero-Shot Object Goal Navigation, *Sixian Zhang, Xinyao Yu, Xinhang Song, Yiyao Wang, Shuqiang Jiang*
- 446 Attention to Neural Plagiarism: Diffusion Models Can Plagiarize Your Copyrighted Images!, *Zihang Zou, Boqing Gong, Liqiang Wang*
- 447 Distilling Parallel Gradients for Fast ODE Solvers of Diffusion Models, *Beier Zhu, Ruoyu Wang, Tong Zhao, Hanwang Zhang, Chi Zhang*
- 448 Advancing Text-to-3D Generation with Linearized Lookahead Variational Score Distillation, *Yu Lei, Bingde Liu, Qingsong Xie, Haonan Lu, Zhijie Deng*
- 449 ImageGem: In-the-wild Generative Image Interaction Dataset for Generative Model Personalization, *Yuanhe Guo, Linxi Xie, Zhuoran Chen, Kangrui Yu, Ryan Po, Guandao Yang, Gordon Wetzstein, Hongyi Wen*
- 450 SuMa: A Subspace Mapping Approach for Robust and Effective Concept Erasure in Text-to-Image Diffusion Models, *Kien Nguyen, Anh Tran, Cuong Pham*
- 451 LATINO-PRO: LATent consisTency INverse sOlver with PRompt Optimization, *Alessio Spagnoletti, Jean Prost, Andrés Almansa, Nicolas Papadakis, Marcelo Pereyra*
- 452 EDiT: Efficient Diffusion Transformers with Linear Compressed Attention, *Philipp Becker, Abhinav Mehrotra, Ruchika Chavhan, Malcolm Chadwick, Luca Morreale, Mehdi Noroozi, Alberto Gil C. P. Ramos, Sourav Bhattacharya*
- 453 Hate in Plain Sight: On the Risks of Moderating AI-Generated Hateful Illusions, *Yiting Qu, Ziqing Yang, Yihan Ma, Michael Backes, Savvas Zannettou, Yang Zhang*
- 454 DC-AE 1.5: Accelerating Diffusion Model Convergence with Structured Latent Space, *Junyu Chen, Dongyun Zou, Wenkun He, Junsong Chen, Enze Xie, Song Han, Han Cai*
- 455 Multimodal LLMs as Customized Reward Models for Text-to-Image Generation, *Shijie Zhou, Ruiyi Zhang, Huaisheng Zhu, Branislav Kveton, Yufan Zhou, Jiuxiang Gu, Jian Chen, Changyou Chen*
- 456 MH-LVC: Multi-Hypothesis Temporal Prediction for Learned Conditional Residual Video Coding, *Huu-Tai Phung, Zong-Lin Gao, Yi-Chen Yao, Kuan-Wei Ho, Yi-Hsin Chen, Yu-Hsiang Lin, Alessandro Gnutti, Wen-Hsiao Peng*

- 16:45 - 17:45** **PAMI TC Meeting** (Exhibit Hall III)
19:00 - 20:00 **Reception** (Hilton Lawn)

- Underwater NeRFs with an AUV, *Selena Sun, Elsa McElhinney, Vassilis Alexopoulos, Felicie Hoffman, Lawton Scaling, Kai Song, Rohan Bhowmik, Angelina Krinos, Ryota Sato, Chisa Ogaki, Julian Shultz, Professor Oussama Khatib*
- Scalable Object Detection in Mixed Reality using Incremental Re-training and One-shot 3D Annotation, *Alireza Taheritajar*
- SOAP: Style-Omniscient Animatable Portraits, *Tingting Liao*
- Vision-guided Shared Autonomy for Smarter Prosthetics
Federico Vasile

Thursday, October 23

7:30 - 14:00 **Registration / Badge Pickup** (HCC Main Lobby)7:00 - 17:00 **Nursing Pod** 7:00 - 17:00 **Prayer / Quiet Room / Mother's Room** (326 A)7:30 - 9:00 **Breakfast** (Exhibit Hall II)8:00 - 9:15 **Oral Session 5A: Content Generation**
(Exhibit Hall III)8:00 - 9:15 **Oral Session 5B: Applications and Evaluation**
(Kalakaua Ballroom)9:15 - 10:15 **Coffee Break** (Exhibit Hall II)9:00 - 9:30 **Poster Setup** (Exhibit Hall I)9:30 - 10:30 **Keynote** (Exhibit Hall III)10:45 - 12:45 **Poster Session 5 & Exhibit Hall** (Exhibit Hall I)10:45 - 12:45 **Demos 5** (Exhibit Hall I)11:00 - 13:00 **Lunch** (Exhibit Hall II / Overflow Seating Rooftop and Third Floor)13:00 - 14:15 **Oral Session 6A: Physical Scene Perception**
(Exhibit Hall III)13:00 - 14:15 **Oral Session 6B: Segmentation and Grouping**
(Kalakaua Ballroom)14:00 - 14:30 **Poster Setup** (Exhibit Hall I)14:30 - 16:30 **Poster Session 6 & Exhibit Hall with Coffee Break**
(Exhibit Hall I)14:30 - 16:30 **Demos 6** (Exhibit Hall I)8:00 - 9:15 **Oral Session 5A: Content Generation**
(Exhibit Hall III)

Session Chairs: Lorenzo Baraldi, Yijun Li

 - Award candidate paper ☆ - Outstanding Reviewer

- 1  LaRender: Training-Free Occlusion Control in Image Generation via Latent Rendering, *Xiaohang Zhan, Dingming Liu*
- 2 MikuDance: Animating Character Art with Mixed Motion Dynamics, *Jiaxu Zhang, Xianfang Zeng, Xin Chen, Wei Zuo, Gang Yu, Zhigang Tu*
- 3 Diffusion Transformer meets Multi-level Wavelet Spectrum for Single Image Super-Resolution, *Peng Du, Hui Li, Han Xu, Paul Barom Jeon, Dongwook Lee, Daehyun Ji, Ran Yang, Feng Zhu*
- 4 LOTS of Fashion! Multi-Conditioning for Image Generation via Sketch-Text Pairing, *Federico Girella, Davide Talon, Ziyue Liu, Zanxi Ruan, Yiming Wang, Marco Cristani*
- 5  FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models, *Vladimir Kulikov, Matan Kleiner, Inbar Huberman-Spiegelglas, Tomer Michaeli*
- 6 LayerTracer: Cognitive-Aligned Layered SVG Synthesis via Diffusion Transformer, *Yiren Song, Danze Chen, Mike Zheng Shou*

8:00 - 9:15 **Oral Session 5B: Applications and Evaluation**
(Kalakaua Ballroom)

Session Chairs: Xiatian Zhu, Maria Vakalopoulou

- 1 ROAR: Reducing Inversion Error in Generative Image Watermarking, *Hanyi Wang, Han Fang, Shi-Lin Wang, Ee-Chien Chang*
- 2 Moto: Latent Motion Token as the Bridging Language for Learning Robot Manipulation from Videos, *Yi Chen, Yuying Ge, Weiliang Tang, Yizhuo Li, Yixiao Ge, Mingyu Ding, Ying Shan, Xihui Liu*
- 3 Automated Model Evaluation for Object Detection via Prediction Consistency and Reliability, *Seungju Yoo, Hyuk Kwon, Joong-Won Hwang, Kibok Lee*
- 4 Counting Stacked Objects, *Corentin Dumery, Noa Etté, Aoxiang Fan, Ren Li, Jingyi Xu, Hieu Le, Pascal Fua*
- 5 MIORe & VAR-MIORe: Benchmarks to Push the Boundaries of Restoration, *George Ciobotariu, Zhuyun Zhou, Zongwei Wu, Radu Timofte*

- 6 Soft Local Completeness: Rethinking Completeness in XAI, ☆ *Ziv Weiss Haddad, Oren Barkan, Yehonatan Elisha, Noam Koenigstein*

9:15 - 10:15 **Coffee Break** (Exhibit Hall II)9:00 - 9:30 **Poster Setup** (Exhibit Hall I)9:30 - 10:30 **Keynote** (Exhibit Hall III)
Linda B Smith, Distinguished Professor,
Indiana University, Secrets in the training data: The visual statistics of infants and children's everyday interactions with objects10:45 - 12:45 **Poster Session 5 & Exhibit Hall** (Exhibit Hall I)* - Highlight paper  - Award candidate paper

○ - Oral Paper ☆ - Outstanding Reviewer

- 1 On the Provable Importance of Gradients for Autonomous Language-Assisted Image Clustering, *Bo Peng, Jie Lu, Guangquan Zhang, Zhen Fang*
- 2 Inter2Former: Dynamic Hybrid Attention for Efficient High-Precision Interactive Segmentation, *You Huang, Lichao Chen, Jiayi Ji, Liujuan Cao, Shengchuan Zhang, Rongrong Ji*
- 3 Vector Contrastive Learning For Pixel-Wise Pretraining In Medical Vision, *Yuting He, Shuo Li*
- 4 OphCLIP: Hierarchical Retrieval-Augmented Learning for Ophthalmic Surgical Video-Language Pretraining, *Ming Hu, Kun Yuan, Yaling Shen, Feilong Tang, Xiaohao Xu, Lin Zhou, Wei Li, Ying Chen, Zhongxing Xu, Zelin Peng, Siyuan Yan, Vinkle Srivastav, Diping Song, Tianbin Li, Danli Shi, Jin Ye, Nicolas Padoy, Nassir Navab, Junjun He, Zongyuan Ge*
- 5 ATCTrack: Aligning Target-Context Cues with Dynamic Target States for Robust Vision-Language Tracking, *Xiaokun Feng, Shiyu Hu, Xuchen Li, Dailing Zhang, Meiqi Wu, Jing Zhang, Xiaotang Chen, Kaiqi Huang*
- 6 HiERO: Understanding the Hierarchy of Human Behavior Enhances Reasoning on Egocentric Videos, *Simone Alberto Peirone, Francesca Pistilli, Giuseppe Averta*
- 7 CaptionSmiths: Flexibly Controlling Language Pattern in Image Captioning, *Kuniaki Saito, Donghyun Kim, Kwanyong Park, Atsushi Hashimoto, Yoshitaka Ushiku*
- 8 Multi-modal Segment Anything Model for Camouflaged Scene Segmentation, *Guangyu Ren, Hengyan Liu, Michalis Lazarou, Tania Stathaki*
- 9 OracleFusion: Assisting the Decipherment of Oracle Bone Script with Structurally Constrained Semantic Typography, *Caoshuo Li, Zengmao Ding, Xiaobin Hu, Bang Li, Donghao Luo, AndyPian Wu, Chaoyang Wang, Chengjie Wang, Taisong Jin, Seven Shu, Yunsheng Wu, Yongge Liu, Rongrong Ji*
- 10 MRGen: Segmentation Data Engine For Underrepresented MRI Modalities, *Haoning Wu, Ziheng Zhao, Ya Zhang, Yanfeng Wang, Weidi Xie*
- 11 An Efficient Hybrid Vision Transformer for TinyML Applications, *Fanhong Zeng, Huanan Li, Juntao Guan, Rui Fan, Tong Wu, Xilong Wang, Rui Lai*
- 12 Graph Domain Adaptation with Dual-branch Encoder and Two-level Alignment for Whole Slide Image-based Survival Prediction, *Yuntao Shou, Xiangyong Cao, Peiqiang Yan, Qiao Hui, Qian Zhao, Deyu Meng*
- 13 DeRIS: Decoupling Perception and Cognition for Enhanced Referring Image Segmentation through Loopback Synergy, *Ming Dai, Wenxuan Cheng, Jiang-jiang Liu, Sen Yang, Wenxiao Cai, Yanpeng Sun, Wankou Yang*
- 14 Pretend Benign: A Stealthy Adversarial Attack by Exploiting Vulnerabilities in Cooperative Perception, *Hongwei Lin, Dongyu Pan, Qiming Xia, Hai Wu, Cheng Wang, Siqi Shen, Chenglu Wen*
- 15 Understanding Personal Concept in Open-Vocabulary Semantic Segmentation, *Sunghyun Park, Jungsoo Lee, Shubhankar Borse, Munawar Hayat, Sungha Choi, Kyuwoong Hwang, Fatih Porikli*
- 16 CoralSRT: Revisiting Coral Reef Semantic Segmentation by Feature Rectification via Self-supervised Guidance, *Ziqiang Zheng, Yuk-Kwan Wong, Binh-Son Hua, Jianbo Shi, Sai-Kit Yeung*

- 17 CNS-Bench: Benchmarking Image Classifier Robustness Under Continuous Nuisance Shifts, *Olaf Dünkler, Artur Jesslen, Jiahao Xie, Christian Theobalt, Christian Rupprecht, Adam Kortylewski*
- 18 Visual Test-time Scaling for GUI Agent Grounding, *Tiang Luo, * Lajanugen Logeswaran, Justin Johnson, Honglak Lee*
- 19 Multi-Schema Proximity Network for Composed Image Retrieval, *Jiangming Shi, Xiangbo Yin, Yeyun Chen, Yachao Zhang, Zhizhong Zhang, Yuan Xie, Yanyun Qu*
- 20 ReferDINO: Referring Video Object Segmentation with Visual Grounding Foundations, *Tianming Liang, Kun-Yu Lin, Chaolei Tan, Jianguo Zhang, Wei-Shi Zheng, Jian-Fang Hu*
- 21 GECKO: Gigapixel Vision-Concept Contrastive Pretraining in * Histopathology, *Saarthak Kapse, Pushpak Pati, Srikar Yellapragada, Srijan Das, Rajarsi R. Gupta, Joel Saltz, Dimitris Samaras, Prateek Prasanna*
- 22 Lumina-Image 2.0: A Unified and Efficient Image Generative Framework, *Qi Qin, Le Zhuo, Yi Xin, Ruoyi Du, Zhen Li, Bin Fu, Yiting Lu, Xinyue Li, Dongyang Liu, Xiangyang Zhu, Will Beddow, Erwann Millon, Victor Perez, Wenhai Wang, Yu Qiao, Bo Zhang, Xiaohong Liu, Hongsheng Li, Chang Xu, Peng Gao*
- 23 DiSCO-3D : Discovering and Segmenting Sub-Concepts from Open-vocabulary Queries in NeRF, *Dorian Petit, Steve Bourgeois, Vincent Gay-Bellile, Florian Chabot, Loïc Barthe*
- 24 ESCNet: Edge-Semantic Collaborative Network for Camouflaged Object Detection, *Sheng Ye, Xin Chen, Yan Zhang, Xianming Lin, Liujuan Cao*
- 25 Growing a Twig to Accelerate Large Vision-Language Models, *Zhenwei Shao, Mingyang Wang, Zhou Yu, Wenwen Pan, Yan Yang, Tao Wei, Hongyuan Zhang, Ning Mao, Wei Chen, Jun Yu*
- 26 Test-time Adaptation for Foundation Medical Segmentation * Model Without Parametric Updates, *Kecheng Chen, Xinyu Luo, Tiexin Qin, Jie Liu, Hui Liu, Victor Ho Fun Lee, Hong Yan, Haoliang Li*
- 27 ResQ: A Novel Framework to Implement Residual Neural Networks on Analog Rydberg Atom Quantum Computers, *Nicholas S. DiBrito, Jason Han, Tirthak Patel*
- 28 What's Making That Sound Right Now? Video-centric Audio-Visual Localization, *Hahyeon Choi, Junhoo Lee, Nojun Kwak*
- 29 Task Vector Quantization for Memory-Efficient Model Merging, *Youngeun Kim, Seunghwan Lee, Aecheon Jung, Bogon Ryu, Sungeun Hong*
- 30 M-Net: MRI Brain Tumor Sequential Segmentation Network via Mesh-Cast, *Jiacheng Lu, Hui Ding, Shiyu Zhang, Guoping Huo*
- 31 Bilateral Collaboration with Large Vision-Language Models for Open Vocabulary Human-Object Interaction Detection, *Yupeng Hu, Changxing Ding, Chang Sun, Shaoli Huang, Xiangmin Xu*
- 32 Moment Quantization for Video Temporal Grounding, *Xiaolong Sun, Le Wang, Sanping Zhou, Liushuai Shi, Kun Xia, Mengnan Liu, Yabing Wang, Gang Hua*
- 33 SVTRv2: CTC Beats Encoder-Decoder Models in Scene Text Recognition, *Yongkun Du, Zhenheng Chen, Hongtao Xie, Caiyan Jia, Yu-Gang Jiang*
- 34 ProbMED: A Probabilistic Framework for Medical Multimodal Binding, *Yuan Gao, Sangwook Kim, Jianzhong You, Chris McIntosh*
- 35 VCA: Video Curious Agent for Long Video Understanding, *Zeyuan Yang, Delin Chen, Xueyang Yu, Maohao Shen, Chuang Gan*
- 36 AIM: Adaptive Inference of Multi-Modal LLMs via Token Merging and Pruning, *Yiwu Zhong, Zhuoming Liu, Yin Li, Liwei Wang*
- 37 InstructSeg: Unifying Instructed Visual Segmentation with Multi-modal Large Language Models, *Cong Wei, Yujie Zhong, Haoxian Tan, Yingsen Zeng, Yong Liu, Hongfa Wang, Yujiu Yang*
- 38 ROVI: A VLM-LLM Re-Captioned Dataset for Open-Vocabulary Instance-Grounded Text-to-Image Generation, *Cihang Peng, Qiming Hou, Zhong Ren, Kun Zhou*
- 39 Aligning Moments in Time using Video Queries, *Yogesh Kumar, Uday Agarwal, Manish Gupta, Anand Mishra*
- 40 S*M: Boosting Semi-Supervised Instance Segmentation with SAM, *Heej Yoon, Heeseong Shin, Eunbeen Hong, Hyunwook Choi, Hansang Cho, Daun Jeong, Seungryoung Kim*
- 41 LVAgent: Long Video Understanding by Multi-Round Dynamical Collaboration of MLLM Agents, *Boyu Chen, Zhengrong Yue, Siran Chen, Zikang Wang, Yang Liu, Peng Li, Yali Wang*
- 42 Pruning All-Rounder: Rethinking and Improving Inference Efficiency for Large Vision Language Models, *Wei Suo, Ji Ma, Mengyang Sun, Lin Yuanbo Wu, Peng Wang, Yanning Zhang*
- 43 Structure-aware Semantic Discrepancy and Consistency for 3D Medical Image Self-supervised Learning, *Tan Pan, Zhaorui Tan, Kaiyu Guo, Dongli Xu, Weidi Xu, Chen Jiang, Xin Guo, Yuan Qi, Yuan Cheng*
- 44 Ensemble Foreground Management for Unsupervised Object * Discovery, *Ziling Wu, Armaghan Moemeni, Praminda Caleb-Solly*
- 45 ARGUS: Hallucination and Omission Evaluation in Video-LLMs, *Ruchit Rawal, Reza Shirkavand, Heng Huang, Gowthami Somepalli, Tom Goldstein*
- 46 Feature Purification Matters: Suppressing Outlier Propagation * for Training-Free Open-Vocabulary Semantic Segmentation, *Shuo Jin, Siyue Yu, Bingfeng Zhang, Mingjie Sun, Yi Dong, Jimin Xiao*
- 47 DiffPS: Leveraging Prior Knowledge of Diffusion Model for Person * Search, *Giyeol Kim, Sooyoung Yang, Jihyong Oh, Myungjoo Kang, ☆ Chanho Eom*
- 48 Mind the Gap: Aligning Vision Foundation Models to Image Feature Matching, *Yuhan Liu, Jingwen Fu, Yang Wu, Kangyi Wu, Pengna Li, Jiayi Wu, Sanping Zhou, Jingmin Xin*
- 49 COIN: Confidence Score-Guided Distillation for Annotation-Free Cell Segmentation, *Sanghyun Jo, Seo Jin Lee, Seungwoo Lee, Seohyung Hong, Hyungseok Seo, Kyungsu Kim*
- 50 LeGrad: An Explainability Method for Vision Transformers via Feature Formation Sensitivity, *Walid Bousseham, Angie Boggust, Sofian Chayboubi, Hendrik Strobelt, Hilde Kuehne*
- 51 UniDxMD: Towards Unified Representation for Cross- * Modal Unsupervised Domain Adaptation in 3D Semantic Segmentation, *Zhengyin Liang, Hui Yin, Min Liang, Qianqian Du, Ying Yang, Hua Huang*
- 52 Two Losses, One Goal: Balancing Conflict Gradients for Semi- * supervised Semantic Segmentation, *Rui Sun, Huayu Mai, Wangkai Li, Yujia Chen, Yuan Wang*
- 53 Learn2Synth: Learning Optimal Data Synthesis Using Hypergradients for Brain Image Segmentation, *Xiaoling Hu, Xiangrui Zeng, Oula Puonti, Juan Eugenio Iglesias, Bruce Fischl, Yaël Balbastre*
- 54 Bridging the Gap Between Ideal and Real-world Evaluation: Benchmarking AI-Generated Image Detection in Challenging Scenarios, *Chunxiao Li, Xiaoxiao Wang, Meiling Li, Boming Miao, Peng Sun, Yunjian Zhang, Xiangyang Ji, Yao Zhu*
- 55 ATAS: Any-to-Any Self-Distillation for Enhanced Open-Vocabulary Dense Prediction, *Juan Yeo, Soonwoo Cha, Jiwoo Song, Hyunbin Jin, Taesup Kim*
- 56 VideoOrion: Tokenizing Object Dynamics in Videos, *Yicheng Feng, Yijiang Li, Wanpeng Zhang, Sipeng Zheng, Hao Luo, Zihao Yue, Zongqing Lu*
- 57 Leveraging Debaised Cross-modal Attention Maps and Code-based Reasoning for Zero-shot Referring Expression Comprehension, *Juntao Chen, Wen Shen, Zhihua Wei, Lijun Sun, Hongyun Zhang*
- 58 ReMP-AD: Retrieval-enhanced Multi-modal Prompt Fusion for Few-Shot Industrial Visual Anomaly Detection, *Hongchi Ma, Guanglei Yang, Debin Zhao, Yanli Ji, Wangmeng Zuo*
- 59 Beyond Simple Edits: Composed Video Retrieval with Dense Modifications, *Omkar Thawakar, Dmitry Demidov, Ritesh Thawkar, Rao Muhammad Anwer, Mubarak Shah, Fahad Shahbaz Khan, Salman Khan*
- 60 Optimal Transport for Brain-Image Alignment: Unveiling Redundancy and Synergy in Neural Information Processing, *Yang Xiao, Wang Lu, Jie Ji, Ruimeng Ye, Gen Li, Xiaolong Ma, Bo Hui*
- 61 Representation Shift: Unifying Token Compression with FlashAttention, *Joonmyung Choi, Sanghyeok Lee, Byungoh Ko, Eunseo Kim, Jihyung Kil, Hyunwoo J. Kim*
- 62 Identity-aware Language Gaussian Splatting for Open-vocabulary 3D Semantic Segmentation, *SungMin Jang, Wonjun Kim*
- 63 ZipVL: Accelerating Vision-Language Models through Dynamic Token Sparsity, *Yefei He, Feng Chen, Jing Liu, Wenqi Shao, Hong Zhou, Kaipeng Zhang, Bohan Zhuang*

- 64 ProSAM: Enhancing the Robustness of SAM-based Visual Reference Segmentation with Probabilistic Prompts, *Xiaoqi Wang, Clint Sebastian, Wenbin He, Liu Ren*
- 65 LaCoOT: Layer Collapse through Optimal Transport, *Victor Quéto, Zhu Liao, Nour Hezbri, Fabio Pizzati, Enzo Tartaglione*
- 66 Zero-Shot Compositional Video Learning with Coding Rate Reduction, *Heeseok Jung, Jun-Hyeon Bak, Yujin Jeong, Gyujeun Lee, Jinwoo Ahn, Eun-Sol Kim*
- 67 DictAS: A Framework for Class-Generalizable Few-Shot Anomaly Segmentation via Dictionary Lookup, *Zhen Qu, Xian Tao, Xinyi Gong, ShiChen Qu, Xiaopei Zhang, Xingang Wang, Fei Shen, Zhengtao Zhang, Mukesh Prasad, Guiguang Ding*
- 68 End-to-End Multi-Modal Diffusion Mamba, *Chunhao Lu, Qiang Lu, Meichen Dong, Jake Luo*
- 69 Vid-Group: Temporal Video Grounding Pretraining from Unlabeled Videos in the Wild, *Peijun Bao, Chenqi Kong, Siyuan Yang, Zihao Shao, Xinghao Jiang, Boon Poh Ng, Meng Hwa Er, Alex Kot*
- 70 G2SF: Geometry-Guided Score Fusion for Multimodal Industrial Anomaly Detection, *Chengyu Tao, Xuanming Cao, Juan Du*
- 71 Balancing Conservatism and Aggressiveness: Prototype-Affinity Hybrid Network for Few-Shot Segmentation, *Tianyu Zou, Shengwu Xiong, Ruilin Yao, Yi Rong*
- 72 Fuzzy Contrastive Decoding to Alleviate Object Hallucination in Large Vision-Language Models, *Jieun Kim, Jinmyeong Kim, Yoonji Kim, Sung-Bae Cho*
- 73 BASIC: Boosting Visual Alignment with Intrinsic Refined Embeddings in Multimodal Large Language Models, *Jianting Tang, Yubo Wang, Haoyu Cao, Linli Xu*
- 74 Counting Stacked Objects, *Corentin Dumery, Noa Etté, Aoxiang Fan, Ren Li, Jingyi Xu, Hieu Le, Pascal Fua*
- 75 LaRender: Training-Free Occlusion Control in Image Generation via Latent Rendering, *Xiaohang Zhan, Dingming Liu*
- 76 NETracer: A Topology-Aware Iterative Tracing Approach for Tubular Structure Extraction, *Chao Liu, Yangbo Jiang, Nenggan Zheng*
- 77 Robust Machine Unlearning for Quantized Neural Networks via Adaptive Gradient Reweighting with Similar Labels, *Yujia Tong, Yuze Wang, Jingling Yuan, Chuang Hu*
- 78 Semantic versus Identity: A Divide-and-Conquer Approach towards Adjustable Medical Image De-Identification, *Yuan Tian, Shuo Wang, Rongzhao Zhang, Zijian Chen, Yankai Jiang, Chunyi Li, Xiangyang Zhu, Fang Yan, Qiang Hu, XiaoSong Wang, Guangtao Zhai*
- 79 Towards Video Thinking Test: A Holistic Benchmark for Advanced Video Reasoning and Understanding, *Yuanhan Zhang, Yunice Chew, Yuhao Dong, Aria Leo, Bo Hu, Ziwei Liu*
- 80 MUG: Pseudo Labeling Augmented Audio-Visual Mamba Network for Audio-Visual Video Parsing, *Langyu Wang, Bingke Zhu, Yingying Chen, Yiyuan Zhang, Ming Tang, Jinqiao Wang*
- 81 Cross-View Isolated Sign Language Recognition via View Synthesis and Feature Disentanglement, *Xin Shen, Xinyu Wang, Lei Shen, Kaihao Zhang, Xin Yu*
- 82 Geo4D: Leveraging Video Generators for Geometric 4D Scene Reconstruction, *Zeren Jiang, Chuanxia Zheng, Iro Laina, Diane Larlus, ☆ Andrea Vedaldi*
- 83 Enhancing Prompt Generation with Adaptive Refinement for Camouflaged Object Detection, *Xuehan Chen, Guangyu Ren, Tianhong Dai, Tania Stathaki, Hengyan Liu*
- 84 Factorized Learning for Temporally Grounded Video-Language Models, *Wenzheng Zeng, Difei Gao, Mike Zheng Shou, Hwee Tou Ng*
- 85 Fix-CLIP: Dual-Branch Hierarchical Contrastive Learning via Synthetic Captions for Better Understanding of Long Text, *Bingchao Wang, Zhiwei Ning, Jianyu Ding, Xuanang Gao, Yin Li, Dongsheng Jiang, Jie Yang, Wei Liu*
- 86 G2PDiffusion: Cross-Species Genotype-to-Phenotype Prediction via Evolutionary Diffusion, *Mengdi Liu, Zhangyang Gao, Hong Chang, Stan Z. Li, Shiguang Shan, Xilin Chen*
- 87 Open-ended Hierarchical Streaming Video Understanding with Vision Language Models, *Hyolim Kang, Yunsu Park, Youngbeom Yoo, Yeeun Choi, Seon Joo Kim*
- 88 MOBIUS: Big-to-Mobile Universal Instance Segmentation via Multi-modal Bottleneck Fusion and Calibrated Decoder Pruning, *Mattia Segu, Marta Tintore Gazulla, Yongqin Xian, Luc Van Gool, Federico Tombari*
- 89 ViT-Linearizer: Distilling Quadratic Knowledge into Linear-Time Vision Models, *Guoyizhe Wei, Rama Chellappa*
- 90 LLM-enhanced Action-aware Multi-modal Prompt Tuning for Image-Text Matching, *Mengxiao Tian, Xinxiao Wu, Shuo Yang*
- 91 The Scalability of Simplicity: Empirical Analysis of Vision-Language Learning with a Single Transformer, *Weixian Lei, Jiacong Wang, Haochen Wang, Xiangtai Li, Jun Hao Liew, Jiashi Feng, Zilong Huang*
- 92 Superpowering Open-Vocabulary Object Detectors for X-ray Vision, *Pablo Garcia-Fernandez, Lorenzo Vaquero, Mingxuan Liu, Feng Xue, Daniel Cores, Nicu Sebe, Manuel Mucientes, Elisa Ricci*
- 93 RhythmGaussian: Repurposing Generalizable Gaussian Model For Remote Physiological Measurement, *Hao Lu, Yuting Zhang, Jiaqi Tang, Bowen Fu, Wenhao Ge, Wei Wei, Kaishun Wu, Yingcong Chen*
- 94 Exploiting Domain Properties in Language-Driven Domain Generalization for Semantic Segmentation, *Seogkyu Jeon, Kibeom Hong, Hyeran Byun*
- 95 Keyframe-oriented Vision Token Pruning: Enhancing Efficiency of Large Vision Language Models on Long-Form Video Processing, *Yudong Liu, Jingwei Sun, Yueqian Lin, Jianyi Zhang, Jingyang Zhang, Ming Yin, Qinsi Wang, Hai Li, Yiran Chen*
- 96 Dynamic-VLM: Simple Dynamic Visual Token Compression for VideoLLM, *Han Wang, Yuxiang Nie, Yongjie Ye, Yanjie Wang, Shuai Li, Haiyang Yu, Jinghui Lu, Can Huang*
- 97 MobileViCLIP: An Efficient Video-Text Model for Mobile Devices, *Min Yang, Zihan Jia, Zhilin Dai, Sheng Guo, Limin Wang*
- 98 Stereo Any Video: Temporally Consistent Stereo Matching, *Junpeng Jing, Weixun Luo, Ye Mao, Krystian Mikolajczyk*
- 99 Dynamic-DINO: Fine-Grained Mixture of Experts Tuning for Real-time Open-Vocabulary Object Detection, *Yehao Lu, Minghe Weng, Zekang Xiao, Rui Jiang, Wei Su, Guangcong Zheng, Ping Lu, Xi Li*
- 100 Beyond Text-Visual Attention: Exploiting Visual Cues for Effective Token Pruning in VLMs, *Qizhe Zhang, Aosong Cheng, Ming Lu, Renrui Zhang, Zhiyong Zhuo, Jiajun Cao, Shaobo Guo, Qi She, Shanghang Zhang*
- 101 Wave-MambaAD: Wavelet-driven State Space Model for Multi-class Unsupervised Anomaly Detection, *Qiao Zhang, Mingwen Shao, Xinyuan Chen, Xiang Lv, Kai Xu*
- 102 MaTVLM: Hybrid Mamba-Transformer for Efficient Vision-Language Modeling, *Yingyue Li, Bencheng Liao, Wenyu Liu, Xinggang Wang*
- 103 Does Your Vision-Language Model Get Lost in the Long Video Sampling Dilemma?, *Tianyuan Qu, Longxiang Tang, Bohao Peng, Senqiao Yang, Bei Yu, Jiaya Jia*
- 104 Visual Textualization for Image Prompted Object Detection, *Yongjian Wu, Yang Zhou, Jiya Saiyin, Bingzheng Wei, Yan Xu*
- 105 ASGS: Single-Domain Generalizable Open-Set Object Detection via Adaptive Subgraph Searching, *Yuxuan Yuan, Luyao Tang, Yixin Chen, Chaoqi Chen, Yue Huang, Xinghao Ding*
- 106 UniConvNet: Expanding Effective Receptive Field while Maintaining Asymptotically Gaussian Distribution for ConvNets of Any Scale, *Yuhao Wang, Wei Xi*
- 107 On the Recovery of Cameras from Fundamental Matrices, *☆ Rakshith Madhavan, Federica Arrigoni*
- 108 SiM3D: Single-instance Multiview Multimodal and Multisetup 3D Anomaly Detection Benchmark, *Alex Costanzino, Pierluigi Zama Ramirez, Luigi Lella, Matteo Ragaglia, Alessandro Oliva, Giuseppe Lisanti, Luigi Di Stefano*
- 109 ReME: A Data-Centric Framework for Training-Free Open-Vocabulary Segmentation, *Xiwei Xuan, Ziquan Deng, Kwan-Liu Ma*
- 110 MultiverSeg: Scalable Interactive Segmentation of Biomedical Imaging Datasets with In-Context Guidance, *Hallee E. Wong, Jose Javier Gonzalez Ortiz, John Guttag, Adrian V. Dalca*
- 111 The Devil is in the Spurious Correlations: Boosting Moment Retrieval with Dynamic Learning, *Xinyang Zhou, Fanyue Wei, Lixin Duan, Angela Yao, Wen Li*
- 112 CABLD: Contrast-Agnostic Brain Landmark Detection with Consistency-Based Regularization, *Soorena Salari, Arash Harirpoush,*

- Hassan Rivaz, Yiming Xiao
- 113 CLIPSym: Delving into Symmetry Detection with CLIP, *Tinghan Yang, Md Ashiqur Rahman, Raymond A. Yeh*
- 114 EVEv2: Improved Baselines for Encoder-Free Vision-Language Models, *Haiwen Diao, Xiaotong Li, Yufeng Cui, Yueze Wang, Haoge Deng, Ting Pan, Wenxuan Wang, Huchuan Lu, Xinlong Wang*
- 115 Describe, Adapt and Combine: Empowering CLIP Encoders for Open-set 3D Object Retrieval, *Zhichuan Wang, Yang Zhou, Zhe Liu, Rui Yu, Song Bai, Yulong Wang, Xinwei He, Xiang Bai*
- 116 Robustifying Zero-Shot Vision Language Models by Subspaces Alignment, *Junhao Dong, Piotr Koniusz, Liaoyuan Feng, Yifei Zhang, Hao Zhu, Weiming Liu, Xinghua Qu, Yew-Soon Ong*
- 117 Leveraging the Power of MLLMs for Gloss-Free Sign Language Translation, *Jungeun Kim, Hyeonwoo Jeon, Jongseong Bae, Ha Young Kim*
- 118 Flash-VStream: Efficient Real-Time Understanding for Long Video Streams, *Haoji Zhang, Yiqin Wang, Yansong Tang, Yong Liu, Jiashi Feng, Xiaojie Jin*
- 119 V2PE: Improving Multimodal Long-Context Capability of Vision-Language Models with Variable Visual Position Encoding, *Junqi Ge, Ziyi Chen, Jintao Lin, Jinguo Zhu, Xihui Liu, Jifeng Dai, Xizhou Zhu*
- 120 OVG-HQ: Online Video Grounding with Hybrid-modal Queries, *Runhao Zeng, Jiaqi Mao, Minghao Lai, Minh Hieu Phan, Yanjie Dong, Wei Wang, Qi Chen, Xiping Hu*
- 121 Enhancing Zero-shot Object Counting via Text-guided Local Ranking and Number-evoked Global Attention, *Shiwei Zhang, Qi Zhou, Wei Ke*
- 122 SegAnyPET: Universal Promptable Segmentation from Positron Emission Tomography Images, *Yichi Zhang, Le Xue, Wenbo Zhang, Lanlan Li, Yuchen Liu, Chen Jiang, Yuan Cheng, Yuan Qi*
- 123 Multi-View Slot Attention Using Paraphrased Texts for Face Anti-Spoofing, *Jeongmin Yu, Susang Kim, Kisu Lee, Taekyoung Kwon, Won-Yong Shin, Ha Young Kim*
- 124 4D-Bench: Benchmarking Multi-modal Large Language Models for 4D Object Understanding, *Wenxuan Zhu, Bing Li, Cheng Zheng, Jinjie Mai, Jun Chen, Letian Jiang, Abdullah Hamdi, Sara Rojas Martinez, Chia-Wen Lin, Mohamed Elhoseiny, Bernard Ghanem*
- 125 Uncertainty-Driven Expert Control: Enhancing the Reliability of Medical Vision-Language Models, *Xiao Liang, Di Wang, Zhicheng Jiao, Ronghan Li, Pengfei Yang, Quan Wang, Tat-Seng Chua*
- 126 Probabilistic Prototype Calibration of Vision-language Models for Generalized Few-shot Semantic Segmentation, *Jie Liu, Jiayi Shen, Pan Zhou, Jan-Jakob Sonke, Efstratios Gavves*
- 127 Reducing Unimodal Bias in Multi-Modal Semantic Segmentation with Multi-Scale Functional Entropy Regularization, *Xu Zheng, Yuanhuiyi Lyu, Lutao Jiang, Danda Pani Paudel, Luc Van Gool, Xuming Hu*
- 128 OuroMamba: A Data-Free Quantization Framework for Vision Mamba, *Akshat Ramachandran, Mingyu Lee, Huan Xu, Souvik Kundu, Tushar Krishna*
- 129 MixA: A Mixed Attention approach with Stable Lightweight Linear Attention to enhance Efficiency of Vision Transformers at the Edge, *Sabbir Ahmed, Jingtao Li, Weiming Zhuang, Chen Chen, Lingjuan Lyu*
- 130 Vamba: Understanding Hour-Long Videos with Hybrid Mamba-Transformers, *Weiming Ren, Wentao Ma, Huan Yang, Cong Wei, Ge Zhang, Wenhui Chen*
- 131 SAMora: Enhancing SAM through Hierarchical Self-Supervised Pre-Training for Medical Images, *Shuhang Chen, Hangjie Yuan, Pengwei Liu, Hanxue Gu, Tao Feng, Dong Ni*
- 132 FE-CLIP: Frequency Enhanced CLIP Model for Zero-Shot Anomaly Detection and Segmentation, *Tao Gong, Qi Chu, Bin Liu, Wei Zhou, Nenghai Yu*
- 133 Referring Expression Comprehension for Small Objects, *Kanoko Goto, Takumi Hirose, Mahiro Ukai, Shuhei Kurita, Nakamasa Inoue*
- 134 AID: Adapting Image2Video Diffusion Models for Instruction-guided Video Prediction, *Zhen Xing, Qi Dai, Zejia Weng, Zuxuan Wu, Yu-Gang Jiang*
- 135 Principles of Visual Tokens for Efficient Video Understanding, *Xinyue Hao, Gen Li, Shreyank N Gowda, Robert B. Fisher, Jonathan Huang, Anurag Arnab, Laura Sevilla-Lara*
- 136 CutS3D: Cutting Semantics in 3D for 2D Unsupervised Instance Segmentation, *Leon Sick, Dominik Engel, Sebastian Hartwig, Pedro Hermosilla, Timo Ropinski*
- 137 Learning Robust Stereo Matching in the Wild with Selective Mixture-of-Experts, *Yun Wang, Longguang Wang, Chenghao Zhang, Yongjian Zhang, Zhanjie Zhang, Ao Ma, Chenyou Fan, Tin Lun Lam, Junjie Hu*
- 138 Text-guided Visual Prompt DINO for Generic Segmentation, *Yuchen Guan, Chong Sun, Canmiao Fu, Zhipeng Huang, Chun Yuan, Chen Li*
- 139 ETVA: Evaluation of Text-to-Video Alignment via Fine-grained Question Generation and Answering, *Kaisi Guan, Zhengfeng Lai, Yuchong Sun, Peng Zhang, Wei Liu, Kieran Liu, Meng Cao, Ruihua Song*
- 140 GEMeX: A Large-Scale, Groundable, and Explainable Medical VQA Benchmark for Chest X-ray Diagnosis, *Bo Liu, Ke Zou, Li-Ming Zhan, Zexin Lu, Xiaoyu Dong, Yidi Chen, Chengqiang Xie, Jiannong Cao, Xiao-Ming Wu, Huazhu Fu*
- 141 Bias-Resilient Weakly Supervised Semantic Segmentation Using Normalizing Flows, *Xianglin Qiu, Xiaoyang Wang, Zhen Zhang, Jimin Xiao*
- 142 Multidimensional Byte Pair Encoding: Shortened Sequences for Improved Visual Data Generation, *Tim Elsner, Paula Usinger, Julius Nehring-Wirxel, Gregor Kobsik, Victor Czech, Yanjiang He, Isaak Lim, Leif Kobbelt*
- 143 MA-CIR: A Multimodal Arithmetic Benchmark for Composed Image Retrieval, *Jaeseok Byun, Young Kyun Jang, Seokhyeon Jeong, Donghyun Kim, Taesup Moon*
- 144 Cracking Instance Jigsaw Puzzles: An Alternative to Multiple Instance Learning for Whole Slide Image Analysis, *Xiwen Chen, Peijie Qiu, Wenhui Zhu, Hao Wang, Huayu Li, Xuanzhao Dong, Xiaotong Sun, Xiaobing Yu, Yalin Wang, Abolfazl Razi, Aristeidis Sotiras*
- 145 STDDNet: Harnessing Mamba for Video Polyp Segmentation via Spatial-aligned Temporal Modeling and Discriminative Dynamic Representation Learning, *Guilian Chen, Huisi Wu, Jing Qin*
- 146 Latent Expression Generation for Referring Image Segmentation and Grounding, *Seonghoon Yu, Joonbeom Hong, Joonseok Lee, Jeany Son*
- 147 Skip-Vision: Efficient and Scalable Acceleration of Vision-Language Models via Adaptive Token Skipping, *Weili Zeng, Ziyuan Huang, Kaixiang Ji, Yichao Yan*
- 148 LV-MAE: Learning Long Video Representations through Masked-Embedding Autoencoders, *Ilan Naiman, Emanuel Ben-Baruch, Oron Anschel, Alon Shoshan, Igor Kviatkovsky, Manoj Aggarwal, Gerard Medioni*
- 149 CogCM: Cognition-Inspired Contextual Modeling for Audio-Visual Speech Enhancement, *Feixiang Wang, Shuang Yang, Shiguang Shan, Xilin Chen*
- 150 Salvaging the Overlooked: Leveraging Class-Aware Contrastive Learning for Multi-Class Anomaly Detection, *Lei Fan, Junjie Huang, Donglin Di, Anyang Su, Tianyou Song, Maurice Pagnucco, Yang Song*
- 151 Adapting In-Domain Few-Shot Segmentation to New Domains without Source Domain Retraining, *Qi Fan, Kaiqi Liu, Nian Liu, Hisham Cholakkal, Rao Muhammad Anwer, Wenbin Li, Yang Gao*
- 152 OmniDiff: A Comprehensive Benchmark for Fine-grained Image Difference Captioning, *Yuan Liu, Saihui Hou, Saijie Hou, Jiabao Du, Shibe Meng, Yongzhen Huang*
- 153 Adaptive Learning of High-Value Regions for Semi-Supervised Medical Image Segmentation, *Tao Lei, Ziyao Yang, Xingwu Wang, Yi Wang, Xuan Wang, Feiman Sun, Asoke K. Nandi*
- 154 COME: Dual Structure-Semantic Learning with Collaborative MoE for Universal Lesion Detection Across Heterogeneous Ultrasound Datasets, *Lingyu Chen, Yawen Zeng, Yue Wang, Peng Wan, Guochen Ning, Hongen Liao, Daoqiang Zhang, Fang Chen*
- 155 MIORe & VAR-MIORe: Benchmarks to Push the Boundaries of Restoration, *George Ciubotariu, Zhuyun Zhou, Zongwei Wu, Radu Timofte*
- 156 MikuDance: Animating Character Art with Mixed Motion Dynamics, *Jiaxu Zhang, Xianfang Zeng, Xin Chen, Wei Zuo, Gang Yu, Zhigang Tu*
- 157 FLOSS: Free Lunch in Open-vocabulary Semantic Segmentation, *Yasser Benigim, AMohammad Fahes, Tuan-Hung Vu, Andrei Bursuc, Raoul de Charette*

- 158 GaussianReg: Rapid 2D/3D Registration for Emergency Surgery via Explicit 3D Modeling with Gaussian Primitives, *Weihaio Yu, Xiaoqing Guo, Xinyu Liu, Yifan Liu, Hao Zheng, Yawen Huang, Yixuan Yuan*
- 159 METEOR: Multi-Encoder Collaborative Token Pruning for Efficient Vision Language Models, *Yuchen Liu, Yaoming Wang, Bowen Shi, Xiaopeng Zhang, Wenrui Dai, Chenglin Li, Hongkai Xiong, Qi Tian*
- 160 Rectifying Magnitude Neglect in Linear Attention, *Qihang Fan, * Huaibo Huang, Yang Ai, Ran He*
- 161 Sparse-Dense Side-Tuner for efficient Video Temporal Grounding, *David Pujol-Perich, Sergio Escalera, Albert Clapés*
- 162 MedSegFactory: Text-Guided Generation of Medical Image-Mask Pairs, *Jiawei Mao, Yuhang Wang, Yucheng Tang, Daguang Xu, Kang Wang, Yang Yang, Zongwei Zhou, Yuyin Zhou*
- 163 CountSE: Soft Exemplar Open-set Object Counting, *Shuai Liu, * Peng Zhang, Shiwei Zhang, Wei Ke*
- 164 Constructing Ophthalmic MLLM for Positioning-diagnosis Collaboration Through Clinical Cognitive Chain Reasoning, *Xinyao Liu, Diping Song*
- 165 Unified Open-World Segmentation with Multi-Modal Prompts, *Yang Liu, Yufei Yin, Chenchen Jing, Muzhi Zhu, Hao Chen, Yuling Xi, Bo Feng, Hao Wang, Shiyu Li, Chunhua Shen*
- 166 DecAD: Decoupling Anomalies in Latent Space for Multi-Class Unsupervised Anomaly Detection, *Xiaolei Wang, Xiaoyang Wang, Huihui Bai, Eng Gee Lim, Jimin Xiao*
- 167 Few-Shot Pattern Detection via Template Matching and Regression, ** Eunchan Jo, Dahyun Kang, Sanghyun Kim, Yunseon Choi, Minsu Cho*
- 168 Hierarchical Event Memory for Accurate and Low-latency Online Video Temporal Grounding, *Minghang Zheng, Yuxin Peng, Benyuan Sun, Yi Yang, Yang Liu*
- 169 UKBOB: One Billion MRI Labeled Masks for Generalizable 3D Medical Image Segmentation, *Emmanuelle Bourigault, Amir Jamaludin, Abdullah Hamdi*
- 170 ILLUME: Illuminating Your LLMs to See, Draw, and Self-Enhance, *Chunwei Wang, Guansong Lu, Junwei Yang, Runhui Huang, Jianhua Han, Lu Hou, Wei Zhang, Hang Xu*
- 171 Learning Yourself: Class-Incremental Semantic Segmentation with Language-Inspired Bootstrapped Disentanglement, *Ruitao Wu, Yifan Zhao, Jia Li*
- 172 Hallucinatory Image Tokens: A Training-free EAZY Approach to Detecting and Mitigating Object Hallucinations in LVLMS, *Liwei Che, Tony Qingze Liu, Jing Jia, Weiyei Qin, Ruixiang Tang, Vladimir Pavlovic*
- 173 Class Token as Proxy: Optimal Transport-assisted Proxy Learning for Weakly Supervised Semantic Segmentation, *Jian Wang, Tianhong Dai, Bingfeng Zhang, Siyue Yu, Eng Gee Lim, Jimin Xiao*
- 174 VRBench: A Benchmark for Multi-Step Reasoning in Long Narrative Videos, *Jiashuo Yu, Yue Wu, Meng Chu, Zhifei Ren, Zizheng Huang, Pei Chu, Ruijie Zhang, Yanan He, Qirui Li, Songze Li, Zhenxiang Li, Zhongying Tu, Conghui He, Yu Qiao, Yali Wang, Yi Wang, Limin Wang*
- 175 Referring to Any Person, *Qing Jiang, Lin Wu, Zhaoyang Zeng, Tianhe Ren, Yuda Xiong, Yihao Chen, Liu Qin, Lei Zhang*
- 176 Aligning Information Capacity Between Vision and Language via Dense-to-Sparse Feature Distillation for Image-Text Matching, *Yang Liu, Wentao Feng, Zhuoyao Liu, Shudong Huang, Jiancheng Lv*
- 177 RA-BUSeg: Relation-aware Semi-supervised Breast Ultrasound Image Segmentation via Adjacent Propagation and Cross-layer Alignment, *Wanting Zhang, Zhenhui Ding, Guilian Chen, Huisi Wu, Jing Qin*
- 178 ExCap3D: Expressive 3D Scene Understanding via Object Captioning with Varying Detail, *Chandan Yeshwanth, David Rozenberszki, Angela Dai*
- 179 DisCo: Towards Distinct and Coherent Visual Encapsulation in Video MLLMs, *Jiahe Zhao, Rongkun Zheng, Yi Wang, Helin Wang, Hengshuang Zhao*
- 180 Neuroverse3D: Developing In-Context Learning Universal Model for Neuroimaging in 3D, *Jiesi Hu, Hanyang Peng, Yanwu Yang, Xutao Guo, Yang Shang, Pengcheng Shi, Chenfei Ye, Ting Ma*
- 181 CT-ScanGaze: A Dataset and Baselines for 3D Volumetric Scanpath Modeling, *Trong Thang Pham, Akash Awasthi, Saba Khan, Esteban Duran Marti, Tien-Phat Nguyen, Khoa Vo, Minh Tran, Son Nguyen, Cuong Tran, Yuki Ikebe, Anh Totti Nguyen, Anh Nguyen, Zhigang Deng, Carol C. Wu, Hien Nguyen, Ngan Le*
- 182 CC-OCR: A Comprehensive and Challenging OCR Benchmark for Evaluating Large Multimodal Models in Literacy, *Zhibo Yang, Jun Tang, Zhaohai Li, Pengfei Wang, Jianqiang Wan, Humen Zhong, Xuejing Liu, Mingkun Yang, Peng Wang, Shuai Bai, Lianwen Jin, Junyang Lin*
- 183 Exploring Probabilistic Modeling Beyond Domain Generalization for Semantic Segmentation, *I-Hsiang Chen, Hua-En Chang, Wei-Ting Chen, Jenq-Neng Hwang, Sy-Yen Kuo*
- 184 Describe Anything: Detailed Localized Image and Video Captioning, *Long Lian, Yifan Ding, Yunhao Ge, Sifei Liu, Hanzi Mao, Boyi Li, Marco Pavone, Ming-Yu Liu, Trevor Darrell, Adam Yala, Yin Cui*
- 185 ViSpeak: Visual Instruction Feedback in Streaming Videos, ** Shenghao Fu, Qize Yang, Yuan-Ming Li, Yi-Xing Peng, Kun-Yu Lin, Xihan Wei, Jian-Fang Hu, Xiaohua Xie, Wei-Shi Zheng*
- 186 Prototypes are Balanced Units for Efficient and Effective Partially Relevant Video Retrieval, *WonJun Moon, Cheol-Ho Cho, Woojin Jun, Taeh Kim, Inwoong Lee, Dongyoon Wee, Minho Shim, Jae-Pil Heo*
- 187 SciVid: Cross-Domain Evaluation of Video Models in Scientific Applications, *Yana Hasson, Pauline Luc, Liliane Momeni, Maks Ovsjanikov, Guillaume Le Moing, Alina Kuznetsova, Ira Ktena, Jennifer J. Sun, Skanda Koppala, Dilara Gokay, Joseph Heyward, Etienne Pot, Andrew Zisserman*
- 188 VideoAds for Fast-Paced Video Understanding, *Zheyuan Zhang, Wanying Dou, Linkai Peng, Hongyi Pan, Ulas Bagci, Boqing Gong*
- 189 Auto-Controlled Image Perception in MLLMs via Visual Perception Tokens, *Runpeng Yu, Xinyin Ma, Xinchao Wang*
- 190 Keep Your Friends Close, and Your Enemies Farther: Distance-aware Voxel-wise Contrastive Learning for Semi-supervised Multi-organ Segmentation, *Haochen Zhao, Jianwei Niu, Xuefeng Liu, Xiaozheng Xie, Li Kuang, Haotian Yang, Bin Dai, Hui Meng, Yong Wang*
- 191 SALAD -- Semantics-Aware Logical Anomaly Detection, *Matic Fučka, Vitjan Zavrtanik, Danijel Skočaj*
- 192 Refer to Any Segmentation Mask Group With Vision-Language Prompts, *Shengcao Cao, Zijun Wei, Jason Kuen, Kangning Liu, Lingzhi Zhang, Jiuxiang Gu, HyunJoon Jung, Liang-Yan Gui, Yu-Xiong Wang*
- 193 From Panels to Prose: Generating Literary Narratives from Comics, *Ragav Sachdeva, Andrew Zisserman*
- 194 When Confidence Fails: Revisiting Pseudo-Label Selection in Semi-supervised Semantic Segmentation, ** Pan Liu, Jinshi Liu*
- 195 MSQ: Memory-Efficient Bit Sparsification Quantization, *Seokho Han, Seoyeon Yoon, Jinhee Kim, Dongwei Wang, Kang Eun Jeon, Huanrui Yang, Jong Hwan Ko*
- 196 Towards a Universal 3D Medical Multi-modality Generalization via Learning Personalized Invariant Representation, *Zhaorui Tan, Xi Yang, Tan Pan, Tianyi Liu, Chen Jiang, Xin Guo, Qifeng Wang, Anh Nguyen, Yuan Qi, Kaizhu Huang, Yuan Cheng*
- 197 Improving SAM for Camouflaged Object Detection via Dual Stream Adapters, ** Jiaming Liu, Linghe Kong, Guihai Chen*
- 198 Triad: Empowering LMM-based Anomaly Detection with Expert-guided Region-of-Interest Tokenizer and Manufacturing Process, *Yuanze Li, Shihao Yuan, Haolin Wang, Qizhang Li, Ming Liu, Chen Xu, Guangming Shi, Wangmeng Zuo*
- 199 Unveiling the Invisible: Reasoning Complex Occlusions Amodally with AURA, *Zhixuan Li, Hyunse Yoon, Sanghoon Lee, Weisi Lin*
- 200 Bridging the Gap between Brain and Machine in Interpreting Visual Semantics: Towards Self-adaptive Brain-to-Text Decoding, *Jiaxuan Chen, Yu Qi, Yueming Wang, Gang Pan*
- 201 MobileIE: An Extremely Lightweight and Effective ConvNet for Real-Time Image Enhancement on Mobile Devices, *Hailong Yan, Ao Li, Xiangtao Zhang, Zhe Liu, Zenglin Shi, Ce Zhu, Le Zhang*
- 202 DisTime: Distribution-based Time Representation for Video Large Language Models, *Yingsen Zeng, Zepeng Huang, Yujie Zhong, Chengjian Feng, Jie Hu, Lin Ma, Yang Liu*
- 203 SmolDocling: An ultra-compact vision-language model for end-to-end multi-modal document conversion, *Ahmed Nassar, Matteo Omenetti, Maksym Lysak, Nikolaos Livathinos, Christoph Auer, Lucas Morin, Rafael Teixeira de Lima, Yusik Kim, A. Said Gurbuz, Michele Dolfi, Peter W. J. Staar*
- 204 WeaveSeg: Iterative Contrast-weaving and Spectral Feature-refining for Nuclei Instance Segmentation, ** Jiajia Li, Huisi Wu, Jing Qin*
- 205 How Can Objects Help Video-Language Understanding?, ** Zitian Tang, Shijie Wang, Junho Cho, Jaewook Yoo, Chen Sun*

- 206 Everything is a Video: Unifying Modalities through Next-Frame Prediction, *G. Thomas Hudson, Dean Slack, Thomas Winterbottom, Jamie Sterling, Chenghao Xiao, Junjie Shentu, Noura Al Moubayed*
- 207 LLaVA-SP: Enhancing Visual Representation with Visual Spatial Tokens for MLLMs, *Haoran Lou, Chunxiao Fan, Ziyang Liu, Yuexin Wu, Xinliang Wang*
- 208 Talking to DINO: Bridging Self-Supervised Vision Backbones with Language for Open-Vocabulary Segmentation, *Luca Barsellotti, Lorenzo Bianchi, Nicola Messina, Fabio Carrara, Marcella Cornia, Lorenzo Baraldi, Fabrizio Falchi, Rita Cucchiara*
- 209 CARIM: Caption-Based Autonomous Driving Scene Retrieval via Inclusive Text Matching, *Minjoo Ki, Daejung Kim, Kisung Kim, Seon Joo Kim, Jinhan Lee*
- 210 Beyond Training: Dynamic Token Merging for Zero-Shot Video Understanding, *Yiming Zhang, Zhuokai Zhao, Zhaorun Chen, Zenghui Ding, Xianjun Yang, Yining Sun*
- 211 Q-Frame: Query-aware Frame Selection and Multi-Resolution Adaptation for Video-LLMs, *Shaojie Zhang, Jiahui Yang, Jianqin Yin, Zhenbo Luo, Jian Luan*
- 212 Normal and Abnormal Pathology Knowledge-Augmented Vision-Language Model for Anomaly Detection in Pathology Images, *Jinsol Song, Jiamu Wang, Anh Tien Nguyen, Keunho Byeon, Sangjeong Ahn, Sung Hak Lee, Jin Tae Kwak*
- 213 Modeling Saliency Dataset Bias, *Matthias Kümmerer, Harneet Singh Khanuja, Matthias Bethge*
- 214 ZeroKey: Point-Level Reasoning and Zero-Shot 3D Keypoint Detection from Large Language Models, *Bingchen Gong, Diego Gomez, Abdullah Hamdi, Abdelrahman Eldesokey, Ahmed Abdelreheem, Peter Wonka, Maks Ovsjanikov*
- 215 HyPiDecoder: Hybrid Pixel Decoder for Efficient Segmentation and Detection, *Fengzhe Zhou, Humphrey Shi*
- 216 Benefit From Seen: Enhancing Open-Vocabulary Object Detection by Bridging Visual and Textual Co-Occurrence Knowledge, *Yanqi Li, Jianwei Niu, Tao Ren*
- 217 Quantifying and Narrowing the Unknown: Interactive Text-to-Video Retrieval via Uncertainty Minimization, *Bingqing Zhang, Zhuo Cao, Heming Du, Yang Li, Xue Li, Jiajun Liu, Sen Wang*
- 218 Beyond Single Images: Retrieval Self-Augmented Unsupervised Camouflaged Object Detection, *Ji Du, Xin Wang, Fangwei Hao, Mingyang Yu, Chunyuan Chen, Jiesheng Wu, Bin Wang, Jing Xu, Ping Li*
- 219 MixA-Q: Revisiting Activation Sparsity for Vision Transformers from a Mixed-Precision Quantization Perspective, *Weitian Wang, Rai Shubham, Cecilia De La Parra, Akash Kumar*
- 220 Advancing Visual Large Language Model for Multi-granular Versatile Perception, *Wentao Xiang, Haoxian Tan, Yujie Zhong, Cong Wei, Dengjie Li, Yujiu Yang*
- 221 Controllable Latent Space Augmentation for Digital Pathology, *Sofïène Boutaj, Marin Scalbert, Pierre Marza, Florent Couzinie-Devy, Maria Vakalopoulou, Stergios Christodoulidis*
- 222 PS3: A Multimodal Transformer Integrating Pathology Reports with Histology Images and Biological Pathways for Cancer Survival Prediction, *Manahil Raza, Ayesha Azam, Talha Kaiser, Nasir Rajpoot*
- 223 MIEB: Massive Image Embedding Benchmark, *Chenghao Xiao, Isaac Chung, Imene Kerboua, Jamie Stirling, Xin Zhang, Márton Kardos, Roman Kolmogorov, Noura Al Moubayed, Kenneth Enevoldsen, Niklas Muennighoff*
- 224 Temperature in Cosine-based Softmax Loss, *Takumi Kobayashi*
- 225 Interpretable point cloud classification using multiple instance learning, *Matt De Vries, Reed Naidoo, Olga Fourkioti, Lucas G. Dent, Nathan Curry, Chris Dunsby, Chris Bakal*
- 226 Zero-Shot Composed Image Retrieval via Dual-Stream Instruction-Aware Distillation, *Wenliang Zhong, Rob Barton, Weizhi An, Feng Jiang, Hehuan Ma, Yuzhi Guo, Abhishek Dan, Shioulin Sam, Karim Bouyarmane, Junzhou Huang*
- 227 DeFSS: Image-to-Mask Denoising Learning for Few-shot Segmentation, *Zishu Qin, Junhao Xu, Weifeng Ge*
- 228 Fine-grained Abnormality Prompt Learning for Zero-shot Anomaly Detection, *Jiawen Zhu, Yew-Soon Ong, Chunhua Shen, Guansong Pang*
- 229 Moto: Latent Motion Token as the Bridging Language for Learning Robot Manipulation from Videos, *Yi Chen, Yuying Ge, Weiliang Tang, Yizhuo Li, Yixiao Ge, Mingyu Ding, Ying Shan, Xihui Liu*
- 230 ROAR: Reducing Inversion Error in Generative Image Watermarking, *Hanyi Wang, Han Fang, Shi-Lin Wang, Ee-Chien Chang*
- 231 Efficient Fine-Tuning of Large Models via Nested Low-Rank Adaptation, *Lujun Li, Cheng Lin, Dezhi Li, You-Liang Huang, Wei Li, Tianyu Wu, Jie Zou, Wei Xue, Sirui Han, Yike Guo*
- 232 Bidirectional Likelihood Estimation with Multi-Modal Large Language Models for Text-Video Retrieval, *Dohwan Ko, Ji Soo Lee, Minhyuk Choi, Zihang Meng, Hyunwoo J. Kim*
- 233 Object-centric Video Question Answering with Visual Grounding and Referring, *Haochen Wang, Qirui Chen, Cilin Yan, Jiayin Cai, Xiaolong Jiang, Yao Hu, Weidi Xie, Stratis Gavves*
- 234 Music Grounding by Short Video, *Zijie Xin, Minquan Wang, Jingyu Liu, Quan Chen, Ye Ma, Peng Jiang, Xirong Li*
- 235 LLaFEA: Frame-Event Complementary Fusion for Fine-Grained Spatiotemporal Understanding in LMMs, *Hanyu Zhou, Gim Hee Lee*
- 236 ResidualViT for Efficient Temporally Dense Video Encoding, *Mattia Soldan, Fabian Caba Heilbron, Bernard Ghanem, Josef Sivic, Bryan Russell*
- 237 Alleviating Textual Reliance in Medical Language-guided Segmentation via Prototype-driven Semantic Approximation, *Shuchang Ye, Usman Naseem, Mingyuan Meng, Jinman Kim*
- 238 Controllable-LPMoE: Adapting to Challenging Object Segmentation via Dynamic Local Priors from Mixture-of-Experts, *Yanguang Sun, Jiawei Lian, Jian Yang, Lei Luo*
- 239 Progressive Test Time Energy Adaptation for Medical Image Segmentation, *Xiaoran Zhang, Byung-Woo Hong, Hyoungseob Park, Daniel H. Pak, Anne-Marie Rickmann, Lawrence H. Staib, James S. Duncan, Alex Wong*
- 240 Implicit Counterfactual Learning for Audio-Visual Segmentation, *Mingfeng Zha, Tianyu Li, Guoqing Wang, Peng Wang, Yangyang Wu, Yang Yang, Heng Tao Shen*
- 241 Revisiting Efficient Semantic Segmentation: Learning Offsets for Better Spatial and Class Feature Alignment, *Shi-Chen Zhang, Yunheng Li, Yu-Huan Wu, Qibin Hou, Ming-Ming Cheng*
- 242 Rethinking Detecting Salient and Camouflaged Objects in Unconstrained Scenes, *Zhangjun Zhou, Yiping Li, Chunlin Zhong, Jianuo Huang, Jialun Pei, Hua Li, He Tang*
- 243 AHCPTQ: Accurate and Hardware-Compatible Post-Training Quantization for Segment Anything Model, *Wenlun Zhang, Yunshan Zhong, Shimpei Ando, Kentaro Yoshioka*
- 244 Pseudo-SD: Pseudo Controlled Stable Diffusion for Semi-Supervised and Cross-Domain Semantic Segmentation, *Dong Zhao, Qi Zang, Shuang Wang, Nicu Sebe, Zhun Zhong*
- 245 GUIOdyssey: A Comprehensive Dataset for Cross-App GUI Navigation on Mobile Devices, *Quanfang Lu, Wenqi Shao, Zitao Liu, Lingxiao Du, Fanqing Meng, Boxuan Li, Botong Chen, Siyuan Huang, Kaipeng Zhang, Ping Luo*
- 246 PathDiff: Histopathology Image Synthesis with Unpaired Text and Mask Conditions, *Mahesh Bhosale, Abdul Wasi, Yuanhao Zhai, Yunjie Tian, Samuel Border, Nan Xi, Pinaki Sarder, Junsong Yuan, David Doermann, Xuan Gong*
- 247 Learning Beyond Still Frames: Scaling Vision-Language Models with Video, *Yiyuan Zhang, Handong Li, Jing Liu, Xiangyu Yue*
- 248 Is CLIP ideal? No. Can we fix it? Yes!, *Raphi Kang, Yue Song, Georgia Gkioxari, Pietro Perona*
- 249 HQ-CLIP: Leveraging Large Vision-Language Models to Create High-Quality Image-Text Datasets and CLIP Models, *Zhixiang Wei, Guangting Wang, Xiaoxiao Ma, Ke Mei, Huaian Chen, Yi Jin, Fengyuan Rao*
- 250 Dynamic Dictionary Learning for Remote Sensing Image Segmentation, *Xuechao Zou, Yue Li, Shun Zhang, Kai Li, Shiyang Wang, Pin Tao, Junliang Xing, Congyan Lang*
- 251 Temporal-aware Query Routing for Real-time Video Instance Segmentation, *Zesen Cheng, Kehan Li, Yian Zhao, Hang Zhang, Chang Liu, Jie Chen*
- 252 WINS: Winograd Structured Pruning for Fast Winograd Convolution, *Cheonjun Park, Hyun Jae Oh, Mincheol Park, Hyunchan Moon, Minsik Kim, Suhyun Kim, Myung Kuk Yoon, Won Woo Ro*
- 253 Bridging Domain Generalization to Multimodal Domain Generalization via Unified Representations, *Hai Huang, Yan Xia, Sashuai Zhou, Hanting Wang, Shulei Wang, Zhou Zhao*

- 254 Free-MoRef: Instantly Multiplexing Context Perception Capabilities of Video-MLLMs within Single Inference, *Kuo Wang, Quanlong Zheng, Junlin Xie, Yanhao Zhang, Jinguo Luo, Haonan Lu, Liang Lin, Fan Zhou, Guanbin Li*
- 255 Towards Fine-grained Interactive Segmentation in Images and Videos, *Yuan Yao, Qiushi Yang, Miaomiao Cui, Liefeng Bo*
- 256 LLM-Assisted Semantic Guidance for Sparsely Annotated Remote Sensing Object Detection, *Wei Liao, Chunyan Xu, Chenxu Wang, Zhen Cui*
- 257 Learnable Retrieval Enhanced Visual-Text Alignment and Fusion for Radiology Report Generation, *Qin Zhou, Guoyan Liang, Xindi Li, Jingyuan Chen, Zhe Wang, Chang Yao, Sai Wu*
- 258 Generalizable Object Re-Identification via Visual In-Context Prompting, *Zhizhong Huang, Xiaoming Liu*
- 259 TAB: Transformer Attention Bottlenecks enable User Intervention and Debugging in Vision-Language Models, *Pooyan Rahmazadehgervi, Hung Huy Nguyen, Rosanne Liu, Long Mai, Anh Totti Nguyen*
- 260 Anomaly Detection of Integrated Circuits Package Substrates Using the Large Vision Model SAIC: Dataset Construction, Methodology, and Application, *Ruiyun Yu, Bingyang Guo, Haoyuan Li*
- 261 Towards Omnimodal Expressions and Reasoning in Referring Audio-Visual Segmentation, *Kaining Ying, Henghui Ding, Guangquan Jie, Yu-Gang Jiang*
- 262 Streaming VideoLLMs for Real-Time Procedural Video Understanding, *Dibyadip Chatterjee, Edoardo Remelli, Yale Song, Bugra Tekin, Abhay Mittal, Bharat Bhatnagar, Necati Cihan Camgoz, Shreyas Hampali, Eric Sauser, Shugao Ma, Angela Yao, Fadime Sener*
- 263 Prompt-driven Transferable Adversarial Attack on Person Re-Identification with Attribute-aware Textual Inversion, *Yuan Bian, Min Liu, Yunqi Yi, Xueping Wang, Shuai Jiang, Yaonan Wang*
- 264 Debaised Curriculum Adaptation for Safe Transfer Learning in Chest X-ray Classification, *Mingyang Liu, Xinyang Chen, Yang Shu, Xiucheng Li, Weili Guan, Liqiang Nie*
- 265 Frequency-Dynamic Attention Modulation For Dense Prediction, *Linwei Chen, Lin Gu, Ying Fu*
- 266 Memory-Efficient 4-bit Preconditioned Stochastic Optimization, *Jingyang Li, Kuangyu Ding, Kim-Chuan Toh, Pan Zhou*
- 267 Cross-Category Subjectivity Generalization for Style-Adaptive Sketch Re-ID, *Zechao Hu, Zhengwei Yang, Hao Li, Zheng Wang, Yixiong Zou*
- 268 FrameFusion: Combining Similarity and Importance for Video Token Reduction on Large Vision Language Models, *Tianyu Fu, Tengxuan Liu, Qinghao Han, Guohao Dai, Shengen Yan, Huazhong Yang, Xuefei Ning, Yu Wang*
- 269 Stepping Out of Similar Semantic Space for Open-Vocabulary Segmentation, *Yong Liu, Song-Li Wu, Sule Bai, Jiahao Wang, Yitong Wang, Yansong Tang*
- 270 CoTMR: Chain-of-Thought Multi-Scale Reasoning for Training-Free Zero-Shot Composed Image Retrieval, *Zelong Sun, Dong Jing, Zhiwu Lu*
- 271 CoHD: A Counting-Aware Hierarchical Decoding Framework for Generalized Referring Expression Segmentation, *Zhuoyan Luo, Yinghao Wu, Tianheng Cheng, Yong Liu, Yicheng Xiao, Hongfa Wang, Xiao-Ping Zhang, Yujiu Yang*
- 272 Aligning Effective Tokens with Video Anomaly in Large Language Models, *Yingxian Chen, Jiahui Liu, Ruidi Fan, Yanwei Li, Chirui Chang, Shizhen Zhao, Wilton W. T. Fok, Xiaojuan Qi, Yik-Chung Wu*
- 273 No More Sibling Rivalry: Debiasing Human-Object Interaction Detection, *Bin Yang, Yulin Zhang, Hong-Yu Zhou, Sibe Yang*
- 274 WSI-LLaVA: A Multimodal Large Language Model for Whole Slide Image, *Yuci Liang, Xinheng Lyu, Wenting Chen, Meidan Ding, Jipeng Zhang, Xiangjian He, Song Wu, Xiaohan Xing, Sen Yang, Xiyue Wang, Linlin Shen*
- 275 Borrowing Eyes for the Blind Spot: Overcoming Data Scarcity in Malicious Video Detection via Cross-Domain Retrieval Augmentation, *Rongpei Hong, Jian Lang, Ting Zhong, Fan Zhou*
- 276 Bridging Local Inductive Bias and Long-Range Dependencies with Pixel-Mamba for End-to-end Whole Slide Image Analysis, *Zhongwei Qiu, Hanqing Chao, Tiancheng Lin, Wanxing Chang, Zijiang Yang, Wenpei Jiao, Yixuan Shen, Yunshuo Zhang, Yelin Yang, Wenbin Liu, Hui Jiang, Yun Bian, Ke Yan, Dakai Jin, Le Lu*
- 277 DASH: Detection and Assessment of Systematic Hallucinations of VLMs, *Maximilian Augustin, Yannic Neuhäus, Matthias Hein*
- 278 Sim-DETR: Unlock DETR for Temporal Sentence Grounding, *Jiajin Tang, Zhengxuan Wei, Yuchen Zhu, Cheng Shi, Guanbin Li, Liang Lin, Sibe Yang*
- 279 ViCTr: Vital Consistency Transfer for Pathology Aware Image Synthesis, *Onkar Susladkar, Gayatri Deshmukh, Yalcin Tur, Gorkem Durak, Ulas Bagci*
- 280 Multi-Modal Multi-Task Unified Embedding Model (M3T-UEM): A Task-Adaptive Representation Learning Framework, *Rohan Sharma, Changyou Chen, Feng-Ju Chang, Seongjun Yun, Xiaohu Xie, Rui Meng, Dehong Xu, Alejandro Mottini, Qingjun Cui*
- 281 DIH-CLIP: Unleashing the Diversity of Multi-Head Self-Attention for Training-Free Open-Vocabulary Semantic Segmentation, *Songsong Duan, Xi Yang, Nannan Wang*
- 282 SignRep: Enhancing Self-Supervised Sign Representations, *Ryan Wong, Necati Cihan Camgoz, Richard Bowden*
- 283 Plug-in Feedback Self-adaptive Attention in CLIP for Training-free Open-Vocabulary Segmentation, *Zhixiang Chi, Yanan Wu, Li Gu, Huan Liu, Ziqiang Wang, Yang Zhang, Yang Wang, Konstantinos Plataniotis*
- 284 Feather the Throttle: Revisiting Visual Token Pruning for Vision-Language Model Acceleration, *Mark Endo, Xiaohan Wang, Serena Yeung-Levy*
- 285 Intermediate Connectors and Geometric Priors for Language-Guided Affordance Segmentation on Unseen Object Categories, *Yicong Li, Yiyang Chen, Zhenyuan Ma, Junbin Xiao, Xiang Wang, Angela Yao*
- 286 Integrating Biological Knowledge for Robust Microscopy Image Profiling on De Novo Cell Lines, *Jiayuan Chen, Thai-Hoang Pham, Yuanlong Wang, Ping Zhang*
- 287 LLaVA-PruMerge: Adaptive Token Reduction for Efficient Large Multimodal Models, *Yuzhang Shang, Mu Cai, Bingxin Xu, Yong Jae Lee, Yan Yan*
- 288 Task-Specific Zero-shot Quantization-Aware Training for Object Detection, *Changhao Li, Xinrui Chen, Ji Wang, Kang Zhao, Jianfei Chen*
- 289 Boosting Multimodal Learning via Disentangled Gradient Learning, *Shicai Wei, Chunbo Luo, Yang Luo*
- 290 CapeLLM: Support-Free Category-Agnostic Pose Estimation with Multimodal Large Language Models, *Junho Kim, Hyungjin Chung, Byung-Hoon Kim*
- 291 AURELIA: Test-time Reasoning Distillation in Audio-Visual LLMs, *Sanjoy Chowdhury, Hanan Gani, Nishit Anand, Sayan Nag, Ruohan Gao, Mohamed Elhoseiny, Salman Khan, Dinesh Manocha*
- 292 HERMES: temporal-coHERent long-forM understanding with Episodes and Semantics, *Gueter Josmy Faure, Jia-Fong Yeh, Min-Hung Chen, Hung-Ting Su, Shang-Hong Lai, Winston H. Hsu*
- 293 HRScene: How Far Are VLMs from Effective High-Resolution Image Understanding?, *Yusen Zhang, Wenliang Zheng, Aashrith Madasu, Peng Shi, Ryo Kamoi, Hao Zhou, Zhuoyang Zou, Shu Zhao, Sarkar Snigdha Sarathi Das, Vipul Gupta, Xiaoxin Lu, Nan Zhang, Ranran Haoran Zhang, Avitej Iyer, Renze Lou, Wenpeng Yin, Rui Zhang*
- 294 Cycle Consistency as Reward: Learning Image-Text Alignment without Human Preferences, *Hyojin Bahng, Caroline Chan, Fredo Durand, Phillip Isola*
- 295 Griffon v2: Advancing Multimodal Perception with High-Resolution Scaling and Visual-Language Co-Referring, *Yufei Zhan, Shurong Zheng, Yousong Zhu, Hongyin Zhao, Fan Yang, Ming Tang, Jinqiao Wang*
- 296 LVBench: An Extreme Long Video Understanding Benchmark, *Weihan Wang, Zehai He, Wenyi Hong, Yean Cheng, Xiaohan Zhang, Ji Qi, Ming Ding, Xiaotao Gu, Shiyu Huang, Bin Xu, Yuxiao Dong, Jie Tang*
- 297 Addressing Representation Collapse in Vector Quantized Models with One Linear Layer, *Yongxin Zhu, Bocheng Li, Yifei Xin, Zhihua Xia, Linli Xu*
- 298 MultiADS: Defect-aware Supervision for Multi-type Anomaly Detection and Segmentation in Zero-Shot Learning, *Ylli Sadikaj, Hongkuan Zhou, Lavdim Halilaj, Stefan Schmid, Steffen Staab, Claudia Plant*
- 299 Debiasing Trace Guidance: Top-down Trace Distillation and Bottom-up Velocity Alignment for Unsupervised Anomaly Detection, *Xingjian Wang, Li Chai, Jiming Chen*
- 300 ODDR: Outlier Detection & Dimension Reduction Based Defense Against Adversarial Patches, *Nandish Chattopadhyay, Amira Guesmi, Muhammad Abdullah Hanif, Bassem Ouni, Muhammad Shafique*
- 301 Similarity Memory Prior is All You Need for Medical Image

- * Segmentation, *Hao Tang, Zhiqing Guo, Liejun Wang, Chao Liu*
- 302 CARL: Causality-guided Architecture Representation Learning for an Interpretable Performance Predictor, *Han Ji, Yuqi Feng, Jiahao Fan, Yanan Sun*
- 303 Diffusion Transformer meets Multi-level Wavelet Spectrum for Single Image Super-Resolution, *Peng Du, Hui Li, Han Xu, Paul Barom Jeon, Dongwook Lee, Daehyun Ji, Ran Yang, Feng Zhu*
- 304 Automated Model Evaluation for Object Detection via Prediction Consistency and Reliability, *Seungju Yoo, Hyuk Kwon, Joong-Won Hwang, Kibok Lee*
- 305 CalliReader: Contextualizing Chinese Calligraphy via an Embedding-Aligned Vision-Language Model, *Yuxuan Luo, Jiaqi Tang, Chenyi Huang, Feiyang Hao, Zhouhui Lian*
- 306 Boosting Vision Semantic Density with Anatomy Normality Modeling for Medical Vision-language Pre-training, *Weiwei Cao, Jianpeng Zhang, Zhongyi Shui, Sinuo Wang, Zeli Chen, Xi Li, Le Lu, Xianghua Ye, Qi Zhang, Tingbo Liang, Ling Zhang*
- 307 CoSMIC: Continual Self-supervised Learning for Multi-Domain Medical Imaging via Conditional Mutual Information Maximization, *Yihang Liu, Ying Wen, Longzhen Yang, Lianghua He, Heng Tao Shen*
- 308 Generalized Few-Shot Point Cloud Segmentation via LLM-Assisted Hyper-Relation Matching, *Zhaoyang Li, Yuan Wang, Guoxin Xiong, Wangkai Li, Yuwen Pan, Tianzhu Zhang*
- 309 Enhancing Partially Relevant Video Retrieval with Hyperbolic Learning, *Jun Li, Jinpeng Wang, Chaolei Tan, Niu Lian, Long Chen, Yaowei Wang, Min Zhang, Shu-Tao Xia, Bin Chen*
- 310 Accelerate 3D Object Detection Models via Zero-Shot Attention Key Pruning, *Lizhen Xu, Xiuxiu Bai, Xiaojun Jia, Jianwu Fang, Shanmin Pang*
- 311 MSA2: Multi-task Framework with Structure-aware and Style-adaptive Character Representation for Open-set Chinese Text Recognition, *Yangfu Li, Hongjian Zhan, Qi Liu, Li Sun, Yu-Jie Xiong, Yue Lu*
- 312 GroundingSuite: Measuring Complex Multi-Granular Pixel Grounding, *Rui Hu, Lianghui Zhu, Yuxuan Zhang, Tianheng Cheng, Lei Liu, Heng Liu, Longjin Ran, Xiaoxin Chen, Wenyu Liu, Xinggang Wang*
- 313 Always Skip Attention, *Yiping Ji, Hemanth Saratchandran, Peyman Moghadam, Simon Lucey*
- 314 Training-Free Class Purification for Open-Vocabulary Semantic Segmentation, *Qi Chen, Lingxiao Yang, Yun Chen, Nailong Zhao, Jianhuang Lai, Jie Shao, Xiaohua Xie*
- 315 SeaS: Few-shot Industrial Anomaly Image Generation with Separation and Sharing Fine-tuning, *Zhewei Dai, Shilei Zeng, Haotian Liu, Xurui Li, Feng Xue, Yu Zhou*
- 316 SC-Captioner: Improving Image Captioning with Self-Correction by Reinforcement Learning, *Lin Zhang, Xianfang Zeng, Kangcong Li, Gang Yu, Tao Chen*
- 317 One Trajectory, One Token: Grounded Video Tokenization via Panoptic Sub-object Trajectory, *Chenhao Zheng, Jieyu Zhang, Mohammadreza Salehi, Ziqi Gao, Vishnu Iyengar, Norimasa Kobori, Quan Kong, Ranjay Krishna*
- 318 Breaking the Encoder Barrier for Seamless Video-Language Understanding, *Handong Li, Yiyuan Zhang, Longteng Guo, Xiangyu Yue, Jing Liu*
- 319 SparseMM: Head Sparsity Emerges from Visual Concept Responses in MLLMs, *Jiahui Wang, Zuyan Liu, Yongming Rao, Jiwen Lu*
- 320 SUB: Benchmarking CBM Generalization via Synthetic Attribute Substitutions, *Jessica Bader, Leander Gierbach, Stephan Alaniz, Zeynep Akata*
- 321 CLIPer: Hierarchically Improving Spatial Representation of CLIP for Open-Vocabulary Semantic Segmentation, *Lin Sun, Jiale Cao, Jin Xie, Xiaoheng Jiang, Yanwei Pang*
- 322 A Token-level Text Image Foundation Model for Document Understanding, *Tongkun Guan, Zining Wang, Pei Fu, Zhengtao Guo, Wei Shen, Kai Zhou, Tiezhu Yue, Chen Duan, Hao Sun, Qianyi Jiang, Junfeng Luo, Xiaokang Yang*
- 323 ReferEverything: Towards Segmenting Everything We Can Speak of in Videos, *Anurag Bagchi, Zhipeng Bao, Yu-Xiong Wang, Pavel Tokmakov, Martial Hebert*
- 324 Continual Multiple Instance Learning with Enhanced Localization for Histopathological Whole Slide Image Analysis, *Byung Hyun Lee, Wongi Jeong, Woojae Han, Kyoungbun Lee, Se Young Chun*
- 325 From Trial to Triumph: Advancing Long Video Understanding via Visual Context Sample Scaling and Self-reward Alignment, *Yucheng Suo, Fan Ma, Linchao Zhu, Tianyi Wang, Fengyun Rao, Yi Yang*
- 326 Cross-Architecture Distillation Made Simple with Redundancy Suppression, *Weijia Zhang, Yuehao Liu, Wu Ran, Chao Ma*
- 327 HiMTok: Learning Hierarchical Mask Tokens for Image Segmentation with Large Multimodal Model, *Tao Wang, Changxu Cheng, Lingfeng Wang, Senda Chen, Wuyue Zhao*
- 328 DC-TTA: Divide-and-Conquer Framework for Test-Time Adaptation of Interactive Segmentation, *Jihun Kim, Hoyong Kwon, Hyeokjun Kweon, Woosong Jeong, Kuk-Jin Yoon*
- 329 FIND: Few-Shot Anomaly Inspection with Normal-Only Multi-Modal Data, *Yiting Li, Fayao Liu, Jingyi Liao, Sichao Tian, Chuan-Sheng Foo, Xulei Yang*
- 330 VISO: Accelerating In-orbit Object Detection with Language-Guided Mask Learning and Sparse Inference, *Meiqi Wang, Han Qiu*
- 331 KDA: Knowledge Diffusion Alignment with Enhanced Context for Video Temporal Grounding, *Ran Ran, Jiwei Wei, Shiyuan He, Zeyu Ma, Chaoning Zhang, Ning Xie, Yang Yang*
- 332 PVChat: Personalized Video Chat with One-Shot Learning, *Yufei Shi, Weilong Yan, Gang Xu, Yumeng Li, Yucheng Chen, Zhenxi Li, Fei Yu, Ming Li, Si Yong Yeo*
- 333 Unsupervised Histopathological Image Semantic Segmentation with Overlapping Patches Consistency Constraint, *Wentian Cai, Weizhao Weng, Zihao Huang, Yandan Chen, Siqun Huang, Ping Gao, Victor C. M. Leung, Ying Gao*
- 334 How Do Optical Flow and Textual Prompts Collaborate to Assist in Audio-Visual Semantic Segmentation?, *Yujian Lee, Peng Gao, Yongqi Xu, Wentao Fan*
- 335 UINavBench: A Framework for Comprehensive Evaluation of Interactive Digital Agents, *Harsh Agrawal, Eldon Schoop, Xinlei Pan, Anuj Mahajan, Ari Seff, Di Feng, Ruijia Cheng, Andres Romero Mier Y Teran, Esteban Gomez, Abhishek Sundararajan, Forrest Huang, Amanda Swearingin, Mohana Prasad Sathya Moorthy, Jeff Nichols, Alexander Toshev*
- 336 Adapt Foundational Segmentation Models with Heterogeneous Searching Space, *Li Yi, Jie Hu, Songan Zhang, Guannan Jiang*
- 337 VIPerson: Flexibly Generating Virtual Identity for Person Re-identification, *Xiao-Wen Zhang, Delong Zhang, Yi-Xing Peng, Zhi Ouyang, Jingke Meng, Wei-Shi Zheng*
- 338 SpikePack: Enhanced Information Flow in Spiking Neural Networks with High Hardware Compatibility, *Guobin Shen, Jindong Li, Tenglong Li, Dongcheng Zhao, Yi Zeng*
- 339 AdsQA: Towards Advertisement Video Understanding, *Xinwei Long, Kai Tian, Peng Xu, Guoli Jia, Jingxuan Li, Sa Yang, Yihua Shao, Kaiyan Zhang, Che Jiang, Hao Xu, Yang Liu, Jiaheng Ma, Bowen Zhou*
- 340 Towards Robustness of Person Search against Corruptions, *Woojung Son, Yoonki Cho, Guoyuan An, Chanmi Lee, Sung-Eui Yoon*
- 341 Toward Long-Tailed Online Anomaly Detection through Class-Agnostic Concepts, *Chiao-An Yang, Kuan-Chuan Peng, Raymond A. Yeh*
- 342 PathFinder: A Multi-Modal Multi-Agent System for Medical Diagnostic Decision-Making Applied to Histopathology, *Fatemeh Ghezloo, Mehmet Saygin Seyfioglu, Rustin Soraki, Wisdom O. Ikezogwo, Beibin Li, Tejoram Vivekanandan, Joann G. Elmore, Ranjay Krishna, Linda Shapiro*
- 343 Beyond [cls]: Exploring the True Potential of Masked Image Modeling Representations, *Marcin Przewięźlikowski, Randall Balestriero, Wojciech Jasiński, Marek Śmieja, Bartosz Zieliński*
- 344 Bringing RNNs Back to Efficient Open-Ended Video Understanding, *Weili Xu, Enxin Song, Wenhao Chai, Xuexiang Wen, Tian Ye, Gaoang Wang*
- 345 Scheduling Weight Transitions for Quantization-Aware Training, *Junghyup Lee, Jeimin Jeon, Dohyung Kim, Bumsub Ham*
- 346 Revisiting Adversarial Patch Defenses on Object Detectors: Unified Evaluation, Large-Scale Dataset, and New Insights, *Junhao Zheng, Jiahao Sun, Chenhao Lin, Zhengyu Zhao, Chen Ma, Chong Zhang, Cong Wang, Qian Wang, Chao Shen*

- 347 Harnessing Vision Foundation Models for High-Performance, Training-Free Open Vocabulary Segmentation, *Yuheng Shi, Minjing Dong, Chang Xu*
- 348 A Visual Leap in CLIP Compositionality Reasoning through Generation of Counterfactual Sets, *Zexi Jia, Chuanwei Huang, Hongyan Fei, Yeshuang Zhu, Zhiqiang Yuan, Ying Deng, Jiawei Zhang, Jinchao Zhang, Jie Zhou*
- 349 AcZeroTS: Active Learning for Zero-shot Tissue Segmentation in Pathology Images, *Jiao Tang, Junjie Zhou, Bo Qian, Peng Wan, Yingli Zuo, Wei Shao, Daoqiang Zhang*
- 350 TinyViM: Frequency Decoupling for Tiny Hybrid Vision Mamba, *Xiaowen Ma, Zhenliang Ni, Xinghao Chen*
- 351 FALCON: Resolving Visual Redundancy and Fragmentation in High-resolution Multimodal Large Language Models via Visual Registers, *Renshan Zhang, Rui Shao, Gongwei Chen, Miao Zhang, Kaiwen Zhou, Weili Guan, Liqiang Nie*
- 352 SweetTok: Semantic-Aware Spatial-Temporal Tokenizer for Compact Video Discretization, *Zhentao Tan, Ben Xue, Jian Jia, Junhao Wang, Wencai Ye, Shaoyun Shi, Mingjie Sun, Wenjin Wu, Quan Chen, Peng Jiang*
- 353 Visual Relation Diffusion for Human-Object Interaction Detection, *Ping Cao, Yepeng Tang, Chunjie Zhang, Xiaolong Zheng, Chao Liang, Yunchao Wei, Yao Zhao*
- 354 Flow-MIL: Constructing Highly-expressive Latent Feature Space For Whole Slide Image Classification Using Normalizing Flow, *Yingfan Ma, Bohan An, Ao Shen, Mingzhi Yuan, Minghong Duan, Manning Wang*
- 355 HarmonySeg: Tubular Structure Segmentation with Deep-Shallow Feature Fusion and Growth-Suppression Balanced Loss, *Yi Huang, Ke Zhang, Wei Liu, Yuanyuan Wang, Vishal M. Patel, Le Lu, Xu Han, Dakai Jin, Ke Yan*
- 356 CompCap: Improving Multimodal Large Language Models with Composite Captions, *Xiaohui Chen, Satya Narayan Shukla, Mahmoud Azab, Aashu Singh, Qifan Wang, David Yang, ShengYun Peng, Hanchao Yu, Shen Yan, Xuwen Zhang, Baosheng He*
- 357 TOGA: Temporally Grounded Open-Ended Video QA with Weak Supervision, *Ayush Gupta, Anirban Roy, Rama Chellappa, Nathaniel D. Bastian, Alvaro Velasquez, Susmit Jha*
- 358 Stable Diffusion Models are Secretly Good at Visual In-Context Learning, *Trevine Oorloff, Vishwanath Sindagi, Wele Gedara Chaminda Bandara, Ali Shafahi, Amin Ghiasi, Charan Prakash, Reza Ardekani*
- 359 FOLDER: Accelerating Multi-Modal Large Language Models with Enhanced Performance, *Haicheng Wang, Zhemeng Yu, Gabriele Spadaro, Chen Ju, Victor Quéru, Shuai Xiao, Enzo Tartaglione*
- 360 Language Decoupling with Fine-grained Knowledge Guidance for Referring Multi-object Tracking, *Guangyao Li, Siping Zhuang, Yajun Jian, Yan Yan, Hanzi Wang*
- 361 Prompt Guidance and Human Proximal Perception for HOT Prediction with Regional Joint Loss, *Yuxiao Wang, Yu Lei, Zhenao Wei, Weiyang Xue, Xinyu Jiang, Nan Zhuang, Qi Liu*
- 362 Prior2Former - Evidential Modeling of Mask Transformers for * Assumption-Free Open-World Panoptic Segmentation, ☆ Sebastian Schmidt, Julius Koerner, Dominik Fuchsgruber, Stefano Gasperini, Federico Tombari, Stephan Günnemann
- 363 Seeing the Unseen: A Semantic Alignment and Context-Aware Prompt Framework for Open-Vocabulary Camouflaged Object Segmentation, *Peng Ren, Tian Bai, Jing Sun, Fuming Sun*
- 364 ViLLa: Video Reasoning Segmentation with Large Language Model, *Rongkun Zheng, Lu Qi, Xi Chen, Yi Wang, Kun Wang, Hengshuang Zhao*
- 365 DynImg: Key Frames with Visual Prompts are Good Representation for Multi-Modal Video Understanding, *Xiaoyi Bao, Chenwei Xie, Hao Tang, Tingyu Weng, Xiaofeng Wang, Yun Zheng, Xingang Wang*
- 366 Object-level Correlation for Few-Shot Segmentation, *Chunlin Wen, Yu Zhang, Jie Fan, Hongyuan Zhu, Xiu-Shen Wei, Yijun Wang, Zhiqiang Kou, Shuzhou Sun*
- 367 Vision-Language Neural Graph Featurization for Extracting Retinal Lesions, *Taimur Hassan, Anabia Sohail, Muzammal Naseer, Naoufel Werghi*
- 368 SSVQ: Unleashing the Potential of Vector Quantization with Sign-Splitting, *Shuai Li, Juncan Deng, Chengxuan Wang, Kedong Xu, Rongtao Deng, Hong Gu, Haibin Shen, Kejie Huang*
- 369 RadGPT: Constructing 3D Image-Text Tumor Datasets, ☆ Pedro R.A.S. Bassi, Mehmet Can Yavuz, Ibrahim Ethem Hamamci, Sezgin Er, Xiaoxi Chen, Wenxuan Li, Bjoern Menze, Sergio Decherchi, Andrea Cavalli, Kang Wang, Yang Yang, Alan Yuille, Zongwei Zhou
- 370 SPA: Efficient User-Preference Alignment against Uncertainty in ☆ Medical Image Segmentation, *Jiayuan Zhu, Junde Wu, Cheng Ouyang, Konstantinos Kamnitsas, J. Alison Noble*
- 371 CoStoDet-DDPM: Collaborative Training of Stochastic and Deterministic Models Improves Surgical Workflow Anticipation and Recognition, *Kaixiang Yang, Xin Li, Qiang Li, Zhiwei Wang*
- 372 LangBridge: Interpreting Image as a Combination of Language Embeddings, *Jiaqi Liao, Yuwei Niu, Fanqing Meng, Hao Li, Changyao Tian, Yinyu Du, Yuwen Xiong, Dianqi Li, Xizhou Zhu, Li Yuan, Jifeng Dai, Yu Cheng*
- 373 Know Your Attention Maps: Class-specific Token Masking for Weakly Supervised Semantic Segmentation, *Joëlle Hanna, Damian Borth*
- 374 VideoMiner: Iteratively Grounding Key Frames of Hour-Long Videos via Tree-based Group Relative Policy Optimization, *Xinye Cao, Hongcan Guo, Jiawen Qian, Guoshun Nan, Chao Wang, Yuqi Pan, Tianhao Hou, Xiaojuan Wang, Yutong Gao*
- 375 SparseVILA: Decoupling Visual Sparsity for Efficient VLM Inference, *Samir Khaki, Junxian Guo, Jiaming Tang, Shang Yang, Yukang Chen, Konstantinos N. Plataniotis, Yao Lu, Song Han, Zhijian Liu*
- 376 LOTS of Fashion! Multi-Conditioning for Image Generation via Sketch-Text Pairing, Federico Girella, Davide Talon, Ziyue Liu, Zanxi Ruan, Yiming Wang, Marco Cristani
- 377 Soft Local Completeness: Rethinking Completeness in XAI, ☆ Ziv Weiss Haddad, Oren Barkan, Yehonatan Elisha, Noam Koenigstein
- 378 Unbiased Region-Language Alignment for Open-Vocabulary Dense Prediction, *Yunheng Li, Yuxuan Li, Quan-Sheng Zeng, Wenhai Wang, Qibin Hou, Ming-Ming Cheng*
- 379 Signs as Tokens: A Retrieval-Enhanced Multilingual Sign Language Generator, *Ronglai Zuo, Rolandos Alexandros Potamias, Evangelos Ververas, Jiankang Deng, Stefanos Zafeiriou*
- 380 Flow4Agent: Long-form Video Understanding via Motion Prior from Optical Flow, *Ruyang Liu, Shangkun Sun, Haoran Tang, Wei Gao, Ge Li*
- 381 ZIM: Zero-Shot Image Matting for Anything, Beomyoung Kim, * Chanyong Shin, Joonhyun Jeong, Hyungsik Jung, Se-Yun Lee, Sewhan Chun, Dong-Hyun Hwang, Joonsang Yu
- 382 An OpenMind for 3D Medical Vision Self-supervised Learning, *Tassilo Wald, Constantin Ulrich, Jonathan Suprijadi, Sebastian Ziegler, Michal Nohel, Robin Peretzke, Gregor Kohler, Klaus Maier-Hein*
- 383 Make Your Training Flexible: Towards Deployment-Efficient Video Models, *Chenting Wang, Kunchang Li, Tianxiang Jiang, Xiangyu Zeng, Yi Wang, Limin Wang*
- 384 OmniSAM: Omnidirectional Segment Anything Model for UDA in * Panoramic Semantic Segmentation, *Ding Zhong, Xu Zheng, Chenfei Liao, Yuanhui Lyu, Jiale Chen, Shengyang Wu, Linfeng Zhang, Xuming Hu*
- 385 LawDIS: Language-Window-based Controllable Dichotomous Image Segmentation, *Xinyu Yan, Meijun Sun, Ge-Peng Ji, Fahad Shahbaz Khan, Salman Khan, Deng-Ping Fan*
- 386 ModalTune: Fine-Tuning Slide-Level Foundation Models with Multi-Modal Information for Multi-task Learning in Digital Pathology, *Vishwesh Ramanathan, Tony Xu, Pushpak Pati, Faruk Ahmed, Maged Goubran, Anne L. Martel*
- 387 VFlowOpt: A Token Pruning Framework for LMMs with Visual Information Flow-Guided Optimization, *Sihan Yang, Runsen Xu, Chenhang Cui, Tai Wang, Dahua Lin, Jiangmiao Pang*
- 388 D-Attn: Decomposed Attention for Large Vision-and-Language Model, *Chia-Wen Kuo, Sijie Zhu, Fan Chen, Xiaohui Shen, Longyin Wen*
- 389 Open-Vocabulary HOI Detection with Interaction-aware Prompt and Concept Calibration, *Ting Lei, Shaofeng Yin, Qingchao Chen, Yuxin Peng, Yang Liu*
- 390 AirCache: Activating Inter-modal Relevancy KV Cache Compression for Efficient Large Vision-Language Model Inference, *Kai Huang, Hao Zou, Bochen Wang, Ye Xi, Zhen Xie, Hao Wang*

- 391 MINERVA: Evaluating Complex Video Reasoning, *Arsha Nagrani, Sachit Menon, Ahmet Iscen, Shyamal Buch, Ramin Mehran, Nilpa Jha, Anja Hauth, Yukun Zhu, Carl Vondrick, Mikhail Sirotenko, Cordelia Schmid, Tobias Weyand*
- 392 MPG-SAM 2: Adapting SAM 2 with Mask Priors and Global Context for Referring Video Object Segmentation, *Fu Rong, Meng Lan, Qian Zhang, Lefei Zhang*
- 393 Multi-Granular Spatio-Temporal Token Merging for Training-Free Acceleration of Video LLMs, *Jeongseok Hyun, Sukjun Hwang, Su Ho Han, Taeoh Kim, Inwoong Lee, Dongyoon Wee, Joon-Young Lee, Seon Joo Kim, Minho Shim*
- 394 Scaling Tumor Segmentation: Best Lessons from Real and Synthetic Data, *Qi Chen, Xinze Zhou, Chen Liu, Hao Chen, Wenxuan Li, Zekun Jiang, Ziyang Huang, Yuxuan Zhao, Dexin Yu, Junjun He, Yefeng Zheng, Ling Shao, Alan Yuille, Zongwei Zhou*
- 395 TAViS: Text-bridged Audio-Visual Segmentation with Foundation Models, *Ziyang Luo, Nian Liu, Xuguang Yang, Salman Khan, Rao Muhammad Anwer, Hisham Cholakkal, Fahad Shahbaz Khan, Junwei Han*
- 396 Teaching AI the Anatomy Behind the Scan: Addressing Anatomical Flaws in Medical Image Segmentation with Learnable Prior, *Young Seok Jeon, Hongfei Yang, Huazhu Fu, Mengling Feng*
- 397 CLIP-Adapted Region-to-Text Learning for Generative Open-Vocabulary Semantic Segmentation, *Jiannan Ge, Lingxi Xie, Hongtao Xie, Pandeng Li, Sun-Ao Liu, Xiaopeng Zhang, Qi Tian, Yongdong Zhang*
- 398 ConformalSAM: Unlocking the Potential of Foundational Segmentation Models in Semi-Supervised Semantic Segmentation with Conformal Prediction, *Danhui Chen, Ziquan Liu, Chuxi Yang, Dan Wang, Yan Yan, Yi Xu, Xiangyang Ji*
- 399 LIRA: Inferring Segmentation in Large Multi-modal Models with Local Interleaved Region Assistance, *Zhang Li, Biao Yang, Qiang Liu, Shuo Zhang, Zhiyin Ma, Liang Yin, Linger Deng, Yabo Sun, Yuliang Liu, Xiang Bai*
- 400 SimMLM: A Simple Framework for Multi-modal Learning with Missing Modality, *Sijie Li, Chen Chen, Jungong Han*
- 401 NAVER: A Neuro-Symbolic Compositional Automaton for Visual Grounding with Explicit Logic Reasoning, *Zhixi Cai, Fucai Ke, Simindokht Jahangard, Maria Garcia de la Banda, Reza Haffari, Peter J. Stuckey, Hamid Rezaatfighi*
- 402 MDP3: A Training-free Approach for List-wise Frame Selection in Video-LLMs, *Hui Sun, Shiyin Lu, Huanyu Wang, Qing-Guo Chen, Zhao Xu, Weihua Luo, Kaifu Zhang, Ming Li*
- 403 Toward Fair and Accurate Cross-Domain Medical Image Segmentation: A VLM-Driven Active Domain Adaptation Paradigm, *Hongqiu Wang, Wu Chen, Xiangde Luo, Zhaohu Xing, Lihao Liu, Jing Qin, Shaozhi Wu, Lei Zhu*
- 404 Incremental Few-Shot Semantic Segmentation via Multi-Level Switchable Visual Prompts, *Maixian Wan, Kaige Li, Qichuan Geng, Weimin Shi, Zhong Zhou*
- 405 TopoTTA: Topology-Enhanced Test-Time Adaptation for Tubular Structure Segmentation, *Jiale Zhou, Wenhan Wang, Shikun Li, Xiaolei Qu, Xin Guo, Yizhong Liu, Wenzhong Tang, Xun Lin, Yefeng Zheng*
- 406 MUSE-VL: Modeling Unified VLM through Semantic Discrete Encoding, *Rongchang Xie, Chen Du, Ping Song, Chang Liu*
- 407 Token-Efficient VLM: High-Resolution Image Understanding via Dynamic Region Proposal, *Yitong Jiang, Jinwei Gu, Tianfan Xue, Ka Chun Cheung, Pavlo Molchanov, Hongxu Yin, Sifei Liu*
- 408 Vision-Language Models Can't See the Obvious, *Ngoc Dung Huynh, Phuc H Le-Khac, Wamiq Reyaz Para, Ankit Singh, Sanath Narayan*
- 409 VideoLLaMB: Long Streaming Video Understanding with Recurrent Memory Bridges, *Yuxuan Wang, Yiqi Song, Cihang Xie, Yang Liu, Zilong Zheng*
- 410 One Polyp Identifies All: One-Shot Polyp Segmentation with SAM via Cascaded Priors and Iterative Prompt Evolution, *Xinyu Mao, Xiaohan Xing, Fei Meng, Jianbang Liu, Fan Bai, Qiang Nie, Max Meng*
- 411 Region-aware Anchoring Mechanism for Efficient Referring Visual Grounding, *Shuyi Ouyang, Ziwei Niu, Hongyi Wang, Yen-Wei Chen, Lanfen Lin*
- 412 VTimeCoT: Thinking by Drawing for Video Temporal Grounding and Reasoning, *Jinglei Zhang, Yuanfan Guo, Rolandos Alexandros Potamias, Jiankang Deng, Hang Xu, Chao Ma*
- 413 Training-Free Industrial Defect Generation with Diffusion Models, *Ruyi Xu, Yen-Tzu Chiu, Tai-I Chen, Oscar Chew, Yung-Yu Chuang, Wen-Huang Cheng*
- 414 Kaputt: A Large-Scale Dataset for Visual Defect Detection, ☆ *Sebastian Höfer, Dorian F. Henning, Artemij Amiranashvili, Douglas Morrison, Mariliza Tzes, Ingmar Posner, Marc Matvienko, Alessandro Rennola, Anton Milan*
- 415 Snakes and Ladders: Two Steps Up for VideoMamba, *Hui Lu, Albert A. Salah, Ronald Poppe*
- 416 ChatReID: Open-ended Interactive Person Retrieval via Hierarchical Progressive Tuning for Vision Language Models, *Ke Niu, Haiyang Yu, Mengyang Zhao, Teng Fu, Siyang Yi, Wei Lu, Bin Li, Xuelin Qian, Xiangyang Xue*
- 417 Images as Noisy Labels: Unleashing the Potential of the Diffusion Model for Open-Vocabulary Semantic Segmentation, *Fan Li, Xuanbin Wang, Xuan Wang, Zhaoxiang Zhang, Yuelei Xu*
- 418 Auto-Vocabulary Semantic Segmentation, *Osman Ülger, Maksymilian Kulicki, Yuki Asano, Martin R. Oswald*
- 419 SIC: Similarity-Based Interpretable Image Classification with Neural Networks, *Tom Nuno Wolf, Emre Kavak, Fabian Bongratz, Christian Wachinger*
- 420 TimeExpert: An Expert-Guided Video LLM for Video Temporal Grounding, *Zuhao Yang, Yingchen Yu, Yunqing Zhao, Shijian Lu, Song Bai*
- 421 Enrich and Detect: Video Temporal Grounding with Multimodal LLMs, ☆ *Shraman Pramanick, Effrosyni Mavroudi, Yale Song, Rama Chellappa, Lorenzo Torresani, Triantafyllos Afouras*
- 422 When Pixel Difference Patterns Meet ViT: PiDiViT for Few-Shot Object Detection, *Hongliang Zhou, Yongxiang Liu, Canyu Mo, Weijie Li, Bowen Peng, Li Liu*
- 423 Hierarchy-Aware Pseudo Word Learning with Text Adaptation for Zero-Shot Composed Image Retrieval, *Zhe Li, Lei Zhang, Zheren Fu, Kun Zhang, Zhendong Mao*
- 424 Player-Centric Multimodal Prompt Generation for Large Language Model Based Identity-Aware Basketball Video Captioning, *Zeyu Xi, Haoying Sun, Yaofei Wu, Junchi Yan, Haoran Zhang, Lifang Wu, Liang Wang, Changwen Chen*
- 425 Synchronizing Task Behavior: Aligning Multiple Tasks during Test-Time Training, *Wooseong Jeong, Jegyeong Cho, Youngho Yoon, Kuk-Jin Yoon*
- 426 Agreement aware and dissimilarity oriented GLOM, *Ru Zeng, Yan Song, Yang Zhang, Yanling Hu, Hui Yu*
- 427 Conditional Latent Diffusion Models for Zero-Shot Instance Segmentation, *Maximilian Ulmer, Wout Boerdijk, Rudolph Triebel, Maximilian Durner*
- 428 Embodied Image Captioning: Self-supervised Learning Agents for Spatially Coherent Image Descriptions, ☆ *Tommaso Apicella, Stefano Rosa, Pietro Morerio, Alessio Del Bue, Lorenzo Natale*
- 429 From Objects to Events: Unlocking Complex Visual Understanding in Object Detectors via LLM-guided Symbolic Reasoning, *Yuhui Zeng, Haoxiang Wu, Wenjie Nie, Guangyao Chen, Xiawu Zheng, Yunhang Shen, Jun Peng, Yonghong Tian, Rongrong Ji*
- 430 C2MIL: Synchronizing Semantic and Topological Causalities in Multiple Instance Learning for Robust and Interpretable Survival Analysis, *Min Cen, Zhenfeng Zhuang, Yuzhe Zhang, Min Zeng, Baptiste Magnier, Lequan Yu, Hong Zhang, Liansheng Wang*
- 431 GenHancer: Imperfect Generative Models are Secretly Strong Vision-Centric Enhancers, *Shijie Ma, Yuying Ge, Teng Wang, Yuxin Guo, Yixiao Ge, Ying Shan*
- 432 Breaking Grid Constraints: Dynamic Graph Reconstruction Network for Multi-organ Segmentation, *Junhao Xiao, Yang Wei, Jingyu Wang, Yongchao Wang, Xiuli Bi, Bin Xiao*
- 433 MaskSAM: Auto-prompt SAM with Mask Classification for Volumetric Medical Image Segmentation, *Bin Xie, Hao Tang, Bin Duan, Dawen Cai, Yan Yan, Gady Agam*
- 434 Large-scale Pre-training for Grounded Video Caption Generation, *Evangelos Kazakos, Cordelia Schmid, Josef Sivic*

- 435 Seeing the Trees for the Forest: Rethinking Weakly-Supervised
 * Medical Visual Grounding, *Ta Duc Huy, Duy Anh Huynh, Yutong Xie,*
 ☆ *Yuankai Qi, Qi Chen, Phi Le Nguyen, Sen Kim Tran, Son Lam Phung,*
Anton van den Hengel, Zhibin Liao, Minh-Son To, Johan W. Verjans,
Vu Minh Hieu Phan
- 436 PBCAT: Patch-Based Composite Adversarial Training against
 Physically Realizable Attacks on Object Detection, *Xiao Li,*
Yiming Zhu, Yifan Huang, Wei Zhang, Yingzhe He, Jie Shi, Xiaolin Hu
- 437 Emulating Self-attention with Convolution for Efficient Image
 * Super-Resolution, *Dongheon Lee, Seokju Yun, Youngmin Ro*
- 438 RareCLIP: Rarity-aware Online Zero-shot Industrial Anomaly
 Detection, *Jianfang He, Min Cao, Silong Peng, Qiong Xie*
- 439 MEH: A Multi-Style Dataset and Toolkit for Advancing Egyptian
 Hieroglyph Recognition, *Maksim Golyadkin, Valeria Rubanova,*
Aleksandr Utkov, Dmitry Nikolotov, Ilya Makarov
- 440 Hybrid-Tower: Fine-grained Pseudo-query Interaction and Generation
 for Text-to-Video Retrieval, *Bangxiang Lan, Ruobing Xie,*
Ruixiang Zhao, Xingwu Sun, Zhanhui Kang, Gang Yang, Xirong Li
- 441 Unbiased Missing-modality Multimodal Learning, *Ruiting Dai,*
Chenxi Li, Yandong Yan, Lisi Mo, Ke Qin, Tao He
- 442 ViM-VQ: Efficient Post-Training Vector Quantization for Visual Mamba,
Juncan Deng, Shuaiqing Li, Zeyu Wang, Kedong Xu, Hong Gu, Kejie Huang
- 443 MolParser: End-to-end Visual Recognition of Molecule Structures
 in the Wild, *Xi Fang, Jiankun Wang, Xiaochen Cai, Shangqian Chen,*
Shuwen Yang, Haoyi Tao, Nan Wang, Lin Yao, Linfeng Zhang, Guolin Ke
- 444 SpLiFormer: Enhancing Spiking Transformers with Lateral Inhibition,
Zeqi Zheng, Yanchen Huang, Yingchao Yu, Zizheng Zhu, Junfeng Tang,
Zhaofei Yu, Yaochu Jin
- 445 B-VLLM: A Vision Large Language Model with Balanced Spatio-
 Temporal Tokens, *Zhuqiang Lu, Zhenfei Yin, Mengwei He,*
Zhihui Wang, Zicheng Liu, Zhiyong Wang, Kun Hu
- 446 Iris: Breaking GUI Complexity with Adaptive Focus and Self-Refining,
Zhiqi Ge, Juncheng Li, Xinglei Pang, Minghe Gao, Kaihang Pan, Wang Lin,
Hao Fei, Wenqiao Zhang, Siliang Tang, Yueting Zhuang
- 447 DM-EFS: Dynamically Multiplexed Expanded Features Set Form for
 Robust and Efficient Small Object Detection, *Aashish Sharma*
- 448 DiffTell: A High-Quality Dataset for Describing Image Manipulation
 Changes, *Zonglin Di, Jing Shi, Yifei Fan, Hao Tan, Alexander Black,*
John Collomosse, Yang Liu
- 449 YOLOE: Real-Time Seeing Anything, *Ao Wang, Lihao Liu, Hui Chen,*
Zijia Lin, Jungong Han, Guiguang Ding
- 450 Mixture-of-Scores: Robust Image-Text Data Valuation via Three
 Lines of Code, *Sitong Wu, Haoru Tan, Yukang Chen, Shaofeng Zhang,*
Jingyao Li, Bei Yu, Xiaojuan Qi, Jiaya Jia
- 451 Allowing Oscillation Quantization: Overcoming Solution Space
 Limitation in Low Bit-Width Quantization, *Weiying Xie, Zihan Meng,*
Jitao Ma, Wenjin Guo, Haowei Li, Haonan Qin, Leyuan Fang, Yunsong Li
- 452 FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained
 Flow Models, *Vladimir Kulikov, Matan Kleiner,*
 Inbar Huberman-Spiegelglas, *Tomer Michaeli*
- 453 LayerTracer: Cognitive-Aligned Layered SVG Synthesis via
 Diffusion Transformer, *Yiren Song, Danze Chen, Mike Zheng Shou*

10:45 - 12:45 Demos 5 (Exhibit Hall I)

- CF3: Compact and Fast 3D Feature Fields, *Hyunjoon Lee,*
Joonyou Min, Jaesik Park
- Stable Signer, *Sen Fang*
- Hi3DGen: High-fidelity 3D Geometry Generation from Images via
Normal Bridging, *Chongjie Ye*

11:00 - 13:00 Lunch (Exhibit Hall II / Overflow Seating Rooftop and Third Floor)

13:00 - 14:15 Oral Session 6A: Physical Scene Perception (Exhibit Hall III)

Session Chairs: Young Min Kim, Nianyi Li

🏆 - Award candidate paper ☆ - Outstanding Reviewer

- 1 SuperDec: 3D Scene Decomposition with Superquadrics
Primitives, *Elisabetta Fedele, Boyang Sun, Leonidas Guibas,*
Marc Pollefeys, Francis Engelmann

- 2 Diffusion Image Prior, *Hamadi Chihaoui, Paolo Favaro*
- 3 Spatially-Varying Autofocus, *Yingsi Qin, Aswin C. Sankaranarayanan,*
🏆 *Matthew O'Toole*
- 4 Towards Foundational Models for Single-Chip Radar,
Tianshu Huang, Akarsh Prabhakara, Chuhan Chen, Jay Karhade,
Deva Ramanan, Matthew O'toole, Anthony Rowe
- 5 Event-based Visual Vibrometry, *Xinyu Zhou, Peiqi Duan,*
☆ *Yeliduosi Xiaokaiti, Chao Xu, Boxin Shi*

13:00 - 14:15 Oral Session 6B: Segmentation and Grouping (Kalakaua Ballroom)

Session Chairs: Nathan Jacobs, Yu-Xiong Wang

- 1 CorrCLIP: Reconstructing Patch Correlations in CLIP for Open-Vocabulary
Semantic Segmentation, *Dengke Zhang, Fagui Liu, Quan Tang*
- 2 E-SAM: Training-Free Segment Every Entity Model,
Weiming Zhang, Dingwen Xiao, Lei Chen, Lin Wang
- 3 Online Reasoning Video Segmentation with Just-in-Time Digital
Twins, *Yiqing Shen, Bohan Liu, Chenjia Li,*
Lalithkumar Seenivasan, Mathias Unberath
- 4 Easy3D: A Simple Yet Effective Method for 3D Interactive
Segmentation, *Andrea Simonelli, Norman Müller, Peter Kontschieder*
- 5 ForestFormer3D: A Unified Framework for End-to-End Segmentation
of Forest LiDAR 3D Point Clouds, *Binbin Xiang, Maciej Wielgosz,*
Stefano Puliti, Kamil Král, Martin Krůček, Azim Missarov,
Rasmus Astrup

14:00 - 14:30 Poster Setup (Exhibit Hall I)

14:30 - 16:30 Poster Session 6 & Exhibit Hall with Coffee Break (Exhibit Hall I)

* - Highlight paper 🏆 - Award candidate paper

○ - Oral Paper

- 1 X2-Gaussian: 4D Radiative Gaussian Splatting for Continuous-time
Tomographic Reconstruction, *Weihao Yu, Yuanhao Cai, Ruyi Zha,*
Zhiwen Fan, Chenxin Li, Yixuan Yuan
- 2 Inverse Image-Based Rendering for Light Field Generation from
Single Images, *Hyunjun Jung, Hae-Gon Jeon*
- 3 HyperGCT: A Dynamic Hyper-GNN-Learned Geometric Constraint
for 3D Registration, *Xiyu Zhang, Jiayi Ma, Jianwei Guo, Wei Hu,*
Zhaoshuai Qi, Fei Hui, Jiaqi Yang, Yanning Zhang
- 4 CounterPC: Counterfactual Feature Realignment for
* Unsupervised Domain Adaptation on Point Clouds,
Feng Yang, Yichao Cao, Xiu Su, Dan Niu, Xuanpeng Li
- 5 AD-GS: Object-Aware B-Spline Gaussian Splatting for Self-Supervised
Autonomous Driving, *Jiawei Xu, Kai Deng, Zexin Fan, Shenlong Wang,*
Jin Xie, Jian Yang
- 6 EvaGaussians: Event Stream Assisted Gaussian Splatting from Blurry
Images, *Wangbo Yu, Chaoran Feng, Jianing Li, Jiye Tang, Jiashu Yang,*
Zhenyu Tang, Meng Cao, Xu Jia, Yuchao Yang, Li Yuan, Yonghong Tian
- 7 Axis-level Symmetry Detection with Group-Equivariant
Representation, *Wongyun Yu, Ahyun Seo, Minsu Cho*
- 8 STD-GS: Exploring Frame-Event Interaction for SpatioTemporal-
Disentangled Gaussian Splatting to Reconstruct High-Dynamic
Scene, *Hanyu Zhou, Haonan Wang, Haoyue Liu, Yuxing Duan,*
Luxin Yan, Gim Hee Lee
- 9 Monocular Semantic Scene Completion via Masked Recurrent
Networks, *Xuzhi Wang, Xinran Wu, Song Wang, Lingdong Kong,*
Ziping Zhao
- 10 ORION: A Holistic End-to-End Autonomous Driving Framework
by Vision-Language Instructed Action Generation, *Haoyu Fu,*
Diankun Zhang, Zongchuang Zhao, Jianfeng Cui, Dingkan Liang,
Chong Zhang, Dingyuan Zhang, Hongwei Xie, Bing Wang, Xiang Bai
- 11 All in One: Visual-Description-Guided Unified Point Cloud Segmentation,
Zongyan Han, Mohamed El Amine Boudjoghra, Jiahua Dong,
Jinhong Wang, Rao Muhammad Anwer
- 12 Bolt3D: Generating 3D Scenes in Seconds, *Stanislaw Szymonowicz,*
Jason Y. Zhang, Pratul Srinivasan, Ruiqi Gao, Arthur Brussee,
Aleksander Holynski, Ricardo Martin-Brualla, Jonathan T. Barron,
Philipp Henzler
- 13 PossLoss: A Reliable and Sensitive Facial Landmark Detection Loss

- Function, *Qikui Zhu*
- 14 Unsupervised RGB-D Point Cloud Registration for Scenes with Low Overlap and Photometric Inconsistency, *Yejun Shou, Haocheng Wang, Lingfeng Shen, Qian Zheng, Gang Pan, Yanlong Cao*
- 15 Semantic Causality-Aware Vision-Based 3D Occupancy Prediction, *Dubing Chen, Huan Zheng, Yucheng Zhou, Xianfei Li, Wenlong Liao, Tao He, Pai Peng, Jianbing Shen*
- 16 U-ViLAR: Uncertainty-Aware Visual Localization for Autonomous Driving via Differentiable Association and Registration, *Xiaofan Li, Zhihao Xu, Chenming Wu, Zhao Yang, Yumeng Zhang, Jiang-Jiang Liu, Haibao Yu, Xiaoqing Ye, Yuan Wang, Shirui Li, Xun Sun, Ji Wan, Jun Wang*
- 17 Benchmarking Burst Super-Resolution for Polarization Images: Noise Dataset and Analysis, *Inseung Hwang, Kiseok Choi, Hyunho Ha, Min H. Kim*
- 18 Group Inertial Poser: Multi-Person Pose and Global Translation from Sparse Inertial Sensors and Ultra-Wideband Ranging, *Ying Xue, Jiayi Jiang, Rayan Armani, Dominik Hollidt, Yi-Chi Liao, Christian Holz*
- 19 DreamCube: RGB-D Panorama Generation via Multi-plane Synchronization, *Yukun Huang, Yanning Zhou, Jianan Wang, Kaiyi Huang, Xihui Liu*
- 20 OcRFDet: Object-Centric Radiance Fields for Multi-View 3D Object Detection in Autonomous Driving, *Mingqian Ji, Shanshan Zhang, Jian Yang*
- 21 GaussianFlowOcc: Sparse and Weakly Supervised Occupancy Estimation using Gaussian Splatting and Temporal Flow, *Simon Boeder, Fabian Gigengack, Benjamin Risse*
- 22 RESCUE: Crowd Evacuation Simulation via Controlling SDM-
* United Characters, *Xiaolin Liu, Tianyi Zhou, Hongbo Kang, Jian Ma, Ziwen Wang, Jing Huang, Wenguo Weng, Yu-Kun Lai, Kun Li*
- 23 SG-LDM: Semantic-Guided LiDAR Generation via Latent-Aligned Diffusion, *ZhengKang Xiang, Zizhao Li, Amir Khodabandeh, Kourosh Khoshelham*
- 24 LookOut: Real-World Humanoid Egocentric Navigation, *Boxiao Pan, Adam W. Harley, Francis Engelmann, C. Karen Liu, Leonidas J. Guibas*
- 25 PointGAC: Geometric-Aware Codebook for Masked Point Modeling, *☆ Abiao Li, Chenlei Lv, Yuming Fang, Yifan Zuo, Jian Zhang, Guofeng Mei*
- 26 Statistical Confidence Rescoring for Robust 3D Scene Graph Generation from Multi-View Images, *Qi Xun Yeo, Yanyan Li, Gim Hee Lee*
- 27 PRM: Photometric Stereo based Large Reconstruction Model, ** Wenhao Ge, Jiantao Lin, Guibao Shen, Jiawei Feng, Tao Hu, Xinli Xu, ☆ Ying-Cong Chen*
- 28 4D Gaussian Splatting SLAM, *Yanyan Li, Youxu Fang, Zunjie Zhu, Kunyi Li, Yong Ding, Federico Tombari*
- 29 Billboard Splatting (BBSplat): Learnable Textured Primitives for *☆ Novel View Synthesis, David Svitov, Pietro Morerio, Lourdes Agapito, Alessio Del Bue*
- 30 Generalizable Non-Line-of-Sight Imaging with Learnable Physical Priors, *Shida Sun, Yue Li, Yueyi Zhang, Zhiwei Xiong*
- 31 Hi3DGen: High-fidelity 3D Geometry Generation from Images via Normal Bridging, *Chongjie Ye, Yushuang Wu, Ziteng Lu, Jiahao Chang, Xiaoyang Guo, Jiaqing Zhou, Hao Zhao, Xiaoguang Han*
- 32 Baking Gaussian Splatting into Diffusion Denoiser for Fast and Scalable *☆ Single-stage Image-to-3D Generation and Reconstruction, Yuanhao Cai, He Zhang, Kai Zhang, Yixun Liang, Mengwei Ren, Fujun Luan, Qing Liu, Soo Ye Kim, Jianming Zhang, Zhifei Zhang, Yuqian Zhou, Yulun Zhang, Xiaokang Yang, Zhe Lin, Alan Yuille*
- 33 Discretized Gaussian Representation for Tomographic Reconstruction, *Shaokai Wu, Yuxiang Lu, Yapan Guo, Wei Ji, Suizhi Huang, Fengyu Yang, Shalayiding Sirejiding, Qichen He, Jing Tong, Yanbiao Ji, Yue Ding, Hongtao Lu*
- 34 SuperMat: Physically Consistent PBR Material Estimation at *☆ Interactive Rates, Yijia Hong, Yuan-Chen Guo, Ran Yi, Yulong Chen, Yan-Pei Cao, Lizhuang Ma*
- 35 RI3D: Few-Shot Gaussian Splatting With Repair and Inpainting Diffusion Priors, *Avinash Paliwal, Xilong Zhou, Wei Ye, Jinhui Xiong, Rakesh Ranjan, Nima Khademi Kalantari*
- 36 Dual-S3D: Hierarchical Dual-Path Selective SSM-CNN for High-Fidelity Implicit Reconstruction, *Luoxi Zhang, Pragyansh Shrestha, Yu Zhou, Chun Xie, Itaru Kitahara*
- 37 FastPoint: Accelerating 3D Point Cloud Model Inference via Sample Point Distance Prediction, *Donghyun Lee, Dawoon Jeong, Jae W. Lee, Hongil Yoon*
- 38 Self-Calibrating Gaussian Splatting for Large Field-of-View Reconstruction, ** Youming Deng, Wengqi Xian, Guandao Yang, Leonidas Guibas, Gordon Wetzstein, Steve Marschner, Paul Debevec*
- 39 RobuSTereo: Robust Zero-Shot Stereo Matching under Adverse Weather, *Yuran Wang, Yingping Liang, Yutao Hu, Ying Fu*
- 40 Transformer-based Tooth Alignment Prediction with Occlusion and Collision Constraints, *Zhenxing Dong, Jiazhou Chen*
- 41 Gaussian Splatting with Discretized SDF for Relightable Assets, *Zuo-Liang Zhu, Jian Yang, Beibei Wang*
- 42 MMGeo: Multimodal Compositional Geo-Localization for UAVs, *Yuxiang Ji, Boyong He, Zhuoyue Tan, Liaoni Wu*
- 43 AdaptiveAE: An Adaptive Exposure Strategy for HDR Capturing in Dynamic Scenes, *Tianyi Xu, Fan Zhang, Boxin Shi, Tianfan Xue, Yujin Wang*
- 44 Large Scene Generation with Cube-Absorb Discrete Diffusion, *Qianjiang Hu, Wei Hu*
- 45 SynAD: Enhancing Real-World End-to-End Autonomous Driving Models through Synthetic Data Integration, *Jongsuk Kim, Jaeyoung Lee, Gyojin Han, Dong-Jae Lee, Minki Jeong, Junmo Kim*
- 46 Benchmarking Egocentric Visual-Inertial SLAM at City Scale, ** Anusha Krishnan, Shaohui Liu, Paul-Edouard Sarlin, Oscar Gentilhomme, David Caruso, Maurizio Monge, Richard Newcombe, Jakob Engel, Marc Pollefeys*
- 47 Robust Unfolding Network for HDR Imaging with Modulo Cameras, *Zhile Chen, Hui Ji*
- 48 Neural Shell Texture Splatting: More Details and Fewer Primitives, *Xin Zhang, Anpei Chen, Jincheng Xiong, Pinxuan Dai, Yujun Shen, Weiwei Xu*
- 49 Gaussian-based World Model: Gaussian Priors for Voxel-Based Occupancy Prediction and Future Motion Prediction, *Tuo Feng, Wenguan Wang, Yi Yang*
- 50 Momentum-GS: Momentum Gaussian Self-Distillation for High-Quality Large Scene Reconstruction, *Jixuan Fan, Wanhua Li, Yifei Han, Tianru Dai, Yansong Tang*
- 51 JointDiT: Enhancing RGB-Depth Joint Modeling with Diffusion Transformers, *Kwon Byung-Ki, Qi Dai, Lee Hyoseok, Chong Luo, Tae-Hyun Oh*
- 52 A Real-world Display Inverse Rendering Dataset, *Seokjun Choi, Hoon-Gyu Chung, Yujin Jeon, Giljoo Nam, Seung-Hwan Baek*
- 53 DAA*: Deep Angular A Star for Image-based Path Planning, *Zhiwei Xu*
- 54 Neural Compression for 3D Geometry Sets, *Siyu Ren, Junhui Hou, Weiyao Lin, Wenping Wang*
- 55 Long-term Traffic Simulation with Interleaved Autoregressive Motion and Scenario Generation, *☆ Xiuyu Yang, Shuhan Tan, Philipp Krähenbühl*
- 56 RCTDistill: Cross-Modal Knowledge Distillation Framework for Radar-Camera 3D Object Detection with Temporal Fusion, *Geonho Bang, Minjae Seong, Jisong Kim, Geunju Baek, Daye Oh, Junhyung Kim, Junho Koh, Jun Won Choi*
- 57 MagicCity: Geometry-Aware 3D City Generation from Satellite Imagery with Multi-View Consistency, *Xingbo Yao, Xuanmin Wang, Hao Wu, Chengliang Ping, Doudou Zhang, Hui Xiong*
- 58 GCRayDiffusion: Pose-Free Surface Reconstruction via Geometric *☆ Consistent Ray Diffusion, Li-Heng Chen, Zi-Xin Zou, Chang Liu, Tianjiao Jing, Yan-Pei Cao, Shi-Sheng Huang, Hongbo Fu, Hua Huang*
- 59 GSRecon: Efficient Generalizable Gaussian Splatting for Surface Reconstruction from Sparse Views, *Hang Yang, Le Hui, Jianjun Qian, Jin Xie, Jian Yang*
- 60 GeoDistill: Geometry-Guided Self-Distillation for Weakly Supervised Cross-View Localization, *Shaowen Tong, Zimin Xia, Alexandre Alahi, Xuming He, Yujiao Shi*
- 61 REPARO: Compositional 3D Assets Generation with Differentiable 3D Layout Alignment, *Haonan Han, Rui Yang, Huan Liao, Jiankai Xing, Zunnan Xu, Xiaoming Yu, Junwei Zha, Xiu Li, Wanhua Li*
- 62 Towards Safer and Understandable Driver Intention Prediction, *☆ Mukilan Karuppasamy, Shankar Gangisetty, Shyam Nandan Rai, Carlo Masone, C V Jawahar*
- 63 Revisiting Point Cloud Completion: Are We Ready For The Real-

- World?, *Stuti Pathak, Prashant Kumar, Dheeraj Baiju, Nicholas Mboga, Gunther Steenackers, Rudi Penne*
- 64 V2XPnP: Vehicle-to-Everything Spatio-Temporal Fusion for Multi-Agent Perception and Prediction, *Zewei Zhou, Hao Xiang, Zhaoliang Zheng, Seth Z. Zhao, Mingyue Lei, Yun Zhang, Tianhui Cai, Xinyi Liu, Johnson Liu, Maheswari Bajji, Xin Xia, Zhiyu Huang, Bolei Zhou, Jiaqi Ma*
- 65 InstaDrive: Instance-Aware Driving World Models for Realistic and Consistent Video Generation, *Zhuoran Yang, Xi Guo, Chenjing Ding, Chiyu Wang, Wei Wu, Yanyong Zhang*
- 66 NormalLoc: Visual Localization on Textureless 3D Models using Surface Normals, *Jiro Abe, Gaku Nakano, Kazumine Ogura*
- 67 EmbodiedSplat: Personalized Real-to-Sim-to-Real Navigation with Gaussian Splats from a Mobile Device, *Gunjan Chhablani, Xiaomeng Ye, Muhammad Zubair Irshad, Zsolt Kira*
- 68 FreeSplat: Pose-free Gaussian Splatting for Sparse-view 3D Reconstruction, *Jiale Xu, Shenghua Gao, Ying Shan*
- 69 UniMLVG: Unified Framework for Multi-view Long Video Generation with Comprehensive Control Capabilities for Autonomous Driving, *Rui Chen, Zehuan Wu, Yichen Liu, Yuxin Guo, Jingcheng Ni, Haifeng Xia, Siyu Xia*
- 70 INSTINCT: Instance-Level Interaction Architecture for Query-Based Collaborative Perception, *Yunjia Xu, Lingzhi Li, Jin Wang, Yupeng Ouyang, Benyuan Yang*
- 71 ForeSight: Multi-View Streaming Joint Object Detection and Trajectory Forecasting, *Sandro Papais, Letian Wang, Brian Cheong, Steven L. Waslander*
- 72 NGD: Neural Gradient Based Deformation for Monocular Garment Reconstruction, *Soham Dasgupta, Shanthika Naik, Preet Savalia, Sujay Kumar Ingle, Avinash Sharma*
- 73 Compression of 3D Gaussian Splatting with Optimized Feature Planes and Standard Video Codecs, *Soonbin Lee, Fangwen Shu, Yago Sanchez, Thomas Schierl, Cornelius Hellge*
- 74 Spatially-Varying Autofocus, *Yingsi Qin,*
 🏆 Aswin C. Sankaranarayanan, Matthew O'Toole
- 75 Event-based Visual Vibrometry, *Xinyu Zhou, Peiqi Duan,*
 🌟 Yeliduosi Xiaokaiti, Chao Xu, Boxin Shi
- 76 Beyond One Shot, Beyond One Perspective: Cross-View and Long-Horizon Distillation for Better LiDAR Representations, *Xiang Xu, Lingdong Kong, Song Wang, Chuanwei Zhou, Qingshan Liu*
- 77 BézierGS: Dynamic Urban Scene Reconstruction with Bézier Curve Gaussian Splatting, *Zipei Ma, Junzhe Jiang, Yurui Chen, Li Zhang*
- 78 Lifting the Structural Morphing for Wide-Angle Images Rectification: Unified Content and Boundary Modeling, *Wenting Luan, Siqi Lu, Yongbin Zheng, Wanying Xu, Lang Nie, Zongtan Zhou, Kang Liao*
- 79 Global Regulation and Excitation via Attention Tuning for Stereo Matching, *Jiahao Li, Xinhong Chen, Zhengmin Jiang, Qian Zhou, Yung-Hui Li, Jianping Wang*
- 80 Global-Aware Monocular Semantic Scene Completion with State Space Models, *Shijie Li, Zhongyao Cheng, Rong Li, Shuai Li, Juergen Gall, Xun Xu, Xulei Yang*
- 81 UniOcc: A Unified Benchmark for Occupancy Forecasting and Prediction in Autonomous Driving, *Yuping Wang, Xiangyu Huang, Xiaokang Sun, Mingxuan Yan, Shuo Xing, Zhengzhong Tu, Jiachen Li*
- 82 Free4D: Tuning-free 4D Scene Generation with Spatial-Temporal Consistency, *Tianqi Liu, Zihao Huang, Zhaoxi Chen, Guangcong Wang, Shoukang Hu, Liao Shen, Huiqiang Sun, Zhiguo Cao, Wei Li, Ziwei Liu*
- 83 MuGS: Multi-Baseline Generalizable Gaussian Splatting Reconstruction, *Yaopeng Lou, Liao Shen, Tianqi Liu, Jiaqi Li, Zihao Huang, Huiqiang Sun, Zhiguo Cao*
- 84 S3R-GS: Streamlining the Pipeline for Large-Scale Street Scene Reconstruction, *Guangting Zheng, Jiajun Deng, Xiaomeng Chu, Yu Yuan, Houqiang Li, Yanyong Zhang*
- 85 HiP-AD: Hierarchical and Multi-Granularity Planning with Deformable Attention for Autonomous Driving in a Single Decoder, *Yingqi Tang, Zhuoran Xu, Zhaotie Meng, Erkang Cheng*
- 86 RayletDF: Raylet Distance Fields for Generalizable 3D Surface
 * Reconstruction from Point Clouds or Gaussians, *Shenxing Wei, Jinxi Li, Yafei Yang, Siyuan Zhou, Bo Yang*
- 87 GAP: Gaussianize Any Point Clouds with Text Guidance, *Weiwei Zhang, Junsheng Zhou, Haotian Geng, Wenyuan Zhang, Yu-Shen Liu*
- 88 Semantic-guided Camera Ray Regression for Visual Localization, *Yesheng Zhang, Xu Zhao*
- 89 SketchSplat: 3D Edge Reconstruction via Differentiable Multi-view Sketch Splatting, *Haiyang Ying, Matthias Zwicker*
- 90 Polarimetric Neural Field via Unified Complex-Valued Wave Representation, *Chu Zhou, Yixin Yang, Junda Liao, Heng Guo, Boxin Shi, Imari Sato*
- 91 High-Precision 3D Measurement of Complex Textured Surfaces Using Multiple Filtering Approach, *Yuchong Chen, Jian Yu, Shaoyan Gai, Zeyu Cai, Feipeng Da*
- 92 OCSplats: Observation Completeness Quantification and Label Noise Separation in 3DGS, *Han Ling, Xian Xu, Yinghui Sun, Quansen Sun*
- 93 VMem: Consistent Interactive Video Scene Generation with
 * Surfel-Indexed View Memory, *Runjia Li, Philip Torr, Andrea Vedaldi, Tomas Jakab*
- 94 AutoScape: Geometry-Consistent Long-Horizon Scene Generation,
 ☆ Jiacheng Chen, Ziyu Jiang, Mingfu Liang, Bingbing Zhuang, Jong-Chyi Su, Sparsh Garg, Ying Wu, Manmohan Chandraker
- 95 From Gallery to Wrist: Realistic 3D Bracelet Insertion in Videos, *Chenjian Gao, Lihe Ding, Rui Han, Zhanpeng Huang, Zibin Wang, Tianfan Xue*
- 96 Street Gaussians without 3D Object Tracker, *Ruida Zhang, Chengxi Li, Chenyangguang Zhang, Xingyu Liu, Haili Yuan, Yanyan Li, Xiangyang Ji, Gim Hee Lee*
- 97 RogSplat: Robust Gaussian Splatting via Generative Priors, *Hanyang Kong, Xingyi Yang, Xinchao Wang*
- 98 HiNeuS: High-fidelity Neural Surface Mitigating Low-texture and
 * Reflective Ambiguity, *Yida Wang, Xueyang Zhang, Kun Zhan, Peng Jia, Xianpeng Lang*
- 99 RGE-GS: Reward-Guided Expansive Driving Scene Reconstruction via Diffusion Priors, *Sicong Du, Jiarun Liu, Qifeng Chen, Hao-Xiang Chen, Tai-Jiang Mu, Sheng Yang*
- 100 Scene Coordinate Reconstruction Priors, *Wenjing Bian,*
 ☆ Axel Barroso-Laguna, Tommaso Cavallari, Victor Adrian Prisacariu, Eric Brachmann
- 101 Diff2I2P: Differentiable Image-to-Point Cloud Registration with Diffusion Prior, *Juncheng Mu, Chengwei Ren, Weixiang Zhang, Liang Pan, Xiao-Ping Zhang, Yue Gao*
- 102 Resonance: Learning to Predict Social-Aware Pedestrian Trajectories as Co-Vibrations, *Conghao Wong, Ziqian Zou, Beihao Xia*
- 103 GaussianUpdate: Continual 3D Gaussian Splatting Update for Changing Environments, *Lin Zeng, Boming Zhao, Jiarui Hu, Xujie Shen, Ziqiang Dang, Hujun Bao, Zhaopeng Cui*
- 104 I2-World: Intra-Inter Tokenization for Efficient Dynamic 4D Scene Forecasting, *Zhimin Liao, Ping Wei, Ruijie Zhang, Shuaijia Chen, Haoxuan Wang, Ziyang Ren*
- 105 InsideOut: Integrated RGB-Radiative Gaussian Splatting for Comprehensive 3D Object Representation, *Jungmin Lee, Seonghyuk Hong, Juyong Lee, Jaeyoon Lee, Jongwon Choi*
- 106 Serialization based Point Cloud Oversegmentation, *Chenghui Lu, Jianlong Kwan, Dilong Li, Ziyi Chen, Haiyan Guan*
- 107 StreamGS: Online Generalizable Gaussian Splatting Reconstruction for Unposed Image Streams, *Yang Li, Jinglu Wang, Lei Chu, Xiao Li, Shiu-Hong Kao, Ying-Cong Chen, Yan Lu*
- 108 RIOcc: Efficient Cross-Modal Fusion Transformer with Collaborative Feature Refinement for 3D Semantic Occupancy Prediction, *Baojie Fan, Xiaotian Li, Yuhao Zhou, Yuyu Jiang, Jiandong Tian, Huijie Fan*
- 109 Lightweight Gradient-Aware Upscaling of 3D Gaussian Splatting Images, *Simon Niedermayr, Christoph Neuhauser, Rüdiger Westermann*
- 110 TeethGenerator: A two-stage framework for paired pre- and post-orthodontic 3D dental data generation, *Changsong Lei, Yaqian Liang, Shaofeng Wang, Jiajia Dai, Yong-Jin Liu*
- 111 Online Language Splatting, *Saimouli Katragadda, Cho-Ying Wu, Yuliang Guo, Xinyu Huang, Guoquan Huang, Liu Ren*
- 112 Probabilistic Inertial Poser (ProIP): Uncertainty-aware Human Motion Modeling from Sparse Inertial Sensors, *Min Kim, Younho Jeon, Sungho Jo*
- 113 Towards Accurate and Efficient 3D Object Detection for Autonomous

- Driving: A Mixture of Experts Computing System on Edge, *Linshen Liu, Boyan Su, Junyue Jiang, Guanlin Wu, Cong Guo, Ceyu Xu, Hao Frank Yang*
- 114 Degradation-Modeled Multipath Diffusion for Tunable Metalens
* Photography, *Jianing Zhang, Jiayi Zhu, Feiyu Ji, Xiaokang Yang, Xiaoyun Yuan*
- 115 GS-Occ3D: Scaling Vision-only Occupancy Reconstruction with Gaussian Splatting, *Baijun Ye, Minghui Qin, Saining Zhang, Moonjun Gong, Shaoting Zhu, Hao Zhao, Hang Zhao*
- 116 MetaScope: Optics-Driven Neural Network for Ultra-Micro
* Metalens Endoscopy, *Wuyang Li, Wentao Pan, Xiaoyuan Liu, Zhendong Luo, Chenxin Li, Hengyu Liu, Din Ping Tsai, Mu Ku Chen, Yixuan Yuan*
- 117 CoLMDriver: LLM-based Negotiation Benefits Cooperative Autonomous Driving, *Changxing Liu, Genjia Liu, Zijun Wang, Jinchang Yang, Siheng Chen*
- 118 A View-consistent Sampling Method for Regularized Training of Neural Radiance Fields, *Aoxiang Fan, Corentin Dumery, Nicolas Talabot, Pascal Fua*
- 119 Free-running vs Synchronous: Single-Photon Lidar for High-flux 3D Imaging, *Ruangrawee Kitichotkul, Shashwath Bharadwaj, Joshua Rapp, Yanting Ma, Alexander Mehta, Vivek K Goyal*
- 120 Mitigating Geometric Degradation in Fast DownSampling via FastAdapter for Point Cloud Segmentation, *Shuofeng Sun, Haibin Yan*
- 121 Noise2Score3D: Tweedie's Approach for Unsupervised Point Cloud Denoising, *Xiangbin Wei, Yuanfeng Wang, Ao Xu, Lingyu Zhu, Dongyong Sun, Keren Li, Yang Li, Qi Qin*
- 122 ArchiSet: Benchmarking Editable and Consistent Single-View 3D Reconstruction of Buildings with Specific Window-to-Wall Ratios, *Jun Yin, Pengyu Zeng, Licheng Shen, Miao Zhang, Jing Zhong, Yuxing Han, Shuai Lu*
- 123 ClaraVid: A Holistic Scene Reconstruction Benchmark From Aerial Perspective With Delentropy-Based Complexity Profiling, *Radu Beche, Sergiu Nedevschi*
- 124 Discontinuity-aware Normal Integration for Generic Central
* Camera Models, *Francesco Milano, Manuel López-Antequera, Naina Dhinra, Roland Siegwart, Robert Thiel*
- 125 Outdoor Monocular SLAM with Global Scale-Consistent 3D Gaussian Pointmaps, *Chong Cheng, Sicheng Yu, Zijian Wang, Yifan Zhou, Hao Wang*
- 126 SEHDR: Single-Exposure HDR Novel View Synthesis via 3D Gaussian
☆ Bracketing, *Yiyu Li, Haoyuan Wang, Ke Xu, Gerhard Petrus Hancke, Rynson W.H. Lau*
- 127 Omni-scene Perception-oriented Point Cloud Geometry Enhancement for Coordinate Quantization, *Wang Liu, Wei Gao*
- 128 SL2A-INR: Single-Layer Learnable Activation for Implicit Neural Representation, *Reza Rezaeian, Moein Heidari, Reza Azad, Dorit Merhof, Hamid Soltanian-Zadeh, Ilker Hacihaliloglu*
- 129 TARS: Traffic-Aware Radar Scene Flow Estimation, *Jialong Wu, Marco Braun, Dominic Spata, Matthias Rottmann*
- 130 DoppDrive: Doppler-Driven Temporal Aggregation for Improved Radar Object Detection, *Yuval Haitman, Oded Bialer*
- 131 FVGen: Accelerating Novel-View Synthesis with Adversarial Video Diffusion Distillation, *Wenbin Teng, Gonglin Chen, Haiwei Chen, Yajie Zhao*
- 132 Simulating Dual-Pixel Images From Ray Tracing For Depth Estimation, *Fengchen He, Dayang Zhao, Hao Xu, Tingwei Quan, Shaoqun Zeng*
- 133 Leaps and Bounds: An Improved Point Cloud Winding Number
☆ Formulation for Fast Normal Estimation and Surface Reconstruction, *Chamin Hwa Koneputugodage, Dylan Campbell, Stephen Gould*
- 134 GVDepth: Zero-Shot Monocular Depth Estimation for Ground Vehicles based on Probabilistic Cue Fusion, *Karlo Koledić, Luka Petrović, Ivan Marković, Ivan Petrović*
- 135 InterGSEdit: Interactive 3D Gaussian Splatting Editing with 3D Geometry-Consistent Attention Prior, *Minghao Wen, Shengjie Wu, Kangkan Wang, Dong Liang*
- 136 From Gaze to Movement: Predicting Visual Attention for Autonomous Driving Human-Machine Interaction based on Programmatic Imitation Learning, *Yexin Huang, Yongbin Lin, Lishengsa Yue, Zhihong Yao, Jie Wang*
- 137 Harnessing Text-to-Image Diffusion Models for Point Cloud Self-Supervised Learning, *Yiyang Chen, Shanshan Zhao, Lunhao Duan, Changxing Ding, Dacheng Tao*
- 138 OD-RASE: Ontology-Driven Risk Assessment and Safety Enhancement for Autonomous Driving, *Kota Shimomura, Masaki Nambata, Atsuya Ishikawa, Ryota Mimura, Koki Inoue, Takayoshi Yamashita, Takayuki Kawabuchi*
- 139 MDP-Omni: Parameter-free Multimodal Depth Prior-based Sampling for Omnidirectional Stereo Matching, *Eunjin Son, HyungGi Jo, Woogyong Kwon, Sang Jun Lee*
- 140 DistillDrive: End-to-End Multi-Mode Autonomous Driving Distillation by Isomorphic Hetero-Source Planning Model, *Rui Yu, Xianghang Zhang, Runkai Zhao, Huaicheng Yan, Meng Wang*
- 141 EDM: Efficient Deep Feature Matching,
* *Xi Li, Tong Rao, Cihui Pan*
- 142 Robust and Efficient 3D Gaussian Splatting for Urban Scene Reconstruction, *Zhensheng Yuan, Haozhi Huang, Zhen Xiong, Di Wang, Guanghua Yang*
- 143 SuperDec: 3D Scene Decomposition with Superquadrics
○ Primitives, *Elisabetta Fedele, Boyang Sun, Leonidas Guibas, Marc Pollefeys, Francis Engelmann*
- 144 CorrCLIP: Reconstructing Patch Correlations in CLIP for Open-
○ Vocabulary Semantic Segmentation, *Dengke Zhang, Fagui Liu, Quan Tang*
- 145 GS-ID: Illumination Decomposition on Gaussian Splatting via Adaptive Light Aggregation and Diffusion-Guided Material Priors, *Kang Du, Zhihao Liang, Yulin Shen, Zeyu Wang*
- 146 NeRF Is a Valuable Assistant for 3D Gaussian Splatting, *Shuangkang Fang, I-Chao Shen, Takeo Igarashi, Yufeng Wang, ZeSheng Wang, Yi Yang, Wenrui Ding, Shuchang Zhou*
- 147 UniGS: Modeling Unitary 3D Gaussians for Novel View Synthesis from Sparse-view Images, *Jiamin Wu, Kenkun Liu, Xiaoke Jiang, Yuan Yao, Lei Zhang*
- 148 Hierarchy UGP: Hierarchy Unified Gaussian Primitive for Large-Scale Dynamic Scene Reconstruction, *Hongyang Sun, Qinglin Yang, Jiawei Wang, Zhen Xu, Chen Liu, Yida Wang, Kun Zhan, Hujun Bao, Xiaowei Zhou, Sida Peng*
- 149 TOTP: Transferable Online Pedestrian Trajectory Prediction with Temporal-Adaptive Mamba Latent Diffusion, *Ziyang Ren, Ping Wei, Shangqi Deng, Haowen Tang, Jiapeng Li, Huan Li*
- 150 AR-1-to-3: Single Image to Consistent 3D Object via Next-View Prediction, *Xuying Zhang, Yupeng Zhou, Kai Wang, Yikai Wang, Zhen Li, Shaohui Jiao, Daquan Zhou, Qibin Hou, Ming-Ming Cheng*
- 151 UnMix-NeRF: Spectral Unmixing Meets Neural Radiance Fields, *Fabian Perez, Sara Rojas, Carlos Hinojosa, Hoover Rueda-Chacón, Bernard Ghanem*
- 152 MaterialMVP: Illumination-Invariant Material Generation via
* Multi-view PBR Diffusion, *Zebin He, Mingxin Yang, Shuhui Yang, Yixuan Tang, Tao Wang, Kaihao Zhang, Guanying Chen, Yuhong Liu, Jie Jiang, Chunchao Guo, Wenhan Luo*
- 153 PerLDiff: Controllable Street View Synthesis Using Perspective-Layout Diffusion Model, *Jinhua Zhang, Hualian Sheng, Sijia Cai, Bing Deng, Qiao Liang, Wen Li, Ying Fu, Jieping Ye, Shuhang Gu*
- 154 7DGS: Unified Spatial-Temporal-Angular Gaussian Splatting, *Zhongpai Gao, Benjamin Planche, Meng Zheng, Anwesa Choudhuri, Terrence Chen, Ziyang Wu*
- 155 StochasticSplats: Stochastic Rasterization for Sorting-Free 3D
☆ Gaussian Splatting, *Shakiba Kheradmand, Delio Vicini, George Kopanas, Dmitry Lagun, Kwang Moo Yi, Mark Matthews, Andrea Tagliasacchi*
- 156 MS3D: High-Quality 3D Generation via Multi-Scale Representation Modeling, *Guan Luo, Jianfeng Zhang*
- 157 DASH: 4D Hash Encoding with Self-Supervised Decomposition for Real-Time Dynamic Scene Rendering, *Jie Chen, Zhangchi Hu, Peixi Wu, Huyue Zhu, Hebei Li, Xiaoyan Sun*
- 158 EmbodiedOcc: Embodied 3D Occupancy Prediction for Vision-based Online Scene Understanding, *Yuqi Wu, Wenzhao Zheng, Sicheng Zuo, Yuanhui Huang, Jie Zhou, Jiwen Lu*
- 159 TurboReg: TurboClique for Robust and Efficient Point Cloud

- Registration, *Shaocheng Yan, Pengcheng Shi, Zhenjun Zhao, Kaixin Wang, Kuang Cao, Ji Wu, Jiayuan Li*
- 160 ConsistentCity: Semantic Flow-guided Occupancy DiT for Temporally Consistent Driving Scene Synthesis, *Benjin Zhu, Xiaogang Wang, Hongsheng Li*
- 161 Efficient Spiking Point Mamba for Point Cloud Analysis, *Peixi Wu, Bosong Chai, Menghua Zheng, Wei Li, Zhangchi Hu, Jie Chen, Zheyu Zhang, Hebei Li, Xiaoyan Sun*
- 162 SpatialSplat: Efficient Semantic 3D from Sparse Unposed Images, *Yu Sheng, Jiajun Deng, Xinran Zhang, Yu Zhang, Bei Hua, Yanyong Zhang, Jianmin Ji*
- 163 CoMoGaussian: Continuous Motion-Aware Gaussian Splatting from Motion-Blurred Images, *Jungho Lee, Donghyeong Kim, Dogyoon Lee, Suhwan Cho, Minhyeok Lee, Wonjoon Lee, Taeh Kim, Dongyoon Wee, Sangyoun Lee*
- 164 Splat-based 3D Scene Reconstruction with Extreme Motion-blur, *Hyeonjoong Jang, Dongyoung Choi, Donggun Kim, Woohyun Kang, Min H. Kim*
- 165 Generalized and Efficient 2D Gaussian Splatting for Arbitrary-scale Super-Resolution, *Du Chen, Liyi Chen, Zhengqiang Zhang, Lei Zhang*
- 166 Visual Surface Wave Elastography: Revealing Subsurface Physical Properties via Visible Surface Waves, *Alexander C. Ogren, Berthy T. Feng, Jihoon Ahn, Katherine L. Bouman, Chiara Daraio*
- 167 GaRe: Relightable 3D Gaussian Splatting for Outdoor Scenes from Unconstrained Photo Collections, *Haiyang Bai, Jiaqi Zhu, Songru Jiang, Wei Huang, Tao Lu, Yuanqi Li, Jie Guo, Runze Fu, Yanwen Guo, Lijun Chen*
- 168 PolarAnything: Diffusion-based Polarimetric Image Synthesis, *Kailong Zhang, Youwei Lyu, Heng Guo, Si Li, Zhanyu Ma, Boxin Shi*
- * LightCity: An Urban Dataset for Outdoor Inverse Rendering and Reconstruction under Multi-illumination Conditions, *Jingjing Wang, Qirui Hu, Chong Bao, Yuke Zhu, Hujun Bao, Zhaopeng Cui, Guofeng Zhang*
- 170 3DRealCar: An In-the-wild RGB-D Car Dataset with 360-degree Views, *Xiaobiao Du, Yida Wang, Haiyang Sun, Zhuojie Wu, Hongwei Sheng, Shuyun Wang, Jiaying Ying, Ming Lu, Tianqing Zhu, Kun Zhan, Xin Yu*
- 171 PCR-GS: COLMAP-Free 3D Gaussian Splatting via Pose Co-Regularizations, *Yu Wei, Jiahui Zhang, Xiaoqin Zhang, Ling Shao, Shijian Lu*
- 172 NuiScene: Exploring Efficient Generation of Unbounded Outdoor Scenes, *Han-Hung Lee, Qinghong Han, Angel X. Chang*
- 173 MinCD-PnP: Learning 2D-3D Correspondences with Approximate Blind PnP, *Pei An, Jiaqi Yang, Muyao Peng, You Yang, Qiong Liu, Xiaolin Wu, Liangliang Nan*
- ☆ ETA: Efficiency through Thinking Ahead, A Dual Approach to Self-Driving with Large Models, *Shadi Hamdan, Chonghao Sima, Zetong Yang, Hongyang Li, Fatma Guney*
- 175 MergeOcc: Bridge the Domain Gap between Different LiDARs for Robust Occupancy Prediction, *Zikun Xu, Shaobing Xu*
- 176 RARE: Refine Any Registration of Pairwise Point Clouds via Zero-Shot Learning, *Chengyu Zheng, Jin Huang, Honghua Chen, Mingqiang Wei*
- 177 Feature Extraction and Representation of Pre-training Point Cloud Based on Diffusion Models, *Chang Qiu, Feipeng Da, Zilei Zhang*
- 178 Occupancy Learning with Spatiotemporal Memory, *Ziyang Leng, Jiawei Yang, Wenlong Yi, Bolei Zhou*
- 179 Towards Open-World Generation of Stereo Images and Unsupervised Matching, *Feng Qiao, Zhexiong Xiong, Eric Xing, Nathan Jacobs*
- 180 DMesh++: An Efficient Differentiable Mesh for Complex Shapes, *Sanghyun Son, Matheus Gadelha, Yang Zhou, Matthew Fisher, Zexiang Xu, Yi-Ling Qiao, Ming C. Lin, Yi Zhou*
- 181 TAD-E2E: A Large-scale End-to-end Autonomous Driving Dataset, *Chang Liu, Mingxu Zhu, Zheyuan Zhang, Linna Song, Xiao Zhao, Qingliang Luo, Qi Wang, Chufan Guo, Kuifeng Su*
- 182 LoD-Loc v2: Aerial Visual Localization over Low Level-of-Detail City Models using Explicit Silhouette Alignment, *Juelin Zhu, Shuaibang Peng, Long Wang, Hanlin Tan, Yu Liu, Maojun Zhang, Shen Yan*
- 183 LANGTRAJ: Diffusion Model and Dataset for Language-Conditioned Trajectory Simulation, *Wei-Jer Chang, Wei Zhan, Masayoshi Tomizuka, Manmohan Chandraker, Francesco Pittaluga*
- 184 EVT: Efficient View Transformation for Multi-Modal 3D Object Detection, *Yongjin Lee, Hyeon-Mun Jeong, Yurim Jeon, Sanghyun Kim*
- 185 OccluGaussian: Occlusion-Aware Gaussian Splatting for Large Scene Reconstruction and Rendering, *Shiyong Liu, Xiao Tang, Zhihao Li, Yingfan He, Chongjie Ye, Jianzhuang Liu, Bin Xiao Huang, Shunbo Zhou, Xiaofei Wu*
- 186 Temporal Overlapping Prediction: A Self-supervised Pre-training Method for LiDAR Moving Object Segmentation, *Ziliang Miao, Runjian Chen, Yixi Cai, Buwei He, Wenquan Zhao, Wenqi Shao, Bo Zhang, Fu Zhang*
- 187 Height-Fidelity Dense Global Fusion for Multi-modal 3D Object Detection, *Hanshi Wang, Jin Gao, Weiming Hu, Zhipeng Zhang*
- * Liberated-GS: 3D Gaussian Splatting Independent from SfM Point Clouds, *Weihong Pan, Xiaoyu Zhang, Hongjia Zhai, Xiaojun Xiang, Hanqing Jiang, Guofeng Zhang*
- 189 S3E: Self-Supervised State Estimation for Radar-Inertial System, *Shengpeng Wang, Yulong Xie, Qing Liao, Wei Wang*
- 190 MamV2XCalib: V2X-based Target-less Infrastructure Camera Calibration with State Space Model, *Yaoye Zhu, Zhe Wang, Yan Wang*
- 191 VideoRFSplat: Direct Scene-Level Text-to-3D Gaussian Splatting Generation with Flexible Pose and Multi-View Joint Modeling, *Hyojun Go, Byeongjun Park, Hyelin Nam, Byung-Hoon Kim, Hyungjin Chung, Changick Kim*
- 192 ReconDreamer++: Harmonizing Generative and Reconstructive Models for Driving Scene Representation, *Guosheng Zhao, Xiaofeng Wang, Chaojun Ni, Zheng Zhu, Wenkang Qin, Guan Huang, Xingang Wang*
- 193 S²M²: Scalable Stereo Matching Model for Reliable Depth Estimation, *Junhong Min, Youngpil Jeon, Jimin Kim, Minyong Choi*
- 194 3DGS-LM: Faster Gaussian-Splatting Optimization with Levenberg-Marquardt, *Lukas Höllein, Aljaž Božič, Michael Zollhöfer, Matthias Nießner*
- 195 ACE-G: Improving Generalization of Scene Coordinate Regression Through Query Pre-Training, *Leonard Bruns, Axel Barroso-Laguna, Tommaso Cavallari, Aron Monszpart, Sowmya Munukutla, Victor Adrian Prisacariu, Eric Brachmann*
- 196 3D Test-time Adaptation via Graph Spectral Driven Point Shift, *Xin Wei, Qin Yang, Yijie Fang, Mingrui Zhu, Nannan Wang*
- ☆ VistaDream: Sampling multiview consistent images for single-view scene reconstruction, *Haiping Wang, Yuan Liu, Ziwei Liu, Wenping Wang, Zhen Dong, Bisheng Yang*
- 198 Towards Visual Localization Interoperability: Cross-Feature for Collaborative Visual Localization and Mapping, *Alberto Jaenal, Paula Carbó Cubero, José Araújo, André Mateus*
- 199 MiDSummer: Multi-Guidance Diffusion for Controllable Zero-Shot Immersive Gaussian Splatting Scene Generation, *Anjun Hu, Richard Tomsett, Valentin Gourmet, Massimo Camplani, Jas Kandola, Hanting Xie*
- 200 AG2aussian: Anchor-Graph Structured Gaussian Splatting for Instance-Level 3D Scene Understanding and Editing, *Zhaonan Wang, Manyi Li, Changhe Tu*
- 201 Hierarchical 3D Scene Graphs Construction Outdoors, *Jon Nyffeler, Federico Tombari, Daniel Barath*
- 202 Spatio-Spectral Pattern Illumination for Direct and Indirect Separation from a Single Hyperspectral Image, *Shin Ishihara, Imari Sato*
- * SDFormer: Vision-based 3D Semantic Scene Completion via SAM-assisted Dual-channel Voxel Transformer, *Yujie Xue, Huilong Pi, Jiapeng Zhang, Yunchuan Qin, Zhuo Tang, Kenli Li, Ruihui Li*
- 204 Adversarial Exploitation of Data Diversity Improves Visual Localization, *Sihang Li, Siqi Tan, Bowen Chang, Jing Zhang, Chen Feng, Yiming Li*
- 205 SU-RGS: Relightable 3D Gaussian Splatting from Sparse Views under Unconstrained Illuminations, *Qi Zhang, Chi Huang, Qian Zhang, Nan Li, Wei Feng*
- 206 GS-LIVM: Real-Time Photo-Realistic LiDAR-Inertial-Visual Mapping with Gaussian Splatting, *Yusen Xie, Zhenmin Huang, Jin Wu, Jun Ma*
- 207 GeoFormer: Geometry Point Encoder for 3D Object Detection with Graph-based Transformer, *Xin Jin, Haisheng Su, Cong Ma, Kai Liu, Wei Wu, Fei Hui, Junchi Yan*
- 208 DrivingGPT: Unifying Driving World Modeling and Planning with

- Multi-modal Autoregressive Transformers, *Yuntao Chen, Yuqi Wang, Zhaoxiang Zhang*
- 209 AlignDiff: Learning Physically-Grounded Camera Alignment via Diffusion, *Liyue Xie, Jiancong Guo, Ozan Cakmakci, Andre Araujo, László A. Jeni, Zhiheng Jia*
- 210 Towards a 3D Transfer-based Black-box Attack via Critical Feature Guidance, *Shuchao Pang, Zhenghan Chen, Shen Zhang, Liming Lu, Siyuan Liang, Anan Du, Yongbin Zhou*
- 211 Tile-wise vs. Image-wise: Random-Tile Loss and Training Paradigm for Gaussian Splatting, *Xiaoyu Zhang, Weihong Pan, Xiaojun Xiang, Hongjia Zhai, Liyang Zhou, Hanqing Jiang, Guofeng Zhang*
- 212 DriveArena: A Closed-loop Generative Simulation Platform for Autonomous Driving, *Xuemeng Yang, Licheng Wen, Tiantian Wei, Yukai Ma, Jianbiao Mei, Xin Li, Wenjie Lei, Daocheng Fu, Pinlong Cai, Min Dou, Liang He, Yong Liu, Botian Shi, Yu Qiao*
- 213 Explaining Human Preferences via Metrics for Structured 3D Reconstruction, *Jack Langerman, Denys Rozumnyi, Yuzhong Huang, Dmytro Mishkin*
- 214 CoopTrack: Exploring End-to-End Learning for Efficient Cooperative Sequential Perception, *Jiaru Zhong, Jiahao Wang, Jiahui Xu, Xiaofan Li, Zaiqing Nie, Haibao Yu*
- 215 VLR-Driver: Large Vision-Language-Reasoning Models for Embodied Autonomous Driving, *Fanjie Kong, Yitong Li, Weihuang Chen, Chen Min, Yizhe Li, Zhiqiang Gao, Haoyang Li, Zhongyu Guo, Hongbin Sun*
- 216 RoCo-Sim: Enhancing Roadside Collaborative Perception through Foreground Simulation, *Yuwen Du, Anning Hu, Zichen Chao, Yifan Lu, Junhao Ge, Genjia Liu, Weitao Wu, Lanjun Wang, Siheng Chen*
- 217 Inverse 3D Microscopy Rendering for Cell Shape Inference with Active Mesh, *Sacha Ichbiah, Anshuman Sinha, Fabrice Delbary, Hervé Turlier*
- 218 E-SAM: Training-Free Segment Every Entity Model, *Weiming Zhang, Dingwen Xiao, Lei Chen, Lin Wang*
- 219 Online Reasoning Video Segmentation with Just-in-Time Digital Twins, *Yiqing Shen, Bohan Liu, Chenjia Li, Lalithkumar Seenivasan, Mathias Unberath*
- 220 Disentangling Instance and Scene Contexts for 3D Semantic Scene Completion, *Enyu Liu, En Yu, Sijia Chen, Wenbing Tao*
- 221 GaussRender: Learning 3D Occupancy with Gaussian Rendering, *Loick Chambon, Eloi Zablocki, Alexandre Boulch, Mickael Chen, Matthieu Cord*
- 222 SA-Occ: Satellite-Assisted 3D Occupancy Prediction in Real World, *Chen Chen, Zhirui Wang, Taowei Sheng, Yi Jiang, Yundu Li, Peirui Cheng, Luning Zhang, Kaiqiang Chen, Yanfeng Hu, Xue Yang, Xian Sun*
- 223 UniVerse: Unleashing the Scene Prior of Video Diffusion Models for Robust Radiance Field Reconstruction, *Jin Cao, Hongrui Wu, Ziyong Feng, Hujun Bao, Xiaowei Zhou, Sida Peng*
- 224 ExploreGS: Explorable 3D Scene Reconstruction with Virtual Camera Samplings and Diffusion Priors, *Minsu Kim, Subin Jeon, In Cho, Mijin Yoo, Seon Joo Kim*
- 225 LaneDiffusion: Improving Centerline Graph Learning via Prior Injected BEV Feature Generation, *Zijie Wang, Weiming Zhang, Wei Zhang, Xiao Tan, Hongxing Liu, Yaowei Wang, Guanbin Li*
- 226 Bridging 3D Anomaly Localization and Repair via High-Quality Continuous Geometric Representation, *Bozhong Zheng, Jinye Gan, Xiaohao Xu, Xintao Chen, Wenqiao Li, Xiaonan Huang, Na Ni, Yingna Wu*
- 227 Video2BEV: Transforming Drone Videos to BEVs for Video-based Geo-localization, *Hao Ju, Shaofei Huang, Si Liu, Zhedong Zheng*
- 228 CuMPerLay: Learning Cubical Multiparameter Persistence Vectorizations, *Caner Korkmaz, Brighton Nuwagira, Baris Coskunuzer, Tolga Birdal*
- 229 SGAD: Semantic and Geometric-aware Descriptor for Local Feature Matching, *Xiangzeng Liu, Chi Wang, Guanglu Shi, Xiaodong Zhang, Qiguang Miao, Miao Fan*
- 230 ScanEdit: Hierarchically-Guided Functional 3D Scan Editing, *Mohamed El Amine Boudjoghra, Ivan Laptev, Angela Dai*
- 231 Supercharging Floorplan Localization with Semantic Rays, *Yuval Grader, Hadar Averbuch-Elor*
- 232 RobustSplat: Decoupling Densification and Dynamics for Transient-Free 3DGS, *Chuanyu Fu, Yuqi Zhang, Kunbin Yao, Guanying Chen, Yuan Xiong, Chuan Huang, Shuguang Cui, Xiaochun Cao*
- 233 End-to-End Driving with Online Trajectory Evaluation via BEV World Model, *Yingyan Li, Yuqi Wang, Yang Liu, Jiawei He, Lue Fan, Zhaoxiang Zhang*
- 234 Planar Affine Rectification from Local Change of Scale and Orientation, *Yuval Nissan, Marc Pollefeys, Daniel Barath*
- 235 ERNet: Efficient Non-Rigid Registration Network for Point Sequences, *Guangzhao He, Yuxi Xiao, Zhen Xu, Xiaowei Zhou, Sida Peng*
- 236 SeqGrowGraph: Learning Lane Topology as a Chain of Graph Expansions, *Mengwei Xie, Shuang Zeng, Xinyuan Chang, Xinran Liu, Zheng Pan, Mu Xu, Xing Wei*
- 237 InvRGB+L: Inverse Rendering of Complex Scenes with Unified Color and LiDAR Reflectance Modeling, *Xiaoxue Chen, Bhargav Chandaka, Chih-Hao Lin, Ya-Qin Zhang, David Forsyth, Hao Zhao, Shenlong Wang*
- 238 CityGS-X: A Scalable Architecture for Efficient and Geometrically Accurate Large-Scale Scene Reconstruction, *Yuanyuan Gao, Hao Li, Jiaqi Chen, Zhengyu Zou, Zhihang Zhong, Dingwen Zhang, Xiao Sun, Junwei Han*
- 239 Doppler-Aware LiDAR-RADAR Fusion for Weather-Robust 3D Detection, *Yujeong Chae, Heejeon Park, Hyeonseong Kim, Kuk-Jin Yoon*
- 240 Egocentric Action-aware Inertial Localization in Point Clouds with Vision-Language Guidance, *Mingfang Zhang, Ryo Yonetani, Yifei Huang, Liangyang Ouyang, Ruicong Liu, Yoichi Sato*
- 241 Epona: Autoregressive Diffusion World Model for Autonomous Driving, *Kaiwen Zhang, Zhenyu Tang, Xiaotao Hu, Xingang Pan, Xiaoyang Guo, Yuan Liu, Jingwei Huang, Li Yuan, Qian Zhang, Xiao-Xiao Long, Xun Cao, Wei Yin*
- 242 DiST-4D: Disentangled Spatiotemporal Diffusion with Metric Depth for 4D Driving Scene Generation, *Jiazhe Guo, Yikang Ding, Xiwu Chen, Shuo Chen, Bohan Li, Yingshuang Zou, Xiaoyang Lyu, Feiyang Tan, Xiaojuan Qi, Zhiheng Li, Hao Zhao*
- 243 MCOP: Multi-UAV Collaborative Occupancy Prediction, *Zefu Lin, Wenbo Chen, Xiaojuan Jin, Yuran Yang, Lue Fan, Yixin Zhang, Yufeng Zhang, Zhaoxiang Zhang*
- 244 Spectral Sensitivity Estimation with an Uncalibrated Diffraction Grating, *Lilika Makabe, Hiroaki Santo, Fumio Okura, Michael S. Brown, Yasuyuki Matsushita*
- 245 Leveraging Local Patch Alignment to Seam-cutting for Large Parallax Image Stitching, *Tianli Liao, Chenyang Zhao, Lei Li, Heling Cao*
- 246 InfiniCube: Unbounded and Controllable Dynamic 3D Driving Scene Generation with World-Guided Video Models, *Yifan Lu, Xuanchi Ren, Jiawei Yang, Tianchang Shen, Zhangjie Wu, Jun Gao, Yue Wang, Siheng Chen, Mike Chen, Sanja Fidler, Jiahui Huang*
- 247 PriorMotion: Generative Class-Agnostic Motion Prediction with Raster-Vector Motion Field Priors, *Kangan Qian, Jinyu Miao, Xinyu Jiao, Ziang Luo, Zheng Fu, Yining Shi, Yunlong Wang, Kun Jiang, Diange Yang*
- 248 MGSR: 2D/3D Mutual-boosted Gaussian Splatting for High-fidelity Surface Reconstruction under Various Light Conditions, *Qingyuan Zhou, Yuehu Gong, Weidong Yang, Jiaze Li, Yeqi Luo, Baixin Xu, Shuhao Li, Ben Fei, Ying He*
- 249 Hydra-NeXt: Robust Closed-Loop Driving with Open-Loop Training, *Zhenxin Li, Shihao Wang, Shiyi Lan, Zhiding Yu, Zuxuan Wu, Jose M. Alvarez*
- 250 IntrinsicControlNet: Cross-distribution Image Generation with Real and Unreal, *Jiayuan Lu, Rengan Xie, Zixuan Xie, Zhizhen Wu, Dianbing Xi, Qi Ye, Rui Wang, Hujun Bao, Yuchi Huo*
- 251 SteerX: Creating Any Camera-Free 3D and 4D Scenes with Geometric Steering, *Byeongjun Park, Hyojun Go, Hyelin Nam, Byung-Hoon Kim, Hyungjin Chung, Changick Kim*
- 252 WIPES: Wavelet-based Visual Primitives, *Wenhao Zhang, Hao Zhu, Delong Wu, Di Kang, Linchao Bao, Xun Cao, Zhan Ma*
- 253 DiffPCI: Large Motion Point Cloud frame Interpolation with Diffusion Model, *Tianyu Zhang, Haobo Jiang, Jian Yang, Jin Xie*
- 254 UPP: Unified Point-Level Prompting for Robust Point Cloud Analysis, *Zixiang Ai, Zhenyu Cui, Yuxin Peng, Jiahuan Zhou*
- 255 ArgMatch: Adaptive Refinement Gathering for Efficient Dense Matching, *Yuxin Deng, Kaining Zhang, Linfeng Tang, Jiaqi Yang, Jiayi Ma*

- 256 RoboTron-Sim: Improving Real-World Driving via Simulated Hard-Case, *Baihui Xiao, Chengjian Feng, Zhijian Huang, Feng Yan, Yujie Zhong, Lin Ma*
- 257 Thermal Polarimetric Multi-view Stereo, *Takahiro Kushida, * Kenichiro Tanaka*
- 258 StealthAttack: Robust 3D Gaussian Splatting Poisoning via Density-Guided Illusions, *Bo-Hsu Ke, You-Zhe Xie, Yu-Lun Liu, Wei-Chen Chiu*
- 259 LongSplat: Robust Unposed 3D Gaussian Splatting for Casual Long Videos, *Chin-Yang Lin, Cheng Sun, Fu-En Yang, Min-Hung Chen, Yen-Yu Lin, Yu-Lun Liu*
- 260 WonderTurbo: Generating Interactive 3D World in 0.72 Seconds, *Chaojun Ni, Xiaofeng Wang, Zheng Zhu, Weijie Wang, Haoyun Li, Guosheng Zhao, Jie Li, Wenkang Qin, Guan Huang, Wenjun Mei*
- 261 Geometric Alignment and Prior Modulation for View-Guided Point Cloud Completion on Unseen Categories, *Jingqiao Xiu, Yicong Li, Na Zhao, Han Fang, Xiang Wang, Angela Yao*
- 262 AccidentalGS: 3D Gaussian Splatting from Accidental Camera Motion, *Mao Mao, Xujie Shen, Guyuan Chen, Boming Zhao, Jiarui Hu, Hujun Bao, Zhaopeng Cui*
- 263 MOSAIC: Generating Consistent, Privacy-Preserving Scenes from Multiple Depth Views in Multi-Room Environments, *Zhixuan Liu, Haokun Zhu, Rui Chen, Jonathan Francis, Soonmin Hwang, Ji Zhang, Jean Oh*
- 264 Coordinate-based Speed of Sound Recovery for Aberration-Corrected Photoacoustic Computed Tomography, *Tianao Li, Manxiu Cui, Cheng Ma, Emma Alexander*
- 265 Sparfels: Fast Reconstruction from Sparse Unposed Imagery, ** Shubhendu Jena, Amine Ouasfi, Mae Younes, Adnane Boukhayma*
- 266 GenFlow3D: Generative Scene Flow Estimation and Prediction on Point Cloud Sequences, *Hanlin Li, Wenming Weng, Yueyi Zhang, Zhiwei Xiong*
- 267 Self-Supervised Sparse Sensor Fusion for Long Range Perception, *Edoardo Palladin, Samuel Brucker, Filippo Ghilotti, Praveen Narayanan, Mario Bijelic, Felix Heide*
- 268 Generative Gaussian Splatting: Generating 3D Scenes with Video Diffusion Priors, *Katja Schwarz, Norman Müller, Peter Kotschieder*
- 269 You Share Beliefs, I Adapt: Progressive Heterogeneous Collaborative Perception, *Hao Si, Ehsan Javanmardi, Manabu Tsukada*
- 270 Curve-Aware Gaussian Splatting for 3D Parametric Curve Reconstruction, *Zhirui Gao, Renjiao Yi, Yaqiao Dai, Xuening Zhu, Wei Chen, Chenyang Zhu, Kai Xu*
- 271 TransiT: Transient Transformer for Non-line-of-sight Videography, *Ruiqian Li, Siyuan Shen, Suan Xia, Ziheng Wang, Xingyue Peng, Chengxuan Song, Yingsheng Zhu, Tao Wu, Shiyang Li, Jingyi Yu*
- 272 Neural Multi-View Self-Calibrated Photometric Stereo without Photometric Stereo Cues, *Xu Cao, Takafumi Taketomi*
- 273 Unified Multi-Agent Trajectory Modeling with Masked Trajectory Diffusion, *Songru Yang, Zhenwei Shi, Zhengxia Zou*
- 274 RayGaussX: Accelerating Gaussian-Based Ray Marching for Real-Time and High-Quality Novel View Synthesis, *Hugo Blanc, Jean-Emmanuel Deschaud, Alexis Paljic*
- 275 SynCity: Training-Free Generation of 3D Worlds, *Paul Engstler, Aleksandar Shtedritski, Iro Laina, Christian Rupprecht, Andrea Vedaldi*
- 276 RadarSplat: Radar Gaussian Splatting for High-Fidelity Data Synthesis and 3D Reconstruction of Autonomous Driving Scenes, *Pou-Chun Kung, Skanda Harisha, Ram Vasudevan, Aline Eid, Katherine A. Skinner*
- 277 Tree Skeletonization from 3D Point Clouds by Denoising Diffusion, *Elias Ariel Marks, Lucas Nunes, Federico Magistri, Matteo Sodano, Rodrigo Marcuzzi, Lars Zimmermann, Jens Behley, Cyrill Stachniss*
- 278 Point Cloud Self-supervised Learning via 3D to Multi-view Masked Learner, *Zhimin Chen, Xuewei Chen, Xiao Guo, Yingwei Li, Longlong Jing, Liang Yang, Bing Li*
- 279 Splat-LOAM: Gaussian Splatting LiDAR Odometry and Mapping, *Emanuele Giacomini, Luca Di Giammarino, Lorenzo De Rebotto, Giorgio Grisetti, Martin R. Oswald*
- 280 Purge-Gate: Backpropagation-Free Test-Time Adaptation for Point Clouds Classification via Token purging, *Moslem Yazdanpanah, Ali Bahri, Mehrdad Noori, Sahar Dastani, Gustavo Adolfo Vargas*
- Hakim, David Osowiechi, Ismail Ben Ayed, Christian Desrosiers*
- 281 AAA-Gaussians: Anti-Aliased and Artifact-Free 3D Gaussian Rendering, *Michael Steiner, Thomas Köhler, Lukas Radl, Felix Windisch, Dieter Schmalstieg, Markus Steinberger*
- 282 SAFT: Shape and Appearance of Fabrics from Template via Differentiable Physical Simulations from Monocular Video, *David Stotko, Reinhard Klein*
- 283 UNIS: A Unified Framework for Achieving Unbiased Neural Implicit Surfaces in Volume Rendering, *Junkai Deng, Hanting Niu, Jiaze Li, Fei Hou, Ying He*
- 284 BridgeDepth: Bridging Monocular and Stereo Reasoning with Latent Alignment, *Tongfan Guan, Jiaxin Guo, Chen Wang, Yun-Hui Liu*
- 285 Neural Inverse Rendering for High-Accuracy 3D Measurement of Moving Objects with Fewer Phase-Shifting Patterns, *Yuki Urakawa, Yoshihiro Watanabe*
- 286 Towards Foundational Models for Single-Chip Radar, ** Tianshu Huang, Akarsh Prabhakara, Chuhan Chen, Jay Karhade, Deva Ramanan, Matthew O'toole, Anthony Rowe*
- 287 Diffusion Image Prior, *Hamadi Chihaoui, Paolo Favaro*
- 288 FlowR: Flowing from Sparse to Dense 3D Reconstructions, ** Tobias Fischer, Samuel Rota Buló, Yung-Hsu Yang, Nikhil Keetha, Lorenzo Porzi, Norman Müller, Katja Schwarz, Jonathon Luiten, Marc Pollefeys, Peter Kotschieder*
- 289 WorldScore: A Unified Evaluation Benchmark for World Generation, *☆ Haoyi Duan, Hong-Xing Yu, Sirui Chen, Li Fei-Fei, Jiajun Wu*
- 290 Perspective-Invariant 3D Object Detection, *Ao Liang, Lingdong Kong, Dongyue Lu, Youquan Liu, Jian Fang, Huaici Zhao, Wei Tsang Ooi*
- 291 CA-I2P: Channel-Adaptive Registration Network with Global Optimal Selection, *Zhixin Cheng, Jiacheng Deng, Xinjun Li, Xiaotian Yin, Bohao Liao, Baoqun Yin, Wenfei Yang, Tianzhu Zhang*
- 292 LightSwitch: Multi-view Relighting with Material-guided Diffusion, *Yehonathan Litman, Fernando De la Torre, Shubham Tulsiani*
- 293 Decoupled Diffusion Sparks Adaptive Scene Generation, *Yunsong Zhou, Naisheng Ye, William Ljungbergh, Tianyu Li, Jiazhi Yang, Zetong Yang, Hongzi Zhu, Christoffer Petersson, Hongyang Li*
- 294 Recover Biological Structure from Sparse-View Diffraction Images with Neural Volumetric Prior, *Renzhi He, Haowen Zhou, Yubei Chen, Yi Xue*
- 295 ReAL-AD: Towards Human-Like Reasoning in End-to-End Autonomous Driving, *Yuhang Lu, Jiadong Tu, Yuexin Ma, Xinge Zhu*
- 296 SpatialCrafter: Unleashing the Imagination of Video Diffusion Models for Scene Reconstruction from Limited Observations, *Songchun Zhang, Huiyao Xu, Sitong Guo, Zhongwei Xie, Hujun Bao, Weiwei Xu, Changqing Zou*
- 297 MonoMVSNet: Monocular Priors Guided Multi-View Stereo Network, *Jianfei Jiang, Qiankun Liu, Haochen Yu, Hongyuan Liu, Liyong Wang, Jiansheng Chen, Huimin Ma*
- 298 HERMES: A Unified Self-Driving World Model for Simultaneous 3D Scene Understanding and Generation, *Xin Zhou, Dingkan Liang, Sifan Tu, Xiwu Chen, Yikang Ding, Dingyuan Zhang, Feiyang Tan, Hengshuang Zhao, Xiang Bai*
- 299 MEGA: Memory-Efficient 4D Gaussian Splatting for Dynamic Scenes, ** Xinjie Zhang, Zhening Liu, Yifan Zhang, Xingtong Ge, Dailan He, Tongda Xu, Yan Wang, Zehong Lin, Shuicheng Yan, Jun Zhang*
- 300 Generative Active Learning for Long-tail Trajectory Prediction via Controllable Diffusion Model, *Daehee Park, Monu Surana, Pranav Desai, Ashish Mehta, Reuben MV John, Kuk-Jin Yoon*
- 301 QuickSplat: Fast 3D Surface Reconstruction via Learned Gaussian Initialization, *Yueh-Cheng Liu, Lukas Höllein, Matthias Nießner, Angela Dai*
- 302 EYE3: Turn Anything into Naked-eye 3D, *Yingde Song, Zongyuan Yang, Baolin Liu, Yongping Xiong, Sai Chen, Lan Yi, Zhaohe Zhang, Xunbo Yu*
- 303 NATRA: Noise-Agnostic Framework for Trajectory Prediction with Noisy Observations, *Rongqing Li, Changsheng Li, Ruilin Lv, Yuhang Li, Yang Gao, Xiaolu Zhang, Jun Zhou*
- 304 SP2T: Sparse Proxy Attention for Dual-stream Point

- Transformer, *Jiaxu Wan, Hong Zhang, Ziqi He, Yangyan Deng, Qishu Wang, Ding Yuan, Yifan Yang*
- 305 Instant GaussianImage: A Generalizable and Self-Adaptive Image Representation via 2D Gaussian Splatting, *Zhaojie Zeng, Yuesong Wang, Tao Guan, Chao Yang, Lili Ju*
- 306 CF3: Compact and Fast 3D Feature Fields, *Hyunjoon Lee, Joonykyu Min, Jaesik Park*
- 307 When Anchors Meet Cold Diffusion: A Multi-Stage Approach to Lane Detection, *Bo-Lun Huang, Zi-Xiang Ni, Feng-Kai Huang, Hong-Han Shuai, Wen-Huang Cheng*
- 308 2D Gaussian Splatting-based Sparse-view Transparent Object Depth Reconstruction via Physics Simulation for Scene Update, *Jeongyun Kim, Seunghoon Jeong, Giseop Kim, Myung-Hwan Jeon, Eunji Jun, Ayoung Kim*
- 309 Learning Neural Scene Representation from iToF Imaging, *Wenjie Chang, Hanzhi Chang, Yueyi Zhang, Wenfei Yang, Tianzhu Zhang*
- 310 No Pose at All: Self-Supervised Pose-Free 3D Gaussian Splatting from Sparse Views, *Ranran Huang, Krystian Mikolajczyk*
- 311 Fusion Meets Diverse Conditions: A High-diversity Benchmark and Baseline for UAV-based Multimodal Object Detection with Condition Cues, *Chen Chen, Kangcheng Bin, Ting Hu, Jiahao Qi, Xingyue Liu, Tianpeng Liu, Zhen Liu, Yongxiang Liu, Ping Zhong*
- 312 Faster and Better 3D Splatting via Group Training, *Chengbo Wang, Guozheng Ma, Yifei Xue, Yizhen Lao*
- 313 Sat2City: 3D City Generation from A Single Satellite Image with Cascaded Latent Diffusion, *Tongyan Hua, Lutao Jiang, Ying-Cong Chen, Wufan Zhao*
- 314 Recovering Parametric Scenes from Very Few Time-of-Flight Pixels, *Carter Sifferman, Yiquan Li, Yiming Li, Fangzhou Mu, Michael Gleicher, Mohit Gupta, Yin Li*
- 315 NeuFrameQ: Neural Frame Fields for Scalable and Generalizable Anisotropic Quadrangulation, *Ying-Tian Liu, Jiajun Li, Yu-Tao Liu, Xin Yu, Yuan-Chen Guo, Yan-Pei Cao, Ding Liang, Ariel Shamir, Song-Hai Zhang*
- 316 FB-Diff: Fourier Basis-guided Diffusion for Temporal Interpolation of 4D Medical Imaging, *Xin You, Runze Yang, Chuyan Zhang, Zhongliang Jiang, Jie Yang, Nassir Navab*
- 317 RTMap: Real-Time Recursive Mapping with Change Detection and Localization, *Yuheng Du, Sheng Yang, Lingxuan Wang, Zhenghua Hou, Chengying Cai, Zhitao Tan, Mingxia Chen, Shi-Sheng Huang, Qiang Li*
- 318 CoDa-4DGS: Dynamic Gaussian Splatting with Context and Deformation Awareness for Autonomous Driving, *Rui Song, Chenwei Liang, Yan Xia, Walter Zimmer, Hu Cao, Holger Caesar, Andreas Festag, Alois Knoll*
- 319 A Differentiable Wave Optics Model for End-to-End Computational Imaging System Optimization, *Chi-Jui Ho, Yash Belhe, Steve Rotenberg, Ravi Ramamoorthi, Tzu-Mao Li, Nicholas Antipa*
- 320 Controllable 3D Outdoor Scene Generation via Scene Graphs, *Yuheng Liu, Xinke Li, Yuning Zhang, Lu Qi, Xin Li, Wenping Wang, Chongshou Li, Xueting Li, Ming-Hsuan Yang*
- 321 CasP: Improving Semi-Dense Feature Matching Pipeline Leveraging Cascaded Correspondence Priors for Guidance, *Peiqi Chen, Lei Yu, Yi Wan, Yingying Pei, Xinyi Liu, Yongxiang Yao, Yingying Zhang, Lixiang Ru, Liheng Zhong, Jingdong Chen, Ming Yang, Yongjun Zhang*
- 322 PolGS: Polarimetric Gaussian Splatting for Fast Reflective Surface Reconstruction, *Yufei Han, Bowen Tie, Heng Guo, Youwei Lyu, Si Li, Boxin Shi, Yunpeng Jia, Zhanyu Ma*
- 323 Driving View Synthesis on Free-form Trajectories with Generative Prior, *Zeyu Yang, Zijie Pan, Yuankun Yang, Xiatian Zhu, Li Zhang*
- 324 ResGS: Residual Densification of 3D Gaussian for Efficient Detail Recovery, *Yanzhe Lyu, Kai Cheng, Xin Kang, Xuejin Chen*
- 325 SEGSLAM: Structure-enhanced 3D Gaussian Splatting SLAM with Appearance Embedding, *Tianci Wen, Zhiang Liu, Yongchun Fang*
- 326 Constraint-Aware Feature Learning for Parametric Point Cloud, *Xi Cheng, Ruiqi Lei, Di Huang, Zhichao Liao, Fengyuan Piao, Yan Chen, Pingfa Feng, Long Zeng*
- 327 PixelStitch: Structure-Preserving Pixel-Wise Bidirectional Warps for Unsupervised Image Stitching, *Hengzhe Jin, Lang Nie, Chunyu Lin, Xiaomei Feng, Yao Zhao*
- 328 MagicDrive-V2: High-Resolution Long Video Generation for Autonomous Driving with Adaptive Control, *Ruiyuan Gao, Kai Chen, Bo Xiao, Lanqing Hong, Zhenguo Li, Qiang Xu*
- 329 FreqPDE: Rethinking Positional Depth Embedding for Multi-View 3D Object Detection Transformers, *Haisheng Su, Junjie Zhang, Feixiang Song, Sanping Zhou, Wei Wu, Junchi Yan, Nanning Zheng*
- 330 Correspondence-Free Fast and Robust Spherical Point Pattern Registration, *Anik Sarker, Alan T. Asbeck*
- 331 NeuraLeaf: Neural Parametric Leaf Models with Shape and Deformation Disentanglement, *Yang Yang, Dongni Mao, Hiroaki Santo, Yasuyuki Matsushita, Fumio Okura*
- 332 ZeroStereo: Zero-shot Stereo Matching from Single Images, *Xianqi Wang, Hao Yang, Gangwei Xu, Junda Cheng, Min Lin, Yong Deng, Jinliang Zang, Yurui Chen, Xin Yang*
- 333 CVFusion: Cross-View Fusion of 4D Radar and Camera for 3D Object Detection, *Hanzhi Zhong, Zhiyu Xiang, Ruoyu Xu, Jingyun Fu, Peng Xu, Shaohong Wang, Zhihao Yang, Tianyu Pu, Eryun Liu*
- 334 Stochastic Gradient Estimation for Higher-Order Differentiable Rendering, *Zican Wang, Michael Fischer, Tobias Ritschel*
- 335 M2EIT: Multi-Domain Mixture of Experts for Robust Neural Inertial Tracking, *Yan Li, Yang Xu, Changhao Chen, Zhongchen Shi, Wei Chen, Liang Xie, Hongbo Chen, Erwei Yin*
- 336 Orchid: Image Latent Diffusion for Joint Appearance and Geometry Generation, *Akshay Krishnan, Xinchun Yan, Vincent Casser, Abhijit Kundu*
- 337 CATSplat: Context-Aware Transformer with Spatial Guidance for Generalizable 3D Gaussian Splatting from A Single-View Image, *Wonseok Roh, Hwanhee Jung, Jong Wook Kim, Seungwan Lee, Innfarn Yoo, Andreas Lugmayr, Seunggeun Chi, Karthik Ramani, Sangpil Kim*
- 338 TPG-INR: Target Prior-Guided Implicit 3D CT Reconstruction for Enhanced Sparse-view Imaging, *Qinglei Cao, Ziyao Tang, Xiaoqin Tang*
- 339 TrackAny3D: Transferring Pretrained 3D Models for Category-unified 3D Point Cloud Tracking, *Mengmeng Wang, Haonan Wang, Yulong Li, Xiangjie Kong, Jiaxin Du, Guojiang Shen, Feng Xia*
- 340 Quadratic Gaussian Splatting: High Quality Surface Reconstruction with Second-order Geometric Primitives, *Ziyu Zhang, Binbin Huang, Hanqing Jiang, Liyang Zhou, Xiaojun Xiang, Shuhan Shen*
- 341 Uncertainty-Aware Diffusion-Guided Refinement of 3D Scenes, *Sarosij Bose, Arindam Dutta, Sayak Nag, Junge Zhang, Jiachen Li, Konstantinos Karydis, Amit K. Roy-Chowdhury*
- 342 VoxelKP: A Voxel-based Network Architecture for Human Keypoint Estimation in LiDAR Data, *Jian Shi, Peter Wonka*
- 343 χ : Symmetry Understanding of 3D Shapes via Chirality Disentanglement, *Weikang Wang, Tobias Weißberg, Nafie El Amrani, Florian Bernard*
- 344 Foresight in Motion: Reinforcing Trajectory Prediction with Reward Heuristics, *Muleilan Pei, Shaoshuai Shi, Xuesong Chen, Xu Liu, Shaojie Shen*
- 345 MAESTRO: Task-Relevant Optimization via Adaptive Feature Enhancement and Suppression for Multi-task 3D Perception, *Changwon Kang, Jisong Kim, Hongjae Shin, Junseo Park, Jun Won Choi*
- 346 Domain-aware Category-level Geometry Learning Segmentation for 3D Point Clouds, *Pei He, Lingling Li, Licheng Jiao, Ronghua Shang, Fang Liu, Shuang Wang, Xu Liu, Wenping Ma*
- 347 Event-boosted Deformable 3D Gaussians for Dynamic Scene Reconstruction, *Wenhao Xu, Wenming Weng, Yueyi Zhang, Ruikang Xu, Zhiwei Xiong*
- 348 ToF-Splatting: Dense SLAM using Sparse Time-of-Flight Depth and Multi-Frame Integration, *Andrea Conti, Matteo Poggi, Valerio Cambareri, Martin R. Oswald, Stefano Mattoccia*
- 349 Joint Semantic and Rendering Enhancements in 3D Gaussian Modeling with Anisotropic Local Encoding, *Jingming He, Chongyi Li, Shiqi Wang, Sam Kwong*
- 350 Unsupervised Imaging Inverse Problems with Diffusion Distribution

- Matching, *Giacomo Meanti, Thomas Ryckeboer, Michael Arbel, Julien Mairal*
- 351 R-LiViT: A LiDAR-Visual-Thermal Dataset Enabling Vulnerable Road User Focused Roadside Perception, *Jonas Mirlach, Lei Wan, Andreas Wiedholz, Hannan Ejaz Keen, Andreas Eich*
- 352 V2XScenes: A Multiple Challenging Traffic Conditions Dataset for Large-Range Vehicle-Infrastructure Collaborative Perception, *Bowen Wang, Yafei Wang, Wei Gong, Siheng Chen, Genjia Liu, Minhao Xiong, Chin Long Ng*
- 353 mmCooper: A Multi-agent Multi-stage Communication-efficient and Collaboration-robust Cooperative Perception Framework, *Bingyi Liu, Jian Teng, Hongfei Xue, Enshu Wang, Chuanhui Zhu, Pu Wang, Libing Wu*
- 354 SC-Lane: Slope-aware and Consistent Road Height Estimation Framework for 3D Lane Detection, *Chaesong Park, Eunbin Seo, Jihyeon Hwang, Jongwoo Lim*
- 355 ForestFormer3D: A Unified Framework for End-to-End Segmentation of Forest LiDAR 3D Point Clouds, *Binbin Xiang, Maciej Wielgosz, Stefano Puliti, Kamil Král, Martin Krůček, Azim Missarov, Rasmus Astrup*
- 356 Easy3D: A Simple Yet Effective Method for 3D Interactive Segmentation, *Andrea Simonelli, Norman Müller, Peter Kotschieder*
- 357 Robust 3D Object Detection using Probabilistic Point Clouds from Single-Photon LiDARs, *Bhavya Goyal, Felipe Gutierrez-Barragan, Wei Lin, Andreas Velten, Yin Li, Mohit Gupta*
- 358 SViM3D: Stable Video Material Diffusion for Single Image 3D Generation, *Andreas Engelhardt, Mark Boss, Vikram Voleti, Chun-Han Yao, Hendrik P. A. Lensch, Varun Jampani*
- 359 HouseCrafter: Lifting Floorplans to 3D Scenes with 2D Diffusion Models, *Yiwen Chen, Hieu T. Nguyen, Vikram Voleti, Varun Jampani, Huaizu Jiang*
- 360 Leveraging BEV Paradigm for Ground-to-Aerial Image Synthesis, *Junyan Ye, Jun He, Weijia Li, Zhutao Lv, Yi Lin, Jinhua Yu, Haote Yang, Conghui He*
- 361 EMD: Explicit Motion Modeling for High-Quality Street Gaussian Splatting, *Xiaobao Wei, Qingpo Wu, Zhongyu Zhao, Zhuangzhe Wu, Nan Huang, Ming Lu, Ningning Ma, Shanghang Zhang*
- 362 Learning Null Geodesics for Gravitational Lensing Rendering in General Relativity, *Mingyuan Sun, Zheng Fang, Jiaxu Wang, Kunyi Zhang, Qiang Zhang, Renjing Xu*
- 363 Kaleidoscopic Background Attack: Disrupting Pose Estimation with Multi-Fold Radial Symmetry Textures, *Xinlong Ding, Hongwei Yu, Jiawei Li, Feifan Li, Yu Shang, Bochao Zou, Huimin Ma, Jiansheng Chen*
- 364 Top2Pano: Learning to Generate Indoor Panoramas from Top-Down View, *Zitong Zhang, Suranjan Gautam, Rui Yu*
- 365 Perspective-aware 3D Gaussian Inpainting with Multi-view Consistency, *Yuxin Cheng, Binxiao Huang, Taiqiang Wu, Wenyong Zhou, Chenchen Ding, Zhengwu Liu, Graziano Chesi, Ngai Wong*
- 366 SparseRecon: Neural Implicit Surface Reconstruction from Sparse Views with Feature and Depth Consistencies, *Liang Han, Xu Zhang, Haichuan Song, Kanle Shi, Yu-Shen Liu, Zhizhong Han*
- 367 SurfaceSplat: Connecting Surface Reconstruction and Gaussian Splatting, *Zihui Gao, Jia-Wang Bian, Guosheng Lin, Hao Chen, Chunhua Shen*
- 368 SAM4D: Segment Anything in Camera and LiDAR Streams, *Jianyun Xu, Song Wang, Ziqian Ni, Chunyong Hu, Sheng Yang, Jianke Zhu, Qiang Li*
- 369 StruMamba3D: Exploring Structural Mamba for Self-supervised Point Cloud Representation Learning, *Chuxin Wang, Yixin Zha, Wenfei Yang, Tianzhu Zhang*
- 370 Representing 3D Shapes with 64 Latent Vectors for 3D Diffusion Models, *In Cho, Youngbeom Yoo, Subin Jeon, Seon Joo Kim*
- 371 Scaling Transformer-Based Novel View Synthesis with Models Token Disentanglement and Synthetic Data, *Nithin Gopalakrishnan Nair, Srinivas Kaza, Xuan Luo, Vishal M. Patel, Stephen Lombardi, Jungyeon Park*
- 372 LINR-PCGC: Lossless Implicit Neural Representations for Point Cloud
- ☆ Geometry Compression, *Wenjie Huang, Qi Yang, Shuting Xia, He Huang, Yiling Xu, Zhu Li*
- 373 CanFields: Consolidating Diffeomorphic Flows for Non-Rigid 4D Interpolation from Arbitrary-Length Sequences, *Miaowei Wang, Changjian Li, Amir Vaxman*
- 374 DriveX: Omni Scene Modeling for Learning Generalizable World Knowledge in Autonomous Driving, *Chen Shi, Shaoshuai Shi, Kehua Sheng, Bo Zhang, Li Jiang*
- 375 Interaction-Merged Motion Planning: Effectively Leveraging Diverse Motion Datasets for Robust Planning, *Giwon Lee, Woosong Jeong, Daehee Park, Jaewoo Jeong, Kuk-Jin Yoon*
- 376 Communication-Efficient Multi-Vehicle Collaborative Semantic Segmentation via Sparse 3D Gaussian Sharing, *Tianyu Hong, Xiaobo Zhou, Wenkai Hu, Qi Xie, Zhihui Ke, Tie Qiu*
- 377 World4Drive: End-to-End Autonomous Driving via Intention-aware Physical Latent World Model, *Yupeng Zheng, Pengxuan Yang, Zebin Xing, Qichao Zhang, Yuhang Zheng, Yinfeng Gao, Pengfei Li, Teng Zhang, Zhongpu Xia, Peng Jia, XianPeng Lang, Dongbin Zhao*
- 378 DATA: Domain-And-Time Alignment for High-Quality Feature Fusion in Collaborative Perception, *Chengchang Tian, Jianwei Ma, Yan Huang, Zhanye Chen, Honghao Wei, Hui Zhang, Wei Hong*
- 379 GauUpdate: New Object Insertion in 3D Gaussian Fields with Consistent Global Illumination, *Chengwei Ren, Fan Zhang, Liangchao Xu, Liang Pan, Ziwei Liu, Wenping Wang, Xiao-Ping Zhang, Yuan Liu*
- 380 Hi-Gaussian: Hierarchical Gaussians under Normalized Spherical Projection for Single-View 3D Reconstruction, *Binjian Xie, Pengju Zhang, Hao Wei, Yihong Wu*
- 381 VisHall3D: Monocular Semantic Scene Completion from Reconstructing the Visible Regions to Hallucinating the Invisible Regions, *Haoang Lu, Yuanqi Su, Xiaoning Zhang, Longjun Gao, Yu Xue, Le Wang*
- 382 TrafficLoc: Localizing Traffic Surveillance Cameras in 3D Scenes, *Yan Xia, Yunxiang Lu, Rui Song, Oussema Dhaouadi, João F. Henriques, Daniel Cremers*
- 383 Towards More Diverse and Challenging Pre-training for Point Cloud Learning: Self-Supervised Cross Reconstruction with Decoupled Views, *Xiangdong Zhang, Shaofeng Zhang, Junchi Yan*
- 384 A Lesson in Splats: Teacher-Guided Diffusion for 3D Gaussian Splats Generation with 2D Supervision, *Chensheng Peng, Ido Sobol, Masayoshi Tomizuka, Kurt Keutzer, Chenfeng Xu, Or Litany*
- 385 Extrapolated Urban View Synthesis Benchmark, *Xiangyu Han, Zhen Jia, Boyi Li, Yan Wang, Boris Ivanovic, Yurong You, Lingjie Liu, Yue Wang, Marco Pavone, Chen Feng, Yiming Li*
- 386 Heatmap Regression without Soft-Argmax for Facial Landmark Detection, *Chiao-An Yang, Raymond A. Yeh*
- 387 Demeter: A Parametric Model of Crop Plant Morphology from the Real World, *Tianhang Cheng, Albert J. Zhai, Evan Z. Chen, Rui Zhou, Yawen Deng, Zitong Li, Kejie Zhao, Janice Shiu, Qianyu Zhao, Yide Xu, Xinlei Wang, Yuan Shen, Sheng Wang, Lisa Ainsworth, Kaiyu Guan, Shenlong Wang*
- 388 3D Gaussian Splatting Driven Multi-View Robust Physical Adversarial Camouflage Generation, *Tianrui Lou, Xiaojun Jia, Siyuan Liang, Jiawei Liang, Ming Zhang, Panjun Xiao, Xiaochun Cao*
- 389 Mixed Signals: A Diverse Point Cloud Dataset for Heterogeneous LiDAR V2X Collaboration, *Katie Z Luo, Minh-Quan Dao, Zhenzhen Liu, Mark Campbell, Wei-Lun Chao, Kilian Q Weinberger, Ezio Malis, Vincent Fremont, Bharath Hariharan, Mao Shan, Stewart Worrall, Julie Stephany Berrio Perez*
- 390 GroundFlow: A Plug-in Module for Temporal Reasoning on 3D Point Cloud Sequential Grounding, *Zijun Lin, Shuting He, Cheston Tan, Bihan Wen*
- 391 RealCam-12V: Real-World Image-to-Video Generation with Interactive Complex Camera Control, *Teng Li, Guangcong Zheng, Rui Jiang, Shuigen Zhan, Tao Wu, Yehao Lu, Yining Lin, Chuanyun Deng, Yegan Xiong, Min Chen, Lin Cheng, Xi Li*
- 392 Exploiting Vision Language Model for Training-Free 3D Point Cloud OOD Detection via Graph Score Propagation, *Tiankai Chen, Yushu Li, Adam Goodge, Fei Teng, Xulei Yang, Tianrui Li, Xun Xu*
- 393 ScenePainter: Semantically Consistent Perpetual

- 3D Scene Generation with Concept Relation Alignment, *Chong Xia, Shengjun Zhang, Fangfu Liu, Chang Liu, Khodchaphun Hirunyaratsameewong, Yueqi Duan*
- 394 FROSS: Faster-Than-Real-Time Online 3D Semantic Scene Graph Generation from RGB-D Images, *Hao-Yu Hou, Chun-Yi Lee, Motoharu Sonogashira, Yasutomo Kawanishi*
- 395 Learning Normals of Noisy Points by Local Gradient-Aware Surface Filtering, *Qing Li, Huifang Feng, Xun Gong, Yu-Shen Liu*
- 396 HUG: Hierarchical Urban Gaussian Splatting with Block-Based Reconstruction for Large-Scale Aerial Scenes, *Mai Su, Zhongtao Wang, Huishan Au, Yilong Li, Xizhe Cao, Chengwei Pan, Yisong Chen, Guoping Wang*
- 397 AMD: Adaptive Momentum and Decoupled Contrastive Learning Framework for Robust Long-Tail Trajectory Prediction, *Bin Rao, Haicheng Liao, Yanchen Guan, Chengyue Wang, Bonan Wang, Jiaxun Zhang, Zhenning Li*
- 398 Unraveling the Effects of Synthetic Data on End-to-End Autonomous Driving, *Junhao Ge, Zuhong Liu, Longteng Fan, Yifan Jiang, Jiaqi Su, Yiming Li, Zhejun Zhang, Siheng Chen*
- 399 BANet: Bilateral Aggregation Network for Mobile Stereo Matching, *Gangwei Xu, Jiaxin Liu, Xianqi Wang, Junda Cheng, Yong Deng, Jinliang Zang, Yurui Chen, Xin Yang*
- 400 Puzzle Similarity: A Perceptually-guided Cross-Reference Metric for Artifact Detection in 3D Scene Reconstructions, *Nicolai Hermann, Jorge Condor, Piotr Didyk*
- 401 Authentic 4D Driving Simulation with a Video Generation Model, *Lening Wang, Wenzhao Zheng, Dalong Du, Yunpeng Zhang, Yilong Ren, Han Jiang, Zhiyong Cui, Haiyang Yu, Jie Zhou, Shanghang Zhang*
- 402 DONUT: A Decoder-Only Model for Trajectory Prediction, *Markus Knoche, Daan de Geus, Bastian Leibe*
- 403 Lidar Waveforms are Worth 40x128x33 Words, *Dominik Scheuble, * Hanno Holzhüter, Steven Peters, Mario Bijelic, Felix Heide*
- 404 Spherical Epipolar Rectification for Deep Two-View Absolute Depth Estimation, *Pierre-André Brousseau, Sébastien Roy*
- 405 ContraGS: Codebook-Condensed and Trainable Gaussian Splatting for Fast, Memory-Efficient Reconstruction, *Sankeerth Durvasula, Sharanshagar Muhunthan, Zain Moustafa, Richard Chen, Ruofan Liang, Yushi Guan, Nilesh Ahuja, Nilesh Jain, Selvakumar Panneer, Nandita Vijaykumar*
- 406 UAVScenes: A Multi-Modal Dataset for UAVs, *Sijie Wang, Siqi Li, Yawei Zhang, Shangshu Yu, Shenghai Yuan, Rui She, Quanjiang Guo, JinXuan Zheng, Ong Kang Howe, Leonrich Chandra, Shrivarshann Srijevan, Aditya Sivasadas, Toshani Aggarwal, Heyuan Liu, Hongming Zhang, Chujie Chen, Junyu Jiang, Lihua Xie, Wee Peng Tay*
- 407 PanoSplat3R: Leveraging Perspective Pretraining for Generalized Unposed Wide-Baseline Panorama Reconstruction, *Jiahui Ren, Mochu Xiang, Jiajun Zhu, Yuchao Dai*
- 408 Seam360GS: Seamless 360° Gaussian Splatting from Real-World Omnidirectional Images, *Changha Shin, Woong Oh Cho, Seon Joo Kim*
- 409 GaussianOcc: Fully Self-supervised and Efficient 3D Occupancy Estimation with Gaussian Splatting, *Wanshui Gan, Fang Liu, Hongbin Xu, Ningkai Mo, Naoto Yokoya*
- 410 GeoSplatting: Towards Geometry Guided Gaussian Splatting for Physically-based Inverse Rendering, *Kai Ye, Chong Gao, Guanbin Li, Wenzheng Chen, Baoquan Chen*
- 411 Wide2Long: Learning Lens Compression and Perspective Adjustment for Wide-Angle to Telephoto Translation, *Soumyadipta Banerjee, Jiaul H. Paik, Debashis Sen*
- 412 LangScene-X: Reconstruct Generalizable 3D Language-Embedded Scenes with TriMap Video Diffusion, *Fangfu Liu, Hao Li, Jiawei Chi, Hanyang Wang, Minghui Yang, Fudong Wang, Yueqi Duan*
- 413 MagShield: Towards Better Robustness in Sparse Inertial Motion Capture Under Magnetic Disturbances, *Yunzhe Shao, Xinyu Yi, Lu Yin, Shihui Guo, Junhai Yong, Feng Xu*
- 414 Focal Plane Visual Feature Generation and Matching on a Pixel Processor Array, *Hongyi Zhang, Laurie Bose, Jianing Chen, Piotr Dudek, Walterio Mayol-Cuevas*
- 415 IM360: Large-scale Indoor Mapping with 360 Cameras, *Dongki Jung, Jaehoon Choi, Yonghan Lee, Dinesh Manocha*
- 416 Leveraging 2D Priors and SDF Guidance for Urban Scene Rendering, *Siddharth Tourani, Jayaram Reddy, Akash Kumbhar, Satyajit Tourani, Nishant Goyal, Madhava Krishna, N Dinesh Reddy, Muhammad Haris Khan*
- 417 CULTURE3D: A Large-Scale and Diverse Dataset of Cultural Landmarks and Terrains for Gaussian-Based Scene Rendering, *Xinyi Zheng, Steve Zhang, Weizhe Lin, Aaron Zhang, Walterio W. Mayol-Cuevas, Yunze Liu, Junxiao Shen*
- 418 Φ-GAN: Physics-Inspired GAN for Generating SAR Images Under Limited Data, *Xidan Zhang, Yihan Zhuang, Qian Guo, Haodong Yang, Xuelin Qian, Gong Cheng, Junwei Han, Zhongling Huang*
- 419 LBM: Latent Bridge Matching for Fast Image-to-Image Translation, ** Clément Chadebec, Onur Tasar, Sanjeev Sreetharan, Benjamin Aubin*
- 420 SparseLaneSTP: Leveraging Spatio-Temporal Priors with Sparse Transformers for 3D Lane Detection, *Maximilian Pittner, Joel Janai, Mario Faigle, Alexandru Paul Condurache*
- 421 Relative Illumination Fields: Learning Medium and Light Independent Underwater Scenes, *Mengkun She, Felix Seegräber, David Nakath, Patricia Schöntag, Kevin Köser*
- 422 GaSLight: Gaussian Splats for Spatially-Varying Lighting in HDR, *Christophe Bolduc, Yannick Hold-Geoffroy, Jean-François Lalonde*
- 423 HumanOLAT: A Large-Scale Dataset for Full-Body Human Relighting and Novel-View Synthesis, *Timo Teufel, Pulkit Gera, Xilong Zhou, Umar Iqbal, Pramod Rao, Jan Kautz, Vladislav Golyanik, Christian Theobalt*
- 424 Super Resolved Imaging with Adaptive Optics, ** Robin Swanson, Esther Y. H. Lin, Masen Lamb, Suresh Sivanandam, Kiriakos N. Kutulakos*
- 425 HVPUNet: Hybrid-Voxel Point-cloud Upsampling Network, *☆ Juhyung Ha, Vibhas Kumar Vats, Soon-heung Jung, Alimoor Reza, David J. Crandall*
- 426 Stealthy Backdoor Attack in Federated Learning via Adaptive Layer-wise Gradient Alignment, *Qingqian Yang, Peishen Yan, Xiaoyu Wu, Jiaru Zhang, Tao Song, Yang Hua, Hao Wang, Liangliang Wang, Haibing Guan*

14:30 - 16:30 Demos 6 (Exhibit Hall I)

- Hybrid Rendering for Multimodal Autonomous Driving: Merging Neural and Physics-Based Simulation, *Máté Tóth, Péter Kovács, Zoltán Bendefy, Zoltán Hortsin, Balázs Teréki, Tamás Matuszka*
- Ubiquitous Ultra-Efficient Mobile Video Super-resolution with Logic Gates, *Felix Petersen, Louis Le Coeur*
- Consensus-Driven Active Model Selection, *Justin Kay, Grant Van Horn, Subhansu Maji, Daniel Sheldon, Sara Beery*

FLOOR PLAN LEVEL 1

Kamehameha Exhibition Hall and Lobby



HAWAII

CONVENTION CENTER

Where Business and Aloha Meet



Hawaii'i Convention Center

1801 Kalākaua Avenue, Honolulu, Hawaii'i 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaiiiconvention.com

Complimentary Wi-Fi Area
Connect to : Hawaii Free Wi-Fi

v2025

FLOOR PLAN LEVEL 3 Meeting Rooms

Complimentary Wi-Fi Area
Connect to : Hawaii Free Wi-Fi

LEGEND	
	Business center
	808 Café
	Escalator (2nd FL parking)
	Escalator (3rd & 4th FL)
	Elevator
	Restroom (Men)
	Restroom (Women)
	Water fountain
	Bottle filler
	Defibrillator
	First aid
	Art display
	Smoking area
	LCD board
	Entrance
	Wi-Fi area
	Vending area



HAWAII
CONVENTION CENTER

Where Business and Aloha Meet



Hawaii'i Convention Center

1801 Kalakaua Avenue, Honolulu, Hawaii'i 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaiiiconvention.com

[f](#) HawaiiConventionCenter [@HawaiiConventionCenter](#)
[t](#) @HICConventionCtr [in](#) Hawaii-Convention-Center

v2025

FLOOR PLAN LEVEL 4

Ballroom and Rooftop Garden



LEGEND

- Escalator (3rd & 4th FL)
- Elevator
- Restroom (Men)
- Restroom (Women)
- Water fountain
- Defibrillator
- Smoking area
- LCD board
- Wi-Fi area

ROOFTOP GARDEN FEATURES:

- 2.5 acre open air rooftop garden
- Floor Load Limit, 100 lbs per net sq. ft.
- Grassy stage area approx. 37 x 42 ft.



Complimentary Wi-Fi Area

Connect to : Hawaii Free Wi-Fi

HAWAII
CONVENTION CENTER

Where Business and Aloha Meet



Hawaii Convention Center

1801 Kalākaua Avenue, Honolulu, Hawaii 96815
tel: 808-943-3500 • fax: 808-943-3599
email: info@hccasm.com • hawaii.convention.com

f HawaiiConventionCenter @HawaiiConventionCenter
@HIConventionCtr in HawaiiConventionCenter

v2025

NURSING POD

The Hawaii Convention Center (HCC) now offers a **private, comfortable and clean space to pump or nurse.**

HCC's **ADA compliant Mamava lactation pod** is spacious, secure and offers a 60" turnaround and grab bars for easy wheelchair access. The pod includes the following features:

- Secure access with Mamava app or access code
- Two benches
- Dual outlet and USB port
- Shelf, mirror, and coat hook
- Ceiling vents and exhaust fan
- Occupancy-activated lighting



Look for the nursing pod located at the **Lobby Level near Exhibit Hall 1.**

To unlock, users may download the Mamava app or use HCC's house phone (located beside the pod) and dial "0" to request for the access code.

For more information about the Mamava app, click [HERE](#).



Information about
the Mamava app

www.hawaiiconvention.com



