CONFERENCE PROGRAM

2026 18TH INTERNATIONAL CONFERENCE ON GRAPHICS AND IMAGE PROCESSING

ICGIP 18th

Wuxi, China | November 6-8, 2026

Co-organized by

南京理工大学 NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

江南大学 JIANGNAN UNIVERSITY

Hosted by

江南大学 JIANGNAN UNIVERSITY

Sponsored by

江苏省计算机学会 JIANGSU COMPUTER SOCIETY

江苏省人工智能学会 JiangSu Association of Artificial Intelligence

TABLE OF CONTENTS

TABLE OF CONTENTS 2
ORGANIZING COMMITTEE 3
WELCOME MESSAGE 6
CONFERENCE VENUE 7
VISITING 8
PRESENTATION GUIDELINES 9
DAILY SCHEDULE 11
KEYNOTE SPEAKER I 19
KEYNOTE SPEAKER II 16
KEYNOTE SPEAKER III 17
KEYNOTE SPEAKER IV 18
INVITED SPEAKER 20
ONSITE SESSION A 28
ONSITE SESSION B 31
ONSITE SESSION C 35
ONSITE SESSION D 39
POSTER SESSION 1 42
POSTER SESSION 2 44
ONLINE SESSION 1 46
ONLINE SESSION 2 50
ONLINE SESSION 3 54
ONLINE SESSION 4 57
ONLINE SESSION 5 61
ONLINE SESSION 6 65
ONLINE SESSION 7 69
ONLINE SESSION 8 73
MEMO 77

ORGANIZING COMMITTEE

Honorary Chair -
Xiaojun Wu, Jiangnan University, China

General Conference Chairs -
Hamido FUJITA, Malaysia-Japan International Instituite of Technology, Universiti Teknologi Malaysia, Malaysia
Ting-Chung Poon, Virginia Polytechnic Institute and State University
Liang Xiao, Nanjing University of Science and Technology, China
Wei Fang, Jiangnan University, China
Pengjiang Qian, Jiangnan University, China

Conference Co-chairs -
Jun Li, China University of Geosciences, Wuhan, China
Jianjun Liu, Jiangnan University, China
Fang Liu, Nanjing University of Science and Technology, China

Program Co-chairs -
Huiyu Zhou, University of Leicester, UK
Shanjun Zhang, Kanagawa University, Japan
Chao Fan, Jiangnan University, China
Tao Zhou, Nanjing University of Science and Technology, China

Publicity Co-chairs -
Cheng Siong Chin, Newcastle University in Singapore, Singapore
Yanhui Guo, University of Illinois Springfield, USA
Shengke Wang, Ocean University of China, China
Meili Wang, Northwest A&F University, China
Jianjun Liu, Jiangnan University, China
Hua Zheng, Zhejiang University Fujian Alumni Association, China

Local Organizing Chairs -
Zhenhua Feng, Jiangnan University, China
Min Jiang, Jiangnan University, China

Web Chair -
Hui Li, Jiangnan University, China

Publication Chair -
Yi Li, Sichuan University, China

Steering Committee -
Hui Yu, University of Glasgow, UK (Steeting Committee Chair)
Zhigeng Pan, Nanjing University of Information Science and Technology, China
Xinhong Hei, Xi'an University of Technology, China
Junyu Dong, Ocean University of China, China
Jianru Xue, Xi'an Jiaotong University, China

Advisory Committees -
Vit Vozenilek, Palacky University, Czech Republic
Ran He, Institute of Automation, Chinese Academy of Sciences (CASIA), China
Cheng Jin, Fudan University, China

Mingyi He, Northwestern Polytechnical University, China
Ruigang Yang, Shanghai Jiao Tong University, China
Hui Kong, University of Macau, China
Ce Zhu, University of Electronic Science & Technology of China
Hongyan Zhang, China University of Geosciences, Wuhan, China

Technical Committees -
Aldo André Díaz Salazar, Federal University of Goiás, Brazil
Alessandro Bruno, IULM University, Italy
Anand Nayyar, Duy Tan University, Vietnam
Asst. Prof. M. Nisanth Kartheek, IIITDM Kancheepuram
Caixia Zheng, Northeast Normal University, China
Chia Kim Seng, Universiti Tun Hussein Onn Malaysia, Malaysia
Chuan Qin, University of Shanghai for Science and Technology, China
Cristian Paul Chioncel, Babe-Bolyai University, Romania
Ghani Albaali, Princess Sumaya University for Technology
Haisheng Li, Beijing Technology and Business University, China
Haruna Abdu, Federal University Lokoja
Jianqing Zhu, Huaqiao University, China
Jiong Zhang, Chinese Academy of Sciences, China
Juanni Liu, National Key Laboratory of Science and Technology on Space Microwave, Xi'an Institute of Space Radio Technology, China
Kannimuthu Subramanian, Karpagam College of Engineering
Li Yuan, University of Science and Technology Beijing, China
Liew Lee Hung, Universiti Teknologi MARA, Malaysia
Linhua Deng, Yunnan Minzu University, China
Mengxi Xu, Nanjing Institute of Technology, China
Ming Xie, Nanyang Technological University, Singapore
Mingyuan Jiu, Zhengzhou University, China
Nguyen Ngoc Phien, Ton Duc Thang University, Vietnam
Prof. Fan Zhang, Beijing University of Chemical Technology, China
Prof. Idy Diop, Cheikh Anta Diop University of Dakar
Prof. Tao Jian, Naval Aviation University, China
Rongchun Li, National University of Defense Technology, China
Ruksar Fatima, KBN College of Engineering
Rushi Lan, Guilin University Of Electronic Technology, China
Shu-Xu Jing, Chang'an University, China
Siddhartha Bhattacharyya, RCC Institute of Information Technology
Sunny Arief Sudiro, STMIK Jakarta STI&K, Indonesia
Tian Ma, Xi 'an University of Science and Technology, China
Tiancheng Xu, Nanjing University of Chinese Medicine, China
Vorapoj Patanavijit, Assumption University, Thailand
Wangsheng Yu, Air Force Engineering University, China
Weiliang Meng, MAIS, CAS Institute of Automation, China
Weirong Liu, Lanzhou University of Technology, China
Xiaoyan Zhang, Shenzhen University, China
Ye Yu, Hefei University of Technology, China
Yew Kee WONG Eric, Hong Kong Chu Hai College, China
Yongfei Zhang, Beihang University, China
Yongqiang Zhao, Northwestern Polytechnical University, China
Yuan - Kai Wang, Fu - Jen Catholic University
Zhenghua Huang, Wuhan University of Science and Technology, China
Ali El-Zaart, Beirut Arab University, Lebanon

Christos Gatzidis, Bournemouth University, UK
Javier Gozalvez, Universidad Miguel Hernandez de Elche, Spain
Jessie R. Balbin, Mapua University, Philippines
Liang Huang, Kunming University of Science and Technology, China
Malik Zawwar Hussain, University of the Punjab, Pakistan
Mohamed Khider, USTHB, Algiers, Algeria
Olarik Surinta, Mahasarakham University, Thailand
Suphongsa Khetkeeree, Mahanakorn University of Technology, Thailand
Xiangyuan Ma, Shantou University, China
Xiaoyu Chen, Nanjing University of Science and Technology, China
Xiwen Zhang, Beijing language and culture university, China
Yanfeng Li, Beijing Jiaotong University, China
Yugen Yi, Jiangxi Normal University, China
Zhaohui Wang, Hainan University, China
Juan Zhang, Shaanxi Normal University, China
Yu Qian, Shandong Women's University, China
Luis Gomez, Universidad de Las Palmas de Gran Canaria, Spain
Witold Pedrycz, University of Alberta, Canada
Chunpeng Wang, Qilu University of Technology, China
Philippe Durand, Conservatoire National des Arts et Métiers, France
Malik Muhammad Nauman, Universiti Brunei Darussalam, Brunei
Jan Kubicek, VSB-Technical University of Ostrava, Czech Republic
Peishun Liu, Ocean University of China, China
Kao Zhang, Nanjing University of Information Science and Technology, China
Przemyslaw Falkowski - Gilski, Gdansk University of Technology, Poland
Chin-Chen Chang, Feng Chia University, Taiwan, China
Darun Kesrarat, Assumption University of Thailand, Thailand
Houjin Chen, Beijing Jiaotong University, China
Jianfeng Lu, Nanjing University of Science and Technology, China
Laiyun Qing, University of Chinese Academy of Sciences, China
Marianne Gino Wata, University of Mindanao, Philippines
Rubén Darío Vasquez-Salazar, Politécnico Colombiano Jaime Isaza Cadavid, Colombia
Sanghun Nam, Changwon National University, South Korea
Shaode Yu, Communication University of China Chaoyang, China
Yoonsik Choe, Yonsei University, South Korea
Youquan Liu, Chang'an University, China
Afaq Ahmad, Sultanate of Oman, Oman
Ching-Nung Yang, National Dong Hwa University
Kamel Hussein Rahouma, Minia University, Egypt
Lifang Wu, Beijing University of Technology, China
Prof. Janaka Rajapakse, Prefectural University of Kumamoto, Japan
Shuang Li, Beijing Information Science and Technology University, China
Suraiya Jabin, Jamia Millia Islamia, India
Xueying Zeng, Ocean University of China, China
Xin Cai, Xinjiang University, China
Mingwei Wang, Hubei University of Technology, China
Liping Yang, Chongqing University, China

WELCOME MESSAGE

You are invited to join 2026 18th International Conference on Graphics and Image Processing (ICGIP 2026) in Wuxi, China during November 6-8, 2026. It's the 18th year of the ICGIP conference. We appreciate the supports from all of you in the past 17 years and you are invited to participate in ICGIP 2026 in Wuxi. ICGIP 2026 is sponsored by Nanjing University of Science and Technology and Jiangnan University; organized by Jiangnan University; and supported by Nanjing University of Information Science and Technology, Jiangsu Computer Society, Jiangsu Association of Artificial Intelligence, Northwest A&F University, Ocean University of China, and China Simulation Federation. The Proceedings of ICGIP 2026 will be published in the SPIE Digital Library and indexed by Scopus, Ei Compendex, Web of Science (CPCI), Inspec, Google Scholar.

ICGIP stands as a premier annual gathering dedicated to the field of Graphics and Image Processing. Its core objective is to showcase the latest research endeavors currently underway. The conference serves as a vibrant platform where scientists, scholars, engineers, and students hailing from universities across the globe, as well as professionals from the industry, can present their ongoing research undertakings. By doing so, it actively promotes the establishment of research - related connections between academia and the industry. ICGIP offers a plethora of opportunities for all participants. Delegates are able to engage in face - to - face exchanges of novel ideas and practical application experiences. This not only facilitates the establishment of business or research partnerships but also enables them to identify potential global collaborators for future joint projects.

On behalf of the conference committee, we thank all the authors, reviewers, and attendees for their contributions and participation in ICGIP 2026. Their dedication and expertise enable us to prepare this high-quality program to make the conference a success. Finally, we wish all the delegates a productive and enjoyable conference.

Conference Committee

2026 18th International Conference on Graphics and Image Processing (ICGIP 2026)

November, 2026

CONFERENCE VENUE



无锡长广溪宾馆

地址：无锡滨湖区太湖新城缘溪道 1 号（江南大学蠡湖校区西北门), 214122

Changguangxi Hotel

Address: No. 1 Yuanxi Road, Taihu New Town, Binhu District, Wuxi (Northwest Gate of Lihu Campus, Jiangnan University), 214122, China

For detail: https://icgip.org/par.html

宾馆地处江南大学校园内西北角，高浪路与缘溪道交界口，与长广溪湿地公园隔河相望，紧邻高速内环公路及无锡城市轨道一号线，地处太湖新城，长广溪湿地公园核心地段，独享蠡湖及长广溪国家湿地公园的双重自然风光。宾馆占地面积 4 万平方米，分为 A 座和 B 座两栋相连的建筑。入住宾馆可共享江南大学所有会议、康体设施和校园的美丽景色。 Wuxi Changguangxi Hotel is located in the core area of Taihu New City, Binhu District. Adjacent to Lihu Campus of Jiangnan University and Changguangxi National Wetland Park, it has complete facilities and an elegant environment, meeting various needs such as business, leisure and study tours.

	Wuxi Shuofang International Airport 22 km by car from the hotel, about 35 minutes
	Wuxi Station 19 km by car from the hotel, about 35 minutes.
	Wuxi East Station 25 km by car from the hotel, about 33 minutes.
	Huishan Station 27 km by car from the hotel, about 50 minutes.

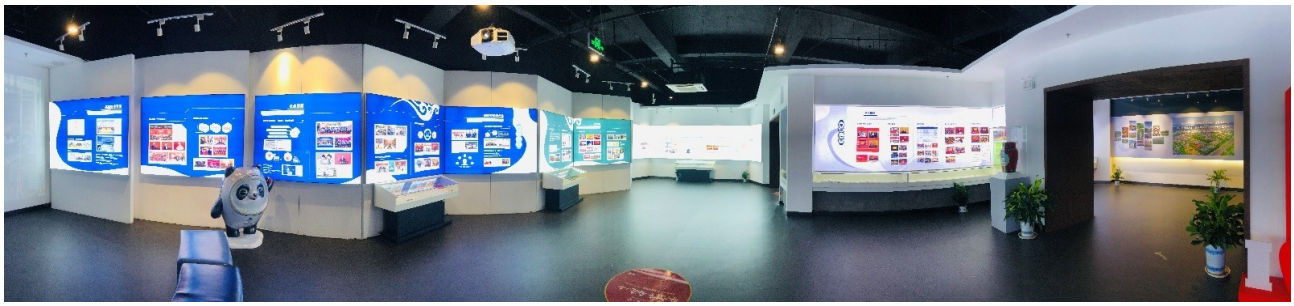


参会人员请从江南大学蠡湖校区**西门**入校，由此可直达长广溪宾馆。

Participants may enter Jiangnan University (Lihu Campus) via the **West Gate**, which provides direct access to Changguangxi Hotel.

VISITING

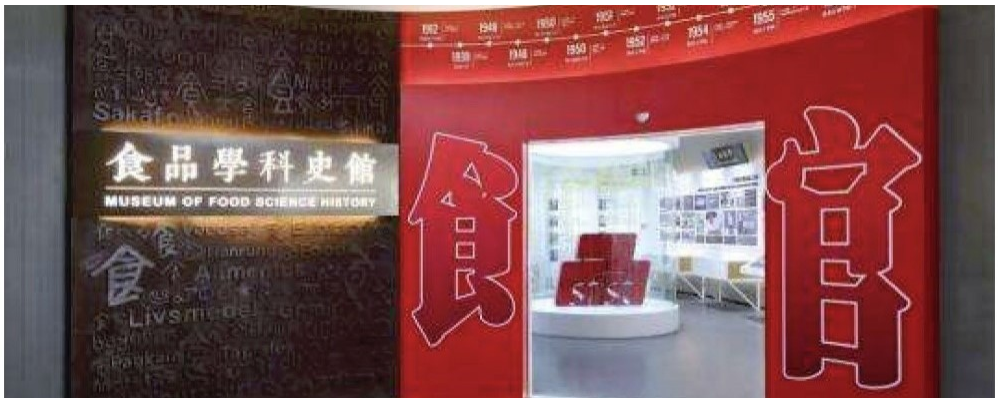
江南大学校史馆-Jiangnan University History Museum



百年沧桑，薪火传承。江南大学的办学始于 1902 年创办于南京堪称"中国师范学堂嚆矢"的三江师范学堂。先后历经两江师范学堂、南京高等师范学校、国立东南大学、国立第四中山大学、国立中央大学、南京大学和南京工学院等时期，1958 年食品工业系迁建无锡轻工业学院。从金陵办学到东迁建校，再到世纪之初组建新的江南大学，学校风风雨雨，一脉相传。寒来暑往，春华秋实，培兰育蕙，硕果累累。

详情：https://jdxsg.jiangnan.edu.cn/bggk/xsgjj.htm

江南大学食品科学馆 | Jiangnan University Food Science Museum



作为全国食品类高校中首家食品学科史馆，该馆不仅是江南大学的文化名片，更是展示江南大学食品学科发展历史，推进食品学科文化传承创新，服务食品行业发展，通过校园文化环境浸润社会的重要载体。该馆藏有相关研究成果、出版物等共计 300 余件展品，包括中国食品系科发展谱系、食品学科领域内著名的图书专著等，充分展示了江南大学食品学科在科学研究、人才培养、产学研合作等方面的重要成果。



注：请参观人员于 11 月 6 日 14:00 自行前往校史馆（图书馆内）门口集合。已登记信息的参观者，信息均已录入进校系统；图书馆临近学校东门。

Note: Registered visitors: please meet at the entrance of the University History Museum (inside the Library) at 14:00, Nov 6. If you already fill the information before, your information has been uploaded to the campus access system. The Library is close to the East Gate.

PRESENTATION GUIDELINES

ORAL PRESENTATION

1. The duration of an oral presentation slot is 15 minutes. Please target your lecture for a duration of about 10 minutes for the presentation plus about 5 minutes for questions from the audience.
2. Your punctual arrival and active involvement in each session will be highly appreciated.
- Get your presentation PPT or PDF files prepared and backed up.
3. Laptops, projector & screen, laser sticks will be provided by the conference organizer.

POSTER PRESENTATION

The size of poster is **59.4 × 84.1cm (A1)**, Posters must be in portrait format(height>width). This cannot be modified.

It's expected that at least one author stands by the poster for (most of the time of) the duration of the poster session. This is essential both to present your work to anyone interest in it and to make sure that your presence is verified by committee.

Meeting rooms	Nov. 7th, 2026
B 座二楼 太湖厅 | Taihu Hall 2F, Building B	Opening Ceremony & Keynote Speeches
A 座 2 楼会议室 2 | Meeting Room 2 (2F), Building A	Onsite Session A, C
A 座 2 楼会议室 5 | Meeting Room 5 (2F), Building A	Onsite Session B, D
A 座 2 楼会议室 6 | Meeting Room 6 (2F), Building A	Onsite Poster Session 1, 2

ONLINE PRESENTATION | Password: 0608

Meeting rooms	Nov. 8th, 2026
Room A: https://us02web.zoom.us/j/81833713809	Online Session 1, 4, 7
Room B: https://us02web.zoom.us/j/87313955014	Online Session 2, 5, 8
Room C: https://us02web.zoom.us/j/81459454849	Online Session 3, 6

Time Zone: **China Standard Time (CST), UTC/GMT+8**

Please make sure that both the clock and the time zone on your computer are set to the correct China standard time.

Presentation Tips

- Each presentation slot is 15 minutes. Please prepare to speak for around 10 minutes, allowing 5 minutes for audience questions.
- Join the meeting room at least 15 minutes before the session begins.

SECURITY

- Please take care of your belongings in public area. For your personal and property safety, delegates are suggested to wear representative card during conference and not to lend it to those

unconcerned to enter event rooms. Conference does not assume any responsibility for loss of personal belongings of participants.

- Don't stay too late in the city, don't be alone in the remote area. Be aware of the strangers who offer you service, signature of charity, etc., at scenic spots. You can search more Tourist Information and Security tips online.
- Emergency Call; Ambulance: 120 Police: 110

Assistant Wechat QR code, remark "ICGIP 2026"	Live Photo Streaming QR Code

DAILY SCHEDULE

Day 1, November 6, 2026	
10:30-16:00	Onsite Sign in and Collect Conference Materials 无锡长广溪宾馆 大堂 | Changguangxi Hotel, Lobby
14:00-15:30	Visiting ➢ 江南大学校史馆-Jiangnan University History Museum ➢ 江南大学食品科学馆-Jiangnan University Food Science Museum For details: https://icgip.org/city.html 详情请参考日程第八页，for details, pls refer to page 8
18:00-19:00	Dinner: Building B, 1F Huaxi Hall | 晚餐: B座，一楼花溪厅
Online Pre-test Timetable and Online Sign	
Room A: https://us02web.zoom.us/j/81833713809 Zoom ID: 818 3371 3809 Password: 0608	
10:00-12:00	Test for keynote speaker, invited speakers, session chairs and technical committees
14:00-15:30	Online Sessions 1, 2, 3, 4
16:00-17:30	Online Sessions 5, 6, 7, 8

Day 2, November 7, 2026

B座二楼 太湖厅 | Building B, 2F Taihu Hall

OPENING CEREMONY

Host: Prof. Zhenhua Feng, Jiangnan University, China

09:00-09:05	**WELCOME MESSAGE** Prof. Weifu Dong, Jiangnan University, China
09:05-09:10	**OPENING REMARKS** Prof. Liang Xiao, Nanjing University of Science and Technology, China
	Group Photo

KEYNOTE SPEECHES

09:10-09:50	**KEYNOTE SPEAKER I** Prof. Christian Sacarea, Babes-Bolyai University, Romania Title: Concept Lattices - The Missing Semantics of Visual Models. A conceptual framework for interlinking Formal Concept Analysis with Graphics and Image Processing
09:50-10:30	**KEYNOTE SPEAKER II (ONLINE)** Prof. Hamido Fujita, Malaysia-Japan International Institute of Technology (MJIIT) Title: Views on Artificial Intelligence and Machine Learning Perspectives: Trustworthiness and Explainability
10:30-10:50	*Coffee Break*

Host: Prof. Tao Zhou, Nanjing University of Science and Technology, China

10:50-11:30	**KEYNOTE SPEAKER III** Prof. Yudong Zhang, Southeast University, China Title: AI-Enabled Colorectal Cancer Research: Progress and Emerging Innovations
11:30-12:10	**KEYNOTE SPEAKER IV** Prof. Josef Kittler, University of Surrey, UK; Jiangnan University, China Title: Intriguing Properties of Foundation Model Representations
12:10-13:30	*Lunch: Building B, 1F Huaxi Hall	晚餐: B座，一楼花溪厅*

AFTERNOON ONSITE PRESENTATIONS		
13:30-15:35	**Onsite Session A: Advanced Imaging, Vision and Graphics** **Chair: Prof. Ching-Nung Yang, National Dong Hwa University** **Invited Speaker- Assoc. Prof. Jing HU** **IP3062, IP3085, IP3079, IP3108, IP3100, IP3124, IP3143**	Meeting Room 2 2F | Building A A座2楼会议室2
13:30-15:35	**Onsite Session B-Computer Vision and Machine Intelligence** **Chair: Prof. Fang Liu, Nanjing University of Science and Technology, China** **Invited Speaker-Assoc. Prof. Jie Guo** **IP3066, IP3093, IP3111, IP3116, IP3117, IP3122, IP3123**	Meeting Room 5 2F | Building A A座2楼会议室5
13:30-15:30	**Poster Session 1- Segmentation, Detection and Tracking for Perception** **Chair: Assoc. Prof. Shuaifeng Zhi, National University of Defense Technology, China** **IP1007, IP1003, IP1020, IP1023, IP1024, IP1028, IP2032, IP2037, IP2040, IP2044, IP2054, IP2058, IP3083, IP3141, IP5001**	Meeting Room 6 2F | Building A A座2楼会议室6
15:35-16:00	*Coffee Break*	
16:00-18:05	**Onsite Session C-Image Processing Fundamentals and Cross-Modal Learning** **Chair: Prof. Jie Guo, Nanjing University, China** **Invited Speaker-Dr. Philippe Durand** **IP1008, IP1011, IP1021, IP2043, IP3078, IP3098, IP3133**	Meeting Room 2 2F | Building A A座2楼会议室2
16:00-18:05	**Onsite Session D- Image Security, Understanding and AI Applications** **Chair: Prof. Linjing Wei, GanSu Agricultural University, China** **Invited Speaker-Assoc. Prof. Shuaifeng Zhi** **IP3086, IP3068, IP3080, IP3090, IP3095, IP3099, IP3140**	Meeting Room5 2F | Building A A座2楼会议室5
16:00-18:00	**Poster Session 2- Multimodal Imaging, Fusion and Reconstruction** **Chair: Prof. Jianjun Liu, Jiangnan University, China** **IP2053, IP1025, IP2041, IP1026, IP2047, IP2048, IP3060, IP2050, IP5002, IP3110, IP3118, IP3112, IP5004, IP3101, IP2059**	Meeting Room 6 2F | Building A A座2楼会议室6
From18:30	*Dinner: Building B, 1F Huaxi Hall | 晚餐: B座，一楼花溪厅*	

Day 3, November 8, 2026

ONLINE PRESENTATIONS

Room A: https://us02web.zoom.us/j/81833713809
Room B: https://us02web.zoom.us/j/87313955014
Room C: https://us02web.zoom.us/j/81459454849
Password:0608

10:00-12:20	**Online Session 1- Small Object Detection and Remote Sensing Perception** **Chair: Prof. Rongchun Li, National University of Defense Technology, China** **Invited Speaker-Prof. Shou Feng** **IP3134, IP1022, IP3102, IP3105, IP3109, IP3130, IP5003, IP3136**	Room A 818 3371 3809
10:00-12:20	**Online Session 2- Reconstruction, Generation and Multimodal Representation** **Chair(s): Prof. Wenhua Qian, YunNan University, China & Assoc. Prof. Ye Yu, Hefei University of Technology, China** **Invited Speaker-Assoc. Prof. Lili Pan** **IP2056, IP3088, IP3091, IP3107, IP3125, IP3129, IP3139, IP1006**	Room B 873 1395 5014
10:00-12:05	**Online Session 3- Segmentation and Boundary-Aware Learning** **Chair: Assoc. Prof. Junfei Shi, Xi'an University of Technology, China** **Invited Speaker-Prof. XiWen ZHANG** **IP2034, IP3061, IP3065, IP3076, IP3082, IP3119, IP3137**	Room C 814 5945 4849
12:05-13:30	***Break Time***	
13:30-15:30	**Online Session 4- Computer Vision and Machine Intelligence (Track 3)** **Chair: Dr. Liew Lee Hung, Universiti Teknologi MARA, Malaysia** **IP1004, IP1019, IP2035, IP2052, IP3075, IP3094, IP3120, IP3138**	Room A 818 3371 3809
13:30-15:30	**Online Session 5- Advanced Deep Learning for Hyperspectral and Remote Sensing Image Processing (Special Session 2)** **Chair: Dr. Huang Nan, Nanjing University of Posts and Telecommunications, China** **IP2038, IP1018, IP2046, IP3071, IP3113, IP3127, IP3072, IP3089**	Room B 873 1395 5014
13:30-16:10	**Online Session 6- Medical Image Analysis and Clinical AI** **Chair: Assoc. Prof. Wei Li, Jiangnan University, China & Assoc. Prof. Shu-Xu Jing, Chang'an University, China** **Invited Speaker- Dr. Ahmad Ali**	Room C 814 5945 4849

	IP1015, IP2036, IP3073, IP3121, IP3126, IP3142, IP5005, IP1016	
15:30-16:00	***Break Time***	
16:00-18:00	**Online Session 7- Computer Vision and Machine Intelligence (Track 3)** **Chair: Dr. Ming Xie, Nanyang Technological University, Singapore** **IP1009, IP1017, IP2039, IP2045, IP3067, IP3081, IP3131, IP3144**	Room A 818 3371 3809
16:00-18:05	**Online Session 8- AI for Earth, Agriculture and Emerging Applications** **Invited Speaker-Prof. Lin Gao** **Chair: Assoc. Prof. Chia Kim Seng, Universiti Tun Hussein Onn Malaysia, Malaysia** **IP3097, IP2031, IP3074, IP3128, IP3087, IP3145, IP3115**	Room B 873 1395 5014

KEYNOTE SPEAKER I



Prof. Christian Sacarea

Babes-Bolyai University, Romania

B 座二楼 太湖厅 | 2F Taihu Hall, Building B | 09:10-09:50, Nov. 7th, 2026

Bio: Christian Săcărea, PhD, is currently holds the position of Vice-Rector. His extensive research interests and contributions lie primarily in the fields of Formal Concept Analysis, Knowledge Representation, Artificial Intelligence, and Data Security. As a scientist, Dr. Săcărea is dedicated to extracting meaningful patterns and structures from complex data, with significant work in knowledge discovery and visualization in domains such as healthcare. He coordinates the Formal Concept Analysis research group at his university, exploring mathematical frameworks for data organization and retrieval. His proficiency in Graph Algorithms and Semantic Web technologies also provides a robust foundation for addressing challenges in 3-D and Surface Reconstruction as well as Computer Vision and Machine Intelligence.

Speech Title: Concept Lattices - The Missing Semantics of Visual Models. A conceptual framework for interlinking Formal Concept Analysis with Graphics and Image Processing

Abstract: Graphics and image processing (GIP) and Formal Concept Analysis (FCA) both work with pictures, but in opposite directions. GIP turns images into descriptions (analysis) and descriptions into images (synthesis). FCA turns data tables into hierarchies of concepts and presents those hierarchies as diagrams for people to reason with. The two fields have more in common than an interest in visual artefacts: both rest on the theory of complete lattices, Galois connections and closure operators, the same theory that underlies mathematical morphology. Each field also lacks something the other has. GIP builds visual models whose conceptual structure, meaning which concepts a model distinguishes and how they are ordered, usually stays implicit. FCA has an explicit, verifiable theory of concepts, but laying out, rendering and exploring its diagrams are still open problems. This keynote develops two things: a theoretical framework assembled from four established bodies of work, and a researcher-developed conceptual framework, the Visual–Conceptual Bridge (VCB). For each ICGIP track we show how the framework guides data collection, analysis and interpretation, and we state the assumptions, limitations and competing perspectives that bound it.

KEYNOTE SPEAKER II



Prof. Hamido Fujita

Malaysia-Japan International Institute of Technology (MJIIT), Malaysia

Zoom ID: 818 3371 3809 Password: 0608 | 09:50-10:30, Nov. 7th, 2026

Bio: Hamido Fujita Life Senior Member IEEE, is Executive Chairman of i-SOMET Incorporated Association, Japan, Professor at Malaysia-Japan International Institute of Technology of Universiti Teknologi Malaysia, and Distinguished Professor at Iwate Prefectural University, Japan, he is also Research Professor at University of Granada, Spain. He is Highly Cited Researcher in Cross-Field for the year 2019 and in Computer Science for the year 2020, 2021, 2022, and 2023 respectively, by Clarivate Analytics. He received Doctor Honoris Causa from Óbuda University, Budapest, Hungary, in 2013 and received Doctor Honoris Causa from Timisoara Technical University, Timisoara, Romania, in 2018, and a title of Honorary Professor from Óbuda University, in 2011. He is Distinguished Research Professor at the University of Granada, and Adjunct Professor with Taipei Technical University, Taiwan, Harbin Engineering University, China and others. He supervised Ph.D. students jointly with the University of Laval, Quebec City, QC, Canada; University of Technology Sydney; Oregon State University, Corvallis, OR, USA; University of Paris 1 Pantheon-Sorbonne, Paris, France; and University of Genoa, Italy. Dr. Fujita is the recipient of the Honorary Scholar Award from the University of Technology Sydney, in 2012. He was the Editor-in-Chief for Knowledge-Based Systems (Elsevier) (2005-2019) and then Emeritus Editor of Knowledge-Based Systems in 2020~. Since 2020 he is currently the Editor-in-Chief of Applied Intelligence (Springer) and the Editor-in-Chief of International Journal of Healthcare Management (Taylor & Francis). He headed a number of projects including intelligent HCI, a project related to mental cloning for healthcare systems as an intelligent user interface between human-users and computers, and SCOPE project on virtual doctor systems for medical applications. He collaborated with several research projects in Europe, and recently he is collaborating in OLIMPIA project supported by Tuscany region on Therapeutic monitoring of Parkinson disease. He has published more than 480 articles mainly in high impact factor journals.

Speech Title: Views on Artificial Intelligence and Machine Learning Perspectives: Trustworthiness and Explainability

Abstract: The hot topics in training in Machine Learning is a crucial aspect that affects the credibility of the system in terms of performance and is employed for robust applications such as in healthcare systems. Machines or algorithms, in wide challengeable applications in security or vision or health care early predictions, learn from data. Nevertheless, in most cases, the extensive and unbalanced data and noise make it unreliable (needs resilient and robustness) in prediction accuracy. Supervised machine learning is and was one of the aspects of providing artificial intelligence-based solutions. However, this is and was limited due to the difficulty of labelling big data and many crucial problems in weak relations and noise in data. Semi-supervised learning, for example, Multiview learning, could assist in solving these problems. In much published research, there are still problems in providing machine learning models that are unbiased and efficient in terms of robustness and resilience in data-driven systems. Multi-class classification still has problems in terms of clear definition in class classification, bias, imbalance and weak relations, making machine learning for multi-class classification insecure for classification or regression analytics. Also I will highlight the explainable AI and it is role to provide trustworthiness in AI solutions.

KEYNOTE SPEAKER III



Prof. Yudong Zhang

Southeast University, China

B 座二楼 太湖厅 | 2F Taihu Hall Building B | 10:50-11:30, Nov. 7th, 2026

Bio: Dr. Yudong Zhang is a Chief Professor and Doctoral Supervisor at Southeast University, and a recipient of a national-level high-level talent award. He was a postdoctoral researcher from Columbia University. He has held full-time positions at universities in seven countries. He has co-authored over 100 papers published in CAS Q1 journals. His Google citation is over 52k with an h-index of 123. He has served as (co-)principal investigator for numerous domestic and international academic projects, including national key R&D projects, National Institutes of Health (NIH) projects, and BBSRC projects.

Speech Title: AI-Enabled Colorectal Cancer Research: Progress and Emerging Innovations

Abstract: Colorectal cancer (CRC) remains a formidable challenge within the global healthcare landscape, necessitating urgent and systemic improvements in its management paradigms. This report outlines the latest advancements in CRC research and clinical practice, focusing on the transformative role of integrated innovations across diagnosis, treatment, and translational medicine—specifically highlighting the pivotal supportive value of AI and image processing technologies. The report underscores the critical strategic significance of technological breakthroughs—including advanced diagnostic methods, computer-aided solutions, AI and image processing technologies, and novel therapeutic strategies—in advancing the diagnosis and treatment of CRC. In particular, AI and image processing technologies facilitate the precise identification of early-stage lesions, the accurate differentiation between benign and malignant conditions, and the dynamic monitoring of disease progression through the intelligent analysis of colorectal endoscopic and pathological images; this significantly enhances diagnostic efficiency and accuracy, thereby providing highly effective support for clinical decision-making. Finally, the report delineates key directions for future research—including the clinical validation, safety assessment, and preclinical translational investigation of AI and image processing technologies—while emphasizing the synergistic interplay between basic research and clinical application, thereby offering a strategic outlook for alleviating the global burden of colorectal cancer.

KEYNOTE SPEAKER IV



Prof. Josef Kittler

University of Surrey, UK

Jiangnan University, China

B 座二楼 太湖厅 | 2F Taihu Hall Building B | 11:30-12:10, Nov. 7th, 2026

Bio: Josef Kittler is Professor of Machine Intelligence at the Centre for Vision, Speech and Signal Processing, University of Surrey. He is also Distinguished Professor at the School of Artificial Intelligence and Computer Science, Jiangnan University. He received his BA, PhD and DSc degrees from the University of Cambridge. He teaches and conducts research in Machine Intelligence, with a focus on Biometrics, Video and Image Database retrieval, Deep Learning and Cognitive Vision. He published a Prentice Hall textbook on Pattern Recognition: A Statistical Approach and several edited volumes, as well as more than 1000 scientific papers, including more than 300 journal publications. He serves on the Editorial Boards of Pattern Recognition, Pattern Recognition Letters and Springer Nature Computer Science. He was Series Editor of Springer Lecture Notes on Computer Science 2004-2016.

Speech Title: Intriguing Properties of Foundation Model Representations

Abstract: The ability of foundation models to produce effective representations for immensely complex problems is absolutely staggering. The information content of these representations is beyond human comprehension and is commonly perceived as magic. It has the capacity to be probed by queries and prompts, revealing plethora of detail about the input visual stimulus. However, the black box nature of their operation hinders explainability of decisions, and this in turn may impede the adoption of impressive solutions in practice. It is pertinent to ask, what the underlying essence of deep representation is. This is not only motivated by the natural curiosity of what's the trick behind the magic? Over and above such human interest, it is a serious scientific question. A deeper knowledge of the nature of deep representations could direct the search for more efficient ways of unlocking and exploiting the rich information content it captures, as well as helping explainability. The talk will explore deep representations generated by AI models and discuss the implications of the findings.

INVITED SPEAKER



Lin Gao, Professor

University of Chinese Academy of Sciences, China

Room B: 873 1395 5014 Password:0608 Nov. 8, 16:00-16:20

Bio: Gao Lin is a Professor and PhD Supervisor at the Institute of Computing Technology, Chinese Academy of Sciences (CAS), Deputy Director of the Ubiquitous Computing System Research Center, and a Professor at the University of Chinese Academy of Sciences. He earned his PhD from Tsinghua University. His research focuses on computer graphics and 3D computer vision. He has published over 100 papers in top venues including SIGGRAPH, TPAMI, and TVCG. His face AIGC application is used by users in more than 180 countries/regions worldwide. He currently/formerly serves as Secretary-General of the Asian Graphics Association and CSIG Intelligent Graphics Committee, Program Committee Member of SIGGRAPH, Area Chair of CVPR, and Associate Editor of IEEE TVCG. As PI, he leads projects including the National Key R&D Program and NSFC Excellent Young Scientist Fund. His awards include the Asian Graphics Young Scholar Award, Wu Wenjun AI Outstanding Young Award, and CCF First Prize for Technical Invention.

Speech Title: Dual-Engine-Driven Digital Space Modeling via Explicit Geometric Representations and Implicit Generative Models

Abstract: This talk centers on a world-model construction paradigm of “explicit–implicit integration with dual-engine driving.” It leverages the high interpretability and geometric/physical fidelity of explicit methods, together with the strong scalability of implicit methods, and employs feed-forward inference and distillation to enable more controllable and realistic generative world models. We first introduce explicit geometric modeling built upon video generative models, and then present geometry-prior-enhanced implicit approaches, including view-controllable multi-trajectory video generation, multi-view multimodal driving video generation, and compressed representation and action modeling for embodied worlds. On this basis, we discuss how explicit geometric constraints can be exploited to improve the consistency and plausibility of generated results, and how generative priors can be used to infer unobserved regions and dynamics, thereby supporting single-image digital human reconstruction and animation, inverse rendering of autonomous driving scenes, and relighting of dynamic scenes. The talk demonstrates a technical pathway toward interpretable, interactive, and physics-compliant digital world construction.



Shou Feng, Professor

Harbin Engineering University, China

Room A: 818 3371 3809 Password:0608 Nov. 8, 10:00-10:20

Bio: Feng Shou, Professor, Ph.D Supervisor of Harbin Engineering University, Deputy Director of the Key Laboratory of Advanced Ship Communication and Information Technology of the Ministry of Industry and Information Technology, IEEE member, Senior member of the Chinese Society of Communications, Member of the Imaging Detection and Perception Committee of the Chinese Society of Image and Graphics, peer review expert of the National Natural Science Foundation of China, academic dissertation review expert of the Ministry of Education, Visiting Scholar of Indiana University Bloomington, Guest Editor of Remote Sensing, an international authoritative journal in the field of remote sensing. Member of the editorial board of international Journal Frontiers in Imaging, American Journal of Remote Sensing, He also serves as a reviewer for many authoritative academic journals such as IEEE TIP, IEEE TNNLS, IEEE TGRS, and etc. In the past three years, he has published 30 academic papers as the first/corresponding author in top journals in the field of Remote Sensing such as IEEE TIP, IEEE TNNLS, IEEE TGRS, and 7 papers have been selected as ESI highly cited papers.

Speech Title: Modality-Adaptive Fractional-Domain Decoupling and Interaction Network for Multi-Source Remote Sensing Image Classification

Abstract: Jointly exploiting the rich spectral information of hyperspectral imagery (HSI) and the geometric and structural information of active sensors (LiDAR/SAR) has attracted increasing attention in the remote sensing classification community. To reduce dependence on large-scale annotations, recent methods incorporate the discrete wavelet transform (DWT) as a frequency domain prior to provide deterministic structural decomposition. However, DWT is confined to the fixed Fourier frequency axis, offering no flexibility to choose a more suitable analysis perspective for signals with different non-stationary characteristics across heterogeneous modalities. To address this issue, we propose a modality-adaptive fractional-domain decoupling and interaction network (MAFrDNet), which introduces modality adaptive fractional-domain decoupling analysis into multi-source remote sensing classification. Experiments on two public datasets indicate that MAFrDNet achieves state-of-the-art classification performance with competitive efficiency.



Jie Guo, Associate Professor

Nanjing University, China

2F Meeting Room 5 | A 座 2 楼会议室 5 | Nov. 7, 13:30-13:50

Bio: Dr. Jie Guo is an associate professor (with tenure) in the School of Computer Science at Nanjing University. He received his PhD from Nanjing University in 2013. His current research interest is mainly in computer

graphics, virtual reality and 3D vision. He has over 100 publications in internationally leading conferences (SIGGRAPH, SIGGRAPH Asia, CVPR, ICCV, ECCV, IEEE VR, etc.) and journals (ACM ToG, IEEE TVCG, IEEE TIP, etc.).

Speech Title: Neural Spatial-temporal Super-sampling for Real-time Rendering

Abstract: Spatio-temporal super-sampling techniques, such as super-resolution and frame interpolation/extrapolation, are currently the most critical post-processing techniques in the field of real-time rendering. With the rapid advancement of AI technologies, such as deep neural networks, AI-assisted spatio-temporal super-sampling methods have gradually become a research focal point due to the superior performance they demonstrate compared to traditional methods. This talk will present our latest research progress in spatio-temporal super-sampling for real-time rendering.



Xiwen Zhang, Professor

Beijing Language and Culture University, China

Room C: 814 5945 4849 Password:0608 Nov. 8, 10:00-10:20

Bio: XiWen Zhang is currently a full professor in the Digital Media Department, School of Information Science, Beijing Language and Culture University. From 2002 to 2007, Prof. Zhang served as an associate professor at the Human-Computer Interaction Laboratory, Institute of Software, Chinese Academy of Sciences; he was a postdoctoral researcher supervised by Prof. Michael R. Lyu in the Department of Computer Science and Engineering, The Chinese University of Hong Kong from 2005 to 2006, and a postdoctoral researcher supervised by Prof. ShiJie Cai in the Department of Computer Science and Technology, Nanjing University from 2000 to 2002. His research interests cover pattern recognition, computer vision, and human-computer interaction, along with their applications in digital images, videos and ink art. Prof. Zhang has published more than 60 peer-reviewed journal and conference papers, including SCI papers in Pattern Recognition, IEEE Transactions on Systems, Man, and Cybernetics – Part B, and Computer-Aided Design, as well as over twenty EI-indexed papers. He obtained his B.E. in Chemical Equipment and Machinery from Fushun Petroleum Institute (renamed Liaoning Shihua University in 2002) in 1995, and his Ph.D. in Mechanical Manufacturing and Automation from Dalian University of Technology in 2000, under the supervision of Prof. ZongYing Ou.

Speech Title: Three views on intelligently extracting and generating information from image

Abstract: Due to pattern recognition and deep learning, various information can be extracted and generated from images. Our work focuses on the proposed hierarchical models, local homogeneity, and adversarial generation. We process diverse digital images, including scans of mechanical paper drawings and paper text, face images, line-drawn portraits, and microscopic bone marrow images. With the proposed hierarchical models, we extract various information: a hierarchical model for engineering drawings is adopted to extract and recognize graphics and their multi-level compound objects from scanned mechanical paper drawings, while a facial model is used to extract faces and their components from photographs. Relying on the proposed local homogeneity, we extract karyocytes and their components from microscopic bone marrow images based on regional color features. We also extract and generate diverse information from images using cycle-consistent adversarial networks, for instance

separating text from grid backgrounds; moreover, conditional generative adversarial networks are employed to generate digital images of traditional Chinese noble lady paintings from line drawings.



Philippe Durand

CNAM, France

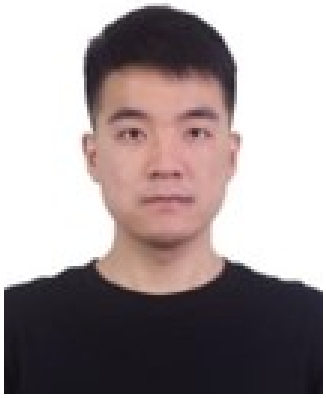
2F Meeting Room 2 | A 座 2 楼会议室 2 | Nov. 7, 16:00-16:20

Bio: Philippe Durand is a Senior Lecturer in the Department of Mathematics and Statistics at the Conservatoire National des Arts et Métiers (CNAM) in Paris, and a member of the Mathematical and Numerical Modelling Department (M2N). His research lies at the interface between mathematical engineering and theoretical mathematics. He has worked on the mathematization of gauge theories in physics and string theory, on tensor analysis applied to networks, and on the application of topological and statistical methods to image processing.

Speech Title: Morphological Granulometry (VARAN) and Topological Extraction (Sentinel-1/2, Le Luc)

Abstract: This invited lecture presents a complete mathematical framework for extracting urban structures built-up areas, residential blocks, and transport networks from highly noisy radar imagery acquired over the city of Le Luc in south-eastern France. Radar images of semi-urban environments are difficult to interpret because of speckle noise, non-Gaussian backscatter distributions, and the multiple scattering mechanisms generated by buildings, roads, and heterogeneous land cover. Our objective is to recover a coherent representation of the urban fabric, comparable to the structures visible in the optical aerial image, while relying on radar data and mathematically robust processing tools. We develop an integrated approach combining three complementary paradigms. The first one is Mathematical Morphology, including Alternating Sequential Filters, granulometry, directional openings, and morphological skeletonization. The second one is Topological Data Analysis (TDA), based on persistent homology, H0–H1 generators, and Wasserstein distances. The third one is an approach inspired by statistical physics, based on binary Ising Markov random fields and simulated annealing, used as a spatially coherent decision model for urban versus non-urban segmentation. Morphology provides a multi-scale geometric description capable of reducing speckle and isolating coherent radar responses associated with buildings and road segments. TDA supplies complemen- tary global information on the organization of urban textures, revealing stable topological patterns loops, cavities, and block structures — that remain robust under noise and local fluctuations. The Ising model introduces an energy-based formulation in which the unary term is built from radar observables such as log-intensity, local contrast, gradient magnitude, and possibly anisotropy descriptors, while the pairwise term enforces spatial coherence between neighboring pixels. Simulated annealing is then used to minimize the corresponding non-convex energy and to avoid poor local minima, producing a regularized binary urban mask that is particularly well suited to noisy radar data. Applied to the Le Luc dataset, this combined methodology enables the recovery of major urban zones, including the historical nuclei VN1, VN2, and VN3, residential housing areas, dense collectivehousing blocks, and the main transport axes such as the highway, major roads, and railway lines. The morphological segmentation is supported by topological signatures, as urban regions exhibit a richer distribution of persistent H1 generators, whereas background areas are

characterized by simpler connectivity profiles dominated by H0 components. The Ising-simulated annealing stage acts as a robust decision layer, transforming local radar descriptors into a spatially stable partition of the scene. This talk highlights the synergy between geometry, topology, and statistical physics for radar image analysis. It shows that the combination of morphological filtering, topological persistence and Ising-type regularization yields an interpretable, noise-tolerant, and mathematically structured pipeline for urban extraction under severe speckle conditions. Perspectives include extensions to multi-temporal radar sequences, polarimetric acquisitions, and integration into near-real-time environmental monitoring systems.



Shuaifeng Zhi, Associate Professor

National University of Defense Technology (NUDT), China

2F Meeting Room 5 | A 座 2 楼会议室 5 | Nov. 7, 16:00-16:20

Bio: Shuaifeng Zhi currently an Associate Professor at the Department of Electronic Science and Technology, National University of Defense Technology (NUDT), China. He received the PhD degree in computing research from the Dyson Robotics Laboratory, Imperial College London in 2021, supervised by Prof. Andrew J. Davison. His current research interests focus on 3D vision, particularly on multi-modal scene understanding and scene representation learning. He is selected for the 10th Youth Elite Scientists Sponsorship Program by CAST and the grantee of HNNSF for the Excellent Young Scientists Fund. He has published over 20 papers in prestigious international journals and conferences including T-PAMI, CUSR, T-GRS, ICCV(Oral), etc. His Google Scholar citations exceed 1,800, with the highest-cited single first-author paper reaching 700+ citations. He serves as an Editorial Board member of The Visual Computer and has regularly served as a reviewer for leading academic venues.

Speech Title: Dense Registration of Optical-SAR Remote Sensing Images under Large Geometric Transformations

Abstract: Registration of optical and synthetic aperture radar (SAR) remote sensing images serves as a critical foundation for image fusion and visual navigation tasks. It is particularly challenging because of their modal discrepancy, primarily manifested as severe nonlinear radiometric differences (NRD), geometric distortions, and noise variations, especially under large geometric transformations. This talk presents our recent progress in establishing dense correspondence for optical-SAR images using optical flow estimation, which outperforms state-of-the-art registration algorithms in terms of accuracy and robustness. We will discuss open challenges and outlooks in achieving generalizable registration of optical-SAR images.



Jing Hu, Associate Professor

Xi'an University of Technology, China

2F Meeting Room 2 | A 座 2 楼会议室 2 | Nov. 7, 13:30-13:50

Bio: Jing Hu received the B.S. degree from Nanchang University, Jiangxi, China, in 2013, the Ph.D. degree from Xidian University, Shaanxi, China, in 2018. She is currently working as an Associate Professor with the School of Computer Science and Engineering, Xi'an University of Technology, China. She is the author or coauthor of more than 20 publications on image processing, including international journal papers, conference papers, and book chapters. She has also been a reviewer of several journals, such as IEEE TPAMI, TIP, TCSVT, TGRS. Her research interests include remote sensing image processing, deep learning and pattern recognition.

Speech Title: Automated Loading Proportion Estimation for Coking Coal Plants via Image Segmentation

Abstract: In coking plants, truck loading is typically monitored manually, leading to inconsistent judgments, reduced efficiency, and potential safety risks. This paper proposes an automated loading proportion estimation method based on dual-camera imaging and semantic segmentation. Images of the truck interior are captured from both sides of the loading area, and a segmentation model is used to estimate the proportion of the loaded region. To address the lack of publicly available datasets for this task, we construct a dedicated dataset of truck loading scenes with manually annotated segmentation masks. We further investigate model configurations involving encoder architectures, attention mechanisms, and multi scale feature extraction, and develop an optimized U-Net-based segmentation model incorporating an EfficientNet-B4 encoder, SCSE attention, ASPP, and a U-Net++ decoder. Experiments demonstrate the feasibility of the proposed model for loading-region segmentation on the constructed dataset. The resulting method provides a practical basis for automated loading monitoring in coking plants.



Lili Pan, Associate Professor

University of Electronic Science and Technology of China, China

Room B: 873 1395 5014 Password:0608 Nov. 8, 10:00-10:20

Bio: Dr. Lili Pan is an associate professor at the University of Electronic Science and Technology of China (UESTC), where her recent research focuses on multi-modal large language models (MLLMs), LLM-driven scientific innovation, and continual learning. She has led several significant national research projects funded by the National Natural Science Foundation of China (NSFC). With over 30 publications in prestigious journals and at conferences such as ICML, CVPR, and ECCV, Dr. Pan has made notable contributions to her field. She obtained her Ph.D. in Information Engineering from UESTC and

was a visiting Ph.D. student at the Robotics Institute, Carnegie Mellon University (CMU), from 2009 to 2011.

Speech Title: Test-Time Compute for Diffusion Models: Choosing the Trajectory, Then Learning from It

Abstract: Spending extra compute at inference time is one of the simplest ways to improve a text-to-image diffusion model for a single prompt without changing its weights. Two ideas dominate: search, which keeps the best trajectory the sampler proposes, and per-prompt test-time training (TTT), which optimizes a reward directly. Both fall short. Search is bounded by what the sampler offers and forgets the good trajectories it finds, while TTT on a single drawn trajectory raises the optimized reward, yet the held-out rewards stall at, or fall below, what plain search reaches. In this talk, I will present two studies that tackle these limits by first deciding which trajectory to workfrom and then learning from it. SEA (Search--Exploit--Amortize) gives noise search a memory: a PUCT tree reuses high-reward champions, and a per-prompt adapter trained by self-distillation amortizes what the search finds, giving the best average rank over four preference metrics at a matched budget of 96 scored candidates. SETT lets TTT choose its trajectory while the image is still forming, by re-noising the latent into candidates, keeping the best preview, and training on that one alone; it ranks first on the optimized reward and on both held-out rewards against nine baselines, with the shortest wall-clock time. The takeaway is that how compute is spread acrosstrajectories matters as much as how each one is optimized.



Ahmad Ali, Doctor Lecturer

Hainan Bielefeld University of Applied Sciences, China

Room C: 814 5945 4849Password:0608 Nov. 8, 13:30-13:50

Bio: Dr. Ahmad Ali received his Ph.D. from Shanghai Jiao Tong University, China, and subsequently completed his postdoctoral research at Shenzhen University, China. He is currently working as a Doctor Lecturer at Hainan Bielefeld University of Applied Sciences, China. His research interests include artificial intelligence, machine learning, deep learning, intelligent transportation systems, spatio-temporal data analytics, and multimodal learning. Dr. Ali has published high-quality research articles in well-recognized international journals, including Neural Networks, Information Sciences, Computer Science Review, Future Generation Computer Systems, IEEE Transactions on Intelligent Transportation Systems, and IEEE Transactions on Consumer Electronics. He has also contributed to reputable CCF-ranked international conferences, including ICPADS and the ACM/IEEE International Conference on Embedded Systems. His current research focuses on developing intelligent, data-driven, and multimodal learning approaches for smart transportation and urban computing applications.

Speech Title: Attention-Driven Dynamic Spatio-Temporal Graph Learning for Multimodal Urban Traffic Flow Prediction

Abstract: Accurate prediction of urban traffic flow is essential for intelligent transportation systems (ITS) and effective smart city management. However, traffic dynamics are strongly influenced by multiple temporal patterns and external conditions, making it challenging for conventional models to fully capture complex spatio-temporal dependencies. To address these challenges, we propose an Attention-

Driven Dynamic Spatio-Temporal Graph Network (Att-DSTGNet) that integrates multimodal information and adaptive attention mechanisms for large-scale traffic forecasting. The proposed framework jointly models four key factors shaping traffic evolution: recent patterns, daily patterns, weekly patterns, and external influences such as weather and events. Att-DSTGNet employs a multi-scale dynamic graph convolutional network to represent evolving spatial dependencies among road segments, while a temporal attention encoder selectively captures fine-grained temporal correlations within and across these factors. To incorporate multimodal context, a Vision Transformer-based module extracts high-level meteorological features, which are adaptively fused with traffic representations through a cross-modal attention mechanism. Extensive experiments on large-scale real-world datasets demonstrate that Att-DSTGNet consistently outperforms state-of-the-art baselines in multi-horizon traffic prediction tasks.

ONSITE SESSION A

- **Topic: Advanced Imaging, Vision and Graphics**
- **Location: Building A, 2F Meeting Room 2 | A 座 2 楼会议室 2**
- **Time: 13:30-15:35, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Prof. Ching-Nung Yang, National Dong Hwa University**

Invited Speaker 13:30-13:50	**Jing Hu, Associate Professor, Xi'an University of Technology, China** **Speech Title: Automated Loading Proportion Estimation for Coking Coal Plants via Image Segmentation**
IP3062 13:50-14:05	Title: Cross-Vocoder Audio Deepfake Detection Based on Layer Adaptive Fusion and Multi-resolution Spectrograms Author(s): Xiaoyan Du, Wenjun Zhang, Jiamei Cai Presenter: Xiaoyan Du, Shaanxi Police College, China Abstract: With the rapid development of speech synthesis and voice cloning technologies, audio deepfakes pose severe threats to personal privacy, financial security, and public trust. Existing detection methods typically rely on the last-layer representation of self-supervised pre-trained models as input features, neglecting the multi-granularity acoustic information encoded across different layers, which leads to insufficient generalization when facing unseen vocoders. To address these issues, this paper proposes an audio deepfake detection method based on Layer Adaptive Fusion (LAF) and multi-resolution spectrograms, with a focus on cross-vocoder generalization scenarios. The method employs a frozen WavLM-base-plus backbone and uses an additive attention mechanism to perform utterance-adaptive weighted fusion of 13 hidden-layer representations, while introducing three-resolution log-mel spectrograms processed by a lightweight CNN to extract complementary time-frequency features. The two streams are concatenated and fed into a fusion classifier for authenticity discrimination. Experiments on the ASVspoof 2019 LA dataset demonstrate that the proposed method reduces the cross-vocoder evaluation set EER from 4.77% (baseline) to 1.27% (73% relative reduction), with the in-domain development set EER reaching 0.43%. Additionally, this paper conducts an exploratory analysis of gradient-reversal adversarial training and supervised contrastive learning as two regularization strategies. Results show that the SupCon variant, while not improving EER, achieves the best min t-DCF of 0.0101, providing a reference value for high-security deployment scenarios.
IP3085 14:05-14:20	Title: Context-bound proactive watermarking via semantic hashing for face-context consistency verification Author(s): Yingxue Sun, Hongjin Yan, Jinshuai Gao, Songjie Wei Presenter: Yingxue Sun, Nanjing University of Science and Technology, China Abstract: Robust watermarking can preserve source identifiers after image processing, yet successful recovery does not establish whether a marked face remains consistent with its surrounding scene. This limitation arises when the watermarked face is retained but the background is replaced. To address this problem, this paper proposes a context-bound proactive watermarking framework that stores a semantic reference to the non-facial scene within the facial region. A compact descriptor derived from masked-scene features is combined with a source trace code and embedded as a single watermark payload. During verification, the recovered reference is compared with a descriptor independently generated from the received background. This process requires neither the original image nor an external hash registry. Experiments examine watermark fidelity and recovery, semantic-hash stability, empirical threshold calibration, and controlled background substitution. The results support cross-region semantic binding as a means of assessing face-context consistency when the marked facial region remains available for verification.
IP3079 14:20-14:35	Title: Region-Selective Normal Constraints for 3D Gaussian Reconstruction in Architectural Scenes Author(s): Xinran Shao, Junping Qin, Dan Wang, Zezhou Shao, Shujia Kong, Yangyang Luo

	Presenter: Xinran Shao, Inner Mongolia University of Technology, China Abstract: In architectural 3D Gaussian reconstruction, weakly textured planes, glass facades, and locally complex structures such as window frames, railings, and roof edges often lack sufficient multi-view constraints, leading to local blur, boundary ambiguity, and structural discontinuities. These problems usually do not appear as complete reconstruction failures, but as local quality degradation within challenging regions. To address this issue, this paper proposes a challenging-region normal-constrained 3D Gaussian reconstruction method for architectural scenes. The proposed method first detects challenging regions from intermediate 3D Gaussian renderings by combining rendering errors, visual representation features, and geometry-related responses, and generates challenging-region masks. Then, a diffusion-based geometry estimation model is used to extract normal priors from input views, which are introduced into subsequent optimization as auxiliary geometric information. During region-selective optimization, the challenging-region masks restrict the effective range of additional normal constraints, allowing external normal priors to mainly participate in the optimization of weakly textured regions and locally unstable structures. Experiments are conducted on the Blume scene from the UAV3D dataset and real drone-captured architectural data. The results show that the proposed method maintains stable global metrics for building-level 3D reconstruction while improving local reconstruction quality within challenging regions, especially in terms of local structural continuity, boundary clarity, and stability in weakly textured regions.
IP3108 **14:35-14:50**	Title: DSTF: Dual-Stream Time-Frequency Fusion with Adaptive Multi-Resolution and Frequency-Structure Modeling for Underwater Acoustic Target Recognition Author(s): LIU Yongkang, REN Lansen, ZHU Yunan, LIU Zhengjun, WEI Linzhe, LI Baoqi Presenter: Liu Yongkang, University of Chinese Academy of Sciences, China Abstract: Passive underwater acoustic target recognition (UATR) is important for marine monitoring and security, but reliable ship classification remains difficult because transient pulses and harmonic lines require different time-frequency (TF) resolu tions and because weak line structure is easily lost under low signal-to-noise ratio (SNR) and recording-domain shift. We propose DSTF, a dual-stream TF fusion framework that learns complementary evidence from adaptive multi-resolution textures and from stable frequency structure. The two representations are combined by a lightweight late-fusion classifier, allowing the model to exploit local transients without sacrificing narrowband frequency cues. Under a unified recording level protocol on DeepShip, DSTF improves segment and recording-group classification, with the largest gains in low-SNR and cross-recording evaluations. Ablation and robustness experiments support the central claim that complementary TF views and explicit frequency structure are more useful than a single fixed-resolution representation.
IP3100 **14:50-15:05**	Title: Frequency-Aware Density Control for 3D Gaussian Splatting with Wavelet-based Visual Primitives Author(s): Linlin Wei, Zhihui Wei Presenter: Linlin Wei, Nanjing University of Science and Technology, China Abstract: 3D Gaussian Splatting (3DGS) consists of four collaborative modules: primitive representation, projective rendering, loss function, and density control. WIPES improves the primitive representation by extending Gaussians to wavelet primitives with explicit frequency parameters, enabling a single primitive to encode richer texture information and expanding the representation space from purely geometric to a joint "geometry-frequency" space. However, for primitives with frequency parameters, existing density control strategy still adheres to the Gaussian primitive framework. While such strategies remain applicable to wavelet primitives, they fail to exploit the frequency dimension—resorting solely to primitive proliferation even when frequency adjustment would be more effective, thus leaving the compactness potential of wavelet primitives largely untapped. This paper fills this gap by proposing a frequency-aware density control framework. Specifically, a rank-normalized position dominance metric, built from percentile ranks of position and frequency gradients, classifies each primitive as spatially or frequency dominant. For frequency-dominant primitives, we postpone densification and delegate optimization to the inner-loop optimizer unless the frequency upper bound is reached. Meanwhile, for primitives whose densification is postponed, we design a stall detection

	mechanism based on Exponential Moving Average (EMA) to prevent frequency optimization from stagnating without progress. We further introduce frequency-aware pruning that identifies frequency-redundant primitives for more aggressive removal. Experiments demonstrate that our method achieves comparable or superior quality while significantly reducing primitive counts—up to 21.0% on texture-rich scenes and 8.4% on geometry-dominant scenes. The frequency-aware pruning further enables 27.6%–34.7% primitive reduction by safely raising the opacity threshold at negligible quality cost.
IP3124 **15:05-15:20**	Title: SRG-DIC: Strain-risk-guided neural displacement refinement for digital image correlation in continuous high-gradient fields Author(s): Shuke Gao, Kexin Li, Pingfang Zhou Presenter: Shuke Gao, Shanghai Jiao Tong University, China Abstract: Digital image correlation (DIC) recovers displacement from speckle-image pairs, but spatial differentiation can amplify small, structured displacement errors into large local strain errors. We investigate whether a predicted strain-risk map can guide local displacement refinement in continuous high-gradient deformation. Strain-risk-guided digital image correlation (SRG-DIC) combines a high-resolution correlation backbone, a strain-risk module, and three shared-weight bounded update steps. The risk module fuses displacement-derivative cues with photometric, feature-matching, and correlation-entropy evidence to localize regions vulnerable to differentiation. The backbone and risk module are fixed during refiner training, and Green-Lagrange strain is computed only after refinement using a fixed polynomial least-squares differentiator. Evaluation uses 300 held-out mild, high, and extreme triplets from five analytical deformation families, with triplets treated as the statistical unit. For extreme cases, over the prespecified 10% of pixels with highest analytical risk, SRG-DIC reduced strain mean absolute error (MAE) by 16.9% relative to the same frozen initial estimator (ratio 0.8311; 95% confidence interval, 0.8237-0.8385), while lowering full-field displacement error and preserving low-risk accuracy. A capacity-matched fixed-weight control supports the contribution of adaptive loss balancing. In a separate frozen-weight mechanism set, shuffling risk-map locations substantially weakened the improvement, whereas removing spatial gate variation had a smaller effect. Public benchmark evaluation further tests transfer beyond the synthetic deformation families. These results support spatially risk-guided displacement refinement for the tested backbone and continuous high-gradient fields.
IP3143 **15:20-15:35**	Title: Shared multi-depth residual reconstruction with physical-depth conditioning for spatial downscaling of ocean temperature and salinity fields Author(s): Hongtao Lv, Yingai Tian, Weipeng Zheng Presenter: Hongtao Lv, Beijing Information Science and Technology University, China Abstract: This study investigates same-day, same-depth horizontal spatial downscaling/reconstruction of multi-depth temperature and salinity fields derived from the LICOM Forecast System (LFS), using 1/4° regular latitude–longitude coarse grids as inputs and the corresponding 1/10° fine-grid fields as target/reference fields. Across 39 selected physical depth levels seen during training, we construct a residual reconstruction framework that shares parameters across depth. Coordinate-aware Bicubic interpolation provides the deterministic target-grid base, same-depth temperature and salinity are used as auxiliary inputs, and the two variables are reconstructed by separate single-target models. We compare an unconditioned auxiliary-input model (AUX), a parameter-matched constant-conditioning control (CONST), and a physical-depth-conditioned model (DEPTH); DEPTH-versus-CONST assesses the performance difference associated with replacing a constant condition by sample-specific physical depth under matched conditioning structure and parameter count. In five-seed matched-control experiments in the Agulhas region, DEPTH relative to CONST reduced root-mean-square error (RMSE) by 0.5058% ± 0.3400% for salinity, with positive gains in 5/5 seeds, and by 0.3704% ± 0.7988% for temperature, with positive gains in 4/5 seeds and greater sensitivity to random-seed/training stochasticity. Because CONST itself also changed performance, the full AUX-to-DEPTH gain cannot be attributed entirely to the physical-depth value. After region-specific retraining in a second region, the positive DEPTH-versus-AUX result reappeared, providing second-region replication. The conclusions are limited to 2023, the two study regions, and the 39 selected depth levels seen during training.

ONSITE SESSION B

- **Topic: Computer Vision and Machine Intelligence**
- **Location: Building A, 2F Meeting Room 5 | A 座 2 楼会议室 5**
- **Time: 13:30-15:35, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Prof. Fang Liu, Nanjing University of Science and Technology, China**

Invited Speaker 13:30-13:50	**Jie Guo, Associate Professor, Nanjing University, China** **Speech Title: Neural Spatial-temporal Super-sampling for Real-time Rendering**
IP3066 13:50-14:05	Title: Illuminating the dark: Event-flow guided deformable alignment for low- light video super-resolution Author(s): Guo Shurui, Zhang Dayong, Hu Liusen, Ye Xiang Presenter: Shurui Guo, Institute of Fluid Physics, China Academy of Engineering Physics, China Abstract: This paper addresses event-guided low-light video super-resolution (EVSR), aiming to restore high-resolution (HR) videos from degraded low-resolution (LR) low-light RGB frames by leveraging auxiliary event stream data. Existing EVSR methods typically rely on discrete frame-based optical flow for cross-modal feature alignment. However, due to severe noise and limited contrast in low-light conditions, these methods struggle to estimate precise motion, resulting in inaccurate feature warping and severe misalignment artifacts, even when utilizing illumination-enhanced models. To overcome these challenges, we propose a novel Event-based Flow-guided Deformable Alignment Video Super-resolution module (EFD-VSR). Unlike previous approaches that directly warp features using rough RGB-based optical flows, our framework introduces a continuous-time motion trajectory model to estimate event optical flow. By utilizing this highly accurate optical flow as a motion prior, we employ bi-directional deformable convolutions to dynamically adjust receptive fields, ensuring robust and structurally consistent cross-modal feature alignment. Extensive experiments on benchmark datasets, including DSEC, demonstrate that the proposed event flow-guided deformable alignment significantly mitigates motion blur and outperforms state-of-the-art methods in both visual quality and quantitative metrics. Notably, our method achieves a 0.34dB gain in PSNR over the SOTA method on the DSEC dataset, demonstrating superior detail recovery under low-light and motion blur conditions.
IP3093 14:05-14:20	Title: PG-DINO:Physics-Guided DINO for Adaptive Underwater Object Detection Author(s): Runping Xi, Feifei Li, Kun Ban, Xuefeng Kou Presenter: Feifei Li, School of Computer Science and Technology, Northwestern Polytechnical University, China Abstract: Underwater object detection remains difficult because of spatially nonuniform attenuation,scattering,scale variation, and category -dependent degradation .Existing methods typically improve image visibility or feature extraction in isolation,leaving physical cues disconnected from downstream decisions .We propose Physics-Guided DINO (PG DINO),an extension of DINO (DETR with Improved deNoising anchOr boxes)that uses a shared transmission proxy to adapt query features and modulate classification scores .A Physics-Guided Transmission Module (PGTM)derives green blue dark-channel estimates at three neighborhood scales and learns their fusion and calibration .Multi -Scale Physics Guided Attention (MSPGA)uses query -local transmission statistics to weight three projected feature branches,enhance small-object responses,and strengthen physical guidance progressively across decoder layers .An Adaptive Physics Guided Gating Network (APGG)routes each query to a category-specific gate or a shared fallback and modulates foreground scores using semantic,transmission,and size cues .PG -DINO achieves 71.8%mAP on DUO and 65.6%on RUOD,exceeding the reproduced DINO baseline by 2.8 and 2 .6 points,respectively;small-object AP reaches 56.3%on DUO and 21.8%on RUOD,improving by 7.1 and 4.9 percentage points,respectively .Ablation and design studies support the proposed fusion,small-object enhancement,decoder schedule,and category-aware gating .These gains require 2.1%more parameters and 2.4%more model FLOPs,while detector-only throughput decreases from 18.95 to 17.98 FPS .The results show that a physically motivated degradation cue can coordinate

	feature learning and prediction adjustment without image restoration or additional physical supervision .
IP3111 **14:20-14:35**	Title: Evidence-driven cross-modal refinement for RGB-infrared object detection Author(s): Xianglong Fan, Liuyang Sun, Yihong Zhang Presenter: Xianglong Fan, Donghua University, China Abstract: Visible and infrared images provide complementary imaging characteristics, but fixed fusion strategies cannot fully exploit response differences between modalities, while fixed upsampling limits the cross-scale propagation of fused features. To address these issues, we propose Evidence-Driven Cross-modal Refinement (EDCR), an RGB-infrared object detection framework built on dual-stream YOLO26. EDCR integrates a Consensus-Residual Evidence Optimization module (CREO), dynamic upsampling with DySample, and an Evidence-Guided Detection Head (EGDetect). At each feature scale, CREO constructs a consensus-candidate response from the sum of the two modality features and a discrepancy response from their absolute difference. A spatial gate learned under detection supervision controls the residual contributions of these two candidates to the fused representation. The resulting cross-modal evidence is also forwarded to EGDetect for subsequent prediction refinement. Evidence is constructed at the P3, P4, and P5 feature scales so that multiscale representations can be refined consistently. DySample adaptively adjusts sampling locations according to local content, strengthening the cross-scale propagation of the fused features in the neck. EGDetect uses the cross-modal evidence to refine the objectness representation and improve detection discrimination. Experiments on the Multi-scenario Multi-Modality Benchmark (M3FD) show that EDCR achieves 89.6% precision, 83.0% recall, 88.9% mean average precision at an intersection over union (IoU) threshold of 0.5 (mAP@0.5), and 60.3% mAP averaged over IoU thresholds from 0.5 to 0.95 (mAP@0.5:0.95), improving the Dual-YOLO26n baseline by 2.7, 4.4, 3.8, and 2.8 percentage points, respectively. Ablation experiments demonstrate that cross-modal evidence modeling, dynamic upsampling, and evidence-guided prediction jointly improve detection performance.
IP3116 **14:35-14:50**	Title: Monocular Relative Pose Estimation of Non-Cooperative Space Targets Using Solar Panel Structural Features Author(s): Hanhan Liu, Ling Wang, Zuoyang Song Presenter: Hanhan Liu, Nanjing University of Aeronautics and Astronautics, China Abstract: Monocular relative pose estimation of non-cooperative space targets is important for space proximity operations such as rendezvous and docking, on-orbit servicing, and active debris removal. However, conventional image feature extraction and matching methods can be sensitive to complex imaging conditions, making reliable relative pose estimation challenging. To address this problem, a monocular relative pose estimation method based on solar panel structural features is proposed. First, the solar panel regions are detected and segmented, and the resulting contours are refined. Stable structural corners are then extracted through quadrilateral fitting according to the structural characteristics of the solar panels. Subsequently, structural feature correspondences are established between different viewpoints, and their geometric relationships are modeled using a homography. Finally, the relative pose between the two views is recovered by exploiting the approximately coplanar structure of the solar panel feature points. The proposed method is evaluated using simulated monocular spacecraft image sequences and semi-physical experiments. Experimental results show that the extracted structural corners achieve good image-plane localization accuracy, while the method can effectively establish structural feature correspondences under target pose variations and obtain relative rotation and translation estimates consistent with the actual motion of the target. The semi-physical experimental results further demonstrate the feasibility of the proposed method for relative rotation estimation under different viewing conditions. These results demonstrate the effectiveness of the proposed method for monocular relative pose estimation of non-cooperative space targets.
IP3117 **14:50-15:05**	Title: High-Dynamic Star Point Detection and Localization Based on Center Heatmap and Subpixel Offset Regression Author(s): Jiaqi Hu, Guangxi Li, Yueyang Ding,, Yu Wang Presenter: Jiaqi Hu, Changchun University, China

	Abstract: To address star-point trailing, missed detection of faint stars, and increased localization errors caused by rapid spacecraft rotation, this paper proposes a star point detection and localization method based on center heatmap and subpixel offset regression. The proposed method employs a multiscale encoder–decoder network for feature extraction, detects star points through the center heatmap, and recovers subpixel coordinates by incorporating two-dimensional offset regression. Comparisons with five methods under the specified experimental settings show that, over the angular-velocity range of 0–12°/s, the proposed method achieves an average Recall, F1-score, and localization RMSE of 0.9707, 0.9775, and 0.3883 pixels, respectively. It also achieves the highest average Recall and F1-score under different stellar magnitudes and background levels. Ablation experiments demonstrate that the center heatmap consistently improves detection performance across all angular velocities, with more pronounced gains under severe trailing conditions. In addition, the subpixel offset branch reduces the average localization RMSE by 55.7%. On an independent test set, the proposed method achieves an F1-score of 0.9649 ± 0.0016. These results validate the robustness of the proposed method for star point detection and localization under complex imaging conditions and demonstrate its ability to provide more complete and reliable observational information for subsequent star matching and star-pattern recognition.
IP3122 **15:05-15:20**	Title: MOSAIC: Motion-guided sparse association with competition-aware track management for online multi object tracking Author(s): Xinming Yang , Minqi Li Presenter: Xinming Yang, Xi'an Polytechnic University, China Abstract: Online multi-object tracking (MOT) in crowded scenes requires efficient data association and reliable identity preservation. However, redundant association candidates increase computational cost, while competition among nearby targets makes unmatched detections unreliable evidence for new track initialization. To address these challenges, we propose MOSAIC, a motion-guided sparse association framework with competition-aware track management. A sparse association graph uses predicted target states to restrict candidate connections and combines geometric overlap with historical motion consistency to evaluate plausible matches. Building on this graph, a local incremental solver updates assignments through augmenting paths, reducing redundant computation while preserving assignment optimality within the retained graph at each association stage. The resulting competition evidence further guides track activation by identifying unmatched detections that remain plausible alternatives for existing identities, thereby suppressing premature initialization. These components unify candidate selection, efficient assignment, and identity confirmation within a shared association framework. Experiments on DanceTrack and MOT17 demonstrate competitive tracking performance. MOSAIC achieves 62.9 HOTA, 64.3 IDF1 and 48.6 AssA on DanceTrack, and 64.8 HOTA, 79.8 IDF1 and 64.6 AssA on MOT17. Ablation studies further demonstrate that, compared to the baseline of the DanceTrack validation set, sparse graph construction and competition-aware management together reduced identity switching by 36.0%, showcasing the algorithm's ability to perceive occlusion in crowded scenes.
IP3123 **15:20-15:35**	Title: SCAD: Scene-Conditioned Anomaly Data Supplementation for Weakly Supervised Video Anomaly Detection Author(s): Xuancan Liu, Xiao Zhang, Zhican Zhong, Yuxin Hong Presenter: Xuancan Liu, South-Central Minzu University, China Abstract: Weakly supervised video anomaly detection (WSVAD) relies on video-level labels but remains constrained by the scarcity and uneven coverage of real abnormal training data. We present SCAD, a data-centric framework for supplementing WSVAD with scene-conditioned generated abnormalities. SCAD uses real surveillance frames as visual conditions, applies multimodal large language model (MLLM) reasoning to assess scene-event admissibility, and constructs an indexed synthetic pool through scene-conditioned generation and manual quality inspection. The retained pool spans all 13 surveillance scenes and eight event categories. Generated clips are selected to match a target scene or event deficiency and incorporated with source-aware classification-loss weighting (SCLW) without changing the detector architecture or inference procedure. On ShanghaiTech with RTFM, $\lambda_s = 0.5$ achieves the highest mean AUC

	among the evaluated mixed-data weighting settings while remaining below real-only training. Across three controlled deficiency settings, matched supplementation achieves higher mean AUC than same-size random and complete-pool supplementation; scene- and bicycle-matched supplementation also improve their missing baselines, while running/chasing provides a boundary case because its removal does not create an observed global-AUC deficit. With equal classification weight, generated-abnormal replacement performs substantially worse than real-only training, and feature-space analysis indicates incomplete synthetic-real alignment. Overall, the results support targeted, weighted use of generated abnormalities as auxiliary data under controlled coverage deficiencies in the evaluated setting.

ONSITE SESSION C

- **Topic: Image Processing Fundamentals and Cross-Modal Learning**
- **Location: Building A, 2F Meeting Room 2 | A 座 2 楼会议室 2**
- **Time: 16:00-18:05, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Prof. Jie Guo, Nanjing University, China**

Invited Speaker 16:00-16:20	**Dr. Philippe Durand, CNAM, France** **Speech Title: Morphological Granulometry (VARAN) and Topological Extraction (Sentinel-1/2, Le Luc)**
IP1008 16:20-16:35	Title: Explainable Deep Learning Framework for Automated Detection and Segmentation of Temporomandibular Joint Disorders Using YOLOv8 with SGD Alignment Analysis Author(s): Shaneth C. Ambat, Jeneffer A. Sabonsolin, Anthony D.Aquino, Gilbert M. Tumibay, Ace C. Lagman, Alejandro D. Magnaye Presenter: Anthony D. Aquino, FEU Institute of Technology, Philippines Abstract: The diagnosis of Temporomandibular Joint Disorders (TMDs) is difficult owing to the complex anatomical structure of the temporomandibular joint (TMJ) and the inherent subjectivity in clinical evaluation. The purpose of the current study is to devise a novel deep learning architecture that can aid in automatic detection and instance segmentation of TMJ structures using Cone-Beam Computed Tomography (CBCT) images. The current study makes use of the TMJ-CBCT-Segmentation-Data dataset and incorporates lessons learned from the Open-Full-Jaw dataset to address the critical problem of analyzing TMJs at the population level. In addition, we have considered an in-depth analysis of the SGD optimization alignment in the loss function of the neural network to highlight how the convergence affects segmentation. Several architectural and computational graphs have been used in the current paper to elucidate the complexity of the forward pass of the model. Moreover, to address the black box problem of deep learning, we incorporate the concept of Explainable AI (XAI) and make use of feature attribution maps to analyze intermediate activations. Experimental results show high precision (0.95) and recall (0.90), along with an excellent mAP@0.5 score.
IP1011 16:35-16:50	Title: Foreground-prior Weighted Uncertainty Estimation for Active NeRF Reconstruction Author(s): Wenjun Liu, Mengmeng Yao, You Li Presenter: Wenjun Liu, Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen, China Abstract: In budget-constrained 3D acquisition, observable views are limited by camera motion, sensing time, and computation. Active view selection for neural reconstruction often ranks candidates by geometric uncertainty or global quality, which can allocate observations to background regions with limited value for target-object tasks. A foreground-prior weighted uncertainty framework is presented for budget-aware active NeRF reconstruction. Starting from a NeRF trained with a small initial view set, geometric uncertainty is estimated from the sensitivity of density predictions to local spatial perturbations. A foreground prior map reweights this uncertainty, yielding a foreground-weighted score for candidate-view ranking. The pipeline performs one-time batch ranking and selects top-K views under different observation budgets, avoiding repeated retraining during selection. This controlled protocol isolates the effect of the scoring mechanism from sequential model updates and enables direct comparison under fixed candidate pools and view budgets, while keeping the evaluation lightweight and reproducible. On a synthetic ShapeNet benchmark with foreground masks, the strategy improves foreground-region PSNR, SSIM, and LPIPS under low and moderate view budgets. The results show a trade-off: foreground-weighted ranking improves target-object regions, while uncertainty-only selection can better optimize global metrics at higher budgets. These results support lightweight object-focused acquisition for robotic inspection, object digitization, and AR/VR content capture.
IP1021 16:50-17:05	Title: Robust Apple Leaf Disease Recognition under Imbalanced Data via Generative Augmentation

	Author(s): Huishan Nie, Xin Cai, Tiantian Du, Xinyuan Nan Presenter: Xin Cai, Xinjiang University, China Abstract: Accurate apple leaf disease recognition in field orchard environments remains challenging because disease images are often limited, class distributions are imbalanced, and lesion appearance varies greatly under complex backgrounds. To address these issues, this paper proposes a generative-augmentation-based recognition framework that combines disease-oriented image synthesis with attention-enhanced convolutional classification. In the data augmentation stage, a multi-scale style-disentangled attention CycleGAN, named MSDA-CycleGAN, is developed to generate disease-specific apple leaf images. By preserving leaf structural information while introducing target disease characteristics, the generated images are used to enrich minority disease categories and construct a more balanced M-Apple dataset. In the classification stage, an EnhancedConvNeXtV2 network is designed by integrating a global attention mechanism and DropBlock regularization into the ConvNeXtV2 backbone. This design strengthens lesion-sensitive feature representation and improves robustness when learning from mixed real and generated samples. Experiments on four apple leaf categories, including healthy leaf, black rot, rust, and scab, show that the proposed framework achieves 98.44% accuracy, 98.35% F1-score, and 98.42% recall, outperforming several representative convolutional neural network baselines. To further evaluate the applicability of the proposed classifier, an additional experiment was conducted on a self-curated Kaggle dataset containing 2,940 disease images from six crop species. EnhancedConvNeXtV2 achieved 85.77% accuracy, 84.92% F1-score, and 84.86% recall on this dataset. Together with the ablation results, these findings demonstrate the effectiveness of the proposed framework and the broader applicability of the classifier to multi-crop leaf disease recognition.
IP2043 17:05-17:20	Title: SAM-Assisted Weakly Supervised Remote Sensing Change Detection with Unchanged Region Feature Alignment Author(s): Zeyu Xia, Fang Liu Presenter: Zeyu Xia, Nanjing University of Science and Technology, China Abstract: Remote sensing image change detection aims to identify target changes from bi-temporal remote sensing imagery. Compared with fully supervised methods, which incur high annotation costs, weakly supervised methods are more practical. However, image-level weakly supervised change detection suffers from issues such as background false positive noise and incomplete prediction boundaries. To address these issues, this paper proposes the SAM-assisted weakly supervised remote sensing change detection with unchanged region feature alignment. The method mainly consists of three modules: a hierarchical Transformer encoder for feature extraction and classification prediction, which also provides CAMs for constructing pseudo-labels; a Multi-Scale Scene Adaptive (MSSA) predictor module that integrates multi-scale contextual information, which enhances the model's ability to recognize and segment landscapes at different scales; and an Unchanged Region Feature Alignment (URFA) module dedicated to explicitly suppressing feature differences in unchanged regions, which leverages the complementary strengths of the backbone and SAM feature maps to precisely identify change regions and refine their boundaries. Experiments on the WHU-CD dataset demonstrate that the proposed method significantly outperforms the baseline method TransWCD-SA, achieving improvements of 3.92% and 4.93% in F1 score and IoU, respectively.
IP3078 17:20-17:35	Title: TSF-DETR: Feature representation refinement for complex wildlife detection Author(s): Dan Wang, Junping Qin, Xinran Shao, Zezhou Shao, Shujia Kong, Yangyang Luo Presenter: Dan Wang, Inner Mongolia University of Technology, China Abstract: Wildlife object detection plays an important role in ecological monitoring and biodiversity conservation. However, wildlife detection in natural environments remains challenging due to complex backgrounds, large variations in object scales, occlusions, and unstable imaging conditions. These factors often lead to weak target responses and insufficient feature representations, especially for small and visually ambiguous wildlife targets. Although Detection Transformer (DETR) enables end-to-end object detection through Transformer-based set prediction, its performance still depends heavily on the quality of input features provided to the Transformer encoder. To address this limitation, we propose TSF-DETR, a feature

	representation refinement framework that improves input features before Transformer encoding without modifying the original DETR architecture. The proposed method incorporates Frequency-aware Dynamic Convolution (FDConv) to strengthen local structural feature extraction, Residual Efficient Channel Attention (Residual ECA) to enhance informative channel responses while suppressing redundant features, and Two-Scale Fusion to integrate complementary spatial and semantic information across different feature resolutions. These modules jointly construct more discriminative visual representations for robust detection in complex wildlife scenes. Experiments on the LoTE-Animal dataset demonstrate that TSF-DETR consistently outperforms the baseline DETR and several representative DETR variants. Furthermore, cross-domain evaluation on the WAID dataset without fine-tuning demonstrates the adaptability of the proposed method under different wildlife imaging conditions. The results indicate that enhancing feature representation before Transformer encoding is an effective strategy for improving the accuracy and generalization capability of DETR-based wildlife detectors.
IP3098 **17:35-17:50**	Title: Geometry-Guided Residual Homography Refinement for Thermal–Visible Image Registration Author(s): Jie Cui, Kaiqi Lang, Chao Liang, Honghai Shen Presenter: Jie Cui, Changchun Institute of Optics, Fine Mechanics and Physics, Chinese Academy of Sciences, China Abstract: Significant non-linear radiometric differences exist between thermal infrared and visible images. In addition, low-texture regions, repetitive structures, and local false correspondences degrade the stability of global homography estimation. To address these issues, this paper proposes a thermal infrared and visible residual homography refinement network (GGRH). The proposed method adopts a fixed configuration comprising the XoFTR matcher and the MAGSAC++ robust estimator to acquire sparse correspondences and compute an initial homography matrix, H_0. Subsequently, match confidence, robust support, two-dimensional reprojection residuals, residual magnitude, and local evidence density are mapped into a six-channel geometric prior. This prior is then utilized to modulate the pre-registered cross-modal appearance features via a bounded gating mechanism. Experimental results on the LLVIP and Thermal-Visible Aerial 2020 datasets demonstrate that GGRH achieves improvements across multiple evaluation metrics, indicating that explicitly encoded geometric prior evidence can accurately guide and correct the residual errors of the initial estimation.
IP3133 **17:50-18:05**	Title: PP-SG: A pepper fruit–peduncle segmentation dataset and context-aware frequency fusion method Author(s): Xiaosheng Lin, Xiaojuan Wang Presenter: Xiaosheng Lin, Northwest Minzu University, China Abstract: With the increasing adoption of deep learning in agricultural robotics, finely annotated crop vision datasets are increasingly important for precision fruit harvesting. Peduncles are critical manipulation regions, yet their slender structures, narrow boundaries, and visual similarity to stems, leaves, and complex backgrounds make accurate segmentation difficult. Existing methods that rely mainly on local appearance often underuse the spatial relationship between fruits and peduncles, leading to missed detections and false positives under foliage interference. To address the shortage of high-quality training data, we construct PP-SG, a publicly available pepper fruit--peduncle dataset based on 671 real greenhouse images with fine-grained polygon-level annotations. After training-set augmentation, the experimental dataset contains 5,277 fruit and peduncle instances. Benchmark experiments with representative detection and segmentation models show that PP-SG supports effective learning of fruit and peduncle features, while peduncle segmentation remains substantially more challenging. We further propose Context-Aware Frequency Fusion (CAFF) and integrate it into the top-down feature fusion stage of YOLO11s. CAFF combines dual-branch frequency residual modeling with global semantic context and local spatial relations to generate adaptive spatial gates, thereby strengthening fruit--peduncle representations and suppressing background interference. Experimental results show that CAFF-YOLO11s achieves 88.72\% fruit IoU, 68.13\% peduncle IoU, and 78.43\% foreground mIoU on PP-SG, with only limited additional computational cost. The proposed dataset and fusion method provide a practical basis for robust fruit--peduncle perception in

	complex greenhouse environments.The dataset is publicly available at: https://huggingface.co/datasets/LXS12310/The_dataset_of_segmented_fruit_stems.

ONSITE SESSION D

- **Topic: Image Security, Understanding and AI Applications**
- **Location: Building A, 2F Meeting Room 5 | A 座 2 楼会议室 5**
- **Time: 16:00-18:05, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Prof. Linjing Wei, GanSu Agricultural University, China**

Invited Speaker 16:00-16:20	**Shuaifeng Zhi, Associate Professor, National University of Defense Technology (NUDT), China** **Speech Title: Dense Registration of Optical-SAR Remote Sensing Images under Large Geometric Transformations**
IP3086 16:20-16:35	Title: Artificial intelligence for tea disease and pest monitoring: datasets, sensing and field validation Author(s): Henghui Mo, Linjing Wei, Meiqi Zhong Presenter: Linjing Wei, Gansu Agricultural University, China Abstract: Tea plantations require timely monitoring of disease and pest damage, yet the tea-health artificial intelligence (AI) literature is difficult to compare because studies use different biological labels, image scales and validation units. This structured critical review links target definition, dataset provenance, sensing platform, model task and field evidence. A documented revision audit combines an archived bibliography, bounded Crossref retrieval and targeted publisher discovery, with a publication cutoff of 23 September 2026 and final checking on 10 October 2026. It yields 23 tea-specific primary reports for detailed comparison. Early work used colour, texture, saliency and convolutional classification on controlled leaf images. Recent studies add natural-scene detection, segmentation, UAV imagery, tiny-pest recognition and lightweight inference. Study-level comparisons distinguish internal image tests from held-out acquisition domains and identify measured edge-device operation without equating it with management benefit. Dataset units and task metrics differ, precluding a pooled architecture ranking. A field-ready system should therefore be judged through an evidence chain: biologically defensible labels, metadata and fixed splits; task-appropriate metrics; external and condition-aware tests; device measurements; uncertainty handling; and a clear link from detections to scouting or integrated pest management (IPM). The review identifies practical reporting requirements and a staged pathway from visual triage to spatial prioritization and validated management support.
IP3068 16:35-16:50	Title: Real-time Detection and Recognition of Inkjet Codes on Transparent Bottles Based on YOLO and Adaptive Preprocessing Author(s): Yuchen Ye, Shutian Jiang, Ziyin Wang, Keyi Zhu, Qing Chang Presenter: Yuchen Ye, ECUST, China Abstract: Transparent curved bottles suffer from coupled specular reflection and cylindrical distortion that severely degrades inkjet code recognition. This paper presents a lightweight edge-deployable pipeline with a coarse-to-fine cascaded strategy. In the coarse stage, YOLOv11n localizes the inkjet-code region and simultaneously acts as a spatial-domain anti-reflection filter by discarding most non-target highlight areas. In the fine stage, a dedicated preprocessing cascade—maximum-value grayscale transform, Otsu thresholding, and Thin-Plate Spline (TPS) interpolation—is applied exclusively to the cropped ROI to suppress residual highlights and rectify geometric warping. PaddleOCR then decodes the rectified ROI. Evaluated on a mixed dataset of 1,500 images (1,130 public images and 370 captured on a production line), the system achieved 95.30% mAP@0.5 and 93.98% recall, boosting OCR character accuracy from 18.39% to 90.23%. The end-to-end latency was 52.4 ms, which is expected to drop to ≤35 ms via TensorRT optimization. This work demonstrates that object detection can serve as the first-line defense against reflection, followed by targeted geometric-optical restoration.
IP3080 16:50-17:05	Title: MedEHA: Entity-aware hierarchical alignment for medical vision-language representation learning Author(s): Xiangmin Kong, Xibin Jia, Dawei Yang, Luo Wang, Kelei Jin, Zhenghan Yang

	Presenter: Xiangmin Kong, Beijing University of Technology, China Abstract: Medical vision-language pre-training (MedVLP) has shown strong potential for learning transferable representations from paired medical images and radiology reports. However, many existing methods encode reports as flat text sequences, making it difficult to explicitly model the compositional relations among pathological findings, anatomical locations, and observation status, and to associate these semantics with local visual evidence. We propose MedEHA, a structured medical vision-language representation learning framework with hierarchical entity modeling and coarse-to-fine cross-modal alignment. MedEHA represents each report as a set of medical entity triples consisting of a medical entity, anatomical location, and existence status. A two-stage hierarchical entity encoder first models the composition within each triple and then captures dependencies among multiple entities in the same report. Based on these structured representations, MedEHA combines global image-report contrastive alignment, prototype-based semantic alignment, and fine-grained entity-aware alignment through language-guided patch grouping. We pre-train MedEHA on 217k paired frontal chest X-ray images and reports from MIMIC-CXR and evaluate the learned representations on classification, segmentation, and object-detection benchmarks including CheXpert, RSNA, COVIDx, and SIIM. MedEHA achieves competitive or superior performance across most evaluation settings, with particularly strong results on localization-sensitive tasks. These results suggest that explicitly modeling structured clinical observations can provide more informative textual supervision and improve fine-grained medical vision-language representation learning.
IP3090 **17:05-17:20**	Title: Cross-view primitive-aware 3D parametric CAD generation via Slot Attention Author(s): Guoji Li, Benbo Zha, Guoqiong Liao, Bokang Li, Peiwen Wang Presenter: GUOji LI, Jiangxi University of Finance and Economics, China Abstract: Generating editable 3D parametric CAD models from 2D engineering drawings remains challenging. Existing vision-language approaches typically feed dense visual tokens into large language models (LLMs), but do not explicitly exploit orthographic projection relations and may introduce substantial background redundancy. We propose a cross-view geometry-constrained visual-token compression framework for parametric CAD generation from front, top, and left views. A frozen Vision Transformer first extracts 768 patch tokens from the three views. Orthographic projection-guided attention then restricts cross-view interactions to projection-consistent candidate regions, after which Slot Attention aggregates the aligned features into 96 slot tokens. Removing two background slots per view leaves 90 visual tokens for the LLM. A lightweight projection module maps these tokens into the LLM embedding space for autoregressive CadQuery generation. On the 2,000-sample test split, the full model attains 68.5% executability, compared with 67.4% for a matched 768-token dense baseline. On the 1,094 samples successfully reconstructed by all variants, its mean normalized Chamfer Distance is 28.75%, versus 23.54% for the dense baseline. The method therefore offers an explicit efficiency–fidelity trade-off: it reduces the LLM-side visual sequence by 8.5× and measured prefill latency by 40.5%.
IP3095 **17:20-17:35**	Title: Direct Log-Radial Models for Optic Cup–Disc Geometry with an Expert-Derived Disc Anchor Author(s): Andrey Krylov, Alexander Khvostikov, Shao Wei Presenter: Andrey Krylov, Lomonosov Moscow State University, Russia Abstract: This study is a single-dataset internal representation benchmark rather than an end-to-end glaucoma-diagnosis experiment. Conditional on an expert-derived optic-disc center and radius, we ask which low-capacity representation best preserves relative cup–disc geometry in fundus photographs. We construct gradient-derived radial edge-response profiles in the dimensionless coordinate u = log(r/R_d) and predict the mean expert area-derived log equivalent-radius target CDR_{log,eq}. This target is not conventional vertical CDR and is used solely as a geometric benchmark. On PAPILA, all 488 eyes from 244 patients are split by patient into fitting, calibration, and a single 49-patient confirmation subset. Under a fixed ridge-regression protocol, compact sector statistics attain the lowest confirmation RMSE point estimate (0.3104) and MAE (0.2401), while an explicit one- versus two-component radial mixture reaches RMSE 0.3241 and the highest correlation (0.664). Secondary grid and anchor-stress analyses do not reverse this ordering. In a separate 64-coordinate spectral ablation,

	matched nonzero Mellin features reach RMSE 0.3162, compared with 0.3255 for Mellin–Hermite, 0.3533 for complex log-Fourier, and 0.3830 for magnitude log-Fourier. Thus, for this PAPILA benchmark and fixed low-capacity learner, direct geometry-matched statistics are highly efficient, while Mellin geometry remains useful within spectral compression. Clinical deployment and automatic disc localization are not evaluated.
IP3099 **17:35-17:50**	Title: Lightweight multi-stream meeting behavior recognition based on RK3588 with auxiliary-evidence reranking Author(s): Xiaohai Yu, Zipeng Chen, Yating Fang, Hua Zheng, Xiaoqin Zhu Presenter: Zipeng Chen, Fujian Normal University, China Abstract: Meeting-room edge systems in government and enterprise environments are often deployed on isolated networks, requiring video to be analyzed locally under limited computing resources. Because these applications are highly sensitive to false alarms, this paper presents a precision-oriented four-stream meeting behavior recognition system based on RK3588. Using YOLO11s, the detector employs a hierarchical 12-class label space mapped to eight internal behavior states. Local classes such as use_phone, call_up, and use_pen provide auxiliary evidence for visually similar person-level behaviors. The confidence of a spatially compatible auxiliary box is fused with that of the primary box to rerank candidates before class-agnostic non-maximum suppression, while a class-specific high-confidence bypass preserves reliable primary detections when auxiliary objects are missed. The deployment pipeline further integrates INT8 distribution focal loss (DFL) decoding, Arm NEON optimization, three parallel Rockchip Neural Network (RKNN) contexts, Kalman filtering with intersection-over-union (IoU) association, and track-level temporal confirmation. Compared with the original class-agnostic NMS, the proposed postprocessing increases macro-averaged anomaly precision by 0.96 percentage points while retaining 78.71% frame-level recall. Across 28.32 min of evaluated recordings, manually reviewed alarms achieve 94.50% event precision and 38.14 false alarms per camera-hour. With 1280 × 1280 input and three neural processing unit (NPU) cores, the four-stream system reaches an aggregate throughput of 12.951 ± 0.020 FPS. These results demonstrate the feasibility of local four-stream meeting behavior analysis on RK3588 under the evaluated conditions.
IP3140 **17:50-18:05**	Title: Synchronization-Based Image Encryption Using a Fractional-Order Reaction-Diffusion Chua's System with Spatiotemporal Chaotic Dynamics Author(s): Mingquan Du, Lulu Fu, Fangyue Chen, Tianyu Ge, Xinhui Yang, Fengyi Liu Presenter: Mingquan Du, Wuxi University, China Abstract: Chaos-based image encryption provides an effective approach to protecting digital image information by exploiting the sensitivity and complex dynamics of chaotic systems. In this paper, we construct a two-dimensional spatial fractional reaction-diffusion Chua's system (2D-FRDCS) by incorporating the Caputo time-fractional derivative, Chua's circuit, and two-dimensional spatial diffusion. Numerical simulations using the Fourier pseudo-spectral method and ETD4RK scheme demonstrate rich spatiotemporal chaotic dynamics. A master-slave synchronization scheme with linear feedback control is then developed, and global Mittag-Leffler synchronization is established using the Lyapunov method. Based on the synchronized chaotic states, we propose an image encryption algorithm integrating dynamic field sampling, pixel scrambling, and two-round bidirectional diffusion. The three synchronized state fields are used to generate permutation indices, diffusion keys, and diffusion control modes, respectively, enabling dynamic key updating across image blocks. Security analyses demonstrate a sufficiently large key space and high key sensitivity, while the ciphertext exhibits a nearly uniform histogram, low adjacent-pixel correlation, and information entropy close to 8 bits. The proposed method provides an effective framework for synchronization-based image encryption using spatiotemporal chaotic dynamics.

POSTER SESSION 1

- **Topic: Segmentation, Detection and Tracking for Perception**
- **Location: Building A, 2F Meeting Room 6 | A 座 2 楼会议室 6**
- **Time: 13:30-15:30, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Assoc. Prof. Shuaifeng Zhi, National University of Defense Technology, China**

1	**IP1007**	Title: Prior-Guided Structure-Detail Wavelet Fusion for Multimodal Remote Sensing Object Detection Author(s): Chunqiang Sun, Peng Tu, Jianjun Liu Presenter: Jianjun Liu, Jiangnan University, Wuxi, Jiangsu, China
2	**IP1003**	Title: Subway Platform Door Gap Shadow and Foreign Object Segmentation Method Based on Improved DeepLabV3+ Author(s): Zixuan Chen, Wen Xue, Qiju Kang, Zihao Xie, Feixue Wu, Taiyang Hu, Zelong Xiao Presenter: Zixuan Chen, Nanjing University of Science and Technology, China
3	**IP1020**	Title: SFLRMamba: Spatial-Frequency Low-Rank State Space Modeling for Remote Sensing Scene Classification Author(s): Yingjie Zhang, Bin Zhou Presenter: Bin Zhou, No.722 Research Institute of China State Shipbuilding Corporation Limited, China
4	**IP1023**	Title: Abdominal Multi-organ Image Segmentation Approach Based On PCE-TRANSUNET Author(s): Yanli ZHAO, Jianye ZHANG, Cai SONG Presenter: Jianye ZHANG, Ningxia Institute of Technology, China
5	**IP1024**	Title: Multi-View Least Squares Representation with Low-Rank and Graph Constraints for Contactless Palmprint Recognition Author(s): Aijun Hao, Mengwen Li, Xiao Li, Chengzhen Xu, Mei Shi Presenter: Aijun Hao, Huaibei Normal University, China
6	**IP1028**	Title: DINOHand-Track: Reconstruction-Guided Hand Reconstruction and Tracking in Continuous Videos Author(s): Qilin Wu, Zhenbo Song, Jianfeng Lu Presenter: Qilin Wu, Nanjing University of Science and Technology, China
7	**IP2032**	Title: An Empirical Study of Lightweight KAN and Boundary-Aware Refinements for Segmenter-Style Semantic Segmentation Author(s): Yunfei Peng, Yancong Deng Presenter: Peng Yunfei, Asia Pacific University of Technology & Innovation (APU), Malaysia
8	**IP2037**	Title: FS-OWOD: Fast-Slow Open-World Object Detection with Hierarchical Unknown Prompting and Online Semantic Registration Author(s): Guanxiang Wang, Lipu Di, Yiming Nie, Weizhong Jiang, Yanxuan Hong, Zhaohui Wang, Wenzhe Li Presenter: Guanxiang Wang, Academy of Military Sciences, China
9	**IP2040**	Title: Symmetry-Aware 3D Gaussian Splatting for Industrial Objects Author(s): Yisong Wang, Weichen Dai, Hua Zhang Presenter: Yisong Wang, Hangzhou Dianzi University, China

10	**IP2044**	Title: SI-SDT： A State-Isolated Supervisory Detector-Tracker for Autonomous Anti-UAV Tracking Author(s): Tinghe Xie, Yan Jin,, Xiaoxi Wu Presenter: Tinghe Xie, AVIC Manufacturing Technology Institute, Beijing, China
11	**IP2054**	Title: Graph Attention Network for Radar/Infrared Target Association Under Incomplete Observations Author(s): Meng Nan, Zhiyong Su, Weiqing Li Presenter: Meng Nan, Nanjing University of Science and Technology, China
12	**IP2058**	Title: PVT-ASCAD: Adaptive Scale Convolutional Attention Decoding Pyramid Vision Transformer for Efficient Medical Image Segmentation Author(s): Jiyue Dai, Yi Cai Presenter: Jiyue Dai, The Hong Kong University of Science and Technology, China
13	**IP3083**	Title: LHNet: Lightweight Feature Descriptors with Multi-Scale Context and Homography Regularization Author(s): Junke Wu, Guoliang Wei Presenter: Junke Wu, University of Shanghai for Science and Technology, China
14	**IP3141**	Title: A Retinal Vessel Segmentation Method Based on Progressive Multi-Scale Fusion and Composite Loss Author(s): Yan Yang, Jiani Zhu, Qihui Yan, Jianyang Feng Presenter: Jiani Zhu, Liaoning Normal University, China
15	**IP5001**	Title: PODad: Prediction-Observation Decoupled Autonomous Driving Author(s): Shaowei Zhang, Zhenbo Song, Jianfeng Lu Presenter: Shaowei Zhang, Nanjing University of Science and Technology, China

POSTER SESSION 2

- **Topic: Multimodal Imaging, Fusion and Reconstruction**
- **Location: Building A, 2F Meeting Room 6 | A 座 2 楼会议室 6**
- **Time: 16:00-18:00, UTC/GMT+8, Nov. 7, 2026**
- **Chair: Prof. Jianjun Liu, Jiangnan University, China**

1	**IP2053**	Title: An Incomplete Multi-View Multi-Label Learning with Universum Author(s): Changming Zhu, Shihao Duan, Xiaopeng Miao Presenter: Changming Zhu, Shanghai Maritime University, China
2	**IP1025**	Title: Feature Memory Bank and Incremental Training for Cloud-Edge Collaborative Monitoring Author(s): Guohai Chen, Yaobin Mao, Wenbo Xiang Presenter: Guohai Chen, Nanjing University of Science and Technology, China
3	**IP2041**	Title: A controllable spectral-spatial attention mechanism for unsupervised hyperspectral band selection Author(s): Yin Yu, Mingwei Wang, Yibing Liu, Jingyan Lu, Jiaxuan Gu Presenter: Yin Yu, Wuhan Technology and Business University, China
4	**IP1026**	Title: Structure-Adaptive Discretized Mamba for Hyperspectral and Multispectral Image Fusion Author(s): Xiaoxuan Xiang, Jingxiang Yang, Liang Xiao Presenter: Xiaoxuan Xiang, Nanjing University of Science and Technology, China
5	**IP2047**	Title: CABACScope: A Context Profiling Framework for Coding Efficiency Analysis in the CABAC of Simplified ECM Video Coding Author(s): Siqi Li, Luheng Jia, Zuhai Zhang, Yutian Liu, Yongkai Huo Presenter: Siqi Li, Beijing University of Technology, Beijing, China
6	**IP2048**	Title: Stable Diffusion with Scene-Adaptive Spectral Guidance for Hyperspectral and Multispectral Image Fusion Author(s): Zhongling Xie, Jingxiang Yang Presenter: Zhongling Xie, Nanjing University of Science and Technology, China
7	**IP3060**	Title: Infrared Shot Non-uniformity Correction Based on Median Residual Extraction and Adaptive Temporal Recursion Author(s): WANG Yujie, LI Lianbing, PANG Huaisheng, ZENG Bangze, ZHOU Yongkang, ZHA Haigang, TAO Zhigang, Bai Yaxin Presenter: Wang Yujie, Kunming Institute of Physics, China
8	**IP2050**	Title: Physics-Guided Degradation Modeling of Multi-View Images for Underwater Depth Estimation Author(s): Yuang Chen, Kun Li, Xiaoyu Chen, Jing Han, Lianfa Bai Presenter: Yuang Chen, Nanjing University of Science and Technology, China
9	**IP5002**	Title: SCAD: Scene-level contrastive adversarial diffusion for multi-label image classification Author(s): Feifei Wang, Lei Sun, Zhenbo Song, Jianfeng Lu Presenter: Feifei Wang, Nanjing University of Science and Technology, China

10	**IP3110**	Title: Black-and-white film simulation using tone-grain decomposition of paired RAW and scanned film images Author(s): Piaoran Li, Yesu Men Presenter: Piaoran Li, Macau University of Science and Technology, China
11	**IP3118**	Title: Efficient non-line-of-sight reconstruction under light-position errors via differentiable 3D Gaussian shadow fields Author(s): Taiping Lu, Jingyuan Zhang, Yi Zhang, Xiaoyu Chen Presenter: Taiping Lu, Nanjing University of Science and Technology, China
12	**IP3112**	Title: Axis-Aware Transition and Detail Enhancement for UAV LiDAR Corridor Segmentation Author(s): Wei Huang, Chenyu Zhao, Jing Hu, Presenter: Chenyu Zhao, Wuhan Polytechnic University, China
13	**IP5004**	Title: NL2SSA: A Natural-Language-Driven Visual Analysis System for Orbital Space Objects Author(s): Huaquan Hu, Xiaohong Murong, Ronghuan Yu Presenter: Huaquan Hu, Space Engineering University, China
14	**IP3101**	Title: Nonstationary Signal Denoising and Forecasting via Improved Multi-Domain Feature Screening and Residual PINNs Author(s): Siqi Zhao, Yongzhi Ning, Yangcheng Ren Presenter: Yongzhi Ning, Dalian Neusoft University of Information, China
15	**IP2059**	Title: Screen Moiré Image Synthesis Method Based on Luminance Modulation and Ambient Light Modeling Author(s): MengKe Liu, Yaobin Mao Presenter: MengKe Liu, Nanjing University of Science and Technology, China
16	**IP3070**	Title: Blockout-driven Controllable Generation of Sci-fi Architectural Assets with Procedural Scene Assembly Author(s): Jie Shang and Huaqun Liu Presenter: Jie Shang, Beijing Institute of Graphic Communication, China

ONLINE SESSION 1

- **Topic: Small Object Detection and Remote Sensing Perception**
- **Room A: 818 3371 3809 Password:0608**
- **Time: 10:00-12:20, China UTC/GMT+8, Nov. 8, 2026**
- **Chair: Prof. Rongchun Li, National University of Defense Technology, China**

Invited Speaker 10:00-10:20	**Shou Feng, Professor, Harbin Engineering University, China** **Speech Title: Modality-Adaptive Fractional-Domain Decoupling and Interaction Network for Multi-Source Remote Sensing Image Classification**
IP3134 10:20-10:35	Title: DF-YOLO: An Improved YOLOv11-Based UAV Aerial Small Target Detection Author(s): Ying Zhang, Hui Liu Presenter: Hui Liu, Shanghai University of Electric Power, China Abstract: The To address the challenges of complex backgrounds, dense small objects, significant multi-scale variations, and interference from complex scenes in object detection within Unmanned Aerial Vehicle (UAV) aerial imagery, this paper proposes an improved small object detection algorithm for UAV images, named DF-YOLO, based on YOLOv11n. A Multi-scale Dynamic Fusion module is designed. By embedding a dual attention mechanism and multi-scale fusion strategy into the C3 architecture, it significantly enhances the feature representation capability for small and occluded objects while maintaining a lightweight structure. An attention-guided dynamic upsampling module, termed AGSample, is constructed. It couples Efficient Channel Attention (ECA) with a depthwise separable channel attention to form an adaptive kernel attention (AKA) design, so that the sampling offsets are predicted from an attention-refined feature at a small parameter cost.Furthermore, Omnidirectional Dynamic Convolution (ODConv) is introduced to replace the standard convolutional layers in the neck network, strengthening the multi-scale feature fusion capability.Experimental results on the VisDrone2019-DET validation set demonstrate that DF-YOLO improves mAP50 and mAP50-95 by 2.2 and 1.3 percentage points, respectively (from 0.300 to 0.322 and from 0.170 to 0.183), compared with the baseline YOLOv11.These improvements validate the effectiveness of the proposed method.
IP1022 10:35-10:50	Title: HQA-YOLO: A Multiscale Feature Refinement Network for UAV Infrared Small-Object Detection Author(s): Mengen Lou, Guoliang Gong Presenter: Mengen Lou, Tianjin University of Science and Technology, China Abstract: Unmanned aerial vehicle (UAV) thermal infrared object detection is important for low-altitude surveillance and emergency rescue. However, targets in high-altitude and long-range imagery are usually small, have weak texture details, and are affected by thermal background interference, which makes small-object detection difficult. This study takes YOLOv8n as the baseline and proposes HQA-YOLO, a feature refinement network for UAV infrared small-object detection. First, an S2SC heterogeneous backbone is constructed by combining SPDConv with ShuffleNetV2-style downsampling to preserve the spatial details of weak small targets while controlling parameter redundancy. Subsequently, an ASF3 adaptive spatial fusion neck is designed to align and adaptively fuse multiscale features, thereby alleviating the effects of scale mismatch and background interference on feature representation. Finally, R-MDSF-Head is constructed to further refine features before prediction, strengthening the responses of low-contrast small targets and improving localization performance in complex backgrounds. Experiments on the HIT-UAV dataset show that HQA-YOLO reduces the number of parameters by 13.3% while improving precision, recall, mAP@0.5, and mAP@0.5:0.95 by 6.47, 3.38, 4.69, and 2.97 percentage points, respectively. These results verify the effectiveness of the proposed method for UAV infrared small-object detection.
IP3102 10:50-11:05	Title: CholSAR: Clutter-Aware Cholesky Gaussian Learning for Oriented Ship Detection in SAR Images Author(s): Xiaojuan Wang, Baoqing Yang, Songlin Chen

	Presenter: Xiaojuan Wang, Harbin Institute of Technology, China Abstract: Oriented ship detection in synthetic aperture radar (SAR) imagery is challenging due to complex background clutter, dense target distributions, and arbitrary orientations. This paper proposes CholSAR, a clutter-aware Gaussian detector built upon GauCho. The network combines an LSKNet-S backbone with a feature pyramid network for multiscale feature extraction and retains the Cholesky-parameterized Gaussian representation of bounding boxes. To account for localization uncertainty in complex backgrounds, a clutter-aware full-geometry uncertainty modeling module couples an auxiliary anisotropic covariance with the predicted geometric covariance and provides auxiliary supervision over target centers, scales, and orientations through an expected Gaussian negative log-likelihood. Furthermore, a localization quality calibration module combines predicted rotated intersection-over-union with classification confidence and centerness to improve candidate ranking. CholSAR achieves mAP75 scores of 35.53% and 19.55% on SRSDD-v1.0 and FAIR-CSAR-v1.0, respectively, outperforming the ResNet-50-based GauCho baseline on both datasets and achieving the highest detection accuracy among the compared methods on SRSDD-v1.0.
IP3105 **11:05-11:20**	Title: Multi-scale structured adaptive query learning for UAV–satellite cross-view geo-localization Author(s): Ruilin Pang, Fulin Liu, Zhenzhong Wei Presenter: Ruilin Pang, Beihang University, China Abstract: Unmanned aerial vehicle (UAV)–satellite cross-view geo-localization (CVGL) estimates UAV locations by retrieving geo-referenced satellite images for onboard observations. Large viewpoint, scale, orientation, and visible-area changes make fixed spatial partitions unreliable. Query-Driven Feature Learning (QDFL) replaces predefined partitions with learnable queries, but its homogeneous query bank repeatedly accesses identical feature memories and may produce redundant responses. We propose Multi-Scale Structured Adaptive Query Learning (MSAQL), which divides learnable queries into independently parameterized groups and routes them through context, refined, and pooled feature memories. The differentiated routes introduce complementary scale and refinement biases while retaining content-adaptive cross-attention. A group-diversity loss discourages correlated responses, and an additional Multi-Similarity objective directly supervises the concatenated descriptor used for retrieval. MSAQL retains a frozen DINOv2 Vision Transformer with lightweight adapters for parameter-efficient cross-view adaptation. On University-1652, it achieves 95.31% Recall@1 (R@1) and 96.07% average precision (AP) for Drone-to-Satellite retrieval, and 97.15% R@1 and 95.01% AP for Satellite-to-Drone retrieval. It also remains competitive across four SUES-200 flight heights and improves average cross-dataset retrieval over QDFL. This design improves the use of complementary scene layout and local structure under large altitude-dependent scale variation. Ablations confirm that query grouping, multi-scale routing, diversity regularization, and final-descriptor supervision jointly improve robust cross-view representation learning.
IP3109 **11:20-11:35**	Title: AIoU: Accumulated IoU Loss for YOLO based Aerial Small Object Detection Author(s): Xinke Zhi, Yunzhi Guo, Junding Sun Presenter: Xinke Zhi, Henan Polytechnic University, China Abstract: Small object detection in aerial images poses significant challenges to the bounding box regression of detectors due to the extreme scarcity of pixel information. In this work, we propose a novel similarity metric named AIoU and its corresponding regression loss function for YOLO detectors. The core of AIoU is the introduction of the ``accumulation'' term, which comprehensively incorporates the IoU of all interpolated boxes between the prediction and the target. The accumulation, integrated with the original IoU via a weighting coefficient, provides a stable and informative similarity measure even for non-overlapping boxes. Consequently, AIoU loss effectively mitigates the gradient vanishing and regression misalignment problems prevalent in small object detection. Evaluated on three aerial image datasets (AI-TOD-v2, DIOR-H, and VisDrone-DET2019) with YOLO11n and YOLOv8n, our AIoU loss achieves the highest AP50-95 for both detectors, surpassing CIoU baselines by 0.2–1.1 percentage points. It also boosts the detection accuracy of very tiny and tiny objects on AI-TOD-v2 with negligible extra computational cost.

IP3130 **11:35-11:50**	Title: Dynamic Prompt Augmentation for Object Detection Author(s): Chenlong Feng, Yu Du, Yan Ma, Hejun Lv, Dongyu Hu, Wentao Liu, Ying Zheng Presenter: Chenlong Feng, Beijing Union University, China Abstract: Recent object detectors have increasingly adopted text prompts to provide category information. However, using a single class name to represent each category may provide limited semantic diversity and restrict the ability of the detector to exploit alternative textual descriptions of the same object concept. In this work, we propose Dynamic Prompt Augmentation (DPA) for enhancing object detection. DPA constructs a prompt-set containing multiple semantically related descriptions for each object category and dynamically selects one prompt at the beginning of each training epoch. The selected prompts are encoded by the text encoder and used in the original visual-textual matching process, while the visual feature extraction and detection architecture remain unchanged. We implement DPA on YOLOE and evaluate it on the DAWN dataset under challenging weather conditions. Experimental results show that DPA improves detection performance, yielding gains of 0.6 percentage points in AP$_{50:95}$ and 3.8 percentage points in AP$_{50}$. Ablation studies demonstrate the benefit of dynamic prompt diversity over a fixed alternative prompt. Moreover, DPA maintains the original model complexity and achieves nearly unchanged inference speed. These results demonstrate that textual prompt diversity provides a simple and effective way to improve object detection without modifying the visual architecture.
IP5003 **11:50-12:05**	Title: S2N-DETR: A Lightweight Spatial-Preserving and Scale-Adaptive Regression Method for Small Object Detection in UAV Imagery Author(s): Mingyang Chu, YuJia Xie, JiaJun Xu, Hang Zhong Presenter: Mingyang Chu, Nanjing University of Finance & Economics, China Abstract: Objects in UAV aerial images are often small and densely distributed. Tiny objects occupy only a few pixels and are easily affected by road, building, and vegetation textures, while the computing and storage constraints of UAV onboard platforms limit detector size. To address these two constraints, this paper develops S2N-DETR based on the RT-DETR framework. The model adopts a C2f feature extraction backbone and uses full SPDConv and LiteSPDConv at the P4 and P5 scale transitions, respectively. The former provides stronger joint modeling of neighboring features, whereas the latter controls parameter growth at the high-channel stage. The S3, S4, and S5 outputs are recalibrated by SimAM before entering the hybrid encoder. For bounding-box regression, an area-adaptive GIoU-NWD hybrid loss assigns a higher NWD weight to small boxes while retaining more GIoU supervision for large boxes. On the VisDrone2019 validation set, the final model achieves 45 .3% mAP@50 and 27.6% mAP@50:95, improving upon the Baseline by 1 .5 and 1 .2 percentage points, respectively. Local evaluation on the test-dev split yields 35 .8% mAP@50 and 20 .7% mAP@50:95 . After Conv–BN fusion, the model contains 10.21M parameters and requires 25 .14 GFLOPs. Compared with RT-DETR-L, it reduces the parameter count by 68 .1% and GFLOPs by 76 .1%, while its mAP@50 is only 0 .9 percentage points lower.
IP3136 **12:05-12:20**	Title: Controlled Multi-Scale Architecture and Degradation-Aware Training for Robust 2D Road-User Detection on KITTI Author: Xiuyuan Wang Presenter: Xiuyuan Wang, University of Birmingham, United Kingdom Abstract: Reliable camera-based road-user detection requires fine-scale localisation and robustness to image degradation within an acceptable inference budget. We report two separate YOLOv8s studies on internal KITTI partitions. Across three seeds, the tested P2–P5 neck and detection head increases three-class Moderate AP from 93.50% to 94.83% on the 749-image final comparison split, with approximately 22% higher model-forward latency. P2 and P2+Coordinate Attention achieve nearly identical AP (94.83% and 94.80%, respectively), but P2 is faster. Size-stratified diagnostics show a gain of 3.499 mAP50–95 points for small objects and a decrease of 1.546 points for large objects. Under a separate 960-pixel protocol, mean degraded mAP50–95 increases from 63.860% for the clean baseline to 73.795% with random-mixed fine-tuning and 73.796% with degradation-calibrated training (DCR). Clean-only, random-mixed and DCR fine-tuning each schedule 11,968 image presentations per epoch, but actual optimizer

	updates were not independently logged. Both degraded-data methods outperform clean-only fine-tuning on these synthetic tests; their near-identical scores do not establish an advantage from curriculum ordering. Each study uses non-overlapping image IDs, but neither partition was grouped by sequence or audited for near duplicates. Hence, the findings concern internal performance rather than independent-scene generalisation, official KITTI test-server results or real-road reliability. Latency measures float32 PyTorch model forwards, excluding preprocessing, postprocessing and full-system overhead. The two protocols differ in resolution, partition and metric; no combined P2+DCR detector was evaluated.

ONLINE SESSION 2

- **Topic: Reconstruction, Generation and Multimodal Representation**
- **Room B: 873 1395 5014 Password:0608**
- **Time: 10:00-12:20, UTC/GMT+8, Nov. 8, 2026**
- **Chair(s): Prof. Wenhua Qian, YunNan University, China & Assoc. Prof. Ye Yu, Hefei University of Technology, China**

Invited Speaker 10:00-10:20	**Lili Pan, Associate Professor** **University of Electronic Science and Technology of China, China** **Speech Title: Test-Time Compute for Diffusion Models: Choosing the Trajectory, Then Learning from It**
IP2056 10:20-10:35	Title: Robust Animal Pose Estimation as a Visual Perception Front End for Pig-Farm Multimodal Large Models Author(s): Jie Ta, Zhou Yang, Jin Liu, Fuchuan NI Presenter: Jie Ta, Institute of Science and Technology Information of Xizang Autonomous Region, China Abstract: Markerless pose estimation can serve as a structured visual-perception front end for downstream pig-farm multimodal large-model reasoning, where camera observations may ultimately be combined with audio, thermal, environmental, production, and veterinary knowledge. Raw video alone is difficult for a large model to interpret consistently because dense animal contact, occlusion, texture interference, and uncertain keypoints are not explicitly represented. This paper formulates V-NSAFormer Enhanced as a robust pose-perception front end that converts animal images into keypoints, confidence values, and entropy-based uncertainty descriptors. Vision-native sparse attention (V-NSA) combines block feature compression, Top-k block selection, and block-position-guided window fusion to suppress cross-instance interference while retaining spatially plausible evidence. Entropy-based noise mitigation (ENM) then identifies uncertain heatmaps and performs fuzzy-cluster-conditioned refinement. On AP-10K and LoTE-Animal, the enhanced base model reaches 72.8 and 51.1 AP, respectively, while the small model reaches 70.1 and 47.9 AP with reduced computation. A pose-packet interface is further defined for downstream fusion with pig-farm audio, thermal, environmental, and production records. The experiments validate the robust pose front end rather than a complete end-to-end disease-diagnosis large model; nevertheless, the resulting anatomical representation and uncertainty signals establish a reproducible interface for behavior understanding, health-risk screening, and human-reviewed management support.
IP3088 10:35-10:50	Title: MATCH: modular alignment with topology-aware compatibility for HDR reconstruction Author(s): Boyu Zhang Presenter: Boyu Zhang, Beijing University of Technology, China Abstract: Multi-exposure high dynamic range (HDR) reconstruction merges differently exposed images to recover detail in saturated and underexposed regions. Most existing methods couple a fixed alignment module to a fixed fusion network and apply the same reconstruction path to every scene, whereas recent agent-driven frameworks gain flexibility by treating pretrained modules as selectable tools. Flexibility alone, however, does not make independently trained modules compatible: a fusion network encounters alignment residuals it was never trained on, so the same warped non-reference features improve some scenes and degrade others. We present MATCH, a modular framework that composes frozen alignment and fusion experts through a lightweight adapter, termed the Receiver, inserted at architecture-specific points inside each frozen fusion network. The Receiver regulates how aligned non-reference features enter attention, cross-exposure interaction, and feature fusion, while both pretrained experts remain unchanged. Given an exposure bracket and a user preference profile, a vision-language planner selects one alignment expert and one fusion expert from nine available alignment–fusion routes and compiles the corresponding executable route. On Kalantari Test15, Receiver-mediated composition improves mean PSNR-μ over direct composition by 0.089 dB for AHDRNet and 0.122 dB for SCTNet, while reducing the number of scenes that fall below the frozen fusion

	expert from 5 to 0 for SCTNet and from 7 to 3 for HDR Transformer, and leaving this count unchanged for AHDRNet. These results indicate that the Receiver improves the reliability of modular composition rather than absolute reconstruction quality, providing a learned interface for scene-adaptive and preference-adaptive HDR reconstruction.
IP3091 **10:50-11:05**	Title: MEDS-GMR: Multi-source Error-aware Directional Spectral Kernel Method for Dynamic Gaussian Mesh Reconstruction Author(s): Wenju Wang, Chen Zhu, Shaokang Zhou, Yunfei Zhao, Yao Zhu, Sen Wang Presenter: Chen Zhu, University of Shanghai for Science and Technology, China Abstract: Existing dynamic Gaussian mesh reconstruction methods still suffer from insufficient deformation representation, inadequate detail preservation, and inconsistent Gaussian-mesh constraints. To address these issues, this paper proposes MEDS-GMR, a multi-source error-aware directional spectral kernel method for dynamic Gaussian mesh reconstruction. MEDS-GMR designs a phase-enhanced multi-plane spatio-temporal feature representation module, which adopts phase-enhanced spatio-temporal plane coordinate construction, multi-scale resolution-aware plane feature fusion, and paired-branch gated fusion to enhance the representation ability of dynamic Gaussians for non-rigid motion. MEDS-GMR develops a direction-adaptive spectral kernel filter, which uses learnable direction-related spectral responses to improve the preservation of local geometric details during explicit mesh reconstruction. MEDS-GMR constructs a multi-source error-aware Gaussian surface refinement module, which enhances the consistency between dynamic Gaussians and explicit mesh surfaces through Gaussian continuous surface projection, multi-source rendering error mapping, and coverage-aware adaptive refinement. Experimental results on the D-NeRF and DG-Mesh datasets demonstrate that MEDS-GMR achieves improved dynamic mesh rendering quality and competitive explicit mesh reconstruction performance. Our source code is available at https://github.com/zhuchen10103-afk/MEDS-GMR.
IP3107 **11:05-11:20**	Title: Deep-learning initialization for fixed-budget ENZ complex-pupil retrieval Author(s): Jiaqi Deng, Wencan Wei, Chonglei Zhang Presenter: Jiaqi Deng, Shenzhen University, China Abstract: Focal-plane complex-pupil retrieval is commonly formulated as a nonlinear optimization problem, but its finite-budget accuracy depends strongly on initialization. We investigate a hybrid strategy in which a fixed ResNet34 predicts 66 complex extended Nijboer-Zernike (ENZ) coefficients from three through-focus point-spread functions, and a physics-based ENZ optimizer subsequently refines only the coefficient vector. The fixed network supplies the initial coefficient vector, allowing the study to quantify how initialization changes convergence under a limited update budget. Using independent numerical test sets, we compare the learned estimate with piston-only, small-random, and prior-scale random initializations under identical Adam settings and iteration budgets from 0 to 200. The learned initialization reduces the mean full-aperture complex NRMSE from 0.1496 before refinement to 0.0494 after 30 updates and 0.00970 after 80 updates, whereas the three data-independent initializations remain at or above 0.374 after 80 updates. Under a complete-complex-coefficient scaling factor of 1.5, the median error decreases from 0.2678 to 0.1375 and 0.05949 after 30 and 80 updates, respectively, with improvement in all 30 paired samples and across radial orders 1 through 10. A matched-sample solver benchmark further shows that learned-initialized Adam-80 reaches a mean error of 0.00963 in 0.1049 s, whereas piston-initialized Adam-400 reaches 0.0278 in 0.5038 s. A longer piston-initialized L-BFGS solve reaches a lower absolute error at a larger time cost. These results show that deep-learning initialization improves the accuracy-time trade-off of finite-budget ENZ retrieval. The 30- and 80-update settings are practical short-budget operating points rather than globally optimal iteration counts.
IP3125 **11:20-11:35**	Title: Vision-Prior Enhanced Dynamic Contrastive Decoding Author(s): Kaixu Jin, Qidong Guo, Xiaoli Hao Presenter: Kaixu Jin, Taiyuan University of Technology, China Abstract: Vision-language models are susceptible to language priors during inference and may generate responses that are inconsistent with image content. To address this issue, this paper proposes Vision-Prior Enhanced Dynamic Contrastive Decoding (VEDC). First, a vision-enhanced

	positive branch improves the use of visual evidence by enhancing visual tokens. Second, a prior-enhanced negative branch uses text-only prompts to explicitly characterize language priors in the absence of visual input. A dynamic contrastive weight is then computed from the Jensen-Shannon divergence between the output distributions of the two branches. The dynamic contrastive weight is computed at each generation step, allowing the model to adjust the strength of contrastive fusion according to the divergence between the two branches. This adaptive mechanism helps suppress language priors when the branches disagree while avoiding excessive intervention when their predictions are consistent. Finally, the predictions of the two branches are fused at the logit level to strengthen visual evidence and suppress interference from language priors. The proposed method requires no additional training and operates entirely at inference time. On the POPE object hallucination benchmark, VEDC improves the mean F1 score over the original decoding method by 1.41 percentage points and outperforms fixed-weight contrastive decoding and the inference-time baselines evaluated in this paper. It also preserves strong multimodal reasoning and general understanding performance on ScienceQA and MMBench, demonstrating its ability to mitigate visual hallucinations while maintaining cross-task adaptability.
IP3129 **11:35-11:50**	Title: DSTFN: Global-Local Complementary Spatiotemporal Modeling for Continuous Valence-Arousal Estimation Author(s): Yulong Yan, Chengju Zhou, Yan Liang Presenter: Yulong Yan, South China Normal University, China Abstract: Continuous valence and arousal estimation from facial videos is a fundamental task in affective computing. However, in unconstrained videos, subtle and transient facial dynamics that carry affective information are easily overwhelmed by global appearance and background variation. Existing approaches predominantly model facial videos through a single global spatiotemporal representation, which may overlook fine-grained local facial dynamics. Moreover, the complementarity between global context and local dynamics is often underexploited, as existing methods typically adopt fixed fusion strategies and provide supervision only for the final representation. To address these challenges, a Dual-branch Spatiotemporal Fusion Network (DSTFN) is proposed for frame-level valence and arousal estimation from facial videos. DSTFN jointly models holistic facial context and fine-grained local dynamics through complementary spatiotemporal branches. The global branch combines a temporal convolutional network with a Transformer encoder to capture long-range spatiotemporal dependencies, while the local branch employs a Local Patch Dynamics Module (LPDM) to characterize subtle facial dynamics from keypoint-centered spatiotemporal patches with an explicit facial mask. To exploit the complementarity between the two branches, an adaptive semantic fusion strategy is introduced that estimates their representational consistency using cosine similarity and dynamically modulates their contributions. Furthermore, a multi-tier supervision strategy jointly supervises the global, local, and fused representations with shared prediction weights, providing dense learning signals and encouraging the branches to learn a common prediction space. On the Aff-Wild2 and AFEW-VA benchmarks, DSTFN achieves the highest arousal and mean concordance correlation coefficients (CCC) among the compared methods. These results demonstrate that explicitly modeling and exploiting the complementarity between global facial context and fine-grained local dynamics can improve continuous affect estimation from facial videos.
IP3139 **11:50-12:05**	Title: Prompt-guided modality augmentation and progressive alignment for visible-infrared person re-identification Author(s): Zengyi Wang, Bo Yin, Zhiqiang Wei, Yiyun Guo Presenter: Zengyi Wang, Ocean University of China, China Abstract: Visible-infrared person re-identification (VI-ReID) remains challenging because of the substantial modality discrepancy between visible and infrared images and the optimization instability introduced by stage-wise crossmodality training. To address these issues, we propose PMAA-Net, a Transformer-based framework that improves cross-modality matching from the input, representation, and training perspectives. First, Dynamic Modality Augmentation (DMA) constructs an adaptive intermediate modality for each visible image by selecting an appropriate transformation according to image statistics, thereby reducing the visible-infrared discrepancy

	at the input level. Second, an instance-adaptive prompt mechanism is incorporated into a shared Vision Transformer to provide explicit token-level guidance for learning modality-robust identity representations from heterogeneous observations. Third, Progressive Modality Alignment Gating (PMAG) establishes a progressive training schedule in which the visible-side input continuously evolves from the DMA-generated intermediate modality toward the original RGB modality, while the optimization objective switches from intra-modality discrimination to crossmodality alignment at a predefined stage. Extensive experiments on SYSU-MM01 and RegDB demonstrate the effectiveness and complementarity of the proposed components. PMAA-Net achieves 69.67% Rank-1 accuracy and 67.44% mAP on the SYSU-MM01 all-search setting, and 89.20% Rank-1 accuracy and 82.59% mAP on RegDB under the visible-to-thermal setting. Ablation studies further confirm that DMA, dynamic prompting, and progressive alignment provide consistent improvements over the PMT baseline across different evaluation protocols.
IP1006 **12:05-12:20**	Title: MambaASPP: Lightweight Polyp Segmentation with Orthogonal Multi-Scale Enhancement Author(s): Ye Xingyan, Li Qichen, Jianyuan Chen, Zhang Zhutao,, Lin Zhang Presenter: Qichen Li, Sun Yat-sen University, China Abstract: Mamba bottlenecks model long-range dependencies, whereas Atrous Spatial Pyramid Pooling (ASPP) explicitly aggregates multi-scale spatial context. We study their combination in MambaASPP, a lightweight polyp-segmentation model with 5.96M parameters, and report a descriptive comparison of 15 configurations. All reported segmentation scores are point estimates from one seed-42 split and one training run per configuration; the same validation partitions were used for epoch selection and reporting. The results are mixed. Relative to BAMPolyp, MambaASPP has higher validation IoU on CVC-ClinicDB (0.8908 vs. 0.8837) but lower IoU on Kvasir-SEG (0.8823 vs. 0.8886) and CVC-ColonDB (0.8376 vs. 0.8609). In cross-domain validation after Kvasir-SEG training, it is higher on CVC-ColonDB (0.6456 vs. 0.6075) but lower on CVC-ClinicDB (0.7595 vs. 0.7726). These observations motivate an Orthogonal Enhancement heuristic: explicit multi-scale context may complement a Mamba bottleneck in some architectures and datasets, but the present single-run evidence neither quantifies capability overlap nor establishes it as a causal mechanism.

ONLINE SESSION 3

- **Topic: Segmentation and Boundary-Aware Learning**
- **Room C: 814 5945 4849 Password:0608**
- **Time: 10:00-12:05, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Assoc. Prof. Junfei Shi, Xi'an University of Technology, China**

Invited Speaker 10:00-10:20	**Xiwen Zhang, Professor** **Beijing Language and Culture University, China** **Speech Title: Three views on intelligently extracting and generating information from image**
IP2034 10:20-10:35	Title: Boundary-Supervised ViT–U-MixFormer for Urban Scene Semantic Segmentation Author(s): Xiang Shi, Yancong Dengb Presenter: Xiang Shi, Shanghai Maritime University, China Abstract: We investigate decoder design and auxiliary boundary supervision for semantic segmentation with isotropic Vision Transformers (ViTs) on Cityscapes. Four intermediate encoder features provide same-resolution inputs to U-MixFormer. In a ViT-B/16 comparison, validation mIoU is 71.09% for a simplified SETR-MLA-style decoder, 71.59% for U-MixFormer, and 73.08% for U-MixFormer with label-derived boundary supervision. Adding boundary supervision also improves thin-class mIoU by 3.02 percentage points (pp) and boundary F1 at a two-pixel tolerance by 2.64 pp. On ViT-S/16, mean mIoU across three training seeds increases from 65.92 ± 0.19% to 68.03 ± 0.32%, with a paired gain of 2.11 ± 0.13 pp (mean ± sample standard deviation). Paired image bootstrap intervals indicate positive gains for the selected ViT-S checkpoints. Earlier ablations identify λb = 3.0 as the best tested supervision weight, with refinement changing mIoU by −0.05 and +0.02 pp in the two tested configurations. With the auxiliary head active, forward latency is 208.94 versus 209.60 ms/image for ViT-B and 100.08 versus 100.86 ms/image for ViT-S. These results support auxiliary boundary supervision for both ViT sizes under the evaluated Cityscapes protocol.
IP3061 10:35-10:50	Title: BRC: Boundary-aware Region Consistency for Semi-Supervised Medical Image Segmentation Author(s): Shumin Li, Huaxiong Yao Presenter: Shumin Li, Central China Normal University, China Abstract: Consistency regularization and pseudo-label learning are mainstream paradigms for semi-supervised medical image segmentation, enforcing prediction invariance under perturbations and performing self-training with pseudo-labels. However, existing methods predominantly focus on overall accuracy while lacking explicit modeling of anatomical boundaries, leading to coarse contours and large surface errors. To address this, we propose Boundary-aware Region Consistency (BRC), a lightweight module integrated into a dual-teacher framework that enforces boundary feature representation and multi-level consistency alignment between labeled and unlabeled data. BRC constructs three semantic regions—whole, center, and edge—from pseudo-labels via morphological operations, applies differential confidence filtering to suppress noisy predictions, and adopts progressive training for region-specific constraints. It also enforces teacher–student alignment in both the probability and deep feature spaces for stable boundary representations. Evaluated on left atrium MRI (LA) and pancreas CT (Pancreas-CT), BRC effectively compensates for the boundary modeling deficiency of conventional semi-supervised methods, achieving superior performance in both overall accuracy and boundary metrics against several state-of-the-art algorithms.
IP3065 10:50-11:05	Title: HFR-Net: A Multi-scale and Attention-based Network for Colon Polyp Segmentation Author(s): Chen Wang, Fubao He Presenter: Chen Wang, Jingdezhen Ceramic University, China Abstract: Colorectal cancer mainly develops from adenomatous polyps, and accurate segmentation of polyps in colonoscopy images is essential for reducing the mortality rate of colorectal cancer. Aiming at the deficiencies of existing segmentation models, including weak

	local–global feature interaction, insufficient multi-scale context modeling, and inadequate cross-level semantic fusion, which easily cause inaccurate segmentation and missed detection of small polyps, this paper proposes a multi-scale and attention-based network(HFR-Net) for colon polyp segmentation. The proposed network adopts ResNet50 as the backbone to construct a U-shaped architecture. Specifically, a Global-Local Coordinate Attention module is embedded in the skip connection branches to integrate global channel semantics and local spatial coordinate information, thereby bridging the semantic gap between encoder and decoder features. A Residual Dense Atrous Spatial Pyramid Pooling module is constructed at the bottleneck to capture multi-scale contextual information using dilated convolutions with multiple dilation rates. Furthermore, a Multi-scale Semantic Distillation module is designed to align and adaptively fuse multi-level features, effectively combining shallow texture details with deep high-level semantic information. Experimental results on three public datasets, including CVC-ClinicDB, Kvasir-SEG, and CVC-ColonDB, show that the proposed HFR-Net achieves Dice coefficients of 0.9383, 0.9153, and 0.9229, along with IoU values of 0.8954, 0.8681, and 0.8624, respectively.
IP3076 **11:05-11:20**	Title: Water-Cloth Stroke Extraction Based on Temporal Analysis and Instance Segmentation Author(s): Shuo Wang, Haifeng Lin, Zhihui Zhang Presenter: Shuo Wang, Beijing Information Technology College, China Abstract: Water-cloth is a reusable fabric writing surface on which characters drawn with a water-dipped brush gradually fade as the water evaporates. This paper investigates its use as an eco-friendly board for classroom teaching. Owing to the transient nature of the strokes, teachers cannot easily preserve instructional content for student review or remote-learning replay. We propose a novel method to extract complete blackboard-style strokes from water-cloth video sequences captured in teaching scenarios. The method combines pixel-level human removal using YOLOv8 instance segmentation, a dynamic-startup mechanism, and a pixel-classification scheme based on the statistical characteristics of gray-value temporal variations for background and stroke pixels. Experiments are conducted on video sequences covering wet, semi-dry, and occluded mixed conditions. The average F1 score reaches approximately 0.91 with IoU of approximately 0.84. For performance optimization the system compares five configurations (three segmentation models and two inference engines); the optimal configuration (YOLOv8n-seg + OpenVINO) processes a frame in approximately 98 ms on an Intel Core Ultra 7 CPU with RTF about 2.94 and effective FPS of about 10, though it cannot fully satisfy the 24-30 FPS live-stream real-time demand. A three-parameter sensitivity analysis and a five-scenario robustness verification further confirm the stability of the scheme. The proposed technique offers a low-cost, feasible technical route for automatic preservation, review, and remote reuse of water cloth content in future digital classrooms.
IP3082 **11:20-11:35**	Title: Retinal-Mamba: Retinal Vessel Segmentation Using Multi-Expert Polygon Scanning and Spatial-Frequency Intensity-Direction Attention Author(s): Yuanyuan Peng, Wen Li Presenter: Wen Li, East China Jiaotong University, China Abstract: Accurate retinal vessel segmentation remains challenging because of vessels span a wide range of scales, exhibit tortuous topology, and often have weak contrast with surrounding tissue. We propose Retinal Mamba, a U-shaped segmentation framework that combines multi-expert polygon scanning gated Mamba (MPS-Mamba) with spatial frequency guided intensity direction collaborative attention (SFIDA). MPS-Mamba constructs a pool of polygon oriented patch scanning paths and uses a mixture of experts router to activate the top-k paths for each input. A lightweight LSTM style gate then fuses routed state features to strengthen continuity along complex vascular structures. In skip connections, SFIDA couples global frequency priors with multi scale gradient magnitude and direction features. Frequency derived modulation calibrates local gradients before bidirectional attention, improving the representation of faint boundaries and fine vessels. Experiments on DRIVE, STARE, and CHASE_DB1 demonstrate competitive IoU and F1-score performance. Ablation studies show that MPS-Mamba and SFIDA independently improve segmentation overlap, while their combined benefits vary across datasets and evaluation metrics.

IP3119 **11:35-11:50**	Title: Brain Tumor MRI Segmentation with Spatial Response Enhancement and Boundary Awareness Author(s): Dan Yu, Hong Fan Presenter: Dan Yu, Shaanxi Normal University, China Abstract: Accurate segmentation of brain tumor regions is important for clinical diagnosis, treatment planning, and disease assessment. However, high-quality annotated cases for model training may be limited due to differences in case sources, scanning conditions, patient populations, and the high cost of pixel-level annotation. Under limited-data conditions, multimodal brain tumor MRI segmentation faces challenges such as small lesion regions, ambiguous boundaries, class imbalance, and interference from redundant background responses. To address these issues, this paper proposes a multimodal brain tumor MRI segmentation method based on spatial response enhancement and a boundary-aware loss. The proposed method adopts a U-Net-style encoder-decoder architecture and introduces a spatial response enhancement module into the encoder to adaptively reweight multimodal joint features, thereby enhancing tumor-related responses and suppressing redundant background information. In addition, a region boundary-aware loss is designed by incorporating a boundary weight map into the BCE and Dice losses to strengthen supervision around tumor boundaries and alleviate missed segmentation of small regions. Experiments are conducted on the BraTS2020 dataset under a limited-data setting and on the BraTS-Africa dataset, with evaluation performed on the whole tumor, tumor core, and enhancing tumor regions. Experimental results show that the proposed method achieves favorable performance in terms of Dice, IoU, and HD95. Ablation studies further demonstrate the effectiveness of the spatial response enhancement module and boundary-aware loss, indicating that feature-level response enhancement and loss-level boundary supervision provide complementary benefits for multimodal brain tumor segmentation with limited training data.
IP3137 **11:50-12:05**	Title: Visual Reference Integration for Reasoning Segmentation in Remote Sensing Imagery Author(s): Weijun Tian, Xiaofeng Wu Presenter: Weijun Tian, Fudan University, China Abstract: Remote sensing reasoning segmentation requires identifying a target implied by a functional or spatial question and delineating its pixels. Spatial prompts can convey the inferred target to a segmenter, but points and boxes do not describe its complete foreground extent. We introduce visual reference integration to supply this missing information from annotated training images and incorporate it into the query prediction. A fixed patch bank stores visual features, foreground occupancy, and a descriptor of each patch's affinity to the prompted target and its surroundings. Retrieval matches both appearance and this prompt-relative descriptor, producing spatial foreground prior maps. One prior adjusts the initial foreground scores with a prediction-dependent weight; a second enters the segmenter as a dense mask prompt. The final prediction combines the adjusted scores with the logit difference between reference-conditioned and original decoding. Training updates the reference-conditioned decoder while preserving the released reasoning model, spatial prompts, feature encoders, and original decoding path. On 1,928 fixed EarthReason test queries, the method achieves 72.578% gIoU and 74.724% cIoU over three training seeds. With matched initialization, training data, query schedules, and update counts, it improves ordinary decoder continuation by 0.682 and 0.957 percentage points, respectively, with gains at every seed. The corresponding gains remain 0.688 and 0.990 points after exact training and development image overlaps are excluded. These results demonstrate the value of combining retrieved foreground annotations with the query-specific segmentation response under reasoning-derived spatial constraints.

ONLINE SESSION 4

- **Topic: Computer Vision and Machine Intelligence (Track 3)**
- **Room A: 818 3371 3809 Password:0608**
- **Time: 13:30-15:30, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Dr. Liew Lee Hung, Universiti Teknologi MARA, Malaysia**

IP1004 **13:30-13:45**	Title: High-Level Guided Feature Fusion with Asymmetric Gating for Lightweight Maritime Object Detection Author(s): Chengqi Zhou;Yancai WU;Yihong zhang Presenter: Chengqi Zhou, Donghua University, China Abstract: With the widespread adoption of drone remote sensing technology, real-time identification of floating objects on the sea surface using deep learning algorithms has become key to assisting maritime search and rescue. However, challenges such as the extremely low pixel occupancy of maritime targets, severe background noise interference, and limited on- board computing resources of drones remain major bottlenecks in this field.To address these issues, this paper proposes HPC- YOLO, a lightweight model for small maritime target detection. Based on YOLOv10s, the model incorporates a High-level Screening-feature Fusion Pyramid Network (HSFPN) into the neck for multi-scale feature fusion. HSFPN applies channel attention to select features at each scale and uses high-level semantic information to guide the screening and fusion of lower-level features. This design combines spatial detail from lower-level features with semantic information from higher-level feature,the C2f- AdditiveBlock-CGLU module is adopted to capture global spatial correlations, and a convolutional gated linear unit (CGLU) is used to adaptively suppress background noise. Finally, the detection head incorporates the C2f-PoolingFormer-CGLU module, which injects a global view at very low computational cost while enhancing multi-scale target parsing.During training, a specific weight initialization strategy is employed, including Gaussian initialization for convolutional layers and normal distribution initialization for batch normalization layers. Experimental results on the seadrones dataset show that HPC- YOLO significantly improves detection precision and recall while maintaining a lightweight architecture, effectively balancing real-time edge deployment with detection performance in complex maritime environments
IP1019 **13:45-14:00**	Title: Lightweight Multi-scale Semantic Detail Fusion Network Author(s): Yiyuan Wang, Yaozong Zhang, Chaojun Li Presenter: Yiyuan Wang, Wuhan Institute of Technology, China Abstract: Existing lightweight semantic segmentation methods have made considerable progress by introducing multi-scale feature aggregation and spatial detail recovery strategies. However, there are still three key problems to be solved, including insufficient local structure preservation during continuous downsampling, limited multi-scale semantic representation under low computational cost, and weak boundary discrimination in cross-scale decoding. To address these challenges, this paper proposes a Lightweight Multi-scale Semantic Detail Fusion Network (LMSDNet) for real-time semantic segmentation. Specifically, we first design a High-frequency Difference Structure-preserving Downsampling (HDSD) module to capture high-frequency structural responses and inject gated detail information into the main downsampling branch, thereby reducing the loss of edges, small objects, and thin structures. Then, we develop a Lightweight Dilated Semantic Extraction (LDSE) module, which adopts dual-branch depthwise dilated convolutions to extract local details and enlarged-context information with low computational cost. Furthermore, we propose a Structure-Boundary Fusion Decoder (SBFD) module to align shallow and deep features and enhance structural continuity and boundary responses through structure enhancement and boundary detail enhancement paths. Finally, the semantic prediction map is generated by a segmentation head. Quantitative and qualitative experiments on the Cityscapes and CamVid datasets demonstrate that LMSDNet achieves an effective balance between accuracy and efficiency, obtaining 75.20\% mIoU and 112.53 FPS on Cityscapes, and 71.90\% mIoU and 125.71 FPS on CamVid, with only 1.73M parameters and 26.89G FLOPs.

IP2035 **14:00-14:15**	Title: Region-Aware and Semantic-Guided Unsupervised Domain Adaptation Semantic Segmentation in Remote Sensing Author(s): Yu Xin, Hao Wu, Yu Sang Presenter: Yu Xin, Liaoning Technical University, China Abstract: Unsupervised domain adaptation (UDA) for the semantic segmentation of remote sensing images aims at learning transferable pixel-level segmentation models by transferring knowledge from labeled source domains to unlabeled target domains. However, domain discrepancies in imaging conditions, geographical distributions, and scene structures often result in unreliable pseudo-labels when source models are applied to target domains. Unlike random pixel-wise noise commonly observed in natural images, pseudo-label errors in remote sensing images exhibit spatially correlated propagation patterns driven by object structures and spectral characteristics. Existing self-training methods mainly rely on pixel-wise confidence for pseudo-label refinement, limiting their ability to capture and correct such structured errors. To address this issue, we propose a pseudo-label refinement framework integrating regional structural and vision-language semantic information. Class-agnostic masks from the Segment Anything Model (SAM) capture object-level structures, while Contrastive Language–Image Pre-training (CLIP) provides semantic priors to improve regional consistency and reduce category ambiguity. To address the heterogeneity among pixel predictions, region masks, and textual semantics, a trainable multi-view mask-text alignment module is introduced to project multi-source information into a unified mask-level semantic space, where a gating mechanism selectively suppresses potentially erroneous pseudo-label updates. Experiments on LoveDA Rural$\rightarrow$Urban and ISPRS Vaihingen$\rightarrow$Potsdam demonstrate that the proposed method consistently improves performance across two adaptation settings and effectively exploits structural and semantic information in remote sensing unsupervised learning.
IP2052 **14:15-14:30**	Title: YOLO-AM： Lightweight Attention Mechanism Enhanced YOLOv11 for Small Object Detection in Complex Automotive Manufacturing Environments Author(s): Xiaomin Zhao, Zhenxuan Liu, Siow Hoo Leong, Beng Yong Lee, Sofianita Mutalib Presenter: Zhenxuan Liu, Xidian University, China Abstract: Automatic detection of industrial workpieces is essential to intelligent manufacturing in automotive production lines. Detection in such environments remains difficult because of large scale variation, the prevalence of small targets under dense stacking, and interference from metallic surface reflections. This paper proposes YOLO-AM, an enhanced small object detection model built upon YOLOv11. The ADown lightweight down-sampling component is integrated into the C3k2 structures to form C3k2_ADown modules in the backbone and neck, improving feature preservation during down-sampling while reducing parameters by 31.7% and computational cost by 35.4%. A Multi-Scale Feed-Forward Network is further embedded before the detection heads, in which local gating modulates two dilated convolutions to enlarge the receptive field and suppress background noise and reflection artefacts. Experiments on a self-constructed automotive workpiece dataset show that YOLO-AM improves Precision by 1.7%, Recall by 2.1%, mAP@0.5 by 2.8%, and mAP@0.5:0.95 by 4.4% over the YOLOv11 baseline, while using 15.6% fewer parameters and operating at 555 FPS. These results indicate that YOLO-AM is a promising solution for vision-based industrial robotic grasping systems. On the public NEU-DET dataset, YOLO-AM raises mAP@0.5:0.95 from 0.417 to 0.422, providing preliminary evidence of cross-dataset transfer.
IP3075 **14:30-14:45**	Title: Interactive Image Panoptic Segmentation based on Spatial and Multi- component Label Perception Author(s): Xin Sun, Tao Wang, Yinghui Sun Presenter: Xin Sun, Nanjing University of Science and Technology, China Abstract: Existing click-based interactive segmentation methods typically encode foreground and background interactions using fixed channels, which restricts their applicability to open panoptic scenarios involving a variable number of target objects. To address this limitation, we propose an interactive panoptic segmentation framework based on a unified “Position+Label” formulation. The framework separately models the spatial and label attributes of user clicks

	while preserving their complementary roles. A multi-object click-map prompt encodes the spatial distribution of current and historical interactions, enabling the network to capture explicit localization cues during iterative refinement. Meanwhile, instance-aware embeddings provide a flexible, channel-independent representation of click labels for multiple objects. To improve feature extraction at arbitrary click locations, a click continuous field refines label encoding across multi-scale feature spaces and is particularly beneficial for selecting small or thin objects. We further introduce a hierarchical label querying module that progressively enhances the multi-component label representation through global interactions between click embeddings and multi-scale image features. The resulting embeddings are aggregated to generate coherent instance-level masks. Experiments on the COCO, SBD, and DAVIS datasets demonstrate competitive performance against state-of-the-art approaches in both single-object and multi-object interactive segmentation, confirming the overall effectiveness and robustness of the proposed spatial and label modeling strategy.
IP3094 **14:45-15:00**	Title: C2F-Net: A progressive channel-to-frequency fusion network for robust infrared-visible object detection Author(s): Junxin Liu, Yuanhang Wang Presenter: Junxin Liu, Dalian University of Technology, China Abstract: Infrared-visible object detection plays an important role in intelligent perception systems, especially under poor illu- mination and adverse weather conditions. However, the task still faces two major challenges. Specifically, spatial-domain fusion methods often lack sufficient cross-modal interaction to compensate for local degradation in either modality before spatial attention estimation, thereby limiting the effectiveness of subsequent feature fusion. Meanwhile, frequency-domain fusion methods often process low and high-frequency components in the same manner, making it difficult to effectively ex- ploit both low-frequency semantic information and high-frequency structural details. To address these issues, a progressive Channel-to-Frequency fusion network is proposed, termed C2F-Net. Firstly, to mitigate the influence of local degradation in either modality, a Cross-Channel Interleaved Spatial Attention (CCISA) module is introduced, which establishes fine- grained interaction between visible and infrared features before spatial attention estimation and uses their complementary information to guide spatial feature enhancement. Secondly, to fully exploit both low and high-frequency information, a Frequency-Decoupled Wavelet Attention (FDWA) module is designed, which separately processes low frequency seman- tic information and high-frequency structural details to improve feature discrimination and object localization. Extensive experiments on the M 3 FD and DroneVehicle datasets show that C2F-Net achieves higher mAP 50 and mAP than the ex- isting comparative methods across multiple backbone networks. Ablation experiments further verify the effectiveness and complementary contributions of the proposed CCISA and FDWA modules.
IP3120 **15:00-15:15**	Title: Port Container Code Recognition via Multi-View Character Fusion and Coding Constraints Author(s): Hua Yang, Yuejuan Yang, Taiyong Deng, Changqing Han, Min Zhou, Yuanyuan Li, Yuanhao Qiu, Zhongke Huang Presenter: Yuejuan Yang, Wuhan Polytechnic University, China Abstract: To address single-view container number recognition errors caused by oblique viewpoints, illumination variations, and character degradation in port environments, this paper proposes a method combining multi-view character information with coding constraints. Images of the same container captured from multiple viewpoints are used as input. YOLO26n locates container number regions, vertically arranged characters are rearranged horizontally, and PaddleOCR obtains recognition results for each view. To resolve character discrepancies across views, weighted voting is performed at corresponding character positions using recognition confidence and image quality to generate a fused candidate. Each original image is treated as an independent source of support during fusion, preventing repeated votes from its preprocessing variants. For final decision-making, individual recognition results and the fused candidate are jointly scored and ranked using cross-view agreement, the container number format specified by ISO 6346, and check-digit rules. Experiments compare the proposed method with confidence- and quality-weighted selection and whole-string majority voting, while ablation experiments examine the contributions of individual processing components. Using container code regions cropped from YOLO-predicted bounding boxes, the proposed method achieves an exact-match

	accuracy of 96.69% on 121 test groups. Removing character-level fusion reduces accuracy to 95.87%, while removing check-digit scoring reduces it to 80.99%. These results show the contributions of character-level fusion and check-digit scoring under the evaluated configuration.
IP3138 **15:15-15:30**	Title: StockSeg: Material-Aware Shape Priors for multimodal Segmentation of Stockpile Point Clouds Author(s): Kean Wei, Ziheng Wang, Xin Fang, Ling Jiang Presenter: Kean Wei, Guangxi University, China Abstract: Accurate segmentation of stockpile point clouds is essential for reliable volume estimation and inventory management, yet remains challenging in outdoor environments because of sparse LiDAR observations, irregular pile geometries, and ambiguous boundaries between stockpiles and surrounding structures. We propose StockSeg, a multimodal segmentation framework that integrates material-aware shape priors and visual context into point-cloud representation learning. StockSeg constructs Anchor Stacks, a library of material-specific canonical templates that capture representative global morphology. Its AnchorAttention module treats the selected anchor as a queryable geometric memory, allowing observed point features to retrieve shape-consistent structural cues. To provide complementary appearance information, MultiModalAttention encodes the corresponding RGB image into a scene-level descriptor that adaptively modulates the anchor-enhanced 3D representation, without requiring dense point-to-pixel correspondence. Experiments on a real-world outdoor stockpile dataset show that StockSeg achieves an mIoU of 0.695, outperforming PointNet++ with multi-scale grouping by 1.9 percentage points. These results demonstrate that combining explicit material-aware geometric priors with lightweight visual modulation improves stockpile segmentation under sparse observations and ambiguous boundary conditions.

ONLINE SESSION 5

- **Topic: Advanced Deep Learning for Hyperspectral and Remote Sensing Image Processing (Special Session 2)**
- **Room B: 873 1395 5014 Password:0608**
- **Time: 13:30-15:30, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Dr. Huang Nan, Nanjing University of Posts and Telecommunications, China**

IP2038 **13:30-13:45**	Title: Hyperspectral Non-Tobacco Materials Detection Based on Spatial-Spectral Information Enhancement Author(s): Tangyao Guo, Yibo Zhang , Huan Wang Presenter: Tangyao Guo, Nanjing University of Science and Technology, China Abstract: Non-tobacco materials (NTMs) detection is a critical procedure to guarantee the purity of cigarette products. In real industrial production scenarios, NTMs have diverse categories and arbitrary morphologies, and some NTMs share extremely similar color characteristics with normal tobacco leaves, making the detection performance of mainstream methods unable to meet industrial quality inspection requirements. This paper proposes an hyperspectral NTMs detection model SSIENet, based on Spatial-Spectral Information Enhancement. It extracts features through a joint spectral-spatial residual module, which explicitly enhances the discriminative capability of spectral-spatial joint features for NTMs at the front end of the network and solves the problem of NTMs feature loss in deep network propagation. Meanwhile, a spectral-spatial mixed augmentation strategy is adopted. By combining physically meaningful spectral region mixing and a class-aware Mixup algorithm, the strategy augments training samples and alleviates the sample imbalance problem. Furthermore, an industrial-oriented hyperspectral dataset of tobacco leaves and NTMs for cigarette production and processing is constructed. The dataset contains 3,435 hyperspectral images, covering three subsets of different NTMs: lubricating oil, textile materials, and mixed debris. The dataset is characterized by rich background textures, diverse NTMs categories, and large-scale target variations, providing reliable data support for subsequent algorithm research. Experimental results on the constructed dataset demonstrate that SSIENet achieves prominent performance improvements in NTMs detection tasks. In particular, on the textile material subset, the mean Intersection over Union (mIoU) increases from 81.06% to 83.24%, the False Positive Rate (FPR) decreases from 21.85% to 15.25%, and the False Negative Rate (FNR) decreases from 10.00% to 6.77%, verifying its excellent segmentation performance.
IP1018 **13:45-14:00**	Title: Tree species recognition under crown source influence in multimodal remote sensing single-tree scenes Author(s): Guopeng Zhang, Yili Zhao Presenter: Guopeng Zhang, Southwest Forestry University, China Abstract: Accurate individual-tree species recognition depends not only on the classifier but also on the crown footprint used to sample hyperspectral pixels. This study isolates crown representation as a controlled upstream variable in a multimodal RGB–CHM–HSI single-tree workflow. On a strictly matched R1 cohort (N=74), GT masks, pure predicted masks, and predicted bounding boxes were evaluated on identical trees using full 98-band mean spectra, fold-wise standardization, LinearSVC (C=10), and leave-one-out out-of-fold predictions. Macro-F1 was 0.5684 (95% CI: 0.4669–0.6644), 0.4559 (0.3687–0.5266), and 0.3770 (0.2905–0.4643), respectively. Relative to GT masks, paired macro-F1 differences were −0.1125 for pure predicted masks and −0.1914 for predicted boxes; the pure-mask-versus-box difference was +0.0789, but its interval crossed zero. We further modified the GT crown footprint directly on the HSI grid by erosion, dilation, and translation, re-extracted the 98-band spectra, and repeated the same classification protocol. These controlled perturbations changed both crown geometry and extracted spectra, but the classification response was type-dependent and nonmonotonic rather than a uniform degradation. The results show that crown representation is an important

	component of the spectral sampling process and that geometric agreement alone does not fully characterize downstream classification utility under the present protocol.
IP2046 **14:00-14:15**	Title: SCAN: Synergistic Cross-Modal Attention Network for Dual-Camera Compressive Hyperspectral Image Reconstruction Author(s): Yi Wang, Yang Xu, Zhihui Wei, Zebin Wu Presenter: Yi Wang, Nanjing University of Science and Technology, China Abstract: Dual-Camera Compressive Hyperspectral Imaging (DCCHI) presents a robust framework for high-resolution spectral sensing. However, the non-diagonal nature of the sensing matrix in DCCHI leads to a significant challenge known as "spectral entanglement", which hinders the effective utilization of the imaging system. This characteristic combines all spectral bands within the data fidelity term, posing a significant challenge for traditional reconstruction methods. In this paper, we present the Synergistic Cross-modal Attention Network (SCAN), which reframes the reconstruction as a dynamic state evolution process. SCAN overcomes the limitations of static unfolding networks by integrating a Recurrent State Unit (RSU) for cumulative memory and a Dynamic Adaptive Regularizer for stage-aware optimization. A pivotal innovation is the Synergistic Attention Block (SAB), which integrates Global Spectral Attention (GSA) for intrinsic band correlation mining and Simplified Grouped-Query Attention (SGQA) for robust cross-modal spatial-spectral fusion. Extensive experimentation has demonstrated that SCAN exhibits a marked superiority over contemporary state-of-the-art methodologies, attaining a peak signal-to-noise ratio (PSNR) of 44.54 dB and a spectral correlation of 0.9996. The qualitative results obtained from this study further corroborate the hypothesis that the proposed method exhibits a superior capacity to suppress artifacts and restore sharp spatial details in real-world applications.
IP3071 **14:15-14:30**	Title: OSFR-Net: Feature Re-Encoding and Local Refinement for High-Precision Optical-SAR Position Matching Author(s): Ce Kang , Xiao Zhou , Longyun Chi Presenter: Ce Kang, Wuhan University of Technology, China Abstract: Optical and synthetic aperture radar (SAR) image matching is widely used in remote sensing image registration, geospatial information updating, disaster assessment, and multisource image fusion. A key task is to accurately locate the region corresponding to a SAR template within a large optical search image. However, the substantially different imaging mechanisms of optical and SAR sensors lead to pronounced nonlinear radiometric differences between the two modalities. Complex scene structures, local texture variations, and background interference further lead to ambiguous responses, confusion among neighboring candidates, and localization errors during fine-grained position matching. To address these problems, this paper proposes an optical-SAR feature refinement network, OSFR-Net. The network introduces a Structure-Guided Cross-Modal Feature Re-Encoder (SCFR) to enhance cross-modal feature representation using structural information; a Local Contrast Response Refinement (LCRR) module to strengthen local discrimination among candidate responses around the true position; and a Candidate-Guided Local Cross-Modal Re-Encoder (CLCR) to selectively enhance local cross-modal features using information from candidate regions. These three modules progressively improve feature representation, response refinement, and local candidate discrimination. Experiments on the SEN1-2 dataset show that OSFR-Net effectively improves optical-SAR position matching performance, particularly under strict pixel-error thresholds. The proposed method achieves matching accuracies of 86.31% and 91.37% under 1-pixel and 2-pixel error thresholds, respectively, with a mean localization error of 1.6600 pixels and a P95 error of 4.4721 pixels, demonstrating improved localization accuracy and stronger discrimination among neighboring candidates.
IP3113 **14:30-14:45**	Title: Dual-stream hierarchical fusion with calibrated worst-acquisition training for SAR change detection Author(s): Yuxia Ma, Bo Pang, Leyi Lin,, Qi Wang Presenter: Yuxia Ma, Hangzhou Dianzi University, China Abstract: Synthetic aperture radar (SAR) change detection is challenged by the coexistence of boundary-scale structural change and acquisition-dependent intensity fluctuation. This paper

	proposes a dual-stream hierarchical fusion network (DSHFNet) with proxy-calibrated worst-acquisition learning (PC-WAL). The architectural contribution is a scale-wise dual-stream representation in which a local path preserves compact structures while a context path aggregates wider spatial support. Spatially enhanced frequency-channel attention (SE-FCA) couples fixed low-frequency channel descriptors with a learned spatial support mask inside the context path before fusion. The residual modulation strengthens frequency responses only where spatial evidence is present, while the ungated local route maintains boundary continuity. Shared temporal weights, absolute feature differences, and hierarchical decoding complete a date-order-invariant detector. The training contribution, PC-WAL, estimates a bounded acquisition domain from stable training windows, synthesizes correlated positive unit-mean speckle with gain variation, and performs a short inner maximization to expose high-loss acquisition states. PC-WAL is removed after training and introduces no inference overhead. On Ottawa, San, and Yellow River, DSHFNet + PC-WAL attains Kappa values of 91.52%, 90.00%, and 91.96%, respectively, with the lowest overall errors among the listed methods. A separate matched training-schedule study shows that PC-WAL improves mean worst-acquisition F1 over independent augmentation by 1.2, 1.3, and 1.2 percentage points. These results demonstrate accurate structural delineation and consistent compensation within the evaluated acquisition-stress domain.
IP3127 **14:45-15:00**	Title: MS-DETR: Mask-guided cascaded restoration for degradation-robust hyperspectral object detection Author(s): Yingshuo Li, Fuxian Wang , Zhijun Chen, Presenter: Zhijun Chen, Beihang University, China Abstract: The discriminative power of hyperspectral object detection comes from fine-grained diagnostic absorption features, so performance is highly sensitive to radiometric fidelity and spectral-shape integrity of input spectra. Transformer-based detectors such as SpecDETR achieve high accuracy on high-quality data but degrade sharply under sensor noise, missing bands, striping, or mixed-pixel dilution. This occurs because encoder-extracted spectral–spatial representations become structurally distorted when diagnostic absorption profiles are damaged. In this paper, we propose a mask-guided cascaded restoration–detection framework named MS-DETR (Mask-Guided Spectral DETR). Without modifying the SpecDETR backbone, we restructure the conventional “data → encode → detect” pipeline into “mask-guided residual restoration → shared re-encoding → detection.” A physics-aware spectral mask produces structured degradations and marks strong- and weak-degradation domains. A degradation-type-specific residual restorer reconstructs structurally missing regions and refines valid observations under domain constraints. End-to-end training adopts only SpecDETR detection loss; thus restoration is optimized for detection performance after re-encoding, instead of pixel-level reconstruction accuracy. Relative to SpecDETR, MS-DETR loses a small amount of clean mAP but remains usable under mixed spectral degradations. The same cascade also holds under degradation types unseen in training.
IP3072 **15:00-15:15**	Title: MFRE-Net: A Multi-Stage Feature and Response Enhancement Network for SAR–Optical Template Localization Author(s): Yu Wang, Xingang Mou, Longyun Chi Presenter: Yu Wang, Wuhan University of Technology, China Abstract: Synthetic aperture radar (SAR) and optical imagery provide complementary information for Earth observation and are widely used in multimodal remote sensing applications. Accurate SAR–Optical template localization is important for cross-modal image analysis, but the substantial radiometric, textural, and structural differences between the two modalities make reliable matching challenging. Current correlation based methods yield poor modality specific feature representation, offer limited capacity for local structural modeling, and cannot fully tap into complementary information across multi channel correlation responses. To address these issues, this paper proposes a Multi-Stage Feature and Response Enhancement Network (MFRE-Net) for SAR–Optical template localization. Built upon a shared Residual U-Net and FFT-based normalized cross-correlation framework, MFRE-Net integrates three complementary modules. A Modality-Specific Residual Adapter (MSRA) performs lightweight modality-specific correction while preserving the shared feature space. A Bilateral Multi-Scale

	Local Enhancer (BMLE) strengthens local structural representations over multiple receptive-field scales before correlation matching. A Multi-Channel Response Refinement Framework (MCRF) further exploits complementary information among correlation-response channels and adaptively fuses original and enhanced responses based on sample-wise response confidence. Experiments on the SEN1-2 dataset show that the proposed method consistently improves localization accuracy over the baseline and achieves competitive performance compared with representative traditional and deep-learning-based methods under multiple pixel-level evaluation thresholds, demonstrating the effectiveness of the proposed multi-stage enhancement strategy for SAR–Optical template localization.
IP3089 **15:15-15:30**	Title: GPSIVNet: Superpixel-Graph Guided Physics-Constrained Spectral Unmixing for Unsupervised Hyperspectral Change Detection Author(s): Wang Yu'Huan, Xia Ji'An, Hu Ling Presenter: Wang Yu'Huan, Nanjing University of Industry Technology, China Abstract: Unsupervised hyperspectral change detection must distinguish scene changes from radiometric variability without change labels for training. We present GPSIVNet, which combines independent date-specific convolutional encoders with superpixel graph learning and a shared spectral decoder. A learnable gate fuses pixel-level and regional abundance-like coefficients. Bounded latent response fields and band-wise gain and bias parameters represent inter-date variability, while pixel and node reconstruction connect the temporal representations to measured spectra. The proposed five-term objective also includes confidence-masked cross-reconstruction, response-free consistency, and an abundance sum penalty. Change intensity is measured by the Euclidean distance between fused coefficient vectors and converted into a binary map using Otsu thresholding. Experiments on hyperspectral scenes and a supplementary multispectral scene compare GPSIVNet with conventional and learning-based change detection methods. GPSIVNet outperforms the evaluated baselines on Farmland and River and achieves competitive results on Hermiston and Texas. These findings support the use of a shared spectral representation with regional spatial context for unsupervised change detection.

ONLINE SESSION 6

- **Topic: Medical Image Analysis and Clinical AI**
- **Room C: 814 5945 4849 Password:0608**
- **Time: 13:30-15:50, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Assoc. Prof. Wei Li, Jiangnan University, China & Assoc. Prof. Shu-Xu Jing, Chang'an University, China**

Invited Speaker 13:30-13:50	**Ahmad Ali, Doctor Lecturer, Hainan Bielefeld University of Applied Sciences, China** **Speech Title: Attention-Driven Dynamic Spatio-Temporal Graph Learning for Multimodal Urban Traffic Flow Prediction**
IP1015 13:50-14:05	Title: MRFMamba: State Space Model for Efficient Magnetic Resonance Fingerprinting Reconstruction Author(s): Yifan Ning, Yajun Pang, Xuebin Sun, Yanwei Pang Presenter: Yifan Ning, Hebei University of Technology, China Abstract: Magnetic Resonance Fingerprinting (MRF) enables efficient multi-parameter mapping, but accelerated nonCartesian acquisitions introduce severe spatiotemporal aliasing and make parameter reconstruction an ill-posed inverse problem. Conventional dictionary matching is limited by discretization error and high computational cost. Deep learning methods provide a promising alternative, yet CNN-based models mainly rely on local receptive fields and therefore struggle to capture the long-range magnetization dynamics embedded in MRF sequences. Transformer-based models can model global dependencies, but their quadratic complexity increases memory consumption and may over-smooth fine anatomical details. To address these challenges, we propose MRFMamba, a state space model-based network for efficient MRF parameter quantification. The proposed framework adopts a factorized spatiotemporal learning strategy that separates temporal dynamics modeling from spatial artifact suppression. Specifically, a Temporal Bidirectional Mamba Block captures forward and backward signal evolution with linear complexity, enabling effective modeling of long MRF trajectories. The resulting temporal features are then refined by Dynamic Kernel Spatial Refinement (DKSR) Blocks, which adaptively fuse multi-scale depth-wise convolutional bases through an attention mechanism to suppress aliasing artifacts while preserving tissue boundaries and high-frequency anatomical structures. Extensive experiments on simulated BrainWeb and in-vivo 3D-MRF datasets demonstrate that MRFMamba consistently outperforms representative CNN- and Transformer-based baselines, achieving improved accuracy and structural fidelity for both T1 and T2 mapping. Code is available at https://github.com/YifanNing6117/YifanNing.
IP2036 14:05-14:20	Title: Sparse-Dense-Aware Generative Augmentation for Imbalanced Medical Image Classification: A Cross-Dataset Empirical Study Author(s): Boyi Zhang Presenter: Boyi Zhang, Beijing University of Chemical Technology, China Abstract: Class imbalance is prevalent in medical image classification tasks. Although traditional image augmentation can expand the training set through operations such as rotation and flipping, it fundamentally relies on geometric transformations of existing samples, making it difficult to supplement missing rare morphologies within classes. Generative augmentation is considered capable of mitigating the scarcity of minority class samples; however, its effectiveness is jointly influenced by the dataset type, intra-class distribution, the quality of generated samples, and the filtering strategy. We construct a generative augmentation pipeline comprising autoencoder-based feature extraction, CBLOF-based sparse/dense sample partitioning, Sparse-Dense-Aware ACGAN generation, and One-Class SVM (OCS) filtering. Comparative experiments are conducted on BloodMNIST and PathMNIST. Experimental results demonstrate that on BloodMNIST, traditional augmentation (AT) achieves the highest Macro F1 score (0.8656), whereas SD ACGAN + OCS yields a Macro F1 of 0.8558. Although lower than AT, this is higher than the standard ACGAN (0.8461), indicating that sparse/dense conditional modeling and OCS filtering are somewhat effective, yet generative augmentation does not

	outperform traditional augmentation in this case. Conversely, on PathMNIST, the Macro F1 of AT drops from the Original baseline of 0.7643 to 0.7064, while SD ACGAN + OCS achieves 0.7642. This performance is comparable to the Original baseline and significantly outperforms AT, demonstrating that traditional geometric augmentation is not universally applicable to all medical imaging tasks. Further experiments on error-aware selective augmentation reveal that using a fixed confidence threshold causes class filtering bias. While class-wise confidence ranking mitigates this issue, it still fails to surpass the full SD ACGAN + OCS pipeline. This paper concludes that generative augmentation should not be regarded as a universal class-balancing strategy; its effectiveness heavily depends on the data distribution, the quality of generated samples, and the applied quality control strategies.
IP3073 **14:20-14:35**	Title: Early fusion of inertial and pulse features for wrist-worn fall detection and graded warning Author(s): Meiqi Zhong, Linjing Wei Presenter: Meiqi Zhong, Gansu Agricultural University, China Abstract: Wrist-worn fall detectors often confuse short, high-impact daily actions with falls, while a binary alarm provides limited information for subsequent feedback. This study evaluates early fusion of synchronized inertial and pulse-event features using class-balanced logistic regression. The public wrist dataset contains 349 labeled activity records from 21 healthy young participants. Each record was represented by a six-second window spanning four seconds before and two seconds after the acceleration peak. Standardization and classifier fitting were restricted to training subjects in 21 leave-one-subject-out folds. The model achieved 98.78% balanced accuracy with a 95% subject-cluster bootstrap confidence interval of 97.57%–99.61%, 100.00% recall, 97.55% specificity, 97.20% F1 score, and 99.87% area under the receiver operating characteristic curve. Compared with a peak-acceleration threshold, the hard-negative false-positive rate decreased from 28.05% to 6.10%, and missed falls decreased from five to zero. Nine feature configurations showed that pulse features improved motion-based recognition, whereas orientation features did not further reduce classification errors after pulse fusion. Four- and eight-second windows achieved balanced accuracies of 97.41% and 97.68%, respectively. Subject-level analysis and training-only sigmoid recalibration quantified variability and probability quality. An illustrative mapping translates event scores into stable, attention, warning, and emergency states with confirmation and notification actions. The results support multimodal event discrimination on this controlled corpus; the warning thresholds and smartwatch interface remain design proposals requiring prospective and user-based validation.
IP3121 **14:35-14:50**	Title: Customized Virtual Simulation of Cholecystectomy Author(s): Deyao Ma, Shiyu Jia Presenter: Deyao Ma, Qingdao University, China Abstract: Laparoscopic cholecystectomy is a commonly performed minimally invasive surgical procedure, but the anatomical structures surrounding the gallbladder vary considerably among patients. Conventional virtual cholecystectomy simulators mainly rely on generic anatomical models and therefore have limited capability to reproduce patient-specific surgical environments. To address this issue, this paper presents a customized virtual cholecystectomy simulation method based on patient-specific anatomical reconstruction and an analytic electrocautery hook model. First, three-dimensional anatomical models are reconstructed from patient medical images and subsequently converted into voxel-based tissue models to construct a patient-specific virtual surgical environment. To support realistic interaction between the surgical instrument and the models, the electrocautery hook is decomposed into three segments: two cylinders and a quarter torus. The surfaces of these segments are discretized into collision points using parametric equations. These points are used for detecting collisions between the electrocautery hook and the models. To simulate cutting using the electrocautery hook, voxels inside the electrocautery hook are detected and deleted. Adjacent voxels are connected by links. Each link with both voxels deleted is also deleted. For each link with only one voxel deleted, its intersection point with the boundary of the electrocautery hook and the normal vector at the intersection point are calculated using binary search and closest point search algorithms. These intersection points and normal vectors are used to construct the cut surface. Experimental results using patient-specific abdominal models demonstrate that the proposed method can

	construct customized surgical environments and support effective electrocautery hook–tissue interaction during virtual cholecystectomy.
IP3126 **14:50-15:05**	Title: Spatial Control Visual Autoregressive Model for Retinal Image Generation Author(s): Yuanwan Huang, Wenbo Zhang, Qiang Chen Presenter: Yuanwan Huang, Nanjing University of Science and Technology, China Abstract: Retinal image analysis is crucial for diagnosing a variety of systemic diseases and ocular pathologies. Visual Autoregressive model (VAR) has refined the theoretical framework of visual autoregressive generation by re-conceptualizing the autoregressive paradigm as "next-scale prediction", providing a high-efficiency path for visual generation. However, existing controllable VAR models encounter significant challenges when processing sparse tree-like structures, such as retinal vessel label maps; they struggle to ensure the uniform distribution of control information across multi-scale residuals and maintain precise spatial correspondence. This paper proposes the Spatial Control Visual Autoregressive (SpaVAR) model, which aims to precisely guide image generation by providing scale-aligned control information. SpaVAR integrates multi-scale feature semantic decoupling, a dynamic convolutional multi-scale condition encoder, and a fine-grained spatial control guidance strategy. These components collectively enhance the performance of controllable VAR models in generating color fundus images. Experiments based on the Kaggle Diabetic Retinopathy Arranged dataset demonstrate that SpaVAR outperforms existing models across multiple performance indicators, achieving fine-grained spatial control over retinal vascular structures.
IP3142 **15:05-15:20**	Title: Memory-Augmented Rank-Aware Learning for Parkinsonian Gait Severity Estimation Author(s): Zhuli Chen, Zhuoqun Huang, Jun Zhanga, Peng Yea, Yan Lianga,, Chengju Zhou Presenter: Zhuli Chen, South China Normal University, China Abstract: Gait impairment is a common motor manifestation of Parkinson's disease (PD), thus, Parkinsonian gait assess- ment is vital for PD monitoring and medical intervention tailoring. Conventionally, gait severity is quantified through the Movement Disorder Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS) by physicians, which inevitably introduces subjective variability. Although vision-based gait analysis provides an automated and objective manner for MDS-UPDRS score assessment, these approaches still suffer from easy confusion among adjacent severity levels due to the lack of explicitly modeling the ordered relationship in representation space. We therefore employ the rank-aware representation learning method Rank-N-Contrast (RNC) to impose ordinal constraints on gait representations. However, RNC constructs its ranking comparisons only within the current mini-batch, which may limit the available diversity of severity relationships during training. To address this limitation, a Memory-Augmented Rank-N-Contrast (MA-RNC) learning method is proposed, in which RNC is augmented with class-specific cross-batch memory to provide a broader set of rank- aware comparisons. Our framework is evaluated on two widely used Parkinsonian gait datasets PGC3D and PD-GaM, using participant- independent protocols. The comparison results and ablation studies demonstrate that the proposed MA-RNC yields a improvement in walk-level macro-F1 and validated the benefit of memory-augmented rank-aware repre- sentation learning for Parkinsonian gait severity estimation.
IP5005 **15:20-15:35**	Title: An Interpretable and Reproducible CT-Native Radial Representation for Cross-Domain Craniofacial Shape Modeling Author(s): Yanqi Xin, Yongze Guo, Yao Lu Presenter: Yanqi Xin, Sun Yat-sen University, China Abstract: Head-and-neck CT contains skull geometry but lacks a compact, spatially inspectable interface for cross-domain information-retention analysis. We define a CT-native 384-direction radial outer-envelope representation and a globally shared, direction-indexed sparse radial interface. Fixed, FPS, random, and learned query subsets are alternative representation variants; the shared index follows the standardized crop frame and spherical directions, not anatomical correspondence or semantic keypoints. SegRap2023 plain CT used a fixed 84/18/18 patient split, with paired contrast CT retained only for same-patient input sensitivity. A locked CQ500 sample comprised 200 sampled patients, 195 complete cases, 98 development cases, and 97 final-test patients. On the locked test, Learned128 + MLP achieved 21.40 mm face-depth MAE

	(95% CI 20.07-22.87), versus 24.85 mm for PCA32 + MLP (23.54-26.22) and 21.62 mm for FPS128 + MLP (20.24-23.15). An independent feature-transform grid gave 21.27 mm for learned raw queries and 21.29 mm for FPS raw queries (paired p=0.466). Post hoc PCA32/64/128 reached 24.49/23.10/22.86 mm; learned Query32/64/128 reached 21.80/22.80/21.27 mm, with matched five-seed audit 7.764 +/- 0.012 mm for learned queries versus 7.762 +/- 0.000 mm for FPS. These results support an interpretable, reproducible radial interface that retains information related to the CT-derived visible-surface proxy.
IP1016 **15:35-15:50**	Title: Alzheimer's Disease Classification Algorithm Based on Mining Pathologically-Driven Brain Network Co-Degeneration Patterns Author(s): Xiang Wang, Gang Wang, Maowen Ba Presenter: Xiang Wang, Ludong University, China Abstract: Early identification of Alzheimer's disease (AD) is important for disease intervention and progression monitoring. To effectively characterize the selective impairment of brain networks and the co-atrophy of key brain regions induced by the spatially specific deposition of tau and Aβ proteins during the progression of Alzheimer's disease, this paper proposes an AD classification algorithm based on mining pathologically-driven co-degeneration patterns of brain networks. which integrates morphological atrophy prior-guided structural-functional fused brain network construction and a positive-negative dual-stream graph convolution fusion mechanism tailored for signed fused brain networks. The method first constructs a morphological covariance prior from structural morphological atrophy features of individual subject and uses it to constrain individual functional-connectivity networks, thereby an individual structural-functional fused brain network is constructed to identifying potential pathways for AD pathological propagation while suppressing weak and noisy connections. Then, the positive and negative connections in the fused brain network are separately fed into a dual-stream graph convolutional network to learn cooperative and antagonistic relationships between brain regions. Finally, combined with the morphological atrophy features of brain region nodes, the dual-stream GCN can fully mine the spatially coordinated variation patterns of pathological burden features within the fused brain network. Furthermore, the mechanisms such as gated fusion and node attention are introduced into the classification framework to enhance the AD-specific network-based neurodegenerative representations and interpretability of the model. Validated on 311 individual subjects from the ADNI dataset, the experimental results demonstrate that the proposed method achieves an Accuracy of 0.7255±0.0560, a Macro-F1 of 0.7209±0.0576, and an AUC of 0.8472±0.0438, outperforming traditional machine-learning methods. Keywords: Alzheimer's disease; co-degeneration pattern; morphological covariance prior; GCN

ONLINE SESSION 7

- **Topic: Computer Vision and Machine Intelligence (Track 3)**
- **Room A: 818 3371 3809 Password:0608**
- **Time: 16:00-18:00, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Dr. Ming Xie, Nanyang Technological University, Singapore**

IP1009 **16:00-16:15**	Title: MedDef-VISTA: Vision Transformer with integrated attention-based distillation for medical image classification Author(s): Enoch Kwateh Dongbo, Sijie Niu, Kun Zhang, Patience Fero, Gerald Canaan Sohn,, John Nagbe Kofa Presenter: Enoch Kwateh Dongbo, University of Jinan, China Abstract: Vision Transformers (ViTs) achieve strong performance in medical imaging, yet maximizing clean screening accuracy under class imbalance remains difficult. Adversarial training can improve attack resistance, but it often reduces clean accuracy and adds compute cost. We propose MedDef-VISTA, which stacks CBAM channel–spatial attention, patch consistency regularisation, and a DefenseModule on a ViT backbone, trained with two-stage high-temperature (T = 50) knowledge distillation. On the TBCR tuberculosis test set (n = 420), MedDef-VISTA reaches 98.10% clean accuracy (+2.62 pp over the undefended baseline after distillation; +1.43 pp from distillation alone) and 90% TB sensitivity (63/70). On the Kermany pneumonia validation set (n = 878; 27% Normal / 73% Pneumonia), the same protocol yields 91.57% accuracy (+10.25 pp over baseline), 93.6% pneumonia recall, F1-macro 0.894, MCC 0.789, and AUC 0.959. Distillation improves clean accuracy on all six TBCR variants (0.95–2.62 pp) and on five of six chest variants (+1.25 to +3.53 pp). Ablations indicate that attention-based refinement accounts for the largest gains. All results are measured on unperturbed images; adversarial robustness is neither optimised nor evaluated.
IP1017 **16:15-16:30**	Title: Hierarchical Autoregressive Planning for Text-to-Motion Generation in 3D Scenes Author(s): Anqi Zhou, Zhibo Zhang, Xiaoqi An, Lin Zhao Presenter: Anqi Zhou, Nanjing University of Science and Technology, China Abstract: Despite significant progress in text-to-motion generation, synthesizing language-guided human motions in realistic 3D scenes remains challenging. The challenge mainly arises from two aspects. First, under the limited scale of language-scene-motion data, it is difficult to jointly model the complex relationships among textual semantics, scene geometry, and human motions. Second, existing scene-aware methods mostly rely on conditional diffusion models or autoregressive models, where global interaction planning and local motion details are often coupled during full-sequence generation. To address these challenges, we propose a scene affordance-guided hierarchical autoregressive motion planning framework. Specifically, we introduce Scene Affordance Tokens as an intermediate representation bringing 3D scene awareness into text-to-motion generation. By compactly representing language-guided scene affordance, these tokens transform the challenging text-scene-motion joint modeling problem into an scene affordance-guided motion planning problem. Conditioned on the Scene Affordance Tokens, we further introduce a hierarchical autoregressive planner that decomposes full-sequence motion generation into multiple scales. Coarse scales perform global motion planning, while finer scales progressively recover local motion details, enabling the generation of high-quality human motions that are aligned with textual semantics, satisfy scene constraints, and exhibit plausible body poses and motion patterns. Extensive experiments on the HUMANISE dataset demonstrate the effectiveness of our framework in generating human motions that are consistent with textual descriptions, interact reasonably with 3D scenes, and maintain rich motion diversity.
IP2039 **16:30-16:45**	Title: WHFFormer: A Wavelet Frequency Transformer for Lightweight Image Super-Resolution Author(s): Wu Lai , Chunjiang Duanmu Presenter: Wu Lai, Zhejiang Normal University,China Abstract: Lightweight image super-resolution (SR) aims to recover high-quality images with

	limited computational resources. Existing Transformer-based SR methods mainly perform feature interaction in the spatial domain, which may weaken the representation of texture-related high-frequency information.In this paper, we propose WHFFormer, a lightweight Transformer framework that introduces wavelet-based frequency learning into SR reconstruction. A Wavelet Frequency Transformer Block (WFTB) is designed to progressively select and enhance frequency features. Specifically, the frequency selection stage decomposes features into different wavelet components and introduces direction-aware weighting to capture complementary high-frequency information. Furthermore, a frequency enhancement module adaptively adjusts spectral responses to preserve texture details during feature transformation.Extensive experiments on benchmark datasets demonstrate that WHFFormer achieves competitive reconstruction performance with low computational complexity.
IP2045 **16:45-17:00**	Title: MCHT: Multi-scale collaborative HVI-feature extraction to approximate Taylor expansion for low-light image enhancement Author(s): Zengjiang Li, Yaozong Zhang, Yuanbao Wu Presenter: Zengjiang Li, Wuhan Institute of Technology, China Abstract: Low-light image enhancement (LLIE) is crucial for intelligent surveillance, autonomous driving, and mobile photography, where degraded illumination often causes low brightness, poor contrast, color distortion, and detail loss, severely affecting both visual perception and downstream vision tasks. Although deep learning methods have achieved remarkable progress, most existing approaches rely on black-box mappings without explicit enhancement modeling, and directly process RGB images where luminance and chrominance are highly coupled, leading to limited interpretability and color distortion. In this paper, we propose MCHT, a low-light image enhancement network based on multi-scale collaborative HVI-feature extraction with Taylor expansion approximation. By formulating enhancement as the weighted combination of a base mapping and high-order derivative compensation terms, MCHT provides an interpretable optimization perspective for illumination restoration. Specifically, we construct a base enhancement layer in the HVI color space, where a histogram-guided intensity curve enhancement (HGICE) module improves illumination and a chrominance prior rectification (CPR) module preserves color consistency. Furthermore, a U-shaped derivative network is developed to estimate high-order compensation terms, incorporating an HVI feature fusion (HFF) module for multi-scale feature interaction and a dual-branch Transformer (DBT) for efficient luminance and chrominance modeling. Extensive experiments on LoLv1 and LoLv2-Real datasets demonstrate that MCHT achieves state-of-the-art performance, obtaining PSNR/SSIM values of 23.99/0.8512 and 22.82/0.8695, respectively, while maintaining only 1.50M parameters and 1.36G FLOPs. Ablation studies further validate the effectiveness of the Taylor expansion strategy and proposed modules.
IP3067 **17:00-17:15**	Title: Distance-Robust 3D Sitting Posture Recognition Using Relative Geometric Features Author(s): Taiyu Jin Presenter: Taiyu Jin, Southwest Jiaotong University, China Abstract: Vision-based sitting-posture recognition can achieve high recognition performance under fixed deployment conditions, but its robustness may deteriorate when the camera configuration changes after deployment. Motivated by the 3D posture representation adopted in SitPose, which combines skeletal-angle features with absolute head depth, this study investigates the influence of camera-relative positional information on posture recognition under deployment-condition shifts. Although absolute positional cues provide useful discriminative information under matched conditions, their direct dependence on the camera coordinate system can make the resulting representation sensitive to changes in sensing geometry. To address this limitation, we propose a 13-dimensional 3D posture representation that combines ten geometric-angle features with three signed directional features. The representation is evaluated under controlled distance and combined camera-configuration changes and through broader frozen cross-session transfer protocols. In a representative cross-distance evaluation, it achieved a Macro-F1 of 0.929, compared with 0.682 for a SitPose-inspired representation using absolute 3D head coordinates. Across 14 frozen transfer protocols, the proposed representation achieved a mean Macro-F1 of 0.9253 compared with 0.8459 for the geometric-only representation. All labeled data were collected from a single participant, so the evaluation targets

	robustness to acquisition-condition changes rather than cross-subject generalization. The method was further integrated into a real-time monitoring system incorporating temporal decision logic, deployment calibration, UDP communication, and embedded feedback. During 30 min of continuous monitoring, the integrated system processed 23,513 frames at an average of 13.06 FPS, with a 99.75% valid-frame rate and a P95 host frame-cycle latency of 87.94 ms.
IP3081 **17:15-17:30**	Title: Reliability-Aware Stereo Vehicle Association for Selective 3D Localization Author(s): Minfeng Huang, Junyue Yu, Jiancai Xu, Mingxian Liang, Jiajia Wu Presenter: Minfeng Huang, Guangzhou City Polytechnic, China Abstract: Object-level stereo localization can fail when independently generated candidate sets contain single-view vehicles or ambiguous counterparts, because forced assignment converts these cases into erroneous disparities and potentially large depth outliers. We formulate vehicle correspondence as gated bipartite assignment with candidate-specific unmatched states, frozen masked-HSV appearance, structural and visibility evidence, and a separate reliability stage for selective 3D localization. Oracle experiments permit only per-view candidate masks; cross-view identity, depth, and pose remain evaluator-only. Predicted experiments use RGB and detector rectangles alone. Evaluation uses five scene-grouped folds on Virtual KITTI 2, with clone-selected transfer to five rendering variants and exact scene-cluster bootstrap intervals. On clone oracle candidates, B3 relative to geometry-only B1 increased Unmatched F1 by 7.52 percentage points and reduced singleton false-acceptance rate by 10.26 points. Its Match F1 change was 0.39 points, with a 95% interval of -0.30 to 0.74 points. The same bounded pattern persisted under clone-selected rendering transfer. With frozen B3 associations and a shared Census localizer, the full risk model obtained normalized RMSE AURC 1.104 and failure-rate AURC 0.084, below the best single-cue values of 1.724 and 0.115. In the predicted panel, an SSDLite MobileNetV3 detector reached only 10.24% recall; B3 did not improve over B2. These results support explicit unmatched handling and multi-cue risk ranking under oracle candidates, while showing that candidate recall and rectangular support dominate the current predicted pipeline.
IP3131 **17:30-17:45**	Title: Image-Audio-Spectrum Direct Weighted Fusion Author(s): Huahao Zhao, Jing Liu, Yuncan Qu, Jiayi Li Presenter: Huahao Zhao, Taishan University, China Abstract: To address the problem of insufficient information for small UAV target detection in complex environments, this paper proposes a UAV detection method based on tri-modal weighted fusion of image, audio, and spectrum. The method uses YOLOv8n as the base framework, while constructing an audio classification sub-model and a spectrum classification sub-model. The three modalities are linearly weighted with adjustable weights to construct a multi-modal UAV detection model (MDDM). Experimental results on a self-built UAV multi-modal dataset show that, compared with traditional pure image recognition and bi-modal fusion methods, the MDDM model can improve the average confidence score of small targets. MDDM also performs well in terms of Precision, Recall, and mAP50. In particular, in small target detection scenarios, the confidence is improved by 0.2418 compared with the pure image baseline. Through systematic weight parameter sensitivity analysis, the strong positive correlation between image weight and Precision, as well as the special erosion effect of spectrum weight on Precision, are revealed, providing a reference for parameter configuration in practical deployment scenarios. However, the small-target AP remains essentially flat compared with the pure image baseline, indicating that direct weighted fusion mainly operates at the score-calibration level.
IP3144 **17:45-18:00**	Title: CAD Primitive Image Recognition Based on Multi-Scale Corner Detection and Topological Feature Fusion Author(s): Xinhua Tang, Bowen Guo, Mingxuan Yang, Weiwei Zhang Presenter: Mingxuan Yang, Shandong Jianzhu University, China Abstract: CAD primitive recognition is a key step for engineering drawing understanding, retrieval and reuse. CAD images are usually characterized by sparse texture, simple structure and clean background. A single local feature tends to produce spurious corners and missed detections, while a single global feature is insufficient to depict fine structures, which limits the recognition accuracy of primitives with similar structures. To address this problem, this paper

	proposes a CAD primitive image recognition method based on multi-scale corner detection and topological feature fusion. First, CAD images are converted to grayscale and normalized to a fixed size, and a multi-scale space is constructed to capture corner responses at different scales. Second, the Shi-Tomasi and FAST responses are computed and weighted-fused at each scale, and a K-value based candidate screening strategy is employed to remove spurious corners caused by straight-line interference. On this basis, corners are classified into five types, i.e., L-type, T-type, X-type, endpoint-type, and other-type corners, according to the gradient orientation distribution in the corner neighborhood, and four categories of refinement local features, including type histogram, spatial distribution, geometric relations and contour shape, are constructed. Meanwhile, global topological features are extracted based on connected-component topological invariants and persistent homology. Finally, the corner and topological features are ranked by random forest importance and concatenated into a fused feature, which is fed into a particle swarm optimized (PSO) classifier to determine the primitive category. Experiments on a dataset of five primitive categories (bed, chair, dining table, toilet and washbasin) show that the PSO-optimized XGBoost model achieves a recognition accuracy of 99.4%, demonstrating the effectiveness of the proposed method.

ONLINE SESSION 8

- **Topic: AI for Earth, Agriculture and Emerging Applications**
- **Room B: 873 1395 5014 Password:0608**
- **Time: 16:00-18:05, UTC/GMT+8, Nov. 8, 2026**
- **Chair: Assoc. Prof. Chia Kim Seng, Universiti Tun Hussein Onn Malaysia, Malaysia**

Invited Speaker 16:00-16:20	**Lin Gao, Professor, University of Chinese Academy of Sciences, China** **Speech Title: Dual-Engine-Driven Digital Space Modeling via Explicit Geometric Representations and Implicit Generative Models**
IP3097 16:20-16:35	Title: A Cotton net nep detection method for carding machines based on improved Faster R-CNN Author(s): Bai Yuwei, Zhang Sanchuan Presenter: Bai Yuwei, Shengda University, China Abstract: In order to improve the product quality of carding machine cotton nets and promote intelligent industrial production, an improved Faster R-CNN model is proposed for detecting cotton neps on carding machine cotton nets. Firstly, ResNet50 is used as the backbone network to prevent gradient vanishing through the residual structure; Secondly, the Feature Pyramid Network achieves effective fusion of low-level features and high-level semantic information by simultaneously employing bottom-up and top-down transmission, which not only effectively controls computational costs but also preserves rich semantic information of features at different scales; Then, RoI Align is used to replace RoI Pooling, accurately locating anchor box positions using floating-point calculations; Finally, the use of the SE (Squeeze and Excitation) channel attention mechanism module enables the network to autonomously enhance feature channels rich in information while suppressing unimportant channels, thereby achieving the goal of improving the model's discriminative capability. Experimental results show that the mean Average Precision (mAP) of the improved Faster R-CNN network model has increased to 97.1%, representing a 4.8% improvement compared to before the improvement, the mAP of the target object Lumpy cotton neps improved to 96.17%, an increase of 5.79% compared to before the improvement; the mAP of Granular cotton neps reached 97.13%, an increase of 5.32% compared to before the improvement; and the mAP of cotton miscellaneous reached 97.92%, an increase of 3.21% compared to before the improvement. And the comprehensive evaluation metric F1 has risen to 0.93, an increase of 0.15 compared to before the improvement. Compared with the YOLOv8, YOLOv7, YOLOv5 and Faster R-CNN network models, the improved algorithm exhibits higher detection accuracy and better overall evaluation performance. This method can provide a reference for detecting cotton neps on carding machine doffer nets and help achieve improvements in doffer net quality.
IP2031 16:35-16:50	Title: RemuCast: Reliable Multi-Scale Diffusion Post-Processing for Extreme Precipitation Nowcasting Author(s): Zexi Wang, Jianxiang Li Presenter: Zexi Wang, Wuhan Polytechnic University, China Abstract: Accurate and reliable nowcasting of extreme precipitation is required for short-term disaster warning, but high- threshold rainfall remains difficult to predict from radar echoes. Deterministic spatiotemporal predictors such as SimVP frequently yield over-smoothed forecasts, attenuating high-reflectivity cores and degrading the detection of heavy-rainfall events. Diffusion-based post-processing methods, represented by PostCast, can recover fine-scale details from blurred predictions; however, for extreme precipitation, structural recovery alone is insufficient because stronger enhancement may also induce false heavy-rainfall alarms. To examine this recovery–reliability trade-off, this paper proposes RemuCast, a reliable multi-scale diffusion post-processing framework for TAASRAD19 radar nowcasting. First, a DDIM sparse reverse trajectory is instantiated so that consistency guidance is applied to fewer but more consequential denoising states, resulting in a compact and relatively conservative sampling path. Second, a stage-aware multi-scale consistency guidance strategy is formulated to impose scale-specific constraints on large-scale rainbands, mesoscale rain clusters, and local heavy-rainfall cores across different sampling stages. Finally, a threshold-aware reliability calibration

	module, termed TARC, interprets predicted RR ≥ 30 mm/h connected components as candidate heavy-rainfall events and suppresses low-reliability false alarms according to their shape, intensity, and consistency with the original SimVP evidence. Ex- periments on TAASRAD19 show that RemuCast preserves P1/P4 heavy-rainfall CSI close to PostCast and reduces P16 FAR from 0.813 to 0.758, where P1, P4, and P16 denote pixel-level, 4-pixel-tolerance, and 16-pixel-tolerance evaluation, respectively. This reliability gain is accompanied by a decrease in P16 CSI from 0.118 to 0.108 and average POD from 0.174 to 0.137, indicating a calibrated trade-off rather than uniform metric improvement. Keywords: extreme precipitation nowcasting, diffusion model, PostCast, DDIM, multi-scale guidance, reliability calibra- tion, false alarm suppression
IP3074 **16:50-17:05**	Title: Physics-anchored few-shot segmentation of PV-EV hosting envelopes across unbalanced distribution feeders Author(s): Henghui Mo, Linjing Wei Presenter: Henghui Mo, Gansu Agricultural University, China Abstract: Photovoltaic (PV) and electric-vehicle (EV) hosting studies evaluate dense operating grids by quasi-static time-series (QSTS) simulation, although the feasible region is naturally a two-dimensional image and high-fidelity labels are expensive. We formulate cross-feeder hosting-envelope estimation as few-shot semantic segmentation of a six-channel response surface. The proposed physics-anchored dual-head network (PA-DHNet) fuses a direct feasibility logit with a reconstructed high-fidelity constraint margin, obtained by adding a learned standardized residual to an eight-snapshot low-fidelity margin. The released corpus contains 84 paired-fidelity 13 x 13 maps (14,196 operating points) from three public unbalanced feeders, including 12 confirmatory maps generated after the final configurations had been frozen. Leave-one-feeder-out evaluation provides 0, 2, 4, or 8 complete target maps for adaptation. At four target maps, PA-DHNet obtains mean intersection over union (IoU) 0.929, boundary F1 0.960, and false-safe rate 0.083. A pixel random forest has higher IoU (0.955) but a higher false-safe rate (0.128) and requires 19.60 rather than 1.11 ms per map. At eight maps, PA-DHNet reaches IoU 0.942, the best boundary F1 (0.967), and the lowest learned false-safe rate (0.058); compact DeepLabv3+ is marginally higher in IoU (0.944). PA-DHNet improves boundary F1 over the random forest by 0.019 (p=0.037) and over a residual U-Net by 0.004 (p=0.040) in paired map-level tests. The evidence supports a fast boundary- and safety-aware surrogate after target adaptation, not universal accuracy dominance or replacement of high-fidelity QSTS.
IP3128 **17:05-17:20**	Title: Visual state-space networks for AI-driven classification and segmentation of large-scale solar eruptions Author(s): Yuxuan Li, Chunhui Wang, Linhua Deng Presenter: Yuxuan Li, Yunnan Minzu University, China Abstract: Coronal mass ejections (CMEs) constitute major triggers of space weather disturbances. Their faint, diffuse signatures in coronagraph images pose substantial challenges to automated detection. This work presents a dual-task framework for SOHO/LASCO C2 running-difference images. For frame-level classification, VMamba-FiLM adopts a VMamba-Tiny backbone augmented with feature-wise linear modulation conditioned on an eight-dimensional descriptor vector. For pixel-level segmentation, Mamba-UNet performs hierarchical state-space encoding under pseudo-mask supervision. Event- and time-aware data partitioning ensures rigorous separation between training and test subsets. On the held-out test set, VMamba-FiLM achieves an F1 score of 0.977, while Mamba-UNet yields a Dice coefficient of 0.951. Evaluated on an independent human-annotated dataset with real masks (N=29), Mamba-UNet demonstrates promising generalization with a Dice coefficient of 0.628. It attains markedly higher recall (0.644 versus 0.502) than heuristic rule-based approaches and successfully recovers faint CME structures missed by rigid thresholding. We further demonstrate that 224×224 downsampling introduces spatial aliasing, which acts as a critical performance bottleneck. The proposed framework retains high-throughput inference with a latency of 81.7 ms per frame. These results validate a robust and scalable solution for automatic CME screening and region extraction.
IP3087 **17:20-17:35**	Title: AI-enabled wheat health monitoring: sensing, model-task alignment, and evidence for field decisions Author(s): Meiqi Zhong, Linjing Wei, Henghui Mo Presenter: Henghui Mo, Gansu Agricultural University, China

	Abstract: Artificial intelligence (AI) for wheat health monitoring is often evaluated as an image-recognition problem, although field use requires an engineering chain from sensing to decision support. This review develops an evidence-maturity framework linking biological target, observation scale, dataset provenance, model task, validation domain, deployment constraint and human decision endpoint. We synthesise evidence from disease, pest, stress and lodging monitoring across close-range, proximal, aerial, multispectral and hyperspectral systems. Across 60 cited sources, deep visual models and spectral learning show progress, but recognition accuracy is often confounded with field robustness. Random image splits, incomplete acquisition metadata, inconsistent severity references and limited external testing restrict comparison across fields, seasons, cultivars and devices. Deployment claims are further weakened when latency, memory, energy, calibration and failure handling are not measured in the intended workflow. We therefore distinguish internal recognition, external robustness, measured deployment and decision support. The synthesis indicates that near-term value lies in systems that prioritise scouting, quantify a defined endpoint and defer uncertain cases rather than automate treatment directly. Progress requires grouped and external validation, calibrated uncertainty, versioned benchmarks, sensor-model-hardware co-design and auditable field feedback. The framework makes explicit the evidence needed to move from a promising detector to a field decision without equating recognition accuracy with intervention benefit.
IP3145 **17:35-17:50**	Title: Cross-Domain Few-Shot Detection of Industrial Structural Defects with Context and Normal Prototypes Author(s): Mingle Wang, Wenhua Zhang Presenter: Mingle Wang, Nanjing University of Science and Technology, China Abstract: Industrial structural defect detection is essential for quality control and safe operation in intelligent manufacturing, yet it must distinguish subtle component conditions with limited labeled examples. Traditional inspection methods often generalize poorly, while conventional deep detectors often require extensive annotations and assume a fixed set of classes. ETS provides a pretrained vision–language detection baseline; industrial structural defect detection nevertheless requires adaptation from a few labeled examples, modeling of component–context relationships, and comparison with normal structures. We build on the ETS Swin-B baseline. To address these challenges, we develop three task-specific components: a Progressive Fine-tuning Strategy (PFT) that schedules parameter-group learning rates for target-domain adaptation; Cross-Scale Structural Relation Attention (CSRA) that models dependencies between candidate components and their surrounding installation regions; and Normal Prototype-guided Residual Encoding (NPR) that retrieves query-relevant normal references and represents deviations from expected structure. We also construct the CDHC dataset with bounding-box annotations for one normal condition and five fine-grained defect conditions, providing a benchmark for few-shot industrial defect detection. Experiments on CDHC and NEU-DET compare the proposed method with the ETS baseline and evaluate CSRA and NPR through a matched ablation built on PFT. At 10 shots, with results averaged over three runs, the method achieves 44.3 AP on CDHC and 28.6 AP on NEU-DET, exceeding the reported ETS baseline by 3.2 and 2.5 AP points, respectively.
IP3115 **17:50-18:05**	Title: Structural properties of NEQR quantum image encodings for cancelable biometrics with post-quantum protection Author(s): Gabriela Mogos, Ho-Pun Lam Presenter: Gabriela Mogos, American University in Bulgaria, Bulgaria Abstract: Quantum image representations such as the Novel Enhanced Quantum Representation (NEQR) encode pixel values into orthogonal qubit subspaces, a structure that has been exploited for watermarking and image processing but never systematically analyzed as a cancelable biometric primitive. In this paper we study, both analytically and empirically, how this orthogonal encoding behaves under the perturbations that biometric matching must tolerate. We prove that any single-pixel change in an NEQR-encoded image produces a statevector displacement of fixed $\ell 2$ norm $\sqrt{2}/N$, independent of the magnitude of the pixel change, and that graph-structured unitary folding preserves this $\ell 2$ distance while altering the Hellinger distance over Born-rule distributions whenever colour collisions occur. To make the analysis concrete we introduce GKSS (Graph-Keyed Statevector Scheme), a cancelable template protection process

	that combines NEQR encoding with NIST-standardized post-quantum primitives: ML-KEM-768 (FIPS 203) for key establishment and ML-DSA-44 (FIPS 204) for template integrity; the security of GKSS as a whole is argued informally, and no game-based proof of the complete construction is claimed. Evaluations on synthetic images, FVC2002 subsets, and the SOCOFing dataset (600 subjects, nine alteration configurations) show that the comparison with classical BioHashing is perturbation-type dependent: the quantum encoding retains an advantage on geometric perturbations (central rotation, Z-cut) while classical random projection remains superior on destructive perturbations. Spectral truncation of the statevector to k=1000 amplitudes reduces template storage 52× (19.5 KB) and improves the equal error rate. We discuss the implications of these findings for designers of quantum-image-based biometric schemes

MEMO