8th 2025 PRAI

Conference Program

2025 the 8th International Conference on Pattern Recognition and Artificial Intelligence

August 15 -17, 2025 | Guiyang, China

SPONSORED BY

贵州大学 GUIZHOU UNIVERSITY IEEE

HOSTED BY

贵州大学 GUIZHOU UNIVERSITY | 大数据与信息工程学院 College of Big Data and Information Engineering

CO-SPONSORED BY

电子科技大学 University of Electronic Science and Technology of China 中国航天 西安微电子技术研究所 Xi'an Microelectronics Technology Institute

贵州师范大学 Guizhou Normal University 西南石油大学 SOUTHWEST PETROLEUM UNIVERSITY 西華師範大學 CHINA WEST NORMAL UNIVERSITY 四川轻化工大学 SICHUAN UNIVERSITY OF SCIENCE & ENGINEERING 中南林业科技大学 ZHONGNAN LINYE KEJI DAXUE

此次会议得到王宽诚教育基金会资助

Table of Content

Welcome Message

On behalf of the Conference Committees, we are pleased to welcome you to the 8th IEEE International Conference on Pattern Recognition and Artificial Intelligence (PRAI 2025), held in Guiyang, China, from August 15 - 17, 2025.

The conference is co-hosted by Guizhou University and IEEE. PRAI 2025 invites authors to submit papers covering diverse aspects of pattern recognition and artificial intelligence. Key areas of interest include, but are not limited to, intelligent manufacturing, smart healthcare, AI applications in aerospace, machine learning, multimodal perception, image and video processing, and the integration of AI.

The conference aims to provide an interactive platform for professionals to explore cutting-edge academic and industrial trends, share the latest research findings and technological breakthroughs, and discuss innovative ideas and practical methods. Its goal is to elevate academic research and industrial application standards, supporting global efforts in technological transformation, while promoting the research, development, and application of related technologies both domestically and internationally.

We are delighted to share that this year, we have received enthusiastic responses from researchers and experts worldwide, including those from China, the USA, Europe, Asia, and beyond. After a rigorous peer-review process, we are proud to announce that a selection of high-quality submissions has been accepted for presentation at PRAI 2025, reflecting the exceptional innovation and research excellence our conference strives to showcase.

We extend our profound gratitude to all who have contributed to making this event possible: the esteemed authors, keynote speakers, peer reviewers, and members of the conference committees, whose dedication and efforts have been instrumental in bringing PRAI 2025 to fruition.

We hope all participants will find this conference insightful and rewarding. Your contributions are vital to advancing the frontiers of pattern recognition and artificial intelligence.

Wishing you a successful and inspiring experience at PRAI 2025!

Conference Organizing Committee

August 15, 2025

Conference Committee

Honorary Chairs

Minyi Guo
Shanghai Jiao Tong University, China

Jinrong Wang
Guizhou University, China

General Chairs

Youliang Tian
Guizhou University, China

Li Zhang
Guizhou University, China

Jinyuan Jia
The Hong Kong University of Science and Technology (Guangzhou), China

Conference Co-chairs

Preshanth Jagannathan
National Radio Astronomy Observatory, USA

Xin Zhang
Northeastern University, China

Yongfeng Huang
Nanjing University, China

Yi Mao
Tsinghua University, China

Xiaosi Qi
Guizhou University, China

Technical Program Committee Chairs

Changqing Zhang
Tianjin University, China

Weiyao Lin
Shanghai Jiao Tong University, China

Yifu Cai
University of Science and Technology of China, China

Technical Program Co-chairs

Aijun Dong
Guizhou Normal University, China

Gauhar Ali
Prince Sultan University, Saudi Arabia

Mingwei Cao
Anhui University, China

Local Organizing Committee Chairs

Yong Chen
Guizhou University, China

Xin Zhang
Guizhou University, China

Yulan Zhang
Guizhou University, China

Publicity Chairs

Zhihui Wang
Dalian University of Technology, China

Jianping Gou
Southwest University, China

Ying Ma
Harbin Institute of Technology, China

Publication Chairs

Zhihan Lv
Uppsala University, Sweden

Fanman Meng
University of Electronic Science and Technology of China, China

Qifei Wang
University of California, USA

Jing Liao
City University of Hong Kong, China

Jun Wan
Macau University of Science and Technology, China

Special Session Chairs

Zhufang Kuang
Central South University of Forestry and Technology, China

Daoming Lv
Sichuan University of Science & Engineering, China

Zhong Ma
Xi'an Microelectronics Technology Institute, China

Jia Gu
City University of Macau, China

Shiqi Wang
City University of Hong Kong, China

Student Volunteers Chairs

Mei Lu
Guizhou University, China

Weihua Ou
Guizhou Normal University, China

Patron and Sponsorship Chairs

Shengke Zeng
Xihua University, China

Zhanglin Cheng
Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Treasures

Yadong Wu
Sichuan University of Science & Engineering, China

Chunlin He
China West Normal University, China

Web Chairs

Shibin Zhang
Chengdu University of Information Technology, China

Guangzhu Chen
Chengdu University of Technology, China

Registration Chairs

Jinlong Shi
Jiangsu University of Science and Technology, China

Yunbo Rao
University of Electronic Science and Technology of China, China

Demo/Posters Chairs

Yue Zhao
Minzu University of China, China

Haisheng Li
Guangxi Normal University, China

Technical Committee Members

Tinghong Gao, Guizhou University, China
Lei Zhao, Tongji University, China
Rui Shi, Sichuan University of Science & Engineering, China
Yang Yang, Nanjing University of Science and Technology, China
Shaoning Zeng, Yangtze Delta Region Institute (Huzhou), UESTC, China
Lu Gan, Yangtze Delta Region Institute (Huzhou), UESTC, China
Hairui Yang, Yangtze Delta Region Institute (Huzhou), UESTC, China
Wei Pan, China West Normal University, China
Xiao He, China West Normal University, China
Mingwei Tang, Xihua University, China
Yuanyuan Huang, Chengdu University of Information Technology, China
Taisong Xiong, Chengdu University of Information Technology, China
Wu He, Sichuan Normal University, China
Zhengcai Wang, GuiZhou Institute of Light Industry, China
Ji Zeng, Panzhihua University, China
Prof. Haosong Gou, China Mobile Communications Group Sichuan Co., Ltd., China
Zairong Wang, Neijiang Normal University, China
Fiasam Linda Delali, Chengdu University, China
Razafindratovolahy Annicet, University of Electronic Science and Technology of China, China
Huarong Chen, Southwest University of Science and Technology, China
Cheng Huang, UT Southwestern Medical Center, USA
Qingsong Lv, Zhejiang University, China
Mutale Albert, University of Electronic Science and Technology of China, China
Junjun Wu, , Harbin Institute of Technology, China
Ya Zhang, YiBin Vocational And Technical College, China
Qixue He, University of Chinese Academy of Sciences, China
Yanhua Shao, Southwest University of Science and Technology, China
Hongying Zhang, Southwest University of Science and Technology, China
Hongyu Zhang, University of Electronic Science and Technology of China, China
Chuangquan Chen, Wuyi University, China
Fanyu Zeng, Nanjing University of Posts and Telecommunications, China
Escobañez, John Clement Suladay, National University - Manila, Philippines
Duoqian Miao, Tongji University, China
Luis Montesinos, Tecnologico de Monterrey, Mexico
Yang Liu, Xi'an Jiaotong-Liverpool University, China
Chuanlin Liu, Southwest Petroleum University, China
Julie Anne Angeles-Crystal, National University - Manila, Philippines
Hossam A. Gabbar, University of Ontario Institute of Technology, Canada
Shu Wang, Xi'an Microelectronics Technology Institute (of CASC), China
Yunjun Ruan, Guizhou Normal University, China
Bo Peng, Southwest Petroleum University, China
Ziyi You, Guizhou Normal University, China
Yuling Chen, Guizhou University, China
Yuan-Kai Wang, Fu Jen Catholic University
Jiajun Ma, Guizhou University, China
Fangfang Zeng, Guizhou University, China
Yinghao Ye, Guizhou University, China

Xi He, The Chinese University of Hong Kong, China
Dong Kang, Qingdao City University, China
Xiaolu Cheng, Suzhou University of Technology, China
Xiwen Zhang, Beijing Language and Culture University, China
Shaomin Wu, University of Kent, UK
Parames Chutima, Chulalongkorn University, Thailand
Marcin Paprzycki, Polish Academy of Sciences, Poland
Chuan Qin, University of Shanghai for Science and Technology, China
Zichao Li, University of Waterloo, Canada
Cheng Peng, University of Electronic Science and Technology of China, Zhongshan Institute, China
Jin Xie, Chinese Academy of Sciences, China
Kazuya Ueki, Meisei University, Japan
Teng Li, Shandong University, China
Zhanhong Du, Chinese Academy of Sciences, China
Chao Shen, Xi'an Technological University, China
Honggui Li, Yangzhou University, China
Xin Wang, China University of Petroleum-Beijing at Karamay, China
Zhonghua Xie, Huizhou University, China
Happy Nkanta Monday, ChengDu University of Technology, China
Koorosh Gharehbaghi, RMIT University, Australia
Bin Zheng, Panzhihua University, China
Guan Sihai, Southwest Minzu University, China
Fangyan Nie, Guizhou University of Commerce, China
Hong Li, Beihang University, China
Kunta Hsieh, Fuzhou University of International Studies and Trade, China
Hongwei Yin, Huzhou University, China
Hongxia Liu, Chizhou Vocational and Technical College, China
Peiyao Wang, Chengdu Jincheng College, China
Shichao Wu, Southwest University of Science and Technology, China
Genjian Yang, CEPREI, China
Yuxia Sheng, Wuhan University of Science and Technology, China
Yuanyuan Zhang, Unit 91475 of PLA, China
Li Xin, Chengdu Jincheng College, China
Yuejiao Wang, Xi'an Microelectronics Technology Institute, China
Haiqing Zhang, Chengdu Jincheng College, China
Dan GUO, Guizhou University, China

Onsite Conference Information

Conference Venue

贵州大学 *Guizhou University, China*

August 15, 2025
Venue: Guizhou Tobacco Training Centerr
地址：贵州烟草培训中心
August 16, 2025
Venue: Chongshi Building, Mingjun Building, West Campus of Guizhou University, Huaxi Avenue, Huaxi District, Guiyang, China
地址：贵州省贵阳市花溪区花溪大道贵州大学西校区崇实楼, 明俊楼
August 17, 2025
Venue: Mingjun Building, West Campus of Guizhou University, Huaxi Avenue, Huaxi District, Guiyang, China
地址：贵州省贵阳市花溪区花溪大道贵州大学西校区明俊楼

Transportation

From Guiyang Longdongbao International Airport

| By Bus (1 hour 10 mins, 4 RMB)

Take Bus Line 254 (toward Ganyintang) at the airport, ride 18 stops to **Huaxi Avenue South Station**. Transfer to Bus Line 210 (toward Qingyan Ancient Town), ride 3 stops to **Guizhou University West Campus Station**. Walk 100 meters east to enter the campus.

| By Taxi (45 mins, 90–110 RMB)

Distance is about 32 km via Jiaxiu South Road. Directly reach the west campus main gate.

From Guiyang Railway Station

| By Bus Line 210 (1 hour 5 mins, 3 RMB)

Walk 400 meters northwest from the station to **Railway Station (Platform 2)**. Take Bus Line 210 (toward Qingyan Ancient Town), ride 16 stops to **Guizhou University West Campus Station**. Walk 100 meters east to the campus.

| By Taxi (35 mins, 35–45 RMB)

Distance is approximately 17 km via Jiaxiu South Road.

From Guiyang North Railway Station

| By Metro & Bus (1 hour 30 mins)

Take Metro Line 1 (toward Xiaohe Industrial Park) from the station, ride 5 stops to **Zhonghua South Road Station**. Exit from Exit B, walk 300 meters to **Zhonghua South Road Bus Stop**. Take Bus Line 210 (toward Qingyan Ancient Town), ride 14 stops to **Guizhou University West Campus Station**. Walk 100 meters east.

| By Taxi (45 mins, 70–90 RMB)

Distance is about 28 km via Jiaxiu South Road.

From Guiyang East Railway Station

| By Metro & Bus (1 hour 40 mins)

Take Metro Line 2 (toward Jinyang South Railway Station) from the station, ride 10 stops to **Pen Shui Chi Station**. Transfer to Metro Line 1 (toward Xiaohe Industrial Park), ride 3 stops to **Zhonghua South Road Station**. Exit from Exit B, walk 300 meters to **Zhonghua South Road Bus Stop**. Take Bus Line 210, ride 14 stops to **Guizhou University West Campus Station**. Walk 100 meters east.

| By Taxi (55 mins, 110–130 RMB)

Distance is approximately 38 km via Jiaxiu South Road.

From Guiyang South Railway Station

| By Bus & Taxi (50 mins, 20–25 RMB)

Walk 500 meters north from the station to **Guiyang South Railway Station Bus Stop**. Take Bus Line 250 (toward Huaxi Park), ride 8 stops to **Huaxi Avenue Middle Station**. Take a taxi from there (5 mins, 10–15 RMB) to Guizhou University West Campus.

| By Direct Taxi (40 mins, 50–60 RMB)

Distance is about 20 km via Jiaxiu South Road.

Key Notes

- **West Campus Location**: Main entrance near the intersection of Jiaxiu South Road and Xueshi West Road, Huaxi District.
- **Campus Shuttles**: Free shuttles connect West Campus with East/South Campuses every 20 minutes (7:00–22:00).

- **Bus Updates**: Check real-time bus arrivals via the "Guiyang Bus" app for Line 210/254/250.

Oral Presentation Tips

✓ The duration of a presentation slot is 15 minutes. Please prepare your presentation for about 12 minutes plus about 3 minutes for questions from the audience;
✓ An LCD projector & computer will be available in every session room for regular presentations;
✓ Presentations MUST be uploaded at the computer at least 15 minutes before the session start.

Poster Presentation Tips

✓ We expect that at least one author stands by the poster for (most of the time of) the duration of the poster session. This is essential to present your work to anyone interest into it.
✓ If you wish to leave your poster to attend a lecture, please leave a message on the board to inform visitors.
✓ Posters should be set-up at least 15 minutes before the session starts and removed at the end of the session. Left behind posters at the end of the session will be disposed of.
✓ The size of poster is 80cm x180cm, Posters must be in portrait format(height>width). This cannot be modified.
✓ Materials to hang the poster will be made available on-site. Our staff will be available onsite to assist you.

Dress Code

✓ All participants are kindly requested to dress formally, as casual wear is discouraged. National formal dress is welcome.

Emergency

✓ Ambulance: 120
✓ Police: 110

Attention Please

✓ Please ensure the safety of your belongings in public areas. For personal and property security, delegates are advised to wear their identification badges during the conference and refrain from lending them to unauthorized individuals. The conference cannot be held responsible for the loss of personal items.
✓ Scan the QR code and send “PRAI 2025-Paper ID” to add conference assistant Wechat.

✓ WeChat QR Code

✓ Live Photo Streaming QR Code

Online Conference Information

Zoom Number	Date	Arrangement
Room A 827 2618 8431	August 15, 16, 17	**Zoom Test for Committee/Session** Chairs/ Speakers Opening Ceremony & Keynote Speech
		Zoom Test Online Sessions 1 & 2 **Online Session** Online Session Special Session 3
Room B 832 0180 3468	August 17	**Zoom Test** Online Sessions 3 & 4 **Online Session** Online Session 1
Room C 849 3724 5301	August 17	**Online Session** Online Session 3
Room D 840 5310 5661	August 17	**Online Session** Online Session 4

Passcode for all ZOOM room: 081517

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.

Platform: Zoom
✓ For General Users
https://zoomus/support/download
✓ For Users from mainland China
https://wwwzoom.com.cn/download

Sign in and Join
✓ Join a meeting without signing in: A Zoom account is not required if you join a meeting as a participant, but you cannot change the virtual background or edit the profile picture.
✓ Sign in with a Zoom account: All the functions are available.

Additional Suggestions
✓ A computer with an internet connection (wired connection recommended)
✓ USB plug-in headset with a microphone (recommended for optimal audio quality)
✓ Webcam (optional): built-in or USB plug-in
✓ Stable Internet Connection
✓ Quiet environment
✓ Proper lighting

Presentation Tips
✓ Each presentation slot is 15 minutes. Please prepare to speak for around 12 minutes, allowing 3 minutes for audience questions.
✓ Join the meeting room at least 15 minutes before the session begins.

Daily Schedule

August 15, 2025 | Friday

For Onsite Participants

Venue: Guizhou Tobacco Training Center, Guizhou University | 1st Floor

地址：贵州烟草培训中心, 贵州大学 | 1 楼

10:00-17:00	现场会议签到&资料领取 Sign in and Collect Conference Materials in the Lobby
14:00-17:00	学术参观 - 国家大数据(贵州)综合试验区展示中心 Academic Visit
18:00-20:00	Dinner 烟草培训中心餐厅-2F

For Online Participants

Passcode: 081517

10:00-12:00	ZOOM Test for Committee / Session Chairs / Speakers	**Zoom A:** 827 2618 8431
14:00-16:00	ZOOM Test for Online Session 1 & 2	**Zoom A:** 827 2618 8431
	ZOOM Test for Online Session 3 & 4	**Zoom B:** 832 0180 3468

August 16, 2025 | Saturday

Venue: Guizhou University | Chongshi Building Room 110

地址：贵州大学 | 崇实楼 110

Online Room A:827 2618 8431 (passcode: 081517)

https://us02web.zoom.us/j/82726188431

Opening Ceremony	
09:00-09:05	Welcome Message
09:05-09:10	Opening Remarks
Keynote Speech	
09:10-09:55	Keynote Speech I: **Di Li**, Tsinghua University, China *Speech Title: The Commensal Radio Astronomy FAST Survey (CRAFTS) and high-throughput data analysis*
09:55-10:40	Keynote Speech II: **Riqing Chen**, Fujian Agriculture and Forestry University, China *Speech Title: Research on Auxiliary Diagnostic Approaches for Neurodevelopmental Disorders Utilizing Spatiotemporal Data Modeling*
10:40-11:05	Group Photo & Break Time
11:05-11:50	Keynote Speech III: **Naveed Ahmad,** Prince Sultan University, Riyadh, Saudi Arabia *Speech Title: AI Based Defense Against OTP Flooding Attacks in Digital Authentication Systems*
11:50-14:00	Lunch 烟草培训中心餐厅-2F

Onsite Author Sessions Venue: Mingjun Building 地址：明俊楼		
14:00-15:15	**Special Session 3-1** Intelligent Data Analysis and Cyber Security Session Chair: **Yunbo Rao**, University of Electronic Science and Technology of China, China PA498, PA668, PA682, PA685-A, PA712	Room 128 1F
14:00-15:15	**Onsite Session 3** Digital Image Analysis and Intelligent Imaging Technology Session Chair: **Chuan Zhao**, Chengdu University of Technology, China PA329, PA695, PA706, PA513, PA699	Room 129 1F
14:00-15:30	**Onsite Session 4** Medical Image Analysis and Processing Methods Session Chair: **Rize Jin,** Tiangong University, China PA463-A, PA350, PA640-A, PA677-A, PA651-A, PA642-A	Room 130 1F
14:00-15:35	**Onsite Session 5** Image-Based Intelligent Detection Technology and Applications Session Chair: **Haoxuan Yang**, Guizhou University, China Zhipan Wu, PA676-A, PA222, PA220, PA480, PA686	Room 131 1F
14:00-15:30	**Poster Session 1** Vision-Based Intelligent Digital Image Processing Technologies and Engineering Applications Session Chair: **Shichao Wu,** Southwest University of Science and Technology, China PA333, PA723, PA348, PA334, PA528, PA226, PA206, PA717, PA639, PA218, PA332, PA654, PA464, PA672, PA343, PA331, PA522, PA664	Room 132 1F
15:30-16:15	Break Time	
16:15-17:50	**Onsite Session 6** Speech/Signal Detection and Measurement Analysis Session Chair: **Mingwei Cao**, Anhui University, China Jin Xie, PA221-A, PA696, PA357, PA674, PA475	Room 129 1F
16:15-17:35	**Onsite Session 7** AI-based Intelligent Information System Design and Data Computing Session Chair: **Zhipan Wu,** Huizhou University, China Chuan Zhao, PA225, PA477, PA353, PA669	Room 130 1F

16:15-17:35	**Onsite Session 8** Natural Language Processing and Information Retrieval Session Chair: **Jiaju Wu,** Institute of Computer Application China Academy of Engineering Physics, China Zhiyong Qiu, PA502, PA347-A, PA511, PA520	Room 131 1F
16:15-18:00	**Poster Session 2** AI-driven Multimodal Intelligent Interactive System Design and Data Computation Session Chair: **Fangyan Nie**, Guizhou University of Commerce, China PA478, PA720, PA481, PA700, PA509, PA724, PA709, PA675, PA514, PA638, PA472, PA219, PA656, PA355, PA352, PA490, PA466, PA708. PA718, PA667	Room 132 1F
18:30-20:30	Banquet 侗家食府侗家工业餐厅店 – 田园南路板桥艺术村	

August 17, 2025 | Sunday		
09:00-12:00	线下论坛 *Onsite Forum*	崇实楼 110 1楼 Chongshi Building Room 110 1F
Online Keynote Speech Online Room A:827 2618 8431 (passcode: 081517) https://us02web.zoom.us/j/82726188431		
10:00-10:45	Keynote Speech IV: **Junchi Yan**, Shanghai Jiao Tong University, China *Speech Title: Solving and Representation for Graph & Combinatorial Problem Instances*	
10:45-11:00	Break Time	
11:00-11:45	Keynote Speech V: **Hamid Reza Karimi,** Politecnico di Milano, Italy *Speech Title: Intelligent Soft Sensing and Prediction for Industrial Applications*	
11:45-14:00	Lunch - 烟草培训中心餐厅-2F	
Onsite Sessions Venue: Mingjun Building 地址：明俊楼		
Time	Activity	Room
14:00-16:15	**Special Session 2** Artificial Intelligence for Aerospace Applications Session Chair: **Ma Zhong**, Xi'an Microelectronics Technology Institute, China PA209-A, PA713, PA1005, PA517, PA665, PA681, PA633, PA725, PA679	Room 128 1F
14:00-15:45	**Onsite Session 9** Image Processing, Computer Vision and Intelligent Sensing Technologies Session Chair: **Yan Yan**, University of Guelph, Canada PA213, PA501, PA526, PA697, PA698, PA504, PA634	Room 129 1F
14:00-15:45	**Onsite Session 10** Multimodal Perception, Audio Processing and Cross-Domain Applications Session Chair: **Hangfei Zhang,** National University of Defense Technology, China PA729, PA346, PA691, PA678, PA728, PA711, PA730-A	Room 130 1F

Online Sessions		
Password: 081517		
13:00- 17:00	**Special Session 3-2** Pattern Recognition and Machine Learning Session Chair: **Binhui Tang**, Hainan Normal University, China PA707, PA690, PA646, PA716, PA689, PA508, PA228, PA692, PA721, PA722, PA714, PA467, PA661	Zoom A: 827 2618 8431
13:00-16:45	**Online Session 1** Object Detection and Detection Models Session Chair: **John Clement S. Escobañez,** Polytechnic University of the Philippines, Philippines PA635, PA227, PA497, PA512, PA527, PA531, PA471, PA705, PA205, PA529, PA521, PA703	Zoom B: 832 0180 3468
13:00- 17:00	**Online Session 2** Language Models and Semantic Analysis Session Chair: PA643-A, PA687, PA644-A, PA645-A, PA701-A, PA673-A, PA496, PA516, PA684, PA204, PA693, PA224, PA653,	Zoom C 849 3724 5301
13:00-16:45	**Online Session 3** Deep Learning, Vision & Intelligent Optimization Applications Session Chair: PA462-A, PA479, PA208, PA663, PA704, PA476, PA217, PA494, PA216, PA344, A683, PA694	Zoom D: 840 5310 5661

Keynote Speaker

Di Li

Tsinghua University, China

Speech Time: 09:10-09:55, Saturday, August 16, 2025
Venue: **Guizhou University | Chongshi Building Room 110**
地址： 贵州大学, 崇实楼 110
Online Room A: 827 2618 8431 (passcode: 081517)

Dr. Li is a radio astronomer. He is a Chair Professor of Tsinghua University and the former Chief Scientist of FAST. He pioneered several observing and data analysis techniques, including HI Narrow Self-Absorption(HINSA) and a new inversion algorithm for solving the dust temperature distribution.These techniques facilitated important measurements of star forming regions , such as their formation time scale and interstellar magnetic field (as a Nature cover article). Dr. Li has led and/or made multiple significant discoveries,including the first detection of interstellar molecular oxygen and the first persistently active fast radio bursts(FRB), the largest FRB bursts set, etc. He is leading the Commensal Radio Astronomy FasT Survey(CRAFTS), which is the world's first commensal survey of HI, pulsars, and FRBs. CRAFTS has discovered more than 200 pulsars, more than 8 FRBs, released 5000 deg2 HI sky maps. He has published more than 350 peer-reviewed journal articles with more than 10000 citations. He won the 2024 Marcel Grossmann Award for “For his groundbreaking contributions to the scientific definition of the most sensitive radio telescope and his numerous innovations in characterizing the dynamic universe”. His numerous other awards and leadership roles include the 2005 National Research Council (US)Resident Research Fellowship based on “his outstanding research capabilities ... as a result of national competition”; the 2022 Distinguished Achievement Award of the Chinese Academy of Sciences(CAS); the 3rd National Innovation Award; the Steering Committee of Australia Telescope National Facility (ATNF); the Square Kilometer Array Science and Engineering Advisory Committee(SEAC); the advisory panel of the Breakthrough Listen initiative, etc.

Speech Title: The Commensal Radio Astronomy FAST Survey (CRAFTS) and high-throughput data analysis

The Five-hundred-meter Aperture Spherical radio Telescope (FAST) is currently the largest single-dish radio telescope in the world with unprecedent high sensitivity and survey speed. The Commensal Radio Astronomy FAST Survey (CRAFTS) is one of the FAST key science projects. Utilizing the world’s-first high cadence noise injection technology, CRAFTS aims to carry out the Galactic neutral hydrogen imaging, neutral hydrogen galaxy search, pulsar search, and fast radio burst (FRB) search simultaneously, doubling the survey efficiency of FAST. The CRAFTS survey faces big challenge from the data transportation, storage, and analysis for huge amount of data collected every day. We have built a high-throughput data real-time processing technology platform, and developed intelligent computational techniques to meet these new requests from FAST surveys.

Riqing Chen

Fujian Agriculture and Forestry University, China

Speech Time: 09:55-10:40, Saturday, August 16, 2025
Venue: **Guizhou University | Chongshi Building Room 110**
地址： 贵州大学, 崇实楼 110
Online Room A: 827 2618 8431 (passcode: 081517)

Chen Riqing, Professor/PhD Supervisor, winner of the National Talent Program, currently serves as the Vice President of Sanming University. With extensive experience in both industry and academia, he has long been engaged in research and development in the fields of the Internet of Things (IoT), cloud computing, information security, big data analytics, and applications. He has filed over 30 patents and published more than 50 academic papers.

He has led multiple national and provincial research projects, including grants from the National Natural Science Foundation of China, the Ministry of Agriculture's Major Agricultural Technology Extension Pilot Program, and Fujian Province's Modern Agriculture "Five New" Demonstration Project. His contributions have earned him several prestigious awards, such as the Fujian Provincial Science and Technology Progress Award, the China Overseas Chinese Contribution Award, and the Fujian Youth Science and Technology Award.

Speech Title: Research on Auxiliary Diagnostic Approaches for Neurodevelopmental Disorders Utilizing Spatiotemporal Data Modeling

Abstract: Despite advances in neuroscience, fundamental limitations persist in biomarker discovery, differential diagnosis, and personalized treatment of autism and other neurodevelopmental disorders. And the construction of its prevention and treatment system remains an urgent challenge in the global neuroscience field. This report explores the mutual feedback mechanism between the joint optimization of brain image registration and segmentation for the construction of a brain development map, which conforms to the time-varying laws of brain development. It focuses on the method of reconstructing brain functional links based on fMRI data, and conducts research on high-dimensional spatiotemporal data modeling for the diagnosis of brain developmental disorders and explores its clinical significance.

Naveed Ahmad

Prince Sultan University, Riyadh, Saudi Arabia

Speech Time: 11:05-11:50, Saturday, August 16, 2025
Venue: **Guizhou University | Chongshi Building Room 110**
地址： 贵州大学, 崇实楼 110
Online Room A: 827 2618 8431 (passcode: 081517)

Naveed Ahmad is arrayed with both Research & Development throughout his career. He is young and energetic, seeking to join an organization where he can contribute to uplifting the economic and social status of people through the intervention of emerging technologies. One aspect of his career is contributing to research by publishing journals, book chapters, and conference papers, as well as supervising PhD and MS students. The second aspect involves bringing in and managing R&D projects at the university to promote the culture of applied research. The third aspect is establishing international linkages with Turkey, China, and the UK for the mutual benefit of the countries in general and the university in particular. The fourth unique aspect is his involvement in consultancy projects with local and international funding agencies, corporate entities, and the government.

Speech Title: AI Based Defense Against OTP Flooding Attacks in Digital Authentication Systems

Abstract: Attackers have changed the way they attack to take advantage of One Time Passwords (OTPs), which are used more and more in digital authentication systems for two-factor authentication and identity verification. One such rising threat is OTP flooding, also known as OTP bombing. In this assault, attackers automate sending repeated OTP requests to overload systems, bother customers, or change how downstream services work (such SMS gateways and billing systems).
When faced with advanced botnets, proxy networks, or low-and-slow assault patterns, traditional countermeasures like CAPTCHA, rate-limiting, and static rules are not enough. These risks not only stop services from working, but they also cost businesses a lot of money and hurt their reputations, especially in the fields of finance, healthcare, and e-commerce.
This session will show an AI-based hybrid strategy that uses machine learning and behavioral analytics to find and stop OTP flooding threats in real time. The answer goes beyond static rate limits by using anomaly detection, supervised learning, and temporal user profiling to create a real-time risk rating engine.

Junchi Yan

Shanghai Jiao Tong University, China

Speech Time: 10:00-10:45, Sunday, August 17, 2025

Online Zoom A: 827 2618 8431 (passcode: 081517)

Junchi Yan is currently a Full Professor (tenured) with Department of Computer Sciencen and Engineering, Shanghai Jiao Tong University, Shanghai, China. Before that, he was a Senior Research Staff Member and Principal Scientist wit h IBM Research where he started his career since April 2011. He obtained the Ph.D. in Electronic Engineering from Shanghai Jiao Tong University. His main research interest is machine learning and its intersection with other areas e.g. operations research and quantum computing. He serves as Associate Editor for Pattern Recognition Journal and ACM Trans. on Probabilistic Machine Learning, and (Senior) Area Chair/Senior PC for ICML, NeurIPS, AAAI, IJCAI, CVPR, etc. He has published over 100 publications in top venues with over 10K google scholar citations.

Speech Title: Solving and representation for graph & combinatorial problem instances

Abstract: In this talk, I will discuss the development of machine learning for combinatorial optimization, not only in general methodology but also particularly generative models for AI4Opt. I will show that how the idea of diffusion models could be introduced to solve the notoriously hard combinatorial problems. I also give some prospective idea on the future of the research directions.

Hamid Reza Karimi

Politecnico di Milano, Italy

Speech Time: 11:00-11:45, Sunday, August 17, 2025

Online Zoom A: 827 2618 8431 (passcode: 081517)

Hamid Reza Karimi is Professor of Applied Mechanics with the Department of Mechanical Engineering, Politecnico di Milano, Milan, Italy and the Honorary Visiting Professor within the School of Computing & Engineering at the University of Huddersfield, UK. Prof. Karimi's original research and development achievements span a broad spectrum within the topic of automation/control systems, and intelligence systems with applications to complex systems such as wind turbines, vehicles, robotics and mechatronics. Prof. Karimi is an ordinary Member of *Academia Europaea (MAE)*, Honorary Academic Member of *National Academy of Sciences of Bolivia*, Distinguished Fellow of *the International Institute of Acoustics and Vibration (IIAV),* Fellow of The *International Society for Condition Monitoring* (ISCM), Fellow of the Asia-Pacific Artificial Intelligence Association (AAIA), Member of Agder Academy of Science and Letters and also a member of the IFAC Technical Committee on Mechatronic Systems, the IFAC Technical Committee on Robust Control, the IFAC Technical Committee on Automotive Control as well as member of the board of Directors of The International Institute of Acoustics and Vibration (IIAV). Prof. Karimi is the recipient of the 2021 BINDT CM Innovation Award, the Web of Science Highly Cited Researcher in Engineering, the 2020 IEEE Transactions on Circuits and Systems Guillemin-Cauer Best Paper Award, August-Wilhelm-Scheer Visiting Professorship Award, JSPS (Japan Society for the Promotion of Science) Research Award, and Alexander-von-Humboldt-Stiftung research Award, for instance. Prof. Karimi is currently the Editor-in-Chief of the Journal of Cyber-Physical Systems, Subject Editor, Technical Editor or Associate Editor for some international journals and Book Series Editor for Springer, CRC Press and Elsevier. He has also participated as General Chair, keynote/plenary speaker, distinguished speaker or program chair for several international conferences in the areas of Control Systems, Robotics and Mechatronics.

Speech Title: Intelligent Soft Sensing and Prediction for Industrial Applications

Abstract: In this speech, we explore innovative methodologies for efficient and cost-effective process optimization in industrial applications. Soft sensing refers to approximating hard-to-measure variables using easily measurable ones at a given instant, whereas prediction extends the focus to future horizons. This presentation highlights several novel soft sensing and prediction frameworks developed for two specific case studies: industrial aluminum electrolysis and undermining systems. These frameworks leverage semi-supervised and self-supervised learning to address few-shot labeled sample scenarios while maintaining performance integrity. The talk concludes with key insights and future directions.

Onsite Oral Sessions

SS03-1: Intelligent Data Analysis and Cyber Security

Chair: **Yunbo Rao**, University of Electronic Science and Technology of China, China

14:00-15:15 | Aug.16, 2025 | Mingjun Building Room 128

TAIK DETAILS	
Time	**Presentation**
14:00-14:15 **PA498**	Title: Vertically Partitioned High-dimensional Data Publish ing with Differential Privacy Author(s): Junming Zhang, Jingru Wang, Shigong Long, Shengguang Ba, Lun Wang Presenter: Junming Zhang and Shigong Long, Guizhou University, China Abstract: Ensuring data security and availability is crucial in high-dimensional data privacy releases, but the effectiveness of dimensionality reduction algorithms and the allocation of privacy budgets often pose challenges. Data usability encompasses accuracy, completeness, and association properties. To address these issues, the rough set-based personalized privacy-preserving Bayesian network algorithm (RSPPrivBayes) is proposed. This algorithm leverages the advantages of rough sets and Bayesian networks: first, it preprocesses the data using rough set theory to retain key correlations while eliminating redundancy for efficient dimensionality reduction; next, according to user privacy requirements, it differentially injects Laplace noise into the categorized data to balance privacy and data correlation; finally, it employs a Bayesian network to construct the released data, thereby protecting privacy while preserving dependency relationships and statistical properties. Experiments confirm that RSPPrivBayes performs well in both privacy protection and data quality, verifying its effectiveness and efficiency.
14:15-14:30 **PA668**	Title: Large-Scale Network Traffic Classification Using Graph Clustering-Enhanced E-GraphSAGE Author(s): Yuxin Ma, Jun Ma, Yan Fan Presenter: Yuxin Ma, Lanzhou University, China Abstract: Network traffic classification is a key technology for intrusion detection systems, but traditional methods face bottlenecks such as high computational complexity and large memory consumption when handling large-scale graph structured data. To address this challenge, this paper proposes the Cluster-E-GraphSAGE framework, which innovatively combines edge-aware graph neural networks with the METIS graph partitioning algorithm to achieve efficient processing of large-scale network traffic graphs through an efficient subgraph partitioning strategy and edge-aware message passing mechanism. The framework supports multi-directed graph structures, completely preserving edge multiplicity and feature attributes, and significantly reduces computational complexity and memory overhead through edge-based subgraph partitioning. Experiments on the standard NF-BoT-IoT and NF-ToN-IoT datasets show that Cluster-E-GraphSAGE performs particularly well in multi-class classification tasks, achieving F1 scores of 92.82% and 81.32% respectively, which are 11.23% and 23.6% improvements over baseline methods. Meanwhile, it reduces memory usage by 4-7 times and GPU memory by 8-11 times. The research results verify that the method maintains high classification accuracy while significantly reducing memory overhead in resource-constrained environments, demonstrating its application potential in real-time intrusion detection scenarios.

Title: One-Stage Mesh Denoising Method Based on Nonlinear Total Generalized Variation Framework

Author(s): Yupeng Wang, Huayan Zhang

Presenter: Yupeng Wang, Tiangong University, China

14:30-14:45
PA682

Abstract: In the paper, we introduce a new one-stage mesh denoising method based on the nonlinear total generalized variation (TGV) framework. The new framework contains three terms: the nonlinear TGV regularization term with gradient operator and nonlinear Laplace operator, a smoothing term, and a fidelity term. The non-differentiable nonlinear optimization problem is solved by an iterative algorithm based on operator splitting and the alternating direction method of multipliers (ADMM). Our denoising method is discussed and compared to several state-of-the-art techniques in terms of denoising results, quantitative comparison, and computational costs. Experimental results indicate that our approach is comparable to state-of-the-art algorithms at reasonable costs. It can produce denoising results with more structures and alleviate the staircase effect. The quantitative errors also verify that the newly proposed algorithm is robust numerically.

Title: Evolutionary Game Theory-Based Verifiable Rational Secret Sharing Scheme with Quantum Resistant

Author(s): Zhimei Yang, Changgen Peng

Presenter: China, Guizhou University, China

14:45-15:00
PA685-A

Abstract: To address the risks of collusion deception and quantum computing threats faced by traditional secret sharing schemes in a rational participant environment, this paper proposes a verifiable rational secret sharing scheme based on evolutionary game theory that is also quantum attack resistant. This scheme incorporates the dynamic stability idea of an evolutionary game into the rational cryptography framework. The scheme converges to the "cooperative reconstruction of the secret" evolutionarily stable strategy in repeated games, thereby inhibiting rational participants' short-term betrayal behavior. To accomplish verifiability, the approach incorporates a zero-knowledge proving process based on lattice cryptography, which allows players to check the authenticity of secret shares without disclosing any extra information. Simultaneously, the approach constructs the secret distribution and reconstruction protocol using post-quantum safe mathematical issues, assuring its long-term security from quantum assaults. This scheme is more resilient and fair in a dynamic rational context, and it fits the security needs of the post-quantum era, making it a viable option for distributed system key management, blockchain threshold signatures, and other applications.

Title: Analysis of User-Privacy, Third-Party Data Sharing and Consent Mechanism on Online Pharmacy Websites in Ontario

Author(s): Atikor Ibodeng, Zuhaibuddin Bhutto, Yan Yan, Sampsa Rauti, Ville Leppänen, Adegboola Adelabu, and Wenjun Lin

Presenter: Yan Yan, University of Guelph, Canada

15:00-15:15
PA712

Abstract: The purpose of this study is to evaluate the privacy practices of online pharmacy websites in Ontario, Canada, focusing on data sharing with third parties, privacy policies, and the effectiveness of consent mechanisms. To achieve this, we conducted a comprehensive analysis using both automatic and manual network traffic analysis methods. We analyzed the traffic flow of 109 pharmacy websites, assessing the presence of third-party services, the transmission of data to third parties, and the availability and clarity of privacy policies and consent forms. Our study indicates that a significant number of pharmacy websites in Ontario share sensitive health information with third-party entities, such as Google Analytics and Facebook, even without explicit user consent. Additionally, many of these websites lack comprehensive privacy policies and

effective consent mechanisms. We aim to create awareness of the limitations of third-party data sharing, the adoption of clearer and more accessible privacy policies, and robust consent mechanisms crucial for protecting the privacy of user data.

Onsite Oral Sessions

T03: Digital Image Analysis and Intelligent Imaging Technology

Chair: **Chuan Zhao**, Chengdu University of Technology, China

14:00-15:15 | Aug.16, 2025 | Mingjun Building Room 129

TAIK DETAILS	
Time	**Presentation**
14:00-14:15 **PA329**	Title: InStitch: An Incremental Approach to Stitching Low-overlap Images Author(s): Ning Li, Haifeng Zhao, Mingwei Cao Presenter: Ning Li, Anhui University, China Abstract: Image stitching is a hot research topic in the fields of computer vision and computer graphics. It is a technology that builds a panoramic image from image collections and is one of the most difficult tasks because the quality of panoramic images is often affected by many potential factors such as camera rotation, lighting, scale changes, and low-overlap ratio among inputted images captured by mobile cameras. In addition, how to visualize a high-resolution panoramic image is also a challenging issue in practical applications. To defend the above-mentioned problems, in this paper, we propose a novel method to stitch low-overlap images to produce a high-resolution panoramic image with a big field of view. The core of this method is stitching low-overlap images by using an incremental approach, called InStitich. The key components of InStitch consist of image matching, incremental registration, and image blending. Moreover, we also developed an end-to-end image stitching system depending on the proposed InStitch. The system can not only produce a high-quality panoramic image but also provide a user-friendly graphical user interface. We have conducted a comprehensive experiment on several benchmarks to evaluate InStitch and make a deep comparison with the state-of-the-art methods. Experimental results show that our method has a superior performance in terms of computational cost and quality of panoramic images.
14:15-14:30 **PA695**	Title: Primary Beam Effect Elimination for Mid-frequency SKA Imaging using a Uformer Network Author(s): Yuxin Xia, Li Zhang, Xinglong Wang, Zujuan Xu Presenter: Yuxin Xia, Guizhou University, China Abstract: As the next generation of radio telescopes, the Square Kilometre Array (SKA) has high resolution and sensitivity to support major scientific tasks such as deep space exploration. However, non-ideal responses in the primary beam can introduce imaging distortions, adversely affecting scientific outputs. Traditional elimination methods struggle to effectively handle the high complexity and dynamic variations inherent in the SKA. Therefore, this paper uses the Uformer model to eliminate the primary beam effect of Mid-frequency SKA. This paper is based on 1500 simulated data in the Galaxy Zoo, and after model training and testing, the experimental results show that the average index of PSNR, RMSE and SSIM of the reconstruction image are 37.87dB, 0.0108, 0.9438. Compared with the dirty image of the primary beam effect, the Uformer method improves the PSNR and SSIM by 18dB and 0.8642, respectively. The results show that the method can effectively improve the image quality, providing algorithm support for high-precision science of SKA.

Title: Observation Simulation and Mitigation of Polarization Effects in Mid-frequency SKA

Author(s): Yang Liu, Li Zhang, Chao Li, Huijie Yang, Jinbo Zhao

Presenter: Yang Liu, Guizhou University, China

14:30-14:45
PA706

Abstract: In modern radio astronomical research, the Square Kilometre Array (SKA) provides highly sensitive continuous measurements for radio interferometric imaging. However, polarization effects are generated during the process of radio interferometric imaging, which can affect the image performance. This paper first conducted simulated observations of polarization effects and analyzed them in detail in combination with the main parameter indicators and two-dimensional power spectral density. To eliminate these effects, this paper applies the DeblurGAN-v2 model to astronomical image reconstruction. The model uses a Feature Pyramid Network (FPN) module, a dual-scale discriminator, and various backbone networks to enhance the performance of astronomical image reconstruction. Experimental results show that, compared with traditional algorithms, the model performs remarkably well in recovering polarization effects and restoring image details. It verifies the unique advantages and application potential of the DeblurGAN-v2 model in the task of astronomical image reconstruction, providing a new technological approach to solving the problem of image quality in the field of radio astronomy.(compared to traditional algorithms, DeblurGAN-v2 improves PSNR by approximately 24 dB and SSIM by 25 %, confirming its effectiveness and offering a new technical pathway for solving image-quality challenges in radio astronomy.)

Title: Augmented Hybrid Cost Volume Network for Stereo Matching

Author(s): Chenglin Dai

Presenter: Chenglin Dai, Guangzhou Huali College, China

14:45-15:00
PA513

Abstract: As a classic estimation task, the Stereo Matching typically uses cost volume constructed by the left and right feature maps to estimate disparity. Although a recent study can make great progress, it is still difficult to achieve the goal of high precision due to the insufficient use of scale information, channel information, and the relevant feature map. Therefore, we firstly present the Two-Scale Hybrid Cost Volume Construction Module (THCVCM), which includes the Two-Scale Hybrid Cost Volume (THCV) to fully use the information of two-scale feature maps. And then we design the Residual-Block-Style Edge Aware Channel Attention Module (REACAM), which is the part of THCVCM applied to better utilize the channel-wise news from the low-scale feature maps. Finally, we adopt the Augmented Residual-Block-Style Attention Hourglass Module (ARAHM) to better aggregate the cost volume by introducing relevant feature maps. In conclusion, we build the Augmented Hybrid Cost Volume Network (AHCVNet) for end-to-end Stereo Matching. Compared to some state-of-the-art models, it is better and more competitive.

Title: Wide-field Effect Correction for Low-frequency SKA Imaging using a Multi-Stage Network

Author(s): Xinglong Wang, Li Zhang, Mei Lu, Zujuan Xu, Yuxin Xia

Presenter: Xinglong Wang, Guizhou University, China

15:00-15:15
PA699

Abstract: The wide-field effect observed in low-frequency Square Kilometre Array (SKA) observations is one of the key factors leading to image quality degradation, and represents a critical challenge in SKA data processing that urgently needs to be addressed. The target sources in this paper are selected from the Galaxy Zoo survey data. By simulating wide-field effects, the corresponding dirty images are generated to construct the dataset. The MPRNet deep learning model is then exploratively applied to the task of correcting wide-field effects in astronomical images. Experimental results demonstrate that the method achieves a PSNR of 45.31 dB and an SSIM of 0.992 on the test set, effectively recovering the source structures with superior restoration quality. This paper presents an initial exploration of deep learning methods for astronomical image

processing, and future work will focus on advancing the network to address more complex coupled degradation effects.

Onsite Oral Sessions

T04: Medical Image Analysis and Processing Methods

Chair: **Rize Jin,** Tiangong University, China

14:00-15:30 | Aug.16, 2025 | Mingjun Building Room 130

TAIK DETAILS	
Time	**Presentation**
14:00-14:15 PA463-A	Title: Three-Dimensional Reconstruction And Classification of Impacted Third Molars Author(s): Cheng Niu, Tingyuan Zeng, Zhenggong Han, Bin Wang, Zhengfen Li, Meiyan Rong, Wei Yang Presenter: Niu Cheng, Guizhou University, China Abstract: Background: The extraction of impacted third molars (ITMs) presents significant surgical challenges due to their morphological diversity, complex anatomical relationships, and proximity to critical structures such as the mandibular nerve canal. Precise preoperative planning is essential to minimize iatrogenic injuries and complications. While digital technologies have been widely adopted in implantology and orthodontics, their application in ITM surgery remains limited. This study introduces a novel approach utilizing three-dimensional (3D) reconstruction and classification methods to optimize ITM extraction. Methods: We established the first medium-sized CBCT dataset dedicated to impacted third molars, enabling segmentation and categorization. (1)Two deep learning models were developed for 3D segmentation and reconstruction of ITMs and adjacent structures.(2) Automated classification based on 3D angulation analysis. The principal axis angles between third and second molars were calculated, with optimal classification thresholds determined through statistical analysis. Results: The first model achieved high segmentation accuracy, with IoU=86.64% and Dice=96.34% when merging second and third molar labels. The second model demonstrated robust performance, with Dice scores of 90.01% (third molar), 87.86% (second molar), and 82.34% (mandibular nerve canal). Automated classification yielded accuracy=81.5%, precision=82.3%, and recall=82.3% at optimal angle thresholds of 25° and 75°. Conclusion: This study presents two deep learning models for precise 3D reconstruction of ITM and associated anatomical structures, along with a threshold-optimized 3D classification system. The framework provides surgeons with intuitive, high-fidelity preoperative visualization, potentially enhancing surgical planning and outcomes in ITM extraction.
14:15-14:30 PA350	Title: MUST: Fast Cardiac MR-to-US Image Conversion via Style Transfer Author(s): Yusong Wang, Bo Peng, Hongmei Zhang, Han Yang Presenter: Yusong Wang, Southwest Petroleum University, China Abstract: With the advancement of deep learning, medical image synthesis has shown great potential in improving ultrasound simulation training. However, traditional generation methods often rely on paired data or extensive annotations, which limits their practical application. To address this challenge, we propose a novel unsupervised image conversion model, MUST, to achieve fast conversion of Magnetic Resonance (MR) images to high-quality Ultrasound (US) images using only unpaired MR and US data. Built upon the CycleGAN framework, MUST incorporates a patch-based multi-layer contrastive learning strategy and integrates Coordinate Attention (CA) and Convolutional Block Attention Module (CBAM) to enhance feature extraction and

	improve image generation stability. Furthermore, we introduce a dedicated ultrasound style loss to reconstruct realistic speckle textures while preserving the structural integrity of the source MR images. Experiments on an unpaired MR-US cardiac image dataset show that the ultrasound images generated by MUST outperform several state-of-the-art unsupervised image-to-image translation methods in terms of structural fidelity, textural realism, and subjective visual quality, highlighting its effectiveness and potential in unsupervised ultrasound image synthesis.
14:30-14:45 **PA640-A**	Title: Deep Learning-Based Multiparametric Fusion of Diffusion MRI for Improved Prediction of Diabetic Kidney Disease Author(s): Jian Liu, Daoyu Yang, Rongpin Wang and Xianchun Zeng Presenter: Jian Liu, Guizhou University, China Background Diabetic kidney disease (DKD) is a severe complication of type 2 diabetes mellitus (T2DM), where early detection plays a critical role in preventing progression. Recent studies have demonstrated the potential of diffusion MRI (DMRI) in assessing DKD. However, effectively combining multiple DMRI parameters to enhance predictive accuracy remains a challenge. Therefore, this study introduces a novel deep learning model designed to integrate different DMRI parameter maps, improving the prediction of DKD in T2DM patients. Methods This study included 282 participants (84 normal controls [NC], 69 with T2DM, and 129 with DKD) between January 2022 and April 2024. All participants were imaged on a 1.5T MRI scanner by diffusion sequence. The diffusion data were analyzed by diffusion tensor model to produce diffusion parameter maps, which were then individually fused with the B0 image ($b = 0$ s/mm^2). The fusion process utilized both spatial and channel information, with a consistency loss function introduced to ensure proper alignment and enhance the robustness of the fusion process. Results The fusion model demonstrated strong performance with a macro-average ROC-AUC of 0.851 ± 0.029. The ROC-AUC values for NC, T2DM, and DKD were 0.886 ± 0.048, 0.795 ± 0.029, and 0.870 ± 0.029, respectively. Notably, the fusion model outperformed the best singleparameter model (B0), which achieved a macro-average ROC-AUC of 0.775 ± 0.052. Conclusions The findings suggest that the integration of multiple DMRI parameter maps with deep learning may serve as a reliable and non-invasive approach for the early detection and monitoring of DKD.
14:45-15:00 **PA677-A**	Title: Artificial Intelligence in Imaging for Obstructive Sleep Apnea: A Systematic Review Author(s): Xiaoxuan Zhang, Xianchun Zeng, Zhenliang Xiong, Daoyu Yang, Yiqi Wu, Rui Zhou Presenter: Xiaoxuan Zhang, Guizhou University Medical College, China Abstract: Obstructive Sleep Apnea (OSA) is a prevalent sleep disorder characterized by recurrent upper airway collapse, leading to significant health risks if undiagnosed or untreated. Polysomnography (PSG), regarded as the gold standard for diagnosis, is complex and costly, thereby limiting its availability. Using artificial intelligence (AI) to analyze medical imaging offers a lower cost solution for large-scale screening, disease severity prediction, and the identification of potential disease phenotypes. This review comprehensively evaluates AI applications across diverse imaging modalities, including computed tomography (CT), cone-beam computed tomography (CBCT), magnetic resonance imaging (MRI), ultrasound, cephalometry, 2D/3D facial imaging, drug-induced sleep endoscopy (DISE), and positron emission tomography (PET). We systematically compare the technical characteristics, clinical applications, and potential for AI integration of these imaging modalities. CT and CBCT provide precise morphological evaluations, whereas MRI offers detailed anatomical and neurological information. Dynamic imaging techniques such as cine MRI, ultrasound, and DISE enable visualization of functional collapse mechanisms. In addition, facial imaging allows non-invasive insights into structural features, while PET reveals

	information at the molecular level. Across modalities, AI improves segmentation, prediction, and classification performance, often outperforming traditional methods. We propose a tiered decision-making framework to guide optimal imaging selection for screening, diagnosis, surgical planning, and research. Despite rapid advancements, challenges persist in dataset standardization, model generalizability, and clinical adoption. Future research should prioritize large scale, multi-center studies, multimodal data fusion, and explainable AI to bridge the gap between research and clinical practice, ultimately improving the diagnosis and treatment of OSA.
15:00-15:15 **PA651-A**	Title: Multi-Scale Dynamic Graph Attention Fusion Network for Obstructive Sleep Apnea Screening Using Resting-State fMRI Author(s): Zhenliang Xiong, Yinglin Zhou, Yan Gong, Rongpin Wang, Xianchun Zeng Presenter: Zhenliang Xiong, Guizhou University, China Abstract: Obstructive Sleep Apnea (OSA) is a prevalent sleep disorder with significant health consequences if undiagnosed. Resting-state functional magnetic resonance imaging (rs-fMRI) provides a non-invasive window into brain functional connectivity patterns, while graph neural networks (GNNs) possess remarkable potential for modeling complex brain network interactions. Previous studies have primarily focused on static single-scale connectivity analysis, lacking both dynamic temporal information integration and multi-scale feature fusion. To address these limitations, we propose a novel Multi-Scale Dynamic Graph Attention Fusion model (MSDGAF) that incorporates: (1) adaptive graph attention modules for dynamic connectivity learning, (2) temporal graph convolution with memory mechanisms, and (3) cross-scale feature alignment between AAL and CC200 brain atlases. We evaluated our framework on rs-fMRI data from 181 subjects (101 OSA patients and 80 healthy controls) collected at a single medical center. Experimental results demonstrate that MSDGAF significantly outperforms comparative models in OSA screening accuracy. The model's superior performance stems from its ability to capture both temporal dynamics and multi-scale spatial patterns in functional brain networks, offering a promising tool for clinical OSA diagnosis.
15:15-15:30 **PA642-A**	Title: An Optimized YOLOv11 Model for Key Bone Region Extraction in Hand X-rays for Pediatric Bone Age Assessment Author(s): Yiqi Wu, Xianchun Zeng, Rui Zhou, Daoyu Yang, Jian Liu and Zhenliang Xiong Presenter: Yiqi Wu, Guizhou University, China Abstract: Bone age assessment based on hand X-ray images provides an accurate determination of adolescents' actual bone age. For precise bone age evaluation, it is crucial to accurately extract the key skeletal regions related to bone age from these X-ray images. Given the uncertain proportions and irregular shapes of these regions, errors in judgment and suboptimal extraction accuracy can significantly reduce the precision of the bone age assessment. This paper proposes the YOLO-RDE model for extracting key skeletal regions from hand X-ray images. The model optimizes the YOLOv11 architecture by incorporating re-parameterized convolutions, a deformable attention module, and Efficient Intersection over Union (EIoU) loss. A dataset of 2, 492 hand X-ray images was used to train and validate the proposed model. The results demonstrate that YOLO-RDE achieves outstanding performance with a mAP@0.90 of 0.909 and a mAP@0.5:0.95 of 0.923. The average detection accuracy of all key skeletal regions reaches 99.9%, outperforming existing models in this task. Moreover, YOLO-RDE is the fastest model among those compared, achieving a Frames Per second (FPS) of 92. This model is capable of accurately extracting the key skeletal regions essential for bone age evaluation while minimizing misjudgments between similar skeletal regions, thereby providing more precise and efficient support for bone age assessment.

Onsite Oral Sessions

T05: Image-Based Intelligent Detection Technology and Applications

Chair: **Haoxuan Yang**, Guizhou University, China

14:00-15:35 | Aug.16, 2025 | Mingjun Building Room 131

Invited Speaker

Zhipan Wu | 14:00-14:20

Huizhou University, China

Title: An Intelligent Detection Method for Lake Water Environments: Identification and Research of Plastic Bottles Based on the YOLO Series

Abstract: As global plastic pollution intensifies, freshwater resources such as lakes face serious threats from plastic bottles and other waste, significantly affecting water quality and ecosystem health. To effectively monitor and manage plastic bottle pollution in lake environments, this paper proposes an intelligent detection method based on YOLOv11. The method innovatively utilizes ChatGPT's DALL·E technology to generate a dataset of plastic bottle images tailored specifically for lake scenarios and integrates it with the deep learning-based object detection algorithm YOLOv11 to achieve real-time identification and localization of plastic bottles. Experimental results demonstrate that the YOLOv11 model exhibits excellent performance in detection accuracy, real-time processing, and adaptability to complex environments, achieving an identification accuracy of 95.6%, significantly reducing false positives and false negatives. Through comprehensive testing and optimization, YOLOv11 has shown strong robustness under various lighting conditions and complex backgrounds, making it particularly suitable for deployment in remote and resource-constrained areas. This research not only provides novel technical support for monitoring water pollution but also offers innovative solutions for future intelligent environmental protection practices.

Time	Presentation
14:20-14:35 **PA676-A**	Title: Emerging privacy computing schemes: challenges and future of inner product functional encryption Author(s): Haoxuan Yang, Changgen Peng Presenter: Haoxuan Yang, Guizhou University, China Abstract: Inner product functional encryption (IPFE), as a novel public key cryptosystem, has gained considerable attention from researchers due to its unique privacy-preserving computation capabilities and robust security features. However, beneath this seemingly promising privacy-preserving ability lie some potential design flaws, which directly lead to unavoidable security issues. Currently, there are no relatively efficient solutions to address these problems. In response to these potential security concerns, we explore them from the perspectives of attributes and identities. We propose targeted solutions, conducting a comprehensive exploration of IPFE from both formal proof and practical application security viewpoints, aiming for an all-round enhancement of the IPFE scheme.
14:35-14:50 **PA222**	Title: Improved YOLOv8n-based algorithm for road object detection in foggy weather Author(s): Wenjie Huang, Zhenyu Chen, Dan Zhang, Fengmin Peng, Mingshen Huang, Guojin Xiao Presenter: Wenjie Huang, Huizhou University, China Abstract: Due to the limited feature representation capability of the YOLOv8n model under complex weather conditions leading to decreased detection accuracy, this paper proposes a novel Pyramid

Pooling Attention with Low-rank Convolution (PPA LC) module based on YOLOv8n. Specifically, we first designed an attention mechanism based on multi-scale pooling, which captures cross-regional information through a hierarchical fusion strategy. This attention mechanism enhances feature rep_x0002_resentation capability and the ability to capture blurred contour features while only adding 0.018M parameters. Additionally, “PPA LC’ adopts the concept of low-rank decomposition to reconstruct standard 7×7 convolution into 1×7 and 7×1 spatially separable convolutions. This maintains a large receptive field while significantly reducing parameters. The PPA LC module employs an embedded design that can seamlessly replace the C2f module in the YOLOv8n structure while maintaining the original network’s computational cost (GFLOPs increased by only 0.8% compared to the original YOLOv8n model). Results show that theoverall accuracy reaches 88.3% (mAP50) and 70.8% (mAP50-95) under two performance evaluation metrics, showing improvement compared to the unmodified YOLOv8n.

14:50-15:05

PA220

Title: A Lightweight YOLOv12n Object Detection Network for Underwater Degraded Environments

Author(s): Fengmin Peng, Qiong Gao, Dan Zhang, Guojin Xiao, Wenjie Huang, Mingshen Huang, Weihao Yang

Presenter: Fengmin Peng, Huizhou University, China

Abstract: Due to the characteristics of light propagation un_x0002_derwater, underwater object detection models face challenges such as blurred object edges that are easily confused with complex backgrounds and difficulties in detecting small objects, which leaves much room for improvement in accuracy. This paper proposes the Convolution-Depthwise Separable Convolution-Max Pooling (CDM) module and develops the enhanced YOLOv12n_x0002_CDM network based on the YOLOv12n architecture.By replacing the C3K2 module in the shallow layers of the original network with the CDM module, YOLOv12n-CDM addresses the limita_x0002_tions of C3K2 in distinguishing blurred object edges from back_x0002_ground interference.The CDM module uses standard convolution and depthwise separable convolution for basic feature extraction and enhancing generalization ability, utilizes max pooling to filter out regions where edges and backgrounds are confused, focuses on prominent object features, and introduces residual connections to retain key features while improving model robustness.This module also enhances attention to small objects.Experimental results show that YOLOv12n-CDM achieves mAP@.5 and mAP@.5: metrics.On the DUO public dataset, it scored 84.4% and 64.8% respectively, which are 0.8% and 1.2% higher than the baseline model.The model has 2.65 million parameters and a computational complexity of 8.2 GFLOPs, maintaining efficient detection performance while effectively controlling complexity.

14:05-15:20

PA480

Title: Collaborative Anomaly Detection Using Agent-Based Modeling in Smart Manufacturing

Author(s): Shuai Yuan, Haoxu Nong, Yunbo Rao

Presenter: Shuai Yuan, University of Electronic Science and Technology of China, China

Abstract: Industrial anomaly detection is critical for ensuring product quality, preventing equipment failure, and maintaining efficient manufacturing processes. Traditional methods often struggle with the diversity and complexity of potential anomalies and require extensive labeled data for specific tasks. Recent advancements in large pre-trained vision models and large language models, such as Grounding DINO for flexible object detection and Segment Anything Model V2 (SAM V2) for precise segmentation and Llama 3 for large corpus, offer new possibilities for more adaptable anomaly detection systems. This paper proposes an agent-based framework that leverages the power of Grounding DINO, SAM V2, and Llama 3 for zero-shot industrial anomaly detection. Our method employs an intelligent agent to first localize potential anomalies using text prompts with Grounding DINO. These prompts are carefully designed to highlight common industrial anomalies to guide the model in detecting specific deviations from the norm. The identified regions are then segmented

precisely using SAM V2, which generates pixel-level masks based on the bounding boxes detected in the localization step. Once the anomaly regions are segmented, the agent generates an anomaly mask that highlights the anomalous regions. Finally, the agent visualizes the anomaly detection results by overlaying the mask onto the original image, allowing human inspectors to identify and address potential defects. We evaluate our approach on the MVTec and VisA anomaly detection dataset. Compared with the mainstream methods, our method has at least increased 1.9% on accuracy, demonstrating its capability to effectively localize and segment various types of industrial anomalies across different object categories without requiring task-specific training data, thus significantly reducing costs.

15:20-15:35
PA686

Title: An End-to-End Multimodal Framework for AI-Driven Educational Video Generation

Author(s): Henghui Meng, Kainan Liu, Chengxiang Li, Yifan Li, Da Shang, Zihao Jia

Presenter: Henghui Meng, University of Sanya, China

Abstract: Online education platforms often lack personalized learning experiences and effective emotional interaction, partly because existing technologies struggle to generate audio and video synchronously, leading to inconsistencies in semantics and emotional expression. To address this challenge, this paper introduces Etutor, an innovative, end-to-end multimodal AI framework for educational video generation that can rapidly create high-quality, personalized instructional videos from a single text input. At the core of Etutor is a three-layer architecture-encompassing semantics, expression, and technology-coordinated by a Large Language Model (LLM). This architecture uniformly interprets user intent, generates instructional content, and orchestrates all modal parameters to ensure a high degree of semantic and emotional consistency. The framework integrates two key technical modules: NMCTTS, a speech synthesis system optimized for Science, Technology, Engineering, and Mathematics (STEM) content that accurately handles professional terminology and formulas; and CF-RTVideo, a video generation system that significantly enhances the clarity of facial details, the vividness of expressions, and the preservation of identity. To comprehensively evaluate system performance, we introduced the Teaching Cognitive Load Synergy Index (TCLSI) as a new evaluation metric. Experimental results show that Etutor achieved a high score of 0.78 on the TCLSI, outperforming all baseline systems that use separate modular components. This result demonstrates Etutor's comprehensive advantage in balancing instructional content delivery, multimodal expressiveness, and learner cognitive load, offering an effective automated solution to the challenges in modern online education.

Onsite Oral Sessions

T06: Speech/signal Detection and Measurement Analysis

Chair: **Mingwei Cao**, Anhui University, China

16:15-17:50 | Aug.16, 2025 | Mingjun Building Room 129

Invited Speaker

Jin Xie | 16:15-16:35

Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China

Title: Precise Analysis of Brain Diseases and Construction of Brain-Controlled Systems by Integrating Artificial Intelligence and Multimodal Signals

Abstract: This study aims to explore the potential value of multimodal neurobiological data through artificial intelligence technologies, thereby providing innovative solutions for the diagnosis, treatment, and neurological function rehabilitation of brain diseases. Traditional diagnosis and treatment methods for brain diseases have limitations in terms of precision and personalization, making it urgent to break through the bottlenecks by means of multimodal data analysis and intelligent algorithms. In the aspect of precise analysis of brain diseases, for major depressive disorder, we constructed a new deep learning framework (MIL-Transformer) to accurately predict the therapeutic effect of antidepressant drugs; for epilepsy, through a new model integrating unsupervised learning and graph convolutional network, we achieved precise localization of epileptic foci and revealed their network-level pathological characteristics. In the precise analysis of EEG signals, we implemented the first Transformer model framework integrating spatiotemporal information that can be applied to MI-BCI neural decoding, achieving breakthrough results in high-precision neural decoding and the mechanism of model-brain alignment. In the construction of brain-controlled systems, we developed an intelligent rehabilitation system integrating brain-computer interface, functional electrical stimulation, and exoskeleton robots. By building a lightweight deep learning model, it realizes accurate recognition of motor intentions and real-time assistance, providing personalized solutions for patients with lower limb dysfunction. The achievements of this study in the precise analysis of brain diseases and the construction of brain-controlled systems through the integration of artificial intelligence and multimodal signals provide important methodological and technical support for promoting the development of precision medicine and improving the diagnosis and treatment effect of brain diseases.

Time	Presentation
16:35-16:50 PA221-A	Title: Decoding FRB energetics and frequency features hidden by observational incompleteness Author(s): Chen-Ran Hu, Yong-Feng Huang, Jin-Jun Geng, Chen Deng, Ze-Cheng Zou, Xiao-Fei Dong, Yi-Dan Wang, Pei Wang, Fan Xu, Lang Cui, Song-Bo Zhang, Xue-Feng Wu Presenter: Chen-Ran Hu, Nanjing University, China Abstract: Fast radio bursts (FRBs) are fierce radio flashes lasting for a few milliseconds from the sky. Although their connection to strongly magnetized neutron stars has been strongly indicated, the exact triggering process and radiation mechanism are still unknown and highly debated. Due to their extremely short duration, the observation of FRBs has long been a difficult task even for large radio telescopes. The difficulty results from the fact that the information obtained in observations is always incomplete, since the telescope always has a limited flux sensitivity and finite operating frequency band. A pressing challenge is to decode the intrinsic features of FRBs from the incomplete

observations. Here we establish an efficient methodology to overcome this problem, aiming at effectively correcting for the fluence and frequency cutoffs. Using this method, inverse modeling is performed on a large number of repeating bursts from FRB 20121102A to recover their intrinsic features. It is found that strong bursts intrinsically tend to concentrate their energy in a narrow band, while the spectral range of weak bursts can be either narrow or wide. However, when a weak burst has a broad spectrum, the wing of the spectrum can easily go undetected, resulting in a very narrow spectrum being observed. The narrow spectrum features observed in repeating FRBs are thus an observational selection effect. Underestimation of the burst energy caused by observational cutoffs is also corrected for, and the intrinsic burst energy distribution is re-constructed. It is also found that the bandwidth increases with the increasing central frequency in the Arecibo sample (1.15-1.73 GHz), but such a correlation is not observed in the FAST (1-1.5 GHz) and GBT (4-8 GHz) sample. It indicates the emission pattern of the FRB source might vary across different active periods and frequency bands.

16:50-17:05
PA696

Title: Broadband Effect Elimination of Low-frequency SKA based on Deep Learning

Author(s): Chao Li, Li Zhang, Yang Liu, Jinbo Zhao

Presenter: Chao Li, Guizhou University, China

The Square Kilometer Array (SKA), the world's largest radio telescope array with the largest comprehensive aperture, provides unprecedented opportunities for lowfrequency astronomical observations with its excellent sensitivity and resolution. However, when SKA performs broadband observation imaging in the low-frequency band, broadband effect are usually triggered, resulting in distortion and blurring of the celestial structure after imaging. The results show that the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) of the used method are 46.10 dB and 0.9921, respectively, which are 32.9 dB and 78.38 % higher than those of the broadband effect dirty images. This verifies that the MambaIRv2 model can effectively improve the imaging quality of low-frequency SKA.

17:05-17:20
PA357

Title: Graph Attention Network for Industrial Equipment Fault Diagnosis Based on Learnable Mel Filters

Author(s): Hong Zhang, Bin Wang, Xi Kuang, Yuanyi Hou, Dawei Zhao, Hangyu Zou, Jianwei Zhang and Dekun Hu

Presenter: Hong Zhang, Chengdu University, China

Abstract: Graph Neural Networks (GNN) have become a powerful geometric learning model capable of learning complex relationships between data and have been successfully applied in mechanical fault diagnosis. However, existing GNN-based diagnostic frameworks that rely on physical state variables such as vibration and temperature have a significant drawback: they all inevitably compromise the original structure of the equipment. In this paper, we propose using audio signals as the subject of study. We introduce a learnable Mel filter to extract adaptive Mel features and combine it with a Graph Attention Network to form the LMF-GAT framework for fault detection, which is designed to learn the complex time-frequency features of audio signals. Experiments on fault detection tasks for equipment such as UnloaderValve and PrimaryAirFan demonstrate that LMF-GAT achieves good performance and exhibits robustness under imbalanced sample conditions, outperforming several baseline models. LMF-GAT also has an extremely low number of parameters, making it suitable for deploying fault detection networks in real industrial production. It can meet the needs of real-time computation and low-power edge devices.

17:20-17:35
PA674

Title: E- NEXT-TDNN: An Enhanced NEXT-TDNN with Dynamic Weighted Aggregation and Attentive Feature Fusion for Speaker Verification

Author(s): Bohua Li, Lin Zhang and Ye Jiang

Presenter: Bohua Li, Nanjing University of Finance & Economics Nanjing, China

Currently, NEXT-TDNN is one of the state-of-the art deep models for speaker verification (SV) tasks. However, its feature aggregation relies on a static mechanism with simple concatenation or addition, lacking adaptive modeling of frame level feature importance and global contextual information. To address these limitations, this paper proposes an enhanced NEXT-TDNN (E-NEXT-TDNN) architecture, which incorporates a Dynamic Weighted Aggregation (DWA) mechanism in the main branch and employs an Attentive Feature Fusion (AFF) strategy within the backbone module. The Dynamic Weighted Aggregation (DWA) enables the model to dynamically prioritize more infor mative channel features across different levels through adaptive weighting. The Attentive Feature Fusion leverages channel-wise attention weights to control the fusion ratio between input and output features, enabling the model to dynamically determine which parts of the input should be retained and which parts of the output should be incorporated. The experimental results show that the method we proposed achieves an EER of 0.787% and a minDCF of 0.0788 on the VoxCeleb1-O dataset, which are improved by 5.95% and 14.13% respectively compared with NEXT-TDNN.

17:35-17:50
PA475

Title: Performance analysis and adaptive behavior of the least linex function-based filter

Author(s): Hui Zhong, Shiwei Yun, Yihui Yang, Liwei Wang, Chuanwu Zhang, Sihai Guan

Presenter: Hui Zhong, Southwest Minzu University, China

Abstract: This paper proposes a novel least Linex adaptive filtering (LLAF) algorithm to address the limitations of conventional methods in handling asymmetric measurement noise distributions. The proposed approach leverages the Linex loss function to develop a robust adaptive filtering framework. The algorithm's performance is evaluated through theoretical analysis, focusing on the mean and steady-state mean square errors (MSE). Furthermore, the computational complexity of the proposed algorithm is systematically analyzed. Extensive simulations in system identification scenarios demonstrate the algorithm's effectiveness, with experimental results aligning well with theoretical predictions. Comparative studies show that the proposed LLAF algorithm outperforms existing robust techniques regarding steady-state performance and robustness against diverse measurement noise conditions.

Onsite Oral Sessions

T07: AI-based Intelligent Information System Design and Data Computing

Chair: **Zhipan Wu,** Huizhou Universtity, China

16:15-17:35 | Aug.16, 2025 | Mingjun Building Room 130

Invited Speaker

Chuan Zhao | 16:15-16:35

Chengdu University of Technology, China

Title: The Essence of AI and the Future Ways of Computing

Abstract: Now AI even has named this age, but what is it? Is it a new thing or an ancient one? This speech should explore the essence of Artificial Intelligence, the status of AI in the view of Science. Use the EPR Paradox structure as a basement to review the development from physic to AI, to Intelligence science (IS) & Information Science (IS); The two steps of AI; The Humes Problem is a difficult one in the history of philosophy of logic and it is the time to really approach and deal with it. 2. It should introduce Phase Theorem that speaker put out in 2005. Use this theorem to explain the essence of language, here we can go on explain NLU/NLP; On the other hand, we have set three NLU systems, one to generate Chinese couplet, one for Chinese ancient poem generation and another for chat. Such long years practice, let us face the the development of ML and try to research the mechanism. During such process we feel the power of CNN, and trouble with how to abstract new rules and add into the systems. NLU even needs to form a new branch as NLUU, to understand the generating things of Chat GPT, Deep Seek and such NLU systems. We should do multilingual joint research to hold full and real meanings of human being's language, and the breakthrough should be study of Chinese.3. Discuss the ways of future computation. As Geoffrey Hinton put out mortal computation, instead of digital computation's linear developing and arrives to its top now, it is time to turn back; Hinton put out such view at early 2025, that "We are rational beings. We are not. We are analogy machines rather than reasoning machines. " so we should think how to add the factor of human being into computation. What is the prospect of computation in future science and technology? Analog computing, embodied computing, human-factor computing, or say human-computer interaction computing and so on should come in to sight, and means new kinds of computers the same. We can look forward a great symphony of kinds computation. That we should have completer strategy to study NLP, AI and IS's, then promote the development of future civilization.

Time	Presentation
16:35-16:50 PA225	Title: ATFNeRF: Attention Fusion-Based Neural Radiance Fields for Sparse View Inputs Author(s): Li Jin, Jiageng Huang, Zhenxing Xu and Wei Gao Presenter: Jiageng Huang, Jiangsu University, China Abstract: Neural radiance fields (NeRF) suffer from significant performance degradation under sparse input conditions due to their heavy reliance on dense, evenly distributed viewpoints. While existing methods leverage cross-scene priors or geometric regularization to mitigate this limitation, they struggle to balance optimization stability, geometric accuracy, and generalization across diverse scenarios. To address these challenges, we propose ATFNeRF, a novel framework that integrates explicit multi-plane representations with implicit neural radiance fields through advanced attention mechanisms. Specifically, a dual-branch planar learning fusion (PLF) module, which enhances 3D geometric

modeling by employing channel attention for feature recalibration and spatial attention for contextual reasoning. In addition, an attention-based feature fusion (AFF) module, which bridges explicit and implicit representations via frequency-aligned multi-head attention, enabling robust dynamic scene modeling through learnable fusion weights. Extensive evaluations on sparse-view static and dynamic NeRF benchmarks demonstrate that ATFNeRF achieves state-of-the-art performance, while ablation studies confirm the effectiveness of each proposed component.

16:50-17:05
PA477

Title: MGACMR: Multi-Granularity Feature Adversarial Cross-modal Retrieval Model

Author(s): Xiaorong Zhang, Sipeng Deng, Jiakang Deng, Cheng Chen and Guijuan Wang

Presenter: Xiaorong Zhang, Southwest University of Science and Technology, China

Abstract: Existing cross-modal retrieval methods based on subspace networks lack mutual guidance and collaborative learning between different modalities when applied to the image-text retrieval of multi-source heterogeneous data. The result is that they fail to fully extract the common semantics and complementary information among multi-source heterogeneous data. To address this issue, we present MGACMR, an multi-granularity adversarial cross-modal retrieval model. In this model, First, we embed the attention mechanism into the adversarial subspace learning network and construct a generative adversarial network for local multi-granularity feature fusion to jointly learn semantic-consistent representations within and across modalities. Second, we designed a cross-modal fine-grained semantic learning network based on the collaborative attention mechanism to enhance the complementary capability of semantic features within the subspace by associative guiding between the different semantics across modalities. The experimental results demonstrate that the proposed approach outperforms similar cross-modal retrieval methods in terms of the Mean Average Precision metric.

17:05-17:20
PA353

Title: Cliva-Med: A Lightweight Vision-Language Model for Medical Visual Question Answering

Author(s): Juni Stefanus Santoso, Xiaohan Ma, Kehui Song, Rize Jin, Xuguang Qiu, Hongbo Zhao

Presenter: Rize Jin, Tiangong University, China

Abstract: While vision-language models (VLMs) have shown promise in healthcare and biomedical fields, challenges persist concerning their computational requirements and accessibility. This paper introduces Cliva-Med, a lightweight medical VLM. It consists of a vision encoder, a lightweight large language model backbone, and a cross-modality projector. It is trained through a process involving the reformulation of a large multimodal biomedical dataset for visual-language alignment and instruction fine-tuning with a smaller dataset to enhance understanding of medical concepts and scenarios. Experimental results demonstrate that Cliva-Med achieves performance comparable to or exceeding current state-of-the-art models in closed-ended medical question answering and image classification tasks, while requiring only 30-50% of their parameters. Qualitative analyses further demonstrate Cliva-Med's strong logical reasoning and enhanced multimodal capabilities. Our code is publicly available at https://github.com/jun1299/Cliva-Med.

17:20-17:35
PA669

Title: Severity Analysis of Depression Based on Text Convolutional Neural Network

Author(s): Xianlong Tan and Shuhua Mao

Presenter: Shuhua Mao, Wuhan University of Technology, China

Abstract: Depression, a prevalent mental disorder, affects over 280 million people globally, necessitating objective and efficient identification methods. Traditional diagnosis relies on subjective assessments, suffering from limitations like misdiagnosis and delayed intervention. This study proposes a text-based severity analysis of depression using a Convolutional Neural Network(CNN). Comments from social platforms (e.g., Weibo, Baidu Post Bar) were collected, preprocessed (removing noise, stopwords), and labeled into three categories: normal, mild, and severe depression. Word2vec was

employed to generate word vectors, and a TextCNN model was trained to classify severity.The results show that the model achieves the best performance with a convolutional kernel size of 3, yielding precision rates of 65.7% for normal, 92.7% for mild, and 74.5% for severe depression. However, challenges remain, including dataset size limitations and classification ambiguities. This approach provides an automated alternative for depression severity analysis,

highlighting the potential of NLP techniques in mental health assessment.

Onsite Oral Sessions

T08: Natural Language Processing and Information Retrieval

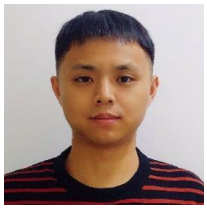
Chair: **Jiaju Wu,** Institute of Computer Application China Academy of Engineering Physics, China

16:15-17:35 | Aug.16, 2025 | Mingjun Building Room 131

Invited Speaker

Zhiyong Qiu | 16:15-16:35

National Key Laboratory of High-End Server Systems, China



Title: An Efficient Hybrid Optimization Framework For Deep Neural Networks

Abstract: Optimization algorithms (optimizers) are crucial for accelerating deep learning training, but they often suffer from high computational complexity and inefficiency. To address these challenges, this study proposes a novel optimization method named Mixer, which integrates both first-order and second-order optimization techniques. The core contributions of Mixer include: (1) designing adaptive algorithm switching conditions; (2) proposing a gradient vector correction strategy to mitigate accuracy degradation; and (3) adopting a lightweight second-order method based on efficient matrix decomposition, significantly reducing computational and memory overhead. Through evaluation on real-world dataset training for networks like ResNet and BERT, and comparison against multiple state-of-the-art methods, the results demonstrate that Mixer delivers superior performance while achieving the same convergence accuracy. Specifically:(1)When training ResNet152, Mixer saves 48.86% and 14.06% end-to-end training time compared to SGD and KAISA, achieving speedups of up to 1.95× and 1.16×, respectively.(2)When training BERT-large, Mixer saves 30.93% and 22.63% training time compared to LAMB and KAISA, achieving speedups of 1.45× and 1.29×, respectively.(3)Simultaneously, Mixer reduces memory footprint by at least 17.42% compared to KAISA.

Time	Presentation
16:35-16:50 PA502	Title: Temporal Knowledge Graph Reasoning with Dynamic Context-Aware Attention and Evolutionary Learning Author(s): Qingkui Chen, Siling Feng*, Mengxing Huang, Yingjie Feng Presenter: Qingkui Chen, Hainan University, Haikou, China Abstract: Temporal Knowledge Graphs (TKGs) are critical for modeling dynamic knowledge with temporal semantics, but existing methods face challenges in capturing fine-grained structural evolutions and long-term temporal dependencies during entity and relation prediction. To address these challenges, this paper introduces DAEL (Dynamic Context-Aware Attention and Evolutionary Learning), a novel model that effectively tackles both dynamic structural dependencies and long-term temporal patterns through two complementary components: the Dynamic Context-Aware Graph Attention Layer (DCAGAT) and an LSTM-GRU joint evolution module. For structural modeling, DCAGAT enables dynamic neighborhood aggregation and multi-perspective feature extraction via adaptive message weighting, dual-relation utilization, and multi-head attention. This design allows the model to prioritize critical neighbors based on temporal context and relational attributes, outperforming static aggregation methods in capturing evolving structural patterns. For temporal modeling, the LSTM-GRU module employs LSTM to model long-term temporal dependencies of relations and GRU to promote co-evolution of entities and relations, to enhance the model's capability to handle complex temporal

dynamics. Experimental results on ICEWS14, ICEWS18, ICEWS05-15, and GDELT datasets show that DAEL achieves MRR improvements over the baseline model RE-GCN of 0.86%, 0.33%, 1.10%, and 0.32% in entity prediction tasks, and 0.83%, 0.22%, 0.98%, and 0.14% in relation prediction tasks, respectively.

16:50-17:05
PA347-A

Title: Research on Intelligent Analysis and Accurate Recommendation System for Policy Text Based on AI Large Language Model: Taking Local Government Digital Economy Development Policy Text as an Example

Author(s): Ding Zhi

Presenter: Ding Zhi, Guizhou University, China

Abstract: The advent of ChatGPT has opened the prelude to the popularization of generative AI, and the AI large language model(AI LLM) has become another focus of industrial development after the Internet. Currently, in the construction of digital government, many local governments are using big language models to promote the improvement of government digitalization and intelligence. The application scenarios are very wide, covering multiple fields such as government office, government services, and centralized management. In practice, local governments have successively launched government AI LLM service platforms, and the digital government field has become one of the main battlefields for dialogue language models. This article is based on the thinking chain, prompt word engineering, and retrieval enhancement technologies of the AI LLM. It designs an intelligent parsing and accurate recommendation system for local government digital economy development policy texts, providing intelligent support for local governments to formulate digital economy policies and promote the integration of big data and the real economy. Firstly, natural language processing methods such as vector space, structural analysis, vocabulary recognition, semantic association, multidimensional analysis, automatic text clustering, and knowledge graph are used to structurally process the digital economy development policy texts crawled from the network, generating a corpus of digital economy policy data. Secondly, a text analysis and precise recommendation software system framework based on the AI LLM was designed for local government users, and a vertical domain policy recommendation intelligent agent was constructed. Once again, a digital economy knowledge intelligent recommendation algorithm based on collaborative filtering, matrix factorization, knowledge graph clustering, deep learning, attention mechanism and other fusion training is provided. Finally, the effectiveness and feasibility of the proposed text parsing and precise recommendation system were verified through experimental evaluation. The results show that the policy text intelligent parsing and precise recommendation system based on AI LLM effectively solves the problem of "information overload" faced by local government users in digital economy policy retrieval and content decomposition, significantly improves the efficiency of policy knowledge services and user experience, and has important technical theoretical value and broad application prospects.

17:05-17:20
PA511

Title: Design and Application of an Evaluation System for Large Language Models

Author(s): Ying Shen, Zhiyuan Cao and Mingang Chen

Presenter: Zhiyuan Cao, Shanghai Development Center of Computer Software Technology, China

Abstract: With the widespread adoption of Large Language Models (LLMs) across various domains, establishing an evaluation system has become increasingly critical. However, existing mainstream evaluation methodologies still suffer from limitations in both evaluation datasets and evaluation methods. To address these limitations, we propose an innovative evaluation system for LLMs. The system comprises four core components: (1) definition of evaluation dimensions, (2) construction of evaluation datasets, (3) automated interaction, and (4) automated evaluation, together forming a complete closed-loop evaluation pipeline. Notably, to bridge the gap between existing evaluation benchmarks and the latest standards, we construct a multi-dimensional evaluation dataset fully aligned with those standards. Based on this dataset, we perform a comprehensive evaluation of multiple LLMs, systematically revealing their performance across dimensions defined by standards.

17:20-17:35
PA520

Title: Multi-information Fusion Knowledge Graph Representation Learning Method Based On Large Language Model

Author(s): Junjie Ma, Ruiyang Shang, Junpeng Bao and Yifeng Huang

Presenter: Ruiyang Shang, Xi'an Jiaotong University, China

Abstract: Knowledge graph has been widely applied in many artificial intelligence fields, but most of the current knowledge graph representation learning methods rely on graph structure information, and do not fully exploit the entity semantic background information and hierarchical structure information contained in the graph. As a result, when we face complex problems, there are issues such as high computational complexity, lack of interpretability, and low reasoning accuracy. Aiming at the problem of how to effectively utilize the entity semantic background information and hierarchical structure information in the knowledge graph, this paper proposes a multi-information fusion knowledge graph representation learning method LHKGE based on a large language model. The LHKGE enriches the semantic representation of entities in the knowledge graph by using the knowledge of the large language model itself. Our experimental results on popular knowledge graph datasets demonstrate that the proposed method significantly improves accuracy in link prediction tasks, showcasing its superiority.

Onsite Oral Sessions

SS02: Artificial Intelligence for Aerospace Applications

Chair: **Ma Zhong**, Xi'an Microelectronics Technology Institute, China

14:00-16:15 | Aug.17, 2025 | Mingjun Building Room 128

Talk Details	
Time	**Presentation**
14:00-14:15 **PA209-A**	Title: VQA-driven Reranking Strategies for Improved Video Retrieval Author(s): Kazuya Ueki Presenter: Kazuya Ueki, Meisei University, Japan Abstract: Recent advancements in video retrieval have demonstrated the effectiveness of multi-modal embeddings in capturing complex video content. In this study, we propose a reranking approach that leverages Visual Question Answering (VQA) models to refine initial retrieval results, further improving retrieval performance beyond embedding-based methods. Using visual understanding driven by VQA, our method optimizes the validity of search ranking by strengthening the alignment between the semantic intent of queries and video content. We evaluated our approach on the TREC benchmark video dataset (AVS task) with more than 1.4 million video samples and achieved a significant improvement in mean average precision, surpassing existing state-of-the-art methods. The experimental results demonstrated that incorporating VQA-based reranking improved cross-modal search performance and enhanced retrieval relevance.
14:15-14:30 **PA713**	Title: Performance Evaluation of YOLOv11 for Ship Recognition in Complex Sea Surfaces Author(s): Hanqing Xue,Chao Chen, Xiaowei Shao, JunXiang Zhao, XiaoMing Dou, Hejia Wang Presenter: Hanqing Xue, Shanghai Jiao Tong University Science and Technology on Space Physics Laboratory, Shanghai China Abstract: To enhance the real-time recognition level of ship targets in remote sensing images under complex scenarios such as variations in lighting, background interference, and changes in target size, this study employs an intelligent recognition algorithm based on the YOLOv11 architecture. Tested and validated on typical maritime ship recognition datasets, the algorithm achieves a 2.3% to 5% improvement in mAP50 compared to classic detection algorithms like YOLOv5 and YOLOv8. Effectively balancing recognition accuracy and speed, the algorithm holds broad application value for the deployment in typical embedded scenarios such as space-based autonomous target recognition, maritime monitoring, and emergency rescue.
14:30-14:45 **PA1005**	Title: Flexible Superpixel-based Graph Neural Network for Hyperspectral Image Classification Author(s): Haojie Hu, Jiaxin Li, Shuanghao Fan, Fang He, Dong Wang Presenter: Haojie Hu, Rocket Force University of Engineering, China Abstract: Recently, Graph Neural Network (GNN) has garnered increasing attention in hyperspectral image classification (HSIC) due to their ability to model complex spatial-spectral relationships. Most GNN-based approaches for HSIC typically rely on manually constructed graphs according to predefined superpixels after segmentation. However, such predefined graph nodes is fixed, and may not optimally capture the discriminative features required for classification, particularly given the high variability in spectral characteristics across different land-cover classes. To address this limitation, we

propose a flexible superpixel-based graph neural network that generates adaptive superpixels. The entire framework is designed to be end-to-end trainable. Experiments on three widely used hyperspectral datasets validate the effectiveness of our method, demonstrating superior performance compared to existing approaches.

14:45-15:00
PA517

Title: DCFSNet: A Feature Significance-Driven Framework forScientific Visualization Color Mapping

Author(s): Jian Zhang, Xiaorong Zhang, De Xing, Fang Wang, Liang Deng

Presenter: Xiaorong Zhang, Southwest University of Science and Technology, China

Abstract: Color mapping plays a crucial role in scientific visualization, with particular emphasis placed on the numerical variation trends and distribution of features within data spaces. As such, differential coloring of feature regions has become a key task. To more effectively and prominently display feature regions, manual adjustment of color mapping is necessary to align with the data characteristics. However, existing methods fail to incorporate the concept of perceptual significance when identifying feature regions. To further enhance the distinguishability of these regions, this paper presents a feature significance-driven framework for scientific visualization color mapping. The framework aims to address the limitations of traditional color mapping in terms of feature region significance, while also enabling the transfer of color styles from a specified image. We formalize the color mapping adjustment as a nonlinear constrained optimization problem. Through robust experiments, effectiveness tests, and comparison experiments, our method demonstrates significant numerical improvements in feature region significance, regional contrast, and feature sensitivity, thereby better supporting scientists in the analysis and exploration of feature regions.

15:00-15:15
PA665

Title: Research on Fault-Tolerant Methods for Matrix Multiplication in GPGPU

Author(s): Yingke Gao, Peidao Wang, Pei Wang, Zheng Yang, Shenglong Li and Bo Liu

Presenter: Yingke Gao, Beijing Institute of Control Engineering, China

Abstract: This paper focuses on the reliability issue of general-purpose computing (GPGPU) in radiation environments and studies a configurable matrix multiplication fault-tolerant algorithm. By comparing and analyzing the performance of the Block Diagonal Matrix Stacking fault-tolerant algorithm (BDMFT) and the checksum-based fault-tolerant algorithm (ABFT) in heavy ion radiation experiments, the fault-tolerant capabilities and limitations of both were quantitatively evaluated. The experimental results show that the ABFT algorithm achieves lightweight protection through the row-column checksum mechanism, has a 100% correction rate for single-row/single-column errors and only introduces linear storage overhead. Although the BDMFT algorithm can shield single-replica errors through trimodal redundancy, its cube-level storage overhead intensifies the hardware load, leading to a significant increase in the risk of hard errors during high-dimensional operations. Based on the above findings, a hierarchical protection optimization scheme is proposed: ABFT is given priority to correct sparse SEU errors; When multiple row/column errors are detected, the BDMFT redundant voting is dynamically activated; If BDMFT fails, ABFT will be automatically executed in blocks. This strategy integrates the low-overhead detection advantage of ABFT with the complex error shielding capability of BDMFT, providing a new idea for high-reliability matrix computing of GPGPU in radiation environments.

15:15-15:30
PA681

Title: Multi-Strategy Improved Whale Optimization Algorithm – Back Propagation Neural Network Damping Prediction Model

Author(s): Yuanyuan Jiang, Xianghua Jiang and Chenhong Du

Presenter: Yuanyuan Jiang, Beihang University, China

Abstract: To address the low damping prediction accuracy of shrouded turbine blades, this study proposes an improved Whale Optimization Algorithm–Back Propagation (IWOA-BP) neural network model. The algorithm incorporates a grouping-based cooperative co-evolution strategy and a good

point set initialization to enhance global search capability and prevent premature convergence. Its effectiveness is validated using 29 CEC2017 benchmark functions. The improved WOA is then used to optimize the initial weights of the BP neural network for predicting the damping behavior of zig-zag shroud blades under varying normal pressures. Experimental results show that the proposed IWOA-BP achieves a test set mean squared error (MSE) of 1.79×10^{-9}-approximately 2351 times lower than the standard BP model-while requiring only 12 seconds of computation. The predicted damping ratio curves closely match the actual values, demonstrating excellent accuracy and strong engineering applicability.

15:30-15:45
PA633

Title: Aero-engin'Fault Diagnosis Based on 1DCNN-BiLSTM

Author(s): Jiaju Wu, Linggang Kong, Ze Xiu, Yao Pan, XiaoDong Wang, Chuan Chen

Presenter: Jiaju Wu, Institute of Computer Application China Academy of Engineering Physics, China

Abstract: Aircraft engines are key components of airplanes. During its lifecycle, an aircraft engine may experience various types of failures due to different flight conditions. By using certain intelligent algorithms to calculate and analyze real-time status monitoring data, real-time diagnosis and prediction of aircraft engine failure can be made, ensuring safe flight of the aircraft and avoiding accidents of aircraft destruction and human casualties. This article proposes a fault mode diagnosis method for aircraft engines based on 1DCNN-BiLSTM. Real-time detection of operational degradation trends during the lifecycle of aircraft engines using 1DCNN-BiLSTM. This method can directly calculate and analyze the degradation of aircraft engines based on QuickAccess Recorder (QAR) and Digital Flight Data Recorder (DFDR) data, without the need to integrate other algorithms to extract fault degradation features. This algorithm fully utilizes the advantages of 1DCNN in extracting local features in the time dimension and BiLSTM in handling nonlinear time series well and utilizing bidirectional contextual information. Finally, fully connected layers are connected to learn the final feature information, and a softmax function is used to diagnose fault categories. The paper categorizes the fault modes into three types: no fault, HPC fault, and HPC&Fan fault. The proposed algorithm has been validated to have higher diagnostic accuracy using the NASA CMAPSS aircraft engine dataset. This fault diagnosis method has practical value in improving the operational reliability of aircraft engines and further predicting the remaining useful life (RUL).

15:45-16:00
PA725

Title: Universal End-to-End Automated Evaluation Framework for Retrieval-Augmented Generation System

Author(s): Shen Li, Haocheng Zhang, Yuejiao Wang, Meijuan Zhou, Zhong Ma, Feng Liang

Presenter: Shen Li, Xi'an Microelectronics Technology Institute Xi'an, China

Abstract: Retrieval-Augmented Generation (RAG) has become one of the key functionalities in current large language model (LLM) applications. It effectively mitigates hallucination issues in LLMs and provides users with the latest domain-specific information by accessing knowledge bases, serving as a cost-efficient method for LLM capability migration. However, users primarily interact with LLMs through application systems like OpenWebUI, where RAG functionality is typically embedded as a "black box." This makes large-scale automated evaluation challenging for existing RAG assessment systems. To address this, this paper proposes an universal end-to-end automated evaluation framework. This framework can evaluate the RAG functionality embedded within various LLM application systems and output quantitative scores. We conducted experimental validation on widely adopted LLM application systems including OpenWebUI, AnythingLLM, and RAGFlow. The results demonstrate that the framework can effectively evaluate RAG performance across different LLM application systems and their configurations.

Title: An Optimization Method for LLM-based Retrieval-Augmented Generation System on OpenWebUI

Author(s): Yuejiao Wang, Haochen Zhang, Shen Li, Meijuan Zhou and Zhong Ma

Presenter: Yuejiao Wang, Xi'an Microelectronics Technology Institute, Xi'an, China

16:00-16:15
PA679

Abstract: Retrieval-Augmented Generation (RAG) systems integrate retrieval mechanisms with generative models to enhance the output of generative models by sourcing relevant information from external knowledge sources. However, redundancy and noise in massive documents can interfere with large models to identify key information, leading to generation errors. This study proposes an optimized RAG framework leveraging OpenWebUI, implementing three key innovations: First, to mitigate semantic deviations in query rewriting, the original query is used as a set of retriever input, both original and rewritten queries are concurrently submitted to the retrieval system. Second, considering the risk that duplicate documents with smaller semantic distance will be filtered out in advance, a rearrangement and deduplication strategy based on semantic distance is designed to give priority to retaining document chunks with smaller distances. Finally, the BGE-M3 model–featuring extended text capacity and higher dimensionality-is deployed as a semantic vector model on the OpenWebUI platform, enabling sophisticated semantic matching through enhanced embedding vector representations. Experimental results demonstrate significant enhancement in retrieval precision and generation quality. The proposed approach exhibits substantial application potential for RAG system architecture and large model deployment, offering strong support for the performance improvement of RAG systems in practical implementations.

Onsite Oral Sessions

T09: Image Processing, Computer Vision and Intelligent Sensing Technologies

Chair: **Yan Yan,** University of Guelph, Canada

14:00-15:45 | Aug.17, 2025 | Mingjun Building Room 129

Talk Details	
Time	**Presentation**
14:00-14:15 **PA213**	Title: Balancing Redundancy Reduction and Feature Retention: A PCA-SVM Approach to Hyperspectral Image Classification Author(s): Mengyao Yang, Uzair Aslam Bhatti, Junfeng Zhang, Nodira Sherjanova, Nodira Sherjanova, Ogabek Sultanov, Yonis Gulzar, Mengxing Huang, Siling Feng Presenter: Mengyao Yang, Hainan University, China Abstract: Hyperspectral imaging (HSI) have multiple bands, and there is a strong correlation between bands, which will cause problems such as redundancy of hyperspectral image information, dimensional disaster, and difficult data classification. Therefore, this paper proposes a classification model consisting of two parts, PCA and SVM. In this method, the acquired hyperspectral images are first preprocessed, and then the preprocessed HSI data reduces the dimension through PCA. After reducing the dimension of the spectral image, the SVM was used to perform the final classification and recognition of the hyperspectral image. The proposed method is experimented on three public datasets: Botswana, Indine Pines and KSC, and the classification accuracy is analyzed under different test sets and training sets ratios. Experimental results show that with the increase of the proportion of the training set, the OA, the AA and the Kappa coefficient are significantly improved, and the feature classes are easier to distinguish, which verifies the effectiveness and practicability of the proposed algorithm.
14:15-14:30 **PA501**	Title: Intrusion Event Recognition in Distributed Fiber-Optic Sensing Systems via Domain Adaptation Learning Author(s): Yuxiang Ni, Linling Wang, Di Wu, Hailong Zhai, Xing Hu, Jing Cheng, Dawei Zhang Presenter: Yuxiang Ni, University of Shanghai for Science and Technology, China Abstract: With the extensive deployment of distributed fiber vibration sensing (DVS) systems in security monitoring and structural health diagnostics, conventional recognition algorithms encounter three critical challenges in cross-domain scenarios: environmental noise discrepancy, domain shift, and prohibitive annotation costs. To overcome these limitations, this study presents a novel cross-domain fiber vibration recognition framework leveraging deep domain adaptation. Our architecture integrates a hybrid distribution alignment mechanism combining CORAL (CORrelation Alignment) and MMD (Maximum Mean Discrepancy) metrics, synergistically enhanced by adversarial training and dynamic optimization strategies to achieve robust spatiotemporal feature learning across indoor/outdoor environments. The proposed hierarchical residual-convolutional network extracts multi-scale time-frequency representations while simultaneously minimizing inter-domain divergence through joint optimization of three objective functions: categorical cross-entropy, adversarial loss, and statistical moment matching constraints. Comprehensive ablation studies demonstrate the superiority of our framework over baseline Softmax classifiers and single-alignment variants, with quantitative evaluations revealing significant performance improvements in cross-domain generalization. This work

	establishes a generalized paradigm for optical vibration monitoring in complex field conditions, advancing the practical implementation of intelligent fiber sensing technologies in industrial applications.
14:30-14:45 PA526	Title: UAG-FusionNet: An Uncertainty and Gradient-Guided Deep Medical Image Fusion Network Author(s): Yingjie Feng, Siling Feng, Mengxing Huang and Qingkui Chen Presenter: Yingjie Feng, Hainan University, China Abstract: Deep learning-based image fusion techniques have been widely applied in multi-modal medical image fusion. However, existing methods often face challenges such as insufficient preservation of image structure and quality, loss of multi-scale structural information, and noise and artifacts during the fusion process. To address these issues, this paper proposes an uncertainty and gradient-guided deep medical image fusion network (UAGFusion), which includes an uncertainty-guided fusion network (UAFN) and a gradient-guided modulation gate refinement network (GGMGFN), with the aim of exploring methods to improve medical image fusion performance. The UAFN incorporates an uncertainty-guided feature enhancement module, cross-scale attention mechanism, and multi-level feature fusion (FF) module to enhance the integration and expression capabilities of multi-modal features. The GGMGFN employs a gradient-guided detail extractor to capture edge and texture information and uses a modulated gated fusion module to integrate multi-source feature streams. To guide network training, this paper designs a hybrid loss function comprising content loss, edge loss, multi-scale structural similarity loss, perceptual loss, and uncertainty perceptual loss to achieve a balance between fusion quality and uncertainty. Experimental results demonstrate that the proposed method achieves competitive performance compared to existing approaches across several objective metrics, while offering improvements in structural information preservation and noise robustness.
14:45-15:00 PA697	Title: Galaxy Image Reconstruction from Sparse Sampling using a Stripformer-based Deep Learning Network Author(s): Zujuan Xu, Li Zhang, Yuxin Xia, Xinglong Wang Presenter: Zujuan Xu, College of Big Data and Information Engineering Guizhou University, China Abstract: In radio interferometry imaging, the sparse sam pling of the telescope array on the UV-plane results in a signifi cant side lobe structure of the point spread function (PSF). The convolution of the real sky brightness and the PSF will introduce severe artifacts in the final imaging. In order to mitigate such artifacts and improve the quality of galaxy images, this paper explores the application of the Stripformer deep learning model to the reconstruction of galaxy images. The model innovatively uses he Strip Attention mechanism to re-weight the image features in the horizontal and vertical directions, effectively capture the long-distance directional dependence, so as to more accurately mitigate the artifacts caused by sparse sampling, and clearly restore the structure of the galaxy spiral arm and the fine details of the galaxy nucleus. The experiment in this paper uses simulated Galaxy Zoo galaxies observation data. The experimental results show that the average indexes of RMSE, PSNR and SSIM of the reconstructed image are 0.006, 45.68 dB and 0.995, respectively, which proves the effectiveness of this method in galaxy structure recovery and artifacts suppression, providing a feasible technique for radio astronomy data processing.
15:00-15:15 PA698	Title: Correction of SKA-low Primary Beam Effect based on Deep Learning Author(s): Jinbo Zhao, Li Zhang, Mei Lu, Chao Li and Yang Liu Presenter: Jinbo Zhao, College of Big Data and Information Engineering Guizhou University, China Abstract: In the actual observation of low frequency band, SKA will produce the primary beam observation effect, which will affect the imaging quality. In this paper, a deep learning DeblurGAN-v2 model is used to correct the primary beam effect of SKA-low, and the astronomical imaging software

OSKAR is used to simulate the SKA-low. The Galaxy Zoo sample is used to construct 3000 datasets for DeblurGAN-v2 for model training and testing. In this paper, the PSNR and SSIM peaks of the reconstructed sky image after the correction of the low-frequency SKA primary beam effect using the DeblurGAN-v2 model reach 35.805 dB and 96.2%, respectively, which is much better than the traditional image domain correction method. The correction of the primary beam effect of the SKA-low array is successfully realized. Experimental results demonstrate the model's superior performance in correcting SKA-low's primary beam effect.

15:15-15:30
PA504

Title: A Lightweight and Adaptive Method for Dynamic Feature Suppression in Visual SLAM

Author(s): Jie Xia, Jianan Jiang, Dechao Chen

Presenter: Jie Xia, Hangzhou Dianzi University, China

Abstract: In dynamic scenarios, the core challenge confronting Simultaneous Localization and Mapping (SLAM) technologies lies in the interference of moving objects with system stability. Existing approaches based on semantic or instance segmentation often suffer from high computational costs, limiting real- time performance on resource-constrained platforms. In this work, a dynamic SLAM framework based on lightweight object detection is presented. The YOLOv8 model is optimized with modules including ADown, Ghost convolution, and C2f-GB to improve multi-scale feature fusion and support edge deployment via the NCNN framework. Object motion states are estimated by integrating optical flow tracking and epipolar constraints. An adaptive dynamic feature point removal strategy based on depth information is further designed to distinguish dynamic points from static background. The proposed system is integrated into ORB-SLAM3, forming a real-time, multithreaded SLAM pipeline that improves localization robustness and accuracy under dy_x0002_namic conditions. Experiments on dynamic datasets show that the proposed method significantly enhances both accuracy and efficiency compared to existing dynamic SLAM systems.

15:30-15:45
PA634

Title: Research On Immersive Meta-Universe Interaction Design Based on User Behaviol Perception

Author(s): LIU MENGYU

Presenter: LIU MENGYU, Beijing City University, China

Abstract: With the rapid development of metaverse platforms, user interactions in virtual environments are becoming increasingly personalized and diverse. Accurately perceiving and responding to user behavior has become a key challenge in enhancing immersion and interaction efficiency. This study proposes a multimodal behavior perception model that integrates body tracking, voice recognition, and eye movement detection. Combined with machine learning algorithms, the model enables real-time recognition of user behavior patterns and adaptive feedback mechanisms. A prototype system was developed using the Unity engine, and empirical user testing demonstrated that behavior-driven interface responses significantly improve immersion, interactivity, and personalization in virtual environments. The study concludes by proposing a scalable design framework for immersive metaverse interactions, offering a reference paradigm for the next generation of multi-sensory intelligent spatial experiences.

Onsite Oral Sessions

T10: Multimodal Perception, Audio Processing and Cross-Domain Applications

Chair: **Hangfei Zhang,** National University of Defense Technology, China

14:00-15:45 | Aug.17, 2025 | Mingjun Building Room 130

Talk Details	
Time	**Presentation**
14:00-14:15 PA729	Title: Ethical Considerations on the Application of Unmanned Systems in Urban Warfare Author(s): Hangfei Zhang, Cheng Xie, Wenan Yi, Rui Zhang and Li Cai Presenter: Hangfei Zhang, National university of defense technology, China Abstract: Urban warfare, characterized by dense civilian populations, complex infrastructure, and asymmetric threats, has become a defining feature of modern conflict. Unmanned systems (US)-including unmanned aerial vehicles (UAVs), ground robots (UGVs), and autonomous platforms-are increasingly deployed in these environments to enhance surveillance, precision strikes, and force protection. While their utility is undeniable, their integration raises profound ethical dilemmas, from violations of the laws of armed conflict (LOAC) to disparities in combat asymmetry and harms stemming from technical limitations. This paper examines the ethical dimensions of US in urban warfare, structured around their general applications, intersection with LOAC, and core ethical challenges (unfairness, technical errors, and adherence to just in bello). Drawing on case studies, it argues that ethical deployment requires aligning technical design with LOAC principles, addressing structural unfairness, and mitigating risks of algorithmic harm. The paper concludes with recommendations for policymakers, military, and engineers to balance innovation with humanitarian imperatives.
14:15-14:30 PA346	Title: Transformer-based Multimodal Spatio-Temporal Fusion Neural Network for Pedestrian Intention Prediction Author(s): Beixuan Lu, Shujuan Huang, Chao Shen, Feng Xiao, Wenjuan Zhang Presenter: Beixuan Lu, Xi'an Technological University, Xi' an, China Abstract: Accurately predicting pedestrian crossing intentions is crucial for urban traffic planning and management in order to maximize traffic signal control, upgrade pedestrian crossing infrastructure, and improve the pedestrian travel experience. In order to increase the accuracy of pedestrian behavior predictions and safeguard pedestrian crossing safety in the presence of human-vehicle interaction, this work attempts to build a pedestrian crossing intention prediction model based on multimodal data. The Transformer architecture, with its strong parallel processing and long-distance dependent capture capabilities, is fully utilized by the model to interpret complicated traffic scene data efficiently. In order to improve the model's characterization ability under multimodal action patterns and mine deep pedestrian behavioral aspects, the Transformer encoder module is added to encode multimodal data, including pedestrians' trajectory, velocity, acceleration, and so forth. Lastly, to achieve reliable pedestrian crossing intention prediction, the intention prediction module combines pedestrian action features with object interaction features. The automated driving dataset (JAAD dataset) is also used to validate the experimental model, and the findings indicate that the model is effective at simulating the intentions and behaviors of pedestrians.

Title: MIDI and Text Conditioned Diffusion Model for Style-Controlled Music Generation

Author(s): Dan GUO, Sha YU

Presenter: Dan Guo, College of Music, Guizhou University, China

14:30-14:45
PA691

Abstract: Although Stable Diffusion-based music generation methods have gained popularity in research and applications, they are known for limited style control and detail preservation issues in generated music. To relieve these issues, this paper presents a new text-to-music framework, combining strengths from four aspects: stable diffusion models, CLAP's text-to-audio alignment, MIDI-cue based music generation condition, and Mel frequency spectrogram as intermediate data. To be specific, a VAE based MIDI-condition mechanism is proposed to enforce controlled styles for generated music, based on specified instruments, music pitches, velocities, rhythms and so on in the prompt. Comparative experiments were conducted against state of-the-art music generation models, along with ablation studies. The results demonstrate that our model outperforms others, achieving higher scores in both quantitative and qualitative assessments of music generation.

Title: A Dynamic Graph Attention Network with Sub-sampling for Wheat Phenotype Prediction

Author(s): Songtao Wu, Jing Xu, Yuxia Sheng

Presenter: Songtao Wu, Wuhan University of Science and Technology, China

14:45-15:00
PA678

Abstract: With the growing global population and the escalating impact of climate change, food security has become an increasingly urgent global concern. As one of the most important staple crops worldwide, wheat plays a vital role in ensuring global food supply and stability. However, its complex genetic structure and trait variation pose significant challenges for breeding superior varieties. Genomic prediction has emerged as a cutting-edge approach to accelerate plant breeding, allowing researchers to uncover the complex relationships between wheat traits and genotypes at the molecular level. This provides both theoretical insights and technical support for precision breeding. In this study, we propose GATv2-RS, a dynamic graph attention network with sub-sampling, designed for wheat phenotype prediction. Our architecture dynamically learns feature representations from raw data while effectively capturing the topological structure of sample-based graphs. Experimental results on real-world wheat datasets demonstrate that GATv2-RS outperforms several advanced methods in predictive accuracy.

Title: Edge-AI Empowered Self-Organizing Drone Swarms for Urban Underground Infrastructure Monitoring

Author(s): Hangfei Zhang, Lin Li, Wenan Yi, Dianyi Song, Liming Liu

Presenter: Cheng Xie, National University of Defense Technology, China

15:00-15:15
PA728

Abstract: Urban underground infrastructure systems, such as sewage networks, utility tunnels, and subway conduits, are critical yet challenging to monitor due to their confined, GPS-denied environments and dynamic structural complexities. Traditional inspection methods are labor-intensive, costly, and often fail to provide real-time insights. This paper proposes a novel theoretical framework for autonomous monitoring using Edge-AI empowered self-organizing drone swarms. By integrating edge computing with distributed artificial intelligence, the framework enables drones to collaboratively sense, process, and act on environmental data in real time without relying on centralized control or external connectivity. The self-organizing swarm leverages adaptive neural networks and decentralized decision-making algorithms to dynamically optimize formation, navigation, and task allocation while overcoming constraints such as limited communication bandwidth, obstacle avoidance, and resource constraints. The theoretical model emphasizes resilience, scalability, and energy efficiency, addressing challenges unique to underground settings through emergent swarm intelligence and lightweight on-

device AI inference. This work lays a foundation for next-generation infrastructure monitoring systems, reducing human risk, operational costs, and downtime while enhancing urban resilience through proactive maintenance.

15:15-15:30
PA711

Title: FDSCNet: Flexible Dynamic Sparse Compression Network for Music Source Separation

Author(s): Chengmi Zhang, Dan Guo and Sha Yu

Presenter: Chengmi Zhang, Guizhou University, China

Abstract: Abstract-Music Source Separation (MSS) is vital for applications like audio processing and music production. Traditional MSS methods using fixed-parameter convolutional neural networks struggle with dynamic music signals, limiting separation accuracy. This paper proposes integrating a dynamic convolution mechanism into an SCNet-based spectral-signal separation framework, enhancing music source decomposition with three adaptive kernels. Additionally, a frequency-sensitive hybrid loss function is designed to address limitations of existing loss functions, improving separation performance in key frequency bands by focusing on critical spectral regions. Experiments on the MUSDB18-HQ dataset demonstrate that the proposed FDSCNet achieves superior average SDR performance across vocals, bass, drums, and other sources compared to existing state-of-the-art methods. Index Terms-Music source separation, frequency-sensitive hybrid loss, adaptive dynamic convolution

15:30-15:45
PA730-A

Title: A Randomized Algorithm for Trace Ratio Problem Of High-Dimensional Tensor Data

Author(s): Yuhang Li, Gang Wu

Presenter: Yuhang Li, China University of Mining and Technology, China

Abstract: The trace ratio optimization of high-dimensional tensor data is widely used in many fields, such as machine learning, signal processing and pattern recognition. The goal is to map original data from a high-dimensional space to a low-dimensional space on the premise of preserving the essential information of the original data as much as possible. At present, there is no closed-form solution to solve the tensor trace ratio problem, and the existing methods generally uses inner-outer iterative algorithms to solve it. Faced with large-scale high-dimensional data, it will encounter the challenges of dimensional disaster, time-consuming, insuffcient memory space and so on. In order to overcome these challenges, this paper establishes a closed-form solution, which replaces the traditional iterative process with one calculation, reduces the computational complexity and improves the effciency. A randomized algorithm is proposed to solve it. And the theoretical error bounds are given to guarantee the algorithm's effectiveness. Traditional multilinear discriminant analysis may encounter limitations such as the curse of dimensionality and interference from redundant features when dealing with high-dimensional tensor data, which leads to suboptimal performance. A sparse tensor discriminant analysis method is proposed to achieve feature selection and dimensionality reduction. This method can identify key discriminative features from high-dimensional data and suppress redundant information. Numerical experiments are performed on some real-world data sets, which verified the effectiveness and feasibility of the proposed methods.

Onsite Poster Sessions

PS01: Vision-based Intelligent Digital Image Processing Technologies and Engineering Applications

Chair: **Shichao Wu,** Southwest University of Science and Technology, China

14:00-15:30 | Aug. 16, 2025 | Mingjun Building Room 132

PAPER DETAILS		
No.	Paper ID	Presentation
1	PA654	Title: DIGRD: Disguised Inertial Gait Recognition Dataset and Method Author(s): Hao Li, Zhaoyang Wu, Wei Wu, Shichao Wu Presenter: Shichao Wu, Southwest University of Science and Technology, China
2	PA333	Title: Large-Parallax Image Stitching with Multi-Scale Feature Fusion and Boundary Awareness Author(s): Su Wang, Mingwei Cao Presenter: Su Wang, Anhui University, China
3	PA723	Title: 3D Reconstruction of Photovoltaic Scenes Based on Aerial Images Author(s): Zilong Wang, Mingwei Cao Presenter: Zilong Wang, Anhui University, China
4	PA348	Title: Architecture and Application Path for On-Device Deployment of Aviation Large Vision Models Author(s): Xilin Zhang, Bo Liu, Ziliang Du, Yang Shao Presenter: Xilin Zhang, Chinese Aeronautical Establishment, China
5	PA334	Title: A Robust Method for Image Registration in Large Parallax and Weak Texture Scenes Author(s): Su Wang, Mingwei Cao Presenter: Qiuju Wang, Anhui University, China
6	PA528	Title: Study on Mango Disease Classification Based on YOLOv8 and Multidimensional Data Enhancement Author(s): Yi Zhang, Bin Zheng, Yinbo Lin, Xiaoming Du Presenter: Bin Zheng, Panzhihua University, China
7	PA226	Title: Enhanced Skin Lesion Classification via Swin Transformer and Data Augmentation Author(s): Yanxiang Mao Presenter: Yanxiang Mao, Shanghai University, China
8	PA206	Title: Decoupling Spatial and Frequency Feature for Superpixel Segmentation Author(s): Guanghui Yan, Haowei Du, Xiaohong Jia Presenter: Haowei Du, Lanzhou Jiaotong University, China
9	PA717	Title: GhostAxSeg: A 0.45M Network with Ghost Backbone and Axial Attention for Real-Time Polyp Segmentation Author(s): Xintao Li, Xuemei Sun Presenter: Xintao Li, Tiangong University, China
10	PA639	Title: CT Image Automatic Recognition Method Based on Two-Dimensional Segmentation and Three-Dimensional Visualization Author(s): Baixiang Zeng, Zhiyu Gao, Haibin Lan, Linhai Xu, Wei Guan, Xiaolong Chen, Lindan Zheng, Qianni Wang, Shuangwei Yu, Changsheng Zhang, Jian Fu Presenter: Baixiang Zeng, Beihang University, China

11	PA218	Title: Video Retrieval Architecture Based on Contrastive Learning and Gaussianized Label Mechanisms Author(s): Kaixuan Sun, Lepeng Wang, Yutao Yao, Jiang Shao, Jiajia Nie, Xueqing Li, Zenghuan Cheng Presenter: Yutao Yao, Shandong University, China
12	PA332	Title: FuseNet: Feature Enhancement-based Multi-frame Depth Estimation Author(s): Rui Ni, Mingwei Cao Presenter: Rui Ni, Anhui University, China
13	PA464	Title: Wavelet-guided Progressive and Dynamic YOLO for SAR Ship Detection in Complex Scenarios Author(s): Xipeng Chen；Ming Hao；Jie Guo；Long Zhuang Presenter: Xipeng Chen, Nanjing Research Institute of Electronics Technology, China
14	PA672	Title: A Tual-Flow Network Based on Attention Mechanism and Global Information for Small Object Detection Author(s): Ning Wang, Meijuan Zhou, Zhong Ma Presenter: Ning Wang, Xi'an Microelectronics Technology Institute, China
15	PA343	Title: A Novel Attention Fusion Approach for Joint Object Detection and Classification in Facial Expression Recognition Author(s): Bohao Li, Cheng Peng, Kun Zou, Wenhui Zhou Presenter: BoHao Li, University of Electronic Science and Technology of China, China
16	PA331	Title: MTL-YOLO: Multi-Task Learning-based YOLO for Detecting Bone Irregular and Discontinuity Author(s): Shifeng Shang, Mingwei Cao, Haifeng Zhao Presenter: Mingwei Cao, Anhui University, China
17	PA522	Title: SPBNet-YOLO: Improving Small Object Detection in UAVImagery Under Diverse Conditions Author(s): Muhammad Noman Tariq, Hongying Zhang, Meerab Rahim Khan, Sijia Zhao Presenter: Muhammad Noman Tariq, Southwest University of Sci ence and Technology, China
18	PA664	Title: A CT Defect Detection Method for Additive Manufacturing Workpieces Driven by Digital Models and Virtual Simulation Author(s): Li Ning, Zhiyu Gao, Haibin Lan, Baixiang Zeng, Linhai Xu, Wei Guan, Xiaolong Chen, Lindan Zheng, Qianni Wang, Bingyang Wang, Changsheng Zhang and Jian Fu Presenter: Zhiyu Gao, Beihang University, China

Onsite Poster Sessions

PS02: AI-driven Multimodal Intelligent Interactive System Design and Data Computation

Chair: **Fangyan Nie,** Guizhou University of Commerce, China

16:15-18:00 | Aug. 16, 2025 | Mingjun Building Room 132

PAPER DETAILS		
No.	Paper ID	Presentation
1	PA478	Title: An Improved Bat Algorithm with Multi-strategy for Multi-robot Path Planning Author(s): Fangyan Nie, Pingfeng Zhang, Ranran Ao, Demin Zhang and Zhenlin Lu Presenter: Fangyan Nie, Guizhou University of Commerce, China
2	PA720	Title: Exploring the Application of Improved GA in Multi-objective Path Planning for Engineering Robots Author(s): Zejun Gong, Hui Zhao, Wei He, Yaoguang Zhao Presenter: Hui Zhao, Hainan Normal University, China
3	PA481	Title: Design of Debris Sorting Robot System on Coal Mine Belt Conveyor Author(s): Lei Wu, Ming Lu, Minghui Zhao Presenter: Minghui Zhao, China Coal Technology & Engineering Group Shanghai Co., Ltd., China
4	PA700	Title: Hierarchical Task Decomposition for Long-Horizon Robotic Autonomy via Multimodal Language Model Synergy Author(s): Genjian Yang, Liuhua Zhang and Zhiwei Gao Presenter: Genjian Yang, CEPREI, China
5	PA509	Title: Design of an Energy-Efficient Hybrid-Precision Neural Network Accelerator with a Unified Memory-Compute Architecture Author(s): Feng Jiao, Siying Bi, Shu Wang, Zhong Ma, Jin Huang, Siwei Xiang, Chen Yang Presenter: Feng Jiao, Xi'an Microelectronics Technology Institute (of CASC), China
6	PA724	Title: An Efficient Swin Transformer Architecture Based on Grouped Query Attention Author(s): Li Wang, Zhong Ma Presenter: Li Wang, Xi'an Microelectronics Technology Institute, China
7	PA709	Title: EmotionChat: Emotional Chain of Thought based MLLM for Dialogue Generation Author(s): Yu Bai, Sirui Zhao, Min Zhang, Shifeng Liu, Yifei Zhu, Tong Xu Presenter: Yu Bai, University of Science and Technology of China, China
8	PA675	Title: Ethical and Security Challenges of Large Language Models in Public Health Systems: A Survey Author(s): Jie He, Fengshi Jing, Deyi Liang, Hao Ren, Qimeng Luo, Yewei Xie, Haobin Zhang and Weibin Cheng Presenter: Jie He, City University of Macau, Macau China
9	PA514	Title: NoNitro: A UAV- and AI-Empowered Integrated System for Soil Nitrate Control Author(s): Emma Tong, Sophia Chen and Jun Wu Presenter: Jun Wu, Michigan State University, USA
10	PA638	Title: Research on Intelligent Scheduling Algorithm for Public Transportation System Based on Neural Network Author(s): Xiaofeng Yan, Hongwei Zhang, Jie Zhao Presenter: Hongwei Zhang, Chongqing University of Technology, Chongqing Medical and Pharmaceutical College, China
11	PA472	Title: Attention-Enhanced CNN-LSTM for Ship Intention Prediction Using AIS Data

		Author(s): Bo Yang, Xiao Liu, Zhenya Wang, Shui Yu, Yunbo Rao Presenter: Xiao Liu, University of Electronic Science and Technology of China, China
12	PA219	Title: Unsupervised Sentence Representation Learning via Rank and Self Distillation with LLM-Augmented Negative Sampling Author(s): Yuyang Wang, Rize Jin, Wei Zhao Presenter: Yuyang Wang, Tiangong University, China
13	PA656	Title: Document-level Cyber Threat Intelligence Relation Extraction Via Relation Directed Graph Author(s): Rui Wang, Yu Yan, Jiaoli Liu, Po Hu and Jianyi Liu Presenter: Jianyi Liu, Beijing University of Posts and Telecommunications, China
14	PA355	Title: MelVQGAN: An Enhanced Neural Codec with Fine-Grained Spectrogram Preservation for High-Fidelity Speech Generation Author(s): Siyuan Zhou, Hongxia Bie Presenter: Siyuan Zhou, Beijing University of Posts and Telecommunications, China
15	PA352	Title: Adaptive Huber based Federated Learning for Sparse Optimization Author(s): Ji Cheng, Hongxin Zhao, Xin Wang, Wanping Zhang, Zhaoqilin Zhao Presenter: Wanping Zhang, University of Petroleum-Beijing at Karamay, China
16	PA490	Title: A Hybrid Informer-LightGBM Framework for Enhanced Carbon Emission Forecasting in Industrial Parks Author(s): Xingyun Yan, Jie Hu, Lingyu Wang, Mingzhu Fang, Hairui Wang, Meixing Guo, Bin Chen, Jin Qi Presenter: Xingyun Yan, Shanghai Jiao Tong University, China
17	PA466	Title: A Bottom-Up Olympic Medal Prediction Framework Based on Multiple Machine Learning and Statistical Methods Author(s): Keyuan Zhang, Chenxing Li, Bin Zheng and Jingdong Zhang Presenter: Zheng Bin, Panzhihua University, China
18	PA708	Title: A Prediction Method for State Variables in Mixing Processes Based on Stacking Model Author(s): Ziyang Song, Faping Zhang, Yiming Lu Presenter: Ziyang Song, Beijing Institute of Technology, China
19	PA718	Title: Third-Party Privacy Data Leak Analysis on Ontario Hospital Websites Author(s): Tarun Kalyani, Justin Stewart, Yan Yan, Sampsa Rauti, Ville Leppänen, Zuhaibuddin Bhutto, and Wenjun Lin Presenter: Yan Yan, University of Guelph, Canada
20	PA667	Title: Research and Implementation of Path Planning Algorithm for Autonomous Vehicles Author(s): Min Yan, Zhang Ling Yan Presenter: Min Yan, Chengdu Jincheng College, China

Online Oral Sessions

SS03: Pattern Recognition and Machine Learning

Chair: **Binhui Tang,** Hainan Normal University, China

13:00-17:00 | Aug.17, 2025 | Room A - 827 2618 8431 (Password: 081517)

Talk Details	
Time	**Presentation**
13:00-13:15 PA707	Title: Forest Fire Object Detection Based on an Improved YOLOv5s Author(s): Zhengji Li, Shichao Xiang, Xinrui Li, Yufan Wen, Peiyao Wang Presenter: Peiyao Wang, Chengdu Jincheng College, China Abstract: Conventional forest fire detection algorithms often suffer from slow inference speed, low accuracy, small target detection failure, and missed detection when targets overlap. To address these challenges, An improved detection method based on YOLOv5s has been proposed in this paper. The lightweight YOLOv5s was selected as the backbone and four targeted enhancements were introduced: (1) Mosaic augmentation in conjunction with CSP blocks to streamline the network and improve cross-channel feature fusion while reducing parameters; (2) the Convolutional Block Attention Module (CBAM) after each C3 block in the neck to enable the network to better focus on salient flame regions and suppress background noise; (3) replacing the original Spatial Pyramid Pooling (SPP) block with an Atrous Spatial Pyramid Pooling (ASPP) module to expand the receptive field without sacrificing feature resolution, thus enhancing multi-scale representation; and (4) substituting the conventional CIOU loss with EIoU loss, which decouples width and height penalties to accelerate convergence and refine bounding box regression. Experiments on a 3, 000-image forest- fire dataset show the proposed method achieves an mAP@0.5 of 89.4%, 10.1% higher than the YOLOv5s baseline, and boosts small-object recall to 84.7%. Improved YOLOv5s method achieves 39 FPS inference on an NVIDIA RTX 3080, which exhibits robust performance in diverse and complex scenes.
13:15-13:30 PA690	Title: Application of Object Detection in Sleep-Disordered Breathing Author(s): Haonan Guo, Qingqing Yan, Li Zhou, Hairu Zhao, Yuxin Mao and Fangyi Zhang Presenter: Haonan Guo, Chengdu Jincheng College, China Abstract: Sleep-disordered breathing (SDB) represents a significant health concern, where conventional detection approaches relying on handcrafted feature extraction and shallow classifiers exhibit limitations such as poor generalization capabilities and suboptimal accuracy. To resolve this challenge, we present a YOLO-series-model-based methodology for automatic detection of snoring and respiratory apnea events.This approach treats physiological signal waveforms as '1D images' and employs end-to-end object detect using YOLOv11n and YOLOv12n models, achieving precise identification and localization of snoring and apnea events. To further enhance model performance, ablation studies were conducted by integrating multiple attention mechanism modules (e.g., SE, Gass, A2, NAM) into the backbone network and detection heads. Experimental results indicate, however, that most attention mechanisms failed to yield significant performance improvements for this specific task. Analysis suggests this may stem from a mismatch between their local attention characteristics and the global temporal modeling required by the task. Notably, the NAM module demonstrated advantages in reducing model complexity. Ultimately, YOLOv11n was selected as the optimal baseline model, with multi-dimensional validation performed through confusion matrices, PR curves, and predictive visualization. This study provides an efficient, lightweight, and deployment-ready novel approach for SDB monitoring.

Title: Research on Classification and Recognition of Traditional Chinese Medicines Based on YOLOv8-MSDA

Author(s): Guihua Zhang, Junchao Dai, Yue Guan, Jianing Li, Yin Liu, Jiashu Sun

Presenter: Zhang Guihua, Chengdu Jincheng College, China

13:30-13:45
PA646

Abstract: Aiming at the recognition difficulties of Chinese herbal medicines due to the variety of species, similar morphology and sparse features of small targets, this study proposes a classification and recognition model based on improved YOLOv8. The traditional recognition methods have the problems of low efficiency and high misclassification rate, and the existing YOLOv8 model has insufficient feature extraction ability for multi-scale targets in complex backgrounds, especially for small targets. To this end, the study introduces the multi-scale null attention mechanism (MSDA) and deformable convolution (DCNv2) to strengthen the model's ability to capture complex features: MSDA realizes multi-scale feature fusion by focusing on local sparse feature interactions while reducing the computational redundancy; and DCNv2 strengthens the efficiency of extracting small targets and geometrically deformed features by adapting to adjusting the convolution kernel offsets. Experiments on a self-constructed dataset of 9709 Chinese herbal medicine images show that the improved algorithm achieves good results in the Chinese herbal medicine dataset, with a 2.4% improvement in mAP compared with the original model, and in specific categories, the recognition accuracy of herbs such as Lycium barbarum and Angelica dahurica is improved by a range of 1.2% to 3.9%. The precision rate and recall rate are all improved. The model effectively enhances the ability to capture complex features while taking into account the computational efficiency, providing technical support for the application of automated recognition of Chinese herbal medicines in scenarios such as quality control of drinking tablets.

Title: Research on the Impact of Digital Game Experience on Willingness for On-site travel--LDA Topic Clustering Analysis Based on Player Reviews of "Black Myth: Wukong"

Author(s): Chen Li

Presenter: Chen Li, City University of Macau, China

13:45-14:00
PA716

Abstract: This paper takes 506, 168 player reviews of Black Myth: Wukong as a sample, and uses the LDA topic clustering algorithm to analyze the influence mechanism of the digital gaming experience on the willingness to travel to real - world destinations. Eight core themes, such as interactive experience, cultural experience, and flow experience, were extracted. It was found that the immersive interactive experience activates tourism imagery, the cultural experience builds cultural identity and tourism willingness, the flow experience strengthens tourism motivation, and the transfer of cross - media experience promotes tourism decision - making. A theoretical model of "digital gaming experience - willingness to travel to real - world destinations" was constructed, revealing the hierarchical influence path. This provides theoretical and practical references for the integration of the digital gaming and cultural tourism industries.

Title: Research on the Application of Intelligent Video Recognition Technology in Pet Management Systems

Author(s): Hongxia Mao and Chunjie Wang

Presenter: Hongxia Mao, Chengdu Jincheng College, China

14:00-14:15
PA689

Abstract: Fueled by the rapid expansion of the pet economy, traditional pet management approaches face significant challenges in efficiency, cost, and real-time data acquisition. To address these limitations, this study investigates the application of intelligent video recognition technologies within pet management systems. We propose and implement a collaborative Cloud-Edge-Device architecture, integrating core technologies including advanced YOLO-series object detection algorithms, DNN-based feature extraction, and behavior analysis models combining LSTM with 3D-CNN. This integrated

system enables individual pet recognition, health monitoring, and intelligent interaction functionalities. Furthermore, we discuss key technical challenges requiring future research, such as achieving robust recognition under complex backgrounds, improving small target detection accuracy, ensuring real-time processing performance, and enhancing data privacy protection mechanisms.

14:15-14:30 PA508	Title: TFDS Analysis and Resolution Strategy of Image Problem Under Train Speed Change Condition Author(s): Sha Wen, Qingmao Ren Presenter: Sha Wen, Sichuan University, China Abstract: In order to improve the image acquisition quality of TFDS based on linear array camera and laser light source technology, This paper analyzes the causes of the image problems of the train vehicle in the actual transportation state such as low speed, variable speed, stop and restart, and puts forward the solutions, and verifies the performance of the improved system detection effect through the way of on-site pick-up. The system not only solves the problems of uneven exposure, poor image effect and poor anti-interference performance of the linear scanning camera. It can also ensure that there is no loss or dislocation of the vehicle image, and finally the image recognition technology is used to synthesize the integrity of the whole train of vehicles. The overall improvement scheme and measures of the system improve the quality of TFDS image acquisition.
Break Time	
15:00-15:15 PA228	Title: LLM-Driven Fire Safety Assistant Author(s): Shan He, Yuwei Hu, Hongyu Li, Ting Li, Yanyan Hu, Fengmei Tang Presenter: Ting Li, Chengdu Jincheng College, China Abstract: This paper introduces a novel fire safety assistant system developed based on the InternLM2.5-chat-20B large language model. The system also incorporates TTS, ASR and digital human technologies. It can provides 24-hour professional fire safety consultation through interactive question-and-answer services. A fire safety knowledge vector database with ten categories was built. In addition, 1, 500 question-answer pairs were created to fine-tune the model. The use of fine-tuning and a RAG mechanism improved the system's overall performance. A scenario-based Chain-of-Thought (COT) prompt engineering framework was also developed. This further enhances the clarity and precision of the system's responses.
15:15-15:30 PA692	Title: A Novel Hybrid Prediction Model of Stock Closing Price Based on CEEMDAN-SE-VMD and ARIMA-CNN-BiLSTM-Attention Author(s): Peiyao Wang, Qiang Dai, Zhongying Qiu and Xinyu Deng Presenter: Peiyao Wang, Chengdu Jincheng College, China Abstract: With the development of the world, the stock mar ket plays an important role in the national economy. Investors ar e still paying attention to effective strategies which can increase in vestment returns and reduce risks. The stock market time series data is dynamic, non-stationary and complex, so predicting the stock closing price is a big challenging task. Therefore, this paper proposes a CEEMDAN-SE-VMD and ARIMA-CNN-BiLSTM-attention novel hybrid prediction model. Firstly, the time series data are decomposed into IMFs through CEEMDAN. The sample entr opy is used to statistically classify each component, dividing it int o high-frequency and low-frequency components. Secondly, for high-frequency components, the VMD is combined to decompose further and extract refined features, and the CNN-BiLSTM-Attention model is used for prediction. For low-frequency components, the ARIMA model is used for prediction. Finally, the final stock closing price prediction value is obtained by integrating the result s. Compared to some classic prediction models in this area, this p aper's experimental results have shown that the new hybrid prediction model exhibits excellent performance in evaluation indicato rs such as RMSE, R2, MAE, MAPE, and 1-WMAPE, reaching 11 12.9924, 0.9876, 838.5309, 1.2040 and 0.9882 respectively, and sig nificantly improves the accuracy of

stock closing price prediction.This method can offer investors a fresh perspective.

15:30-15:45
PA721

Title: Student learning portrait based on AI agent and multimodal data fusion

Author(s): Binhui Tang, Renwei Liu, Qianqian Liu, Keke Zha, Yishan Lu

Presenter: Binhui Tang, Hainan Normal University, China

Abstract: Student learning profiles are a core component of educational digital transformation and personalized education strategies. They aim to accurately characterize individual learning characteristics and enable dynamic tracking by integrating students' multimodal, cross-dimensional learning data, thereby promoting the optimization of personalized education. However, current personalized education faces three core challenges: limitations of static data, insufficient knowledge collaboration, and lagging feedback mechanisms. To address these issues, this paper proposes a student learning profile framework based on multi-agent collaboration and multimodal data fusion. This framework consists of three core agents: multimodal data processing, knowledge graphs, and student learning profiles, aiming to achieve more precise and dynamic learning assessment and intervention. Within this framework, the multimodal data processing agent employs late-stage multimodal data fusion technology, cross-modal attention mechanisms, and dynamic time warping (DTW) technology to convert multi-source heterogeneous data streams such as text, images, videos, and audio into high-frequency updated structured representations; The knowledge graph agent dynamically constructs and evolves individual student learning knowledge graphs, connecting knowledge points, skills, and careers, and compares and analyzes them with predefined standard knowledge graphs; The student learning profile agent generates student learning profiles, visualizes subject-specific skill radar charts and career inclination network diagrams, and promptly displays student feedback information. This learning profile framework is based on the Manus multi-agent architecture (planning, memory, and tool modules), supporting real-time updates at the minute level. It breaks through the cognitive bottleneck of “surface visibility but no deep understanding, ‘enabling precise alignment between academic development and career needs. Experimental validation demonstrates that this framework can provide dynamic, explainable, and actionable decision support for personalized education, driving the efficient implementation of personalized education initiatives.

15:45-16:00
PA722

Title: Laser Holographic Image Segmentation and Recombination Processing Method Integrating Artificial Intelligence Technology

Author(s): Hong Zhou, Chunqing Yang

Presenter: Hong Zhou, Chengdu Jincheng College, China

Abstract: This paper proposes a laser holographic image processing method integrating artificial intelligence technology, focusing on improving image quality through an optimized segmentation-recombination framework. The core innovation lies in the introduction of the Adaptive Genetic Algorithm (AGA), which dynamically adjusts crossover and mutation probabilities during image segmentation. Furthermore, the Otsu's method (maximum between-class variance method) is employed to determine the optimal threshold for holographic image segmentation. Finally, by combining the scale difference value with the spatial neighborhood edge energy fusion method, the image pixel sequence is extracted, and the positional difference between the processed image and the original image is corrected. The contour points with the maximum gray value are obtained, and the recombination of laser holographic images is realized through pixel information fusion. Experimental results show that compared with traditional methods and deep learning-based techniques, this method achieves better performance in segmentation accuracy (peak signal-to-noise ratio (PSNR) of 65.4 dB) and recombination efficiency (average registration error rate < 0.15%). Computational complexity analysis indicates the core steps have a time complexity of O(N), and GPU acceleration enables real-time processing (33 frames/s). Dataset validation (500 images covering medical, industrial, and natural scenes) confirms generalizability. These advancements verify the effectiveness of the proposed method in laser holographic imaging applications.

Title: Agent-Based Traffic Monitoring and Regulation System for SDN

Author(s): Xiaodan Guo, Weihao Du, Falong Zhou, Xun Zhu, Mengdie Deng, Juming Bao

Presenter: Xiaodan Guo, Chengdu Jincheng College, China

16:15-16:30
PA714

Abstract: The management of modern networks is becoming increasingly complex, and while Software-Defined Networking (SDN) offers programmability, existing control mechanisms often lack flexibility and operational intuitiveness due to heavy reliance on manual configurations or opaque machine learning models. To address this limitation, we propose the indirect integration of a Large Language Model (LLM)-powered conversational agent into the SDN control loop, implementing the ReAct (Reasoning + Acting) paradigm as our core innovation to synergize natural language reasoning with concrete network operations. This agent utilizes a specialized toolkit that translates human conversational intents both into SQL queries (via a Text-to-SQL engine) for deep network state analysis and into REST API calls (through OpenFlow rule modifications) for real-time traffic regulation, validated through comprehensive emulation-based evaluation.

Title: Secure Storage Strategies for Massive Data in Network Computing Environments

Author(s): Yan Xu, Hongmei Liu, Xinrui Gong, Zeren Fan

Presenter: Yan Xu, Chengdu Jincheng College, China

16:30-16:45
PA467

Abstract: In recent years, the advancement of computer technology has significantly enhanced data computing capabilities in network environments. The advent of the big data era has transformed information in daily life, work, and study into digital correlation chains. Network-based resource sharing simplifies complex learning tasks and improves work/study efficiency. However, traditional network storage strategies frequently encounter issues such as data overflow leaks and storage server security vulnerabilities. To address these challenges, this study proposes a secure large-capacity data storage strategy for network computing environments, which dynamically incorporates Data Correction Units (DCU), Logic Compensation Units (LCU), and Node Encryption Units (NEU) to effectively resolve limitations of conventional approaches. Simulation experiments demonstrate that this strategy features fast response times, strong penetration resistance, high logical consistency, robust data security, and user-friendly operation

Title: The precise portrayal of multi-dimensional students' growth trajectories in the informatization construction of colleges and universities

Author(s): Qiuqiong Hu, Wenxuan Hou, Zihao Yang

Presenter: Qiuqiong Hu, Chengdu Jincheng College, China

16:45-17:00
PA661

Abstract: In the process of informatization construction in colleges and universities, technologies such as network security data analysis and mining, big data security and privacy protection, threat intelligence text mining, natural language processing and mobile big data computing are integrated, and machine learning algorithms are introduced to construct a model for characterizing students' growth trajectories. Taking five universities as research samples, 100, 000 student data covering dimensions such as academic performance, behavior, and social interaction were collected. After parsing the text information through natural language processing and integrating multi-source data through mobile big data computing, the data was input into the model for training and analysis. The results show that the prediction accuracy of academic trajectory (grade trend) based on LSTM reaches 89.2%, and the recognition accuracy of behavioral trajectory (activity scenarios) based on random forest has increased to 91.5%. This effectively helps colleges and universities accurately understand the development trend of students, provides data support for the formulation of personalized education strategies, and promotes the in-depth synergy between the informatization construction of colleges and universities and the growth of students.

Online Oral Sessions

ON01: Object Detection and Detection Models

Chair: **John Clement S. Escobañez,** Polytechnic University of the Philippines, Philippines

13:00-16:45 | Aug.17, 2025 | Room B - 832 0180 3468 (Password: 081517)

Talk Details	
Time	**Presentation**
13:00-13:15 PA635	Title: Occlusion-Aware Instance Segmentation for Scene-level Sketches Author(s): Chenghao Li, Shiping Dong and Shizhe Zhou Presenter: Shiping Dong, Hunan University, China Abstract: Sketching is a simple and effective way for humans to express their perception of the world. Sketch instance segmentation plays a crucial role in sketch understanding and is widely used in areas such as sketch recognition, sketchbased image retrieval or editing. Existing image segmentation methods often fail to achieve optimal results due to the modal differences between natural images and sketches, which ignore the sparsity and abstraction of sketches. Existing sketch segmentation methods mainly focus on semantic segmentation of single instance sketches. In this paper, we propose a new maskedregion convolutional neural network based on a large-size first layer convolutional kernel for instance segmentation of scenelevel sketches. Our experiments have shown that our method has higher segmentation accuracy for scene level sketches and better segmentation performance for scene level sketches with occlusion.
13:15-13:30 PA227	Title: Impact of Different Rain Removal Preprocessing Methods on Rainy Object Detection Performance Author(s): Jie Dong, Dong Su, Siyi He, Haoyi Tang, Lingjun Liu, Zhonghua Xie Presenter: Jie Dong, Huizhou University, China Abstract: As the core research direction of computer vision technology, object detection performance optimization under complex meteorological conditions has become the focus of common attention in industry and academia. Although YOLO series models show excellent detection efficiency and accuracy in conventional scenes, their performance in rainy environments still faces significant challenges, mainly due to edge blurring, texture information loss and contrast reduction caused by precipitation interference. In this study, an image preprocessing framework based on multi-model fusion is proposed, which combines the detail recovery capability of MPRNet, the long-range dependent modeling feature of Restormer, and the multi-scale feature fusion mechanism of MFDNet to build an enhanced preprocessing process for rain and fog degradation. Through comparative experiments on the standard test set, the indexes such as mAP@0.5, mAP@0.5:0.9 and Precision are used for quantitative evaluation. This research provides a scalable technical solution for visual perception system under bad weather conditions, and has practical significance for night application scenarios such as automatic driving and intelligent security
13:30-13:45 PA497	Title: Brain Tumor Detection Using Advanced Deep Learning Models - Vision Transformer and EfficientNetV2 Author(s): Yang Liu and Li-Jun Yu Presenter: Yang Liu, Xi'an Jiaotong-Liverpool University, China

Abstract: Advanced deep learning model - EfficientNetV2 is reported that the model has achieved a better performance than the Vision Transformer (ViT) for some simple tests in the original research recently. However, it is still now clear whether the Efficient NetV2 performs better than the ViT for practical applications. We will study the two advanced deep learning models for brain tumor image detection. In addition, we will use one robust machine learning model – eXtreme Gradient Boosting (XGBoost) as benchmark to study how much the two deep learning models perform better than the benchmark.

13:45-14:00
PA512

Title: GTWNet: A High-Accuracy and Lightweight Framework for Rice Pest Detection

Author(s): Shilei Deng, Gaowei Wu. Wensheng Zhang

Presenter: Shilei Deng, Hainan University Haikou, China

Abstract: Rice is one of the most essential staple crops worldwide, with its yield and quality critical to global food security. However, rice is highly susceptible to pest infestations during its growth cycle, which severely impact both yield and quality. To address this, we propose GTW Net, an innovative lightweight framework designed to achieve high-precision pest detection while optimizing computational efficiency. Specifically, GTW Net integrates Ghost Conv and C3Ghost modules for efficient linear feature extraction, significantly accelerating both training and inference. The architecture incorporates a specialized Tiny-Neck structure to enhance the model's attention to key pest features and improve multi-scale feature fusion efficiency. Furthermore, the Wise-IoU loss function is introduced to optimize label prioritization by adaptively balancing gradients from medium quality samples while suppressing low-quality outliers, thereby accelerating model convergence and improving detection robustness. Extensive experiments conducted on the self-built dataset IP13 (containing 13 major rice pests) demonstrate GTW Net's superiority over mainstream detectors, achieving state-of-the-art metrics of 72.4% precision (+4.4%), 71.4% recall (+1%), 74.9% mAP@50 (+1.5%), and 50.2% mAP@50-95 (+1.8%). Notably, GTW Net reduces parameters to 7.9M (-29.3%) and computational costs to 16.5G FLOPs (-42.5%), establishing an optimal accuracy-efficiency tradeoff for rice pest detection. Consequently, GTW Net delivers an accurate, lightweight solution for real-time rice pest identification, enabling reliable crop health monitoring and yield optimization.

14:00-14:15
PA527

Title: Unified Face Attack Detection via Multi-scale CNN-ViT andGraph Convolutional Networks

Author(s): Xinyu Chen, JunLiu Zengxi Huang, Dazhi Zhang

Presenter: Xinyu Chen, Xihua University, China

Abstract: Face recognition (FR) systems are integral to applications like access control and online payments but face significant threats from diverse physical and digital attacks. Existing face anti-attack methods often target limited attack types, rendering them less effective against composite threats in real-world scenarios. To address this, we propose Uni GCT, a unified face attack detection framework that integrates a dual-stream CNN-ViT architecture with linear attention and Graph Convolutional Networks (GCN). The multi-scale CNN-ViT stream effectively extracts and fuses diverse attack features, while the GCN-based stream models higher-order semantic correlations among attack types, providing cross-category discriminative prior knowledge. By combining these streams through matrix multiplication, Uni GCT achieves robust feature integration and enhances detection performance. Experimental results show that Uni GCT outperforms state-of-the-art methods, significantly reducing miss detection rates while maintaining computational efficiency, offering a scalable and effective solution for securing FR systems against evolving threats.

14:15-14:30
PA531

Title: A Lightweight Object Detection Algorithm for Complex Traffic Environments

Author(s): Yi Wang, Nian Jia, Zengxi Huang, Ziping Liu

Presenter: Yi Wang, Xihua University, China

Abstract: Robust vehicle detection is paramount for urban traffic management in Intelligent Transportation Systems (ITS). However, practical deployment faces significant challenges: dense vehicle clusters, diverse illumination, and the need for lightweight algorithms on embedded systems. While advanced detectors like the YOLO series have improved inference speed and accuracy, balancing high detection performance with stringent model lightweighting in complex traffic remains a critical problem.

This paper introduces SSL-YOLO, a novel lightweight object detection model designed for challenging traffic scenarios. Our approach integrates StarNet as the backbone, leveraging its high-dimensional feature mapping to enhance feature extraction without increasing computational complexity. The C3K2_Star module further optimizes feature localization and reduces overhead. To address occlusion, a Spatial Enhancement Attention Module (SEAM) is embedded within the detection head, mitigating response loss via spatial and multi-scale feature fusion.

On the UA-DETRAC dataset, SSL-YOLO demonstrates superior performance: 30\% fewer parameters than YOLOv11, while simultaneously boosting mAP@0.5 by 5.2\% and mAP@0.5:0.95 by 3.5\%. This validates SSL-YOLO's effectiveness in balancing detection accuracy and model lightweighting for real-world applications.

Break Time

15:00-15:15
PA471

Title: Power Line Foreign Object Invasion Detection Algorithm Based on RGB-D Multimodal Images

Author(s): Jianfeng Wu, Yaosheng Huang, Shangbing Yang

Presenter: Yaosheng Huang, Jiangmen Power Supply Bureau of Guangdong Power Grid Co., Ltd, China

Abstract: With the rapid expansion of global power infrastruc ture and the continuous increase in power demand, the intelligent operation and maintenance of power grids have become crucial. Transmission lines, as the backbone of the power system, directly affect the reliability of power supply and the stability of the power grid. Foreign object intrusion on power lines is one of the most serious threats to power transmission safety, which may lead to power outages, equipment damage, and even catastrophic failures. The traditional manual inspection method is inefficient and costly, and it is difficult to meet the requirements of real-time monitoring. To improve the detection efficiency and accuracy, this study proposes a multimodal power line foreign object intrusion detection network based on the improved DETR. Firstly, based on the DETR framework, the effective layers of RGB and depth features are extracted by using the improved symmetric backbone network structure, and the RGB-D data feature fusion module is introduced to effectively extract the feature representations of RGB and depth images. Meanwhile, a 3D position coding combined with depth information is designed to achieve feature interaction in the spatial and channel dimensions. Aiming at the problem of high computational complexity of the DETR model, by optimizing the encoder-decoder structure and attention mechanism of the Transformer, the computational overhead has been significantly reduced. Experiments on the self-built foreign object dataset of power lines show that the detection accuracy rate of the improved model reaches 95.12% , demonstrating superior performance in complex background and small target detection scenarios. The method proposed in this study not only improves the detection accuracy, but also has good engineering applicability, providing an effective technical solution for the intelligent inspection of power lines.

15:15-15:30
PA705

Title: Improved Hyperdimensional Computing Strategies for Epileptic Seizure Detection

Author(s): Zhen Zhang

Presenter: Zhen Zhang, Huizhou University, China

Abstract: Epileptic seizure detection plays a pivotal role in improving the quality of life for epilepsy patients. However, the dearth of practical EEG-based signal detection tools for usual use is a significant challenge, which necessitates the development or enhancement of existing detection

methods for epileptic seizures. Hyperdimensional computing, a category of neuromorphic computing methodologies, has recently been employed in various signal processing tasks due to its unique characteristics. Nevertheless, in seizure tasks, the efficacy of standard hyperdimensional computing falls short of other sophisticated algorithms, especially in imbalance datasets.

15:30-15:45
PA205

Title: Intelligent warning system for the elderly living alone Based on Open Pose and YOLOv8

Author(s): Zhen Zhang

Presenter: Zhen Zhang, Huizhou University, China

Abstract: This study replicates the standard hyperdimensional computing method and its subsequent key improvement studies, introducing four enhanced hyperdimensional computing methods following a strategic combination. These methods are refined under identical preconditions and evaluated against the same benchmarks as advanced machine learning methods. The results demonstrate that one of the enhanced iterative compressed multi-Class hyperdimensional computing methods can achieve a median performance value of 94.3% on an unbalanced CHB-MIT dataset which simulates a real seizure scenario. The above result surpasses the random forest method used as a benchmark. This research underscores the potential of hyperdimensional computing in epileptic seizure detection, setting the stage for future investigations and practical implementations.

15:45-16:00
PA529

Title: FusionTrack: An Adaptive Hybrid Tracking Framework Overcoming Rapid Object Motion, Frequent Occlusions, Dynamic Scene Transitions, and Appearance-Scale Variations

Author(s): Akhil Penta, Vaibhav Adwani, Ankush Chopra

Presenter: Akhil Penta, Tredence Analytics, Bangalore, India.

Abstract: Visual tracking in dynamic winter sports presents unique challenges, including fast athlete movements, intricate multi-camera setups, and various occlusions. We introduce FusionTrack, a novel single-object tracking (SOT) framework. It provides reliable skier tracking in demanding broadcast sports videos and similar datasets. Our approach strategically adapts multi-object tracking (MOT) principles, such as appearance-based re-identification and intelligent detection fusion, into a specialized Tracking-by-Detection (TbD) pipeline for single skier. Moving beyond traditional Kalman filter-based methods, our main contributions include: 1) A deep LSTM sequence model for improved bounding box predictions, enhancing temporal motion modeling; 2) A novel adaptive association strategy that thoughtfully switches from a harmonic mean blend of motion and appearance cues to solely an appearance-based metric during sudden camera changes, detected via simple pixel-wise scene change; and 3) A filtered Exponential Moving Average (EMA) for appearance embeddings, which maintains identity consistency despite noise or occlusions. Tested extensively on the challenging SkiTB dataset, FusionTrack achieved an average F1-score of 88.5%, outperforming the previous best of 87.7%. For the Freestyle Skiing (FS) discipline, our F1-score reached 84.9%, against an existing best of 78.9%. This represents an overall absolute improvement of 8% compared to our team's earlier transformer-based baseline, which had an average F1-score of 80.5%.

16:15-16:30
PA521

Title: Impact of Rural Virtual Tourism Experience on On-SiteTravel Intention Mediatied by Place Attachment

Author(s): Hua Dong, Qiang Wei

Presenter: Hua Dong, Jianghan University, China

Abstract: This study examined how rural virtual tourism experiences in-fluence on-site travel intentions, focusing on place attachment's mediating role. Using social cognitive theory and the SOR framework (N=393), we found all four experience dimensions (entertainment, education, aesthetics, escapism) directly in-creased travel intentions. Escapism uniquely enhanced place attachment (β =0.367,

p<0.001), mediating its effect (indirect effect=0.175, 95% CI [0.073, 0.283]). Virtual tourism creates affective bonds that convert digital experiences into physical travel actions.

16:30-16:45
PA703

Title: Two-stage Feature Selection Algorithm Based on an Improved Grey Predictive Evolutionary Algorithm for Depression Detection

Author(s): Mengyuan Li, Shuhua Mao

Presenter: Shuhua Mao, Wuhan University of Technology, China

Abstract: Depression has emerged as a high-profile mental health issue, particularly among students. This study obtained multisource heterogeneous data from the Kaggle open-source platform. During the data preprocessing phase, an effective data dimensionality reduction method was proposed: a two stage feature selection approach combining significance statistics and an Improved Binary Grey Predictive Evolutionary Algorithm (IBGPEA). The proposed method employs a feature hierarchy mechanism in the first stage by integrating statistical concepts with the 'jump degree" principle, while the second stage introduces a novel feature selection strategy based on grey predictive evolutionary algorithm to enhance prediction accuracy and accelerate convergence. Based on the selected features, a depression risk prediction model was constructed using Tab Net. The model leverages attention mechanisms and sequential decision-making processes to effectively capture complex relationships among features. Interpretability analysis reveals that the feature "Have you ever had suicidal thoughts' demonstrates significant contribution to depression risk prediction. Compared with Random Forest, SVM, KNN, Multi-Layer Perceptron, and Adaboost, the Tab Net model shows remarkable advantages in various performance indicators, enabling more accurate and reliable prediction of depression risk.

Online Oral Sessions

ON02: Language Models and Semantic Analysis

13:00-17:00| Aug.17, 2025 | Room C -849 3724 5301 (Password: 081517)

Talk Details	
Time	**Presentation**
13:00-13:15 **PA643-A**	Title: A Two-Stage Framework for Legal Summarization via Case-Based Retrieval and Structural Refinement Author(s): Yulin Zhou, Yongbin Qin, Ruizhang Huang Presenter: Yulin Zhou, Guizhou University, China Abstract: Automatic summarization of judicial documents is essential for enhancing the efficiency and accessibility of legal information processing. Yet, current approaches often neglect the interpretative value of precedent cases and fail to utilize the structured nature of legal texts. In this paper, we introduce CaReS, a novel two-stage summarization framework that seamlessly integrates Case-Based Retrieval and Structural Refinement. In the first stage, the model retrieves the most semantically relevant precedent case and uses it as a one-shot exemplar to prompt a large language model (LLM) for generating an initial summary of the target legal document. In the second stage, the initial summary is evaluated against the document's predefined structural elements-such as parties' claims, legal reasoning, and judgment-using semantic similarity measures. If any key components are missing or underrepresented, the system dynamically supplements the content through guided prompting to produce a complete and coherent final summary. Experiments on legal summarization datasets demonstrate the effectiveness of CaReS in generating structurally consistent and legally meaningful summaries, highlighting its value in legal natural language processing applications.
13:15-13:30 **PA687**	Title: An Assisted Medical Diagnosis Question Answering System Based on Knowledge Graphs Author(s): Liu Fang, Chen Yang and Chen Yingjie Presenter: Liu Fang, Neusoft Institute Guangdong, China Abstract: With the rapid advancement of artificial intelligence technology, knowledge graphs are seeing increasingly widespread application in the medical field, offering new technological means for assisted medical diagnosis. However, current research still faces challenges such as difficulties in natural language understanding and insufficient knowledge reasoning capabilities. This paper designs and implements a knowledge graph-based question answering system for assisted medical diagnosis, aiming to enhance diagnostic efficiency and accuracy through structured medical knowledge. The system first uses natural language processing techniques to analyze user-input medical questions, and then performs semantic reasoning by leveraging entities such as diseases, symptoms, medications, and medical tests, along with their relationships in a medical knowledge graph, to generate reliable diagnostic suggestions. Through knowledge graph embedding and relationship path mining, the system achieves multi-hop reasoning and ensures the scientific validity of responses based on authoritative medical databases. Experimental results demonstrate that this system achieves high accuracy and practicality in diagnostic question answering tasks for common diseases. It can provide efficient decision support assistance for both doctors and patients while reducing the risk of misdiagnosis.

Title: Sharpening Semantic Gradient in a Planarized Sentence Representation

Author(s): Caiwei Yang, Yanping Chen and Shuai Yu

Presenter: Caiwei Yang, Guizhou University, China

13:30-13:45
PA644-A

Abstract: Mapping a sentence into a two-dimensional representation has the advantage of unfolding nested semantic structures in a sentence and encoding the interaction between tokens. In the planarized sentence representation, neighboring elements denote overlapped linguistic units in a sentence. An important phenomenon is that the semantic information of a true linguistic unit may penetrate neighboring elements, which blurs the semantic edge of a linguistic unit and disturbs the planarized sentence representation. Therefore, sharpening the semantic gradient helps aggravate semantic information from neighborhoods and depressing noises in neighboring elements. This paper reveals the mechanism of sharpening semantic gradient in the planarized sentence representation. Our method is evaluated on six evaluation datasets. The results show impressive improvement on three information extraction tasks. The success indicates that representing and processing sentences in a two-dimensional representation has a great potential to decode the sentential semantic structure and support sentence-level information extraction.

Title: PMDS: Progressive Multi-Document Summarization with Iterative Summary Integration

Author(s): Shuai Yu, Caiwei Yang, Yongbin Qin and Ruizhang Huang

Presenter: Shuai Yu, Guizhou University, China

13:45-14:00
PA645-A

Abstract: The proliferation of textual information in the digital age has made multi-document summarization (MDS) a critical tool for efficient information access. Traditional MDS approaches often struggle with input length constraints, redundancy, and coherence. In this work, we propose a novel sequential summarization paradigm: instead of generating a summary for the entire document set simultaneously, we iteratively summarize documents by integrating the current document with the previously generated summary. This progressive strategy enables incremental synthesis and alleviates token-length bottlenecks in large language models. To mitigate error accumulation and preserve factual consistency, we introduce a modular framework based on a fine-tuned pre-trained language model, augmented with lightweight auxiliary models for content selection and verification. We conduct extensive experiments on nine public MDS datasets spanning diverse domains (e.g., news, scientific literature, legal texts, clinical trials), evaluated by seven automatic metrics (ROUGE-1/2/L, BERT Score, FactCC, BLEU, and BARTScore) and human assessments. Our method achieves new state-of-the-art results on multiple benchmarks. Ablation studies confirm the effectiveness of each module, and zero-/few-shot evaluations on unseen datasets demonstrate strong generalization ability. Our findings establish the effectiveness and scalability of progressive multi-document summarization for synthesizing coherent and informative summaries across domains.

Title: Dynamic Document Clustering through Topic-Level Continuity and Cluster Propagation

Author(s): Jingjing Xue, Ruina Bai and Ruizhang Huang

Presenter: Jingjing Xue, Guizhou University, China

14:00-14:15
PA701-A

Abstract: Existing deep document clustering methods are mainly designed for static data and rarely consider temporal dynamics. Even in dynamic settings, prior approaches typically extract topic information at the document level and then perform clustering, which is inefficient and overlooks the continuity of topic semantics. To address this, we propose a dynamic clustering framework that propagates topic-level information across time slices. By learning and transferring cluster-level representations over time, our method enhances clustering quality and enables better tracking of topic evolution and emerging trends. Experiments on real-world datasets demonstrate the effectiveness of our approach.

14:15-14:30 PA673-A	Title: A Resilient Generative Model in Few-shot Question Answering Author(s): Anqi Zou, Yongbin Qin and Yanping Chen Presenter: Anqi Zou, Guizhou University, China Abstract: In few-shot question answering (QA), only a limited number of training cases are available to tune pre-trained language models (PLMs) for fitting to the task objective. Because a PLM often contains a large number of parameters optimized by a pretraining objective, it is difficult to bridge the gap between the two optimization objectives with a few training cases, which leads to the overfitting problem. Furthermore, in the decoding process of a generative model, sometimes rebuilding an answer to a question is more challenging than directly choosing from an input, because the ability to rebuild requires more informative knowledge and may generate hallucinatory outputs. Motivated by the above discussions, we propose a resilient deep model to support the few-shot question answering tasks. Instead of ambitiously optimizing a full generative model with a few training cases, the resilient model contains a partial learning module and an answer choosing module to reduce the problems caused by a lack of training cases. This has the advantage to enhance the few-shot learning in question answering. Our method was evaluated using several few-shot QA public evaluation datasets. It achieves a state-of-the-art performance, considerably outperforming related works.
Break Time	
15:00-15:15 PA496	Title: Research of a Medical Knowledge Extraction Algorithm Based on Graph Neural Networks Author(s): Hongxia Liu, Hui Tan, Hui Li, Xiangben Hu Presenter: Xiangben Hu, Chizhou Vocational and Technical College, China Abstract: With the breakthrough research of deep learning in various fields, medical knowledge extraction has also seen new developments. However, unstructured utterances have some difficulties in medical knowledge extraction due to the complex semantic information. In addition, the overlapping of entity relations reduces the accuracy of knowledge ex traction. To solve the aforementioned issues, a graph neural network-based medical knowledge extraction algorithm named MKEAGNN is proposed in this paper. Specifically, the model captures semantically dependent structural information and sequence information of sentences by combining graph convolutional neural network (GCN) and bidirectional long short-term memory neural network (Bi-LSTM). Then, pruning operations are introduced to eliminate the noise in the depen dency tree. In addition, the multi-head attention mechanism is used to assign different weights to the entities. Finally, the experiments compare the method proposed in this paper with LSTM, BiLSTM, BiLSTM-GCN. Based on the experimental results, the proposed method in this paper achieves 89.15% in terms of knowledge extraction accuracy. This result shows the effectiveness and usefulness of the MKEAGNN algorithm. Index Terms-Knowledge Extraction; GCN; Unstructured utterances; Bi-LSTM; Semantic Information
15:15-15:30 PA516	Title: Incremental Multi-View Clustering with Knowledge Distillation and Contrastive Learning Author(s): Hong Peng Chen, Hong Wei Yin and Wen Jun Hu Presenter: Hongpeng Chen, Huzhou University, China Abstract: The prevalence of multiview data in real-world applications highlights the need to effectively capture interview correlations for tasks like clustering. While most existing Multiview clustering methods assume simultaneous access to all views, practical scenarios often involve views arriving sequentially due to delays, storage, or privacy constraints. To address this issue, an Incremental Multiview clustering with knowledge distillation and contrastive learning IMKC is proposed in this paper. IMKC incrementally integrates new views by distilling semantic knowledge from previous ones, using a cross view distillation module to retain valuable information and filter noise. A contrastive learning strategy further enhances consistency between history and common representations. Experimental results on five benchmark datasets show that our method outperforms existing incremental MVC methods.

Title: Temporal and Causal-Aware Summarization of Legal Public Opinion with Large Language Models

Author(s): Wei Zhang, Yang Kou

Presenter: Wei Zhang, Guizhou Tobacco Company, China

15:30-15:45
PA684

Abstract: Automatic summarization of legal public opinion texts poses unique challenges due to complex temporal dynamics and nuanced causal relationships inherent in legal events discussed in news and social media. Existing summarization approaches, including powerful large language models (LLMs), often fail to capture these critical temporal sequences and causal links, resulting in summaries that lack coherence and narrative logic. To address this, we propose a novel summarization framework that explicitly integrates temporal and causal relationship modeling into an LLM-based architecture, specifically fine-tuning the Qwen language model with extracted event timelines and causal connections. Evaluations conducted on the publicly available CAIL-2022 legal public opinion dataset, demonstrate that our proposed method significantly outperforms traditional summarizers and baseline LLMs across metrics such as ROUGE-1, ROUGE-2, ROUGE-L and BERTScore. Experimental analysis further confirms that the generated summaries exhibit enhanced chronological coherence and accurately reflect causal event structures. This study provides compelling evidence that embedding temporal and causal reasoning into LLM-based summarization substantially improves the quality and informativeness of summaries in the legal public opinion domain, offering promising directions for future research and practical applications in public opinion analysis.

Title: AG-Net: Category-level 6D Pose Estimation Network Based on Attention Mechanism and Global Enhancement

Author(s): Bowen Shi, Jinlong Shim Suqin Baim Ruixue Gong

Presenter: Bowen Shi, Jiangsu University of Science and Technology, China

15:45-16:00
PA204

Abstract: How to effectively enhance feature extraction is a challenge faced by current 6D pose estimation methods. To address this issue, we propose a novel 6D pose estimation network based on VI-Net, shortened as AG-Net, which uses ECA_block and Global enhancev module to enhance feature extraction: ECA_block embeds a channel attention mechanism into the convolutional layers, replacing the fully connected layers with 1x1Conv to capture relationships between different channels and improves the performance of feature extraction. The Global enhancev module further processes the received information and enhances feature extraction by fusing global features, effectively balancing performance and speed, and better estimating the translation and size of objects. We applied the proposed AG-Net to category-level 6D pose estimation tasks and tested it on the REAL275 and CAMER25 datasets using IOU 3D intersection and n◦mcm evaluation metrics. The results showed that AG-Net outperformed current state-of-the-art methods. Our code and models are available at https://github.com/AFESDTTM/AG-Net Index Terms-6D pose estimation, RGB-D information, feature extraction, attention mechanism

Title: Chest X-ray Diagnosis with Joint Report Generation and Lesion Localization Using Multimodal Vision–Language Models

Author(s): Houxuan Zhang, Kuan Huang

Presenter: Houxuan Zhang, Kean University, USA

16:15-16:30
PA693

Abstract: Lung diseases remain a major global health concern, and chest X-ray imaging is a critical tool for early diagnosis and treatment. To support radiologists, automatic computer-aided diagnosis (CAD) systems must detect abnormalities and accurately localize them within the image. However, developing such systems faces key challenges: annotated bounding box labels are scarce, and many models lack clear interpretability and visual grounding to support clinical trust. To address these gaps, we propose a unified multimodal diagnostic framework that combines a DenseNet-121 classifier with Grad-CAM-based pseudo-labeling and a powerful vision-language model (Qwen2.5-VL-7B-Instruct). The DenseNet classifier is trained on pneumonia and pneumothorax cases from the MIMIC-CXR-JPG dataset, with Grad-CAM used to extract pseudo bounding boxes by thresholding activation maps. These pseudo-labels, combined with limited expert annotations, supervise the fine-tuning of Qwen2.5-VL, enabling it to generate structured radiology reports alongside spatial coordinates for pathological findings. Experimental results demonstrate that incorporating pseudo-labeled data significantly boosts localization performance, achieving 26.09% recall at IoU >= 0.25 compared to 4.35% using real labels alone. Meanwhile, report generation quality remains strong, with BLEU-4 and ROUGE-L scores comparable to state-of-the-art systems and with faster inference than models like LLaVA. This work illustrates how Grad-CAM-based pseudo-labeling and multimodal vision-language models can enhance interpretability and efficiency, providing a practical pathway toward integrated, explainable AI for chest X-ray diagnosis.

Title: An Interpretable Image Classification Approach Using Prototype-Based Deep Belief Rules

Author(s): Jiawei Wu, Lianmeng Jiao and Quan Pan

Presenter: Jiawei Wu, Northwestern Polytechnical University, China

16:30-16:45
PA224

Abstract: Belief rule-based systems, despite their interpretability in classification tasks, face two critical limitations in image analysis: the curse of dimensionality when handling high dimension data and incompatibility with non-tabular formats, which restrict their applicability to interpretable image classification applications. To address these challenges, we propose a prototype-based deep belief rule reasoning methodology for interpretable image classification. The core methodology involves replacing decision layer in deep neural networks with a deep belief rule base, enabling joint optimization of classification performance and decision transparency. The deep belief rule base construction comprises two critical phases: anchoring initial class boundaries through belief cluster algorithm based on decision graph, and generating comprehensive deep belief rule base, in which the basic principle for constructing the deep belief rule base is to rely on the distance between prototypes, instead of distinguishing the actual categories of prototypes. We conduct extensive experiments on multiple benchmark datasets, demonstrating that our approach achieves superior interpretability while maintaining high classification performance.

Title: Boundary Recognition System for Intelligent Lawn Mowing Robot Based on mmWave radar

Author(s): Aiqing Wang, Yu Qin, Zhiguo Shi and Yong Wang

Presenter: Aiqing Wang, College of Engineers Zhejiang University, China

16:45-17:00
PA653

Abstract: To address the critical challenges of low boundary recognition accuracy, high hardware costs, and computational complexity in wire-free intelligent lawn mowing robot operating in complex environments, this paper proposes a novel boundary detection solution utilizing cost-effective 60 GHz mmWave radar. The system implements a multipath suppression strategy through non-coherent accumulation of 4-channel receiver signals, effectively suppressing noise in the acquired radar data. By

fusing multidimensional features, the proposed method significantly enhances the distinguishability between grass, water surfaces, and various types of hard pavements. A lightweight random forest classifier was developed, achieving 97.7% material recognition accuracy on the training set and 94.0% on the test set for five material types under various environmental conditions. This work demonstrates the engineering feasibility of combining millimeter-wave radar with machine learning for boundary detection, offering advantages in cost-efficiency, high precision, and computational lightness. The proposed system provides crucial technical support for the evolution of smart lawn mowers from physically restricted operation to fully autonomous navigation.

Online Oral Sessions

ON03: Deep Learning, Vision & Intelligent Optimization Applications

13:00-16:45 | Aug.17, 2025 | Room D - 840 5310 5661 (Password: 081517)

Talk Details	
Time	**Presentation**
13:00-13:15 PA462-A	Title: Guided 3D Generation through Depth Image Fusion Utilizing Diffusion Model Author(s): Yifan Wei, Xingwang Zhao Presenter: Yifan Wei, Guizhou University, China Abstract: Recent advances in text-to-3D generation have achieved impressive results, yet maintaining multi-view consistency remains a significant challenge. Existing methods often struggle with geometric coherence due to the limitations of 2D based priors or the scarcity of high-quality 3D training data. In this work, we introduce a deph-guided text-to-3D generation framework that leverages depth maps as structural priors to enhance 3D consistency. Depth maps naturally encode spatial information, serving as an effective constraint to mitigate geometric inconsistencies across different viewpoints. Our approach integrates depth priors into the optimization process, complementing 3D representations such as Gaussian Splatting or Neural Radiance Fields (NeRF). By refining geometry through depth guided constraints and enhancing details with 2D diffusion priors, our method significantly improves both structural integrity and multi-view consistency in generated 3D assets. Experimental results demonstrate that our approach achieves high-fidelity 3D reconstructions with improved coherence across different perspectives, providing a promising direction for efficient and high-quality text-to-3D synthesis.
13:15-13:30 PA479	Title: GrapeCPNet: A Data-Driven Deep Learning Approach for Automated 3D Phenotyping and Predictive Analytics in Grape Cultivation Author(s): Sumit Mamtani Presenter: Sumit Mamtani, New York University, USA Abstract: Accurate phenotypic analysis of grape bunches at the berry level is crucial for yield estimation and quality control in vineyard management. Traditional manual measurement methods are labor-intensive, time-consuming, and prone to errors. Advances in data analytics, deep learning, and 3D computer vision have enabled automated, high-throughput phenotyping solutions. This study introduces GrapeCPNet, a self- supervised deep learning pipeline that integrates 3D point cloud analytics with instance segmentation and shape completion to reconstruct and analyze individual grape berries. The proposed method employs the SoftGroup deep learning network for berry segmentation, followed by a novel self-supervised point cloud completion network that reconstructs missing berry surfaces caused by occlusions. Morphological traits such as berry count, radius, and volume are extracted using data-driven statistical modeling. Experimental validation on four grape varieties achieved over 85% R2 for radius prediction and 97% for volume estimation, demonstrating the effectiveness of the proposed analytics framework in agricultural phenotyping. This study highlights the potential of AI-powered data analytics and predictive modeling in precision agriculture, enabling more accurate yield forecasts and improved crop management strategies. While validated on four table grape varieties in controlled indoor environments, further study is needed to confirm generalizability to outdoor and vineyard conditions.

Title: An Ecological Footprint Image Dataset for Automatic Identification of Paleontology Discovery

Author(s): Yinghao Cai, Shaoning Zeng

Presenter: Yinghao Cai, South China Normal University, China

13:30-13:45
PA208

Abstract: Paleontological relics are precious and rare, and non-specialist paleontological explorers can bring a wealth of information to the field. However, there are quite a few of shortcomings, including misdiagnosis and corruption of information. Until now, no image dataset is collected to assist the recognition of ecological footprints, especially using the machine learning techniques. We initially collect an ecological dataset, containing two kinds of images with 2385 samples from the Internet, as well as the scientific publications of paleontology for computer vision. Furthermore, we utilize the Frequency-tuned Salient Region Detection (FT) and Local Contrast based Salient Region Detection (LC) algorithms to preprocess the data. These methodologies facilitate the identification of the striking parts of an image to help the recognition algorithms. Subsequently, we deploy six classical deep neural networks for the purpose of classification, which have garnered extensive application across a spectrum of image recognition and processing tasks. As a result, the dataset can be effectively used scientifically for automated identification.

Title: Multi-Objective Optimization Model and Enhanced Genetic Algorithm for Fresh Food E-Commerce Distribution under Time Window Constraints

Author(s): Fang Fugui, Yang Jian

Presenter: Fang Fugui, Chengdu Jincheng Coll, Sch Comp & Software, China

13:45-14:00
PA663

Abstract: Driven by increasing consumer demand for fresh food products and elevated expectations for fast, high-quality delivery, efficient logistics systems have become a critical challenge for the fresh food e-commerce industry. This study focuses on the Vehicle Routing Problem with Time Windows (VRPTW) in the context of fresh food e-commerce distribution. A multi-objective optimization model is developed to simultaneously minimize total costs, including distribution expenses, fixed vehicle costs, penalties for time window violations, and costs related to product freshness deterioration. The key constraints considered are vehicle capacity limits and customer-specific delivery time windows. To address this complex logistical problem, an enhanced Genetic Algorithm (GA) integrated with an elite preservationstrategy is proposed. Simulation experiments validate the effectiveness of the model and algorithm. The results show that the optimized solution achieves a substantial 19.18% reduction in total distribution costs compared to baseline scenarios. The proposed approach demonstrates strong performance and practical applicability, offering a promising solution for improving logistics efficiency in the fresh food e-commerce sector.

Title: An Intelligent Approach for Arch Height Index Measurement Based on Machine Vision

Author(s): Wenjie Chen, Xiaoming Li and Zhengrui Zhang

Presenter: Zhen Zhang, Huizhou University, China

14:00-14:15
PA704

Abstract: The arch height is a crucial indicator for assessing foot health, and the Arch Height Index (AHI) is a widely recognized method for measuring arch height. The current AHI measurement method mainly involves using calipers to measure the foot, but this method has issues such as complex operation and difficulty in implementation. Therefore, this study proposes a new automatic AHI measurement method based on image segmentation. This method uses a semantic segmentation model to extract key structures of foot images, automatically obtains key parameters such as arch height and truncated foot length, and directly calculates the AHI value. This study collected 46 side images of the foot to verify the algorithm. By comparing with the measurement results of professional instruments, it was found that the AHI value calculated by this intelligent method and the actual AHI value had a Pearson correlation coefficient of 0.9911, and the p-value was calculated by t-test to be 0.945, which was greater than the significance level of 0.05, indicating that the AHI value calculated by

this intelligent algorithm is very close to the real AHI value. The above results show that the AHI intelligent measurement method based on image segmentation proposed in this study can effectively obtain the AHI value of the foot, providing a reliable basis for clinical diagnosis and treatment. This study provides a new method for the AHI measurement field, which has the advantages of simple operation and high accuracy, and is worth further promotion and application in the field of telemedicine.

14:15-14:30 **PA476**	Title: Multi-scale feature fusion UNetFormer: Improving the accuracy of remote sensing urban scene image segmentation Author(s): Shichun He, Xiaodong Yu, Desong Chen, Maolin Zhang and Xi Kuang Presenter: Shichun He, Chengdu University, China Abstract: Due to the diversity and complexity of objects in remote sensing images, remote sensing semantic segmentation still faces many challenges. This study improves the UNet Former network structure and proposes an enhanced remote sensing image semantic segmentation model. In view of the limitations of UNet Former in the feature fusion stage, an improved feature fusion mechanism is designed and introduced, which effectively improves the model's ability to express multi-scale complex object features. Experimental results show that on the remote sensing dataset ISPRS Vaihingen, the proposed method achieves 83.6% performance, significantly outperforming the original UNet Former and other existing lightweight models. In addition, while maintaining a high inference efficiency, the improved model further improves the segmentation accuracy of complex urban scenes, verifying its effectiveness and feasibility in practical applications.

Break Time

15:00-15:15 **PA217**	Title: Optimization of sepsis drug delivery based on continuous state space deep reinforcement learning Author(s): Zhiyu Ren, Yiwei Liu, Peiyu Fang, Mengting Ni Presenter: Zhiyu Ren, Beijing University of Posts and Telecommunications Beijing, China Abstract: Sepsis is one of the leading causes of death worldwide, and optimization of its treatment strategies is essential to improve patient outcomes. Treating sepsis is extremely challenging because individual patients respond very differently to medical interventions, and there is no universally recognized treatment for sepsis. In this work, we present a novel approach to derive optimal treatment strategies for patients with sepsis through the use of continuous state space models and deep reinforcement learning. By introducing Norepinephrine Equivalent Dose (NEE) as a unified quantitative standard, the model effectively integrates the dose adjustment of multiple vasoactive drugs to achieve precise treatment in complex and variable clinical scenarios. We demonstrate that, on average, the treatments recommended by the model are more valuable and reliable than those recommended by clinicians. In a large validation dataset, we found that patients whose clinician's actual dose matched the AI decision had the lowest mortality. Our model provides personalized and clinically explainable treatment decisions for sepsis to improve patient care
15:15-15:30 **PA494**	Title: A Novel Task Allocation Method for UAVs to Maximize the Number of Assigned Tasks Author(s): Jihua Xu Presenter: Jihua Xu, Jiangsu Automation Research Institute, China Abstract: The inherent flexibility of Unmanned Aerial Vehicles (UAVs) combined with their line-of-sight (LoS) communication capabilities has established them as an effective solution for emergency service provision in numerous practical applications. This paper proposes a LEPI algorithm, developed through systematic enhancements to the conventional PI algorithm. The proposed LEPI algorithm comprises three distinct operational phases: 1) task inclusion phase, 2) conflict removal phase, and 3) task reorganization phase, which explicitly incorporates critical constraints including task deadlines and

	UAVs' limited flight endurance. Comprehensive numerical simulations demonstrate that the implemented LEPI algorithm significantly improves task allocation efficiency, achieving an increase in successfully assigned tasks compared to baseline methods under equivalent operational conditions.
15:30-15:45 **PA216**	Title: A Gated Convolutional Network for Autism Spectrum Disorder Classification and Functional Connectivity Analysis Author(s): Hua Zhou; Jianbin Zhong, Liwei Fan, Yan Tang Presenter: Liwei Fan, Software Engineering Institute of Guangzhou, Guangzhou, China Abstract: To address the critical limitation that conventional CNN-based methods fail to adequately exploit the inherent characteristics of functional connectivity matrices constructed from autism spectrum disorder (ASD) data, this study proposes a novel Gated Functional Connectivity Network (GFCNet) framework. It utilizes functional connectivity features from resting-state functional magnetic resonance imaging (rs-fMRI) data for accurate and interpretable ASD classification. GFCNet integrates gated convolutional modules and transfer learning strategies to effectively extract discriminative features from functional connectivity matrices. Additionally, Grad-CAM visualization is employed to identify key brain regions related to ASD classification decisions. Experiments on the multi-site ABIDE dataset show that our method outperforms mainstream CNN architectures, such as ResNet and DenseNet, achieving 73.5% accuracy and 87.3% sensitivity, surpassing several recent ASD studies. Additionally, the model successfully identifies abnormal functional connectivity in key brain regions, including the medial prefrontal cortex (mPFC), posterior cingulate cortex (PCC), and insular cortex, in ASD patients. Further experiments indicate that the Craddock 200 (CC200) atlas provides the best brain parcellation scheme for ASD diagnosis.
15:45-16:00 **PA344**	Title: DFEC former: An iTrans former-Based Model with Variable Differential and Frequency Enhanced Channel Attention for Multivariate Time Series Prediction Author(s): Dear Desong Chen, Xi Yu, Mengji Yang, Xiaodong Yu and Shichun He Presenter: Desong Chen, Chengdu University, China Abstract: Time series forecasting is of crucial importance in optimising resource allocation and decision-making across various domains, including finance, healthcare, and industrial au tomation. However, extant transformer-based models frequently encounter two significant challenges: interference from irrelevant variables, which results in ineffective attention allocation, and insufficient modelling of frequency dependencies, limiting their ability to capture complex patterns in multivariate time series data. These limitations have a substantial impact on the models' performance in real-world forecasting tasks, where variable inter actions and frequency characteristics are of critical importance. In order to address these challenges, we propose DFECformer, an enhanced iTransformer-based model that incorporates dif ferential attention in order to suppress noise and emphasise key features, along with frequency-enhanced channel attention to capture the frequency dependencies between channels. A comparison of experimental results with those obtained from other mainstream forecasting models on the publicly available ECL dataset demonstrates that DFECformer achieves significant improvements over the benchmark iTransformer. The mean squared error (MSE) is reduced by 4.7%, and the mean absolute error (MAE) is reduced by 3.8%, the results have been averaged over all four prediction horizons (96, 128, 336, and 720 steps), thus highlighting the model's consistent improvement across dif ferent time frames. This finding indicates the substantial efficacy of DFECformer in multivariate forecasting tasks, suggesting that it is a promising approach for general multivariate time series forecasting.

16:15-16:30
PA683

Title: Expert-Novice Disparities in Eye Movement Patterns During Primary Flight Training

Author(s): Yuanyuan Zhang, Sicheng Tong, Xiaoguang Zhou, Yuan Zhou and Quanchao Wei

Presenter: Sicheng Tong, Unit 91475 of PLA, China

Abstract: Effective gaze behavior is critical for flight safety, yet novice pilots often develop inefficient habits early in training. This study leveraged high-fidelity simulation and 120Hz eye-tracking to capture gaze patterns of flight instructors (n=19) and trainees (n=40) during climb turns, cruise, and descent turns. Applying n-gram modeling-a core computational technique used in artificial intelligence for sequence analysis- to data from 12 areas of interest (AOIs), we identified quantifiable differences: instructors exhibited highly regular scan sequences (e.g., "attitude indicator - airspeed indicator - attitude indicator" with 20% higher recurrence) and prioritized core instruments (35-48% longer dwell time, d=0.78-1.15), while trainees showed fragmented patterns (43% unique sequences) with over-reliance on immediate feedback instruments (58-68% dwell increase, d=0.69-0.91) and horizon references. These n-gram-derived behavioral signatures provide the essential structured data foundation for developing AI-powered applications, such as real-time gaze assessment algorithms, adaptive virtual instructors for diagnosing scan inefficiencies, and personalized attention training protocols in next-generation aviation curricula.

16:30-16:45
PA694

Title: A Hybrid Simulated Annealing Genetic Algorithm for Bi-Objective Optimization

Author(s): Wenchao Fang

Presenter: Wenchao Fang, Guangdong Polytechnic of Industry and Commerce, China

Abstract: With the development of the economy, enterprises are paying more and more attention to sustainable development. Sustainable development of enterprises requires them to actively undertake corporate social responsibility while obtaining economic benefits. Therefore, achieving the bi-objective optimization of corporate social responsibility and economic benefits has become a current research hotspot. This paper applies the hybrid simulated annealing genetic algorithm to optimize the bi-objective of corporate social responsibility and economic benefits. Corresponding computation results are obtained through numerical simulations and analyzed. The research results show that compared with the standard genetic algorithm, the optimal solution obtained by the hybrid simulated annealing genetic algorithm is better. On this basis, this paper explores the impact of two different cooling strategies, an aggressive-to-conservative cooling schedule strategy and a conservative-to-aggressive cooling schedule strategy, on algorithm solving performance. Simulation results show that the convergence speed of the optimal solution obtained by using the aggressive-to-conservative cooling schedule strategy is faster, but the optimal solution obtained by using the conservative-to-aggressive cooling schedule strategy is better.

Note