CONFERENCE PROGRAM

The 7th International Conference on Control and Computer Vision (ICCCV 2025)

The 5th International Conference on Robotics and Intelligent Systems (ICRIS 2025)

March 28-30, 2025 / Wuhan, China

Sponsored by

湖北工業大學 HUBEI UNIVERSITY OF TECHNOLOGY

Organized by

湖北工業大學 HUBEI UNIVERSITY OF TECHNOLOGY | 机械工程学院 School of Mechanical Engineering | 智能制造产业学院 School of Intelligent Manufacturing Industry

现代制造质量工程湖北省重点实验室 Key Lab Of Modern Manufacture Quality Engineering

Supported by

湖北省仪器仪表学会

湖北省机械工程学会无损检测专业委员会
武汉无损检测学会
中国机械工程学会无损检测分会湖北省无损检测人员培训中心

HBMES

TUTE

天津市自动化学会

IKL 天津市信息传感与智能控制重点实验室

天津市人工智能学会

HIMT 湖北省计量测试技术研究院

智能输送技术与装备
湖北省重点实验室

HBSEI 湖北特种设备检验检测研究院 Hubei Special Equipment Inspection and Testing Institute

湖北汽车工业学院 HUBEI UNIVERSITY OF AUTOMOTIVE TECHNOLOGY

Table of Content

Welcome Message ..... - 5 -
Organizing Committee ..... - 6 -
Organizing Committee ..... - 7 -
Onsite Conference Notice ..... - 8 -
Online Conference Notice ..... - 12 -
Keynote Speakers ..... - 14 -
Invited Speakers ..... - 15 -
Invited Speakers ..... - 16 -
Simple Program ..... - 17 -
Detailed Program ..... - 21 -
Opening Remark ..... - 21 -
Keynote Speech I ..... - 22 -
Keynote Speech II ..... - 23 -
Keynote Speech III ..... - 24 -
Keynote Speech IV ..... - 25 -
Invited Speech I ..... - 27 -
Invited Speech II ..... - 28 -
Invited Speech III ..... - 29 -
Invited Speech IV ..... - 30 -
Invited Speech V ..... - 31 -
Invited Speech VI ..... - 33 -
Invited Speech VII ..... - 34 -
Invited Speech VIII ..... - 35 -
Invited Speech IX ..... - 36 -
Invited Speech X ..... - 37 -
Invited Speech XI ..... - 38 -
Invited Speech XII ..... - 39 -
Keynote Speech V ..... - 40 -
Keynote Speech VI ..... - 41 -
Invited Speech XIII ..... - 42 -
Invited Speech XIV ..... - 43 -
Invited Speech XV ..... - 44 -
Invited Speech XVI ..... - 45 -
Invited Speech XVII ..... - 46 -
Session 1 ..... - 47 -
CC0047 ..... - 47 -
CC0048 ..... - 47 -
CC0053 ..... - 48 -
CC0054 ..... - 48 -
CC0035 ..... - 49 -
CC0043 ..... - 49 -
Session 2 ..... - 50 -

CC0074 ..... - 50 -
CC0076 ..... - 50 -
CC0077 ..... - 51 -
CC2013 ..... - 51 -
CC0040 ..... - 52 -
CC0069 ..... - 52 -
Session 3 ..... - 54 -
CC0067 ..... - 54 -
CC0032 ..... - 54 -
CC0057 ..... - 54 -
CC2007 ..... - 55 -
CC5001 ..... - 55 -
Poster Session ..... - 57 -
CC0017 ..... - 57 -
CC2003 ..... - 57 -
CC0027 ..... - 58 -
CC0033 ..... - 58 -
CC0045 ..... - 58 -
CC0052 ..... - 59 -
CC5002 ..... - 60 -
CC0038 ..... - 60 -
CC0065 ..... - 60 -
CC0072 ..... - 61 -
CC2002 ..... - 61 -
CC2004 ..... - 62 -
CC2005 ..... - 62 -
Session 4 ..... - 63 -
CC0004 ..... - 63 -
CC0012 ..... - 63 -
CC0019 ..... - 64 -
CC0025-A ..... - 64 -
CC5005 ..... - 64 -
Session 5 ..... - 66 -
CC0016 ..... - 66 -
CC0018 ..... - 66 -
CC0021 ..... - 67 -
CC0063 ..... - 67 -
CC0064 ..... - 68 -
Session 6 ..... - 69 -
CC0007 ..... - 69 -
CC0008 ..... - 69 -
CC0031 ..... - 70 -
CC5006 ..... - 70 -
CC0078 ..... - 70 -

Session 7 ..... - 72 -
CC0011 ..... - 72 -
CC0014 ..... - 72 -
CC0015 ..... - 73 -
CC0022 ..... - 73 -
CC0026 ..... - 73 -
CC0061 ..... - 74 -
CC0073 ..... - 74 -
CC2011 ..... - 75 -
Session 8 ..... - 76 -
CC0009 ..... - 76 -
CC0044 ..... - 76 -
CC0071 ..... - 77 -
CC0062 ..... - 77 -
CC0051 ..... - 78 -
CC0049 ..... - 78 -
Notes ..... - 79 -

Welcome Message

On behalf of the organizing committees, it is our honor to extend a heartfelt welcome to all attendees of The 7th International Conference on Control and Computer Vision (ICCCV 2025) and its workshop: The 5th International Conference on Robotics and Intelligent Systems (ICRIS 2025), taking place in Wuhan, China during March 28-30, 2025.

ICCCV 2025 promises to be enriching platforms for the exchange of ideas, insights, and innovations at the intersection of Control and Computer Vision.

After more than one year' s preparation, we received more than 90 submissions from China, Vietnam, India, Indonesia, Peru, Germany and other countries. More than 130 Technical Program Committee Members participated in the review process. Thanks for their great efforts and excellent work.

ICCCV 2025 has 6 keynote speeches, 17 invited speech and 9 technical sessions. We' re confident that over the two days you'll get the theoretical grounding, practical knowledge, and personal contacts that will help you build long-term, profitable and sustainable communication among researchers and practitioners working in a wide variety of scientific areas with a common interest in Control and Computer Vision.

We sincerely would like to thank all the authors as well as the technical program committee members and reviewers. Their high competence, enthusiasm, time and expertise knowledge enabled us to prepare the high-quality final program and helped to make the conference become a successful event.

ICCCV 2025 General Chair
Prof. Xiaochun Song
Hubei University of Technology, China
March 2025

Organizing Committee

Honorary Chair

Tielin Shi, Huazhong University of Science and Technology, China

General Chairs

Xiaochun Song, Hubei University of Technology, China

Ting-Chung Poon, Virginia Tech, Virginia, USA

Advisory Chairs

Guanglan Liao, Huazhong University of Science and Technology, China

Lei Nie, Hubei University of Technology, China

Program Chairs

Zhongsheng Zhai,Hubei University of Technology, China

Qiang (Shawn) Cheng, University of Kentucky, USA

Wenlong Lu, Huazhong University of Science and Technology, China

Ning Sun, Nankai University, China

Local Organizing Chairs

Zhi Xiong, Hubei University of Technology, China

Jun Tu, Hubei University of Technology, China

Publicity Chairs

Wei Feng, Hubei University of Technology, China

Xiaobing Dai, Hubei University of Technology, China

Xiwen Zhang, Beijing Language and Culture University, China

Souad BEZZAOUCHA REBAÏ, Électrique Informatique et AutomatiqueLa Rochelle, France

David Li, University of Glasgow, UK

Yanbiao Sun, Tianjin University, China

Bin Xue, Tongji University, China

Qijian Tang, Shenzhen University, China

Lin Sun, Xi'an Jiaotong University, China

Yanan Yu, Tianjin University of Technology and Education, China

Organizing Committee

Publication Chairs

Xiaofeng Wang, University of South Carolina, USA

Weitian Wang, Montclair State University, USA

Yang Yue, Xi'an Jiaotong University, China

Technical Committee Chairs

Mukesh Prasad, University of Technology Sydney, Australia

Qinghua Wu, Hubei University of Technology, China

Jianglei Di, Guangdong University of Technology, China

Xixing Li, Hubei University of Technology, China

Zhiyang Deng, Hubei University of Technology, China

Zili Lei, Hubei University of Technology, China

Xiao Chen, Hubei University of Technology, China

Yang Zhang, Hubei University of Technology, China

Onsite Conference Notice

1 Conference Venue | 会议地点

Hubei University of Technology, China
湖北工业大学

No.28, Nanli Road, Hong-shan District, Wuhan, Hubei Province, China

地址：湖北省武汉市洪山区南李路28号

2 Transportation | 交通

• How to Reach Hubei University of Technology, China
https://www.travelchinaguide.com/cityguides/hubei/wuhan/getting-there.htm

From Wuhan Railway Station 从武汉站出发
—By Public Transportation (56 minutes)
Take Metro Line 4 (Berlin direction) at Wuhan Railway Station and get off at Wuchang Railway Station. Transfer to Metro Line 7 (Qinglongshan Metro Town direction) and get off at Hugong University Station (Exit B). Finally, walk 400 meters to Hubei University of Technology.
在武汉火车站乘坐地铁 4 号线（柏林方向）在武昌火车站下车，换乘地铁 7 号线（青龙山地铁小镇方向）在湖工大站下车（B 口出），最后步行 400 米到达湖北工业大学。
—By Taxi (21km, 37 minutes)

From Wuchang Railway Station 从武昌站出发
—By Public Transportation (25 minutes)
在武昌火车站乘坐地铁 7 号线（青龙山地铁小镇方向）在湖工大站下车（B 口出），最后步行 400 米到达湖北工业大学。
Take Metro Line 7 (Qinglongshan Metro Town direction) at Wuchang Railway Station and get off at Hugong University Station (Exit B). Finally, walk 400 meters to reach Hubei University of Technology.
—By Taxi (7km, 20 minutes)

From Wuhan East Station 从武汉东站出发
—By Public Transportation (58 minutes)
Take Metro Line 2 (Tianhe Airport direction) at Wuhan East Station and get off at Crab Cape Station. Transfer to Metro Line 7 (Qinglongshan Metro Town direction) and get off at Hugong University Station (Exit B). Finally, walk 400 meters to Hubei University of Technology.
在武汉东站乘坐地铁 2 号线（天河机场方向）在螃蟹岬站下车，换乘地铁 7 号线（青龙山地铁小镇方向）在湖工大站下车（B 口出），最后步行 400 米到达湖北工业大学。
—By Taxi (12km, 24 minutes)

From Hankou Railway Station 从汉口火车站出发

—By Public Transportation (50 minutes)

在汉口火车站乘坐地铁 2 号线（佛祖岭方向）在螃蟹岬站下车，换乘地铁 7 号线（青龙山地铁小镇方向）在湖工大站下车（B 口出），最后步行 400 米到达湖北工业大学。

Take Metro Line 2 (Fozuling direction) at Hankou Railway Station and get off at Crab Cape Station. Transfer to Metro Line 7 (Qinglongshan Metro Town direction) and get off at Hugong University Station (Exit B). Finally, wa lk 400 meters to Hubei University of Technology.

—By Taxi (24km, 37 minutes)

From Tianhe Airport 从武汉天河机场出发

—By Public Transportation (1 hour and 34 minutes)

在天河机场站乘坐地铁 2 号线（佛祖岭方向）在螃蟹岬站下车，换乘地铁 7 号线（青龙山地铁小镇方向）在湖工大站下车（B 口出），最后步行 400 米到达湖北工业大学。

Take Metro Line 2 (Fozuling direction) at Tianhe Airport Station and get off at Crab Cape Station. Transfer to Metro Line 7 (Qinglongshan Metro Town direction) and get off at Hugong University Station (Exit B). Finally, walk 400 meters to Hubei University of Technology.

—By Taxi (45km, 47minutes)

3 Conference Rooms | 会议室信息

Rooms	Activities	Time
1st Floor -Teaching Building 7A 教七楼 A 座 1 楼	Registration	10:00-16:00 March 28th
Library Lecture Hall 图书馆报告厅	Keynote Speeches	9:00-12:10 March 29th
Teaching Building 7-A714 教七楼 A 座 714	Invited Speech I& IV & VII & X Session 1 & 4	13:30-17:55 March 29th
Teaching Building 7-A715 教七楼 A 座 715	Invited Speech II & V & VIII & XI Session 2 & 5	13:30-17:55 March 29th
Teaching Building 7-A717 教七楼 A 座 717	Invited Speech III & VI & IX & XII Session 3 & 6	13:30-17:55 March 29th
1st Floor Lecture Hall -Teaching Building 7 教七楼 1 楼报告厅	Closing and Award Ceremony	17:55-18:20 March 29th

4 Onsite Presentation | 现场报告

• Timing: a maximum of 15 minutes total, including speaking time and discussion. Please make you're your presentation is well timed.
• Each speaker is required to meet her / his session chair in the corresponding session rooms 10 minutes before the session starts and copy the slide file (PPT or PDF) to the computer.
• It is suggested that you email a copy of your presentation to your personal in box as a backup. If for some reason the files can't be accessed from your flash drive, you will be able to download them to the computer from your email.
• Please note that each session room will be equipped with a LCD projector, screen, point device, microphone, and a laptop with general presentation software such as Microsoft Power Point and Adobe Reader.
• Poster Presenters should bring your poster to the conference venue and put it on designated place.

5 Name Badge | 代表证

For security purposes, delegates, speakers, exhibitors and staff are required to wear their name badge to all sessions and social functions. Lending your participant card to others is not allowed. Entrance into sessions is restricted to registered delegates only. If you misplace your name badge, please ask the staff at the registration desk to arrange a replacement.

6 Gentle Reminder | 温馨提示

- Please ensure that you take all items of value with you at all times when leaving a room. Do not leave bags or laptops unattended. The conference organizer does not assume any responsibility for the loss of personal belongings of the participants.
- Accommodation is not provided. Delegates are suggested make early reservation.
- Please show the badge and meal coupons when dining.

7 Recommend Hotel | 住宿推荐

- **JI Hotel Wuhan Hubei University of Technology**
- **全季酒店（武汉湖北工业大学店）**

https://hk.trip.com/hotels/wuhan-hotel-detail-112540422/quan-ji-jiu-dian/?crn=1&adult=2&children=0

If you need to book this hotel, you can contact the conference staff: 18149749902 (same as the WeChat)
Notes: ICCCV 2025
Conference Agreement Price is: 350 CNY/Night (Including Breakfast)

Other Hotels:
Atour Hotel Wuhan University of Technology Jiedaokou, 5km
Wuhan University of Technology Intercity Hotel, 6.7km
Hampton by Hilton Wuhan Baishazhou Avenue Hotel, 5.9km
Baili Aishang Hotel（Wuhan Nanhu Poly Central Mansion), 2.7km

Online Conference Notice

1 Platform: Zoom

Download Link: https://zoom.us/download

2 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.

3 Time Zone | 时区

GMT+8

***You're suggested to set up the time on your computer in advance.**

4 Online Room Information | 线上会议号信息

Zoom ID: 835 3930 0002

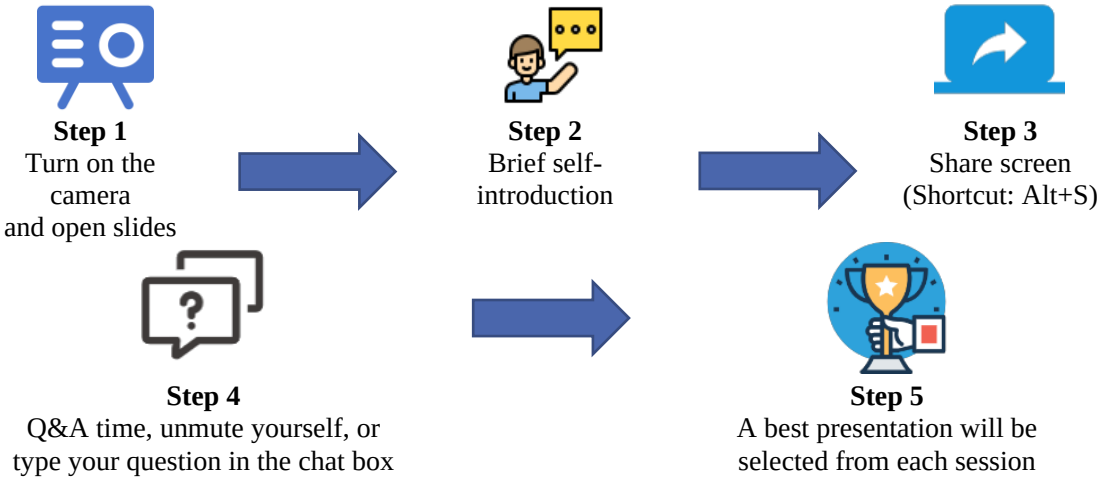
Zoom Link: https://us02web.zoom.us/j/83539300002

You can scan QR code to enter：

1. You can download the virtual background here.
2. Prior to the formal conference, presenter shall join the test room to make sure everything is on the right track
3. Note: Please rename your Zoom Screen Name in below format before entering meeting room.

Role	Format	Example
Conference Committee	Position-Name	Conference Chair-Name
Keynote/ Invited Speaker	Position-Name	Keynote/Invited Speaker-Name
Author	Session Number-Paper ID-Name	S1-CC0001-Name
Delegate	Delegate-Name	Delegate-Name

5 Presentation Process by Zoom Meeting | 报告流程



6 About Presentation | 线上报告

• Every presenter has 15 minutes, including Q & A. Each presentation should have at least 10 minutes.
• The best presentation certificate and all authors' presentation certificates will be sent after conference by email.
• It is suggested that the presenter email a copy of his / her video presentation to the conference email box as a backup in case any technical problem occurs.

7 Environment & Equipment Needed | 环境及设备要求

• A quiet place; Stable Internet connection; Proper lighting and background
• A computer with internet and camera; Earphone

8 Conference Recording | 会议录制

• We'll record the whole conference. If you do mind, please inform us in advance. We'll stop to record when it's your turn to do the presentation.
• The whole conference will be recorded. It is suggested that you should dress formally and we appreciate your proper behavior.
* The recording will be used for conference program and paper publication requirements. It cannot be distributed to or shared with anyone else, and it shall not be used for commercial nor illegal purpose.

Keynote Speakers

Prof. Tielin Shi

Huazhong University of Science and Technology, China

Prof. James Kwok

Hong Kong University of Science and Technology, China

Prof. Bo Du

Wuhan University, China

Prof. Haitao Zhang

Huazhong University of Science and Technology, China

Prof. Ting-Chung Poon

Virginia Tech, Virginia, USA

Prof. Anand Asundi

d'Optron Pte Ltd, Singapore

Invited Speakers

Prof. Seokwon Yeom

Daegu University, South Korea

Prof. Baoqing Guo

Beijing Jiaotong University, China

Prof. Xiang Peng

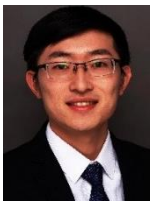
Shenzhen Anhua Optoelectronics Technology Co., Ltd, China

Assoc. Prof. Zhiyu Jiang

Northwestern Polytechnical University, China

Prof. Hui Liu

Kunming University of Science and Technology, China



Assoc. Prof. Wenzheng Zhai

Huazhong University of Science and Technology, China

Assoc. Prof. Gaurav Gupta

Wenzhou-Kean University, China

Assoc. Prof. Yuanlong Xie

Huazhong University of Science and Technology, China

Assoc. Prof. Minglei Yuan

Anhui University of Finance and Economics Hefei Institute for Advanced Research, China

Invited Speakers

Assoc. Prof. Hanzhong Wu

Huazhong University of Science and Technology, China

Assoc. Prof. Qijian Tang

Shenzhen University, China

Dr. Nenglun Chen

Nanjing University of Information Science and Technology, China

Prof. Chiharu Ishii

Hosei University, Japan

Assoc. Prof. Xie Ming

Nanyang Technological University, Singapore

Assoc. Prof. Jiaxin Cai

Xiamen University of Technology, China

Prof. Pascal Lorenz

University of Haute-Alsace, France

Assoc. Prof. Souad BEZZAOUCHA REBAÏ

Électrique Informatique et AutomatiqueLa Rochelle, France

Simple Program

March 28th (Friday)

Onsite Registration & Materials Collection for Onsite Participants 现场参会人员签到注册
Time: 10:00-16:00 (GMT+8)
Venue: **1st Floor-Teaching Building 7A, Hubei University of Technology, China** **湖北工业大学教七楼 A 座 1 楼**
Address: **No.28, Nanli Road, Hong-shan District, Wuhan, Hubei Province, China** **湖北省武汉市洪山区南李路 28 号**
Registration Steps:
1. Arrive at the **1st Floor-** Teaching Building 7, Hubei University of Technology, China 到达湖北工业大学教七楼 A 座 1 楼；
2. Inform the conference staff of your paper ID; 告诉会务人员您的论文 ID；
3. Sign your name on the Participants list; 在参会人员名单上签名；
4. Sign your name on Lunch & Dinner requirement list; 在午餐和晚餐需求表上签名；
5. Check your conference kits (1 program, 1 lunch coupon, 1 dinner coupon, 1 Name badge, 1 bag, 1 U disk); 检查你的会议包:(一份会议日程，一份午餐券，一份晚餐券，一张代表证，一个包，一个 U 盘)
6. Finish registration. 完成签到

Zoom Test for Online Participants 线上参会人员 Zoom 测试	
Zoom: 835 3930 0002	
Zoom Link: https://us02web.zoom.us/j/83539300002	
9:00-9:10	Prof. Ting-Chung Poon, Virginia Tech, Virginia, USA
9:10-9:20	Prof. Anand Asundi, d'Optron Pte Ltd, Singapore
9:20-9:30	Prof. Chiharu Ishii, Hosei University, Japan
9:30-9:40	Assoc. Prof. Xie Ming, Nanyang Technological University, Singapore
9:40-9:50	Assoc. Prof. Jiaxin Cai, Xiamen University of Technology, China
9:50-10:10	Session 7
10:10-10:30	Session 8
16:00-16:10	Prof. Pascal Lorenz, University of Haute-Alsace, France
16:10-16:20	Assoc. Prof. Souad BEZZAOUCHA REBAÏ, Électrique Informatique et AutomatiqueLa Rochelle, France
You can attend the test in another session if you cannot manage it in your given time.	

March 29th (Saturday)

Onsite Speeches and Sessions

Morning Sessions		
Room: Library lecture hall 丨图书馆报告厅		
	Host	**Prof. Guanglan Liao**, Huazhong University of Science and Technology, China
9:00-9:10	**Opening Remark**	**Prof. Xiaochun Song**, Hubei University of Technology, China
9:10-9:50	**Keynote Speech I**	**Prof. Tielin Shi**, Huazhong University of Science and Technology, China
9:50-10:10	**Group Photo & Coffee Break**	
10:10-10:50	**Keynote Speech II**	**Prof. James Kwok,** Hong Kong University of Science and Technology, China
10:50-11:30	**Keynote Speech III**	**Prof. Bo Du**, Wuhan University, China
11:30-12:10	**Keynote Speech IV**	**Prof. Haitao Zhang**, Huazhong University of Science and Technology, China
12:10-13:30	**Lunch Time** Venue: Chufengyuan Hotel	

Afternoon Sessions			
	A714	**A715**	**A717**
13:30-13:50	**Invited Speech I** **Prof. Seokwon Yeom** Daegu University, South Korea	**Invited Speech II** **Prof. Baoqing Guo** Beijing Jiaotong University, China	**Invited Speech III** **Prof. Xiang Peng** Shenzhen Anhua Optoelectronics Technology Co., Ltd, China
13:50-14:10	**Invited Speech IV** **Assoc. Prof. Zhiyu Jiang** Northwestern Polytechnical University, China	**Invited Speech V** **Prof. Hui Liu** Kunming University of Science and Technology, China	**Invited Speech VI** **Assoc. Prof. Wenzheng Zhai** Huazhong University of Science and Technology, China
14:10-15:40	**Session 1** **Computer Vision and Artificial Intelligence** **Session Chair:** Assoc. Prof. Yanan Yu, Tianjin University of Technology and Education, China (CC0047, CC0048, CC0053, CC0054, CC0035, CC0043)	**Session 2** **Remote Sensing and Computer Vision** **Session Chair:** Assoc. Prof. Jie Li, Hubei University of Technology, China (CC0074, CC0076, CC0077, CC2013, CC0040, CC0069)	**Session 3** **Optical 3D Imaging and Measurement** **Session Chair:** Assoc. Prof. Qijian Tang, Shenzhen University, China (CC0067, CC0032, CC0057, CC2007, CC5001)
15:40-16:00	**Break Time**		
	Poster Session **Object Detection and Algorithms** **Session Chair:** Assoc. Prof. Zhi Xiong, Hubei University of Technology, China (CC0017, CC2003, CC0027, CC0033, CC0045, CC0052, CC5002, CC0038, CC0065, CC0072, CC2002, CC2004, CC2005)		

16:00-16:20	**Invited Speech VII** **Assoc. Prof. Gaurav Gupta** Wenzhou-Kean University, China	**Invited Speech VIII** **Assoc. Prof. Yuanlong Xie** Huazhong University of Science and Technology, China	**Invited Speech IX** **Assoc. Prof. Minglei Yuan** Anhui University of Finance and Economics Hefei Institute for Advanced Research, China
16:20-16:40	**Invited Speech X** **Assoc. Prof. Hanzhong Wu** Huazhong University of Science and Technology, China	**Invited Speech XI** **Assoc. Prof. Qijian Tang** Shenzhen University, China	**Invited Speech XII** **Dr. Nenglun Chen** Nanjing University of Information Science and Technology, China
16:40-17:55	**Session 4** **Image Analysis and Computational Model** **Session Chair:** Prof. Hui Liu, Kunming University of Science and Technology, China (CC0004, CC0012, CC0019, CC0025-A, CC5005)	**Session 5** **Intelligent Detection Technology and Engineering Application** **Session Chair:** Assoc. Prof. Yuanlong Xie, Huazhong University of Science and Technology, China (CC0016, CC0018, CC0021, CC0063, CC0064)	**Session 6** **Vision-Based Signal Analysis and Intelligent Control Technology** **Session Chair:** Assoc. Prof. Yang Zhang, Hubei University of Technology, China (CC0007, CC0008, CC0031, CC5006, CC0078)
17:55-18:20	**Closing and Award Ceremony** Venue: 1st Floor Lecture Hall -Teaching Building 7		
18:20-20:00	**Dinner Time** Venue: Chufengyuan Hotel		

March 30th (Sunday)

Online Speeches and Sessions

Morning Sessions		
Zoom ID: 835 3930 0002		**Zoom Link: https://us02web.zoom.us/j/83539300002**
9:00-9:40	**Keynote Speech V**	**Prof. Ting-Chung Poon**, Virginia Tech, Virginia, USA
9:40-10:20	**Keynote Speech VI**	**Prof. Anand Asundi**, d'Optron Pte Ltd, Singapore
10:20-10:30	**Break Time**	
10:30-10:50	**Invited Speech XIII**	**Prof. Chiharu Ishii**, Hosei University, Japan
10:50-11:10	**Invited Speech XIV**	**Assoc. Prof. Xie Ming**, Nanyang Technological University, Singapore
11:10-11:30	**Invited Speech XV**	**Assoc. Prof. Jiaxin Cai**, Xiamen University of Technology, China
11:30-13:00	**Break Time**	
Afternoon Sessions		
Zoom ID: 835 3930 0002		**Zoom Link: https://us02web.zoom.us/j/83539300002**
13:00-15:00	**Session 7**	**Intelligent Image Detection Model and Application** (CC0011, CC0014, CC0015, CC0022, CC0026, CC0061, CC0073, CC2011)
15:00-15:20	**Invited Speech XVI**	**Prof. Pascal Lorenz**, University of Haute-Alsace, France
15:20-15:40	**Invited Speech XVII**	**Assoc. Prof. Souad BEZZAOUCHA REBAÏ**, Électrique Informatique et AutomatiqueLa Rochelle, France
15:40-18:10	**Session 8**	**Image-Based Intelligent Measurement System and Anomaly Detection** (CC0009, CC0044, CC0071, CC0062, CC0051, CC0049)

Detailed Program

Opening Remark

Time	**9:00-9:10, March 29th**
Room	**Library lecture hall**

Prof. Xiaochun Song

President of Hubei University of Technology, China

Keynote Speech I

Time 9:10-9:50, March 29th
Room Library lecture hall

Prof. Tielin Shi

Huazhong University of Science and Technology, China

Prof. Tielin Shi is a professor and doctoral supervisor at the School of Mechanical Science and Engineering of Huazhong University of Science and Technology, member of the Changjiang Scholars Program of the Ministry of Education and standing council member of the Chinese Vibration Engineering Society. Shi won the First-class Award for Natural Science from the Ministry of Education, the First-class Award for Scientific and Technological Progress from the State Education Commission, the First-class Award for Scientific and Technological Progress from the Ministry of Machinery Industry, the Third-class Award for National Scientific and Technological Progress and the China Youth Science and Technology Award. Shi has presided or participated in more than 40 research projects, including the National Basic Research Program of China (973 Program), the National High-Tech Research and Development Program of China (863 Program) and the National Natural Science Foundation of China. In recent years, Shi focused his research on fields such as microsystems and micro-manufacturing, instruments and equipment, monitoring and fault diagnosis. He has published more than 300 papers, over 150 of which are included in the Science Citation Index (SCI). He has applied for more than 80 national invention patents, over 50 of which have been authorized.

Speech Contents

Application of Artificial Intelligence Technology in the Fields of IC Packaging and Medical-Engineering Interdisciplinary Research

Abstract: With the rapid development of technology, both the IC packaging and medical-engineering fields are facing the need to improve efficiency, precision, and find innovative solutions. The introduction of artificial intelligence technology has brought new opportunities to address these issues. In the field of IC packaging, we proposed a new non-destructive testing method for IC packaging defects based on active infrared thermography and high-frequency ultrasonic scanning. By combining artificial intelligence technology, we can quickly and accurately identify various defects in the IC packaging process, such as solder balls missing, cracks, and bubbles, significantly improving the detection efficiency and accuracy while reducing labor costs. In the medical-engineering interdisciplinary field, we combined artificial intelligence technology with microfluidics and high-content microscopic imaging technology. We developed a high-precision microfluidic cell sorting system, which solves problems such as insufficient cell viability and aerosol contamination in traditional flow cytometry cell sorting. We also developed a high-content intelligent microscopic analysis system, which provides a key tool for dynamic monitoring and analysis of living cells. Both systems play important roles in fields such as biomedical research, drug development, and clinical diagnosis.

Keynote Speech II

Time 10:10-10:50, March 29th
Room Library lecture hall

Prof. James Kwok

Hong Kong University of Science and Technology, China

James Kwok is a Professor in the Department of Computer Science and Engineering, Hong Kong University of Science and Technology. He is an IEEE Fellow.
Prof Kwok received his B.Sc. degree in Electrical and Electronic Engineering from the University of Hong Kong and his Ph.D. degree in computer science from the Hong Kong University of Science and Technology. He then joined the Department of Computer Science, Hong Kong Baptist University as an Assistant Professor. He returned to the Hong Kong University of Science and Technology and is now a Professor in the Department of Computer Science and Engineering. He is serving / served as an Associate Editor for the IEEE Transactions on Neural Networks and Learning Systems, Neural Networks, Neurocomputing, Artificial Intelligence Journal, International Journal of Data Science and Analytics, and on the Editorial Board of Machine Learning. He is also serving as Senior Area Chairs of major machine learning / AI conferences including NeurIPS, ICML, ICLR, IJCAI, and as Area Chairs of conferences including AAAI and ECML. He is on the IJCAI Board of Trustees. He is recognized as the Most Influential Scholar Award Honorable Mention for "outstanding and vibrant contributions to the field of AAAI/IJCAI between 2009 and 2019". Prof Kwok is the IJCAI-2025 Program Chair.

Speech Contents

Vision-Language Models: Pre-Training, Fine-Tuning and Trustworthiness

Abstract: Vision-language models (VLMs) are now widely used in various vision-language modeling tasks. However, there are still a number of challenges. First, cross-modal masked language modeling is often used to learn the vision-language associations. However, existing masking strategies are insufficient in that the masked tokens can sometimes be simply recovered with only the language information, ignoring the visual inputs. Second, during fine-tuning, multiple models with various hyperparameter configurations are often created, but typically only one of these models is actually utilized in the downstream task. Third, vision-language models are more vulnerable to jailbreak attacks than their LM predecessors.

To address the first issue, we use a masking strategy based on the saliencies of language tokens to the image. For the second issue, we consider the learned soup, which combines all fine-tuned models with learned weighting coefficients. While this can significantly enhance performance, it is also computationally expensive. We propose to mitigate this by formulating the learned soup as a computationally-efficient hyperplane optimization problem and employing block coordinate gradient descent to learn the mixing coefficients. Finally, to construct robust VLMs, we propose a training-free protecting approach that exploits the inherent safety awareness of LLMs, and generates safer responses via adaptively transforming unsafe images into texts to activate the intrinsic safety mechanism of pre-aligned LLMs in VLMs.

Keynote Speech III

Time	**10:50-11:30, March 29th**
Room	**Library lecture hall**

Prof. Bo Du

Wuhan University, China

Bo Du received the Ph.D. degree from State Key Lab of Information Engineering in Surveying, Mapping and Remote Sensing, Wuhan University, Wuhan, China in 2010.

He is currently a Hongyi Chair Professor with the School of Computer Science and Institute of Artificial Intelligence, Wuhan University. He is also the director of the National Multimedia Software Engineering Technology Research Center, dean of the School of Computer Science at Wuhan University. His major research interests spread across subareas in artificial intelligence (AI), including computer vision, deep learning, image processing, medical AI and data science. His research results have expounded in 500+ publications at prestigious journals and proceedings in prominent conferences, such as IEEE TPAMI, TIP, TCSVT, TMI, TNNLS, TCYB, TGRS, IJCV, IJCAI, AAAI, CVPR, ICCV, ECCV, NeurIPS, ICML, ICLR, and KDD, etc. Fourty-five of them are ESI hot papers or highly cited papers.

He is currently a senior member of IEEE. He serves as associate editor for Neural Networks, Pattern Recognition, Neurocomputing and Neural Processing Letters. He also serves as a reviewer of 20 Science Citation Index (SCI) magazines including IEEE TPAMI, TCYB, TGRS, TIP, JSTARS, and GRSL. He regularly serves as senior PC member of IJCAI and AAAI. He served as area chair for ICPR.

His influence spans various IEEE domains. His citation record is impressive: 28773 citations, h-index: 86 (as of 03/11/2024). He won the Highly Cited Researcher (2019-2024) by the Web of Science Group. He won IEEE Geoscience and Remote Sensing Society Transactions Prize Paper Award. He won the IJCAI (International Joint Conferences on Artificial Intelligence) Distinguished Paper Prize, IEEE Data Fusion Contest Champion, and IEEE Workshop on Hyperspectral Image and Signal Processing Best Paper Award.

Speech Contents

Progress of Large Language Models and the Medical Applications

Abstract: This presentation outlines the latest progress in large language model (LLM) research by the Sensing Intelligence and Machine Learning Lab at Wuhan University, focusing on addressing challenges in pretraining, fine-tuning, and deployment. Key innovations include E2S2 for encoding-enhanced pretraining, Self-Evolution Learning to dynamically identify under-explored tokens, and ScTD for efficient semantic-preserving pretraining. For fine-tuning, PANDA achieves SOTA performance with minimal parameter updates via knowledge distillation, while FSAM optimizes generalization using Fisher information. Deployment solutions like ZSAQ enable zero-shot quantization, ATKD improves knowledge distillation, and ROSE enhances output safety. The lab's Vega-v1/v2 models set new benchmarks on GLUE (91.3) and SuperGLUE (91.3), surpassing global competitors. Additionally, they developed Chinese Medical LLMs (7B/13B) for medical dialogue and the first billion-parameter remote sensing LLM, achieving 82.78% mAP on DOTA_v1.0. These advancements balance theoretical breakthroughs with practical applications, positioning the lab as a leader in advancing LLM capabilities across diverse domains.

Keynote Speech IV

Time 11:30-12:10, March 29th
Room Library lecture hall

Prof. Haitao Zhang

Huazhong University of Science and Technology, China

Prof. Hai-Tao Zhang is the Deputy Dean of the School of Artificial Intelligence and Automation at Huazhong University of Science and Technology (HUST), an Endowed Chair Professor at Huazhong, a recipient of the State Council Special Allowance, a recipient of Natioinal Science Fund for Distinguished Young Scholars, and Natioinal Science Fund for Excellent Young Scholars. He is the Director of the Ministry of Education Engineering Center for Autonomous Intelligent Unmanned Systems, and the Chief Scientist of the National Key R&D Program on New-Generation Artificial Intelligence under the National Science and Technology Innovation 2030 initiative. He obtained his Ph.D. from the University of Science and Technology of China in 2005, conducted postdoctoral research at the University of Cambridge in 2007.

Zhang specializes in areas such as swarm intelligence, autonomous unmanned surface vehicle fleets coopeartion, and multi-robot collaborative manufacturing. He has led several major national-level projects, including the National Science and Technology Innovation 2030 Key Projects, two joint key projects funded by the National Natural Science Foundation of China, and GF Key Projects. He has published over 150 SCI-indexed papers, including more than 100 papers in prestigious journals such as Nature Machine Intelligence, Nature Communications, National Science Review, Automatica, and IEEE Trans/Journal/Magazine. He has authored three English monographs published by Springer. Zhang has received two First Prizes for Natural Science in Hubei Province and one First Prize for Technological Invention in Guangdong Province, all as the first awardees. He has been listed among Stanford University's top 2% of global scientists. His theoretical achievements in swarm intelligence have been highlighted by Nature Physics, and his applied research results have been reported by the State Council. He holds 40 authorized invention patents, 6 of which have been transferred to major enterprises, and has contributed to one national standard. His research achievements have been commercialized in organizations such as China State Shipbuilding Corporation and Guangzhou Shipyard International. He is a board member of the Chinese Association of Command and Control and has served as an editorial board member or associate editor for journals including IEEE Trans. SMC-Systems, IEEE Trans. CAS II, Engineering , Unmanned Systems, Control Theory and Applications, and Robot. Additionally, he has been on the program committees of important international conferences such as IEEE CDC, ACC, and IFAC.

Speech Contents

Autonomous Unmanned Surface Vehicle Fleet-Unmanned Aerial Vehicle Swarm Cross-Domain Cooperative Coverage Detection and Confrontation Gaming

Abstract: High efficiency, ultra-stability, and high precision represent the pinnacle of autonomous unmanned surface vehicle (USV) fleet cooperation technology, which has long been an international challenge. This seminar introduces the latest advancements made by Prof. Zhang's team under the support of major national projects such as the National 2030 Initiative, the National Science Fund for Distinguished Young Scholars, and other key funding programs. These advancements include key technologies in USV-UAV swarm cooperative coverage, cross-domain cooperative takeoff and landing for USV-UAV fleets, patrol and trailing, and maritime channel

confrontation gaming. The presentation also covers the team's research and development efforts in core functional components such as collective coverage perception, SLAM, cooperative takeoff and landing, and confrontation gaming, as well as the development of a complete set of swarm equipment, including 12 types of autonomous USVs and various types of USV-borneUAVs. Finally, the presentation highlights the phased application achievements of the aforementioned core technologies and equipment in areas such as real-time multi-point synchronous monitoring of large-scale marine facilities in the Guangdong-Hong Kong-Macao Greater Bay Area and electromagnetic exploration of oil and gas resources in the South China Sea.

Invited Speech I

Time 13:30-13:50, March 29th
Room Teaching Building 7-A714

Prof. Seokwon Yeom

Daegu University, South Korea

Seokwon Yeom has been a faculty member of Daegu University since 2007. He has a Ph.D. in Electrical and Computer Engineering from the University of Connecticut in 2006.
He has been a guest editor of Applied Sciences and Drones in MDPI since 2019. He has served as a board member of the Korean Institute of Intelligent Systems since 2016, and a member of the board of directors of the Korean Institute of Convergence Signal Processing since 2014. He has been program chair of several international conferences. He was a vice director of the AI homecare center and a head of the department of IT convergence engineering at Daegu University in 2020-2023, a visiting scholar at the University of Maryland in 2014, and a director of the Gyeongbuk techno-park specialization center in 2013.
His research interests include Intelligent image and optical information processing, deep and machine learning, and target tracking.

Speech Contents

Small Drone based Thermal Target Tracking with Track Association

Abstract: This invited talk addresses multiple target tracking with a small drone based on the thermal imaging. First, the YOLO detection model detects thermal objects and extracts position information. Second, the targets are individually tracked based on the positions measured by YOLO to generate their own trajectories. The tracks are estimated by Kalman filter or IMM filter. Track association and fusion select the fittest track for the track and fuses them. Track segment association is developed to connect broken track segments over time. In the experiments, three hikers who lost in their way in the mountains were captured by the thermal imaging camera mounted on a drone. Robust tracking results were obtained in terms of total track life, mean track life, and track purity.

Invited Speech II

Time	13:30-13:50, March 29th
Room	Teaching Building 7-A715

Prof. Baoqing Guo

Beijing Jiaotong University, China

Baoqing Guo is a professor and doctoral supervisor at Beijing Jiaotong University. He has been recognized as an Outstanding Teacher by Baosteel and serves as the chief professor in the field of intelligent perception at the National Key Laboratory of Autonomous Operation for Advanced Rail Transit. He is also the chief scientist for a National Key R&D Program. Additionally, he is a committee member of the Intelligent Operation and Maintenance Division of the Chinese Society of Mechanical Engineering.

His research focuses on monitoring, detection, and intelligent perception of rail transit operating environments. In this field, he has led over 20 projects, including National Key R&D Programs and National Natural Science Foundation projects. He has published more than 50 SCI/EI-indexed papers and has been awarded one Second Prize for Beijing Science and Technology Progress and two First Prizes from the China Railway Society for Scientific and Technological Achievements.

Speech Contents

Advanced Rail transit Autonomous Operation based on Environment Perception

Abstract: In this lecture, I will introduce the principle of rail transit train operation control and the special requirements of rail transit in terms of environmental perception compared with automobiles. On this basis, I will put forward several key technologies that need to be broken through in terms of intelligent perception of the operating environment, equipment status monitoring, reliable communication, and autonomous tracking. As perception is the premise of autonomous train operation, some progress in all-weather perception in complex train driving environment will be introduced, including the self registration and fusion detection method of on-board multi-source sensors, the extraction method of railway scene dynamic gauge, the intrusion target generation method based on generative AI, the detection of unusual targets based on anomaly detection and the risk situation analysis method based on target location relationship, as well as the application effect of the above methods in railway field.

Invited Speech III

Time	13:30-13:50, March 29th
Room	Teaching Building 7-A717

Prof. Xiang Peng

Shenzhen Anhua Optoelectronics Technology Co., Ltd, China

Prof, Xiang Peng graduated from the School of Precision Instrument and Opto-Electronics Engineering, Tianjin University in 1982. From 1990 to 1992, he worked as Alexander von Humboldt Fellow at University of Stuttgart. From 1992 to 2002, he worked at the School of Precision Instrument and Opto-Electronics Engineering. In January 2003, he was transferred to Shenzhen University as a professor and has been selected as Highly Cited Scholar in China by Elsevier (2022), and top 2% scientists with a life-time influential impact on science by Stanford and Elsevier. Now he is the chief scientist and laboratory director of Shenzhen Anhua Optoelectronics Technology Co., Ltd. Prof. Peng's research activities have been supported by the Natural Science Foundation of China (NSFC), Sino-German Center for Research Promotion (SGCRP), and Industrial Sectors.

Speech Contents

Making Mobile 3D Scanning More Smart

Abstract: In this talk l would focus on a topic of 3D scanning of complex shapes with a mobile device. When scanning an object with a handheld 3D scanner, we must deal with two key issues: sensor tracking and re-localization in order to provide a good initial guess, for the registration of multi-view range images, with use of the ICP strategy. l shall discuss the technical route of sensor tracking based on an unsupervised learning approach, following by a route of re-localization with help of fast point feature histogram (FPFH). Furthermore, we propose a supervised learning-based approach to learn feature descriptor on the set of point clouds, which shows promising for making a mobile 3D scanning more smart. In the meanwhile, l shall also give a brief introduction to R&D activities at Shenzhen Anhua Optoelectronic Technology Co. Ltd. to showcase several prototypes of 3D imaging and modeling.

Invited Speech IV

Time 13:50-14:10, March 29th
Room Teaching Building 7-A714

Assoc. Prof. Zhiyu Jiang

Northwestern Polytechnical University, China

Zhiyu Jiang is Associate Professor with the Northwestern Polytechnical University. Previously he received the Ph.D degree in Signal and Information Processing from the University of Chinese Academy of Sciences, Beijing. China.
He focuses on computer vision and remote sensing. More than 15 papers have been accepted/published on IEEE Trans. and other top conferences. He has won the IEEE best student paper award in 2014. He served as Guest Editor, reviewers of 20+ International Journals and Transactions, such as IEEE T-CYB/IEEE T-GRS. He also served as Program Co-Chair, Steering Committee, Track Chair, TPC Member of International Conference for 50+ times.

Speech Contents

Cross-Domain Visual Semantic Perception

Abstract: Cross-domain visual semantic perception is a critical area in computer vision that focuses on developing models capable of performing tasks across different domains without extensive retraining. This presentation will cover key advancements in scene-agnostic object detection, scene classification, and semantic segmentation. Scene-agnostic object detection involves creating robust algorithms that can identify objects consistently across various environments. Scene classification focuses on accurately categorizing different visual scenes, which is essential for applications such as surveillance. Semantic segmentation takes this a step further by labeling each pixel in an image, providing a detailed understanding of the visual content. The discussion will highlight recent breakthroughs, current limitations, and future research directions in these interconnected fields, offering valuable insights into how cross-domain visual semantic perception is reshaping the landscape of computer vision applications.

Invited Speech V

Time 13:50-14:10, March 29th
Room Teaching Building 7-A715

Prof. Hui Liu

Kunming University of Science and Technology, China

Liu Hui, Ph.D., professor at KUST, doctoral supervisor, member of the 93 Society, visiting scholar at the University of Ottawa in Canada, young top talent of the "Xingdian Talent Support Program" in Yunnan Province, leader of the first batch of national first-class courses, peer review expert of the National Natural Science Foundation, senior member of the Chinese Computer Society, expert in the national graduate education evaluation and monitoring expert database, vice president of science and technology at Yunnan Kunming Iron and Steel Electronic Information Technology Co., Ltd., teaching master at Kunming University of Science and Technology, serving as a young editorial board member for journals such as "Metallurgical Automation" and "Journal of Iron and Steel Research", and reviewer for journals such as "Control and Decision making" and "Optical Engineering".Mainly engaged in theoretical and applied research in machine learning, soft sensing, image recognition, etc., leading 2 projects funded by the National Natural Science Foundation, 2 general projects funded by the Yunnan Provincial Department of Science and Technology, 1 science and technology special project for key industries served by universities in Yunnan Province, and 1 youth talent special project supported by the Yunnan Xingdian Talent Support Program. In Engineering Applications of Artificial Intelligence Computers & Chemical Engineering、Chemometrics and Intelligent Laboratory Systems、 More than 30 journal articles have been published in SCI/EI, covering topics such as control and decision-making, control theory and applications. In terms of teaching, I have led more than 10 educational reform projects, including the construction of national first-class courses, high-quality courses for graduate students in Yunnan Province, the construction of a case library for graduate courses in Yunnan Province, research projects on undergraduate education and teaching reform in Yunnan Province, and the first batch of Yunnan Province's expansion plan projects. Won the first prize in the finals of the National College Electronic Information Young Teacher Teaching Competition, the special prize in the 6th and 7th Yunnan Provincial College Teacher Teaching Competition, the second prize in the 3rd and 4th National College Teacher Teaching Innovation Competition, the "Smart Teaching Star" of the Ministry of Education's online education, and the first prize in the National UOOC Alliance's blended teaching excellent case, and more than ten other teaching awards. The MOOC "Introduction to Intelligent Control" responsible for its construction has been launched on platforms such as Strong Learning Nation, Chinese University MOOC, Xuetang Online (Chinese and International), Zhizhi Tree, Xueyin Online, and Youke Alliance.

Speech Contents

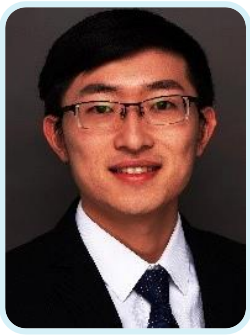
Difficulty Analysis and Technology Summary and Outlook in the Process of End-Point Control of Converter Steelmaking

Abstract: The endpoint determination in converter steelmaking is a crucial and pivotal step during the blowing process, with its main challenge lying in the accurate and real-time measurement of carbon content and temperature within the molten pool. Focusing on the methods and principles of endpoint carbon and temperature measurement in converter steelmaking, this paper delves into various aspects including the production process of converter steelmaking, traditional manual endpoint determination methods, contact sensor methods, detection

theories and methods based on spectral radiation, carbon and temperature regression methods based on flame image processing and recognition, as well as data-driven soft measurement modeling methods for endpoint carbon and temperature. It summarizes the research content and ideas of existing methods, analyzes areas that still require further improvement, and proposes future research directions in this field, aiming to serve as a catalyst for continued research and development in this area.

Invited Speech VI

Time	13:50-14:10, March 29th
Room	Teaching Building 7-A717



Assoc. Prof. Wenzheng Zhai

Huazhong University of Science and Technology, China

Zhai Wenzheng, a doctoral supervisor and associate professor, is a Humboldt Scholar, a Huazhong Distinguished Scholar, an Innovative Talent of Hubei Province, and a Wuhan Talent. He serves as a member of the Extreme Manufacturing Branch of the Chinese Mechanical Engineering Society, the Deputy Secretary-General of the Hubei Special Processing Association, and the Deputy Editor-in-Chief of JMMP.

He has presided over more than 10 national-level projects, including key projects of GF, National Natural Science Foundation projects, and key research and development tasks. He has researched multi-energy field assisted additive manufacturing technology focusing on certain key model load-bearing components, and the fatigue performance tests of typical structural components have been passed in the main engine factory. Among them, certain components manufactured by multi-energy field assisted additive technology passed the flight live-fire test in 2023, winning high praise from the application unit.

He has been consecutively selected into the list of the world's top 2% scientists in 2023 and 2024. His work has been cited more than 4,000 times on Google Scholar, and his H-index is 31.

Speech Contents

Coaxial Laser-Arc Hybrid Additive Manufacturing Technology

Abstract: The performance, cost-effectiveness, and operation and maintenance efficiency of key components in fields such as aerospace, wind power, and marine directly affect the green and healthy development of the national economy and are also crucial to national security. However, for a long time, the manufacturing of core key components has faced bottleneck problems such as long manufacturing cycles, low performance, and short service life. There is an urgent need to establish an independently innovative intelligent laser additive manufacturing technology system featuring high efficiency, high performance, and long service life to meet the intelligent manufacturing needs of modern high-end equipment and other fields.

The coaxial laser-arc hybrid additive manufacturing technology has the advantages of high deposition efficiency and non-directionality. By using the coaxial laser-arc hybrid additive manufacturing technology for the regulation of the microstructure and properties, and simultaneously regulating the behavior of the molten pool through the oscillating laser composite energy field and regulating the stress of the solidified microstructure with pulsed laser, the purpose of controlling defects and residual stress can be achieved. During the forming process, the wire material is uniformly heated, resulting in high forming accuracy and low residual stress. This technology is a key technology for manufacturing complex components of high-end equipment, especially in the aerospace field and other related fields.

Invited Speech VII

Time	16:00-16:20, March 29th
Room	Teaching Building 7-A714

Assoc. Prof. Gaurav Gupta

Wenzhou-Kean University, China

Dr. Gaurav Gupta is currently an Associate Professor of Mathematical Sciences (Data Analytics) at Wenzhou-Kean University in Wenzhou, China. With over 16 years of teaching and research experience, Dr. Gupta has a strong academic background. Gupta's research interests are primarily centered around Applied Mathematics, Computer Vision, and Machine Learning. Between 2007 and 2010, he contributed his expertise to the Indian Space Research Organization (ISRO). Notably, he has secured more than $150,000 in grant funding from the Wenzhou Education Bureau, highlighting his dedication to advancing educational initiatives. He has an impressive track record of publication with 60 research papers published in well-regarded journals and conferences. Furthermore, he has played a pivotal role in mentoring students, supervising two PhD students, advising nine Master's Dissertations, and overseeing three undergraduate projects.

Dr. Gupta's commitment to academic and research excellence extends to his active participation in various conferences and workshops. He has contributed as a keynote speaker, technical committee member, and session chair, making significant contributions to the academic community.

Speech Contents

The Future of Diagnostics: AI-Powered Medical Image Analysis

Abstract: AI-powered medical image analysis is transforming diagnostics by enhancing accuracy, efficiency, and accessibility. Leveraging machine learning and deep learning, AI systems can analyze X-rays, MRIs, CT scans, and ultrasounds to detect diseases, segment anatomical structures, and support personalized treatment. While challenges like data privacy, annotation requirements, and validation remain, the potential for AI to revolutionize diagnostics is immense. This talk explores key applications, current trends, and future directions, highlighting how AI is shaping the future of medical imaging.

Invited Speech VIII

Time	16:00-16:20, March 29th
Room	Teaching Building 7-A715

Assoc. Prof. Yuanlong Xie

Huazhong University of Science and Technology, China

Dr. Xie Yuanlong is an Associate Professor and Master's Supervisor at Huazhong University of Science and Technology (HUST). He earned his Bachelor's degree in Electrical Engineering and Automation in 2014 and his Ph.D. in Mechatronic Engineering in 2018, both from HUST. During his studies, he was a visiting scholar at the University of Leeds, UK.

Dr. Xie's research focuses on intelligent control, mobile robotics, smart manufacturing, and autonomous systems. He has published over 140 SCI/EI-indexed papers, including 40 as first or corresponding author, and holds over 80 patents. He has led projects funded by the National Natural Science Foundation of China and the Postdoctoral Science Foundation.

As an IEEE Senior Member, Dr. Xie serves as a guest editor for journals like Sensors and Applied Sciences and has organized sessions at major conferences, including IEEE AIM and CCC.

Speech Contents

Robust High-Precision Motion Control for Wheeled Omnidirectional Mobile Robots

Abstract: Wheeled omnidirectional Mobile Robots serve as critical components in industrial manufacturing for coordinated operational tasks. However, such systems encounter multi-source uncertain working environments caused by obstacle interference, complex ground conditions, and abrupt end-effector load variations, severely compromising critical performance metrics including operational safety, control stability, and grasping precision. This report addresses the challenges of robust high-precision control for wheeled omnidirectional Mobile Robots to enhance the system's adaptability, accuracy, and stability under heterogeneous uncertainties, thereby advancing the operational reliability of mobile manipulators in complex industrial environments.

Invited Speech IX

Time	16:00-16:20, March 29th
Room	Teaching Building 7-A717

Assoc. Prof. Minglei Yuan

Anhui University of Finance and Economics Hefei Institute for Advanced Research, China

Dr. Minglei Yuan holds a Ph.D. in Computer Science and Technology from Nanjing University and is currently an Associate Professor at both the Department of Computer Science and Technology at Anhui University of Finance and Economics and the Hefei Higher Research Institute. His research primarily focuses on machine learning and pattern recognition. Dr. Yuan has successfully led key Natural Science Research projects funded by the Anhui Provincial Department of Education in 2016 and 2023. His findings have been published in internationally renowned journals such as Pattern Recognition, Frontier in Plants Science, and Expert Systems with Applications, and he has presented his work at major academic conferences including IEEE international conferences, ICPR, and IJCNN.

Speech Contents

Cross-Domain Generalization and Uncertainty Modeling in Few-Shot Image Classification

Abstract: This report reviews two approaches in few-shot image classification. The first method employs a Forget-Update Module (FUM) to address cross-domain challenges by mitigating distribution differences between training and testing data. The second method introduces an Uncertainty-Based Network (UCN) that leverages data augmentation and mutual information to model prediction uncertainty in a transductive inference framework.

Invited Speech X

Time	16:20-16:40, March 29th
Room	Teaching Building 7-A714

Assoc. Prof. Hanzhong Wu

Huazhong University of Science and Technology, China

Hanzhong Wu, Associate Professor, member of the Geometric Quantities Professional Committee of Chinese Society for Metrology and Testing, member of the organizing committee of CLEO-PR, editor of Optics Express, and a peer-review expert for the National Natural Science Foundation of China. He was awarded the Outstanding Doctoral Dissertation Award by Chinese Instrument and Control Society. He has been the principal investigator of more than ten projects, including the National Key R&D Program projects, National Natural Science Foundation of China, 173 Fund, Enterprise Joint Fund, etc. He has been engaged in the research of ultra-precision measurement technology based on optical frequency combs, and the development of new types of optical combs and their scientific instruments. He has published more than 50 papers in journals such as Nature Communications, Advanced Photonics Nexus, Optics Letters, Applied Physics Letters, etc., and applied for more than 20 invention patents. The graduate students he supervised have received awards such as the National Scholarship, Outstanding Graduate Student, Outstanding Academic Report Award, etc.

Speech Contents

Dual Multi-Soliton Microcomb Ranging

Abstract: Laser frequency combs based on the chip-scale microresonators (microcombs) have been giving rise to the new opportunities for the fields of optical ranging, due to the compact footprint, broad spectral band, and high repetition frequency. However, the power conversion efficiency of the coherent single-soliton state is low, often at ~1% from the pump laser. Chaotic microcombs with much higher power conversion efficiency also allow for distance measurement, but the high-noise nature can not bring high-precision measurement. Here we, for the first time, demonstrate that multi-soliton microcomb can be harnessed to perform ultrafast and precise distance measurement by dual multi-soliton interferometry. Multi solitons carry the higher power conversion efficiency and are easy to reach, in contrast to the single soliton, and offer nanometric-precision distance measurement, in comparison with the chaos. Our experimental results show the measurement uncertainty can be within ±17 nm in the lab, and the precision can reach 2.12 nm at 2 μs, and 5.48 pm at 500 μs, when using 3 solitons in the signal comb and 3 solitons in the local comb. We also show the multifunctional applications using this method, including the vibration monitoring, the spinning disk measurement, and the unmanned aerial vehicle detection. Further, we explorer the photon-level ranging when the optical power is at fW level. Based on the technique of time-correlated single-photon counting, the measurement uncertainty can be below ±9.5 μm, and the Allan deviation can achieve 3.57 μm at 1 s, and 202 nm at 50 s, with 5 solitons in the signal comb and 1 soliton in the local comb. Additionally, we carry out the long-distance measurement of about 270 m out of lab, and the non-line-of-sight imaging. Multi-soliton ranging features higher precision, faster speed, higher power conversion efficiency, compared to the single-soliton ranging. Our work is able to inspire new trains of thought for the multi-soliton microcomb based applications, e.g., optical ranging, communications, spectroscopy, and optical time transfer.

Invited Speech XI

Time	16:20-16:40, March 29th
Room	Teaching Building 7-A715

Assoc. Prof. Qijian Tang

Shenzhen University, China

Qijian Tang, Associate Professor at Shenzhen University, Shenzhen's high-level talents. He obtained a Ph.D. from Tianjin University in 2015. His research focuses on optical 3D imaging and measurement, Fourier ptychographic imaging, computational optical imaging and measurement. He has led various projects funded by the National Natural Science Foundation of China, the National Key Research and Development Program, and provincial and municipal science and technology programs. He has published over 20 academic papers, applied for 10 invention patents, and has received the Second Prize of the Guangdong Province Science and Technology Progress Award and the Second Prize of the Shenzhen Science and Technology Progress Award.

Speech Contents

3D Reconstruction Base on Line Encoding

Abstract: Fringe projection profilometry has the advantages of full-field measurement, high precision, strong anti-interference ability, and simple system structure, and it is widely applied in fields such as industrial measurement and cultural relics digitization. However, when reconstructing the surface of highly reflective or high dynamic range objects, due to the limited dynamic range of the camera, it leads to low reconstruction accuracy in overexposed or underexposed areas and incomplete data information. Aiming at the above problems, we propose two line encoding schemes, and finally achieve more efficient and complete three-dimensional reconstruction.

Invited Speech XII

Time	16:20-16:40, March 29th
Room	Teaching Building 7-A717

Dr. Nenglun Chen

Nanjing University of Information Science and Technology, China

Chen Nenglun is currently a Lecturer with Nanjing University of Information Science. He received his Ph.D. degree from the University of Hong Kong in 2022. He mainly engages in research related to computer graphics and computer vision. His research interests include multimodal learning, digital geometric modeling, medical image processing, self-supervised learning, scene understanding, and shape analysis. He has published over 15 papers in top-tier international conferences and journals. He is also a reviewer for multiple top-tier conferences and journals, including CVPR, ICCV, SIGGRAPH, and TVCG.

Speech Contents

Self-supervised Learning for Multimodal Structure-Aware Scene Understanding and 3D Shape Analysis

Abstract: Self-supervised learning, as an effective way of learning with un-labeled data, has been widely studied in various research fields. Due to the lack of labeled data, researchers in this field usually explore various kinds of task-specific priors and incorporate the priors into the training paradigms. This talk is mainly about our previous works regarding self-supervised learning for multimodal scene understanding and 3D shape analysis. In our preliminary work, we have primarily focused on two key research problems, including shape structural representation and cross-modal knowledge transfer. We have developed a series of methodological frameworks and demonstrated their effectiveness in scene understanding and shape analysis tasks. We have also investigated using multimodal large-scale models on these two categories of tasks, showcasing their effectiveness in both task domains.

Keynote Speech V

Time	**9:00-9:40, March 30th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Prof. Ting-Chung Poon

Virginia Tech, Virginia, USA

Ting-Chung Poon is Professor of Electrical and Computer Engineering at Virginia Tech, USA. His current research interests include Information Optics and Optical Scanning Holography (OSH). Dr. Poon is the author of Optical Scanning Holography with MATLAB (Springer, 2007), and co-author of several textbooks, including Modern Information Optics with MATLAB (Cambridge University Press and Higher Education Press, China, 2023). He served as the General Chair of the Optica Annual Meeting (Frontiers in Optics, FiO 2021,2022). He is a Fellow of Optica, the SPIE, and the IoP, and a Life Fellow of the IEEE. He was honored with SPIE's Dennis Gabor Award in 2016 and received Optica's 2024 Emmett N. Leith Medal.

Prof. Yaping Zhang (Co-author)

Kunming University of Science and Technology, China

Yaping Zhang is Professor and Director of the Yunnan Provincial Key Laboratory of Modern Information Optics, Kunming University of Science and Technology, China. Dr. Zhang is an academic leader in optics in Yunnan Province and a member of the Steering Committee on Opto-Electronic Information Science and Engineering, Ministry of Education, China. She is the first author of several books, including the textbook Modern Information Optics with MATLAB, co-published by Cambridge University Press (U.K.) and Higher Education Press (China). She served as the General Chair of the Optica Topical Meeting on Digital Holography and Three-Dimensional Imaging (DH) in 2023 and was a Subcommittee Chair for the Optica Annual Meeting (Frontiers in Optics, FiO 2021,2022). She is a Senior Member of the IEEE.

Speech Contents

Layer-Mesh-Based Computer-Generated Holography

Abstract: Computer-generated holography deals with the methods for digitally generating holographic interference patterns. The resulting interference patterns are called computer-generated holograms (CGHs). The use of CGHs allows for greater flexibility and precision in creating holographic images. There are three primary approaches in computer-generated holography: point-based, polygon-based, and layer-based. In this presentation, we will begin with a brief overview of these methods. Following that, we will introduce a novel approach called layer-mesh-based computer-generated holography. Additionally, we will showcase some computer simulations and optical experiments to demonstrate this innovative method.

Keynote Speech VI

Time	**9:40-10:20, March 30th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Prof. Anand Asundi

d'Optron Pte Ltd, Singapore

Anand Asundi (安顺泰) has over 40 years of experience in the field of Optical Engineering/Metrology, Photomechanics and Optical NDT. He got his PhD from Stony Brook University and was a research fellow at Virginia Tech. He was Professor at the University of Hong Kong (HKU) (1983-1996) and Nanyang Technological University (NTU) (1996-2019). He has published over 450 SCI indexed papers with over 15,000 citations and H Index of 55. He has also chaired and organized numerous international conferences held in Asia, USA and Europe, notably the icOPEN series which he initiated. He was Editor of Optics and Lasers in Engineering for over 20 years and is Fellow of SPIE and was on their board for two years. He is the recipient of the SPIE, Chandra Vikram award for Optical Metrology in 2024. He was founding Director of the Centre for Optical and Laser Engineering, NTU and founding Chairman of Optics and Photonics Society of Singapore. He is currently founding director and CEO of d'Optron Pte Ltd, one of three companies that he founded.

Speech Contents

Cameras: Then and Now

Abstract: The evolution of cameras from their early origins to the modern digital age represents a significant technological and cultural transformation. This paper explores the development of camera technology, highlighting key innovations and shifts that have shaped how we capture and interact with images. By examining historical and contemporary cameras, including the impact of artificial intelligence (AI), 3D imaging, and their applications in inspection and measurement, we gain a comprehensive understanding of the profound changes in photography and their implications for media, industry, and society.

Invited Speech XIII

Time	**10:30-10:50, March 30th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Prof. Chiharu Ishii

Hosei University, Japan

Chiharu Ishii received his PhD in Mechanical Engineering from Sophia University, Japan in 1997. He worked at Ashikaga Institute of Technology between 1997 and 2002, at Kogakuin University between 2002 and 2009, and at Shibaura Institute of Technology between 2009 and 2010. He has been working at Hosei University since 2010, and currently working as a Professor with the Department of Mechanical Engineering, Faculty of Science and Engineering at Hosei University. Dr. Chiharu Ishii has received several awards such as The Best Paper Award in the area of Tactile and Haptic Interfaces at the 4th International Conference on Human System Interaction (IEEE HSI 2011); Best Paper Award at the 1st International Conference on Computer Science, Electronics and Instrumentation (ICCSE 2012); Best Presentation Award at the International Conference on Intelligent Mechatronics and Automation (ICIMA 2013); Excellent Oral Presentation Award at the 4th International Conference on Soft Computing & Machine Intelligence (ISCMI 2017); 3rd Prize, Excellent Paper Award at the 2021 IEEE 3rd Global Conference on Life Sciences and Technologies (LifeTech 2021). He is currently a member of IEEE, SICE, JSME, RSJ, IEEJ and JSCAS. His research interests are in medical robotics and assistive technology.

Speech Contents

Study on Robotic Palpation System

Abstract: In this talk, a robotic palpation system developed in our laboratory is discussed. The robotic palpation system is teleoperated by leader-follower control method. The reaction force of the touched object is measured by 6-axis force/torque sensor and is presented to the operator's arm through the haptic input device Omega.7. In addition, tactile stimulation is presented to operator's fingertip by the tactile feedback device. In robotic surgery, when palpation is performed under use of the trocar, it is difficult to realize force or tactile sense of the touched object because the forceps is affected by contact with the trocar. Therefore, cancellation of the influence of the trocar on the forceps was attempted by using the neural network. In order to verify effectiveness of the developed palpation system, location identification experiments were carried out. The results showed the effectiveness of the neural network, and identification error of location of the imitation tumor was less than 0.7mm.

Invited Speech XIV

Time	**10:50-11:10, March 30th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Assoc. Prof. Xie Ming

Nanyang Technological University, Singapore

Xie Ming received the B.Eng degree in control and automation engineering from East-China Institute of Textile Technology (now, under the name of Donghua University, Shanghai, China). Subsequently, as a recipient of the nation's prestigious overseas scholarship of Chinese government, he has completed the postgraduate studies and doctorate research works, and has received the Master degree from the University of Valenciennes (France) in 1986 as well as the PhD degree from the University of Rennes (France) in 1989. Since 1986, he has worked as Research Assistant at IRISA-INRIA Rennes, Expert Engineer at INRIA Sophia-Antipolis, Lecturer/Senior Lecturer/Associate Professor of Nanyang Technological University, Fellow of Singapore-MIT Alliance (SMA) (Affiliated with Innovation in Manufacturing Systems and Technology Program), Guest Professor of Huazhong University of Science and Technology (2002, 2006), Professor awarded by China's Jiangsu Provincial Government (2014), and Dean of College of Electrical Engineering and Control Science at Nanjing Tech University (2014-2016). He was the General Chair of 2007 International Conference on Climbing and Walking Robots (CLAWAR), the General Chair of 2009 International Conference on Intelligent Robotics and Applications (ICIRA), the Co-founder of the International Journal of Humanoid Robotics (SCI/SCIE indexed), Co-founder of Singapore-China Association for Advancement of Science and Technology, Co-founder of Robotics Society of Singapore. He has taught the courses such as Robotics, Artificial Intelligence, Applied Machine Vision, Measurement and Sensing Systems, Microprocessor Systems, and University Physics. In terms of scientific research, he has authored three books in English, two books in Chinese, and two edited books in English. He has published several book chapters, over 10 patents of invention, over 40 research papers in scientific journals and over 100 research papers in international conferences. He was the recipient of one best conference paper award from World Automation Congress, the recipient of one best conference paper award from CLAWAR, the recipient of one outstanding paper award from International Journal of Industrial Robot, the recipient of one Gold Prize (S$8K) from CrayQuest, the recipient of one Grand Champion Prize (S$15K) from CrayQuest, the recipient of one A-Star's Best Research Idea Prize (S$5K), the recipient of one Silver Medal from Dragon Design Foundation.

Speech Contents

KnowNet: A Large Knowledge Model

Abstract: With the rise of Artificial Intelligence, we are fortunate to witness the transition from achieving machine's automation to achieving machine's autonomy. On one hand, the success of Artificial Intelligence is guaranteed by the availability of big data which is the result of the formation of large systems that are interconnected by various networks. On the other hand, the importance of Artificial Intelligence is due to the urgent need of having self-intelligence by robots and machines of tomorrow. Interestingly, the critical step toward achieving machine's self-intelligence is the ability of designing large knowledge models instead of improving existing databases. In this invited talk, I will share our research works which aim at providing a general guiding principle for the design of a large knowledge model under the new paradigm of AI 3.0. The published research findings could be found inside 1) Xie M., *Jayakumar K. S. and *Chia H. F., 2004, Meaning-centric Framework for Natural Text/Scene Understanding by Robots, International Journal of Humanoid Robotics, Vol. 1, No. 2, pp. 375-407, and 2) Xie M., 2024, Top-down Design of Human-like Teachable Mind, Special Issue in Celebrating IJHR's 20th Year Anniversary, International Journal of Humanoid Robotics.

Invited Speech XV

Time	**11:10-11:30, March 30th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Assoc. Prof. Jiaxin Cai

Xiamen University of Technology, China

Jiaxin Cai received his Ph.D. degree in Information and Computation Science from Sun Yat-Sen University in 2014. He also received his M.S. degree and B.Sc. degree in Bio-medical Engineering from Southern Medical University in 2011 and 2008 respectively. Currently, he is an associate professor in the School of Mathematics and Statistics at Xiamen University of Technology. He has authored over 36 peer-reviewed papers at academic journals and conferences. His current research interests include machine learning, computer vision and bio-medical engineering.

Speech Contents

Later Temporal Attention in Computer Aided Medical Diagnosis

Abstract: The clinical course of COVID-19, as well as the immunological reaction, is notable for its extreme variability. Identifying the main associated factors might help understand the disease progression and physiological status of COVID-19 patients. The dynamic changes of the antibody against Spike protein are crucial for understanding the immune response. This work explores a temporal attention (TA) mechanism of deep learning to predict COVID-19 disease severity, clinical outcomes, and Spike antibody levels by screening serological indicators over time. We use feature selection techniques to filter feature subsets that are highly correlated with the target. The specific deep Long Short-Term Memory (LSTM) models are employed to capture the dynamic changes of disease severity, clinical outcome, and Spike antibody level. We also propose deep LSTMs with a TA mechanism to emphasize the later blood test records because later records often attract more attention from doctors. Risk factors highly correlated with COVID-19 are revealed. LSTM the highest classification accuracy for disease severity prediction. Temporal Attention Long Short-Term Memory (TA-LSTM) achieves the best performance for clinical outcome prediction. For Spike antibody level prediction, LSTM achieves the best permanence. The experimental results demonstrate the effectiveness of the proposed models. Simple factors like LDH, Mono%, ALB, LYMPH%, DM, and Sex are critical factors in disease severity. LDH, Neu#, hs-CRP, PLT, and Urea are critical factors in clinical outcomes. We further find that Age, RDW_CV, PLT, LDH, eGFR (CKD-EPI), LYMPH#, RDW_SD, PCT, and TCHO are the Top-9 significant predictors of the Spike antibody level. The proposed models can provide a computer-aided medical diagnostics system by simply using time series of serological indicators.

Invited Speech XVI

Time	**15:00-15:20, March 29th**	**Zoom ID**	**835 3930 0002**
Room	**N/A**	**Zoom Link**	**https://us02web.zoom.us/j/83539300002**

Prof. Pascal Lorenz

University of Haute-Alsace, France

Pascal Lorenz (lorenz@ieee.org) received his M.Sc. (1990) and Ph.D. (1994) from the University of Nancy, France. Between 1990 and 1995 he was a research engineer at WorldFIP Europe and at Alcatel-Alsthom. He is a professor at the University of Haute-Alsace, France, since 1995. His research interests include QoS, wireless networks and high-speed networks. He is the author/co-author of 3 books, 3 patents and 200 international publications in refereed journals and conferences. He was Technical Editor of the IEEE Communications Magazine Editorial Board (2000-2006), IEEE Networks Magazine since 2015, IEEE Transactions on Vehicular Technology since 2017, Chair of IEEE ComSoc France (2014-2020), Financial chair of IEEE France (2017-2022), Chair of Vertical Issues in Communication Systems Technical Committee Cluster (2008-2009), Chair of the Communications Systems Integration and Modeling Technical Committee (2003-2009), Chair of the Communications Software Technical Committee (2008-2010) and Chair of the Technical Committee on Information Infrastructure and Networking (2016-2017), Chair of IEEE/ComSoc Satellite and Space Communications Technical (2022-2023), IEEE R8 Finance Committee (2022-2023), IEEE R8 Conference Coordination Committee (2023). He has served as Co-Program Chair of IEEE WCNC'2012 and ICC'2004, Executive Vice-Chair of ICC'2017, TPC Vice Chair of Globecom'2018, Panel sessions co-chair for Globecom'16, tutorial chair of VTC'2013 Spring and WCNC'2010, track chair of PIMRC'2012 and WCNC'2014, symposium Co-Chair at Globecom 2007-2011, Globecom'2019, ICC 2008-2010, ICC'2014 and '2016. He has served as Co-Guest Editor for special issues of IEEE Communications Magazine, Networks Magazine, Wireless Communications Magazine, Telecommunications Systems and LNCS. He is associate Editor for International Journal of Communication Systems (IJCS-Wiley), Journal on Security and Communication Networks (SCN-Wiley) and International Journal of Business Data Communications and Networking, Journal of Network and Computer Applications (JNCA-Elsevier). He is senior member of the IEEE, IARIA fellow and member of many international program committees. He has organized many conferences, chaired several technical sessions and gave tutorials at major international conferences. He was IEEE ComSoc Distinguished Lecturer Tour during 2013-2014.

Speech Contents

Architectures of Next Generation Wireless Networks

Abstract: Internet Quality of Service (QoS) mechanisms are expected to enable wide spread use of real time services. New standards and new communication architectures allowing guaranteed QoS services are now developed. We will cover the issues of QoS provisioning in heterogeneous networks, Internet access over 5G networks and discusses most emerging technologies in the area of networks and telecommunications such as IoT, SDN, Edge Computing and MEC networking. We will also present routing, security, baseline architectures of the inter-networking protocols and end-to-end traffic management issues.

Invited Speech XVII

Time	15:20-15:40, March 29th	Zoom ID	835 3930 0002
Room	N/A	Zoom Link	https://us02web.zoom.us/j/83539300002

Assoc. Prof. Souad BEZZAOUCHA REBAÏ

Électrique Informatique et AutomatiqueLa Rochelle, France

Souad BEZZAOUCHA REBAÏ graduated from Ecole Nationale Polytechnique, Algeria, where she received an Engineer Diploma in Automation Control (Valedictorian with highest class) and a Master of Science in Electrical Engineering in 2005 and 2007, respectively.
She received a Master degree in Computer Science Engineering from Ecole Polytechnique, France in 2010 (Excellence Scholarships from the Foundation of Ecole Polytechnique) and spent one year from 2007 to 2008 as a research student at the Graduate School of Engineering, Kyoto University, Japan (Japanese Government-Monbukagakusho Scholarship). She received her PhD from Université de Lorraine, France, in 2013. From January 2014 to May 2015 she was a postdoctoral fellow and lecturer at Laboratoire de l'Intégration du Matériau au Système (IMS), Bordeaux, France and Bordeaux Polytechnic National Institute (INP) Enseirb-MatMeca. From June 2015 to May 2020 she was a Research Associate at the Interdisciplinary Center for Security, Reliability and Trust (SnT), University of Luxembourg, where she worked on various robotic and cyber-security projects. Since March 2021 she was appointed as Associate Professor at EIGSI-La Rochelle and she is attached to the MIA laboratory- La Rochelle University as associate researcher since Jan 2023. Her research interests are in robust control theory, observers design, nonlinear Systems, robotics and cyber security.

Speech Contents

Linear Matrix Inequalities (LMIs) in Control Engineering: An introduction to LMI-based approach

Abstract: In control theory, LMIs are widely used for stability analysis, controller design, and optimization of dynamic systems. LMIs provide convex constraints, making them computationally efficient to solve using numerical methods. In the following presentation, we will discuss the use of LMIs, interests of convex optimization in control and application cases. Indeed, control problems often involve constraints on system stability, performance, and robustness. When
formulated as convex problems, they can be solved efficiently, providing global optimal solutions, making the problem efficiently solvable. An illustrative example application will be presented in order to highlight the approach efficiency.

Session 1

March 29th, 2025
Time Zone: GMT+8

Topic: Computer Vision and Artificial Intelligence
Time: 14:10-15:40 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A714
Session Chair: Assoc. Prof. Yanan Yu, Tianjin University of Technology and Education, China

Onsite

CC0047

Multi-Stage Human Motion Prediction Algorithm Based on Spatiotemporal Graph Convolution

Zongli Liu[1], Yanan Yu[1], Hongli Zhao[2] and Ke Du[1]
1. Tianjin University of Technology and Education, China
2. Tianjin Boshen Ruichuang Tech-nology Co., Ltd., China

Abstract-Human motion prediction is a classic computer vision task, with the core objective of predicting future motion sequences based on given historical motion sequences. Multi-stage prediction strategies have achieved significant success in the field of human motion prediction. However, existing network models struggle to effectively integrate multi-stage frameworks, leading to limitations in dynamic ad-justment of key features and long-term dependency modeling, ultimately constraining prediction accuracy and robustness. To address these issues, this paper proposes a multi-stage network model that fuses spatiotemporal features. Spatiotemporal convo-lution is utilized to enhance the long-term dependency of spatio-temporal information in motion sequences, while a dynamic gating mechanism is introduced to adaptively adjust feature weights. Finally, the proposed network model is integrated into each stage of multi-stage prediction to further optimize the efficiency of spatio-temporal feature extraction. Experimental results demonstrate that the proposed method outperforms existing approaches on multiple datasets, improving performance by 4% on Human3.6M and 3% on CMU-MoCap.

CC0048

Sparse Bidirectional Two-Dimensional Principal Component Analysis for Feature Extraction

Wenwen Liu, **Xinyuan Kang** and Weijia Feng
Tianjin Normal University, China

Abstract-The use of computer vision technology to achieve recognition has become an important research topic in the field of artificial intelligence, and feature extraction is an important step in automatic recognition using computer vision technology. The spatio-temporal feature descriptor extracts a feature vector with excessively high dimensionality, which increases the computational complexity of various classification tasks and extends the processing time. PCA aims to map high dimensional data to low-dimensional space through linear transformation and is widely used in the fields of dimensionality reduction and feature extraction, data compression, and pattern recognition. However, PCA has shortcomings, such as losing the inherent two-dimensional features of the image and producing principal components that are unfavorable for interpretation. To address these issues, this paper proposes a sparse bidirectional two-dimensional principal component analysis algorithm (S2D2PCA). This algorithm is developed by applying sparse constraints to the principal components of the bidirectional two-

dimensional principal component analysis algorithm (2D2PCA). The S2D2PCA algorithm extracts features directly from two-dimensional images and preserves their two-dimensional features; improves the interpretability of the principal components, reduces redundancy, and improves the efficiency of feature extraction. In this paper, experimental validation is conducted on the ORL and Weld datasets. The results demonstrate that the classification performance of the S2D2PCA algorithm is superior to that of PCA and 2D2PCA.

CC0053

Occluded Pedestrian Re-identification Based on Pose-guided Feature Enhancement

Lang Yang[1], Yanan Yu[1], Hongli Zhao[2], Erjie Li[1] and Zhengxuan Guo[1]
1. Tianjin University of Technology and Education, China
2. Tianjin Boshen Ruichuang Technology Co., Ltd., China

Abstract-Pedestrian re-identification (Re-ID) is a key technology that uses artificial intelligence to solve core problems in the field of public safety, aiming at quickly recognizing specific individuals across images or videos captured by cameras. Still, the reality of complex occlusion environments can significantly degrade its performance. In order to effectively solve the problems of over-reliance on external information and neglect of key features, limited global and local feature expression ability, and weak anti-interference ability against noise in current occluded pedestrian re-identification methods, a pose-guided feature enhancement model is proposed. Firstly, a patch enhancement module is introduced into ViT (Vision Transformer) to enhance the original patch sequence itself and optimize the feature extraction paths with the aid of a joint pose estimation network, then the feature expression accuracy is improved by the feature aggregation module, and finally, the learning ability of the model is further enhanced by the introduction of BIN normalization. Experiments demonstrate that the model obtains 71.1% Rank-1 and 62.0% mAP on the standard data set Occluded-Duke, which improves 6.9% Rank-1 and 6.3% mAP compared with the baseline model, indicating remarkable performance improvements.

CC0054

Robust Subspace Learning Via Binary Weights for Face Recognition

Nan Wang, **Jiayi An** and Weijia Feng
Tianjin Normal University, China

Abstract-In extreme noise environments, existing subspace learning algorithms applied to face recognition tasks exhibit limitations in feature extraction capabilities, robustness, and anomaly detection. To address these issues, this paper proposes a robust two-dimensional principal component analysis (2DPCA) method based on binary weights, aimed at effectively identifying and eliminating outliers, thereby enhancing the robustness and efficacy of face recognition systems in practical applications. By incorporating a binary weight strategy and optimal means, our model maintains a high capability for low-dimensional representation across varying noise levels. We conducted extensive experiments by artificially adding 40% and 60% occlusion noise into four publicly available facial datasets: Aberdeen, YALE, CaltechFaces, and GT. The experimental results demonstrate that the proposed algorithm achieves satisfactory outcomes even when handling severely corrupted data, particularly under conditions of 60% noise, where its reconstruction performance significantly surpasses that of the comparison algorithms, underscoring its superior robustness and adaptability.

CC0035

Feature-Enhanced Small Object Detection Algorithm for Surface Defects of Cigarette Pack

Xuesong Ding, Jianbin Gu, Jinlong Geng, Xiang Liu, Haotian Li, Jiaqi Luo and Xiao Wang
Hebei Baisha Tobacco Co., Ltd., China

Abstract-The surface defects of cigarette pack are characterized by varying scales and complex backgrounds. Small target detection algorithms, as a potential solution, have been proven to have the ability to identify such surface defects. However, the existing small target detection algorithms still have many limitations, especially in terms of detection accuracy, model complexity, and computing speed. To address these challenges, we propose a lightweight cigarette pack defect detection algorithm, YOLOv5s-FDE. In the backbone network, we introduce the GSConv structure, which incorporates channel shuffle and depthwise separable convolution to construct the feature enhancement module (FEM) with a cross-stage architecture. It aims to reduce information loss and enhance feature utilization. In the feature fusion stage, we introduce the DySample upsampling module, which effectively preserves target detail features through a dynamic sampling mechanism, and reduces the memory usage. Finally, the E-Triplet cross-dimensional attention module is proposed, which leverages a multi-branch structure to enhance cross-dimensional information interaction and suppresse irrelevant background information. Experimental results demonstrate that, compared to the YOLOv5s baseline algorithm, YOLOv5s-FDE achieves an average precision of 95.6%. The number of model parameters and computational cost are reduced by 0.88M and 3.5G, respectively. The proposed method outperforms recent state-of-the-art detection algorithms, whiles providing theoretical support for intelligent defect detection of cigarette pack.

CC0043

Size Measurement Method of Cigarette Pack Based on Key Point Detection and Stereo Matching

Jianli Bu, Jianbin Gu, Jinlong Geng, Xiang Liu, Haotian Li, Xiao Wang and Jiaqi Luo
HeBei Baisha Tobacco Co., Ltd, China

Abstract-In modern production, the process of gluing and forming the outer package of cigarette packs has been fully automated. However, due to the influence of equipment operation deviation, the molded cigarette packs may be shifted during the gluing process, causing the final outer packaging to fail quality standards. Currently, cigarette factories mainly rely on manual measurement to inspect critical dimensions of the cigarette. To reduce human resource consumption and enhance inspection accuracy, a method for measuring the part dimensions of the cigarette pack based on key point detection and stereo matching is proposed. Specifically, the YOLOv8 algorithm with excellent real-time performance is first used to recognize the Lotus cigarette pack in the picture. The Lite-HRNet algorithm, which has the ability of high-resolution feature extraction, is used to detect the key points of the recognized cigarette pack, thus obtaining the 2D pixel coordinate information of the parts to be detected. Next, utilizing the depth information provided by the CREStereo algorithm, the 2D pixel coordinates of the keypoints are converted to 3D camera coordinates. Finally, the spatial distances between key points are calculated based on the Euclidean distance formula to produce accurate measurements. In the actual measurement of 20 packs of cigarettes, the average absolute errors of the distances between the left and right side silver waistlines were 2.68 mm and 2.34 mm, respectively. This paper presents a method that can effectively meet the practical needs for measuring the dimensions of cigarette pack parts.

Session 2

March 29th, 2025
Time Zone: GMT+8

Topic: Remote Sensing and Computer Vision
Time: 14:10-15:40 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A715
Session Chair: Assoc. Prof. Jie Li, Hubei University of Technology, China

Onsite

CC0074

LPDDN: An Embedded Real-Time Pavement Defect Detection Model Under Computationally Constrained Conditions

Jiaxi Guan[1], **Chenxiang Li[2]**, Yi Xiao[1], Yuan Xia[1], Huafang Ou[1] and Linyao Zhou[1]
1. Hubei Institute of Measurement and Testing Technology, China
2. Wuhan University of Science and Technology, China

Abstract-Road transport is critical to economic development, yet pavement defects like cracks and potholes pose significant safety risks. Traditional detection methods are inefficient and resource-intensive, while existing deep learning models often fail to meet real-time requirements on edge devices. To address these challenges, we propose LPDDN (Lightweight Pavement Defect Detection Network), a novel embedded real-time detection model. LPDDN integrates a Lightweight Information Sharing Detection Head (LISD Head) to reduce model complexity, a WIoU loss function to enhance detection accuracy, and a Layer-Adaptive Magnitude-based Pruning (LAMP) method to optimize computational efficiency. Experiments on the China-MotorBike dataset demonstrate LPDDN's superior performance, achieving 94.9% mean average precision (mAP) with only 858,941 parameters. Deployment on NVIDIA Jetson Xavier NX further proves its real-time capability with detection times under 50 ms. LPDDN offers an efficient solution for real-time pavement defect detection, balancing accuracy and computational efficiency for practical applications.

CC0076

CPRNet: Class Prototype Reweight Neural Network for Few-Shot Object Detection

Shuhong Shan[1], Haoran Li[1], Yunfan Chen[2] and **Wenqi Zheng[1]**
1. Harbin Engineering University, China
2. Hubei University of Technology, China

Abstract-Few-Shot Object Detection (FSOD) addresses the challenge of detecting and localizing target objects in a test set when only a limited number of training samples are available. In few-shot learning, the support set provides reference samples to guide the model, while the query set evaluates its performance on unseen data. While the many FSOD models struggle to fully capture contextual information and fine-grained details during feature extraction, which can lead to overfitting due to limited training diversity. To address these limitations, we propose a framework, named Class Prototype Reweight Network (CPRNet), that introduces a dual-strategy improvement based on DEVIT module. First, the feature enhance (FE) module enhances feature extraction through a spatial attention mechanism and an adaptive feature enhancement algorithm. Second, the prototype

reweight (PR) module refines class prototype construction using a dual-feature aggregation layer and a class prototype optimization algorithm. Experimental results on COCO datasets show that our proposed CPRNet achieves significant performance gains, demonstrating its effectiveness in few-shot object detection tasks.

CC0077

Research On Monitoring Methods For Construction Waste Disposal Site Based On Matched Point Cloud

Guanghuan He[1], Bihong Li[2], Jing Rong[1], Bo Liu[3], **Yan Zhang[4]** and Jinlong Meng[1]
1. Guangxi Polytechnic of Construction, China
2. Nanning Survey and Design Institute Group Co., Ltd, China
3. Henan Institute of Science and Technology, China
4. China University of Geosciences, China

Abstract-Dynamic monitoring of urban construction waste dumps is an important part of urban construction and development work, and this paper proposes a monitoring method of construction waste dumps based on ground-matched point cloud. Taking the construction waste disposal site in Qingxiu District, Nanning City, China as an example, a drone is used to collect 4-phase aerial survey data and generate the ground-matched point cloud through data processing; for the hole problem of the filtered ground point cloud, Kriging, IBW, RBF, and LSSVM algorithms are used to repair the holes in the ground point cloud, and the repairing accuracy of the models is analyzed in comparison with each other, so as to construct the most suitable Digital Elevation Model (DEM). ; combined with the actual values obtained by GNSS-RTK, the volume of the DEM of the 4-phase building waste disposal site is checked for accuracy; the DEM of the 4-phase building waste disposal site is used for differential processing to obtain the volume change values of the 4-phase DEM, and combined with the Digital Orthophoto Map (DOM) generated based on the matched point cloud for monitoring the building waste disposal site. The results show that: 1) The accuracy of point cloud hole repair using the LSSVM algorithm improved by 22.6%, 17.8%, and 20.3% compared to the traditional Kriging, IBW, and RBF algorithms, respectively; 2) The volume differences between the DEMs of the four-phase construction waste disposal sites constructed based on the LSSVM algorithm and the actual volumes measured by GNSS-RTK were 1.81%, 2.02%, 1.73%, and 1.96%, respectively; 3) Multi-phase DEMs and DOMs can effectively monitor construction waste disposal sites. This study aims to provide technical support for the dynamic operational management and safe production of urban construction waste disposal sites.

CC2013

Enhanced Sensitivity and Durability of f-MXene/MXene/PU Sponge-Based Pressure Sensors for Real-Time Health Monitoring

Zhong Zheng and Soukaina Benchaouia
Hubei University of Technology, China

Abstract- The development of flexible and wearable pressure sensors has gained significant attention due to their potential applications in health monitoring, robotics, and human-machine interfaces. However, existing pressure sensors face challenges such as limited sensitivity, narrow detection ranges, and difficulties in scalable fabrication. In this study, we propose a novel multilayer f-MXene/MXene/PU sponge-based flexible pressure sensor, designed to address these limitations. The sensor was fabricated using a dip-coating method, where functionalized MXene (f-MXene) and MXene nanosheets were sequentially deposited onto a polyurethane (PU) sponge substrate. This structure enhances the sensor's mechanical flexibility, electrical conductivity, and air permeability while

preserving its compressibility. Experimental results demonstrate that the optimized multilayer f-MXene/MXene/PU sensor exhibits a high sensitivity of 7.4 kPa^{-1} over a broad pressure sensing range of 58 kPa. Additionally, the sensor features rapid response and recovery times (340 ms and 300 ms, respectively) and maintains stable performance over 200 consecutive loading-unloading cycles, indicating strong durability. Application tests further validate its capability to detect a wide range of pressures, from subtle touch interactions (e.g., finger pressing) to high-force applications (e.g., foot stepping). These findings highlight the potential of MXene-based composite sensors in wearable electronics and real-time pressure monitoring. Future research should focus on enhancing the long-term stability of MXenes against oxidation and integrating wireless communication modules for smart sensor applications.

CC0040

Research on Mobile Robot Path Planning Based on Improved RRT Algorithm

Xiang Wang[1], **Songhang Xie[1]**, Haichuan Li[2] and YuKe Xie[1]
1. Hebei University of Science and Technology, China
2. Meteorological Observation Centre Cangzhou Meteorological Administration Cangzhou, China

Abstract-To address the limitations of the traditional Rapidly Exploring Random Tree* (RRT*) algorithm, such as slow convergence and poor adaptability to dynamic environments, an improved dynamic path planning method is proposed by integrating the RRT* algorithm with the Dynamic Window Approach (DWA) to achieve efficient obstacle avoidance for mobile robots in complex dynamic environments. The improved RRT* algorithm generates a globally optimal and safe path based on known environmental information. The proposed method optimizes sampling points to ensure path optimality while reducing search time. It introduces the concept of maximum path nodes to eliminate redundant nodes in the extended tree, enhancing algorithm efficiency. Additionally, the safety of the global path is ensured by removing dangerous nodes generated by the RRT* algorithm and applying a greedy algorithm to remove redundant nodes, shortening the global path length.

CC0069

Path Planning for Robot Arms: Leveraging an Enhanced RRT-Connect Algorithm

Qun Zhao, Yan Shang, Yu Zhou, Xingwei Zhu and **Rui Yan**
Hebei University of Science and Technology, China

Abstract-In a complex three-dimensional environment, the path planning of a robot arm is the key to ensuring its efficient and safe completion of tasks. Traditional path planning algorithms such as Dijkstra and A* are challenging in meeting the requirements of real-time and high efficiency in high-dimensional complex space. Although the RRT-CONNECT (Rapidly-exploring Random Tree-Connect) algorithm shows high efficiency in path planning, its fixed step parameters lead to insufficient searching ability in narrow areas, and the generated path has redundancy and redirection problems. To solve these problems, this paper proposes a path-planning method for the robot arm based on an improved RRT-CONNECT algorithm. Firstly, we introduce the target bias strategy and optimize the sampling method to make the distribution of sampling points near the target point more natural and uniform and therefore improve the search efficiency. Secondly, the adaptive step size technology is used to subdivide the path segment and detect the collision point by point, which improves the accuracy of collision detection and the reliability of path planning. In addition, pruning technology and B-spline fitting technology are also introduced in this paper to remove redundant nodes and roundabout sections in the path, making the path smoother and more straightforward and further improving the robot arm's path quality and

navigation efficiency. The experimental results show that the improved algorithm can generate high-quality paths more efficiently in complex three-dimensional environments, significantly improve the path planning performance of the robot arm, and meet the real-time and high-efficiency requirements in practical applications.

Session 3

March 29th, 2025
Time Zone: GMT+8

Topic: Optical 3D Imaging and Measurement
Time: 14:10-15:40 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A717
Session Chair: Assoc. Prof. Qijian Tang, Shenzhen University, China

Onsite

CC0067

A Method of Separating Interreflection in Digital Fringe Projection Profilometry

Yu Fei Liu[1], Qi Yun Zeng[1], Xiao Yu[2], Xiao Li Liu[1], Xiang Peng[1] and Qi Jian Tang[1]
1. Shenzhen University, China
2. Shenzhen Anhua Optoelectronics Technology Co., Ltd Shenzhen, China

Abstract-Optical 3D surface measurement techniques, such as Fringe Projection Profilometry (FPP), are commonly used methods for reconstructing object surfaces. However, when traditional FPP is applied to measure regions involving interreflection, the measured phase can be distorted, leading to reduced accuracy in 3D shape measurement or even measurement failure. In this study, we propose to employ high-frequency binary masks to modulate phase-shifting and Gray code patterns and eliminate the interreflection among object surfaces. Experimental results demonstrate that this method can effectively separate direct and indirect illumination and suppress interreflection to enhance the accuracy of 3D object measurement.

CC0032

Hybrid Iterative Constrained Hologram Generation Method For Multi Plane Imaging

Qinyang Li[1], Zhongsheng Zhai[2], Zhen Zeng[2], Zhi Xiong[2], Wei Feng[2], Qinghua Lv[2] and Xuanze Wang[2]
1. Huazhong University of Science and Technology, China
2. Hubei University of Technology, China

Abstract-This paper presents a Computer Generated Holography (CGH) algorithm, fusing the Gerchberg-Saxton (GS) algorithm, angular spectrum method, and double phase amplitude coding method. Tackling challenges of speckle noise and interlayer crosstalk, the algorithm optimizes multi-image reconstruction quality through constraint factors. The algorithm structure combines sequential GS method and uses double phase amplitude coding method to reduce interlayer crosstalk and speckle noise. Validated through simulations and experiments, the algorithm demonstrates superior reconstruction, making it applicable in diverse fields such as biology, medicine, and engineering design, offering a robust solution for multi-plane image reconstruction challenges.

CC0057

Surface Defect Detection of Photodiodes Based on FPGA

Xinfang Zhao, Qinghua Lyu, **Hui Zeng**, Benyuan Chen, Zhongsheng Zhai, Hui Lyu and Yuting Long
Hubei University of Technology, China

Abstract- Surface defect detection of photodiodes is crucial for ensuring the stability and reliability of device performance. It effectively prevents potential failures, improves product quality, and enhances production efficiency. This study addresses the surface defect detection of photodiodes by constructing a multi-class defect dataset and proposing a hardware-accelerated detection method. First, a 16-bit fixed-point quantization and batch normalization (BN) layer fusion strategy were employed to optimize and train the YOLOv4-tiny model, improving its computational efficiency. Next, based on the ZYNQ7020 hardware platform, hardware architectures for the convolution, upsampling, and pooling modules were designed. The efficient hardware acceleration of the neural network was achieved using multi-channel parallel convolution operations and multi-path parallel data transmission techniques. Experimental results demonstrate that the proposed design achieved excellent detection performance on a test set of 100 photodiodes surface defects, with an average precision of 90.4%. The system can perform photodiodes surface defect detection on FPGA with a frame rate of 3.5 frames per second and a power consumption of only 3.08W.

CC2007

TF-AMSKPconv: Spatio-temporal Feature Fusion Based Semantic Segmentation Of Dynamic Point Clouds For Autonomous Driving Lidar

Zihan Zhou, Yi Wang and Junyao Gao
Chongqing University of Technology, China

Abstract-In autonomous driving, dynamic point cloud semantic segmentation is crucial for environmental perception and dynamic object detection. This study proposes the TF-AMSKPconv Net, a time feature-adaptive multi-scale KPconv network, to enhance dynamic object segmentation. The network integrates a time feature extraction module (TF) to capture dynamic object features and combines them with the point-based spatial feature extraction module (AMSKPconv) to improve segmentation performance. First, spherical projection transforms point cloud data from consecutive frames into 2D images, capturing dynamic changes over time. Inter-frame difference features are then used to extract motion features, which are processed with multi-scale strip convolution kernels to capture dynamic features at various scales. A channel attention mechanism is applied to weigh the features, enhancing the model's focus on important details. Finally, dynamic and spatial features are fused through weighted normalization, and a classification layer is used for semantic segmentation. Our experimental results demonstrate that the time-information-based projection method significantly outperforms traditional spatial feature-based approaches for dynamic point cloud semantic segmentation, especially in complex environments. The proposed TF-AMSKPConv Net achieves a competitive performance, attaining a mean Intersection over Union (mIoU) of 57.98% on the SemanticKITTI dataset for multi-frame point cloud semantic segmentation.

CC5001

Research on Epilepsy Detection Method Based on Spatiotemporal Feature Fusion and LASSO Optimization

Junqi Yang[1], Chunjun Zheng[1], Pengfei Ma[2], Junqi Li[3], Xinjing Song[1], Caixia Wu[1] and Boyang Li[1]
1. College of Computer and Software, Dalian Neusoft University of Information, China
2. Modern Industrial College of Health Management, Jinzhou Medical University, China
3. College of Land Resources and Environment, Jiangxi Agricultural University, China

Abstract-In the age of artificial intelligence, BCI has become very significant, but still, the epileptic signals' detection in EEG requires higher accuracy because of relying on single features only. To improve this challenge,

the paper presents a new epilepsy detection technique called TSL by combining spatiotemporal features and feature optimization with LASSO. Feature classification is done through SVM. Experimental results of CHB-MIT are carried out using proposed methodology and the methodology used previously to establish comparison. The results obtained from the MIT dataset show that TSL achieves an accuracy of 0.9779, a sensitivity of 0.9853, and a specificity of 0.9732, performing better than traditional methods using either temporal or spatial features. These reflect very strong performance for TSL and thus underscore its potential for early detection and timely intervention in clinical practice.

Poster Session

March 29th, 2025
Time Zone: GMT+8

Topic: Object Detection and Algorithms
Time: 15:40-16:00
Session Chair: Assoc. Prof. Zhi Xiong, Hubei University of Technology, China

Onsite

CC0017

Open-Set Environment Activity Recognition Based On Cross-Layer Selected Class Anchor Clustering Learning With An Acoustic-Seismic Sensor System

Jiakuan Wu, Nan Wang, Huajie Hong, Wei Wang, Kunsheng Xing and Yujie Jiang
National University of Defense Technology, China

Abstract-The acoustic and seismic sensor system, with its advantages of low cost, low power consumption, ease of deployment, strong obstruction resistance, and high concealability, is particularly well-suited for the development of environmental activity recognition systems. However, existing classification methods for acous-tic and seismic signals are generally constrained to closed-set scenarios, where the test-phase input classes must align with the training-phase classes. This limi-tation significantly hinders the deployment of such sensor systems. To overcome this challenge, open-set recognition is essential. In this paper, we propose an open-set recognition algorithm, Cross-layer Selected Class Anchor Clustering Learning, which leverages selective learning based on CAC Loss and incorpo-rates a cross-entropy constraint in the penultimate layer of the network. To evalu-ate the performance of our algorithm, we have constructed a comprehensive acoustic and seismic dataset that includes typical environmental activities, and compared our method with other state-of-the-art algorithms. The results demon-strate that the proposed algorithm not only achieves effective classification of closed-set samples but also delivers high recognition accuracy for open-set un-known samples.

CC2003

Research on Adaptive Lane Line Detection Algorithm Based on OpenCV-Python

Jingjing Zhang, Wenjun Ding, Yi Xu and Xu Yao
Anhui Vocational and Technical College, China

Abstract-Accurate and efficient detection and recognition of lane lines is one of the prerequisites for achieving autonomous driving in automobiles. To eliminate the influence of traditional algorithms on complex road conditions such as lane line wear and shadow interference, a color segmentation method and adaptive sliding window lane line detection algorithm based on OpenCV-python are proposed. The color segmentation method uses python to segment the L channel in the color space HLS and the b channel in the color space Lab, which are used as white and yellow lane lines respectively, and merge them to form a new bird's-eye view, eliminating the interference of other edges and noise; Adaptive sliding window is an improved algorithm for extracting lane lines from histogram sliding windows. The minimum effective pixel value is pre-set, and the number of lane line pixels in the first rectangular window is counted. By comparing with the minimum effective pixel value, the position of the sliding window is adaptively adjusted to accurately extract lane lines. Real car recording of video data under

different environmental road conditions, using this algorithm for lane line detection and recognition of video frame images, shows good performance.

CC0027

Perceptual Encoding Optimization Algorithm Guided by Human Eye Focusing Mechanism

Xinran Wan, Huiping Deng and Sen Xiang
Wuhan University of Science and Technology, China

Abstract-Perceptual coding algorithms play a crucial role in eliminating visual redundancy in video compression. Despite their promising results, there remains potential for further optimization, particularly in leveraging human visual characteristics for rate control and effectively applying just noticeable distortion (JND)—a threshold defining human visual sensitivity to changes.In this paper, we propose a novel HEVC rate control scheme inspired by the focusing mechanism of the human visual system (HVS). By integrating saliency and focal distance information, we construct a focusing factor and incorporate it into the encoding framework. This focusing factor serves as a key parameter for rate control, optimizing LCU-level bit allocation and enhancing video coding efficiency.Furthermore, we combine the focusing factor with a JND-based model to design a Gaussian filter that smooths non-focused areas, reducing background redundancy while enhancing visual quality. The proposed algorithm is implemented on HM16.7, and experimental results show bitrate savings and PSNR improvements ranging from 0.1 to 0.5 dB compared to state-of-the-art algorithms. Additionally, subjective evaluations confirm significantly enhanced visual quality at equivalent bitrates.

CC0033

6D Pose Estimation And Grasping Based On Deep Learning With MBM

Yan Ziqingying and Yang Yongda
Xiangtan University, China

Abstract-Aiming at the problem of low recognition rate and poor robustness of traditional robotic manipulation visual recognition algorithm, a algorithm based on Minimum Bounding Model(MBM) is introduced based on the existing researches of deep learning 3D vision in this paper to accurately detect and identify the target. This method is based on the improved YOLO algorithm and is trained with data made by a new data method. Through a single RGB image, the object can be directly identified and the 6D pose information can be estimated simultaneously. Based on the above process, the path planning algorithm can be used to capture the object. Experimental test is proved that the method can accurately identify object and perform pose estimation. The grasping experiment is carried out on the Co602a manipulator. The results show the effectiveness of the method.

CC0045

A Method for Clustering Targets in Sonar Images

Chaolong Liu, Juan Cheng and Shuai Shao
University of Information Engineering, China

Abstract-Sonar image targets clustering is an important research topic that requires further exploration. In this paper, to address the issue of poor clustering performance when applying clustering methods designed for optical images to sonar images, due to the significant differences between sonar and optical image targets as well as the uneven distribution of sonar image targets classes, an effective clustering method for sonar image targets is

proposed based on the advanced Contrastive Clustering (CC) method. Firstly, Focal loss is introduced into the loss function of the instance-level contrastive head to mitigate the issue of uneven distribution of classes instances. Subsequently, the data augmentation method is redesigned to accommodate the characteristics of sonar image targets. The experimental results demonstrate significant improvements across all three metrics: Normalized Mutual Information (NMI), Accuracy (ACC), and Adjusted Rand Index (ARI), with maximum relative gains of 16% in NMI, 4% in ACC, and 16% in ARI.

CC0052

SF-YOLO: A Lightweight and Precise Approach for Small Object Detection in Aerial Images

Tajrian Abm Shafayet
Huazhong University of Science and Technology, China

Abstract- Detecting small objects in aerial data presents a persistent challenge due to their dense distribution, limited feature representation, and large-scale variations. For example, detecting small objects is even harder for aerial data than standard object detection tasks. Traditional object detection techniques tend to struggle with maintaining detailed features, achieving balanced feature aggregation, and meeting computation requirements for low-end devices. The YOLO model proposed in this article only draws on the general frame work structure of the YOLO model (after YOLOv3), namely the backbone network, neck, and detection head. This article redesigns these three parts to improve their detection performance for small-scale targets. YOLO-based models are leading models of real-time object detection and have become an industry standard in applications like surveillance or traffic monitoring where speed takes precedence. Because of the fast, efficient performance of YOLO models, they are well suited for use in resource-constrained applications like object detection on UAVs (Unmanned Aerial Vehicles) or embedded systems. Nevertheless, although YOLO excels for general-purpose detection, it has weak accuracy for small objects, which is especially problematic for aerial imaging applications, where small instances of objects will be close to get her in a scene, significantly altering the distribution and expected heights of the bounding boxes, particularly from large variances in distance due to the highly variable nature of aerial image resolution. To avoid these constraints, we introduce a lightweight design integrating Rep VGG Block, Selective Fusion Block (SF Block), Spatial Pyramid Pooling Fast with Width and Depth (inferred) (SPPF-WD) module, and Specialized Small-Object Detection Branch. This increases feature extraction for multi scale feature fusion and small-object detection while maintaining a reduced architecture, which is beneficial in real-time scenarios on limited computational capacity platforms. The proposed model, SF-x, does better on the SIMD dataset than the best existing models, with mAP@50 and mAP@50-95 that are 1.2% and 3.6% higher, respectively respectively. Furthermore, it demonstrates efficacy by significantly improving performance at every object scale, thereby solidifying its state-of-the-art status in terms of precision and robustness. The model demonstrates a 14.8% increase in APS and a 10.5% reduction in false positives compared to baseline models. Importantly, the lightweight SF-n configuration (2.99M, 8.7 GFLOPs) is designed for edge devices. On the other hand, the scalable configurations (SF n, SF-m, SF-l, SF-x) are able to achieve the best trade-off between accuracy, computational cost, and inference speed. The findings highlight the efficiency and adaptability of the proposed lightweight paradigm, paving the way for it to become a yardstick solution for small object detection in aerial and remote sensing images. Compared to similar state-of-the-art solutions, if you prioritize accuracy and real-time, you can use it in UAV surveillance, traffic, environmental assessment, or anything similar in a resource-constrained environment. The code and trained weights for the proposed framework can be found at this link: GitHub Repository: SF YOLO.

CC5002

Research on Artificial Intelligence Empowering Innovation in Traditional Chinese Culture

Chenggong Zhai[1], Liming Du[1], Han Qiao[1], YanXiang Zhang[1], XinGuang Yuan[1] and Hongri Zhu[1]
1. Army Logistics Academy, China

Abstract- This article explores the application and prospects of artificial intelligence technology in the innovation of traditional Chinese culture. Through literature research and case analysis, this article elaborates on the specific applications of artificial intelligence in cultural heritage, artistic creation, educational dissemination, and industrial integration. Research has found that artificial intelligence technology can effectively promote the digital protection, innovative transformation, and modern dissemination of traditional culture. At the same time, this article also points out the challenges faced in the process of technological application and proposes corresponding countermeasures and suggestions. The research results indicate that the deep integration of artificial intelligence and traditional culture will open up new paths for the inheritance and development of excellent traditional Chinese culture.

CC0038

Light Field Measurement with Space Consistency

Sibo Huang[1,2] and Zewei Cai[2]
1. Nanjing University of Science and Technology, China
2. Shenzhen University, China.

Abstract-Focal scanning light field imaging uses a camera to capture images at different fo-cal planes to form an image stack, from which full-pixel-resolution light field in-formation can be computationally recovered. However, the depth mapping be-tween the camera’ s image and object space is nonlinear, which leads to a discrep-ancy between image-space capture and object-space reconstruction. In this paper, a consistent light field measurement method is proposed. This allows the image stack in image space to be transformed into object space, with imaging distortions corrected to ensure space consistency for light field reconstruction. Subsequently, the transmission distance prior in object space can be obtained to reconstruct the object-space light field from the corrected image stack. After camera calibration, the reconstructed light field becomes measurable, allowing the three-dimensional measurement with camera parameters. Experimental results demonstrate that the proposed method can achieve high-precision light field measurement.

CC0065

Lightweight UAV-Based Road Disease Detection Algorithm

Jinbo Guo, Shenghuai Wang, Fenghua Xu, Xiaohui Chen, Chen Wang and Wei Zhang
Hubei University of Automotive Industry, China

Abstract-To enable efficient operation of drone detection models on resource-constrained devices while maintaining high accuracy, an improved lightweight UAV pavement damage detection algorithm, named LHP-YOLO, is proposed. This algorithm enhances YOLOv8n by focusing on two aspects: lightweight design and improved detection accuracy. First, the RG-C2f module is introduced to replace the original C2f module in YOLOv8, reducing redundant computations in the model. Additionally, a lightweight detection head, Detect-T3G, is designed to minimize the number of parameters. To further improve the model's performance in complex backgrounds and small object detection, Content-Guided Attention Fusion (CGA Fusion) is incorporated into the

feature fusion structure of the YOLOv8n neck network. Moreover, Channel-Wise Knowledge Distillation (CWD) is applied to boost the accuracy of the lightweight model. Experimental results on the publicly available pavement damage dataset RDD 2022 show that, compared to the original YOLOv8n model, LHP-YOLO achieves a 3.4% increase in mAP50 while reducing the model size and computation by 41.3% and 51.8%, respectively, with parameters totaling 1.76M and 3.9 GFLOPs. These improvements meet the real-time performance requirements for UAV-based pavement damage detection.

CC0072

Image De-Raining Method Based on PReNet and Diffusion Model

Maoping Ran[1], **Yuwen Wang[1]**, Jiaxi Guan[2], Bin Wu[2], Chenxiang Li[1] and Pu Liu[1]
1. Wuhan University of Science and Technology, China
2. Hubei Institute of Measurement and Testing Technology, China

Abstract-Existing image de-raining methods often suffer from incomplete rain streak removal and the disappearance of background information during the de-raining process. This paper suggests a novel de-raining approach based on the composition of PReNet and diffusion models. By improving the PReNet network structure, incorporating a CA Layer attention mechanism module, and enhancing the recursive layers, the feature extraction capabilities are strengthened. The noise estimation network in the diffusion model is replaced with the improved PReNet structure to leverage its denoising abilities for effective rain removal. Network training is conducted on synthetic datasets, and both qualitative and quantitative results are obtained. Quantitatively, on the Rain100H and Rain100L datasets, the proposed method achieves average PSNR values of 26.51 dB and 33.24 dB, and SSIM values of 0.8310 and 0.9633 respectively, outperforming current state-of-the-art methods. Qualitatively, the suggested approach retains more image details shows good generalization to various kinds of rain streaks, and avoids image distortion.

CC2002

Research on Traffic Sign Detection Algorithm under Complex Road Conditions

Zhou Huang, Ruijia Yao, Runyang Xiao and Ziqun Du
Hubei University of Technology, China

Abstract-Traffic sign recognition under complex road conditions is an important research content in the field of automatic driving, for the current traffic sign recognition of omission, false detection, and poor generalization under complex road conditions, this paper proposes a traffic sign recognition algorithm based on improved YOLOv5. Firstly, the inverse perspective transformation is performed on the im-age captured by the onboard camera to convert the image from the fish-eye view to the normal perspective view to make the feature extraction elements more com-plete. To deal with multi-scale targets, a feature pyramid structure is introduced in YOLOv5. By adding additional convolutional layers to the top layer of the fea-ture extractor, multiple feature maps with different resolutions can be obtained, representing the target information to varying scales of the image, so that the model can effectively recognize multi-scale targets. The experimental results show that compared with the original YOLOv5 algorithm, the improved algo-rithm improves the recognition speed and accuracy, is easy to implement, and can meet the accuracy and real-time requirements of traffic sign recognition under complex road conditions.

CC2004

Outdoor Unmanned Vehicle System Design and Patrol Algorithm Research

Zhou Huang, Runyang Xiao, Ruijia Yao, Zhengwang Xu and Ziqun Du
Hubei University of Technology, China

Abstract-In modern unmanned vehicle technology, outdoor patrolling algorithms play a crucial role in enabling autonomous navigation. To enhance the performance of unmanned vehicles in complex outdoor environments, researchers have been continuously refining and optimizing these algorithms. This paper presents a detailed overview of our approach, including the design of the autonomous navigation camera algorithm, vision processing, and recognition technologies. To address challenges such as limited computational resources, environmental interference in line patrolling, and the prolonged debugging time required for the line patrolling algorithm, we propose innovative solutions. These include multi-threading, segmented adaptive binarization, the edge inheritance method, the least squares approach, and enhancements to the YOLOv5-Lite model. Together, these advancements significantly improve the vehicle's adaptability to diverse lane conditions.

CC2005

A Method For Detecting ST Segment Anomalies Using An Inverted Transformer With Incept Fusion Attention

Feiyan Zhou[1,2], **Yawen Wang[1,2]**, Yuhao Sun[1,2] and Huimin Zhang[1,2]
1. Guangxi Normal University, China
2. Key Lab of Education Blockchain and Intelligent Technology, Ministry of Education, China

Abstract-The presence of abnormal ST segments in electrocardiograms (ECGs) is associ-ated with a variety of prevalent medical conditions. However, current technolo-gies for automatic arrhythmia detection predominantly focus on QRS complexes, and there is a notable lack of research dedicated to identifying anomalous ST segments. To effectively and autonomously identify unusual cardiac activity based on ECGs, we propose a novel methodology for ST anomaly detection that combines a modified sequential transformer with convolutional techniques. This approach integrates an inverted transformer and channel prior attention into a uni-fied network, resulting in high detection performance with a reduced error rate while utilizing single-lead diagnostics. On our collected dataset, comprising over 180,000 authentic clinical records, the overall detection accuracy reached an im-pressive 96.14%, with an outstanding F1 score of 96.17%. For the PTB-XL da-tabase, we achieved an accuracy of 88.58% and an F1 score of 85.02%. These remarkable findings surpass the accuracy and generalizability of existing method-ologies, thereby equipping clinicians with an advanced tool for detecting abnor-malities in ST segments.

Session 4

March 29th, 2025
Time Zone: GMT+8

Topic: Image Analysis and Computational Model
Time: 16:40-17:55 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A714
Session Chair: Prof. Hui Liu, Kunming University of Science and Technology, China

Onsite

CC0004

DSS6D: Dual Stream Fusion Network for 6D Pose Regression of Symmetric Objects

Zerui Jiang[1], Guangjian Zhang[1] and Yawen Zheng[2]
1. ChongQing University of Technology, China
2. Open University of LuoYuan, China

Abstract-This paper proposes a novel framework for 6D pose estimation. In the case of symmetrical objects, objects may have multiple ground-truth poses,this one-to-many relationship may lead to estimation errors. To address this problem, we introduce a symmetry-invariant pose distance metric called the mean (maximum) grouped primitive distance or a (M)GPD. The A(M)GPD loss enables the model to converge to the correct pose so that all minima on the A(M)GPD loss can be mapped to the correct pose. We use a fully convolutional network to extract point-wise features from RGB-D data. By driving the connection between features through an attention mechanism, regress the 6D pose without post-refinement. Compared with other frameworks, our framework is more suitable for the 6D pose of symmetrical objects. We conduct experiments on the YCB-Video and TLESS datasets, and the results demonstrate that the model outperforms the state-of-the-art methods when dealing with symmetrical objects, demonstrating that the framework has high accuracy and low computational cost.

CC0012

RIBTRACK: Rib-Based Tracking for Multi-Planar Reconstruction

Huasheng Zhuo[1], **Yue Zhou**[1], Donglin Wen[1], Dongwen Luo[1], Yuran Gu[2], Gang Wu[1], Zhiyong Liu[1], Guanglan Liao[1] and Tielin Shi[1]
1. Huazhong University of Science and Technology, China
2. China Medical University, China

Abstract-Rib fractures are common injuries that pose significant risks, necessitating rapid and accurate diagnosis. Traditional chest CT radiographic diagnostic techniques are highly dependent on the manual reading of images. The process is labor-intensive, highly experience-dependent, and less accurate, prompting the need for advanced methods based on artificial intelligence. To solve this problem, this study demonstrates a novel automated rib multiplanar reconstruction method that integrates object detection and tracking algorithms, specifically YOLO-X and ByteTrack. A dataset of 200 patients from Tongji Medical College, including 88 with rib fractures and 112 normal, is utilized to train and validate the algorithm model. Superior performance with high precision (0.993), recall (0.995), and mAP (0.994) can be achieved by YOLOv10-ByteTrack integration. The tracking algorithm improves the success rate by over 30%, with a statistical error reduction of nearly 70%, greatly enhancing the

detection accuracy. This study represents a significant step forward in computerized tomography reconstruction, provides a reliable solution for rib fracture diagnosis, and addresses limitations of traditional MPR, contributing to both scientific research and clinical practice.

CC0019

MFRes-UNet: A ResUNet Fusing Multiscale Features for Accurate Spine CT Images Segmentation

Hanlin Zhao[1], Liyuan Zhang[1] and Xiongfeng Tang[2]
1. Changchun University of Science and Technology, China
2. Orthopedic Medical Center, Jilin University Second, China

Abstract-Accurate vertebrae segmentation in spine CT images is essential for diagnosing and treating spinal diseases. However, existing methods face challenges due to the extensive span of the spine, the complex shape of its curvature, and the high similarity between adjacent vertebrae in CT sagittal plane images. These factors often lead to limitations in capturing the detailed features and contextual information necessary for accurate segmentation. This paper presents a ResUNet model that incorporates both a multiscale feature extraction module and a feature refinement module. The multiscale feature extraction module utilizes multiscale and dilated convolution techniques to broaden the model's receptive field, thereby enhancing the extraction of global semantic information. Additionally, the feature refinement module employs a decoupling strategy with large convolution kernels and a spatial kernel selection mechanism to enable the model to capture richer, detailed features. The proposed model has been validated on the CTSpine1k dataset, and its performance was rigorously evaluated. The experimental results indicate that the model achieves an average Dice score of 88.89% and an average HD95 distance of 4.17 mm. Compared to other state-of-the-art models, the proposed method exhibits a significant advantage in accurately processing the details of vertebral edges.

CC0025-A

Esophageal Segmentation Based On Fusion Module Of CNN And Vit

Xiao Lou, Youzhe Zhu, Fengsong Ye and Yong Chen
Lishui People's Hospital, China

Abstract-Image segmentation of the esophagus itself and its radiotherapy target could effectively assist oncologists in the delineation of the target area. Previously, there was few works focusing on the segmentation of esophageal lesions such as gross tumor volume (GTV) and clinical tumor volume (CTV). This paper proposed a network——PCV-UNet, which based on the fusion module of CNN and ViT, to solve the esophageal segmentation problem of GTV and CTV. The CNN-ViT fusion module was used as the backbone of PCV-UNet. The advantage of PCV-UNet was to capture the local and global attention in parallel by CNN and ViT, respectively. The SAM Encoder was employed for transfer learning, which improved the training efficiency. The prompt learning module improved the accuracy of prediction and switched the neural network to an interactive mode. Feature selection and enhancement were carried out with enhanced high-frequency unit. Compared with relevant networks, PCV-UNet had a superior performance in the criteria of MSD, HD, and DC, which were MSD=6.79(3.85, 8.29), HD=13.46 ± 9.37, and DC=67.52 ± 16.55%, respectively.

CC5005

Effective IR-RGB Image Registration and Fusion Method for Railway Scene

Feng Zhang[1], **Xingfang Zhou**[2,4], Baiju Feng[3], Xinlong Xu[4] and Baoqing Guo[2,4]

1. Guoneng Shuohuang Railway Development Co., Ltd. Cangzhou, China
2. State Key Laboratory of Advanced Rail Autonomous Operation, Beijing Jiaotong University, China
3. School of Physics, Engineering and Technology, University of York, UK
4. School of Mechanical, Electronic and Control Engineering, Beijing Jiaotong University, China

Abstract-Railway perimeter protection is crucial for ensuring railway operation safety. Video surveillance technology is widely used in it, but visible light cameras often suffer from poor imaging quality under nighttime or adverse weather conditions, which affects the effectiveness of protection. The fusion of RGB and infrared (IR) images can improve image quality, thereby enhancing subsequent railway intrusion detection. To fully utilize the characteristics of railway scenes to achieve accurate image registration, a coarse-to-fine image registration method based on line features is designed. For the registered image pairs, to ensure the fused image contains both the texture and contour details of RGB image and the temperature information of IR image, an image fusion method based on Non-Subsampled Contourlet Transform (NSCT) is proposed, specifically, a new saliency measurement (NSM) to measure the multi information of image, including structure, sharpness, and brightness is for low-frequency sub-band fusion, fourth-order correlation coefficients (FOCC) matching is for high-frequency sub-band fusion. Experiments demonstrate that the proposed method performs better in several evaluation metrics compared with other four traditional algorithms, displaying richer information and high resolution.

Session 5

March 29th, 2025
Time Zone: GMT+8

Topic: Intelligent Detection Technology and Engineering Application
Time: 16:40-17:55 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A715
Session Chair: Assoc. Prof. Yuanlong Xie, Huazhong University of Science and Technology, China

Onsite

CC0016

ELS-YOLO: A Lightweight Algorithm for Strip Steel Surface Defect Detection

Hongwei Lu and Huihui Zhou
Hubei University of Automotive Technology, China

Abstract-To address the issues of defect leakage, inaccurate detection, and the high computational expense linked to the YOLOv10n model in the context of strip steel defect identification, we introduce an improved defect detection algorithm called ELS-YOLO. Firstly, we introduce the Efficient Multiscale Perception Network (EMSPN), a lightweight yet effective network that reduces computational complexity while improving detection precision. Secondly, we integrate the Large Separable Kernel Attention (LSKA) mechanism within the SPPF module, resulting in the creation of the SPPF-LSKA module that improves the identification of multiscale strip steel defects. Additionally, we design a lightweight detection head, the SEAM-Head, which effectively reduces the model's parameters and computational demands. Results from experiments conducted on the NEU-DET dataset demonstrate that the ELS-YOLO model attains a precision of 75.3%, while utilizing merely 2.03M parameters and incurring a computational expense of 5.6G, surpassing mainstream algorithms. Furthermore, tests on the DeepPCB dataset demonstrate that ELS-YOLO achieves comparable performance to YOLOv10n while consuming fewer resources and parameters, highlighting its efficiency and robustness.

CC0018

Deep Learning-Assisted Label-Free Glioma Cell detection with droplet Microfluidics

Chunhua He[1], **Huasheng Zhuo[1]**, Jianxing Wang[1], Zihan Li[2], Zhiyong Liu[1], Lei Nie[2], Guanglan Liao[1] and Tielin Shi[1]
1. Huazhong University of Science and Technology, China
2. Hubei University of Technology, China

Abstract-Glioma, a primary malignant tumor of the central nervous system, presents sig-nificant challenges in intraoperative identification due to its infiltrative growth pat-tern. Traditional diagnostic methods, such as frozen sections and imprint cytolo-gy, are hindered by complex sample preparation and subjective interpretation. This study introduces an innovative system integrating droplet microfluidics and deep learning technologies to address cell overlap and unstable observation envi-ronments that have historically hindered accurate cell identification. Our opti-mized YOLOv8 architecture incorporates strategically placed attention mecha-nisms at layers 8, 9, and 15, enhancing feature extraction capabilities for small object detection while maintaining computational efficiency. The system's robust multi-focal analysis capability enables accurate distinction between glioma cells and normal

cells across different focal planes, surpassing current imaging-based diagnostic methods. Through validation against conventional fluorescence detec-tion methods, this work demonstrate superior stability in cell counting without threshold dependence and real-time processing with minimal latency. This inte-grated approach advances intraoperative diagnostics with a practical solution that can improve precision in glioma surgery and patient outcomes.

CC0021

Visual Storage Battery Inspection via Large Vision-language Model

Cheng Wang, Can Zhang, **Bin Hu**, Bao Tang, Shikang Wang and Xinggang Wang
Huazhong University of Science and Technology, China

Abstract-Storage batteries are widely used across various industries, and their failure can lead to severe consequences. With the advancement of artificial intelligence, particularly computer vision, visual data of storage batteries can be captured through monitoring cameras and processed using vision-centric convolutional neural networks (CNNs) or transformers for storage battery inspection. However, traditional vision-centric methods often rely on large amounts of annotated data, which is a significant limitation in the field of storage battery inspection. Additionally, storage batteries typically operate in complex environments with severe noise, leading to lower recognition accuracy in conventional approaches. To address these challenges, a specialized dataset tailored for storage battery inspection is created. We then introduce VLMSBI, a novel vision-language model-based method, to perform storage battery inspection. Our approach demonstrates superior performance compared to traditional vision-centric CNN- and transformer-based methods, while utilizing only minimal annotated data for supervised fine-tuning (SFT). Notably, VLMSBI's versatile architecture allows for seamless adaptation to other industrial component inspection tasks by simply modifying the fine-tuning data and corresponding input prompts.

CC0063

Top-Forging Crack Detection for Cold-Heading Steel

Long Wu, Yu-Xiu Wu and Hongwei Yang
Anhui University of Technology, China

Abstract-Cold heading steel top forging cracks are challenging to capture due to their complex shapes and varying structural dimensions. Additionally, differences in understanding standards and reference images can lead to inconsistent judgment results in manual inspection. This paper introduces a detection and classification algorithm for cold-heading steel top forging cracks based on an enhanced YOLOv8n model, incorporating multiple attention mechanisms to address these challenges. The proposed algorithm first integrates receptive field attention convolution (RFAConv) and CBAM (concentration-based attention module) into the backbone network to improve the convolutional modules. Moreover, DAttention is added to the backbone network for dynamically allocating attention weights and speeding up model training. In the neck portion of the network, C2f and Conv modules are replaced with improved C2f_RFCBAM and RFCBAMConv modules to optimize the model and further increase detection precision. Finally, a new loss function, SioU, is introduced to accelerate convergence and improve regression accuracy. Experimental results show that the improved YOLOv8n model applied to a self-annotated cold heading steel dataset, achieves increases of 7.3% in mAP@0.5, 2.3% in mAP@0.5:0.95, and 10.3% in recall rate.

CC0064

Surface Defect Detection Model for Quick Connectors

Lin Zhou, Wang Chen, Shuai Yang and Peng Huang
Hubei University of Automotive Industry, China

Abstract-In response to issues such as missed detection of small defects, low detection efficiency for certain defects, and insufficient overall automation in the automated detection of Quick-Connect Fittings (QCF), this paper proposes an intelligent detection model based on the collaborative use of supervised and unsupervised learning. First, a supervised training approach is conducted using an improved YOLOv8 model (QCF-YOLO), which optimizes the feature map representation capability, significantly improving the detection accuracy of small defects. This model has the advantages of fewer parameters and easy deployment, meeting real-time detection requirements. Secondly, unsupervised secondary training is performed in the cloud using detection results and raw images collected from edge devices. By comparing the detection results of supervised and unsupervised models, the model's ability to detect unknown defect categories is progressively enhanced, effectively avoiding false detections caused by background interference in unsupervised training. Finally, based on a "cloud-edge-end collaborative architecture, " the model is deployed at the edge. Through the collaborative work of cloud training, edge detection, and terminal data collection, the automation level and detection efficiency of Quick-Connect Fitting defect detection are significantly improved.

Session 6

March 29th, 2025
Time Zone: GMT+8

Topic: Vision-Based Signal Analysis and Intelligent Control Technology
Time: 16:40-17:55 (Duration for Each Presentation: 15 minutes)
Room: Teaching Building 7-A717
Session Chair: Assoc. Prof. Yang Zhang, Hubei University of Technology, China

Onsite

CC0007

Design Method of Aircraft Control Software Architecture Based on Hypergraph and Hypernetwork

Xianlong Ma, Xinpeng Xu and **Qian Zhang**
Northwestern Polytechnical University, China

Abstract-Software-defined Aircraft (SDA) is an aerospace innovation that draws on the concept of Software-defined Networking (SDN) to improve the functionality, flexibility and adaptability of aircraft through software and programmable technologies. With the continuous progress of aircraft technology, software-defined aircraft (SDA), as a new aircraft architecture design method, has gradually become a research hotspot in the aviation field. The current research focus is mainly on the design of aircraft architecture by using the software-defined idea, but there is a lack of systematic theoretical guidance. Aiming at the weak basic theory in the research of software-defined aircraft, this paper proposes the software architecture design theory based on hypergraph and hypernetwork on the basis of analyzing the design requirements of software-defined aircraft model architecture, designs a new representation method of aircraft architecture model, and analyzes the support that this theory can provide to the software architecture design and optimization of aircraft. Finally, the feasibility of the theory is verified by simulation, and the accuracy and reliability of the model are evaluated by comparing the model prediction results with the actual aircraft performance.

CC0008

CLAVS: A Novel Continual Learning Model and Benchmark for Audio-visual Saliency Prediction

Jiaqi Wang[1], Nana Zhang[1], Ming Zhu[1], Dandan Zhu[2] and Kun Zhu[3]
1. Donghua University, China
2. East China Normal University, China
3. Tongji University, China

Abstract-Modeling and predicting human attention in audio-visual circumstance is crucial to understanding and simulating human visual attention mechanism in real multi-modal environment. In practical scenarios, data collected from various scenes often arrives sequentially, resulting in continuous shifts in the data distribution. However, the majority of existing audio-visual saliency approaches are developed based on the assumption of complete data availability and a stationary data distribution. Consequently, this assumption makes them inevitably suffer from the issue of catastrophic forgetting. To address the aforementioned issue, we are the first to propose a novel continual learning-based audio-visual saliency (CLAVS) model. This model is designed to preserve previously acquired knowledge through the compression and replaying of samples from earlier tasks, effectively

mitigating the issue of model forgetting. Specifically, we first introduce a new multi-modal world model to capture task-specific audio-visual dynamic information in a Gaussian mixture latent space. Second, we devise an experience replay method to generate the old knowledge, which can effectively address the issue of catastrophic forgetting. Furthermore, we construct two new task-based benchmarks according to video content and dataset membership. Extensive experimental results on two benchmarks demonstrate the superiority of the proposed CLAVS model and outperform the naive combinations of previous state-of-the-art in saliency prediction and continual learning.

CC0031

Optimization Of AOA Localization Station Distribution Based On Improved Black-Winged Kite Algorithm

Wen Ke and Xiao Zhou
Wuhan University of Technology, China

Abstract-Aiming at the problem of station distribution for AOA cooperative localization, an improved black-winged kite optimization algorithm is proposed, which integrates a random differential mutation strategy to generate new individuals and increase population diversity, making the algorithm less likely to fall into local optimum. Combined with the local search ability of the simplex method, the optimization results are improved. Through comparison with traditional algorithms and traditional station placement methods, the superiority of the improved algorithm in optimization ability and the feasibility of optimized station placement are verified. In the case of positioning a target in a designated area, the algorithm can be used to rationally arrange the positions of observation stations and improve the positioning accuracy of the target.

CC5006

Powerline Inspection using Event-Triggered Multi-UAV Formation Control

Tua A. Tamba[1] and Benedictus C.G. Cinun[1]
1. Parahyangan Catholic University, Indonesia

Abstract-This paper proposes a multi-agent system approach to robust formation control scheme of multi-UAV for powerline inspection applications. A backstepping multi-agent sliding mode control (BMA-SMC) method that is equipped with radial basis function neural network estimator of wind disturbances is developed for the multi-UAV system to ensure precise and robust tracking of the powerline cables and towers. An event-triggered control execution scheme is also proposed to ensure efficient utilization of each UAV's computational and communication resources when implementing the developed formation control scheme. The simulation results of multi-UAV formation when tracking conventional powerline cable network configurations are reported to illustrate the effectiveness of the proposed robust multi-UAV control scheme.

CC0078

Study on Risk Assessment Methods of Multi-Hazard in Coastal Urban Areas

Wenjie Zhao[1], Xiangang Luo[1], Xiaokang Tu[1], Yan Zhang[1], Bo Liu[2], Jingmin Tu[3] and BoYu Zhao[1]
1. China University of Geosciences, China
2. Henan Institute of Science and Technology, China
3. Hubei University of Technology, China

Abstract-To achieve an effective assessment of the multi-hazard chain risks in coastal cities, this study selects Sanya, a typical island city in China affected by the typhoon-geological disaster-flood disaster chain, as the research area. Based on the intrinsic characteristics and mechanisms of urban multi-hazard occurrences, and integrating the causal logic of mathematical models and machine learning algorithms, this paper constructs a set of hazard assessment index systems suitable for typhoons, geological disasters, and floods, utilizing multi-source data analysis. The typhoon hazard is assessed using probabilistic statistics and a combination of subjective and objective weighting analysis, while the hazards of geological disasters and floods are evaluated using the Random Forest(RF) model. Considering the interrelationships among multiple hazards, a multi-hazard risk matrix is employed to build a comprehensive hazard assessment model, which measures the likelihood of disaster occurrences in coastal urban areas, classifies hazard levels, and analyzes the assessment results. The findings indicate: 1) The typhoons passing through Sanya predominantly move in a north-south direction, with the Tianya District and Jiyang District being the most severely affected; 2) The RF-based models for assessing geological and flood hazards achieved AUC values of 0.904 and 0.898, respectively. High-risk areas for geological disasters are mainly concentrated in the central part of Sanya, whereas flood-prone areas are primarily located along the southern coast and at river confluences in the main urban area; 3) The spatial distribution characteristics of comprehensive hazard risks are notably evident, with high- risk zones predominantly located in the central-southern parts of Tianya and Jiyang Districts, and the southeastern coastal area of Haitang District, closely aligning with the typhoon hazard zoning results, thereby illustrating the correlation between typhoons and geological as well as flood disasters.based models for assessing geological and flood hazards achieved AUC values of 0.904 and 0.898, respectively. High-risk areas for geological disasters are mainly concentrated in the central part of Sanya, whereas flood-prone areas are primarily located along the southern coast and at river confluences in the main urban area; 3) The spatial distribution characteristics of comprehensive hazard risks are notably evident, with high-risk zones predominantly located in the central-southern parts of Tianya and Jiyang Districts, and the southeastern coastal area of Haitang District, closely aligning with the typhoon hazard zoning results, thereby illustrating the correlation between typhoons and geological as well as flood disasters.

Session 7

March 30th, 2025
Time Zone: GMT+8

Topic: Intelligent Image Detection Model and Application
Time: 13:00-15:00 (Duration for Each Presentation: 15 minutes)
Zoom Link: https://us02web.zoom.us/j/83539300002
Session Chair:

Online

CC0011

Coarse-to-fine Automatic Segmentation and Labeling for Urban MLS Point Clouds

Hongjiang Ye[1], Shaohui Mai[1], Mingchun Wang[1], Min Gao[2] and **Yuanyuan Fei[2]**
1. Guangdong Power Grid Co., Ltd. Guangzhou Nansha Power Supply Bureau, China
2. Digital Life Research Center, Guangzhou HKUST Fok Ying Tung Research Institute, China

Abstract-In this paper, we propose a novel method for automatic segmentation and labeling of urban MLS point clouds. Considering the scarcity of labeled data for semantic segmentation models and the impracticality of manual labeling for large-scale point clouds, an automatic labeling process is in great need. Our method commences by extracting ground and walls through RANSAC. Subsequently, a coarse-to-fine segmentation strategy is implemented, followed by feature analysis and label matching. In experiments, we trained a KPFCNN semantic segmentation model on Toroto-3d dataset using the proposed method. The results successfully affirm this method's effectiveness, with a notably high accuracy for key classifications.

CC0014

Wavelet-based Frequency Decoupling Refinement Network for Remote Sensing Image Super Resolution

Jinyang Yu and Yong Dou
National University of Defense Technology, China

Abstract-Deep learning methods have demonstrated remarkable capabilities in image super-resolution. However, remote sensing images present unique challenges due to the heterogeneous content, multi-scale characteristics, and abundance of high-frequency details. In light of recent advancements, we rethink the architecture design for super-resolution of remote sensing images, introducing a wavelet-based frequency decoupling refinement network (WFDRN) to reconstruct lost information from low resolution images. Specifically, we leverage discrete wavelet transform (DWT) to decouple the deep features into low-frequency (LF) and high-frequency (HF) component, which are subsequently processed by specialized network architectures incorporating self-attention mechanisms and vanilla convolutions, respectively. Furthermore, we cascade multi-level wavelet units to expand the receptive field and capture multi-scale contextual information, activating more pixels to contribute to reconstruction. Quantitative and qualitative experiment results on two remote sensing datasets demonstrate that our method effectively restores details, and surpasses traditional single image super-resolution methods.

CC0015

Text-Guided Image Generation Method Based on Cross-Modal Fashion Style Learning

Wanhui Zhang, Li Liu, Xiaodong Fu, Lijun Liu and Wei Peng
Kunming University of Science and Technology, China

Abstract-In the interdisciplinary field spanning computer vision and art, text-guided image generation techniques have garnered significant attention, particularly in the realms of virtual try-ons and fashion design. Current approaches to text-guided image generation primarily utilize latent space encoding and alignment of text-image feature spaces, which prove effective in general scenarios. However, when tasked with generating images possessing specific fashion attributes, these methods often result in semantic inconsistencies and missing details in the generated clothing images. To address this challenge, this work presents an effective text-guided image generation method based on cross-modal fashion style learning. First, the fashion text encoding module is constructed to represent the input text description as global-local text descriptors to obtain fashion features. Then, a clothing image style representation module is designed to extract coarse-grained, fine-grained, and hierarchical latent features from the input texture patterns and clothing images. Finally, a cross-modal fashion style learning block, integrating the fashion mapping network and the generative adversarial network, is utilized to reinforce semantic consistency and yield clothing images with intricate details. Extensive experiments conducted on the Viton-HD dataset validate that the proposed method significantly outperforms other state-of-the-art approaches, achieving an initial IS score of 5.22 and a Fréchet Inception Distance (FID) of 16.35. Ablation studies and comparative analysis further demonstrate the potential application of our method, specifically in improving the semantic consistency between fashion texts and generated images, as well as enhancing fashion details in the generated images.

CC0022

Improving Small Object Detection with Convolutional Attention Module

An Wen[1], Kai Yu[2], Lu Wan[1], Yizhou Pan[1], Nengji Li[1], **Jiapeng Zhang[2]** and Qiang Wang[2]
1. Zhejiang Huayun Information Technology Co., Ltd., China
2. Zhejiang University, China

Abstract-Addressing the issue of poor detection performance of small targets by the YOLO algorithm, and also to further enhance the detection accuracy of YOLO, this paper first substitutes the backbone network with a composite backbone network CB-YOLO. On this basis, a CBA-YOLO network model with a convolutional attention block module is designed. By adding convolutional attention modules at specific locations in the neck network of CB-YOLO, the model assigns more attention weights to small targets during the feature fusion process in the neck network, enhancing the feature expression ability of small targets and making the model more focused on the detection of small targets, thereby improving the model's detection accuracy for small targets. The paper conducts experiments on two publicly available datasets of different scales, verifying that the CBA-YOLO model enhances the detection accuracy of small targets without sacrificing the detection accuracy of medium and large targets.

CC0026

Infrared Image Detection Algorithm Of Electrical Equipment Based On Improved Yolov8

Dongfeng Liu and Xiaoyang Ju
Taiyuan University of Science and Technology, China

Abstract-There are difficulties in the infrared image detection of substation equipment, such as large equipment size span, diverse types, and low detection accuracy of small-sized equipment. So A detection algorithm based on the YOLOv8 model, ET-YOLO (ELA-HSFPN-TADDH), is proposed to address those challenging problems. ELA-HSFPN, the Enhanced Lightweight Attention High Selective Feature Pyramid, is adopted to improve the model's feature fusion ability, fuse high-level and low-level feature maps, and obtain rich detailed information. Then the Lightweight Task Alignment Dynamic Detection Head (TADDH) is employed to combine scale, space, and task attention mechanisms to reduce model size and improve detection efficiency. For the ET-YOLO algorithm, the evaluating indicator mAP@0.5 reaches 90.9% and is better than other tested models. The parameter and computational complexity of ET-YOLO were reduced by 31.25% and 17.43% respectively compared to the YOLOv8 model. The experimental results show that the ET-YOLO algorithm can meet the requirements of accurate real-time identification of substation equipment and lay the foundation for subsequent substation equipment fault diagnosis.

CC0061

Multi-modal Dynamic Fusion Network for Intent Understanding in Assisted Older Adults

Zhaoyu Li, Likai Dong, Tao Xu, Jiahui Sun and Guangze Zhu
University of Jinan, China

Abstract-Intention understanding plays a crucial role in human-robot interaction. Most methods use multimodal fusion to obtain robust and efficient results. However, most methods rely on predefined input modalities and employ separate networks for each, which leads to increased interaction latency. We propose a network named Dynamic Multimodal Path Selection (DMPS), which consists of a gating module that recognizes input data and a multimodal fusion module. The gating module dynamically selects the appropriate fusion path and integrates multiple algorithms into a single network. It also distinguishes whether the input gesture is static or dynamic. The multimodal fusion module employs a Transformer and CNN-based feature extractor to enhance the accuracy of intent recognition. To verify the effectiveness of our method, we created the DMPS Multimodal Dataset and built an experimental environment designed to assist older adults. The experimental results demonstrate that the proposed DMPS network outperforms existing methods, achieving an accuracy of 94.00% in the fetch operation.

CC0073

MMMT-Net: Multi-Modal Multi-Temporal Network for Liver Tumor Segmentation from CT Images

Ri Jin[1], Hu-Ying Tang[2], Qian Yang[3] and Wei Chen[2]
1. University of Electronic Science and Technology of China, China
2. The First Affiliated Hospital of the Army Medical University (Southwest Hospital), China
3. Sichuan Provincial People's Hospital, Chengdu, China

Abstract-Accurate liver tumor segmentation is critical for the diagnosis and treatment of liver cancer, yet remains challenging due to image quality and tumor heterogeneity. Existing methods predominantly focus on single-phase and single-time-point CT scans, neglecting complementary multi-phase information and longitudinal correlations. In this paper, we propose the first multi-modal multi-temporal network, which integrates diverse information from arterial phase and longitudinal data to facilitate liver tumor segmentation of venous phase. Our approach mainly comprises three novel blocks: (1) A polynomial residual block that captures abundant tumor characteristics via nonlinear hierarchical feature learning; (2) A longitudinal phase attention block that refines input features from

channel and spatial dimensions, and recalibrates the target feature based on inter-phase and inter-temporal dependencies; (3) A progressive up-sampling block that preserves structural details during resolution restoration, and forms an effective decoder architecture. Experiments on a large-scale CT dataset of longitudinal multi-phase liver tumors demonstrate that the proposed network achieves promising liver tumor segmentation and outperforms state-of-the-art methods. In addition, an ablation study validates the functions of the proposed blocks.

CC2011

YOLOv11-SV: Small Target Detection in Visible Light via Multi-Scale Feature Extraction and Frequency domain enhancement

Shize Huang[1], Yifeng Suo[1], Zehao Shen[1], Tang Chen[2] and Ping Zhang[1]
1. University of Electronic Science and Technology of China, China
2. Wenzhou-Kean University, China

Abstract-Small target detection in visible light plays a crucial role in applications such as traffic monitoring, security surveillance, and autonomous driving. However, traditional YOLO models often struggle with feature loss, high false detection rates, and unstable detection performance in low-light, complex backgrounds, and low-contrast environments. To address these challenges, this paper proposes an improved model, YOLOv11-SV, which enhances small target detection accuracy and stability by optimizing multi-scale feature extraction and high-frequency information retention strategies. Specifically, our model designs two key modules. The first module replaces the feature extraction component in the YOLOv11 backbone with a wavelet transform-based structure, which Enhance high-frequency information. The second module proposes an advanced feature aggregation framework to perform multi-scale pooling, enhancing feature representation across different scales. Experimental results demonstrate that YOLOv11-SV outperforms other YOLO models in key metrics such as mAP, precision, and inference speed, particularly excelling in low-contrast environments.

Session 8

March 30th, 2025
Time Zone: GMT+8

Topic: Image-Based Intelligent Measurement System and Anomaly Detection
Time: 15:40-18:10 (Duration for Each Presentation: 15 minutes)
Zoom Link: https://us02web.zoom.us/j/83539300002
Session Chair:

Online

CC0009

Implicit Fine-Grained Alignment For Cross-Modal Retrieval

Hui Liu[1,2], Xiao-Ping Chen[1], Rui Hong[1,2], Yan Zhou[1,2], Tian-Cai Wan[2] and Tai-li Bai[1]
1. University of Electronic Science and Technology of China, China
2. Avic Chengdu Aircraft Design and Research Institute, China

Abstract-Cross-modal retrieval has always been one of the essential tasks in multimodal learning, with the aim of searching for the most relevant cross-modal samples based on a given query. Taking image-text retrieval as an example, a direct and efficient approach is to use visual semantic embedding technology to bring positive samples closer and push away negative samples in the shared subspace. However, these methods often overlook the potential implicit alignment relationships in text. In this letter, we enhance representation by revisiting potential alignment relationships in text and propose a new method called IACR that exploits an Implicit Alignment Module (IAM), thereby capturing the fine-grained local associations. Specifically, IAM captures implicit alignment through a multi-head attention mechanism to establish fine-grained alignment between image representations and partial words in the text. This approach avoids the additional reasoning costs caused by fine-grained modeling and maintains local correlation to improve performance. To verify the superiority of this cross-modal alignment paradigm, we conduct experiments on two large-scale image-text retrieval datasets, i.e., Flickr30K and MS-COCO, to evaluate the proposed IACR and provide analysis to verify its effectiveness.

CC0044

Detection Of Key Components Of Transmission Lines Defect Identification Studies

Jianchao Wang, **Zijing Chen**, Shuhai Wang and Shaojie Li
Hebei University of Science and Technology, China

Abstract-Due to the fact that the transmission line is in the outdoor environment for a long time, it is easy to be affected by bad weather, environmental pollution and other factors, resulting in defects in line equipment, which may lead to accidents.This section focuses on the identification of defects in key components of transmission lines[1].The transmission line targets were extracted, and the SSD algorithm and the YOLO target detection algorithm were combined with the strategy, and the defects of insulators, shockproof hammers and spacers were identified into the improved YOLOv5 algorithm network, and the detection model was evaluated through the evaluation index. Compared with the mainstream algorithms, and ablation experiments were carried out, it was found that we propose an improved YOLOv5 algorithm. By incorporating K-means++ adaptive anchor box optimization, we enhance the model's adaptability to target size variations, resulting in a 3.7% mAP improvement.

We design the Gating Layer Block to better capture contextual information of small target defects, enhancing the network's receptive field and feature extraction capability, leading to a 1.9% mAP improvement. We introduce the CSAtt attention mechanism module to broaden the model's focus range. We also design the Spatially Adaptive Convolution with variable dilation convolution to adapt to objects of different scales, resulting in a 9.1% precision improvement. Finally, we improve the polynomial loss function to enhance model performance and save computational resources and time.

CC0071

LaMHA: Efficient Multimodal Malicious Meme Classification via LoRA-Tuned Adaptation and Attentive-MLP Fusion

Meng Qi and Chung-Lun Wei
City University of Macau, Macau, China

Abstract-To address insufficient cross-modal semantic alignment and limited dynamic discrimination in multi-modal hate content detection, this paper proposes a lightweight architecture integrating Low-Rank Adaptation (LoRA) and Attention-gated Multilayer Perceptron (aMLP). The architecture achieves efficient and accurate hate detection through a two-stage cascade: 1) The LoRA module embedded in CLIP's text encoder performs lightweight text feature projection via low-rank matrix decomposition, enabling efficient cross-modal adaptation with only 0.3% parameters fine-tuned; 2) The aMLP module at the classification layer dynamically fuses multi-modal features using cross-modal attention weights and suppresses noise via a gating mechanism. Experiments on HMC and HarMeme datasets demonstrate superior performance in Acc (77.74%/83.72%), F1 (69.79%/79.82%), and AUROC (84.01%/92.16%), with a notable 4.33% AUROC gain on cultural metaphors. Ablation studies confirm nonlinear synergistic effects (11.53% F1 improvement on HMC). This work provides an efficiency-precision balanced solution for multi-modal content safety.

CC0062

Winter Jujube Grading Method Based On Two-Stream Structure And Weakly Supervised Regional Partitioning

Xinrang Tian[1], Tianrui Wang[1], Zian Liu[2], Rufeng Zhang[3], Peiqi Miao[1] and Shixin Cen[1]
1. Tianjin University of Traditional Chinese Medicine, China
2. Hebei University of Technology, China
3. Hebei University of Water Resources and Electric Engineering, China

Abstract-Winter jujube is a nutritious fruit that often experiences bruises, wrinkles, and maturity issues during harvesting. Therefore, its precise classification is crucial for optimizing economic value and ensuring quality control. In recent years, although CNN and Transformer models have shown great potential in fruit classification in the field of deep learning, they face challenges in handling complex image relationships and dealing with high computational demands. This paper proposes a novel classification method for Winter jujube that synergistically combines the strengths of CNN and Transformer models and incorporates a weakly supervised region partitioning for enhanced feature extraction. This method stacks the CNN backbone and Transformer, where CNN shows strength of capturing features and Transformer exhibits global relationship in each feature. This approach specifically addresses the challenges of distinguishing subtle variations within Winter jujube varieties. Demonstrated through rigorous testing on Winter jujube datasets, this method achieved a notable accuracy of 93.09%, surpassing several established baseline models. Overall, the classification strategy developed in this study

significantly enhances the precision and efficiency of Winter jujube classification, marking a substantial advancement for quality control in agricultural practices.

CC0051

Research on the Contact Arc of High-speed Pantograph with Active Optimization Control Strategy

Zhixin Ou
Anhui Communications Vocational & Technical College, China

Abstract-The way for high-speed trains to obtain stable current is through the contact between the pantograph and the overhead contact system, as the train speed increases, the vibration frequency between the pantograph and catenary increases, and the frictional arc will cause the quality of the current to deteriorate. The contact wire is a guide wire fixed above the steel rail, which is not easily subjected to complex design and modification, mainly based on the active control of the pantograph to adjust the current collection effect. A method based on active optimization control is proposed in the article to solve the phenomena of pantograph fluctuation and contact friction arc. Firstly, establish a system model for the pantograph and make structural model decisions; Secondly, combine parameter optimization algorithms with active control pantograph models for strategy training, such as vibration waveform, wear amplitude, arc frequency, etc. Finally, simulation experiments have shown that the active optimization algorithm and control strategy based on the pantograph model have good control effects, the proportion of decrease in error rate, volatility, and arc current detection is significant.

CC0049

Research on Real Verification Technology of Relay Protection Cabinet Diagram Based on Deep Learning

Gang Chen, Mengxuan Yan and Zhoubin Yu
Jiangmen Power Supply Bureau of Guangdong Power Grid Co., Ltd., Jiangmen 529000, China

Abstract-Aiming at the problems of low efficiency and insufficient accuracy of existing relay protection cabinet drawing verification methods, this paper proposes a multi-scale convolutional neural network (MSCNN-SAM) method based on wavelet convolution kernel and semantic attention mechanism. The method uses the multi-scale feature of wavelet transform to construct a convolutional kernel, effectively extracts multi-scale features in the drawing image, and combines with the semantic attention mechanism to focus on the key semantic information regions in the drawings, improving the ability to capture detailed features and important parts. Compared with traditional convolutional neural networks, MSCNN-SAM achieves higher accuracy and recall in the drawing verification task. The experimental results show that the method can significantly improve the efficiency and accuracy of relay protection cabinet drawing verification, which provides strong support for ensuring the safe operation of power systems.

Notes