

2022 14th International Conference on Graphics and Image Processing (ICGIP 2022)

WORKSHOP

2022 4th International Conference on Virtual Reality and Intelligent System (VRIS 2022)
【Nanjing, China | October 21-23| Virtual Conference】

Published by

SPIE. DIGITAL LIBRARY

Sponsored by



南京理工大学

NANJING UNIVERSITY OF SCIENCE & TECHNOLOGY

Organized by



计算机科学与工程学院
School of Computer Science and Engineering



江苏省计算机学会
JIANGSU COMPUTER SOCIETY



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

TABLE OF CONTENTS

01

WELCOME MESSAGE

02

USEFUL INFORMATION

03

DAILY SCHEDULE

04

KEYNOTE SPEAKERS

05

INVITED SPEAKERS

06

PARALLEL SESSIONS

WELCOME MESSAGE

2022 14th International Conference on Graphics and Image Processing (ICGIP 2022) in Nanjing, China during October 21-23, 2022. It is sponsored by Nanjing University of Science and Technology, China; Organized by School of Computer Science and Engineering, Nanjing University of Science and Technology, China; Jiangsu Computer Society; Jiangsu Association of Artificial Intelligence;

ICGIP is the main annual Graphics and Image Processing conference aimed at presenting current research being carried out.

ICGIP 2022 conference aims at serving as a forum for researchers and scientists from areas of visual computing, image processing and analysis, computer graphics and artificial intelligence, etc., to present, discuss, and exchange ideas on enabling technologies, system designs, applications and practice experiences. Benefiting from this forum, the attendees communicate with each other and reach a better understanding of different approaches as well as of their similarities. This conference provides opportunities for the delegates to exchange new ideas and application experiences face to face, to establish potential collaboration between researchers and to find global partners for future collaboration.

We are very pleased to announce this year's Keynote speeches will be given by

- Prof. Xu-Cheng Yin, University of Science and Technology Beijing, China
- Prof. Xiaochun Cao, University of Chinese Academy of Sciences, China
- Prof. Ir Dr Yiu-Ming Cheung, Hong Kong Baptist University, China

As coronavirus spreads worldwide, considering the safety of everyone, the organizing committee decided to convert ICGIP 2022 into a full online conference. During August 18-20, online sessions are organized by domain experts within the technical scopes of the conference.

All of the papers were subjected to peer-review by conference committee members and international reviewers. We feel deeply grateful to all that have contributed to make this event possible: authors who contributed papers, the conference steering committee, the conference speakers, and the peer reviewers. Thanks are also extended to the conference administrative committee and the supporters for their tireless efforts throughout the course of the conference.

Once again, on behalf of the conference committee, we welcome you all to Kunming! We wish you a very pleasant stay in this beautiful city. And hope you will find the ICGIP 2022 experience a memorable one.

Conference Committee

USEFUL INFORMATION

Online Presentation Instruction

- Please unmute audio and start video while your presentation.
- It's suggested to use headset with microphone or earphone with microphone
- Please prepare the slide file of your presentation on your laptop in advance.
- Duration of each Presentation: about 12 Minutes of Presentation and 3 Minutes of Q&A.
- E-certificate will be sent to presenters after conference by email. An excellent presentation will be selected from each session and announced on the website after conference. An excellent presentation certificate will be sent after conference by email.
- The conference secretary will capture a group photo at the end of each session, and you can download the picture after the conference on website after conference.

Please rename yourself before entering the conference room as below (in English):

Author: Paper ID-Name; Listener: Listener-Name; Keynote Speaker: Keynote-Name; Conference Committee: Position-Name

Zoom Meeting Instruction

- **Zoom Meeting Room Password: 102123**
- Conference rooms will be open 30 mins before scheduled time.
- Please enter the room 10-15 minutes in advance and be prepared.

ZOOM Download: (oversea authors): URL: <https://zoom.us/>

1. Room A: <https://us02web.zoom.us/j/84697177419>
2. Room B: <https://us02web.zoom.us/j/87279824425>
3. Room C: <https://us02web.zoom.us/j/83168271473>



Follow us on We-chat

Wechat Account (remark "ICGIP2022+Paper ID")



Room A: https://us02web.zoom.us/j/84697177419 Password: 102123	
10:00-16:00	Test time for Keynote Speakers, Invited Speakers and Committee Members
11:00-12:00	Test time for Session Chairs
14:00-14:50	Session 1: P1050, P1370, P1388, P1389, P1391, P1393, P1395, P1402, P1425 Session 2: G408, P0032, P1055, P1062, P1064, P1372, P1373, P1380, P1401 Session 3: P0013, P0025, P0027, P1054, P1060, P1385, P1386, P1400, P1903 Session 4: P0006, P0014, P0020, P0039, P1049, P1374, P1376, P1413, P1907 Session 5: P0011, G5001, P0012, P0026, P0034, P1046, P1383, P1411, P1901
15:00-15:50	Session 6: P0029, P1045, P1048, P1057, P1066, P1371, P1405, P1410, P1427 Session 7: P0016, P0028, P1042, P1053, P1375, P1390, P1416, P1426, P1429 Session 8: P0001, P0024, P1043, P1058, P1384, P1394, P1404, P1420, P1421 Session 9: P0010, P1428, P1378, P0002, P1061, P1059, P1056, P0040, P0019 Session 10: G402, G404, G410, P0038, P1052, P1063, P1369, P1430, P1908
16:00-16:50	Session 11: G405, G411, P0021, P1065, P1397, P1408, P1424, P1431, P1533 Session 12: P0007, P0008, P0033, P0035, P1387, P1396, P1398, P1409, P1415 Session 13: G406, P0017, P0018, P0022, P0036, P0041, P1047, P1381, P1392 Session 14: P0003, P1044, P1382, P1406, P1412, P1414, P1532, P2017

DAY 2- CONFERENCE AGENDA

October 22, 2022, Beijing Time

Room A: https://us02web.zoom.us/j/84697177419 Password: 102123	
Host: Prof. Gui Guan, Nanjing University of Posts and Telecommunications, China	
09:00-09:05	Welcome Regards Professor Lu Jianfeng, Vice Leader, School of Computer Science and Engineering, NJUST
09:05-09:10	Opening Remarks Professor Xiao Liang, School of Computer Science and Engineering, NJUST
09:10-09:50	Keynote Speaker I Professor Xu-Cheng Yin, University of Science and Technology Beijing, China <i>Speech Title: Open-Set Text Recognition: Definition, Framework and Algorithm</i>
09:50-10:30	Keynote Speaker II Professor Xiaochun Cao, University of Chinese Academy of Sciences, China <i>Speech Title: "ILL-Posed" Computer Vision Tasks</i>
10:30-10:40	Group Photo & Break Time
10:40-11:20	Keynote Speaker III Professor Ir Dr Yiu-Ming Cheung, Hong Kong Baptist University
11:20-11:40	Invited Speaker 1 Professor Xiaorong Xue, Liaoning University of Technology <i>Speech Title: Remote sensing image fusion based on MobileViT and spatial detail reconstruction</i>
11:40-12:00	Invited Speaker 2 Associate Professor Jungang Yang, National University of Defense Technology, China <i>Speech Title: Camera Array Computational Imaging Technology and Application</i>
12:00-13:30	Break Time

DAY 2- CONFERENCE AGENDA

October 22, 2022, Beijing Time

Password 102123	Room A https://us02web.zoom.us/j/84697177419	Room B https://us02web.zoom.us/j/87279824425	Room C https://us02web.zoom.us/j/83168271473
13:30-15:45	Session 1: Pattern Recognition P1050, P1370, P1388, P1389, P1391, P1393, P1395, P1402, P1425	Session 2: Image Segmentation G408, P0032, P1055, P1062, P1064, P1372, P1373, P1380, P1401	Session 3: Target Detection and Defect Detection P0013, P0025, P0027, P1054, P1060, P1385, P1386, P1400, P1903
15:45-16:00	Break Time		
16:00-18:15	Session 4: Image Analysis and Calculation P0006, P0014, P0020, P0039, P1049, P1374, P1376, P1413, P1907	Session 5: Image Classification P0011, G5001, P0012, P0026, P0034, P1046, P1383, P1411, P1901	Session 6: Image Enhancement and Denoising P0029, P1045, P1048, P1057, P1066, P1371, P1405, P1410, P1427

DAY 3- CONFERENCE AGENDA

October 23, 2022, Beijing Time

Room A: https://us02web.zoom.us/j/84697177419 Password: 102123	
Section Host: Associate Prof. Liu Fang, Nanjing University of Posts and Telecommunications, China	
09:00-09:20	Invited Speaker 3 Professor Yanlang Hu, Xi'an Institute of Space Radio Technology, China <i>Speech Title: Onboard model compression and acceleration</i>
09:20-09:40	Invited Speaker 4 Associate Professor Zhen Ye, Chang'an University, China <i>Speech Title: Garbage Classification Model Integrating Attention Mechanism</i>
09:40-10:00	Invited Speaker 5 Lecturer Kun Jiang, Xi'an University of Technology, China <i>Speech Title: Adaptive Locality Sensitive Analysis Representation Learning via K-SVD Algorithm</i>
10:00-10:15	<i>Break Time</i>

Password 102123	Room A https://us02web.zoom.us/j/84697177419	Room B https://us02web.zoom.us/j/8727982442	Room C https://us02web.zoom.us/j/8316827147
10:15-12:30	Session 7: Image Reconstruction and 3D model P0016, P0028, P1042, P1053, P1375, P1390, P1416, P1426, P1429	Session 8: Intelligent Detection Technology and Algorithm P0001, P0024, P1043, P1058, P1384, P1394, P1404, P1420, P1421	Session 9: Optical imaging System and Computer Photography P0010, P1428, P1378, P0002, P1061, P1059, P1056, P0040, P0019

DAY 3- CONFERENCE AGENDA

October 23, 2022, Beijing Time

Password 102123	Room A https://us02web.zoom.us/j/84697177419	Room B https://us02web.zoom.us/j/87279824425	Room C https://us02web.zoom.us/j/83168271473
13:30-15:45	Session 10: Computer Aided Design and Image Processing G402, G404, G410, P0038, P1052, P1063, P1369, P1430, P1908	Session 11: Computer Model and Image Application G405, G411, P0021, P1065, P1397, P1408, P1424, P1431, P1533	Session 12: Image Analysis and Method P0007, P0008, P0033, P0035, P1387, P1396, P1398, P1409, P1415
15:45-16:00	Break Time		
16:00-18:15	Session 13: Image Detection and Recognition G406, P0017, P0018, P0022, P0036, P0041, P1047, P1381, P1392	Session 14: Image Fusion P0003, P1044, P1382, P1406, P1412, P1414, P1532, P2017	



Professor Xu-Cheng Yin
University of Science and Technology Beijing

Room A: <https://us02web.zoom.us/j/84697177419>

Password: 102123

Xu-Cheng Yin is a full professor and the dean of School of Computer and Communication Engineering, University of Science and Technology Beijing, China. He received the PhD degree from the Institute of Automation, Chinese Academy of Sciences, in 2006. He was a visiting professor in the College of Information and Computer Sciences, University of Massachusetts Amherst, USA, for three times (Jan 2013 to Jan 2014, Jul 2014 to Aug 2014, and Jul 2016 to Sep 2016).

His research interests include pattern recognition, computer vision, information retrieval, and document analysis and recognition. He has published more than 100 research papers (IEEE T-PAMI, IEEE T-IP, CVPR, ICDAR, etc.). From 2013 to 2019, his team had won the first place of a series of text detection and recognition competition tasks for 15 times in ICDAR Robust Reading Competition. He is supported by National Science Fund for Outstanding Young Scholars (China).

Speech Title: Open-Set Text Recognition: Definition, Framework and Algorithm

Abstract: There are a variety of novel data, novel patterns and novel labels emerged for pattern recognition and computer vision in open-set scenarios, e.g., autonomous driving and Internet multimedia understanding. In this talk, we introduce a new task, open-set text recognition, which can spot unknown (novel) characters or words, and recognize them when the corresponding side-information is provided. First, the definition and the generic framework of open-set text recognition are described. Next, a representative algorithm for open-set text recognition, i.e., character-context decoupling (separating word-contextual information and character-visual information), is reported. Finally, some potential open-set text recognition applications, e.g., text recognition for minority languages, are also discussed.



Professor Xiaochun Cao, IET Fellow
University of Chinese Academy of Sciences, China

Room A: <https://us02web.zoom.us/j/84697177419>

Password: 102123

Xiaochun Cao is a Professor of School of Cyber Science and Technology, Sun Yat-sen University. He received the B.E. and M.E. degrees both in computer science from Beihang University (BUAA), China, and the Ph.D. degree in computer science from the University of Central Florida, USA, with his dissertation nominated for the university level Outstanding Dissertation Award. After graduation, he spent about three years at ObjectVideo Inc. as a Research Scientist. From 2008 to 2012, he was a professor at Tianjin University. Before joining SYSU, he was a professor at Institute of Information Engineering, Chinese Academy of Sciences. He has authored and coauthored over 200 journal and conference papers. In 2004 and 2010, he was the recipients of the Piero Zamperoni best student paper award at the International Conference on Pattern Recognition. He is on the editorial boards of IEEE Transactions on Image Processing and IEEE Transactions on Multimedia, and was on the editorial board of IEEE Transactions on Circuits and Systems for Video Technology.

Speech Title: "ILL-Posed" Computer Vision Tasks

Abstract: Computer vision tasks range from the simple perspective projection matrix estimation in a traditional camera calibration application to the large-scale foundation model fitting in a contemporary object detection cloud service. One may solve most computer vision tasks through fitting functions mapping the dense, if not continuous due to quantization, visual input to a discrete and meaningful output space, including categories, bounding boxes, and depths. Due to the significant difference in cardinalities of the domain and codomain, these mapping functions fail to meet one of the three Hadamard criteria for being well-posed. In other words, the unstable computer vision solution does not depend continuously on the parameters or input data. Many researchers are trying their best to design or learn computer vision algorithms being sufficiently robust to complex perturbations such as occlusion, smoke, rain, and fog. There are also scholars looking for dedicated but powerful adversarial perturbations. Does there exist an invariant backdoor perturbation that is capable to push an arbitrary image across the decision boundary in a classification task? Are all perturbations adversarial? In this talk, I will introduce these questions our team is exploring and briefly outline some of the progress. However, much still remains unclear in spite of our efforts, and we reiterate that there might not have the answers we're looking for before AI undergoes a brand new paradigm shift.



Professor Ir Dr Yiu-Ming Cheung (FIEEE, FAAAS, FIET, FBCS) Hong Kong Baptist University

Room A: <https://us02web.zoom.us/j/84697177419>

Password: 102123

Yiu-ming Cheung (SM'06--F'18) received Ph.D. degree from Department of Computer Science and Engineering at The Chinese University of Hong Kong in 2000. He joined the Department of Computer Science in Hong Kong Baptist University (HKBU) in 2001, and became a full professor since 2012. He is an IEEE Fellow, AAAS Fellow, IET Fellow, and British Computer Society (BCS) Fellow. His research interests include Machine Learning, Intelligent Visual Computing, Pattern Recognition, Data Mining, Watermarking, and Optimization. He has published over 250 articles in the high-quality conferences and journals. He was ranked the World's Top 1% Most-cited Scientists in the field of Artificial Intelligence and Image Processing by Stanford University in 2020. His three co-authored papers are selected as ESI Highly Cited Papers (i.e. listed in Top 1% globally in the corresponding discipline). He has been granted one Chinese patent and two US patents. In particular, selected from 1000 new inventions and products of 700+ competition teams from 40 countries, he was awarded two most prestigious prizes: (1) the Gold Medal with Distinction (i.e. the highest grade in Gold Medals) and (2) Swiss Automobile Club Prize, in the 45th International Exhibition of Invention, Geneva, Switzerland, on March 29-April 2, 2017, in recognition of his innovative work. Also, he was the Gold Award Winner of Hong Kong Innovative Invention Award in the Seventh Hong Kong Innovative Technologies Achievement Award 2017. In addition, he won the Gold Medal with Congratulations of Jury (i.e. the highest grade in Gold Medals) at the 46th International Exhibition of Inventions of Geneva 2018. Prof. Cheung is the recipient of: (1) 2011 Best Research Award in Department of Computer Science, HKBU, (2) Best Paper Awards in SEAL'2017, ISICA'2017, ICNC-FSKD'2014, and IEEE IWDVT'2005, respectively, (3) Best Student Paper Award in ISMIS'2018, (4) Best in Theoretical Paper Award in WI-IAT'2020, and (5) 2017 IETI Annual Scientific Award. Also, he is the recipient of 2017-2019 Albert Nelson Marquis Lifetime Achievement Award. He has been elected as an IEEE CIS Distinguished Lecturer, and the Changjiang Chair Professor awarded by Ministry of Education of China. He is the Founding Chairman of IEEE (Hong Kong) Computational Intelligence Chapter, and the Chair of Technical Committee on Intelligent Informatics (TCII) of IEEE Computer Society. He has served in various capacities (e.g., Organizing Committee Chair, Program Committee Chair, Program Committee Area Chair, and Financial Chair) at several top-tier international conferences, including IJCAI'2021, ICPR'2020, ICDM'2017 & 2018, WCCI'2016, WI-IAT'2012, ICDM'2006 & WI-IAT'2006, to name a few. He is an Associate Editor for several prestigious journals, including IEEE Transactions on Neural Networks and Learning Systems, IEEE Transactions on Cybernetics, IEEE Transactions on Emerging Topics in Computational Intelligence, IEEE Transactions on Cognitive and Developmental Systems, Pattern Recognition, and Neurocomputing, as well as the Guest Editor in several international journals. Currently, he is an Engineering Panel member of Research Grants Council, Hong Kong, and a Fellow Committee member of IEEE Computational Intelligence Society.



Professor Xiaorong Xue, Liaoning University of Technology, China

Bio: Xiaorong Xue received the PhD degree in Computer Science and Technology from Northwestern Polytechnical University, Xi'an, China, in 2004. From 2004 to 2006, he was a Post-Doctoral Research Fellow in Remote Sensing Image Processing, Peking University, Beijing, China. From 2014 to 2015, he was a Visiting Scholar in intelligent information processing, George Mason University, Virginia, USA. He has authored or coauthored more than forty research papers in China and abroad. He is currently a Professor with the School of Electronics & Information Engineering, Liaoning University of Technology, Jinzhou, China. His research interests include image processing, pattern recognition, and remote sensing.

Speech Title: Remote sensing image fusion based on MobileViT and spatial detail reconstruction

Abstract: The fusion of remote sensing images has made important progress in recent years, and the fusion of low-resolution hyperspectral images and high-resolution multispectral images into high-resolution hyperspectral images is also a popular direction at present. How to make the spectral information fidelity and spatial detail enhancement of the fused images achieve better results is the key problem to be solved, and the methods of convolutional neural network and deep learning have achieved good results in this area in recent years. In this paper, MobileViT is introduced based on neural network to improve the accuracy of the image during initial fusion, which is divided into two parts, the feature extraction part is added to MobileViT after fusion by neural network to achieve a better optimization effect; the reconstruction part after fusion restores the image information by constraining the space and detail loss. The Experiments result on two commonly used hyperspectral datasets show that the method in this paper has better performance.



Associate Professor Jungang Yang, National University of Defense Technology, China

Bio: Jungang Yang received the B.E. and Ph.D. degrees from National University of Defense Technology (NUDT), in 2007 and 2013, respectively. He was a visiting Ph.D. student with the University of Edinburgh, Edinburgh from 2011 to 2012. He is currently an associate professor with the College of Electronic Science and Technology, NUDT. His research interests include computational imaging, image processing, infrared small target detection and compressive sensing. Dr. Yang received the New Scholar Award of Chinese Ministry of Education in 2012, the Youth Innovation Award and the Youth Outstanding Talent of NUDT in 2016, the Excellent Doctoral Dissertation of Hunan Province Award in 2016, the Excellent Doctoral Dissertation of China Education Society of Electronics Award in 2017, and the Young Talents Project of Hunan in 2020.

Speech Title: Camera Array Computational Imaging Technology and Application

Abstract: Optical imaging system is one of the most important type of sensors to acquire information. Although currently used optical imaging sensor uses CCD or CMOS instead of chemical sensitive film, which used in early optical imaging system, the basic scheme of optical imaging system is still unchanged. The currently used optical imaging scheme belongs to the direct imaging technology, i.e. use a set of reflector mirrors or lens to obtain a focused image of the scene, then use chemical sensitive film or CCD or CMOS to record the image. Although currently used optical imaging system has made significant development and has high performance, there still exist some limitations and drawbacks, such as dependence of physics and manufacture technology, high cost of volume, weight, and price, performance limitation, and inflexible working scheme. This talk will show a novel camera array computational imaging technology. Compared to traditional imaging technology, the characteristics and advantages of camera array computational imaging technology are shown as follows:

1. Computational imaging is the combination of optic and signal processing (i.e. computation).
2. A high performance imaging result can be obtained by synthesize a set of low-cost and simple sensors. And the volume, weight, and price can be reduced significantly.
3. The performance limitations can be broken through by using this new imaging scheme.
4. The parameters of imaging system can be easily adjusted during the computation process. And many advanced algorithms can be used to enhance the performance.

This talk will discuss the limitations of traditional optical imaging technology, the basic theories of computational imaging technology, the preliminary results of camera array computational imaging technology, and the future perspectives of computational imaging.



Professor Yanlang Hu, Xi'an Institute of Space Radio Technology, China

Bio: Yanlang Hu is a professor at Xi'an Institute of Space Radio Technology. He is the top talent of China Aerospace Science and Technology Corporation, Outstanding Young talent of the China Academy of Space Technology. His research interests include high speed satellite data transmission and onboard artificial intelligence processing technology. Prof. Hu is the recipient of 6 scientific research awards: (1) the third prize of the National Defense Science and Technology Invention Award, (2) the first prize of the Creation Achievement Award of the China Creative Studies Institute, (3) the Second Prize of the National Defense Science and Technology Progress in Shaanxi Province, (4) the third prize of Science and Technology Progress Award of China Aerospace Science and Technology Corporation, (5) the Second Prize of the Creation Achievement Award of SHANXI Creative Studies Institute, (6) and the bronze medal of Shaanxi Science and Technology Workers Innovation and Entrepreneurship Competition. He was granted more than 50 national invention patents and published more than 10 academic papers. He has presided over/participated in many national key pre-research projects, the National Natural Science Foundation of China, etc.

Speech Title: Onboard model compression and acceleration

Abstract: Due to its outstanding performance on numerous tasks, deep neural network (DNN) has a wide range of applications in space missions such as remote sensing, communication, deep space exploration, and in-orbit maintenance. However, it suffers from problems such as heavy model parameters and large computational costs, which are particularly severe in the space environment. This report describes the compression and acceleration techniques for convolutional neural networks when deployed onboard based on the space application requirements. This report includes the following four parts: 1) the applications of convolutional neural networks in space, 2) the methods and techniques for compression of deep convolutional models, 3) techniques for the acceleration of deep models onboard, and 4) our works in this field. To address the problems of DNN applications in the space environment, the study of efficient onboard models and their acceleration techniques is not only of theoretical importance but also a guide to the deployment of DNNs onboard.



Associate Professor Zhen Ye, Chang'an University, China

Bio: She received the B.S. degree in Electronic & Information Engineering M.S. and the Ph.D. degree in information & communication engineering from Northwestern Polytechnical University, China, in 2007, 2010 and 2015, respectively. Meanwhile, she spent one year as a co-training Ph. D student from September, 2011 to October, 2012 in Mississippi State University, USA.

She is currently an Associate Professor with the School of Electronics and Control Engineering, Chang'an University, Xi'an. Her research interests include hyperspectral image analysis, pattern recognition and machine learning.

Speech Title: Garbage Classification Model Integrating Attention Mechanism

Abstract: In this paper, we propose a garbage classification model that integrates the attention mechanism and multiple network optimization methods. First, we construct a four-category primary network for recyclable garbage, kitchen garbage, hazardous garbage, and other garbage. And then, four secondary networks are constructed to map the above 4 primary categories to 40 secondary classes. Both the primary and secondary networks take Resnet101 as the main backbone network and integrate attention mechanism, Focal loss function, and warm-up learning rate. The experimental results prove that the proposed model has a high classification performance for the HUAWEI cloud garbage classification dataset.



Lecturer Kun Jiang, Xi'an University of Technology, China

Bio: Kun Jiang received his B.S., M.S., and Ph.D. degrees in Computer Science and Technology at College of Computer, National University of Defense Technology (NUDT), China, in 2008, 2011, and 2015, respectively. He held a postdoc research position at Xi'an Jiaotong University from 2016 to 2018. Now he has been working as a lecturer in the Faculty of Computer Science and Engineering, Xi'an University of Technology, China. His current research interest includes big data processing, coding optimization theory, and applications in complex network, multimedia data compression and recommendation systems et., al.

Speech Title: Adaptive Locality Sensitive Analysis Representation Learning via K-SVD Algorithm

Abstract: Recent years, analysis representation learning and its applications for classification have been well explored and applied, due to its exible representation ability and low classification complexity. With a learned analysis dictionary, test samples can be transformed into a robust low-dimensional subspace for classification efficiently. However, the underling geometrical structure of data samples has rarely been explored and reliably married with analysis representation learning to enhance the model discrimination. To remedy these deciciencies, we propose a novel adaptive locality sensitive analysis representation learning model for pattern classification (ALAR). Specifically, we rst impose an adaptive weighted constrained graph regularization into analysis representation learning model to uncover the intrinsic structure of data samples. Through the learned analysis dictionary, we transform the samples into a robust and discriminative subspace representation where the structural information can be preserved extremely. Moreover, the ALAR model is iteratively solved by the synthesis K-SVD algorithm and gradient technique with closed-form solutions. Experimental results on four benchmark databases demonstrate the classification performance superiority of our ALAR model over the state-of-the-art methods.

Session Topic: Pattern Recognition

Time: 13:30-15:45

Session Chair: Professor Xiwen Zhang, Beijing Language and Culture University, China

<p>P1050 13:30-13:45</p>	<p>Joint Multi Half-orientation Features Learning for Contactless Palmprint Recognition Mengwen Li, Huaiyu Liu, Qianqian Meng, Qing Wang Presenter: Mengwen Li, Huaibei Normal University, China</p> <p>Abstract: Contactless palmprint recognition attracted much attention in recent years for it is more user-friendly and sanitary compared with contact palmprint recognition. However, due to the lack of restrictions on the position of the palms when collecting images, there are severe translation and rotation in contactless palmprint images, which will seriously affect the recognition accuracy. Conventional palmprint recognition methods based on the hand-craft features mainly focus on the characteristics of palmprint images, but the correlations among samples are usually neglected. Therefore, it is urgent that extracting the stable and discriminative features to improve the recognition performance. To solve this problem, a joint multi half-orientation features learning method (JMHOFL) was proposed in this article. First, we extracted the orientation features using banks of half-Gabor filters, and constructed the multi half-orientation features (MHOF) of the palmprint image. To overcome the effects of translation and rotation, MHOF obtained multi orientation codes and performed block-wise statistics on these orientation codes. Afterwards, a joint low-rank inter-class sparsity least squares regression (JLRICS_LSR) was proposed to study more stable and discriminative features from MHOF. JLRICS_LSR takes into account the structure between samples, and reduces the influence of noises. Lastly, Euclidean distance is used for feature matching. Experiments on CASIA, IITD, and Tongji palmprint databases showed the promising performance of the proposed method.</p>
<p>P1370 13:45-14:00</p>	<p>Classroom Quality Analysis Based on Facial Expression Recognition Jin Zhuang, Yuan Bo, Lei QinChuan, Yue Yixin Presenter: Zhuang Jin, College of Computer, Guangdong University of Science and Technology, China</p> <p>Abstract: Affected by the post-epidemic epidemic, most colleges and universities have chosen to adopt the "classes suspended but learning continues" classes. As a result, teachers are unable to obtain timely feedback from students in the teaching process, resulting in a decrease in the quality of teaching. This paper used the Tensorflow framework to implement a class evaluation method based on facial expression recognition. This method can obtain the facial expression of each student in real time, return the analyzed learning situation to the teacher. The teacher can conduct classroom teaching according to the real-time feedback data. Adjust for better classroom teaching. Finally, through the class test, it is found that in the class with data analysis prompts, teachers can better grasp the overall situation of the classroom, thereby improving students' attention and classroom learning quality.</p>

SESSION 1

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

P1388 14:00-14:15	<p>Prior-Guided Dense Up-down Sampling Network for Face Super-Resolution and Recognition Simiao Wang, Yunan Liu, Chunpeng Wang, Mingyu Lu Presenter: Yunan Liu, Dalian Maritime University, Dalian, Liaoning</p> <p>Abstract: Identifying low-resolution (LR) face images in the wild is still an open and challenging task, as discriminative feature is hard to learn from LR image with blurry appearance. To address this problem, many face super-resolution (SR) methods have been proposed to reconstruct high-resolution (HR) face images from the observed LR images. However, most of existing methods have two drawbacks: they tend to generate over-smoothed results, and they largely ignore to recover the facial identity information. In this paper, we propose a new multi-task framework for very small face image SR and recognition. First, we propose the Dense Up-down Sampling Unit (DUSU), which is able to effectively represent the nonlinear LR-to-HR mapping via an error-correcting feedback mechanism. Secondly, we introduce an effective and robust facial prior knowledge, the high frequency sub-bands of Non-Subsampled Contourlet Transform (NSCT), to enhance the texture details of super-resolved images. Third, we introduce an evaluation network with perceptual loss to recover the identity information of reconstructed face images. Extensive experiments demonstrate that our method not only achieves more appealing results than the state-of-the-art methods in terms of traditional SR metrics, but also significantly improve recognition accuracy of very small faces in the wild.</p>
P1389 14:15-14:30	<p>Max-margin class incremental learning with Mixup augmentation Daihu Ye, Ke Xu, Chunyan Xu, Zhen Cui Presenter: Daihu Ye, Nanjing University of Science and Technology, China</p> <p>Abstract: The most famous dilemma for class incremental learning(CIL) is the catastrophic forgetting phenomenon, where the trained model quickly forgets to recognize the old classes when learning novel ones. To cope with this problem, we propose a novel max-margin class incremental learning with less forgetting, which can well discriminate the relationship between classes and prevent knowledge forgetting in the incremental learning process. Specifically, we introduce a max-margin objective function to explicitly enforce the maximum distance between different classes larger than a predefined margin, which can avoid the ambiguity between old and new classes and then increase the discriminative ability of the learned classifier. Further, we adopt a mixup augmentation mechanism by performing the mixup operation on the reserved images of old classes and the incremental images of new classes, especially to reduce knowledge forgetting. Comprehensive evaluations and comparisons on three public datasets (including CIFAR-100, ImageNet-Subset and ImageNet) well demonstrate that our proposed CIL approach can effectively improve performance compared to the existing CIL methods.</p>
P1391 14:30-14:45	<p>Automatic License Plate Recognition Using Transformer Teng Zhang, Wei Jia Presenter: Wei Jia, Hefei University of Technology, China</p>

SESSION 1

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

	<p>Abstract: Automatic license plate recognition (ALPR) is widely used in Intelligent Traffic System (ITS). Deep learning has become the mainstream technology for ALPR. At present, the most advanced ALPR technologies are all designed based on convolutional neural networks (CNNs). From 2020, the visual Transformer technology has been introduced into the computer vision (CV) field, and has achieved excellent performance in many CV tasks. However, there are few researches on ALPR using visual transformer technology. In order to study the license plate detection and recognition using visual transformer, we evaluate the performance of two representative object detection visual Transformer methods including DETR and Deformable DETR on three license plate datasets. Experimental results show that the visual transformer methods can achieve good detection results, but their detection and recognition performance is worse than that of some SOTA object detection methods based on CNNs. In the future, with the emergence of new visual Transformer methods and the construction of some large-scale license plate datasets, visual transformer will achieve better performance of license plate detection and recognition.</p>
<p>P1393 14:45-15:00</p>	<p>Person Re-identification with IBN Layer and Channel Attention Module for Indoor scenarios Hao Wang, Guoan Cheng, Yongdong Li, Guiyan Cai, Lu Sun and Shengke Wang Presenter: Guoan Cheng, Qingdao Harbour Vocational & Technical College, Qingdao</p> <p>Abstract: Person re-identification technology is being utilized increasingly frequently in autonomous processing and analysis of surveillance video jobs as a result of recent advancements in deep learning, particularly with safety precautions and smart transportation. As a result of the issues with inadequate illumination and reflection in indoor settings, At the moment, much of the related research on human re-identification concentrates on outside situations, with little attention paid to indoor scenarios. These make the process of person re-identification in complicated indoor scenarios very difficult. The indoor person re-identification algorithm is investigated in this research in order to increase the precision of person recognition in indoor settings. The IBN layer is an addition to the Resnet50 backbone network that uses a combination of instance normalization (IN) and batch normalization (BN) to eliminate individual appearance difference while retaining the feature difference of different individuals to address the issues with the obvious difference of light and shade in person images taken by indoor monitoring. To enhance the expressiveness capability of individual features, the attention module based on feature channel is added to the residual network. In specifically, the learning approach automatically determines the value of each channel in a person's attributes in order to amplify the important qualities and suppress the unnecessary ones. On the other hand, to address the issue of it being challenging to distinguish between similar people caused by more interference factors such as occlusion and reflection in indoor scene, we introduce triple loss in the model training process, which can make the model better learn the details of persons. The three primary validation data sets utilized in this study are Market1501, OUC365, and DukeMTMC-reID. The indoor style and high definition in the OUC365 data set are more obvious, the noise is more obvious in the Market1501 data set, and there is a significant difference in the number of photos among various people in the DukeMTMC-reID data set. The proposed method is tested on several data sets in this paper, and successful results are obtained.</p>
<p>P1395</p>	<p>Width-Resolution Multiplier Lightweight Network for Person Re-identification</p>

SESSION 1

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

15:00-15:15	<p>Jiatao Li, Cheng Feng, Yujie Li, Wei Fan, Qinghua Zhao* Presenter: Jiatao Li, Qinghua Zhao, Nanjing University of Finance & Economics, China</p> <p>Abstract: Person re-identification is a key sub-problem in image retrieval. It is a technique to judge whether there is a specific pedestrian in an image or video by using computer vision technology, that is, a monitoring pedestrian image is given to retrieve a cross-device pedestrian image. With the development of technology, many excellent algorithms have successfully achieved excellent result in video surveillance and security fields. However, which many ReID algorithm rely on large backbone network, such as ResNet, which are designed for image classification instead of ReID. This paper presents a combined depth space (CDS) network algorithm based on width and resolution. The algorithm is mainly designed by efficient calculation for pedestrian re recognition in the current backbone network, which focus on the combined pattern information in the pedestrian image by reproducing a pedestrian re recognition network based on the combined depth space architecture search through CDs. The low-cost search strategy of Top-k sample search strategy is used to make full use of the search space and avoid falling into local optimization. Furthermore, an effective Fine-grained Balance Neck (FBLNect) is presented to balance the effects of triplet loss and SoftMax loss during the training process. Finally, the experiments on three different data sets: Market1501, DukeMTMC and MSMT17, the proposed algorithm achieves better performance when the parameters are much smaller than those of other mainstream networks.</p>
P1402 15:15-15:30	<p>Semi-supervised learning for tongue constitution recognition Yichao Ma, Chunhong Wu, Tian Li Presenter: Yichao Ma, University of Science and Technology Beijing, China</p> <p>Abstract: Constitution recognition plays an important role in the prevention and treatment of diseases in Traditional Chinese Medicine (TCM). Using AI technology to analyze the characteristics of tongue images for automatic constitution identification can greatly improve the efficiency. In practice, the acquirement of tongue image constitution labels relies on the professional TCM doctors, which is costly and difficult. This causes the total amount of tongue images with constitution labels in the existing datasets is relatively small. In order to make full use of the information in the datasets, semi-supervised learning (SSL) is introduced in this paper with a large number of unlabeled tongue images to assist the training of the network model. In addition, focal loss is introduced to tackle the unbalanced distribution of tongue images, and attention mechanism is also added in the process of feature extraction. Experiments verified the effectiveness of our method.</p>
P1425 15:30-15:45	<p>Unsupervised Person Re-Identification Based On Intermediate Domains Haijie Jiao, Mengyuan Ding, and Shanshan Zhang Presenter: Haijie Jiao, Nanjing University of Science and Technology, China</p> <p>Abstract: Unsupervised domain adaptive person re-identification (UDA re-ID) aims to transfer knowledge learned from the labeled source domain to the</p>

SESSION 1

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

unlabeled target domain. Most recent methods focus on narrowing down the domain gap between the source and target domains while ignore the bridge between them. %Specifically,
In this work, we explicitly model appropriate intermediate domains and construct two adaptation pairs (“source-intermediate” and “intermediate-target”) instead of the original “source-target” one pair adaptation. The purpose is to ease the adaptation difficulties caused by large domain gaps, making the adaptation process more smooth. To generate the intermediate domain, we use image-to-image translation methods which generate images that have the same contents and ID labels shared with the source domain and similar style to the target domain. When evaluated on standard benchmarks, our proposed methods outperforms the state of the arts by a large margin on the target domains.

Session Topic: Image Segmentation

Time: 13:30-15:45

Session Chair: Associate Professor Lili Nurliyana Abdullah, Universiti Putra Malaysia, Malaysia

<p>G408 13:30-13:45</p>	<p>Semantic Segmentation of High Spatial Resolution Remote Sensing Imagery Based on Weighted Attention U-Net Yue Zhang, Leiguang Wang, Ruiqi Yang, Nan Chen, Yili Zhao, Qinling Dai* Presenter: Yue Zhang, Southwest Forestry University, China</p> <p>Abstract: In recent years, with the development of deep learning and attention mechanism, more research has been carried out to realize semantic image segmentation based on deep learning integrated attention mechanisms. However, the current semantic segmentation methods have low segmentation accuracy, high computation cost, and serious loss of detailed information. In this paper, a lightweight designed attention gate model was introduced to reduce the computation cost. And because it can suppress irrelevant regions in the input image, while highlighting the salient features of specific tasks, the combination of the two weighting factors input features (x^l) and gating signal (g) in this structure can improve segmentation accuracy and reduce loss of detail. Therefore, this study used the weighted attention U-Net network to perform semantic segmentation on the GID dataset and finally evaluated it on the four indicators of Precision, Recall, F1-Sorce, and mIoU. This result shows that different weight values have a more significant impact on the experimental results. The attention U-Net with the best weight combination compared with the traditional U-Net network, Precision, Recall, F1-Sorce, and mIoU are increased by 0.88%, 1.4%, 1.13%, and 1.2%, respectively. Compared with the original attention U-Net, Precision, Recall, F1-Sorce, and mIoU are increased by 0.86%, 1.24%, 1.04%, and 1.75%, respectively.</p>
<p>P0032 13:45-14:00</p>	<p>Research on thyroid CT image segmentation based on U-shaped convolutional neural network Yunzhi Zeng, Yanfen Zhang, Ning Gong, Mei Li, Meili Wang Presenter: Yunzhi Zeng, Northwest A&F University, China</p> <p>Abstract: Computer-aided thyroid CT image segmentation aims to provide imaging physicians and clinicians with auxiliary diagnostic suggestions and improve the efficiency of physicians in diagnosing the thyroid region. However, it is still a challenging task to distinguish the thyroid from other surrounding tissues due to adhesions in thyroid CT images caused by thyroid disease. To achieve accurate segmentation of thyroid CT images under the intervention of different types of thyroid nodules, we proposed a thyroid segmentation network named ResUnet by introducing the residual learning idea to UNet. Our network controls the gradient dispersion by incorporating a batch normalization operation and an intermediate layer regularization operation, then solves the degradation problem by introducing the residual connections into the convolution operation. Moreover, our ResUnet network can converge faster with the same number of layers, thus supporting a deeper design of the network. Extensive experiments also validated the high accuracy (94.10%), specificity (98.94%), and sensitivity (96.34%) of the proposed ResUnet for the segmentation of thyroid nodules, which can assist CT</p>

SESSION 2

Room B: <https://us02web.zoom.us/j/87279824425>

October 22, 2022, Beijing Time

	physicians in the diagnosis of the thyroid gland.
P1055 14:00-14:15	<p>A correntropy-based local additive bias-field-corrected image fitting model for image segmentation Haoming Chen, Bo Chen, Yuru Zhang, Wensheng Chen, Yuwen Jiang, Binbin Pan Presenter: Haoming Chen, Shenzhen University, China</p> <p>Abstract: Segmenting images with intensity inhomogeneity is a difficult problem in the field of image processing. In order to deal with it, this paper presents a novel local additive bias-field-corrected image fitting model using correntropy criterion. Firstly, a local additive bias-field-corrected fitting image model in the level set formulation is built by using information of bias field and reflection ratio simultaneously. Secondly, an energy function is introduced by minimizing the difference between the fitting image and original input image in a neighborhood, which makes it effective in segmenting images with intensity inhomogeneity. Thirdly, by incorporating the correntropy criterion into the new energy, the proposed method can reduce the impact of noise on segmentation results. Finally, the local energy is integrated with respect to the neighborhood center and a global result of image segmentation in the whole domain is obtained. Experiments show that our method is robust to different kinds of noises, and the computational efficiency is better than the existing bias field correction model.</p>
P1062 14:15-14:30	<p>Metal Surface Defects Segmentation Method using Cycle Generative Adversarial Networks on Small Datasets Chuxin Yang, Zhenglin Li, Longping Liu Presenter: Chuxin Yang, Guangxi University of Science and Technology, China</p> <p>Abstract: Aiming at the problems of small metal surface defect samples in industrial production and the difficulty of data annotation in supervised segmentation algorithms, a background reconstruction method based on Cycle generative adversarial networks is proposed to realize metal surface defect segmentation in combination with the traditional threshold segmentation algorithm. Firstly, the corresponding defect-free template is reconstructed from the defect image using Cycle generative adversarial networks, and the defect image and the reconstructed template are subjected to the differential subjected to eliminate the influence of the background texture of the defect sample. Finally, the segmentation process is performed using the adaptive thresholding segmentation method. In order to adapt to the small sample training as well as to improve the performance of background reconstruction of the generative network, the U-Net network structure is used as the generator, and the attention mechanism is also introduced. Meanwhile, the L1 loss and Multi-scale SSIM loss are combined to design the cycle consistency loss for training. The experimental results show that the method in this paper can accomplish good defect segmentation results using a small number of defective samples.</p>
P1064 14:30-14:45	<p>A 3D Self-Adjustable Organ Aware Deep Network for Abdominal Segmentation in CT Images Shenhai Zheng, Haiguo Zhao, Hong Wang and Laquan Li, Presenter: Haiguo Zhao, Chongqing University of Posts and Telecommunications, China</p>

	<p>Abstract: CT scan is an important reference means of disease diagnosis. Automatic segmentation of organ regions can save a lot of time and labor costs, and allow doctors to produce more intuitive observations of the internal organization of the human body structure. However, automatic multi-organ segmentation remains challenging due to the complicated anatomical structures and low tissue contrast in CT images. Traditional segmentation methods are relatively inefficient for organ segmentation and the traditional network architectures are rarely designed to meet the requirements of lightweight and efficient clinical practice. In this paper, we propose a novel network named Self-Adjustable Organ Attention U-Net (SOA-Net) to overcome these limitations. The SOA-Net includes multi-branches feature attention (MBFA) module and the feature attention aggregation (FAA) module. These two modules have multiple branches with different kernel sizes to adaptively adjust their receptive field size based on multiple scales of the target organs. The SOA-Net is a 3D self-adjustable organ aware deep network which can adaptively adjust receptive fields sizes based on multiple scales of the target organs to realize the efficient segmentation of multiple abdominal organs. We evaluate our method on a challenging abdominal multi-organ CT dataset, and the final experiments proved that our model achieves the best segmentation performance compared with the state-of-the-art segmentation networks.</p>
<p>P1372 14:45-15:00</p>	<p>Semantic Segmentation and Image Quality Assessment of Anterior Segment Images for Smartphones Yuanhao Chen, Kun Huang, Wen Fan and Qiang Chen Presenter: Nanjing University of Science and Technology</p> <p>Abstract: Eye diseases have always been a threat to public health worldwide. Many people suffer from various eye diseases, but there are not enough skilled ophthalmologists to meet the demand for medical care. Thus, finding a method to perform ophthalmic examinations automatically and conveniently is necessary. Although many well-designed ophthalmic diagnosis systems have been proposed to diagnose ophthalmic disorders using artificial intelligence algorithms, they tend to depend on high-quality anterior segment images to perform appropriately. In order to capture high-quality anterior segment images simply with a smartphone, we proposed a system including a semantic segmentation model and an image quality assessment method for anterior segment images. Our proposed segmentation model, namely the multi-task anterior segment image semantic segmentation (MT-ASISS) model, has a designed multi-task learning network structure and achieves an accuracy of 92.63% in Dice and a processing speed of 138ms per frame on smartphones. Our anterior segment image quality assessment method, namely Mixed-Parameters Quality Assessment (MPQA) method, has an accuracy of 92.6% in mean average precision (mAP). The system can help reduce the demand for professional image collecting equipment, share the burden of choosing satisfactory images manually and improve the efficiency of acquiring anterior segment images.</p>
<p>P1373 15:00-15:15</p>	<p>MFCTrans-net: a Multi-scale Fusion and Channel Transformer net for retinal vessel segmentation Zhuo Li, Biyuan Li, Jun Zhang, Jianqiang Mei, Binghui Li Presenter: Zhuo Li, Tianjin University of Technology and Education, China</p>

	<p>Abstract: Automatic segmentation of retinal blood vessels is a nontrivial task due to the complexity of retinal fundus image. In this paper, a new network named MFCTrans-net is proposed for retinal blood vessel segmentation. The MFCTrans-net is an improvement over original U-Net, which can be summarized as (1) to better fuse the encoder and decoder features and reduce the semantic gap, we replace the skip connection in the original U-net by the Channel-wise Cross Fusion Transformer(CCT); (2) two side paths are added to the U-Net which allow the network to capture features at multiple scales; (3) a novel loss function is also proposed which focuses on the topology integrity of vessel meanwhile maintaining pixel segmentation accuracy. The proposed network has been developed and evaluated in the DRIVE, CHASE-DB1 and IOSTAR datasets, which offer a manual segmentation of the vascular tree by each of its images. The performance of our method is evaluated in terms of visual effects and quantitative evaluation metrics on these four publicly available datasets with comparison to several representative methods. Furthermore, we use the proposed method to segment a collection of the experimentally obtained retinal blood vessel images with poor quality. The experimental results demonstrate the performance of our proposed method.</p>
<p>P1380 15:15-15:30</p>	<p>Dual-Attention Deep Fusion Network for Multi-modal Medical Image Segmentation Shenhai Zheng, Xin Ye, Jiabin Tan, Haoyuan Li and Laquan Li Presenter: Xin Ye, Chongqing University of Posts and Telecommunications, China</p> <p>Abstract: Multi-modal medical image segmentation plays a vital role in clinical applications such as auxiliary diagnosis and surgical planning. However, it is still a challenging task to extract and fuse complementary information crossing modalities. Focusing on this target, this paper proposes a dual-attention deep fusion network for multi-modal medical image segmentation. Generally, this proposed network follows the typical encoder-decoder and skip connection workflow. Unlike existing methods, it applies two attention modules to deeply fuse multi-modal features twice at the encoder and skip connections, respectively. Specifically, we propose a cross-modal complementary feature learning (CMFL) attention module to deeply extract complementary feature information hidden between two modalities, and another multi-scale dual attention (MSDA) module to highlight important effective fusion features and suppress redundant features. Furthermore, the residual connections designed in the decoder can reduce the loss of semantic information caused by the upsampling process. Finally, the proposed network is tested on the Prostate dataset and BraTS dataset and compared with state-of-the-art segmentation networks. Experimental results demonstrate that our proposed method can enhance the cross-modal feature representation, make full use of the complementary information between different modalities, and obtain more accurate segmentation results.</p>
<p>P1401 15:30-15:45</p>	<p>Episode-based training strategy for zero-shot semantic segmentation Bo Xiong, Jianming Liu, Zhuoxun Jing Presenter: Bo Xiong, Jiangxi Normal University, China</p> <p>Abstract: We introduce episode-based training into zero-shot semantic segmentation (ZS3) for the first time. In particular, the model is trained on a set of simulated ZS3 tasks. The model gains the ability to predict simulated unseen classes over multiple episodes, which generalizes well to true unseen</p>

SESSION 2

Room B: <https://us02web.zoom.us/j/87279824425>

October 22, 2022, Beijing Time

classes after training on multiple episodes. On the basis of this training framework, we propose a visual semantic alignment network named VSAN as our basic model, which mainly includes a feature extractor and semantic projection network. This base model constrains the visual-semantic distribution of the same class by a distance measure between visual features and semantic prototypes. In the inference stage, the trained projection network can generate corresponding semantic prototypes for all classes, and predict the segmentation results of the entire image by measuring the distance between visual features and semantic prototypes. The base model VSAN is called EB-VSAN after using the episode-based training strategy. Our model is a discriminative model, as opposed to generative methods, our model does not need retraining the classifier when a new class emerges, while avoiding multi-stage training. Our extensive ZS3 experiments on the benchmark dataset show that the EB-VSAN model outperforms current state-of-the-art methods, specifically, our hIoU metric outperforms state-of-the-art methods by an average of 1.7% on PASCAL VOC, while Average 2.5% improvement on PASCAL Context.

Session Topic: Target Detection and Defect Detection

Time: 13:30-15:45

Session Chair: Professor Peishun Liu, Ocean University of China, China

P0013 13:30-13:45	<p>Research on PCB solder joint defect detection method based on machine vision Yuanpei Chang, Ying Xue, Yu Zhang, Jiajun Ma, Guangjie Li, Dandan Wu, Qiang Zhan, Jiancun Zuo Presenter: Yuanpei Chang, Shanghai Polytechnic University, China</p> <p>Abstract: Printed circuit boards (PCBs) are an essential component of electronic products, and detecting solder joint defects is critical in the PCB production process. Machine vision technology allows detection with high efficiency and cost-effectiveness. Therefore, this paper summarizes the basic principles of image processing-based and machine learning-based methods for defect detection and compares the advantages and disadvantages of both methods with relevant performance evaluation indicators. Finally, this paper contains a summary and an outlook.</p>
P0025 13:45-14:00	<p>An Improved YOLOX for Remote Sensing Image Object Detection Zhou Fang, Lin He, Yingqi Li Presenter: Zhou Fang, South China University of Technology, China</p> <p>Abstract: Recent studies have shown that the attention mechanism can further improve the detection accuracy of YOLOX algorithm in remote sensing images, and coordinate attention can well solve the long-range dependencies problem of the previous attention modules, but its attention weights is too redundant in the channel dimension. At the same time, there is a problem of example imbalance in the training of YOLOX algorithm for remote sensing image object detection. To solve the above problems, an improved YOLOX algorithm is proposed, which combines the improved coordinate attention and focal loss. The former not only further adopts pooling and convolution operations to make the attention weights no longer contain redundant channel information when it still has the potential to capture long-range dependencies, but also introduces 1D convolution layers to obtain the final attention weights in three different directions to make the model pay more attention to the effective parts of the features of remote sensing images. The latter optimizes the quality of the gradients, which makes the training more effective and improves the detection accuracy. Training and testing with open remote sensing image dataset. The detection results show the effectiveness and superiority of our method.</p>
P0027 14:00-14:15	<p>Multi-Scale Feature Fusion Attention Network for Infrared Small Target Detection Yidan Zhang, Chunlei Li, Yundong Liu, Zhoufeng Liu, and Ruimin Yang Presenter: Yidan Zhang, Zhongyuan University of Technology, China</p> <p>Abstract: Compared with other target detection tasks, infrared small target detection has the problem of feature information loss in deep networks due</p>

	<p>to fewer target pixels and the lack of color and texture features. To address aforementioned issue, a Multi-Scale Feature Fusion Attention Network (MSFFA) is proposed to better utilize shallow edge features and deep semantic features. Its main components contain Convolutional Block Attention Module (CBAM), Multi-Scale Receptive Field Feature Fusion Module (R3FM), and Bidirectional Feature Aggregation Network (BFANet). CBAM is designed to calculate the importance of each feature map and enhance useful features from the channel and spatial dimensions. R3FM is proposed to characterize the global context information of deep layers feature map to enlarge the network's receptive field for small targets detection with a larger range of location information. BFANet is developed to shorten the path of information exchange between different layers and reinforce the utilization of shallow features in the network. Moreover, the K-means clustering algorithm is adopted to optimize the width to height ratio of the bounding anchor, and it can better match the positive samples to improve the training performance. Extensive experiments on public infrared small target detection dataset demonstrate that the proposed method achieves better performance compared to the other state-of-the-art methods.</p>
<p>P1054 14:15-14:30</p>	<p>Improved YOLOv5 Network for Agricultural Pest Detection Yan Yu, Tian Sun, Jin Yan Presenter: Yan Yu, Nanjing University of Science and Technology, China</p> <p>Abstract: Agriculture is the source of food, clothing, and the foundation of human existence. In recent years, various countries have been increasing investment in agricultural production and actively carrying out agricultural pest control work. Therefore, how effectively realizing pest identification is a top priority at present. Traditional identification methods have disadvantages, such as time-consuming and laborious, untimely diagnosis, and limited diagnosis range. With the development of modern digital technology, image processing technology develops rapidly, which opens up a new way for pest identification. This paper proposes an improved network model for pest identification using the YOLOv5 target detection algorithm. Firstly, the data sets of pests and diseases are collected and marked. Then, an improved anchor frame size is proposed to make it more suitable for the data set used in this paper. Finally, an improved network structure of YOLOv5 is proposed, which improves the ability of the network to capture characteristic information. The experimental results show that the mean Average Precision (mAP@0.5) of the improved network model reaches 79.7%. At the same time, compared with Faster R-CNN, Dynamic R-CNN, Double-Head R-CNN, YOLOv3, and YOLOv5, the mAP@0.5 is improved by 8.5%, 8.1%, 8.8%, 9.3%, and 3.7% respectively.</p>
<p>P1060 14:30-14:45</p>	<p>Effective fabric defect detection using contrastive learning and layered fusion network Huosheng Xie, Weijie Xu Presenter: Weijie Xu, FuZhou University, China</p> <p>Abstract: Defect detection is a very important topic in industrial inspection. Due to the rapid expansion of computer computing power, manual-based defect detection approaches is no longer satisfactory, and deep learning-based detection techniques are increasingly being applied to defect detection</p>

SESSION 3

Room C: <https://us02web.zoom.us/j/83168271473>

October 22, 2022, Beijing Time

	<p>tasks. contrastive learning is better for defect detection tasks due to the advantages of its label-free training method. So the Efficient Fabric Defect Detector (EFDD), a high-efficiency defect detector based on contrastive learning approach and layered fusion network, is proposed in our paper. First, we propose the Layered Fusion network to encode defect types with different scale sizes and improve the feature extraction ability of the model. Second, a universal structure with an attention module is introduced, this module can be integrated into ResNet with little overhead. Third, we investigate the impact of parallel enhancement strategies on model training in the data imbalance, different combinations can affect the model to different degrees. Finally, we evaluate the model using three classical public datasets and the average improvement of EFDD over the three datasets is AP 2.1%, AR 1.5% compared to other contrastive learning methods. Compared to the supervised methods, our model accuracy is close to or even exceeds that of the supervised methods.</p>
<p>P1385 14:45-15:00</p>	<p>Gated Bidirectional Pyramid Context Network for Infrared Maritime Target Detection Guangrui Lv, Lili Dong, Junke Liang, and Wenhai Xu Presenter: Guangrui Lv, Dalian Maritime University, China</p> <p>Abstract: Detecting infrared small targets in complex marine environments is an important technology in maritime distress target search and tracking systems. To enrich the feature representation of maritime targets and suppress background noise, a gated bidirectional pyramid context network (GBPC-Net) is proposed. Firstly, a hierarchical feature extraction backbone is constructed to generate multi-scale feature, and then a gated bidirectional connection module (GBCM) is designed to aggregate hierarchical features of infrared maritime targets and eliminate complex background interference. Among them, the channel-based GBCM adopts the directions of top-to-down and bottom-to-up to aggregate multi-scale features from different layers into semantic-assisted detail features and detail-assisted semantic features. While the spatial-based GBCM further hierarchically modulates channel aggregation features in different directions to generate multi-scale gated aggregation features. Next, an adaptive pyramid context module (APCM) is introduced to learn the similarity between the local detail and the context information of different scales, which can emphasize the difference between small maritime targets and complex backgrounds. Subsequently, the features from APCM are used to guide the fusion of detailed features in lower-layer networks, and the aggregated feature map is applied to infrared maritime target detection. Finally, a target detection dataset derived from the real marine environment is constructed, and a series of comparative experiments are conducted on this dataset and the results show that our method can more accurately detect infrared maritime targets than some state-of-the-art methods.</p>
<p>P1386 15:00-15:15</p>	<p>Lightweight Human Pose Estimation with Attention Mechanism Chu Xiaoshuai, Ji Ruirui, Dong Danyang, Xi Yuzhuo Presenter: Chu Xiaoshuai, Xi'an University of Technology, China</p> <p>Abstract: There are certain inevitable challenges in human pose estimation tasks based on deep learning methods, such as large amount of network parameters and high computational complexity. This paper proposes a lightweight network to reduce the scale of model parameters and</p>

	<p>computational complexity, meanwhile improve the accuracy of the human pose estimation task. The new method takes the high-resolution networks HRNet-32 as the basic framework and replaces the basic module with the MBConv lightweight module. The attention mechanism is incorporated into the network to model the context information, so as to improve the perception ability and the feature extraction ability of the module and improve the accuracy of human pose estimation. The experimental results on COCO2017 show that the proposed network can detect human key points with high precision even when the amount of parameters is reduced by 56%, which verifies that the proposed method has good lightweight performance.</p>
<p>P1400 15:15-15:30</p>	<p>A Multi-Attention Fusion Mechanism for Collaborative Industrial Surface Defect Detection Xiaoli Yue, Guoqiang Zhong Presenter: Xiaoli Yue, Ocean University of China, China</p> <p>Abstract: The surface defects of industrial materials can seriously affect product quality, so that the industry has the demand for high precision defect detection algorithms. Towards this issue, we investigate the methods of defect detection accuracy improvement and collaborative training. This paper first innovatively proposes the multi-attention fusion mechanism (MAF), which integrates both channel and space dimensions, and embeds the spatial pyramid structure into the attention module. It alleviates the problem of inconspicuous defective features and enhances the feature extraction ability. Secondly, this paper proposes the mixForm data augmentation algorithm to transform the target defects in space and shape to tackle the problem of few samples. The detection model's ability to recognize defects of multiple types and small objects is simultaneously improved. Thirdly, the split federated learning (SFL) framework enables collaborative training of industrial surface defect detection models with a low resource cost. Our scheme improves model training efficiency and achieves high accuracy detection for small amounts of defect samples. Finally, the experimental results show that MAF with the aid of mixForm achieves 82.91 mAP on the NEU-DET dataset. Using MAF, the defect detection algorithm achieves at least 1.89 mAP improvement over using other attention mechanisms. The experiments also demonstrate that SFL achieves faster convergence and higher detection performance than traditional federated learning approaches.</p>
<p>P1903 15:30-15:45</p>	<p>An Aircraft Surface Damage Inspection Method Based on Improved SSD Hongjun Qiu, Daitao Wang and Wenjing Yu Presenter: Hongjun Qiu, Guangzhou Civil Aviation College, China</p> <p>Abstract: Aiming at the current weakness of aircraft airline maintenance and the requirement of developing aircraft surface inspection system based on machine vision: high precision, real-time and portability, an automatic aircraft surface damage inspection method based on improved SSD was studied. Firstly, the characteristics of aircraft surface damage data are analyzed, the types of aircraft surface damage based on image feature are specified, a feasible aircraft surface damage sample standard is established. And then, based on basic SSD, combined with MobileNet and FPN, a multi-scale convolutional neural network model, SSD_MobileNet, is constructed to intend to perform the automatic aircraft surface damage inspection. Experimental results show that its mAP can reach 65.6%, FPS can reach 25.9, and only 25M of program size, and the automatic location and classification of aircraft</p>

SESSION 3

Room C: <https://us02web.zoom.us/j/83168271473>

October 22, 2022, Beijing Time

	surface damage can be realized. The designed inspection method achieves a better comprehensive performance in the inspection precision, speed and program size, and is suitable for deployment on portable aircraft maintenance equipment.
--	--

Session Topic: Image Analysis and Calculation

Time: 16:00-18:15

Session Chair: Associate Professor Juan Zhang, Shaanxi Normal University, China

<p>P0006 16:00-16:15</p>	<p>An lp-nonconvex regularization method for image smoothing Guoliang Zhu, Xiaoguang Lv, Xueman Sun, Biao Fang Presenter: Guoliang Zhu, Jiangsu Ocean University, China</p> <p>Abstract: Image smoothing techniques are widely used in computer vision and graphics applications, such as detail enhancement, artifact removal, image denoising and high dynamic range (HDR) tone mapping. In this paper, an ℓ_p-nonconvex minimization model is presented to achieve diverse smoothness of edges. To induce sparsity more strongly than the ℓ_1 norm regularization, we take the nonconvex arctangent penalty function of the image gradient as the regularization term. To make the model more flexible and effective, we use the ℓ_p norm function as the fidelity term. The majorization-minimization (MM) algorithm is employed for the proposed nonconvex optimization model. We discuss the convergence of the resulting MM algorithm. Comprehensive experiments and comparisons show that the proposed method is effective in a variety of image processing tasks.</p>
<p>P0014 16:15-16:30</p>	<p>Research on Realistic Effect Generation Algorithm of Rendering Images Based on GAN Runmin Gan, Hu Su Presenter: Runmin Gan, Southwest Jiaotong University, China</p> <p>Abstract: Saving algorithmic overhead for realistic effect generation is always a hot topic in graphics research. The algorithm in this paper applies the method of the Generative Adversarial Network (GAN) to the study of realistic computer graphics, proposing a new solution to reduce the algorithm overhead. We adopt conditional GAN and add an additional module before the generator to provide the lighting information of the current scene. This extra module represents the illuminating direction with grayscale gradient maps of different angles helping the network get lighting information through image input. In addition, we optimize the loss function by adding the L1 loss and the feature perception loss to improve the generative effect of our network. In the feature perception loss, we use a pre-trained VGG network to calculate the detail feature gap between images, to help the model generate images with better light and shadow effects. Our algorithm can add realistic effects to the existing coarse-rendered image according to the lighting conditions and obtain the corresponding fine-rendered image in single light scenes. The extensive experimental results show that our algorithm has a good post-processing effect on realistic rendering, and the time overhead of the algorithm is independent of the complexity of the scene. From the robustness test, we can know that our network also has a good generalization ability.</p>
<p>P0020</p>	<p>Combined Regional Homography-Affine Warp for Image Stitching Xinyi Li, Lin He, Xinguo He</p>

SESSION 4

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

16:30-16:45	<p>Presenter: Xinyi Li, South China University of Technology, China</p> <p>Abstract: Image stitching is the process of combining multiple images with narrow fields of view into a panoramic image with high resolution. Conventional global warp-based image stitching algorithm has limited alignment accuracy and causes shape distortion while spatially-varying warp-based ones have high computational complexity. To address these problems, we proposed a novel regional warp which adopts various transformation models to handle different areas of the image. Images can be divided into overlapping and non-overlapping regions based on the distribution of matched features. For the overlapping area, two kinds of projective transformation are adopted in combination to warp each pixel in this region. For the non-overlapping area, it is further partitioned into two regions where a projective transformation and an affine transformation are utilized separately. Experimental results show that the proposed warp not only provides good alignment accuracy but also avoids severe shape distortion.</p>
P0039 16:45-17:00	<p>Joint Generative Learning and Super-Resolution for Real-World CameraScreen Degradation Guanghao yin, Shouqian Sun, Chao li, Xin Min Presenter: Guanghao Yin, Zhejiang University, China</p> <p>Abstract: In real-world applications, there are many scenarios where people need to capture digital screens. While the quality of digital images captured by cameras and mobile phones is constantly being improved, taking high-resolution images of digital screens remains challenging. Except for the camera sensor, the display screen also involves more complicated degradations, such as noise, color distortion, etc. However, few studies of single image super-resolution (SISR) have focused on the camera-screen degradation. In this paper, we build the first camera-screen degraded dataset (Cam-ScreenSR), where HR images are original ground truths from the previous DIV2K dataset and corresponding LR images are camera-captured versions of HRs displayed on the screen. Moreover, we propose a joint two-stage model which consists of the downsampling degradation GAN(DD-GAN) and the dual residual channel attention network (DuRCAN). Specifically, DD-GAN first learns the real degradation to synthesize more various LR images, and then DuRCAN learns to recover the mixed real and synthetic LR images supervised with paired HR ground-truths. We also use a Laplacian loss to sharpen the high-frequency edges. Extensive experiments validate that our proposed method outperforms existing SOTA models in both synthetic and real-degraded datasets. Moreover, in real captured photographs, our model also delivers the best visual quality with sharper edge, fewer artifacts, and especially appropriate color enhancement, which has not been accomplished by previous methods.</p>
P1049 17:00-17:15	<p>A CABAC pre-coding based and Lossless Recompression Method for JPEG Images Zimin Jiang, Changcai Lai, Qinghua Sheng, Jie Jiang Presenter: Zimin Jiang, Hangzhou Dianzi University, China</p>

SESSION 4

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

	<p>Abstract: As the number of application of images on the Internet increases, how to store and transmit these images becomes a big challenge. JPEG as the most widely used image compression format on the Internet is often applied to pictures compression. However, just using JPEG alone to compress images is not enough now. In hence, some methods use improved entropy coding to further recompress JPEG images losslessly or process the images on DCT domain for lossy recompression. These methods are useful and work for various images. But there is no special design for fixed surveillance applications. Depending on the feature of images generated by a same fixed surveillance camera, a JPEG image lossless recompression method based on CABAC pre-coding, residual coefficients between JPEG image group and simplified context prediction is proposed by us. With a little reduction of decoding time as well as little increase of encoding time, average 27% bits saving can be achieved in the experiment.</p>
P1374 17:15-17:30	<p>Deep Non-Convex Low-Rank Subspace Clustering Weixuan Luo, Xi Zheng, Min Li Presenter: Weixuan Luo, Shenzhen University, China</p> <p>Abstract: Subspace clustering is an important research topic in computer vision. The traditional low-rank model has potentiality in clustering for images with the linear subspace assumption. However, for images that do not meet the assumption, the manipulation of approximating rank function with nuclear norm in the above methods always results in the over-penalization for large singular values. Therefore, a non-convex low-rank subspace clustering method involved deep learning is proposed in this paper, which realizes the extraction of non-linear features of the image by introducing convolutional self-encoding and self-representation layer. Specially, for the low-rank attribute of the data, the non-convex Schatten-p ($0 < p < 1$) norm is used to characterize the matrix rank, which can bring more accurate solution than the traditional low-rank methods. Using the decomposition formula of non-convex Schatten-p norm, this paper also gives the corresponding optimization algorithm. Numerous numerical experiments show that the combination of auto-encoder with self-representation layer and non-convex Schatten-p norm can enhance the subspace clustering ability of nonlinear data.</p>
P1376 17:30-17:45	<p>An Rendering Algorithm based on Real-image Conversion for Chinese Ink Painting Style Yang Zhao, Guowu Yuan, Hao Wu, Wenhua Qian Presenter: Yang Zhao, Yunnan Normal University, China</p> <p>Abstract: Chinese ink painting is an art with a long history in China, and the simulation of ink painting has become one of the most challenging topics in the field of computer graphics and digital image processing. This paper realizes a simulation algorithm of Chinese ink painting by constructing a hybrid two-dimensional fluid physical model. Firstly, we extract a set of coherent, smooth lines of the input image. The edge contours are used as the ink stroke drawn by the user, which is the initial parameter of ink diffusion. Secondly, we propose a three-layer model of ink particle migration and transport. The hydrodynamic theory is introduced to simulate the migration and transport of ink particles in shallow water layer, ink particle deposition layer and</p>

SESSION 4

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

	<p>capillary layer. Thirdly, we design a gray texture generation method of Chinese ink painting. By using this method, we can obtain a gray-scale background image simulating the diffusion of the ink, which would be fused to the ink diffusion image obtained through the hybrid physical model we designed to produce a final picture of the Chinese ink effect. Finally, an interactive digital Chinese ink painting system is also designed and implemented in this paper. Experiments show that the system can automatically generate ink painting effects with different degrees of abstraction and diffusion effects.</p>
P1413 17:45-18:00	<p>Preoperative Identification of Microvascular Invasion in Hepatocellular Carcinoma based on Multi-modal and Multi-response Convolutional Neural Network Xin Dai, Sue Cao, Yongze Guo, Wenqi Shi, Lina Zhang, Mengsi Li, Jingbiao Chen, Jin Wang and Yao Lu Presenter: Xin Dai, Sun Yat-sen University, China</p> <p>Abstract: Microvascular invasion (MVI) is a reliable predictor of the survival of patients with hepatocellular carcinoma (HCC). Accurate preoperative MVI assessment is essential to determine the appropriate surgical approach and management strategy to decrease the HCC recurrence rate. In this study, a preoperative evaluation method was proposed based on a convolutional neural network (CNN) model. Using Computed Tomography (CT) volume data, the relationship between CT volume data and MVI can be explored based on a multi-modal and multi-response CNN via an end-to-end model. A total of 400 patients were included in this study. First, the arterial phase (AP) and venous phase (VP) volume data were used as the inputs of the model; The size of the input was arbitrary and the inputs was converted to the same size by spatial pyramid pooling (SPP) behind. Then, these features were merged by the multi-modal network. The features of the AP and VP were combined through the multi-modal fusion of decision-making layers. Of the 400 patients, 215 (53.75%) and 185 (46.25%) are MVI-positive and MVI-negative cases, respectively. The areas under the receiver operating characteristic curves of the three-dimensional (3D) CNN model corresponding to the training and testing sets were 0.904 and 0.893, respectively. In the test set, 88.89% of the MVI-negative cases (16/18) and 86.37% of the MVI-positive cases (19/22) were detected. The evaluation results indicated a considerable potential feature correlation between CT volume data and MVI. The proposed multi-modal and multi-response CNN model had positive effect on the preoperative evaluation of MVI.</p>
P1907 18:00-18:15	<p>An improved slanted-edge method for measuring Modulation Transfer Function based on edge-preserving filter Saisai Nie, Xiukai Ruan, Guihua Cui, Weikuan Lu Presenter: Saisai Nie, Wenzhou University, China</p> <p>Abstract: ISO 12233 provides a method for measuring Modulation Transfer Function (MTF) of digital input devices based on a slanted-edge method. However, the results measured by the method are greatly affected by image processing procedures, especially sharpening, and false MTF is often given. In this article, an improved slanted-edge method is proposed where an edge-preserving filter with optimal parameters is introduced to smooth the image before implementing ISO 12233 method. Particularly, three edge-preserving filters, bilateral filter, guided filter and adaptive weighted least square filter,</p>

SESSION 4

Room A: <https://us02web.zoom.us/j/84697177419>

October 22, 2022, Beijing Time

and one general filter, Gaussian filter, are compared in mitigating the effect of sharpening algorithm on the measurement of digital camera MTF. The experimental results indicate that the effect of sharpening algorithm on the measured MTF can be effectively reduced by using an edge-preserving filter, and significantly improvement on the MTF accuracy at Nyquist frequency is achieved. Especially, the slanted-edge method based on the guided filter and the weighted least squares filter can give more real MTF for a digital camera.

Session Topic: Image Classification

Time: 16:00-18:15

Session Chair: Professor Yao Lu, Sun Yat-sen University, China

<p>P0011 16:00-16:15</p>	<p>A lightweight convolution network with self-knowledge distillation for hyperspectral image classification Hao Xu, Guo Cao, Lindiao Deng, Lanwei Ding, Ling Xu, Qikun Pan, Yanfeng Shang Presenter: Hao Xu, Nanjing University of Science and Technology, China</p> <p>Abstract: Recently, using convolutional neural networks (CNNs) to extract spectral-spatial features has become an effective way for HSI classification. However, complex CNN models require many training parameters and floating-point operations (FLOPs). This usually means longer training and testing times. Furthermore, deep networks become prone to overfitting when the labeled samples of hyperspectral data are limited. In this article, a lightweight convolution network with self-knowledge distillation (SKDLCN) is developed for HSI classification, and it has two crucial elements, including a dual-path convolution network and a self-knowledge distillation module. At first, a method called 3-D transformation is performed for data augmentation to alleviate the overfitting problem. Then, the proposed network consists of small 1×1 convolutions with a residual path and a density path. Specifically, it can efficiently complete the extraction of spectral and spectral-spatial features sequentially from HSI. Self-knowledge distillation can be explained within the knowledge distillation framework as students become teachers, which gradually extracts knowledge of the model itself during training. Specifically, the target is adaptively adjusted by combining the ground truth of the model itself and past predictions. Experiments on two public HSI datasets demonstrate that the proposed method is significantly superior to some state-of-the-art methods with limited training samples.</p>
<p>G5001 16:15-16:30</p>	<p>Embedding BN Layers into AlexNet for Remote Sensing Scene Image Classification Dongfu Dai, Weiheng Xu, Shaodong Huang Presenter: Dongfu Dai, Southwest Forestry University, China</p> <p>Abstract: To improve model convergence speed and accuracy of the AlexNet model on high-resolution remote sensing scene image classification, the Batch Normalization (BN) layer were used to replace the Local Response Normalization (LRN) to normalize the features of each channel in the convolution layers. In addition, we replaced the filling method of all layers in the AlexNet model with the "SAME" method to reduce the loss of image edge information in convolution. Moreover, we add dropout strategy after each pooling layer to prevent model overfitting. Finally, three remote sensing scene datasets including NWPU-RESISC45, AID, and UCM were used for accuracy and convergence speed verification. The overall accuracies(OA) of our improved model were 96.10%, 96.80%, and 97.14% of on the three datasets, respectively, which were 14.19%, 13.68%, and 10.47% higher than those of AlexNet, respectively. Meanwhile, compare with other models, this study model has higher OA for remote sensing scene image classification. Therefore,</p>

	the improved model of this study can accurately identify scene categories.
P0012 16:30-16:45	<p>Multi-scale Attention-based Few-shot Hyperspectral Images Classification Lanwei Ding, Guo Cao*, Ling Xu, Lindiao Deng, Hao Xu, Qikun Pan, Yanfeng Shang Presenter: Lanwei Ding, Nanjing University of Science and Technology, China</p> <p>Abstract: In recent years, deep learning-based hyperspectral image classification techniques have developed rapidly. Many effective deep learning models have been proposed in academia, such as 3D-CNN and some other CNN-based methods, which have achieved high accuracy in hyperspectral image classification. These excellent methods rely on large number of labeled samples for their effectiveness. In practice, labeling pixels of hyperspectral images is expensive (time-consuming and labor-intensive), so it is often difficult to obtain enough labeled samples for training deep neural network models. To address this problem, we propose a multiscale attention-based few-shot learning (MAFSL) method using only a few labeled samples for each category in this paper. First, few-shot learning is performed on mini-ImageNet to obtain prior knowledge, and then the knowledge is transferred to the hyperspectral dataset. Before embedding features, multiscale attention-based feature extraction with reconstruction loss is applied to the hyperspectral image. Then, the obtained features are input into the spatial feature extraction network and the spectral extraction network, respectively. Finally, the embedded features are put into the metric space for classification. The proposed model can get a higher classification accuracy because the extracted features have less correlation with each other. Experimental results show that our MAFSL outperforms many existing supervised learning methods when only a small number of labeled samples are used.</p>
P0026 16:45-17:00	<p>A multi-view feature decomposition deep learning method for lung cancer histology classification Heng Gao, Minghui Wang, Haichun Li, Zhaodi Liu, Wei Liang and Ao Li. Presenter: Heng Gao, University of Science and Technology of China, China</p> <p>Abstract: Accurate classification of squamous cell carcinoma (SCC) and adenocarcinoma (ADC) using computed tomography (CT) images is of great significance to guide treatment for patients with non-small cell lung cancer (NSCLC). Although existing deep learning methods have made promising progress in this area, they do not fully exploit tumor information to learn discriminative representations. In this study, we propose a multi-view feature decomposition deep learning method for lung cancer histology classification. Different from existing multi-view methods that directly fuse features extracted from different views, we propose a feature decomposition module (FDM) to decompose the features of axial, coronal and sagittal views into common and specific features through an attention mechanism. To constrain this feature decomposition, a feature similarity loss is introduced to make common features obtained from different views to be similar to each other. Moreover, to assure the effectiveness of feature decomposition, we design a cross-reconstruction loss which enforces each view to be reconstructed according to its own specific feature and other view's common features. After the above feature decomposition, comprehensive representations of tumors can be obtained by efficiently integrating common features to improve the</p>

	classification performance. Experimental results demonstrate that our method outperforms other state-of-the-art methods.
P0034 17:00-17:15	<p>Hyperspectral image classification based on spectral spatial feature extraction and deep rotation forest ensemble with AdaBoost Lindiao Deng, Guo Cao, Ling Xu, Hao Xu, Qikun Pan, Lanwei Ding, Yanfeng Shang Presenter: Lindiao Deng, Nanjing University of Science and Technology, China</p> <p>Abstract: In recent years, deep learning methods have been widely applied to hyperspectral image classification. However, these deep learning methods need lots of training samples to tune abundant parameters which induce a heavy computation burden. In this paper, we propose a classification model based on spectral spatial feature extraction and deep rotation forest ensemble with AdaBoost (SSDRA). First, linear discriminant analysis (LDA) and extended morphological attribute profile (EMAP) are used to extract features from hyperspectral images. In this way, the useful features of hyperspectral images can be integrated to a great extent while reducing the dimension of hyperspectral images. Then, the features of joint regions combining patches and superpixels are input into the classification model for training. Next, a deep rotation forest ensemble with AdaBoost (DRA) is designed for classification, so that our method can achieve superior performance with a small number of training samples. Finally, to optimize the classification results, superpixel smoothing is performed. The final results are obtained by using majority voting on the classification results within superpixels and among superpixels of different scales. To verify the effectiveness of the proposed method, experiments are performed using two public hyperspectral datasets. The experimental results demonstrate that the proposed method achieves satisfactory classification results.</p>
P1046 17:15-17:30	<p>Hyperspectral Image Classification Aided by LiDAR Data Zheng Deng, Genping Zhao, Shihui Zhao, Li Wang, Zhuowei Wang, Heng Wu, Lianglun Cheng Presenter: Zheng Deng, Guangdong University of Technology, China</p> <p>Abstract: Spectral variability and shadow effects can limit the hyperspectral image (HSI) classification performance. Compared with HSI, the LiDAR data is an excellent complement with its abundant elevation information. In this study, a procedure including pre-processing, deep residual network classification and post-processing is investigated for classification of HSI aided by the LiDAR data to release the problem of identifying shaded objects and spectral variability. Specifically, three aspects with respect to spectral band selection using Archetypal Analysis (AA), feature level fusion based classification by deploying a residual network associated approach and label correction utilizing the elevation information, are explored to realize more accurate classification. Experiments on three public multi-source (hyperspectral and LiDAR) remote sensing datasets show more promising classification can be achieved via fusion two-source of remote sensing data than that using only independent hyperspectral image. In particular, on the Houston 2017 dataset, OA and Kappa achieved significant gains of 2.47% and 2.79% respectively after incorporating LiDAR information. Moreover, the results demonstrate the elevation information used independently in the post-processing stage can help with effective refinement of classification results.</p>

<p>P1383 17:30-17:45</p>	<p>Multiscale Semantic Alignment Graph Convolution Network for Single-Shot Learning based Hyperspectral Image Classification Haojie Gong, Abdolraheem Khader, Ghulam Farooque and Liang Xiao, Presenter: Haojie Gong, Nanjing University of Science and Technology, China</p> <p>Abstract: Graph convolutional networks (GCN) can extract features from non-Euclidean space very effectively, and it has been successfully applied in various fields of hyperspectral images (HSIs). However, due to the limited labeled HSI data, GCN often performs not well and encounters over-smoothing problems as the number of network layers increases. Furthermore, building a GCN adjacency matrix for HSI classification directly is computationally complex. This paper proposes a multiscale semantic alignment graph convolutional network (MSAGCN) for HSI classification to solve the problems mentioned above. The proposed method mainly consists of three parts, superpixel segmentation, semantic alignment and multiscale graphs. Firstly, superpixel segmentation is performed on the original HSI, and each superpixel region contains similar spatial and spectral information. Secondly, semantic features of labeled nodes are extracted using identity aggregation with fixed receptive fields. The class-center similarity is adopted using these semantic features to align nodes semantically. This semantic alignment technique alleviates over-smoothing in GCN. Finally, the multiscale technique enables MSAGCN to obtain different scale spectral-spatial features. Experimental results show that our proposed model exhibits competitive results on open source hyperspectral datasets with only one labeled sample per class.</p>
<p>P1411 17:45-18:00</p>	<p>Fastformer: Transformer-Based Fast Reasoning Framework Wenjuan Zhu, Ling Guo, Tianxiang Zhang, Feng Han, Yi Wei, Xiaoqing Gong, Pengfei Xu, Jing Guo Presenter: Wenjuan Zhu, Northwest University, China</p> <p>Abstract: Action recognition is one of the main tasks of computer vision. However, due to the large amount of video data generated a large amount of information in the process of depth computation, and there is a large amount of information redundancy, which does not play a role in the process of action recognition, but reduces the speed of model inference. In order to solve this problem, most of the existing methods improve the accuracy at the cost of speed. In this paper, we propose Fastformer: a transformer-based structure for fast inference video classification to further improve model inference speed while maintaining accuracy. In order to achieve the balance of speed and accuracy, we solve the inter-frame and intra-frame redundancy of video and design a new self-attention network, which uses the improved highway network to make the model realize the same function as the traditional self-attention module, while greatly reducing the amount of calculation and the number of required parameters. We conduct a large number of experiments to verify the effect of our model. Overall, Fastformer significantly outperforms existing vision transformers with regard to the speed versus accuracy trade-off. For example, at 76.4% Keyframes-400 top-1 accuracy, Fastformer is 28% faster than TimeSformer.</p>

SESSION 5

Room B: <https://us02web.zoom.us/j/87279824425>

October 22, 2022, Beijing Time

P1901
18:00-18:15

Classification of Breast Cancer Pathological Images Combining Fine-Grained Region Location

MA Shi-hao YANG Bei TIAN Gui-lan LI Xiao-yu

Presenter: MA Shi-hao, Zhengzhou University, China

Abstract: Classification algorithm based on deep learning is the main technology for computer-aided intelligent diagnosis of breast pathological images. The existing deep learning algorithms rarely pay attention to multi-scale information in breast cancer pathological images, and can not extract key regions from pathological images for auxiliary diagnosis. To address this issue, this paper proposes a novel method for classifying breast cancer pathological images which incorporates a fine-grained region location mechanism. This method is realized by a dual-branch architecture with global network and local network. Global and local features are extracted from the whole image and local key image, respectively. The final classification results can be obtained by integrating both the global and local network analysis. In order to locate the most predictive area in the whole image, this paper designs utilizes a fine-grained region localization mechanism and combines the above two branch networks. Extensive experiments on the BreakHis data set are conducted to verify the effectiveness of the proposed algorithm. The empirical results show that this method improves the classification accuracy by comparing the performance with that of several typical convolutional neural network and state-of-the-art algorithms.

Session Topic: Image Enhancement and Denoising

Time: 16:00-18:15

<p>P0029 16:00-16:15</p>	<p>Object Detection in Infrared Images Using Modified YOLOv4 Models and an Image Enhancement Module Dan Wang, Huiqian Du*, Zhifeng Ma Presenter: Dan Wang, Beijing Institute of Technology, China</p> <p>Abstract: Deep learning-based object detection approaches have shown excellent performance in RGB images. However, when used to detect objects from infrared images, the accuracy may reduce significantly due to low contrast, obscure textures and strong noise of infrared images. To alleviate the problem, we design a detail enhancement module involving spatial attention mechanism to enhance the textures and details of images. The output of the proposed module is fed into modified YOLOv4. We introduce Alpha-IoU loss and Weighted-NMS to YOLOv4 to enhance geometric factors in both bounding box regression and Non-Maximum Suppression, leading to notable gains of average precision. The experiment results show that compared with YOLOv4, mAP0.5 and mAP0.5:0.95 of our model are improved by 1.1% and 3.5% respectively, effectively improving the detection accuracy.</p>
<p>P1045 16:15-16:30</p>	<p>Densely Residual Attention based Low-Light Image Enhancement Sami Ul Rehman, Ammar Hawbani, Xingfu Wang, Muhammad Hamza, Liang Zhao, Saeed H. Alsamhi, Majjed Al-Qatf Presenter: Sami Ul Rehman, University of Science and Technology of China, China</p> <p>Abstract: The visual quality of nighttime photographs diminishes greatly due to low contrast and high noise. We need a robust image enhancement methodology to improve such low-light images close to standard daylight images. Due to deteriorated conditions of uneven light and noise, this image enhancement problem becomes ill-posed. Our paper has proposed a Densely Residual Attention Network (DRANet), an end-to-end attention base densely residual network. The architecture of DRANet consists of sub-modules convolution block (CB) and densely residual feature - convolutional block attention module (DRF-CBAM). DRF-CBAM also has sub-modules, deep residual feature block (DRFB), and convolutional block attention module (CBAM). Using the most recent results from attention and deep residual-based convolution networks in number of computer vision problems, we have used DRFB to enhance the features in-depth by using its dense and residual skip connections. Similarly, features in both spatial and channel axis have been extracted by using a lightweight CBAM attention module. Contrast, luminosity, and noise of the enhanced images have been balanced by additionally implementing a color balancing function at the end of the proposed network. Furthermore, we have used a combination of LLab, LSSIM and LMAE loss functions to make the proposed network stable and recover both contextual and local details while training. MAE, PSNR, SSIM, MS-SSIM, FSIM, Cosine Similarity, and deltaE2000 have been used as referenced while NIQE as non-referenced based image quality assessment (IQA) metric. Experiment results showed that our proposed methodology is very effective with higher referenced and lower non-referenced image IQA metrics values. Furthermore, the effectiveness</p>

SESSION 6

Room C: <https://us02web.zoom.us/j/83168271473>

October 22, 2022, Beijing Time

	<p>of our method has also demonstrated by the visual and perceptual quality of enhanced images.</p>
P1048 16:30-16:45	<p>Complementary Features-aware Attentive Multi-Adapter Network for Hyperspectral Videos Tracking Shaoqi Ma, Abdolraheem Khader, Liang Xiao Presenter: Shaoqi Ma, Nanjing University of Science and Technology, China</p> <p>Abstract: Hyperspectral object tracking aims to estimate the bounding box for the given target using hyperspectral data. Different from traditional color videos, hyperspectral videos have more abundant band information for their capacity to capture the reflectance spectrum of the target at a wider range of wavelengths provides important capabilities and opportunities, which provides new capabilities for discriminating targets in complex scenes, but also presents new challenges. The limited dataset and the high dimensionality of hyperspectral data are two new challenges in constructing hyperspectral trackers, resulting in existing hyperspectral tracking methods based mainly on correlation filters. This paper proposes a new Complementary Features-aware Attentive Multi-Adapter Network (CFA-MANet), which can train a neural network well and achieve high performance for Hyperspectral Object tracking just using the limited dataset. Specifically, we add a complementary features-aware module to the multi-adapter network, which employs two different strategies to reduce the dimensionality of hyperspectral data from two complementary perspectives, and the joint implementation of these two strategies results in a reduction in the amount of computed data and parameters of the designed neural network while achieving competitive results. Moreover, spatial and channel attention modules are used to learn a wider range of contexts and improve the representation of different semantic features, respectively. Cross-attention is used to learn complementary information and thus generate more discriminative representations. Experimental results on hyperspectral datasets show that our method achieves the best results compared to several recent hyperspectral tracking methods.</p>
P1057 16:45-17:00	<p>Low-Rank and Spectral-Spatial Variation Regularized Hyperspectral Image Denoising Algorithm Yanhui Liu, Weiguo Wang Presenter: Yanhui Liu, Ocean University of China, China</p> <p>Abstract: Hyperspectral image (HSI) is an important carrier for recording, transmitting, and storing information. Denoising is an indispensable step in HSI processing, which is the process of restoring noisy images to high-quality images that can reflect the objective world. It is crucial to establish a reasonable denoising method. Most recent studies regard HSI as a three-dimensional tensor and then establish the low-rank method, or use the spatial information of the image to establish a total variation method. In this paper, we combine the spectral-spatial structure and low-rank properties of HSI to construct the regularization term and propose a low-rank spectral-spatial adaptive total variation (LRSSAHTV) model. Then we define a separable soft thresholding (SST) operator to minimize the spectral-spatial adaptive total variation (SSAHTV) regularization problem, which is the basic theory of our algorithm. A linearized alternating direction method of multipliers (LADMM) algorithm is proposed to solve the composite model. To verify the effectiveness of our method, we perform the simulated noisy HSI and the real polluted HSI denoising experiments in section 4. We also compare the</p>

SESSION 6

Room C: <https://us02web.zoom.us/j/83168271473>

October 22, 2022, Beijing Time

	denoising effect of the proposed method with the total variation method and low-rank method, which shows that our method can obtain the most precise restored image.
P1066 17:00-17:15	<p>A Non-local Image Denoising Method Based on TV-L1 with Variable Exponents Hai Geng, Zhixin Gu, Lingyan Weng, Jian Yu Presenter: Hai Geng, Nanjing Institute of Tourism and Hospitality, China</p> <p>Abstract: Among various image denoising methods, the total variation (TV) has offered considerable performance. However, it is prone to the staircasing effect. To alleviate this problem, we propose a novel TV-based denoising method that utilizes the variable exponents as the regularization constraints and the L1-norm as the fidelity term. To further improve the robustness of our method, we develop it under a non-local framework to avoid overlooking the global information of the image. Extensive experiments on real-world images have validated the effectiveness of the proposed method.</p>
P1371 17:15-17:30	<p>V-channel adaptive defogging with low illumination images based on optimized retinex model Mincong Chen; Yawen Pan Presenter: Mincong Chen, Nanjing institute of technology, China</p> <p>Abstract: The defogging effects of different retinex algorithms, including single-scale retinex (SSR), multiscale retinex (MSR) and multiscale retinex with color restoration (MSRCR), are compared in this paper. It was found that some images treated by the above methods were dark. This phenomenon is more obvious when processing foggy images with low light. Contrast limited adaptive histogram equalization (CLAHE) is applied to improve the image brightness and contrast. Further analysis of the V-channel in HSV space shows that when the normalized histogram distribution of the V-channel is concentrated below 0.5 and the image has the component of the highlighted region, the images are dark after processing by the traditional retinex algorithms. Based on this, the V-channel adaptive enhancement method is proposed to improve the overall image brightness. The experiments show that the proposed algorithm works better when combined with both the modified MSR algorithm and CLAHE. The overall brightness of the image is improved, and the information entropy of the image is also increased.</p>
P1405 17:30-17:45	<p>An improved OCT retinal image denoising algorithm based on variational image decomposition Zhuo Li, Biyuan Li, Jun Zhang, Jianqiang Mei, Xinchun Zhao, Binghui Li Presenter: ZhuoLi, Tianjin University of Technology and Education, China</p> <p>Abstract: Optical Coherence Tomography (OCT) has become an important auxiliary diagnostic technology in the field of fundus disease detection due to its advantages of high resolution and high penetration depth. However, speckle noise exists in retinal images acquired by OCT. It is a difficult problem for</p>

SESSION 6

Room C: <https://us02web.zoom.us/j/83168271473>

October 22, 2022, Beijing Time

	<p>OCT image processing technology to keep the middle layer structure information in the process of OCT image denoising. A novel OCT retinal image denoising method based on Gaussian mixture model and variational image decomposition is proposed. Firstly, the proposed BL-G-BM3D variational image decomposition model is used to initially denoise OCT images, and then the Gaussian mixture model is used to cluster the initial denoising results to obtain binary masks that can distinguish background and structure. Finally, the final denoising results are obtained by multiplying the mask image with the initial denoising results. An OCT retinal image with high noise and low contrast was tested and compared with five commonly used denoising methods. The results show that the proposed method can achieve both de-noising effect and laminar structure preservation for high-noise OCT retinal images.</p>
P1410 17:45-18:00	<p>A Lightweight Convolutional Network Based on Pruning Algorithm for YOLO Guanyu Liu, Yuzhao Li, Yuanchen Song, Yumeng Liu, Xiaofeng Xu, Ruiheng Zhang Presenter: Ruiheng Zhang, Beijing Institute of Technology, China</p> <p>Abstract: With the rapid development of deep learning, neural network models have become increasingly complicated, leading to larger storage space requirements and slower reasoning speed. These factors make it difficult to be deployed on resource-limited platforms. To alleviate this problem, network pruning, an effective model compression method, is commonly performed in a deep neural network. However, traditional pruning methods simply set redundant weights to zero, thus failing to achieve the acceleration effect. In this paper, a channel-wise model scaling method is proposed to reduce the model size and speed up reasoning by structurally removing the redundant filters in convolutional layers. To make the residual block more sparse, we develop a pruning method for residual cells. Experimental results on the YOLOv3 detector show that our proposed approach achieves a 70.6% parameter compression ratio without compromising accuracy.</p>
P1427 18:00-18:15	<p>An Adaptive Mean Denoising Algorithm for Pepper and Salt Noised Image Fengqing Qin, Liang Yuan, Lilan Cao, Lihong Zhu, Chaorong Li, Yilan Xue Presenter: Fengqing Qin, Yibin University, China</p> <p>Abstract: To improve the quality of the “pepper & salt” noised image, an adaptive mean filter algorithm is proposed. Firstly, the appropriate filter window size is automatically selected according to the intensity of noise in the image. Secondly, whether each pixel is a “pepper & salt” noise point is detected by using minimum-maximum principle. The noised point outputs the mean gray value of the unnoised pixels in the current window. The non-noised point outputs the original value. Thirdly, the denoising effect of the three algorithms was compared by simulation experiments, and the influence of different “pepper & salt” noise density on denoising effect was tested. From the objective and subjective judgement, the propose method achieves higher PSNR and better visual effect.</p>

Session Topic: Image reconstruction and 3D model

Time: 10:15-12:30

Session Chair: Professor Lifang Wu, Beijing University of Technology, China

P0016 10:15-10:30	<p>Multi-Scale Transformer 3D Plane Recovery Fei Ren, Qingling Chang, Liu Xinglin and Yan Cui Presenter: Fei Ren, Wuyi University, China</p> <p>Abstract: The 3D plane recovery and reconstruction method of a single image aim to identify plane instance regions and estimate plane parameters and is widely used in technologies such as AR/VR(Augmented Reality/Virtual Reality). In recent years, significant progress has been achieved in single-image 3D plane recovery approaches based on deep learning, however much of the research has focused on overall plane segmentation performance rather than the accuracy of small-scale plane segmentation. Given the lack of an accurate definition of the segmentation of small-scale plane regions in existing methods, we propose a novel multi-scale transformer-based plane recognition and recovery model, which can accurately identify the edges of small-scale plane regions. The network branches composed of the model are used to detect plane and non-planar areas respectively. Different input feature scales make the two network branches have different global feature extraction capabilities. Finally, the two branches are strengthened to recognize plane areas in the same scene through mutual information loss. The consistency makes the two branch networks have the ability to share parameters. The experimental results in the Scannet and NYU V2 datasets show that the model can accurately identify small-sized plane areas, and the detection and recovery accuracy reaches the state-of-the-art effect.</p>
P0028 10:30-10:45	<p>Fast System Matrix Iterative Computation Algorithm for PET Image Reconstruction Yibing Li, Huabin Wang, Dailei Zhang, Yonglin Chen , Xuejun Li Presenter: Yibing Li, Anhui University, China</p> <p>Abstract: In positron emission tomography (PET) studies, iterative method is usually used for iterative reconstruction of PET data, in which the system matrix reflects the mapping between image space and projection space, which is the key of iterative reconstruction algorithm. The previous orthogonal distance ray tracing algorithm is computationally complex and inefficient. To improve its computational speed and imaging quality, we propose a new algorithm. Firstly, incremental thinking was introduced on the basis of Siddon algorithm to directly solve the neighborhood where the current voxel and the upper voxel do not repeat, and accelerate the calculation of voxel coordinate index. Secondly, the distance between the neighborhood voxel and LOR line was iteratively solved based on the distance between the voxel and LOR line, which further improved the calculation speed. Finally, the probability value of the voxel which is completely covered by the detector is set as a constant value, while the probability value of other voxels decreases with the</p>

SESSION 7

Room A: <https://us02web.zoom.us/j/84697177419>

October 23, 2022, Beijing Time

	<p>distance from LOR line, which improves the imaging quality of the algorithm. A large number of evaluation experiments were performed on the resolution prosthesis model and the line-derived prosthesis model to verify the effectiveness of our method.</p>
P1042 10:45-11:00	<p>Global accurate multi-view point cloud registration based on manifold clustering and thermal gradient method Hui Chen, Yibo Wang, Heping Huang and Muhammad Ilyas Menhas Presenter: Hui Chen, Shanghai University of Electric Power, China</p> <p>Abstract: Distances are the basic measures used in data clustering. The selection of the distance has a significant impact on the final result. Euclidean distances do not accurately reflect the actual geometrical properties of complex shape and structure, while the geodesic distance may give full consideration to their geometry. Therefore, in the registration process, geodesic distance is more appropriate in such cases. In this study, a multi-view point cloud registration method is proposed to solve the problem of low registration accuracy of objects with complex shape and structure. The method transforms the registration problem into a clustering problem, and the rigid transformation is updated while updating the centroids. Clustering and rigid transformation are alternately applied to the point cloud to achieve final registration results. The proposed method is validated by experiments using data from Stanford University's public datasets. Experimental results show better accuracy and robustness of the proposed method.</p>
P1053 11:00-11:15	<p>A method for extracting flash point cloud feature on the surface of wind turbine blade based on difference of normal vector Yangzhi Chen, Lvwen Huang, Wang Liu Presenter: Lvwen Huang, South China University of Technology, China</p> <p>Abstract: The surface automatic processing of large wind turbine blades has always been an industry problem. The RGBD camera can be used to scan the blade surface to obtain a 3D point cloud of complex curved surfaces. However, the point cloud contains a lot of invalid information and is unorganized data, which makes it difficult to segment the point cloud of the blade. In this paper, we propose a method based on difference of normal vector to extract the blade flash feature of wind turbine blade. Firstly, the point cloud obtained by the RGBD camera is filtered to obtain a pure wind turbine blade point cloud. Then, the difference of normal vector of the wind turbine blade point cloud is calculated to separate the flash feature from the blade surface. Finally, a pure flash point cloud is segmented by clustering, which provides guidance data for automatic grinding of flash. Through experiments, it is proved that this method can accurately segment the blade flash feature on huge wind turbine blade without a large amount of data for model training. And the separation effect is not affected by the scanning angle, sampling noise, etc. It has strong robustness.</p>
P1375 11:15-11:30	<p>End-to-End Mesh Reconstruction from Partial Point Cloud based on Continuous Implicit Function Jiawei Yu, Xiaoshui Huang, Tao Chen, Yazhou Yao, Qiong Wang Presenter: JiaWei Yu, Nanjing University of Science and Technology, China</p>

	<p>Abstract: The existing mesh reconstruction methods rely on the complete point cloud input. However, the point cloud collected in the actual environment is always the partial point cloud of an object/scene. The direct mesh reconstruction from the partial point cloud is a challenging problem but there is currently no end-to-end method for mesh reconstruction from partial point clouds. To solve this challenge, this paper proposes an end-to-end mesh reconstruction from the partial point cloud method based on continuous implicit function. Specifically, the continuous implicit function of a complete point cloud is learned by combining the local and global features of a partial point cloud. Then we sample a point cloud from the continuous implicit function and reconstruct the mesh by leveraging the deformation network. A loss function based on the point cloud normal vector is proposed to further optimize mesh reconstruction. Experiments on the ShapeNet-55 dataset show that the loss of our method in three different incomplete degrees of the point cloud is reduced by 35%, 38%, and 38% respectively compared with NMF, with an average reduction of 38%.</p>
<p>P1390 11:30-11:45</p>	<p>SCGRNet: Shape Completion-guided Registration Network for Face Point Cloud Zeyu Jia, Yichao Zhou, Xiyuan Hu, Zhenmin Tang Presenter: Zeyu Jia, Nanjing University of Science and Technology, China</p> <p>Abstract: Point cloud registration is a fundamental problem in computer vision and robotics, and has been widely used in various applications, including 3D reconstruction, simultaneous localization and mapping, and autonomous driving. Over the last decades, numerous researchers have devoted themselves to tackling this challenging problem. With the success of deep learning in geometric vision tasks, various types of deep learning-based point cloud registration methods have been proposed, and some of them exhibit good performance. However, their performance degrades rapidly when the overlap between point clouds is reduced. In this paper, we propose a shape completion-guided registration network for face point cloud registration, which takes face shape as a prior knowledge to guide the network to learn a complete face features at the low overlap. Another noteworthy point is that we use the Regularized Projective Manifold Gradient (RPMG) layer to bridge the gap between the neural network output measured in Euclidean space and the rotation parameters described in the $SO(3)$ manifold. In addition, our proposed method infers the translation parameters from the two complementary point clouds and the predicted rotation rather than estimates them directly. For the problem of low overlap point cloud registration, our method can effectively improve the registration accuracy compared to the current state-of-the-art methods. The rotation error and translation error are, respectively, less than 2 degrees and 2mm on the FaceScape-based low overlap face registration dataset.</p>
<p>P1416 11:45-12:00</p>	<p>Active scene reconstruction via self-rotation driven by optimized information theory Yao Chen, Chenyang Zhu Presenter: Yao Chen, National University of Defense Technology, China</p> <p>Abstract: With the increased demands of digital twins and virtual reality, autonomous scene reconstruction has gradually become an essential direction</p>

	<p>in the field of 3D vision. However, efficient and accurate reconstruction with RGBD sensors remains challenging due to the high nonlinearity of camera pose optimization caused by the complexity of the exploration trajectory. Inspired by the working principle of Lidar and panorama, we introduce self-rotation for robust reconstruction. This design can reduce the optimization freedom and expand the visible range of the robot in a specific area, which can significantly improve reconstruction accuracy. To ensure accuracy and efficiency simultaneously, we present a two-branch framework based on the information theory to drive the robot for active scene reconstruction. The scanning branch finds the optimal goal position for self-rotation scanning, and the exploration branch is optimized to find an efficient exploration path which helps the scanning branch find more global optimal. Many experimental results show that our method effectively improves the quality and efficiency of scene reconstruction compared to the current SOTA method [1] (the average integrity is from 47.28% to 84.96%, and the average time consuming is from 315.4s to 189s).</p>
<p>P1426 12:00-12:15</p>	<p>A Systematic Registration Method for Cross-source Point Clouds Based on Cross-view Image Matching Guanghan Chu, Dazhao Fan, Ming Li and Haijun Zhang Presenter: Guanghan Chu, PLA Strategic Support Force Information Engineering University, China</p> <p>Abstract: Refined 3D model reconstruction of wide-area cities usually requires registration of multi-source data collected by different platforms and various sensors. Few studies discuss the problem of registration from cross-source image point clouds. This registration task is challenging due to the large variation in the density of point clouds generated from images of different resolutions, the extremely large view differences, the uncertain scale differences of point clouds in arbitrary coordinate systems, and the noise points caused by the low image quality. In this study, we propose a robust point cloud registration method based on cross-view image matching to solve these problems mentioned above. Firstly, the method uses the deep learning cross-view image matching algorithm to obtain 2D image matching points. They are then mapped to 3D space using depth information. Secondly, the dual quaternion is introduced to solve the spatial transformation model. Finally, the ICP fine-registration algorithm is used for optimization. To analyze the performance of the proposed method, experiments are tested on a public dataset in Dortmund, Germany. The experimental results show that the proposed method is not only able to overcome large coordinate system scale differences but is also immune to noise points and outliers. Compared with other point cloud registration methods, it greatly improves the efficiency and accuracy.</p>
<p>P1429 12:15-12:30</p>	<p>Improving Reference-driven Undersampled MRI Reconstruction via Iterative Data Correction Guisong Wang and Xiaofeng Du Presenter: Guisong Wang, Xiamen University of Technology, China</p> <p>Abstract: When large training datasets are unavailable in real clinical scenarios, researchers turn to unsupervised learning for under-sampled magnetic resonance image reconstruction. However, unsupervised learning methods suffer from insufficient a priori knowledge. We introduce self-consistency constraint with the calibration and acquisition data to tackle these issues. Specifically, we propose an iterative data correction operator to ensure high</p>

SESSION 7

Room A: <https://us02web.zoom.us/j/84697177419>

October 23, 2022, Beijing Time

	fidelity of the reconstructed MR data. Experiments shows that the method is flexible and can reconstruct data from arbitrary k-space sampling patterns and easily incorporates additional image priors.
--	---

Session Topic: Intelligent detection technology and algorithm

Time: 10:15-12:30

<p>P0001 10:15-10:30</p>	<p>Vehicle Detection in the Infra-red Thermal Images based on a Sparse Neural Network Xunxun Zhang, Lang Peng, Xiaoyu Lu Presenter: Xunxun Zhang, Xi'an University of Architecture and Technology, China</p> <p>Abstract: Recently, infra-red thermal images have drawn wide publicity for vehicle detection in traffic monitoring because of their resistances to insufficient light and hidden occasions. However, there are two significant challenges compared with vehicle detection in visible images. One is the insufficiency of the labelled infra-red thermal samples, and the other is less texture information. These two points lead to fewer parameters for the vehicle detection network in the infra-red thermal images. Therefore, to conquer these difficulties in infra-red thermal images, we construct a sparse neural network based vehicle detection method with sparse parameters (weights) and high generalization ability. Therein, to avoid over-fitting in the training stage, we eliminate some redundant parameters via constraints and on-line pruning. Finally, to verify the proposed vehicle detection method in infra-red thermal images, we introduce VIVID, VOT2019, and RGBT234 data sets. The vehicle detection results show that the proposed method presents superior results compared to several mainstream approaches.</p>
<p>P0024 10:30-10:45</p>	<p>Research on Face Recognition Based on Fusion Detection and Pose Estimation Rong Jiang Presenter: Rong Jiang, Weihai Vocational College, China</p> <p>Abstract: This paper discusses the research of face recognition based on fusion detection and pose estimation. The auxiliary network is added in face detection, and the SSD based on face pose prediction method is used in pose estimation to improve the accuracy and speed of face recognition. The generated lightweight model is deployed to the embedded system. The experimental results show that it is feasible.</p>
<p>P1043 10:45-11:00</p>	<p>Crowded People Detection for Occluded Classroom Surveillance Scenes Based on Relation Model Yuping Li, Fengqin Yao, Xuan Guo and Shengke Wang Presenter: Yuping Li, Ocean University of China, China</p> <p>Abstract: People detection is an important task in video surveillance. Due to the people's similar characteristics and occlusion, crowded people detection for occluded classroom surveillance scenes is challenging. In this paper, a new detection framework based on the relation model method is proposed to detect crowded people in occluded classroom surveillance scenes. Our method is mainly to predict a box set of related objects and then use the positive</p>

SESSION 8

Room B: <https://us02web.zoom.us/j/87279824425>

October 23, 2022, Beijing Time

	<p>boxes to refine the noisy boxes. Specifically, a new box set selector is designed to select positive boxes prone to generating accurate predictions, and then the rest occluded boxes are refined through the relation model module. To demonstrate the effectiveness of our proposed method, a new classroom video surveillance dataset ICDU is made, and we conduct extensive experiments on this classroom video surveillance dataset and the public dataset CrowdHuman. Experiment results show that our proposed method performs excellently on our ICDU dataset and CrowdHuman dataset.</p>
P1058 11:00-11:15	<p>A Fatigue Driving Detection Algorithm Based On YOLOv5 LI Zhanli, JIA Ni, JIN Hongmei Presenter: Xi'an University of Science and Technology, China</p> <p>Abstract: Aiming at the problems of slow detection speed and low detection accuracy in existing fatigue driving detection algorithms, a fatigue driving detection algorithm based on YOLOv5 is proposed. In order to improve the feature extraction ability of the network, the convolution module is used to replace the slice structure in Backbone; the algorithm combines the data features of fatigue driving images to simplify and optimize the Neck structure of the YOLOv5 model, which will be suitable for detecting 19×19 features of larger size objects. Graph branch pruning to reduce model complexity and improve real-time detection. Finally, on the basis of facial feature extraction, the algorithm determines the state of fatigue features according to PERCLOS and POM parameters combined with thresholds and outputs the results. The experimental results on the NHTU-DDD data set show that the accuracy of the detection model reaches 95.25%, the model size is only 10MB, and the single-frame detection speed is 9ms, which is 21.5% higher than the original YOLOv5 algorithm. At the same time, the model parameters are greatly reduced, which can better meet the real-time requirements of fatigue detection application scenarios.</p>
P1384 11:15-11:30	<p>Medical waste detection base on improved YOLOv5-s Shaowen Zhang, Weiya Shi, Jingfeng Yang Presenter: Shaowen Zhang, Henan University of Technology, China</p> <p>Abstract: Aiming at the characteristics of high detection accuracy and fast detection speed in the task of medical waste detection, this paper proposes an improved YOLOv5-s object detection model. Firstly, the Convolutional Block Attention Module (CBAM) is added to the original YOLOv5-s model backbone network to enhance the attention of the network model to medical waste; then the Ciou loss is used to replace the IoU loss to enhance the positioning accuracy and accelerate the convergence speed of the network model. This paper uses a variety of data enhancement methods to expand the experimental data set, a large number of experiments show that the mean average precision of this improved model reaches 90.73%, compared to the original YOLOv5-s model improved by 2.66%, detection speed reaches 115.7frame/s, the detection effect is better than the current mainstream object detection model.</p>

SESSION 8

Room B: <https://us02web.zoom.us/j/87279824425>

October 23, 2022, Beijing Time

P1394 11:30-11:45	<p>Ellipse shape prior based anti-noise network for parathyroid detection Wanling Liu, Zhongsheng Cai, Fei Chen, Bo Wang, Wenxin Zhao, Wenhuan Lu Presenter: Wanling Liu, Tianjin University, China</p> <p>Abstract: Endoscopic thyroid surgery is a common treatment method for thyroid diseases. However, due to the small size of parathyroid gland, its unstable position during surgery, and the similar color and texture of the parathyroid gland to its surrounding tissues and organs, parathyroid glands are extremely difficult to be visualized and detected during the operation, leading to accidental damages which result in abnormal parathyroid hormone secretion after surgery. Therefore, detecting the parathyroid glands and preventing damage during endoscopic thyroid surgery is critical. The current target detection technology using SOAT algorithm can help detecting the small parathyroid area, but the accuracy and efficiency are often unsatisfied because of the high image noise. In this research, we propose a PGNet model by applying an anti-noise feature extraction module to prevent the sharp drop in the accuracy of the model under the noisy environment. The anti-noise feature extraction module is based on residual structure, and hence the capability of parathyroid shape detection can be significantly improved. To further improve the detection accuracy of the parathyroid gland, an ellipse anchor frame and an improved IoU loss function have been applied in this PGNet model. The detection accuracy of PGNet model has been verified using data obtained in real-time surgical scenes, and the results show that the PGNet model can achieve 89.3% parathyroid detection accuracy when the IoU threshold is 0.5, which is better than the current SOTA CNN model.</p>
P1404 11:45-12:00	<p>Multiple Feature Fusion Algorithm for Human Fall Detection in Intelligent Monitor Video TANG Yize, QIU Lanxin, WANG Yiliang, LING Zhi, XIANG Hui , HE Chen Presenter: TANG Yize, Zhejiangxintong, China</p> <p>Abstract: Human fall detection is main abnormal event in intelligent power distribution network monitor scenario. Due to diverse body postures and human actions before human fall, it is difficult to detect human fall action effectively. To solve this problem, we studied five human fall-related features and proposed multiple feature fusion algorithm to detect human fall action in intelligent monitor video. Experiment results and field test both demonstrated its effectiveness of the proposed algorithm.</p>
P1420 12:00-12:15	<p>Frequency Domain Deepfake Detection Based on Two-Stream Neural Network Xu Yijia, Zhang Dong Dong Presenter: Xu Yijia, Tongji University, China</p> <p>Abstract: Benefiting from the progress of deep learning driven generative models, face forgery technologies have rapidly become mature, thus raising</p>

	<p>public concerns about the illegal usage of these technologies. Despite the fact that fake images and videos are often unrecognizable to human eyes, recent work has found that hidden artifacts can be exposed in frequency domain. Our method introduces the idea of separating human face area using landmarks before mining the forgery patterns in frequency domain. This we believe can help the network learn more discriminative features, also, a two-stream learning framework combining single-frame pathway and multi-frame pathway is developed to mine frequency clues. Compared with previous methods, our approach showed good results while using a few training data and little time. The effectiveness of our method is shown on different versions of the FF++ and Celeb-DF dataset.</p>
P1421 12:15-12:30	<p>Real-time Clothing Detection Networks For Surveillance Videos Zhiying Li, Hao Chen, Weibin Chen, Shuyuan Lin, Rong Zhao, Yufeng Qian, Zefan Wang, Yuer Yang and Feiran Huang Presenter: Zhiying Li, Jinan University, China</p> <p>Abstract: Real-time garments detection is a useful technique for describing individuals from the video stream captured by surveillance cameras. In most cases, the surveillance cameras are shot from a high angle and a long distance, so the garments objects in the captured images are small and blurred. Traditional object recognition models are designed to locate obvious and clear objects and then classify them into different categories. They are not sensitive to detect or even neglect small and blurry garments in surveillance camera images. To this end, we propose an effective real-time garments detection model- Multi-Scale Tiny attention-based networks, called MuST. We also collect 5000 surveillance camera images and annotate garments items into 112 categories to build an accurate benchmark dataset for garments recognition. In particular, we design and add a special tiny decoupled prediction head that helps to locate small and fuzzy garments more accurately. Moreover, the garments are usually protected or affected by distracting environments, which may negatively affect the classification accuracy. Therefore, we introduce the novel multi-scale concatenation module to integrate and contrast-analyze the information of the garment objects and their local environment. Finally, MuST can better localize and classify small and fuzzy garments. Experimental results on real-world collected garment recognition datasets prove that MuST achieves the best garment recognition accuracy among all real-time object recognition models. In addition, MuST achieves optimal inference speed among all real-time and non-real-time object recognition models.</p>

Session Topic: Optical imaging system and computer photography

Time: 10:15-12:30

<p>P0010 10:15-10:30</p>	<p>DTEA: Optical Flow Estimation with Deep Taylor Expansion Approximation Network Zifan Zhu, Chen Huang, Zhicheng Wang, Wenduo Xu, and Zhenghua Huang Presenter: Zifan Zhu, Wuhan Institute of Technology, China</p> <p>Abstract: Optical flow is one of the most challenging tasks in computer vision for recovering the three-dimensional structure and motion of objects. Traditional hand-crafted priori methods can generate robust results at the cost of highly computational complexity. While deep optical flow estimation methods are impressive for their performance, especially a fast testing speed, but fail in a good interpretability. To cope with these issues, this paper proposes a novel optical flow estimation method, namely deep Taylor expansion approximation (DTEA). We firstly discuss the relationship between optical flow and Taylor unfolding, then introduce the proposed DTEA network in detail, and finally present experimental results. Quantitative as well as qualitative results of experiments on the FlyingChairs dataset validate the proposed DTEA network is effective, of which the performance can be extensively improved when the depth of the unfolding Taylor approximation is increasing.</p>
<p>P1428 10:30-10:45</p>	<p>Study of Spectral Reflectance Recovery Based on Color Constancy Yuheng Wu, Guangyuan Wu, Chuanyang Zang Presenter: Yuheng Wu, Qilu University of Technology, China</p> <p>Abstract: A procedure for spectral recovery method based on color constancy is proposed from camera response values. The proposed method standardizes camera response values under different light sources by using different color constancy algorithms, and applied the pseudo inverse operation with the camera responses expansion to achieve the spectral reflectance recovery. To reduce the influence of external light source, the crossover combination of light sources and color constancy algorithms are calculated to determine the optimal selection of processing. The effectiveness of the color constancy algorithm was verified by using the spectral accuracy comparison and image similarity. The experimental results show that color constancy can significantly improve the spectral recovery accuracy, and the grey edge has the best performance.</p>
<p>P1378 10:45-11:00</p>	<p>Eliminating Illumination Influence via Gaussian-based Local Sensitive Histogram Kaijun Qi, Li Cao Presenter: Kaijun Qi, Nanjing University of Aeronautics and Astronautics, China</p> <p>Abstract: Eliminating the influence of illumination is one of the important preprocessing methods to improve the stability of image processing</p>

SESSION 9

Room C: <https://us02web.zoom.us/j/83168271473>

October 23, 2022, Beijing Time

	<p>algorithms. Shengfeng He provided an effective method to eliminate the influence of illumination by using local sensitive histogram. This method considers that the farther the distance between two pixels in the image, the smaller the influence of illumination on each other, which is an exponential relationship. However, this assumption has very low sensitivity for pixels at different distances from the center point, which leads to feature loss in images after eliminating the influence of illumination. Therefore, this paper applies Gaussian distribution to redefine the local sensitivity histogram, which effectively improves the sensitivity difference of pixel distance. The experimental comparison shows that in the case of large differences in lighting conditions, the algorithm proposed in this paper can retain more image features and is more robust in eliminating the influence of illumination.</p>
P0002 11:00-11:15	<p>Analysis of measuring accuracy for planar and non-planar scenes in photogrammetry Kai Guo, Hu Ye, Junhao Gu, Xin Gao Presenter: Kai Guo, Northwest Institute of Nuclear Technology, China</p> <p>Abstract: Photogrammetry with stereo vision is widely used in computer vision and SLAM (simultaneous localization and mapping), whose key steps are calibration and intersection measurement. Calibration is to obtain the intrinsic and extrinsic parameters, including the principal point, focal length and pose. Intersection measurement is to obtain the 3D information after calibration, including position, velocity and rotation. In some cases, such as visual monitoring cameras (VMCs), photogrammetry uses large field of view, and has the characteristics of long distance from camera to target and wide measuring range, which increase the difficulty of calibration and is unable to place 3D control points arbitrarily. What's more, the distance from the target area to 3D control point area has a great influence on the measuring accuracy of intersection measurement. In this paper, we proposed a new method to place 3D control points, including planar and non-planar scenes and this method can distinguish the two scenes. Then the planar and non-planar methods can be used to calibrate in different cases respectively. In addition, we analyzed the layout of 3D control points to obtain relation between the measuring accuracy and the distance from the target area to 3D control point area. Experimental results show the longer the distance, the greater the measuring error in synthetic data and real images, and to improve the measuring accuracy, the 3D control points should be planar or non-planar strictly, not quasi-planar.</p>
P1061 11:15-11:30	<p>RVSRT: Real-time Video Super Resolution Transformer Linlin Ou, Yuanping Chen Presenter: Linlin Ou, CNIC, China</p> <p>Abstract: Video super-resolution is the task of converting low-resolution video to high-resolution video. Existing methods with better intuitive effects are mainly based on convolutional neural networks (CNNs), but the architecture is heavy, resulting in a slow inference structure. Aiming at this problem, this paper proposes a real-time video super-resolution. Real-time video super resolution transformer (RVSRT) can quickly complete the super-resolution task while considering the visual fluency of video</p>

SESSION 9

Room C: <https://us02web.zoom.us/j/83168271473>

October 23, 2022, Beijing Time

	<p>frame switching. Unlike traditional methods based on CNNs, this paper does not process video frames separately with different network modules in the temporal domain, but batches adjacent frames through a single UNet-style structure end-to-end Transformer network architecture. Moreover, this paper creatively sets up two-stage interpolation sampling before and after the end-to-end network to maximize the performance of the traditional CV algorithm. The experimental results show that compared with SOTA TMNet [1], RVSRT has only 50% of the network size (6.1M vs 12.3M, parameters) while ensuring comparable performance, and the speed is increased by 80% (26.2 fps vs 14.3 fps, frame size is 720*576).</p>
P1059 11:30-11:45	<p>Fast Line Segment Matching Based on Point-Line Affine Invariants and Pairwise Constraints Haowei Zhang, Zhihe Zhang, Zhaoyan Yan, and Yan Li* Presenter: Haowei Zhang Lanzhou University , China</p> <p>Abstract: This paper we proposed a method about line segment matching based on point-line invariants. We use ORB and EDlines, two efficient and stable methods, to extract point and line features respectively. Then we introduce the implementation details of our matching algorithm. It includes affine invariance of the ratio of the distance from two coplanar points to the line and pairwise constraint of the geometric relationship between two lines. In order to eliminating mismatches, we use a series of methods to optimize the result of our algorithm. We set up a scoring mechanism for candidate matches and the final matches will be given by evaluating the voting matrix. Its performance is evaluated by extensive experiments. The results show that our proposed methods outperform the mainstream methods, and are robust to rotation, scale, blur and other transformations.</p>
P1056 11:45-12:00	<p>Self-Perceptual Generative Adversarial Network for Synthetic Aperture Sonar Image Generation Yuxiang Hu, Wu Zhang, Baoqi Li, Jiyuan Liu, Haining Huang Presenter: Yuxiang Hu, University of Chinese Academy of Sciences, China</p> <p>Abstract: Due to the shortage of Synthetic Aperture Sonar (SAS) image datasets, the development of many underwater tasks is hindered. To tackle this problem, coupling optical rendering and image-to-image translation is a novel and feasible way. However, because of the big gap between simulated optical images and real SAS images, the performances of existing works are not desired and have plenty of room for improvement. In this paper, we introduce a Self-Perceptual Generative Adversarial Network (SPerGAN) which can controllably generate SAS images with high fidelity. It utilizes a kind of self-perceptual loss to generate high-quality and diverse SAS images. Moreover, we introduce a novel evaluation method of SAS images that accords closely with human cognition. To evaluate the performance of our method, we first compare it against recent outstanding image-to-image translation methods on qualitative and quantitative aspects. Then we make ablation studies to explore the effects of different cycle consistency losses and hyper-parameters. The results show that our method can generate diverse and realistic SAS images.</p>
P0040	<p>Multi-Task Learning Using Optical Flow for Motion Magnification Li Chen, Bingchao Zhao, Xiaoyue Liu, Zhaozhou Cai, Cong Peng</p>

SESSION 9

Room C: <https://us02web.zoom.us/j/83168271473>

October 23, 2022, Beijing Time

12:00-12:15	<p>Presenter: Li Chen, Nanjing University of Aeronautics and Astronautics, China</p> <p>Abstract: The human visual systems cannot directly discern subtle but meaningful variations, such as the facial micro-expressions and the structural vibrations. Motion magnification techniques allow the naked eyes to perceive these variations. Current approaches on motion magnification generally belong to the Lagrangian and Eulerian perspectives. Nevertheless, the methods either require heavy computation or cannot extract tiny but valuable variations from noise disturbance. This paper utilizes multi-task learning to unite the Lagrangian and Eulerian perspectives for a novel motion magnification method. First, the method develops a multi-task network, which utilizes optical flow estimation to support motion magnification with accurate motion extraction. The homoscedastic uncertainty is then applied to balance task relation in the loss functions. In order to support multi-task learning, a simulated dataset is synthesized with real images from public datasets. Finally, the experimental results demonstrate that the proposed method exceeds the previous ones and optical flow can effectively support motion magnification.</p>
P0019 12:15-12:30	<p>Miss distance estimation using shadow and single view Kai Guo, Hu Ye, Ye Tian, Xin Gao and Junhao Gu Presenter: Northwest Institute of Nuclear Technology, Xi'an 710024, China</p> <p>Abstract: The stereo vision is generally used to obtain the 3D information in traditional three-dimensional measurement. At least two cameras are calibrated in advance and then resection is performed to obtain the three-dimensional coordinates. It can be seen that obtaining 3D information needs at least two cameras (or two views) because single camera can only obtain 2D information. In addition, only the 3D spatial position when the image is captured by camera can be obtained. When we measure 3D information of miss distance of the weapon with high-velocity motion, such as missile, it is hard to capture the image when the weapon touches on the target because the limitation of camera fps (frames per second). Hence, we can only obtain the position of the moment before the weapon touches on the target and this would bring error for miss distance estimation. In this paper, a fast miss distance estimation method is proposed using shadow and single view (i.e., single camera). This proposed method only uses one camera and uses the characteristic that the intersection of the axes of weapon and its shadow is the actual image projection of the moment when the weapon touches on the target. The proposed method dose not need to capture the image of moment when the weapon touches on the target and hence not need high fps, then would extend the range of choice for camera. Experimental results indicate our proposed method has better performance in terms of accuracy, numerical stability and computational speed for miss distance estimation, compared with the traditional stereo vision.</p>

Session Topic: Computer Aided Design and Image Processing

Time: 13:30-15:45

Session Chair: Professor Malik Zawwar Hussain, University of the Punjab, Pakistan

G402 13:30-13:45	<p>A Construction Method of Biofeedback Training System Based on Virtual Reality Technology Nan Chao, Wei Huang Presenter: Nan Chao, Tsinghua University, China</p> <p>Abstract: Modern people are facing increasingly serious mental health problems under the COVID-19 pandemic. Biofeedback training, as a powerful self-regulation method for stress management and relaxation training, is in need of improvement and optimization in digital era. With the aim of promoting the integration and development of biofeedback training and virtual reality, and improving users' motivation as well as engagement in biofeedback training, we come up with a novel construction method of integrating virtual reality and biofeedback training system based on psychoengineering paradigm. The method can be applied to various categories of biofeedback training including heart rate biofeedback, breath training biofeedback, stress intervention biofeedback and so on. Results of the research demonstrate that the construction principle can promote the integration and development of virtual reality and biofeedback training, and have the potential to improve user's motivation as well as engagement of biofeedback training.</p>
G404 13:45-14:00	<p>Research on the Influence of Using Augmented Reality Technology in Industrial Display Design Wei Li, Shi jing Cheng Presenter: Wei Li, Neijiang Normal University, China</p> <p>Abstract: Background: This study mainly discusses the change of traditional display design and the impact on visitors' experience by applying augmented reality technology to large-scale industrial display design. Methods: literature research, case study and comparative analysis were used method. First, it studies various literatures about augmented reality, and then designs and makes practical business cases. Finally, the impact of visitors on augmented reality experience is investigated. After the hypothesis is established through SPSS data analysis, the data is collected from the visitors through questionnaire survey, and the experience before and after using augmented reality is compared and analyzed by T-test analysis. Finally, the conclusion is drawn. Results: firstly, the application of augmented reality technology makes the visual effect of display design more intuitive, vivid, and interesting, and brings more effective publicity for the promotion of enterprises and brands. Secondly, users feel the near real product experience, receive rich and colorful information, and can interact with each other in real time, which brings visual, auditory, and tactile experience to the audience.</p>

SESSION 10

Room A: <https://us02web.zoom.us/j/84697177419>

October 23, 2022, Beijing Time

G410 14:00-14:15	<p>Application of Augmented Reality Technology in Train Maintenance Training System Hu Su, Chuan Yun Presenter: Endong Han, School of Electrical Engineering, Southwest Jiaotong University, China</p> <p>Abstract: Aiming at the lack of low-cost visual simulation tools for train maintenance, which can realize the failure simulation of various parts of the train, a train maintenance training simulation system based on augmented reality technology is proposed. Firstly, the overall design scheme of the train maintenance training simulation system is introduced. Secondly, using Unity3D as the driving engine, the stability of the virtual model display and the interaction with the virtual model during the maintenance process are designed. The Eazy AR interface and image processing technology are used to realize the The roaming of the train maintenance scene, the fault maintenance of the train equipment and the corresponding human-computer interaction function are given, and the experimental results are given.</p>
P0038 14:15-14:30	<p>VP-GAT: Vector Prior Graph Attention Network for Automated Segment Labeling of Coronary Arteries Tianqi Zhang, Tao Han, Yining Wang, Jingfan Fan, Yucong Lin, Deqiang Xiao, Jian Yang Presenter: Tianqi Zhang, Beijing Institute of Technology, China</p> <p>Abstract: Automatic segment labeling of the coronary artery tree is important for computer-aided diagnosis (CAD) of cardiovascular disease. High individual variability among human bodies makes the task very difficult. State-of-the-art methods generally rely on the location information of coronary main branches and image information in a small range, which adversely affects the labeling effect of side branches. We propose a vector prior graph attention network (VP-GAT), which uses image features of organs around the coronary arteries as anatomical prior knowledge, considering the position and direction relationships between segments and surrounding organs. VP-GAT consists of three main parts: image prior GAT, full-vector filed extractor, and image domain prior knowledge extractor. We first extract the anatomical information of the coronary arteries as a full vector field, and then extract the image domain prior knowledge through the hybrid model of ResUnet and Transformer. Finally, we feed the two into the image prior GAT to label the segments. Our method is evaluated on real clinical datasets achieving an F1 score of 95.5%. Extensive experiments show that VP-GAT significantly outperforms state-of-the-art methods in labeling the side branches of coronary arteries.</p>
P1052 14:30-14:45	<p>CSF-Net: Color Space Fusion Network for Color Constancy Quanhua Wang, Yinwei Zhan, Sheng Kong Presenter: Quanhua Wang, Guangdong University of Technology, China</p> <p>Abstract: Color constancy is the ability of the human visual system to correctly perceive colors in a scene under different illuminants. To mimic this ability, computational color constancy has been proposed to estimate illuminant color and correct the color bias for a given image. Existing</p>

	<p>computational color constancy approaches focus on the widely used RGB color space. However, RGB color space could not explicitly reflect color characteristics, so it is necessary to investigate other color spaces that better express color features. HSV color space separates the color channel from others, and hence helps us to distinguish different colors. In order to better predict illuminant colors, we design a color space fusion network CSF-Net that integrates the RGB and HSV color spaces at both the feature level and the prediction level. At the feature level, color features extracted from RGB and HSV are combined to estimate an illuminant color. At the prediction level, two illuminant colors are predicted respectively by the color features from RGB and HSV. By merging all the three predicted colors, the illuminant color is finally predicted. Experiments conducted on the public Gehler-Shi and NUS 8-camera datasets show that the proposed CSF-Net performs better than most of the prior works in illuminant color estimation and achieves better color correction in the case of complex illumination.</p>
P1063 14:45-15:00	<p>A Novel Vision-Based Scheme to Levelness Estimation of DLP Projector Feng Yang, Lidong Zhao, Zening Men, Zhi Zhao, Lifang Wu Presenter: Feng Yang, Beijing University of Technology, China</p> <p>Abstract: Digital Light Processing (DLP) 3D printing is implemented by projecting the image of the slice of the model on the exposure surface, and the accuracy of the projected image will influence the quality of the printed object. It is necessary to maintain the accuracy of the projected image. A prerequisite for high-precision printing is that the projector is horizontal. The existing methods calibrate the levelness of the projector by observing the level-meter; the calibration precision cannot be guaranteed. In this paper, a vision-based scheme is proposed to evaluate the levelness of a projector and to estimate the rotation angles. Firstly, the hardware is designed to capture the projected image on the exposure surface. A pixel in the projected image is about 74*74 pixels in the captured image. Secondly, the possible variation of pixel relation in different situations is analyzed in theory. Then, the variation of pixel distance (or slope) by tilt angle is further analyzed. The corresponding variation curves are obtained by curve fitting. Finally, based on the corresponding curves, the tilt direction and angle of the projector can be estimated from the captured image. The experimental results demonstrate that the average error of the estimated tilt angle is smaller than 0.2 degrees.</p>
P1369 15:00-15:15	<p>Research on Personalized Recommendation from the Perspective of Staff-Position Matching Chao Deng Presenter: Chao Deng, Guangdong University of Science and Technology, China</p> <p>Abstract: In the context of the COVID-19 epidemic, the mutual selection between job seekers and enterprises mostly relies on online publishing of recruitment information and the delivering resumes online. How can job seekers obtain suitable employment information quickly and effectively, and how companies can accurately recruit people who are suitable for their positions, these are very challenging problems. In this paper, taking personnel-job matching as a theoretical guide, based on the latent semantic model, the correlation model between the job seeker and the applicant's</p>

SESSION 10

Room A: <https://us02web.zoom.us/j/84697177419>

October 23, 2022, Beijing Time

	recruitment requirements was built, after comparing the similarity values of similar companies and job seekers. Realized the purpose of recommending data information with a high degree of matching to job seekers and enterprises.
P1430 15:15-15:30	<p>Design and Implementation of Virtual Intelligent Doctor Consultation System Based on Unity3D Zhibin Liu, Zhigeng Pan, Guangyv Wei, Yilu Zhao, Chuangxin Cai, Xu Lang, Hebei Gao, Hao Dai Presenter: Zhibin Liu, Nanjing University of Information Science and Technology, China</p> <p>Abstract: At present, due to the impact of COVID-19, more and more industries are implementing online office and online teaching. With the continuous development of virtual reality technology and the arrival of 5G era, computer technology brings users more convenient and immersive medical services. Based on the Unity real-time rendering development platform, this paper applies 3DSMAX modeling software to complete the production of virtual scenes and character models, combined with various open services provided by Baidu AI and OLAMI artificial intelligence platform, to complete the development of the virtual intelligent doctor consultation system. Taking the COVID-19 epidemic as an example, the article allows users to talk with virtual doctors and finally complete the whole consultation process by simulating the form of question and answer between patients and doctors in real life, which shows that the virtual intelligent doctor consultation system designed and developed in this paper has certain interest and practicability in the medical consultation application scenario.</p>
P1908 15:30-15:45	<p>An improved single constant Kubelk-Munk method for pigment unmixing of Chinese paintings SUN Yu-tong, LYU Shu-qiang, HUANG Chun-hao, HOU Miao-le, Presenter: Beijing University of Civil Engineering and Architecture, China</p> <p>Abstract: Since the mineral pigments on the surface of Chinese paintings are small in particle size and the pigment layer is too thin to cover the paper completely, their spectral mixing often shows nonlinearity. In this paper, aimed to accurately estimate the different types of pigments abundance on the surface of Chinese paintings, we proposed an improved single-constant Kubelk-Munk (KM) method for pigments unmixing algorithm. First, we selected Malachite, Azurite and Cinnabar pigments and mixed them in pairs in different abundances. They were added glue and then painted on the rice paper to produce the samples. The spectra of pure pigments and their mixed samples were acquired by the instrument of ASD FieldSpec 4. Second, we calculated the absorption and scattering ratio of the samples, pure pigments and rice paper. The rice paper and pure pigments were regarded as endmembers together for the pigment layer is too thin to cover the paper completely. The abundances were figured out base on the Fully Constrained Least Squares (FCLS) method endmember. Third, the abundances for different pigment endmember were re-calculated after removing the paper endmember. The experimental results show that the improved single-constant KM method has higher unmixing accuracy compared with the traditional spectral unmixing methods, and has good prospects for the application of unmixing composite pigments on the surface of Chinese paintings.</p>

Session Topic: Computer model and image application

13:30-15:45

Session Chair: Professor Jihua Ye, Jiangxi Normal University, China

<p>G405 13:30-13:45</p>	<p>Research on Poetry Generation Based on Bi-LSTM Networks and Attention Mechanism Taozheng Zhang, Jiaqi Guo, Yan Zhang Presenter: Jiaqi Guo, Communication University of China, China</p> <p>Abstract: Poetry is a common language for studying automatic text generation by neural network. Using neural network model to realize automatic writing of poetry is meaningful to the study of automatic text generation and the inheritance of traditional Chinese culture. This topic designs and trains a deep learning model based on bidirectional long short-term memory network (Bi-LSTM) and attention mechanism (Attention). The function of generating acrostic with customizable poet's style and artistic conception are realized. Under the Encoder-Decoder framework, the model improves the traditional encoder based on recurrent neural network (RNN) and makes it based on Bi-LSTM. It can better capture the semantic dependencies of the context, and effectively overcome gradient disappearance and gradient explosion in RNN. The decoder is improved based on the long short-term memory network (LSTM) and the attention mechanism, so that the model can better solve the long-term dependence problem in the traditional decoder. The addition of the attention mechanism enables the model to better connect contextual information, grasp the central semantics, and achieve better generation effects. This subject uses the Tang poetry corpus to train the above model. It achieves a good effect of generating acrostic, but the acrostics also have problems such as word repetition. The follow-up research direction of the subject is how to use more advanced models, such as ERNIE-GEN, to solve the above problems and achieve better generation effects.</p>
<p>G411 13:45-14:00</p>	<p>Prominence Convergence in the Strategy Coordination of Crowdsourcing Workers Jinwei Zhang, Jiuchuan Jiang, Jinpeng Wei, Huiju Zhao Presenter: Jiuchuan Jiang, Nanjing University of Finance and Economics, China</p> <p>Abstract: With the increasing complexity of tasks in crowdsourcing, researchers begin to employ how a group of workers complete complex tasks collaboratively. In most works, workers often complete tasks independently and the final results are integrated by the requester or the crowdsourcing platform. We observe that the results of workers' task execution may be affected by their social relationships in crowdsourcing platforms; in the process of task execution, workers may adaptively adjust and coordinate their own strategies according to their preferences and the interactions with other workers. Based on the above findings and inspired by the coordination mechanism in multi-agent system, this paper studies the strategy coordination mechanism of crowdsourcing workers, i.e., workers have different social statuses and strategies, and they will change their strategy through the</p>

	<p>interaction with other workers. This paper firstly defines the synchronization ability of workers and proposes a crowdsourcing worker synchronization network model. Thus, the final strategies of workers are the result of continuous coordination in the synchronization network. By using this model, this paper also explores the phenomenon of workers' prominence strategy, and finally conduct experimental tests in different environments. The experimental results demonstrate the correctness of the model proposed in this paper.</p>
<p>P0021 14:00-14:15</p>	<p>Application of XGBoost and TrajGRU to improve the accuracy of ECMWF wind forecasts Wei Zhang, Yueyue Jiang, Xiaojiang Song, Boyu Guoan, Renbo Pang Presenter: Yueyue Jiang, Ocean University of China, China Abstract: Sea surface wind is the main research object in the field of marine meteorology, and it is also one of the main reasons for marine disasters. Accurate sea surface wind forecast data is of great significance for marine disaster detection and early warning. The western North Pacific (WNP) has the greatest number of tropical cyclones of any sea in the world, with typhoons occurring virtually every month, so more accurate wind forecast data for this region is important for coastal residents and seafarers. This paper uses the European Centre for Medium-Range Weather Forecasts(ECMWF) fine grid forecast data and ERA-5 reanalysis data, and uses the TrajGRU network and the XGBoost algorithm to make rolling corrections to the ECMWF wind forecast in the future 0-120h. The experimental results show that after using the TrajGRU model to correct, the average absolute errors of wind speed and wind direction at all correction moments are reduced by about 10.3% and 4% respectively, which is better than the XGBoost method as a whole. In addition, the TrajGRU model can more accurately correct the regions with large errors in wind speed and wind direction, so that the ECMWF forecast data can be better applied to practice.</p>
<p>P1065 14:15-14:30</p>	<p>A Portrait Image Recommendation Method based on Collaborative Filtering Yuwei Hu, Xueyuan Zheng, Ping Zong Presenter: Yuwei Hu, Nanjing Skytech Co., Ltd, China Abstract: Portrait recognition is a key task in public security area. Most present portrait recognition methods lay more emphasis on extracting facial features by different algorithms for a variety of scenarios, but focus less on fusion of multiple recognition results. In this paper, we propose a portrait image recommendation method based on collaborative filtering, for which the main task is to improve the hit accuracy by utilizing the portrait recognition results from multiple feature-based models. Considering these models preferences and historical scores on portrait images, user-based collaborative filtering is applied to calculate the similarity between the models. Meanwhile, due to the difference of similarity criteria, an improved comprehensive similarity model based on Auto-Encoder is established to synthesize similarity values calculated by different feature-based models to predict the rating scores of the candidate portrait images. Experiments on portrait images in the wild show that the hit accuracy reaches 85.48% and is 1.11% higher than that best results of the single feature-based models, which illustrates that the proposed method is effective and practical for recommendation of portrait images.</p>

P1397 14:30-14:45	<p>GCN-based group level analysis of brain functional connectivity in fMRI Xiangyuan Ma, Zhisheng Deng, Zibi Xu, Yaqin Zhang Presenter: Zhisheng Deng, Shantou University, China</p> <p>Abstract: Brain functional network describes the functional connectivity (FC) between brain regions, and hence provides a crucial way for analyzing brain diseases. In order to explore neural mechanism of a brain disease, statistical test method is usually used to obtain the FC differences between normal group and abnormal group. However, it is difficult for statistical test method to utilize features from brain region nodes and brain connection edges simultaneously. In this study, we develop a method based on graph convolution network (GCN) for brain functional connectivity analysis in functional magnetic resonance imaging (fMRI). Graph convolution is used to extract the features from brain region nodes and brain connection edges simultaneously, and the interpretability of GCN is applied to obtain the FC differences between different groups. The proposed method is able to analyze the brain functional connectivity more comprehensively and can be a supplement to traditional statistical test method. A task-state public fMRI data set including healthy group, severe traumatic brain injury (TBI) patient group was used for training and testing of the models. And a statistical test method was used as baseline in the performance evaluation. The results showed that the proposed GCN-based method outperformed the statistical baseline method. This method has potential to find more useful FC when we analyzing the neural mechanisms of brain diseases.</p>
P1408 14:45-15:00	<p>An Improved GhostNet for Unsafe Driving Behavior Algorithm Shuyin Tang, Huasheng Zhu, Yang Yang, Zhanxin Sun, Yongjian Li Presenter: Shuyin Tang, Nanchang Institute of Technology, China</p> <p>Abstract: The existing unsafe driving behavior detection algorithms are difficult to meet the requirements of good real-time performance and high precision. This paper proposes an improved GhostNet unsafe driving behavior detection algorithm. The algorithm uses the Ghost convolution operator to reduce the complexity of the algorithm and improve the real-time performance. In addition, a dual-scale time adaptive module (DCTAM) is designed to extract the temporal information. Through fusing spatial and temporal information to improve the accuracy of detecting unsafe driving behaviors. The experimental results show that the algorithm in this paper reduces the computational complexity and stabilizes the detection accuracy.</p>
P1424 15:00-15:15	<p>A comparative study on continuum removal for Chang'E-4 VNIS hyperspectral data and its impact on elements retrieval Chao Wang, Xiangsui Zeng, Jia Li, Liyan Shang*, Jiong Feng Presenter: Xiangsui Zeng, Tongji University, China</p> <p>Abstract: Spectral absorption features are important parameters for lunar mineral identification and its abundance inversion. The spectral absorption features of lunar regolith around 1000 nm (Band I) and 2000 nm (Band II) are significant weakened due to long-history space weathering. Continuum</p>

SESSION 11

Room B: <https://us02web.zoom.us/j/87279824425>

October 23, 2022, Beijing Time

	<p>removal can effectively enhance the spectral absorption features of lunar surface materials, which is particularly applicable for processing the in-situ hyperspectral data that obtained by the Visible and Near Infrared Spectrometer (VNIS) onboard Chang'E-4 Yutu-2 Rover that landed in Von Kármán crater, a geologically old place. In this manuscript, three types of functions, i.e. linear, the second-order parabola polynomial and the piecewise polynomial functions, are used to fit the continuum of VNIS spectra. Their influences on calculating the parameters of spectral absorption features are accessed. It is found that the piecewise polynomial continuum performs better in enhancing the spectral absorption features of VNIS data in Band I and Band II. Both endpoints of the continuum can be found. The derived absorption features, such as band depth and centers are highly consistent with those obtained with ENVI software, who use full convex hull to determine spectral continuum. The difference between the piecewise polynomial method and full convex hull method in absorption band center and absorption band area ratio is less than 1%. In addition, we test the retrieval of W_0 and F_s with the spectral feature parameters from different continuums. It is found that continuum composed by piecewise linear equation is an appropriate.</p>
P1431 15:15-15:30	<p>RLC-Servo: A Full-Automatic Hand-Eye Cooperative Servo Model Based on Reinforcement Learning Minhao Li, Xinxin Zhang Presenter: Minhao Li, National University of Defense Technology, China</p> <p>Abstract: Hand-eye cooperative servo refers that the robot can perform corresponding action adjustments by analyzing images taken from the camera to achieve the purposes like object tracking and grasping in a calibrated visual system. The process of analyzing images includes a series of algorithms such as feature extraction, feature matching. The traditional visual servo model provides with a generally effective solution but cannot control the motion velocity of robotic arms efficiently, which brings two obvious drawbacks: The too fast motion speed may cause the object falling out of the FOV of camera, while the too slow motion speed will cost a long time to convergence. In this paper, we propose an upgraded servo algorithm by introducing Reinforcement Learning (RL) into the process of controlling. We define the servo error based on feature points and divide it into a certain number of sections to construct the state space. The agent is trained to adjust the robotic action according to the specific servo error, thereby accelerating the servo convergence process while maintaining the feature points visible. We first perform extensive experiments in the virtual simulated scene and then build a real-world visual system for evaluation. Results show that the RL-based servo model is capable of reducing 59.6% of iterations in the virtual scene and 16.9% of iterations in the real scene by contrast with the traditional methods, which illustrates the robustness and the efficiency of our proposed model.</p>
P1533 15:30-15:45	<p>A Parallel Method of NURBS Inverse Evaluation for 3D CAD Model Quality Testing Wu Liu, Ziheng Bao, Chunhui Yang Presenter: Wu Liu, Fifth Electronic Research Institute of Ministry of Industry and Information Technology (MIIT), China</p> <p>Abstract: Given a CAD model, a preprocessing step of checking its geometrical and topological correctness, as well as the consistency between them, is usually carried out before actual working. To support such checking, computing geometric properties of the model's NURBS surfaces is indispensable.</p>

SESSION 11

Room B: <https://us02web.zoom.us/j/87279824425>

October 23, 2022, Beijing Time

Among the many geometric properties, the inverse evaluation, which calculates a given 3D query point's corresponding UV parameter values, is the most fundamental. Existing relevant methods consistently employ a serial, iterative searching strategy, which suffers from efficiency issues, especially for tasks involving enormous query points. This paper proposes a parallel, subdivision-based strategy to increase query speed. It uses geometric subdivision to quickly cull out potential parameter regions where a query point's UV parameters reside, then utilizes the Gauss-Newton method to march to the precise UV values. All steps are done completely on GPU: the geometric subdivision is shared across all query points, and the Gauss-Newton marching is done in parallel. Experimental results show that a significant increase of at least 150x can be attained using the proposed method.

Session Topic: Image analysis and method

Time: 13:30-15:45

Session Chair: Professor Yanlang Hu, Xi'an Institute of Space Radio Technology, China

<p>P0007 13:30-13:45</p>	<p>Capsule Attention Module-Based CapsNet for Hyperspectral Image Classification Xinsheng Zhang, Zhaohui Wang Presenter: Xinsheng Zhang, Hainan University, China</p> <p>Abstract: Hyperspectral image (HSI) classification aims to assign each pixel with a proper land-cover label. Over the past few years, HSI classification using convolutional neural networks (CNNs) has progressed significantly. In spite of their effectiveness, CNN is not efficient in capturing the hierarchical structure of the entities in the images and does not fully consider the spatial information that is important to classification. Capsule network (CapsNet) preserves the hierarchy between different parts of the entity in an image by replacing scalar representations with vectors which has become an active area in the classification field in the past years. In this article, a capsule attention module-based CapsNet (CAM-CapsNet) is proposed which is not only employed to improve the performance of HSI classification but also to reduce the computation cost of the model. Specifically, 3-D convolutional layers are used to extract higher level spatial and spectral features. The local connection dynamic routing is proposed to reduce the number of parameters in the network. For the sake of boosting the representational capacity of CapsNet for spectral-spatial HSI classification, the network is improved by discriminating the significance of different spectral bands. A capsule attention module is designed to adaptively recalibrate spectral bands by selectively emphasizing informative bands and suppressing the less useful ones. The CAM-CapsNet was trained on three HSI datasets and achieved higher accuracy by comparing with some of the state-of-the-art models.</p>
<p>P0008 13:45-14:00</p>	<p>A Network for Acute Bilirubin Encephalopathy Classification Based Upon Attention Mechanism and 3D Convolution Kernels Haoyu Zhang, Zhaohui Wang Presenter: Hainan University, Hainan, China</p> <p>Abstract: Neonatal hyperbilirubinemia is a disease of bilirubin metabolism disorder, which is a common in newborns. Without timely medical attention, neonatal hyperbilirubinemia may develop into acute bilirubin encephalopathy, resulting in serious long-term neurological deficits. Magnetic resonance imaging, as a non-invasive imaging technique, is widely used in the diagnosis of acute bilirubin encephalopathy in newborns. However, the T1-weighted images of magnetic resonance imaging of newborns with normal myelin development and newborns with acute bilirubin encephalopathy have similar high signal intensity, making it difficult to make a clinical diagnosis based on the conventional radiological reading. As an important computer-aided diagnosis method, deep convolutional neural network has been widely used to improve the work efficiency of radiologists. In this paper, a convolutional</p>

	<p>neural network based on classification network for acute bilirubin encephalopathy is proposed. It contains a feature fusion section and a fairly deep Resnet classification network. Experimental results show that the performance of the proposal is better than those of deep learning models in discussion.</p>
<p>P0033 14:00-14:15</p>	<p>Rethinking the effect of image preprocessing in deep learning for semantic segmentation Xiangdong He, Hua Luo, Zhiqing Li , Yan Diao , Xianglong Li, Yuliang Feng Presenter: Xiangdong He, Sichuan University, China</p> <p>Abstract: When using deep learning for image analysis and understanding, preprocessing is a generally pre-requisite important step to improve the accuracy of neural networks, but it also adds an amount of computation. This is a trade-off that needs to be made for our speed-sensitive implementation of machine vision for vitreous injection surgery robots. However, the role and effectiveness of image preprocessing vary widely in current research in areas such as motion vision guidance and medical imaging, so rigorous experiments are needed to demonstrate this. We constructed a 1x1 convolutional preprocessing feature fusion layer with learnable weights, blended the preprocessed images with the original ones, and fed them into a neural network for training to systematically investigate the effects of common preprocessing methods on semantic segmentation of eye images. Based on SegFormer network architectures, we evaluated the effects of 16 frequently-used preprocessing techniques, including Laplacian, Unsharp Mask, Canny, Gabor, LBP, SFS, Haralick, Entropy, Felzenszwalb, SLIC, QuickShift, Watershed, Median Filter, Wavelet (BayesShrink and VisuShrink), Mumford Shah. These methods were divided into four groups depending on their functionality. It was found that none of the above algorithms had statistically significant positive effects in terms of accuracy. But the group of sharpening and edge extraction was both more robust to natural corruptions and more adaptive to fewer samples. The experiments show that image preprocessing is valuable only when the noise is strong or the training samples are small, but when the quantity and noise quality are normal, removing preprocessing can instead improve real-time performance.</p>
<p>P0035 14:15-14:30</p>	<p>Generative Adversarial Networks and Spatial Uncertainty Sample Selection Strategy for Hyperspectral Image Classification Wenyue Yu, Sijie Niu, Xizhan Gao, Kun Liu, Jiwen Dong Presenter: Wenyue Yu, University of Jinan, China</p> <p>Abstract: Hyperspectral image classification is widely used in agriculture, atmospheric environment and other fields. In recent years, deep learning has achieved remarkable success in hyperspectral image classification. However, supervised deep learning largely depends on training sets with high-quality labels, and obtaining large-scale data with high-quality labels is difficult, expensive and time-consuming. Therefore, in response to the problem of insufficient training samples, this paper proposes a hyperspectral image classification method based on generative adversarial networks and spatially uncertain sample selection strategy. By designing two generation networks composed of Autoencoder, the real spectral bands and spatial patches are input into the generation network to generate spectral and spatial information respectively. In order to extract more discriminative features, the discriminator uses different convolution kernels to fuse features and extract joint spatial spectral features. In addition, this paper adopts spatial</p>

	<p>uncertainty sample selection strategy, which selects more representative and informative samples for labeling. The network designed in this paper is combined with the sample selection strategy to further improve the recognition ability of the discriminator. Experimental results on three hyperspectral image datasets show that compared with several existing methods, this method is less sensitive to the number of training samples and has higher classification performance in the case of limited training samples.</p>
<p>P1387 14:30-14:45</p>	<p>Semantic segmentation of road scene based on the mechanism of adversarial attention TAN Jia-bin、 LEI Jing-sheng Presenter: TAN Jia-bin, Shanghai University of Electric Power, China</p> <p>Abstract: Semantic segmentation plays an increasingly important role in the field of computer vision. The current semantic segmentation algorithms are mainly based on full convolutional network downsampling for feature extraction, however, the reduction of image resolution due to downsampling will inevitably lead to the loss of image information and the process is irreversible. Although features can be reused to compensate for the loss, the problem of lost pixel correlation and inaccurate segmentation results still occur. In order to solve these problems, the attention mechanism and the generative adversarial network are fused into the semantic segmentation network, and a semantic segmentation network is proposed. The attention mechanism is added to the generator and discriminator modules and trained successively, then the generator module with rich information and global information is extracted, and its output is fused with the feature graph sampled above, in order to reduce the loss of local information and complete the task of image segmentation, an interpolation method based on image edge is introduced to sample the image, which improves the bilinear interpolation edge blur. Experiments on Cityscapes dataset show that the proposed network model is effective and reliable.</p>
<p>P1396 14:45-15:00</p>	<p>Vessel classification algorithm based on the convolutional attention module for natural maritime images Shijia Zhao, Jiachun Zheng, Lei Zhang, Yanduo Hu Presenter: Shijia Zhao, Jimei University, China</p> <p>Abstract: In recent years, to ensure the stable development of the marine economy, the analysis and processing of maritime vessel targets combined with natural light images have played a significant role. However, in scenes such as ports, there are sometimes background noises similar to vessel features, so the classifier based on the convolutional neural network cannot achieve good classification results. To pay more attention to effective information when classifying natural maritime images, we propose an efficient channel attention module (ECAM) and a re-parameterized spatial attention module (RSAM) through one-dimensional convolution and re-parameterization method. This paper combines ECAM and RSAM to present an efficient re-parameterized convolutional block attention module (ER-CBAM) to classify natural maritime images. Besides, in response to the current lack of large-scale natural marine image datasets, we established a natural maritime vessel image (NMVI) dataset. Experiments on NMVI show that by combining the proposed attention module, ResNet50 can achieve a top-1 accuracy score of 85.32% with almost no extra computational consumption and a 3.02% improvement</p>

	<p>compared to the previous 82.30%, which suggests that the proposed method is suitable for maritime scenarios.</p>
<p>P1398 15:00-15:15</p>	<p>Adaptive scale based U-shape transformer network for ischemic stroke lesion segmentation in CTP images Huiling Zhang, Wencong Zhang, Yingjia Chen, Zibi Xu, Xiangyuan Ma Presenter: Huiling Zhang, Shantou University, China</p> <p>Abstract: Ischemic stroke lesion segmentation in Computed Tomography Perfusion (CTP) images is crucial for the quantitative diagnosis of stroke. However, it remains a challenging problem due to the poor image quality of CTP and the complex appearance of the lesions. In this study, we develop a U-shape transformer network with an adaptive scale for ischemic stroke lesion segmentation in CTP images. The state-of-the-art nnU-Net structure is used as the backbone, and a transformer block with self-adapting scale is introduced. The proposed network adopts the advantage of transformer in capturing global information and retains the advantage of convolutional neural network (CNN) in extracting local correlation features. In order to obtain better adaptation of transformer block to ischemic stroke segmentation task, we propose a self-adapting scale selection strategy that offer better patch size and window size to assist the transformer block capture more global information and avoid semantic information being corrupted. Five-fold cross-validation was used in training of the networks, and nnUNet was used as a baseline model in the performance evaluation. The results showed that after involving the proposed method, the mean DICE of the segmentation improved from 0.72 to 0.78 in the ISLES public dataset. For the independent test set, the proposed method achieved a mean DICE of 0.48, a mean precision of 0.60, and a mean recall of 0.46, compared to 0.46, 0.57 and 0.43 by the baseline model. The proposed framework has the potential for improving diagnosis and treatment of ischemic stroke in CTP.</p>
<p>P1409 15:15-15:30</p>	<p>Research on Superpixels Segmentation of Cloud Remote Sensing Images Based on Density Features Yang Yang, Xinchao Yin, Qi Zhang, Yaxing Sun Presenter: Yang Yang, Nanjing University of Science and Technology, China</p> <p>Abstract: Remote sensing images are widely used in many applications of earth observation. However, the existence of clouds seriously affects the interpretation of remote sensing images. In order to improve the accuracy of cloud detection, it is usually necessary to complete the segmentation of the cloud boundary before the operation of cloud detection. Based on the SLIC algorithm, an improved SLIC superpixels segmentation method based on density feature is proposed to realize the image segmentation of cloud remote sensing images. First, in the initial clustering center generation, the density peak clustering method is used instead of the uniform setting method. Then, in the calculation of distance measurement, the local density distance term is added. Finally, we get the superpixels segmented image by iteration. Four remote sensing images with different underlying surfaces were selected as the test data. Compared with the SLIC algorithm, the qualitative and quantitative results and calculation speed show the effectiveness of our method.</p>

P1415
15:30-15:45

A Novel Image Scaling Algorithm Based on Wavelet Transform and Polyphase Filtering

Junwei Yang, Linhuang Wu, Yewang Chen, Weijie Yu, Shiyu Lin

Presenter: Junwei Yang, Fuzhou University, China

Abstract: Due to the poor performance of the traditional interpolation algorithms in terms of the video image scaling effect and the low algorithm complexity required by existing hardware implementations, a scaling algorithm based on wavelet reconstruction and polyphase filtering implementation is proposed. The method first extracts the pre-generated weight coefficients according to the scaling requirements that including scaled image quality, size, and processing speed. Furthermore, the filter coefficients are interpolated with the source image pixel points mapped by the target image through a polyphase filter. Moreover, the resulting interpolated image is selectively decomposed and reconstructed by Haar wavelet transform to generate super-resolution images. The experimental results show that the algorithm improves the scaling quality of the images both subjectively and objectively, while enhancing the processing speed of the images and achieving non-integer scaling.

Session Topic: Image detection and recognition

Time: 16:00-18:15

Session Chair: Professor Qian Zhang, Taishan University, China

G406 16:00-16:15	<p>Effects of Hyper-parameters Setting in Bi-LSTM-CRF on Chinese Named Entity Recognition Taozheng Zhang, Pingping Ma Presenter: Pingping Ma, Communication University of China, China</p> <p>Abstract: Named entity recognition is a basic task in the field of natural language processing. It also plays a role that can not be underestimated in the era of big data. This experiment will use the Bi-LSTM-CRF model to extract information from the input text to achieve the function of named entity recognition. In this experiment, we first select a suitable data set and perform vectorization, then build and train the Bi-LSTM-CRF model. At the same time, the dropout mechanism is added to assist. The optimal hyper-parameters will be found by constantly changing the parameter settings, so that the model shows the accuracy and robustness in the NER task. Each evaluation index reaches the optimal value. After the optimized model is obtained, the visualization of the model is carried out. All the entity parts are extracted from the input text and then output, showing the effect of the named entity recognition of the model and realizing a high level of named entity recognition.</p>
P0017 16:15-16:30	<p>Adaptive Amplification of Image Texture Boundaries Jian Wei, Jiaqi Guo, Guolan Ma, Jingyan Wei, Qi Wang, Shigang Wang, Yan Zhao, Xuejun Wang Presenter: Jian Wei, Jilin University, China</p> <p>Abstract: The texture boundaries normally have frequent co-occurrence in natural images, but existing image enhancement techniques for the most parts focus on sharpening the edges, i.e., intensity discontinuities. Moreover, these approaches often suffer from noise over-emphasis and extra artifact production. In this paper, we propose an adaptive texture boundary boosting algorithm. The proposal exploits the co-occurrence filter for dual-layer decomposition and pixel-wise amplification factor calculation for image synthesis. This leads to a various-scale enhancement framework, which allows our method to only highlight those co-occurrent features while avoiding unsalient edges and noise, gradient reversals, as well as halos. Both subjective and objective comparisons using different detail stretching schemes demonstrate the effectiveness of the presented approach.</p>
P0018 16:30-16:45	<p>Research on the comparison of FCN and U-Net in Remote Sensing Image Change Detection Xing Zhou, Xiaorong Xue and Guangna Qu Presenter: Xing Zhou, Xiaorong Xue, Liaoning University of Technology, China</p>

	<p>Abstract: Remote sensing image change detection has a wide range of applications in urban planning, disaster monitoring, environmental protection and other fields. Since fully convolutional neural network has a good performance in image processing, it is widely used in remote sensing image change detection, among which U-Net and FCN are two important fully convolutional neural networks. After a comparative analysis of the two neural network structures, it is proposed that the FCN structure has a better ability to extract changed informations. At the same time, a skip connection method CSC is proposed which can enhance the feature extraction ability of FCN. The computational complexity of FCN is almost unchanged after CSC is applied. The change detection capability of CSC-FCN exceeds that of U-Net when the computational complexity is much lower than that of U-Net. It is concluded that the FCN structure has better change detection ability in dealing with multi-channel data containing complex timing information.</p>
<p>P0022 16:45-17:00</p>	<p>An Effective Unsupervised Method for Change Detection in SAR Images Qianqian Liu, Xiaorong Xue, Kejun Liu and Haiyan Liu Presenter: Qianqian Liu, Xiaorong Xue, Liaoning University of Technology, China</p> <p>Abstract: Synthetic aperture radar (SAR) image change detection has a very important and wide applications, and the complexity of SAR image data makes the accurate identification of change regions a challenging task. Many Convolutional Neural Networks (CNN) have been successfully applied to change detection tasks, but most of them have complex structures and heavy computation, so this paper proposes an unsupervised lightweight Convolutional Neural Network (DNNet) for change detection. We introduce conditionalized convolution (CondConv) and ConvNeXt block into the change detection task to improve the performance of the network in recognizing complex objects. In this paper, the bi-temporal images are first subjected to differential analysis, and then the convolutional neural network is trained with pseudo-labels generated by clustering. The effectiveness of DNNet is verified on three real SAR datasets and compared with several state-of-the-art methods. Experiments results show that the proposed method has better effectiveness.</p>
<p>P0036 17:00-17:15</p>	<p>A multi-attention based fMRI feature extraction method for brain states recognition Chong Wang, Hongmei Yan, Tao Liu, Wei Sheng, Yun-Shuang Fan, Rong Li, Huafu Chen Presenter: Chong Wang, University of Electronic Science and Technology of China, China</p> <p>Abstract: Identifying the experimental conditions from brain activity has been a focus direction in recent functional magnetic resonance imaging (fMRI) studies, which advances our understanding of the brain mechanism and can be applied in brain computer interface systems. However, task fMRI signals contain a lot of noise that is irrelevant to the task (e.g., spontaneous brain activity and physiological noise), and these task-independent components limit the performance of brain states recognition. In this work, we proposed a multi-attention neural network (MANN) to extract task-dependent components of the fMRI data and recognize the task conditions. We employ three attention modules in MANN (temporal, spatial and relational attention modules) to</p>

SESSION 13

Room A: <https://us02web.zoom.us/j/84697177419>

October 23, 2022, Beijing Time

	<p>describe the brain activation from multiple dimensions and extract task-related fMRI features. We evaluate the proposed model using the emotion task fMRI data from the Human Connectome Project dataset, in which more than 1000 participants are adopted. The MANN achieves a classification accuracy of 99.51% between different task conditions (shape and face). To further investigate how the attention modules work, we visualize the attention weights and perform ablation studies. Our result indicate that the attention modules can learn biologically meaningful brain representations and contribute to the improvement of the classification accuracy. Our model offers a powerful tool for brain states recognition, and has the potential application in clinical diagnosis.</p>
P0041 17:15-17:30	<p>Saliency-based Fabric Defect Detection Network with Feature Pyramid Learning and Refinement Zhoufeng Liu, Bo Tian, Xiao Li, Chunlei Li, Yan Dong Presenter: Bo Tian, Zhongyuan University of Technology, China</p> <p>Abstract: Fabric defect detection is an essential step of quality control in the textile manufacturing industry. The fabric image texture and defects are complex and diverse, which result in poor detection results and low efficiency of the traditional fabric defect detection algorithm. Visual saliency model can quickly outstand the salient object from the complex background, and has been proven applicable in fabric defect detection. However, the existing saliency detection models still confront great challenges in boundary refinement and lineshaped defect detection. Therefore, a novel saliency-based fabric defect detection network with feature pyramid learning and refinement module is proposed to powerfully characterize features and refine boundary, in which a scale-correlated feature pyramid module (SCFPM) with cross-level connections is proposed to effectively characterize the multi-scale features from the backbone network. Moreover, an auxiliary refinement module (ARM) is designed to further refine and strengthen the input features. Finally, we incorporated the hand-crafted saliency priors to guide the network to generate the accurate saliency maps. Extensive experiments on the built fabric image datasets demonstrate that our proposed model performs favorably against most state-of-the-art methods under different evaluation metrics.</p>
P1047 17:30-17:45	<p>A new database of Houma Alliance Book ancient handwritten characters and its baseline algorithm Xiaoyu Yuan, Zhibo Zhang, Yabo Sun, Zekai Xue, Xiuyan Shao, Xiaohua Huang Presenter: Xiaohua Huang, Nanjing Institute of Technology, China</p> <p>Abstract: The Houma Alliance Book is one of the national treasures of the Museum in Shanxi Museum Town in China. It has great historical significance in researching ancient history. To date, the research on the Houma Alliance Book has been staying in the identification of paper documents, which is inefficient to identify and difficult to display, study and publicize. Therefore, the digitization of the recognized ancient characters of Houma League can effectively improve the efficiency of recognizing ancient characters and provide more reliable technical support and text data. This paper proposes a new database of Houma Alliance Book ancient handwritten characters and a multi-modal fusion method to recognize ancient handwritten characters. In the</p>

	<p>database, 297 classes and 3,547 samples of Houma Alliance ancient handwritten characters are collected from the original book collection and by human imitative writing. Furthermore, the decision-level classifier fusion strategy is applied to fuse three well-known deep neural network architectures for ancient handwritten character recognition. Experiments are performed on our new database. The experimental results first provide the baseline result of the new database to the research community and then demonstrate the efficiency of our proposed method.</p>
<p>P1381 17:45-18:00</p>	<p>ROBOT Arm Gripping Pose Estimation Algorithm Based on Binocular Camera ZHAO Jiao, GE Wan-Yi, XU Zhe, JING Shu-Xu Presenter: Jiao Zhao, Chang'an University, China</p> <p>Abstract: The robot can work stably and reliably on the assembly line with fixed position. However, when the position of the robot is not fixed and the pose of the working object is unknown, the robot needs to obtain the pose of the working object and the relative position relationship between the robot and the working object to grasp the target. Aiming at the above problems, this paper designs a pose estimation method for the manipulator to grasp the target. Firstly, the backbone network ResNet50 is replaced by a lightweight, efficient and less computationally intensive MobileNet network to improve the Mask R-CNN network. Secondly, the improved neural network was trained by transfer learning based on image acquisition and COCO dataset. Finally, the pose estimation method of binocular data source based on line features is designed to obtain the pose information of grasping objects. Experimental results show that the operating speed of MobileNet Mask R-CNN network is faster than Mask R-CNN network, and the speed of target recognition is also increased by about 4%. At the same time, high precision pose information can be obtained, and it has high efficiency and robustness.</p>
<p>P1392 18:00-18:15</p>	<p>MR Image Reconstruction via Non-Local Attention Networks Liu Zhou, Minjie Zhu, Dongping Xiong, Lijun Ouyang, Yan Ouyang, Xiaozhi Zhang Presenter: Liu Zhou, University of South China, China</p> <p>Abstract: As an advanced medical imaging technology, magnetic resonance imaging (MRI) has great advantages and application potentials in medical clinical diagnosis. However, since the long scanning time and the artifacts caused by patient movements, the imaging results are always not satisfactory. Therefore, accelerating MRI and improving the imaging quality are the key problems. In this work, we propose a novel deep network that combines the U-net architecture with non-local attention blocks for MRI reconstruction. We employ the U-net to construct the basic network. The non-local attention is exploited to capture the remote dependencies in MRI images which calculates the weighted average of the remaining multiple location features as the value of the response location. The U-net has limitations in capturing long-term dependencies, however, the non-local attention can solve this problem well. Furthermore, we develop the residual module to better retain the detail information. The proposed model is compared with some recent leading MRI reconstruction methods, including the state-of-the-art deep learning-based methods. Compared with these methods, the proposed residual non-local attention network provides superior MRI reconstruction results and retains better perceptual image details.</p>

Session Topic: Image fusion

Time: 16:00-18:00

Session Chair: Professor Hongjian Shi, BNU-HKBU United International College, China

<p>P0003 16:00-16:15</p>	<p>Multiscale fusion and convolution spatial propagation networks for deep complementation of outdoor scenes Hui Chen, Shuqi Liu, Heping Huang and Muhammad Ilyas Menhas Presenter: Shuqi Liu, Shanghai University of Electric Power, China</p> <p>Abstract: In autonomous driving and other robotics, rich depth perception is critical for 3D reconstruction tasks of outdoor landscapes. Many neural networks combine sparse depth maps with high-quality RGB images to generate a dense effect, resulting in dense depth maps. However, they frequently combine LiDAR and RGB image data by conducting feature concatenation or element addition, which results in the loss of some features as well as changes to the depth values of the original sparse data. To address these issues, this article proposes that the relationship between spatial and channel attention be used to link local and global features in order to accurately complete and correct sparse input so that the RGB images can better lead the depth completion job. We also use an affinity matrix to keep the original depth values in order to make the RGB image simply act as a guide without modifying the original pixel depth. We tested and assessed our algorithm on the KITTI dataset, and it outperformed the existing network in outdoor situations.</p>
<p>P1044 16:15-16:30</p>	<p>Medical image fusion based on multi-scale transform and sparse representation Qiaoqiao Li, Weilan Wang, Shi Yan Presenter: Qiaoqiao Li, Northwest Minzu University, China</p> <p>Abstract: Multi-modality image fusion is an effective technique to fuse the complementary information from multi-modality images into a fused image. In this paper, a new fusion method is proposed based on multi-scale transform and sparse representation. In the sparse representation-based method, the key point of it is to construct a good dictionary. In this method, a set of adaptive sub-dictionaries are constructed instead of a single dictionary. First, the source images are decomposed by nonsubsampled contourlet transform (NSCT) to obtain the low-pass coefficients and high-pass coefficients. Second, the sparse representation-based method is applied to fuse the low-pass coefficients. In this processing, low-pass images are divided into classified patches which are based on the patch oriented gradient. A set of sub-dictionaries are constructed from the classified patches. For the high-pass coefficients, we choose the "max-absolute" rule to fuse them. Third, combining the fused low-pass coefficients and the fused high-pass coefficients we can obtain the fused coefficients. In the last, the fused image is reconstructed by using inverse NSCT. Experimental results have shown the effectiveness of the proposed method.</p>

P1382 16:30-16:45	<p>Exploring Affective Image Representation with Visual Attention and Aesthetic Fusion Jixiong Ma, Hao Zhang, Kangjian He, Dan Xu Presenter: Jixiong Ma, Yunnan University, China</p> <p>Abstract: Affective image analysis aims to understand the sentiment of different images. The challenge is to develop a discriminative representation that bridges the affective gap between low-level features and high-level emotions. Most existing studies bridge the gap by designing deep models carefully to learn global representations in one shot directly or identify image emotion by extracting features at different levels in the model. They ignore that both local regions of an image and relationships between them impact emotional representation learning. This paper develops an affective image analysis method based on the aesthetic fusion hybrid attention network (AFHA). A modular hybrid attention block is designed to extract image emotion features and model long-range dependencies of images. By stacking hybrid attention blocks in ResNet-style, we obtain an affective representation backbone. Furthermore, considering that image emotion is inseparable from aesthetics, we employ a modified ResNet to extract image aesthetics. Finally, through a fusion strategy, the image's emotion is considered with the aesthetics conveyed. Experiments demonstrate the close relationship between emotion and aesthetics, and our plan has an excellent competitive effect compared with existing methods on the image sentiment analysis dataset.</p>
P1406 16:45-17:00	<p>Underwater Stereo Matching based on Multilevel Recurrent Field Transforms with Iterative Attentional Feature Fusion Jiaqi Leng, Ying Gao, Zhijie Xie, Yanhai Gan, Qingxuan Lv, Hao Fan Presenter: Jiaqi Leng, Ocean University of China, China</p> <p>Abstract: Stereo matching is the key to reconstruct depth map. Recent works focus on deep learning to improve accuracy. Most networks encountered the difficulty of poor generalization ability and high computational cost especially on high resolution images. Transformed from the network for optical flow, Recurrent All-Pairs Field Transforms for Stereo (RAFT-Stereo) achieved extraordinary accuracy and cross-dataset generalization in atmosphere. However, it depends on the similarity between the pretrained dataset and the target dataset. Due to the difficulty of obtaining underwater scenarios, high-quality underwater datasets are insufficient and consequently, the accuracy in underwater dataset drops significantly.</p> <p>In this paper, we revise the residual block in its feature extractor to improve the performance of RAFT-Stereo in underwater scenarios. We choose an iterative Attentional Feature Fusion module to utilize the global information in feature fusion. At last, we test our networks on ETH3D benchmark and our own underwater dataset. The metrics show that our method improved comprehensively.</p>
P1412 17:00-17:15	<p>Image Restoration Method based on Adaptive Multiple Priors Fusion in Scattering Scenes Rui Ruan, Weihui Zeng, Yu Lei, Yangyang Guo, Zheng Liang Presenter: Rui Ruan, Anhui University, China</p>

	<p>Abstract: Image is usually characterized by low contrast, blurry detail and distorted color due to complex imaging mechanism of scattering scene, which veils many valuable image information. To improve images quality in scattering scenes, we propose a novel image restoration method based on adaptive multiple priors fusion, which mainly includes two components: a multiple priors constraint-based transmission estimation component and an adaptive multiple priors fusion-based backscattered light estimation component. Firstly, we proposed a scoring formula by fusing lightness prior, contrast prior and saturation prior to locate the backscattered light, which can effectively avoid the limitation of using any single priors. Afterward, we explore a new prior called extended dark channel prior (EDCP), and then adaptive combining EDCP, dark channel prior (DCP) and underwater dark channel prior (UDCP) to estimate the transmission robustly. Finally, Extensive experiments on images of different scattering scenes demonstrate that the proposed method is effective and superior for image restoration.</p>
<p>P1414 17:15-17:30</p>	<p>Siamese Network Algorithm Based on Multi-Scale Channel Attention Fusion and Multi-Scale Depth-Wise Cross Correlation Qingjun Chen, Hua Zheng, Hao Pan, Xiaoqi Liao and Hongkai Wang Presenter: Hua Zheng, Fujian Normal University, China</p> <p>Abstract: The research takes the feature extraction network and depth-wise cross correlation learning method of the Siamese network as the starting point. Firstly, the regression strategy of the proposed framework is anchor-free, and the residual network ResNet50 is chosen as the backbone network, and add the channel attention mechanism SENet. The SE-MSCAM multi-scale channel attention model is proposed to make up for the lack of local feature extraction ability of the feature extraction network on the basis of SENet. On this basis, the attention fusion module AFFN is added to enhance the soft selection of attention. Combined with the SE-MSCAM multi-scale attention model and the attention fusion module AFFN, the ResNet50-AFFN multi-scale channel attention fusion network is proposed. Secondly, regarding the limitation of single-scale learning of SiamRPN++ depth-wise cross correlation, the MS-DWXCORR multi-scale depth-wise cross correlation is proposed which increases the diversity of learning feature scales to improve the efficiency of tracking network similarity learning. The experimental results show that, on the VOT2018 benchmark, the EAO of our method outperforms 4.0% of the mainstream algorithm SiamCAR, the tracking accuracy is improved by 3.4% and the tracking speed of our method maintains 40 FPS; the tracking success rate is improved by 2.0% and the tracking accuracy rate is improved by 3.2% compared to the mainstream algorithm SiamCAR. It has higher accuracy and robustness in dealing with occlusion, deformation, illumination variation, fast motion, and other scenarios of visual tracking, and has better tracking performance.</p>
<p>P1532 17:30-17:45</p>	<p>Single Image Snow Removal via Multi-Scale Dual Domain Decomposition and Fusion Zhang Yunpeng, Zhou Pucheng, Xue Mogen Presenter: Zhang Yunpeng, PLA Army Academy of Artillery and Air Defense, China</p>

	<p>Abstract: Under snowy weather conditions, cameras are prone to the interference of snow and can severely reduce the quality of the captured images, which will affect the computer vision performance greatly. Since no temporal information can be exploited, snow removal from single image is a challenging problem. In this paper, a novel snow removal method from single image was proposed by designing a kind of multi-scale image processing framework both in the spatial and frequency domain. Firstly, the input snowy image was decomposed into detailed sub-images and approximate parts by the Laplacian pyramid transform. Secondly, the approximate part is decomposed again into the background and detailed sub-image by the edge-preserving and structure-preserving image smoothing filter. After that, the non-subsampled shearlet transform was introduced to detect snowflakes within the frequency domain of the detailed sub-images, while mathematical morphological filtering was adopted to remove the labeled snowflakes within their spatial domain. Finally, the desnowing image was obtained by the inverse Laplacian pyramid transform. Experiments on real-world snowy images show that the proposed method produces better results than those of other state-of-the-art methods.</p>
P2017 17:45-18:00	<p>Instance-level image synthesis method based on Multi-Scale style transformation Jihong Yang, Haiyan Shao, Huabiao Qin, Yixiang Xie and Linyi Huang Presenter: Jihong Yang, South China University of Technology, China</p> <p>Abstract: Semantic image synthesis is to synthesize photorealistic images according to the given input semantic layout. Existing methods try to build a single-scale style encoder based on semantic regions, which inject style simply at a single level, constraining its application in scenes with multiple styles. Especially for different instance objects in the same semantic region, single-scale networks tend to generate the same style and control style ineffectively. To cope with this issue, we propose Multi-Scale Instance-level image synthesis method (MSIN). In order to learn more discriminative representation from different feature levels in instance, a multi-scale style encoder is designed to extract more details instead of traditional single-scale style encoder, which adopts a "pyramid" structure to contact contextual information. In addition, in order to synthesize visually pleasing and photorealistic images, MSIN leverages the region-style fusion mechanism in adaptive normalization layer, which realizes instance-wise object-to-object multi-style generation simultaneously. Compared with the previous methods, our method can generate images with fine details successfully and control style in instance object, whose semantics are more reasonable and diverse of different instance objects. The experimental results demonstrate the superiority of MSIN on dealing with semantic image synthesis tasks and outperforms existing methods in terms of instance objects and diverse generation.</p>