



**V**ALSE 2023 无锡

**视觉与学习**青年学者研讨会  
VISION AND LEARNING SEMINAR

**会刊**

**主办单位**

中国人工智能学会 中国图象图形学学会

**承办单位**

江南大学 无锡国家高新技术产业开发区管理委员会

**协办单位**

江苏省人工智能学会 无锡市计算机学会

中国图象图形学学会青年工作委员会

2023年6月10日—12日

## 目录

VALUE 2023 欢迎您 .....	1
大会组委会 .....	2
VALUE 2023 会场分布 .....	3
VALUE2023 会议总体日程 .....	6
Tutorial 与 Workshop 日程一览 .....	10
大会报告及讲者简介 .....	21
年度进展评述 (APR) 及讲者简介 .....	25
Tutorial 报告及讲者简介 .....	21
Workshops 报告及讲者简介 .....	35
Workshop 1: 大模型对 CV/PR 的挑战与机会 .....	35
Workshop 4: 多模态认知计算 .....	49
Workshop 5: 语言文字识别与理解 .....	52
Workshop 6: ChatGPT 与计算机视觉 .....	56
Workshop 7: 机器人具身智能 .....	60
Workshop 8: 物联网与工业应用 .....	65
Workshop 9: 混合现实中的视觉与学习技术 .....	70
Workshop 10: 目标检测与分割 .....	74
Workshop 11: 大模型背景下的自动驾驶与车路协同 .....	78
Workshop 12: 多模态大模型与提示学习 .....	82
Workshop 14: 视觉内容生成 .....	90
Workshop 15: 自监督视觉表征学习 .....	94
Workshop 16: 人工智能与科学计算 .....	78
Workshop 19: 优秀学生论坛 .....	111
Poster 交流论文一览表 .....	122
合作单位简介 .....	140
部分组织单位简介 .....	150
VALUE——学术华尔兹 .....	151
VALUE 在线活动参与方法介绍 .....	153
VALUE 2022-2023 年度 Webinar 报告一览表 .....	154
VALUE 2022-2023 年度 Student Webinar 报告一览表 .....	160

<b>VALSE 各委员会</b> .....	162
<b>会场及酒店</b> .....	162
<b>会场位置</b> .....	166
<b>交通信息</b> .....	166
<b>酒店信息</b> .....	169

# VALSE 2023 欢迎您

欢迎大家来到无锡，享受VALSE给大家带来的学术盛宴。

VALSE发起于2011年，是Vision And Learning SEminar的简写，取法语“华尔兹舞”之意。旨在为计算机视觉、图像处理、模式识别与机器学习研究领域的华人青年学者提供一个自由、平等、低成本的深度学术交流舞台。在这个舞台上，我们恪守并倡导理性批判、勇于探索、实证、创新等科学精神；在这个舞台上，我们倡导自由平等原则下、理性而纯学术的百家争鸣和思想交锋；在这个舞台上，我们期望欣赏到国内青年学者越来越优美的学术华尔兹（VALSE）。通过这个舞台，我们期望促进国内青年学者的思想交流和学术合作，从而在相关领域做出重量级的学术贡献，提升中国学者在国际学术舞台上的学术和影响力。

围绕上述目标，过去十二年来，VALSE逐渐形成了自己的特色社区文化、找准了自己的使命，包括：

- 1) 创造深层次学术交流与合作的新模式；
- 2) 搭建经济实用的在线学术交流舞台；
- 3) 构筑连接学术界和工业界间的桥梁；
- 4) 践行国际学术规范，倡导先进科研理念。

第十三届VALSE大会于2023年6月10-12日在无锡举行。届时将延续历年传统，呈上3个大会主旨报告、4个大会特邀报告、12个APR报告、4场Tutorial、20场Workshop、200+篇顶会顶刊Poster，合计共有百余位知名青年学者共同带来视觉与学习等人工智能领域的一次学术盛会。此外，在会场合作单位展区还将展示合作单位精彩的演示。

鉴于VALSE不向参会者收取任何费用，故特别感谢OPPO、图森未来、AutoDL、华为、百度、字节跳动、阿里妈妈、合合信息、蚂蚁技术研究院、极市平台、奥比中光、真格基金、地平线、金山办公、融科联创（天津）、思腾合力（天津）、美团、格灵深瞳、美图、超集信息、爱诗科技、趋动科技、商汤科技等企业对本次会议提供的赞助与大力支持。

VALSE 2023 组委会

# 大会组委会

## 大会主席 (General Chairs)

吴小俊 江南大学  
程明明 南开大学  
陈卫 江南大学

## Tutorial Chairs

刘夏雷 南开大学  
郑乾 浙江大学

## 展览主席

侯淇彬 南开大学  
李超 江南大学

## Finance Chairs

程一 中科院计算所  
胡聪 江南大学

## Publicity Chairs

丛润民 山东大学  
李辉 江南大学  
张鼎文 西北工业大学

## Workshop Chairs

苏航 清华大学  
郭裕兰 国防科技大学  
崔兆鹏 浙江大学  
谢凌曦 华为  
刘偲 北京航空航天大学  
周晓巍 浙江大学  
马超 上海交通大学  
洪晓鹏 哈尔滨工业大学  
胡鹏 四川大学  
陈涛 复旦大学

## 本地主席

王锐 江南大学  
罗晓清 江南大学

## 程序委员会主席 (Program Chairs)

徐天阳 江南大学  
王楠楠 西安电子科技大学  
贾伟 合肥工业大学

## Poster Chairs

王福田 安徽大学  
刘哲 江南大学  
刘建军 江南大学

## Website Chair

郑海永 中国海洋大学

## Registration Chairs

郭宗辉 中科院计算所  
陆恒杨 江南大学

## Sponsorship Chairs

山世光 中科院计算所  
姬艳丽 电子科技大学

## APR Chairs

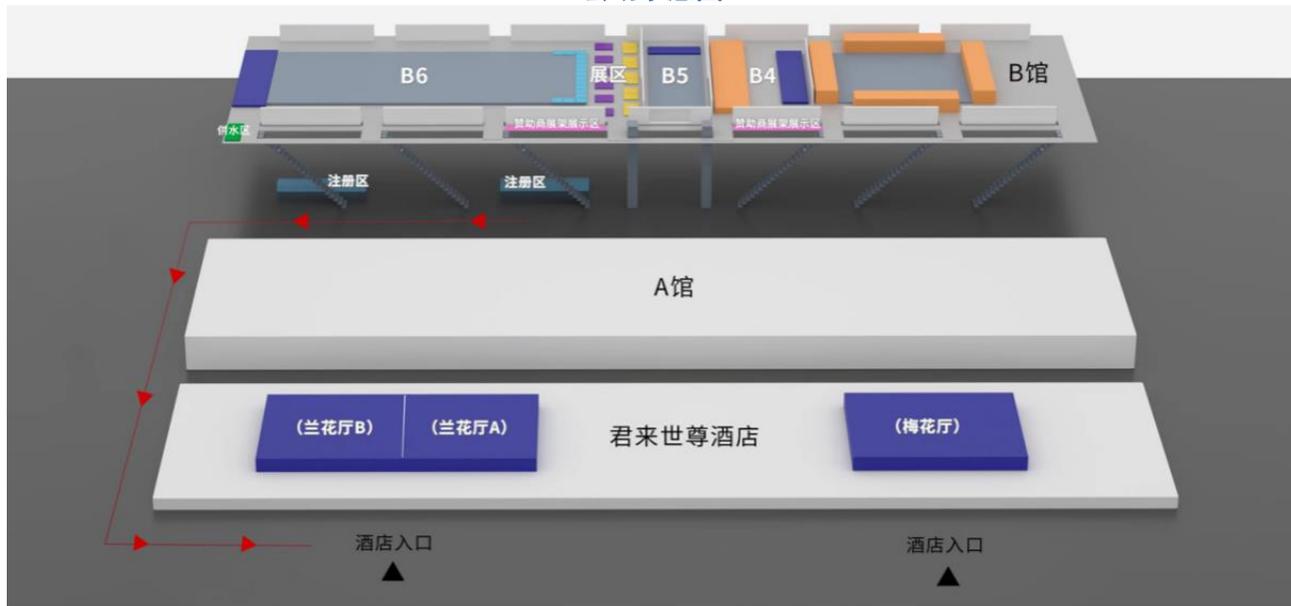
白翔 华中科技大学  
江波 安徽大学  
严骏驰 上海交通大学

## 本地主席

张东霖 江南大学  
朱书伟 江南大学

# VALSE 2023 会场分布

## 会场示意图

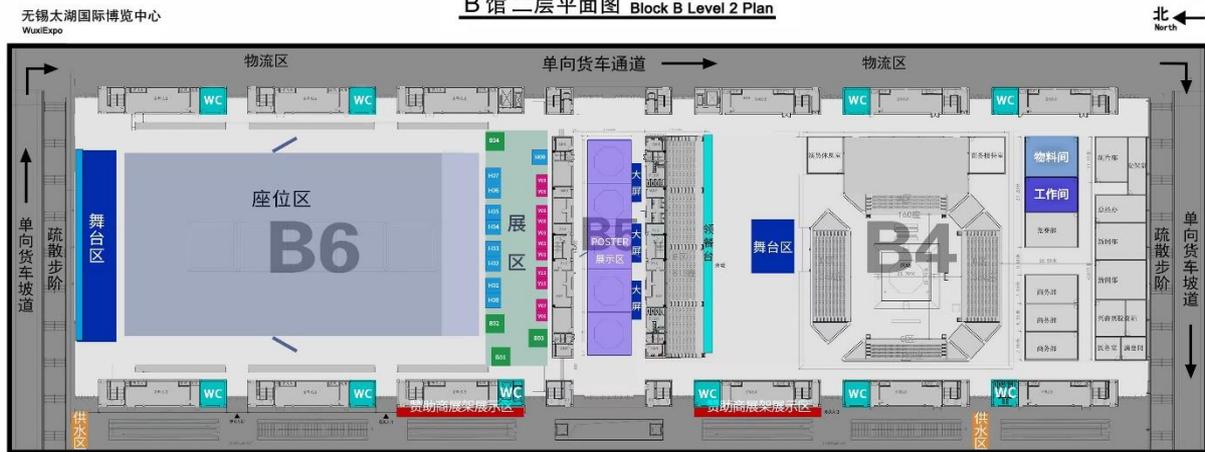


## 开幕式--布局点位图

### 6月10号布局图

无锡太湖国际博览中心  
WuxiExpo

B馆二层平面图 Block B Level 2 Plan



## 第二天--布局点位图

# 6月11号布局图

### B 馆二层平面图 Block B Level 2 Plan

无锡太湖国际博览中心  
WuxiExpo

北  
North



## VALSE2023 会议总体日程

6月10日		
时间	内容	地点
08:15-09:00	<b>开幕式</b>	B6 馆
09:00-09:40	大会主旨报告-1 报告人：高文（北京大学） 报告题目：特征编码与数字视网膜	
09:40-10:15	大会主旨报告-2 报告人：焦李成（西安电子科技大学） 报告题目：下一代深度学习的思考与若干问题	
10:15-10:50	大会主旨报告-3 报告人：陈熙霖（中国科学院计算技术研究所） 报告题目：计算机视觉-从孤立到系统性方法	
10:50-11:10	铂金企业宣讲  图森未来宣讲报告： 题目：图森未来最新产品与业务介绍 宣讲人：王乃岩  华为宣讲报告： 题目：华为在 CV 和 AI 领域的探索实践分享 宣讲人：姚添宇	
11:10-12:10	年度进展评述（一）	
	讲者：吴建鑫（南京大学） 题目：神经网络模型轻量化设计	
	讲者：刘焯斌（清华大学） 题目：基于 NeRF 的三维视觉	
	讲者：张兆翔（中国科学院自动化研究所） 题目：类脑智能	
12:10-13:30	Poster & 午休	
13:30-14:00	大会特邀报告-1 报告人：朱军（清华大学） 报告题目：扩散概率模型的前沿进展	B6 馆
14:00-14:30	大会特邀报告-2 报告人：王井东（百度） 报告题目：CV 大模型 VIMER	
14:30-15:00	大会特邀报告-3 报告人：章国锋（浙江大学） 报告题目：基于隐式神经表示的三维重建、渲染与定位	

15:00-15:30	大会特邀报告-4 报告人：刘偲（北京航空航天大学） 报告题目：开放视觉感知	
15:30-15:50	铂金企业宣讲 OPPO 宣讲报告： 题目：OPPO 影像主张及技术创新的机遇与挑战 宣讲人：罗俊  AutoDL 宣讲报告： 题目：破解用卡难复现难新方案 宣讲人：余佳	
15:50-17:50	年度进展评述（二）	
	讲者：侯淇彬（南开大学） 题目：开放域目标检测/识别	
	讲者：王利民（南京大学） 题目：视频动作分析与识别	
	讲者：胡瀚（微软亚洲研究院） 题目：视觉自监督学习	
	讲者：程臻（西北工业大学） 题目：遥感目标检测	
	讲者：邓伟洪（北京邮电大学） 题目：人脸感知计算	
	讲者：邓成（西安电子科技大学） 题目：多模态融合感知	
	讲者：魏哲巍（中国人民大学） 题目：图表示学习	
	讲者：刘成林（中国科学院自动化研究所） 题目：文档智能	
17:50-18:00	合作单位颁奖	
18:30-20:00	VIP 晚宴	长广溪宾馆
<b>6月11日</b>		
8:30-12:15	Workshop 6: ChatGPT 与计算机视觉 讲者：吴飞（浙江大学）、白翔（华中科技大学）、车万翔（哈尔滨工业大学）、鲁伟明（浙江大学）、郑志彤（OPPO）、魏龙辉（华为）	B6 馆
	Workshop 7: 机器人具身智能 讲者：孙富春（清华大学）、蒋树强（中国科学院计算技术研究所）、卢策吾（上海交通大学）、丛杨（中国科学院沈阳自动化研究所）、付春江（优必选科技）、徐婧林（北京科技大学）、龚月（格灵深瞳）	B4 馆
	Workshop 10: 目标检测与分割 讲者：程明明（南开大学）、张兆翔（中国科学院自动化研究所）、张磊（重庆大学）、徐天阳（江南大学）、元玉慧（微软亚洲研究院）	B5 馆

	Workshop 3: 围绕手机的计算影像学 讲者: 张磊 (OPPO)、戴玉超 (西北工业大学)、贾旭 (大连理工大学)、薛天帆 (香港中文大学)、张健 (北京大学)	兰花厅 A
	Workshop 17: 视觉知识和多重知识表达 讲者: 王树徽 (中科院计算所)、付彦伟 (复旦大学)、王文冠 (浙江大学)、谢伟迪 (上海交通大学)、孙奕帆 (百度)	兰花厅 B
	Tutorial 4: 计算神经科学 讲者: 吴思 (北京大学)	梅花厅
12:15-13:30	午休	
13:30-18:00	Workshop 1: 大模型对 CV/PR 的挑战与机会 讲者: 田永鸿 (北京大学)、田奇 (华为)、刘静 (中国科学院自动化研究所)、黄高 (清华大学)、董胤蓬 (清华大学)、朱锡洲 (商汤科技)、张祥雨 (旷视科技)	B6 馆
	Workshop 14: 视觉内容生成 讲者: 卢志武 (中国人民大学)、王杨 (合肥工业大学)、齐勇刚 (北京邮电大学)、古纾旸 (微软亚洲研究院)、易冉 (上海交通大学)	B5 馆
	Workshop 8: 物联网与工业应用 讲者: 杨旸 (特斯联)、何源 (清华大学)、杨树森 (西安交通大学)、谢磊 (南京大学)、董玮 (浙江大学)、王晓飞 (天津大学)、丁二锐 (百度)	B4 馆
	Workshop 9: 混合现实中的视觉与学习技术 讲者: 许威威 (浙江大学)、王贝贝 (南京理工大学)、刘利斌 (北京大学)、潘烨 (上海交通大学)、许岚 (上海科技大学)	兰花厅 A
	Workshop 18: 女性研究者论坛 讲者: 邬霞 (北京师范大学)、景丽萍 (北京交通大学)、杨欣 (华中科技大学)、王珊珊 (中国科学院深圳先进技术研究院)、付莹 (北京理工大学)、霍静 (南京大学)、张曼 (北京邮电大学)	兰花厅 B
	Tutorial 1: 从 Transformer 到 GPT 讲者: 邱锡鹏 (复旦大学)、刘鹏飞 (上海交通大学)	梅花厅
	<b>6 月 12 日</b>	
8:30-12:15	Workshop 12: 多模态大模型与提示学习 讲者: 左旺孟 (哈尔滨工业大学)、余宙 (杭州电子科技大学)、王云鹤 (华为)、邵婧 (商汤科技)、朱霖潮 (浙江大学)	B6 馆
	Workshop 13: 三维视觉技术前沿 讲者: 任博 (南开大学)、杨佳琪 (西北工业大学)、申抒含 (中国科学院自动化研究所)、王鹤 (北京大学)、刘玉身 (清华大学)	B4 馆
	Tutorial 3: 讲者: 俞扬 (南京大学)	B5 馆

	Workshop 11: 大模型背景下的自动驾驶与车路协同 讲者: 李弘扬 (上海人工智能实验室)、李鸿升 (香港中文大学)、王乃岩 (图森未来)、赵行 (清华大学)、赵恒爽 (香港大学)、张骞 (地平线)	兰花厅 A
	Workshop 5: 语言文字识别与理解 金连文 (华南理工大学)、殷绪成 (北京科技大学)、姚聪 (阿里巴巴达摩院)、殷保才 (科大讯飞研究院)、苏劲松 (厦门大学)、刘禹良 (华中科技大学)	兰花厅 B
	Workshop 16: 人工智能与科学计算 讲者: 谢凌曦 (华为)、欧阳万里 (上海人工智能实验室)、严骏驰 (上海交通大学)、杨跃东 (中山大学)、苏航 (清华大学)	梅花厅
12:15-13:30	午休	
13:30-18:00	Workshop 15: 自监督视觉表征学习 讲者: 叶齐祥 (中国科学院大学)、纪荣嵘 (厦门大学)、张拯 (微软亚洲研究院)、孔涛 (字节跳动)、谢晋 (南京理工大学)	B6 馆
	Workshop 4: 多模态认知计算 讲者: 汪萌 (合肥工业大学)、李玺 (浙江大学)、郑锋 (南方科技大学)、武宇 (武汉大学)、周航 (百度)	B4 馆
	Tutorial 2: 扩散模型 讲者: 李崇轩 (中国人民大学)	B5 馆
	Workshop 2: 遥感影像解译 讲者: 孟德宇 (西安交通大学)、吴泽彬 (南京理工大学)、张洪艳 (中国地质大学(武汉))、陈思伟 (国防科技大学)、邹征夏 (北京航空航天大学)、姚西文 (西北工业大学)、杨学 (上海人工智能实验室)	兰花厅 A
	Workshop 19: 优秀学生论坛 讲者: 张博航 (北京大学)、许牧天 (香港中文大学-深圳)、陈涵晟 (同济大学)、杨啸 (清华大学)、董伯文 (哈尔滨工业大学)、杨木李 (西安电子科技大学)、黄振宇 (四川大学)、方浩树 (上海交通大学)、张曦予 (西北工业大学)	兰花厅 B
	Workshop 20: 开放域的学习方法与视觉任务 讲者: 代季峰 (清华大学)、段楠 (微软亚洲研究院)、张伟楠 (上海交通大学)、罗平 (香港大学)、张幸幸 (启元实验室)、王文海 (上海人工智能实验室)、梁小丹 (中山大学)	梅花厅

## Tutorial 与 Workshop 日程一览

时间地点	报告与讲者信息
<b>Tutorial 1: 从 Transformer 到 GPT</b> 6月11日 13:30-18:00 梅花厅	讲者: 邱锡鹏 (复旦大学) 题目: 从 Transformer 到 GPT
	讲者: 刘鹏飞 (上海交通大学) 题目: 从 Transformer 到 GPT
<b>Tutorial 2: 扩散模型</b> 6月12日 13:30-18:00 B4 馆	讲者: 李崇轩 (中国人民大学) 题目: 扩散模型
<b>Tutorial 3: Learning Environment Models for Reinforcement Learning</b> 6月12日 8:30-12:15 B4 馆	讲者: 俞扬 (南京大学) 题目: Learning Environment Models for Reinforcement Learning
<b>Tutorial 4: 计算神经科学</b> 6月11日 8:30-12:15 梅花厅	讲者: 吴思 (北京大学) 题目: 计算神经科学
<b>Workshop 1: 大模型对 CV/PR 的挑战与机会</b> 6月11日 13:30-18:00 B6 馆	主席: 代季峰 (清华大学) 张兆翔 (中国科学院自动化研究所) 操晓春 (中山大学)、王文海 (上海人工智能实验室)
	讲者: 田永鸿 (北京大学) 题目: 国产 E 级智算平台上视觉大模型训练及应用
	讲者: 田奇 (华为) 题目: 在大模型背景下, 重新审视计算机视觉的发展方向
	讲者: 刘静 (中国科学院自动化研究所) 题目: 多模态预训练的进展回顾与展望
	讲者: 黄高 (清华大学) 题目: 视觉数据多样性带来的机遇与挑战

时间地点	报告与讲者信息
	<p>讲者：董胤蓬（清华大学） 题目：视觉基础模型的安全性与可信性</p> <p>讲者：朱锡洲（商汤科技） 题目：面向多模态多任务的通用视觉感知大模型</p> <p>讲者：张祥雨（旷视科技） 题目：基于解耦学习的可逆纵列神经网络</p> <p>Panel 嘉宾：田奇（华为）、朱锡洲（商汤科技）、张祥雨（旷视科技）、刘静（中国科学院自动化研究所）、黄高（清华大学）、董胤蓬（清华大学）、田永鸿（北京大学）、操晓春（中山大学）、张兆翔（中国科学院自动化研究所）</p>
<p><b>Workshop 2: 遥感影像解译</b> 6月12日 13:30-18:00 兰花厅 A</p>	<p>主席：白翔（华中科技大学）程臻（西北工业大学）李军（中国地质大学（武汉））、孙显（中国科学院空天信息创新研究院）、方乐缘（湖南大学）</p> <p>讲者：孟德宇（西安交通大学） 题目：基于噪声和图像先验建模的高光谱图像复原研究</p> <p>讲者：吴泽彬（南京理工大学） 题目：云边协同的多源遥感融合高效处理</p> <p>讲者：张洪艳（中国地质大学（武汉）） 题目：基于弱监督深度学习框架绘制中国首幅 1 米分辨率土地覆盖一张图</p> <p>讲者：陈思伟（国防科技大学） 题目：极化 SAR 图像精细解译与应用</p> <p>讲者：邹征夏（北京航空航天大学） 题目：物理驱动的遥感图像合成与解译</p> <p>讲者：姚西文（西北工业大学） 题目：高分遥感图像弱监督目标检测</p> <p>讲者：杨学（上海交通大学） 题目：视觉有向目标的高精度检测</p> <p>Panel 嘉宾：孟德宇（西安交通大学）、吴泽彬（南京理工大学）、张洪艳（中国地质大学）、陈思伟（国防科技大学）、邹征夏（北京航空航天大学）、姚西文（西北工业大学）、杨学（上海交通大学）</p>
<p><b>Workshop 3: 围绕手机的计算影像学</b> 6月11日 8:30-12:15 兰花厅 A</p>	<p>主席：程明明（南开大学）王立君（大连理工大学）卢湖川（大连理工大学）、左旺孟（哈尔滨工业大学）、郭春乐（南开大学）</p> <p>讲者：张磊（OPPO） 题目：当 A (academia) 碰到 I (industry)：图像复原和增强任务中 AI 的困境和机遇</p>

时间地点	报告与讲者信息
	讲者：戴玉超（西北工业大学）
	题目：卷帘快门相机：模型、优化与学习
	讲者：贾旭（大连理工大学）
	题目：影像的时空分辨率增强
	讲者：薛天帆（香港中文大学）
题目：如何设计一个真正意义上的智能相机	
讲者：张健（北京大学）	
题目：Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model	
Panel 嘉宾：张磊（OPPO）、戴玉超（西北工业大学）、贾旭（大连理工大学）、薛天帆（香港中文大学）、张健（北京大学）、曹汛（南京大学）、杨文翰（鹏程实验室）、顾舒航（电子科技大学）	
<b>Workshop 4: 多模态认知计算</b> 6月12日 13:30-18:00 B5 馆	主席：姬艳丽（电子科技大学）胡迪（中国人民大学）田亚鹏（德克萨斯大学达拉斯分校）、张正（哈尔滨工业大学（深圳））
	讲者：汪萌（合肥工业大学）
	题目：基于智能交互的心理与情绪计算分析
	讲者：李玺（浙江大学）
	题目：多模态视觉结构建模和特征学习
	讲者：郑锋（南方科技大学）
	题目：音视文融合感知与生成
讲者：武宇（武汉大学）	
题目：多模态感知与跨模态生成	
讲者：周航（百度）	
题目：真人虚拟人的语音智能驱动	
Panel 嘉宾：汪萌（合肥工业大学）、李玺（浙江大学）、郑锋（南方科技大学）、李冠彬（中山大学）、武宇（武汉大学）、周航（百度）、张正（哈尔滨工业大学（深圳））、张晗（西北工业大学）	
<b>Workshop 5: 语言文字识别与理解</b> 6月12日 8:30-12:15 兰花厅 B	主席：白翔（华中科技大学）侯淇彬（南开大学）袁粒（北京大学）
	讲者：金连文（华南理工大学）
	题目：弱监督文字识别新思考
	讲者：殷绪成（北京科技大学）
	题目：开放集文字识别技术
讲者：姚聪（阿里巴巴达摩院）	
题目：面向文档理解的视觉-语言联合建模	
讲者：殷保才（科大讯飞研究院）	
题目：复杂场景下图文识别关键技术及产业化	

时间地点	报告与讲者信息
	<p>讲者：苏劲松（厦门大学） 题目：基于图文语义融合的多模态神经机器翻译研究</p> <p>讲者：刘禹良（华中科技大学） 题目：浅析多模态大型模型在文字图像智能领域的应用现状、挑战及未来展望</p> <p>Panel 嘉宾：金连文（华南理工大学）、殷绪成（北京科技大学）、姚聪（阿里巴巴达摩院）、殷保才（科大讯飞研究院）、苏劲松（厦门大学）、刘禹良（华中科技大学）</p>
<p><b>Workshop 6: ChatGPT 与计算机视觉</b></p> <p>6月11日 8:30-12:15 B6 馆</p>	<p>主席：谢凌曦（华为）张东霖（江南大学）李翔（南开大学）</p> <p>讲者：吴飞（浙江大学） 题目：ChatGPT：chat 的计算合成与涌现机理</p> <p>讲者：白翔（华中科技大学） 题目：再谈 ChatGPT 那些事儿</p> <p>讲者：车万翔（哈尔滨工业大学） 题目：ChatGPT——大语言模型向人类期望的对齐之路</p> <p>讲者：鲁伟明（浙江大学） 题目：多模型协同的计算机视觉处理：以 HuggingGPT 为例</p> <p>讲者：郑志彤（OPPO） 题目：借鉴 ChatGPT 思想，小布助手在 AIGC 上的探索</p> <p>讲者：魏龙辉（华为） 题目：图文多模态预训练和生成技术进展</p> <p>Panel 嘉宾：吴飞（浙江大学）、白翔（华中科技大学）、车万翔（哈尔滨工业大学）、鲁伟明（浙江大学）、杨振宇（OPPO）、魏龙辉（华为）</p>
<p><b>Workshop 7: 机器人具身智能</b></p> <p>6月11日 8:30-12:30 B5 馆</p>	<p>主席：弋力（清华大学）王永才（中国人民大学）杨旭（中国科学院自动化研究所）</p> <p>讲者：孙富春（清华大学） 题目：物理特性驱动的机器人认知学习</p> <p>讲者：蒋树强（中国科学院计算技术研究所） 题目：视觉导航：从模拟到现实</p> <p>讲者：卢策吾（上海交通大学） 题目：具身三维视觉 PIE 方案</p> <p>讲者：丛杨（中国科学院沈阳自动化研究所） 题目：3D 视觉引导的机器人自主操作</p> <p>讲者：付春江（优必选科技） 题目：人形机器人多模态应用</p>

时间地点	报告与讲者信息
	讲者：徐婧林（北京科技大学） 题目：视频人体动作的细粒度分析与理解
	讲者：龚月（格灵深瞳） 题目：速度与“机”情——智能巡检机器人在列车运维中的应用
	Panel 嘉宾：孙富春（清华大学）、蒋树强（中国科学院计算技术研究所）、卢策吾（上海交通大学）、丛杨（中国科学院沈阳自动化研究所）、徐婧林（北京科技大学）、龚月（格灵深瞳）、付春江（优必选科技）
<p><b>Workshop 8:</b> <b>物联网与工业应用</b> 6月11日 13:30-18:00 B4 馆</p>	<p>主席：吴帆（上海交通大学）陈旭（中山大学）黄倩怡（中山大学）</p> <p>讲者：杨旸（特斯联） 题目：面向用户算力需求特征的 6G 网络智能架构</p> <p>讲者：何源（清华大学） 题目：从应用视角看物联网技术发展——射频计算初探</p> <p>讲者：杨树森（西安交通大学） 题目：边云场景大数据计算：算法、平台与应用</p> <p>讲者：谢磊（南京大学） 题目：“工业视觉与边缘智能”助力新型电力系统</p> <p>讲者：董玮（浙江大学） 题目：大语言模型下的物联网低代码开发</p> <p>讲者：王晓飞（天津大学） 题目：云边缘泛在协同人工智能</p> <p>讲者：丁二锐（百度） 题目：数字人/物/场生成的产业化实践</p> <p>Panel 嘉宾：杨旸（特斯联）、何源（清华大学）、杨树森（西安交通大学）、谢磊（南京大学）、董玮（浙江大学）、王晓飞（天津大学）、丁二锐（百度）、陈旭（中山大学）</p>
<p><b>Workshop 9:</b> <b>混合现实中的视觉与学习技术</b> 6月11日 13:30-18:00 兰花厅 A</p>	<p>主席：周晓巍（浙江大学）胡瑞珍（深圳大学）</p> <p>讲者：许威威（浙江大学） 题目：神经网络渲染技术进展报告</p> <p>讲者：王贝贝（南京理工大学） 题目：真实感材质的表示与恢复</p> <p>讲者：刘利斌（北京大学） 题目：多模态和物理驱动的数字人动画</p> <p>讲者：潘焯（上海交通大学） 题目：混合现实环境中 3D 角色交互技术研发</p>

时间地点	报告与讲者信息
	<p>讲者：许岚（上海科技大学）            题目：神经辐射场和虚拟数字人</p> <p>Panel 嘉宾：许威威（浙江大学）、王贝贝（南京理工大学）、刘利斌（北京大学）、潘焯（上海交通大学）、许岚（上海科技大学）</p>
<p><b>Workshop 10: 目标检测与分割</b>            6月11日            8:30-12:10            B4 馆</p>	<p>主席：贾伟（合肥工业大学）李成龙（安徽大学）黄岩（中国科学院自动化研究所）丛润民（山东大学）</p> <p>讲者：程明明（南开大学）            题目：粒度自适应的图像感知技术</p> <p>讲者：张兆翔（中国科学院自动化研究所）            题目：基于多传感器融合的视觉物体检测与分割</p> <p>讲者：张磊（重庆大学）            题目：开放环境视觉感知之开放域目标检测</p> <p>讲者：徐天阳（江南大学）            题目：信息牵引下的视觉目标定位与识别</p> <p>讲者：元玉慧（微软亚洲研究院）            题目：统一计算机视觉识别任务的 DETR 网络结构</p> <p>Panel 嘉宾：程明明（南开大学）、张兆翔（中国科学院自动化研究所）、张磊（重庆大学）、徐天阳（江南大学）、元玉慧（微软亚洲研究院）</p>
<p><b>Workshop 11: 大模型背景下的自动驾驶与车路协同</b>            6月12日            8:30-12:15            兰花厅 A</p>	<p>主席：马超（上海交通大学）李镇（香港中文大学(深圳)）、张瑞茂（香港中文大学(深圳)）、李镇（牛津大学）</p> <p>讲者：李弘扬（上海人工智能实验室）            题目：道路结构认知 2.0 - 感知与高精地图之争</p> <p>讲者：李鸿升（香港中文大学）            题目：Towards More Accurate and Robust 3D Scene Perception with LiDAR Point Clouds</p> <p>讲者：王乃岩（图森未来）            题目：基于稀疏特征图的超远距离感知技术</p> <p>讲者：赵行（清华大学）            题目：视觉为中心的自动驾驶：3D 占用感知，运动预测与地图构建</p> <p>讲者：赵恒爽（香港大学）            题目：3D Scene Understanding: from Fundamental Design to Representation Learning and Future Frontiers</p> <p>讲者：张骞（地平线）            题目：BEV 新范式下的高阶自动驾驶技术</p>

时间地点	报告与讲者信息
	Panel 嘉宾: 李弘扬(上海人工智能实验室)、李鸿升(香港中文大学)、王乃岩(图森未来)、赵行(清华大学)、赵恒爽(香港大学)、张骞(地平线)
<b>Workshop 12: 多模态大模型与提示学习</b> 6月12日 8:30-12:15 B6馆	主席: 张利军(南京大学) 卢志武(中国人民大学) 牛玉磊(哥伦比亚大学)、丁明宇(加州大学伯克利分校)
	讲者: 左旺孟(哈尔滨工业大学) 题目: 预训练模型和语言增强的零样本视觉学习
	讲者: 余宙(杭州电子科技大学) 题目: 知识增强的多模态预训练和提示学习
	讲者: 王云鹤(华为) 题目: 多模态交织: 高效模型架构和融合方法
	讲者: 邵婧(商汤科技) 题目: 基于LLM的多模态提示学习: 框架、提示数据及评测标准
	讲者: 朱霖潮(浙江大学) 题目: 模态鸿沟与交互式提示学习
	Panel 嘉宾: 左旺孟(哈尔滨工业大学)、余宙(杭州电子科技大学)、王云鹤(华为)、邵婧(商汤科技)、朱霖潮(浙江大学)
<b>Workshop 13: 三维视觉技术前沿</b> 6月12日 8:30-12:15 B5馆	主席: 高盛华(华中科技大学) 戴玉超(天津大学) 王龙光(思谋科技)、崔兆鹏(浙江大学)
	讲者: 任博(南开大学) 题目: 含透明与反射物体的三维场景内容生成
	讲者: 杨佳琪(西北工业大学) 题目: 点亮传统方法: 几何三维点云配准方法新探索
	讲者: 申抒含(中国科学院自动化研究所) 题目: 大规模复杂场景三维重建与理解
	讲者: 王鹤(北京大学) 题目: 基于三维视觉的泛化物体操纵
	讲者: 刘玉身(清华大学) 题目: 三维智能处理与重建
	Panel 嘉宾: 任博(南开大学)、杨佳琪(西北工业大学)、申抒含(中国科学院自动化研究所)、王鹤(北京大学)、刘玉身(清华大学)
<b>Workshop 14: 视觉内容生成</b> 6月11日 13:30-18:00 B5馆	主席: 李崇轩(中国人民大学) 章国锋(浙江大学) 于茜(北京航空航天大学) 盛律(北京航空航天大学)
	讲者: 卢志武(中国人民大学) 题目: 多模态通用生成模型的基本框架与最新进展

时间地点	报告与讲者信息
	<p>讲者：王杨（合肥工业大学） 题目：基于知识蒸馏场景下的样本生成</p> <p>讲者：齐勇刚（北京邮电大学） 题目：跨模态生成视角下的抽象视觉理解</p> <p>讲者：古纾旸（微软亚洲研究院） 题目：From Paint by Example to Generalized Image Editing</p> <p>讲者：易冉（上海交通大学） 题目：面向艺术肖像画的媒体艺术生成与评估</p> <p>Panel 嘉宾：卢志武（中国人民大学）、王杨（合肥工业大学）、齐勇刚（北京邮电大学）、古纾旸（微软亚洲研究院）、易冉（上海交通大学）、盛律（北京航空航天大学）</p>
<p><b>Workshop 15:</b> <b>自监督视觉表征学习</b> 6月12日 13:30-18:00 B6 馆</p>	<p>主席：沈为（上海交通大学）王兴刚（华中科技大学）胡瀚（微软亚洲研究院）</p> <p>讲者：叶齐祥（中国科学院大学） 题目：视觉表征模型结构设计——从局部全局特征耦合的 Conformer 到全预训练模型 iTPN</p> <p>讲者：纪荣嵘（厦门大学） 题目：从 ChatGPT 到国产多模态基础大模型</p> <p>讲者：张拯（微软亚洲研究院） 题目：MIM 的理解，局限，提升与扩展</p> <p>讲者：孔涛（字节跳动） 题目：自监督表示学习与机器人操作</p> <p>讲者：谢晋（南京理工大学） 题目：自监督点云场景流估计</p> <p>Panel 嘉宾：卢策吾，王鹤，高阳，蒋树强，刘华平，郭彦东</p>
<p><b>Workshop 16:</b> <b>人工智能与科学计算</b> 6月12日 8:30-12:15 梅花厅</p>	<p>主席：王楠楠（西安电子科技大学）白磊（上海人工智能实验室）陈涛（复旦大学）</p> <p>讲者：谢凌曦（华为） 题目：盘古气象大模型：3D 全球高分辨率气象预报方法</p> <p>讲者：欧阳万里（上海人工智能实验室） 题目：从计算机视觉到 AI4Science-挑战与机遇</p> <p>讲者：严骏驰（上海交通大学） 题目：量子图机器学习及其化学应用初探</p> <p>讲者：杨跃东（中山大学） 题目：融合 HPC 和 AI 的药物分子设计</p>

时间地点	报告与讲者信息
	讲者：苏航（清华大学） 题目：非均匀物理数据分布的偏微分方程（PDEs）的通用求解算子 Panel 嘉宾：谢凌曦（华为）、欧阳万里（上海人工智能实验室）、严骏驰（上海交通大学）、杨跃东（中山大学）、苏航（清华大学）
<b>Workshop 17: 视觉知识和多重知识 表达</b> 6月11日 08:30-12:15 兰花厅 B	主席：朱霖潮（浙江大学）丛润民（山东大学）张鼎文（西北工业大学） 讲者：王树徽（中科院计算所） 题目：跨媒体多重知识获取与可信学习推理 讲者：付彦伟（复旦大学） 题目：先验信息引导的图片内容生成与编辑 讲者：王文冠（浙江大学） 题目：知识与数据双驱动的视觉感知 讲者：谢伟迪（上海交通大学） 题目：基于知识驱动的多模态表征学习 讲者：孙奕帆（百度） 题目：Exploring the Prototype and Scope of Visual Concept Panel 嘉宾：王树徽（中国科学院计算技术研究所）、付彦伟（复旦大学）、王文冠（浙江大学）、谢伟迪（上海交通大学）、孙奕帆（百度）
<b>Workshop 18: 女性研究者论坛</b> 6月11日 13:30-18:00 兰花厅 B	主席：雷柏英（深圳大学）张姗姗（南京理工大学）贾晔焯（山东大学） 讲者：邹霞（北京师范大学） 题目：时-空共变脑功能分析算法研究 讲者：景丽萍（北京交通大学） 题目：认知启发的目标跟踪探索 讲者：杨欣（华中科技大学） 题目：基于视觉的深度估计与定位 讲者：王珊珊（深圳先进技术研究院） 题目：知识驱动的机器学习智能医学成像与分析 讲者：付莹（北京理工大学） 题目：暗光图像视频增强 讲者：霍静（南京大学） 题目：跨模态图像合成关键技术 讲者：张曼（北京邮电大学） 题目：步态识别技术和应用

时间地点	报告与讲者信息
	Panel 嘉宾: 邹霞 (北京师范大学)、景丽萍 (北京交通大学)、杨欣 (华中科技大学)、王珊珊 (深圳先进技术研究院)、付莹 (北京理工大学)、霍静 (南京大学)、张曼 (北京邮电大学)
<b>Workshop 19: 优秀学生论坛</b> 6月12日 13:30-18:00 兰花厅 B	主席: 董胤蓬 (清华大学) 盛律 (北京航空航天大学) 邓志杰 (上海交通大学)
	讲者: 张博航 (北京大学) 题目: Rethinking the Expressive Power of GNNs via Graph Biconnectivity
	讲者: 许牧天 (香港中文大学-深圳) 题目: 基于遮蔽建模的三维室内场景自监督预训练方法及大规模多视角图片数据集
	讲者: 陈涵晟 (同济大学) 题目: 基于单阶段扩散 NeRF 的 3D 生成与重建
	讲者: 杨啸 (清华大学) 题目: Towards Effective Adversarial Textured 3D Meshes on Physical Face Recognition
	讲者: 董伯文 (哈尔滨工业大学) 题目: Self-Promoted Supervision for Few-Shot Transformer
	讲者: 杨木李 (西安电子科技大学) 题目: Towards Distribution-Agnostic Novel Class Discovery
	讲者: 黄振宇 (四川大学) 题目: Noisy Correspondence: A New Paradigm for Learning with Noisy Label
	讲者: 方浩树 (上海交通大学) 题目: 人类级别的通用机器人抓取
	讲者: 张曦予 (西北工业大学) 题目: 3D Registration with Maximal Cliques
<b>Workshop 20: 开放城的学习方法与视觉任务</b> 6月12日 13:30-18:00	Panel 嘉宾: 张博航 (北京大学)、许牧天 (香港中文大学-深圳)、陈涵晟 (同济大学)、杨啸 (清华大学)、董伯文 (哈尔滨工业大学)、杨木李 (西安电子科技大学)、黄振宇 (四川大学)、方浩树 (上海交通大学)、张曦予 (西北工业大学)
	主席: 苏航 (清华大学) 徐航 (华为诺亚方舟实验室) 张磊 (重庆大学) 官辰 (南京理工大学) 讲者: 代季峰 (清华大学) 题目: 开放域物体识别算法

时间地点	报告与讲者信息
梅花厅	讲者：段楠（微软亚洲研究院） 题目：从视觉内容生成到视觉任务自动化
	讲者：张伟楠（上海交通大学） 题目：面向开放域的虚实迭代强化学习
	讲者：梁小丹（中山大学） 题目：多模态开放域检测大模型及应用
	讲者：罗平（香港大学） 题目：混合大模型
	讲者：张幸幸（启元实验室） 题目：动态开放环境的具身持续学习方法
	讲者：王文海（上海人工智能实验室） 题目：大规模基础模型在开放世界任务中的研究与应用
	Panel 嘉宾：代季峰（清华大学）、段楠（微软亚洲研究院）、张伟楠（上海交通大学）、梁小丹（中山大学）、罗平（香港大学）、张幸幸（启元实验室）、王文海（上海人工智能实验室）

## 大会报告及讲者简介

**高文 北京大学**

**报告题目：**特征编码与数字视网膜



**讲者简介：**中国工程院院士，鹏城实验室主任，北京大学信息与工程科学部主任、博雅讲席教授，国际电气和电子工程师协会会员（IEEE Fellow）、美国计算机协会会员（ACM Fellow）。现任第十四届全国人大代表，曾任第十届、十一届、十二届全国政协委员，国务院学科评议组计算机学科成员，中国计算机协会理事长，计算机学报主编。以第一完成人身份，一次获得国家技术发明一等奖、一次获得国家技术发明二等奖、五次获得国家科技进步二等奖。荣获全国五一劳动奖章（2023年）、何梁何利基金科学与技术进步奖

（2022年）、广东省南粤突出贡献奖（2021年）、“2005中国十大教育英才”称号和中国计算机学会王选奖。主要从事人工智能应用和多媒体技术、计算机视觉、模式识别与图像处理、虚拟现实方面的研究，主要著作有《数字视频编码技术原理》、《Advanced Video Coding Systems》等。在本领域国际期刊上发表论文 300 余篇。

**焦李成 西安电子科技大学**

**报告题目：**下一代深度学习的思考与若干问题

**报告摘要：**尽管深度学习技术得到了长足的发展，并在诸多领域取得了显著成果。然而发展过程中，存在诸多理论问题需要进一步深入的研究和关注。因此，本报告着重和大家一起探讨深度学习基础理论相关的研究。首先，回顾了深度学习的思想起源与发展历程。紧接着，讨论了对深度学习再认识与再思考，从而引出应突破的基础理论。然后，从类脑启发、物理启发和进化启发等三个方面讨论了深度学习的表征、学习与优化理论。最后，给出了对下一代深度学习的一些思考。



**讲者简介：**焦李成，欧洲科学院外籍院士，俄罗斯自然科学院外籍院士，IEEE Fellow。现任西安电子科技大学华山杰出教授、计算机科学与技术学部主任、人工智能研究院院长、智能感知与图像理解教育部重点实验室主任、教育部科技委学部委员、教育部人工智能科技创新专家组专家、国家级领军人才首批入选者、教育部长江学者计划创新团队负责人、“一带一路”人工智能创新联盟理事长，陕西省人工智能产业技术创新战略联盟理事长，中国人工智能学会第六-七届副理事长，IEEE/IET/CAA/CAA/CIE/CCF Fellow，连续八年

入选爱思唯尔高被引学者榜单。主要研究方向为智能感知与图像理解、深度学习与类脑计算、进化优化与遥感解译。曾获国家自然科学基金二等奖、吴文俊人工智能杰出贡献奖、霍英东青年教师奖、全国模范教师称号、中国青年科技奖、及省部级一等奖以上科技奖励十余项。

**陈熙霖 中国科学院计算技术研究所**

**报告题目：**计算机视觉-从孤立到系统性方法

**报告摘要：**在开放世界中通过观察积累、类比模仿到推理解、交互尝试直至干预反馈是

高等动物从外界获得经验与知识的基本手段与过程。在 AI 领域，很长时间以来的研究范式是以孤立算法为核心的单点研究，同时，现实世界中广泛存在着样本分布不均、任务多样性等问题。对以往的孤立研究范式而言，这些问题显然是难以克服的困难，因此需要从系统化的角度探索融合多模态信息，构建从感到探、从被动到主动的系统性学习体系。本报告将介绍我们近期在这方面的一些思考和尝试，探索从系统性的角度实现连续积累和学习的路径。



**讲者简介：**陈熙霖，中国科学院计算技术研究所研究员，ACM/IEEE / IAPR Fellow，中国计算机学会会士。主要研究领域为计算机视觉、模式识别、多媒体技术以及多模式人机接口。先后主持多项自然科学基金重大、重点项目、国家杰出青年基金、973 计划课题等的研究。现任/曾任 IEEE Trans. on Image Processing 和 IEEE Trans. on Multimedia 的 Associate Editor、Journal of Visual Communication and Image Representation 的 Senior Associate Editor，以及计算机学报副主编、人工智能与模式识别副主编等。担任过 IEEE FG 2013 / IEEE FG 2018 / IEEE VCIP 2022 等大会主席，并十多次担任 CVPR、ICCV、ECCV、NeurIPS 等会议的领域主席。研究工作先后获得国家自然科学二等奖 1 项，国家科技进步二等奖 4 项，在相关领域的重要刊物和会议上发表论文 300 多篇。

## 朱军 清华大学

**报告题目：**扩散概率模型的前沿进展

**报告摘要：**AIGC 发展迅速，扩散概率模型是 AIGC 的关键技术之一，在文图生成、3D 生成等方面取得显著进展。该报告将介绍扩散概率模型的若干进展，包括扩散概率模型的基础理论和高效算法、大规模多模态扩散模型以及 3D 生成等内容。



**讲者简介：**朱军，清华大学计算机系 Bosch AI 教授、IEEE Fellow，清华大学人工智能研究院副院长，曾任卡内基梅隆大学兼职教授。2001-2009 年获清华大学学士和博士学位，主要从事机器学习研究，担任国际著名期刊 IEEE TPAMI 的副主编，担任 ICML、NeurIPS、ICLR 等（资深）领域主席 20 余次。获求是杰出青年奖、科学探索奖、中国计算机学会自然科学一等奖、吴文俊人工智能自然科学一等奖、ICLR 国际会议杰出论文奖等，入选万人计划领军人才、中国计算机学会青年科学家、MIT TR35 中国先锋者等。

## 王井东 百度

**报告题目：**CV 大模型 VIMER

**报告摘要：**本报告主要内容包括 CV 大模型 VIMER 以数据为中心的算法和应用。首先，介绍自监督表征学习算法 Context Autoencoder (CAE) 算法及其推广。然后，讲述基于 CAE 的工业视觉大模型、OCR 文字识别大模型、人体大模型等。最后分享图文对比预训练大模型在自动驾驶数据挖掘中的应用和基于多任务学习的交通感知大模型。



**讲者简介:** Jingdong Wang is Chief Scientist for computer vision with Baidu. Before joining Baidu, he was a Senior Principal Researcher at Microsoft Research Asia from September 2007 to August 2021. His areas of interest include vision foundation models, self-supervised pretraining, OCR, neural architecture design, human pose estimation, semantic segmentation, image classification, object detection, and large-scale indexing. He has been serving/served as an Associate Editor of IEEE TPAMI, IJCV, IEEE TMM, and IEEE TCSVT, and an area chair of leading conferences in vision, multimedia, and AI, such as CVPR, ICCV, ECCV, ACM MM, IJCAI, and AAAI. He was elected as an ACM Distinguished Member, a Fellow of IAPR, and a Fellow of IEEE, for his contributions to visual content understanding and retrieval.

### 章国锋 浙江大学

**报告题目:** 基于隐式神经表示的三维重建、渲染与定位

**报告摘要:** 近年来,以 NeRF 为代表的基于隐式神经表示的三维建模与渲染取得了很大的成功,引起了业界的广泛关注。随着技术的不断发展,基于 NeRF 的神经表达和渲染也从单一的场景重建和新视角生成,逐渐拓宽到场景中物体解耦的建模和可编辑渲染、精细化的场景/物体编辑,乃至全场三维理解、光照自适应建模等多个方面,而且也出现了基于隐式神经表示的 SLAM 方法,同时兼顾了稠密重建和定位。本报告将重点汇报我们团队近两年在这方面的研究进展,并对未来发展方向进行探讨和展望。



**讲者简介:** 章国锋,浙江大学教授,博导,国家优秀青年科学基金获得者。主要从事三维视觉与增强现实方面的研究,尤其在同步定位与地图构建(SLAM)和三维重建方面取得了一系列重要成果,研制了一系列相关软件

(<http://www.zjucvg.net>),并开源了一系列相关软件

(<https://github.com/zju3dv/>),是 OpenXR Lab 扩展现实开源平台(<https://openxrlab.org/cn/>)的主要发起人之一。获全国百篇优秀博士学位论文奖、计算机学会优秀博士学位论文奖、教育部高等学校科学研究优秀成果

奖科学技术进步奖一等奖(排名第4)、浙江省自然科学奖一等奖(排名第2)、浙江省技术发明奖一等奖(排名第4)以及混合现实和增强现实领域国际顶级会议 ISMAR 2020 唯一最佳论文奖。目前为《Virtual Reality & Intelligent Hardware》和《应用科学学报》编委,《中国图象图形学报》青年编委,中国图象图形学学会三维视觉专委会副主任,浙江省人工智能学会增强现实分会副会长;担任 VALSE 2019、2021 大会程序委员会主席, VALSE 2022 大会主席, ChinaVR 2021 大会程序委员会主席, CVPR 2021、2023 领域主席, ISMAR 2019-2022 以及 VR 2021-2023 程序委员会委员。

### 刘偲 北京航空航天大学

**报告题目:** 开放视觉感知

**报告摘要:** 随着视觉-语言大模型的快速发展,视觉感知技术逐渐将应用领域拓展到开放

环境，以突破限定数据标签的限制，提升对未知类别物体的感知能力。本报告主要介绍开放环境下目标检测、分割、跟踪、关系检测等感知技术的相关进展及潜在应用。



**讲者简介：**刘懿，北航教授，博导。主持国家优秀青年科学基金。博士毕业于中科院自动化所，曾于新加坡国立大学任 RA 和博后，曾任微软亚洲研究院铸星计划研究员。研究方向是跨模态智能分析、目标检测和跟踪。共发表了 CCF A 类论文 80 余篇。Google Scholar 引用 10000+次。获 CCF-A 类会 ACM MM 2012 最佳技术演示奖，ACM MM 2013、ACM MM 2021 最佳论文奖，以及 IJCAI 2021 最佳视频奖。担任中国图象图形学学会理事、副秘书长。多次担任 ICCV、CVPR、ECCV、

NeurIPS、ACM MM 等会议 AC。担任 IEEE TMM、IEEE TCSVT、CVIU 等期刊 AE。获得 10+项 CVPR、ICCV、ACL 等国际顶级竞赛冠军。

## 年度进展评述（APR）及讲者简介

**刘成林 中国科学院自动化研究所**

**报告题目：**文档智能

**报告摘要：**文档智能是对各种文档图像进行分析、识别和理解的技术，包括文档图像处理、版面分析、内容识别、语义理解等研究内容和任务。内容识别又包括文本检测与识别、图形符号（流程图、公式、图表等）识别、表格分析等任务。文本识别有悠久的历史，近年来方法和性能还在不断发展，其他任务受到越来越多的关注。尤其是结合文本语言信息和文档结构信息的结构化文本识别、信息抽取、文档问答等任务（这些属于文档理解范畴）成为新的研究热点，并产生了一些相应的语言预训练模型或跨模态预训练模型。本报告主要介绍过去两年在文本检测与识别、版面分析、公式和流程图识别等文档识别任务和表格分析、信息抽取、文档问答等文档理解任务上的研究进展，并分析未来发展趋势。



**讲者简介：**刘成林，中国科学院自动化研究所副所长，模式识别国家重点实验室主任，研究员、博士生导师，中国科学院大学人工智能学院副院长。1989年、1992年、1995年分别在武汉大学、北京工业大学、中国科学院自动化研究所获学士、硕士和博士学位。1996年至2004年先后在韩国科学技术院、日本东京农工大学、日立中央研究所从事博士后和研发工作。2005年起在中国科学院自动化研究所任研究员。2008年获得国家杰出青年科学基金。研究兴趣包括模式识别、机器学习、文字识别与文档分析等。在国内外期刊和学术会议上发表论文300余篇，合著英文专著一本。现任 Pattern Recognition 期刊和《自动

化学报》的副主编，以及多个期刊的编委。任国际模式识别学会副主席，中国人工智能学会副理事长、会士，中国自动化学会会士、模式识别与智能系统专委会主任，中国图象图形学学会常务理事。美国电气电子工程师协会会员（IEEE Fellow）、国际模式识别学会会士（IAPR Fellow）。

**吴建鑫 南京大学**

**报告题目：**神经网络模型轻量化设计

**报告摘要：**在深度神经网络取得各种突破的同时，其高存储和计算成本也要求对神经网络的推理进行加速，并对模型本身瘦身（即轻量化）。本年度回顾将介绍近期神经网络模型在推理加速与模型轻量化方面的进展，包括本方向中相对较新的研究问题，以及模型结构、知识蒸馏等相关技术在这一年的进展。



**讲者简介：**吴建鑫，南京大学人工智能学院教授，研究方向为计算机视觉与深度学习，尤其是计算资源与数据资源受限情景中的深度学习，如模型压缩与加速、小样本学习、半监督物体定位与检测等。曾任或现任 PAMI 等期刊的编委及 CVPR 等会议的程序主席、资深领域主席、领域主席。

## 刘焯斌 清华大学

**报告题目：**基于 NeRF 的三维视觉

**报告摘要：**神经辐射场(NeRF)是一种以隐式场和体渲染为基础的三维表征，以其端到端可微、高质量视点生成等特性得到广泛关注。自 NeRF 被提出以来，学者对其隐式场本身或体渲染过程进行了诸多改进，以实现加速推理和训练、几何与表观解耦、材质和光照编辑乃至稀疏视点下的动静态和多尺度场景建模。与此同时，通过结合多元表征和生成式模型，NeRF 在三维视觉领域的应用层出不穷。本报告将回顾过去一年神经辐射场的重要研究成果，涵盖其表征基础的优化及代表性应用研究，重点将围绕 NeRF 现存的两大挑战，包括在轻量化采集条件下的高质量三维重建与渲染，以及将 NeRF 拓展到时空动态场景的高效四维表征来进行探讨与展望。



**讲者简介：**刘焯斌，清华大学自动化系特聘教授，国家基金委杰青。研究方向为三维视觉，主要从事数字人重建与生成研究。提出了三维数字人的多元表征模型与协同优化机理，突破了单一表征难以兼顾实时性、精确性和便捷性的矛盾，研制了多套国际领先的单一视点或多视点实时动态三维重建与生成系统，发表 TPAMI/SIGGRAPH/CVPR/ICCV 等论文近 70 篇，成果对 8 家知名企业技术许可。担任 IEEE TVCG 及 CGF 副主编，CVPR 2021、ICCV 2021、ECCV 2022 领域主席，SIGGRAPH Asia 2019, 2020

技术委员会委员。任中国图象图形学会三维视觉专委会副主任。获 2012 年国家技术发明一等奖（排名 3），2019 年中国电子学会技术发明一等奖（排名 1）。

## 张兆翔 中国科学院自动化研究所

**报告题目：**类脑智能

**报告摘要：**近年来，人工智能取得长足进展，尤其是以大数据、大算力、大模型为主要代表的方法在诸多感知与学习问题上取得突破，对整个学科带来重要冲击。与此同时，其能耗问题、标注问题、长尾问题等依旧存在局限，与人类智能相比依旧不可同日而语。向人脑学习，开展类脑智能研究，具有重要理论意义与应用价值。本报告对 2022 年以来的类脑智能相关方面，包括脉冲神经网络、类脑学习算法、类脑方法新应用等进行梳理，对类脑智能相关的进展进行回顾，并对未来进行展望。



**讲者简介：**张兆翔，中国科学院自动化研究所研究员、博士生导师，入选教育部长江学者，国家万人计划青年拔尖人才等。主要研究方向包括脑启发的神经网络建模、视觉认知学习、面向开放环境的场景感知与理解，在本领域 TPAMI、IJCV、JMLR 等顶刊和 CVPR、ICCV、ICLR、NeuralPS 等顶会发表论文 100 余篇，授权专利 20 余项，承担了国家自然科学基金重点项目、国家自然科学基金企业联合重点项目、国家重点研发项目等一系列国家级科研项目，是 IEEE 高级会员，中国计算机学会 CCF 杰出会员、中国人工智能学会 CAAI 杰出会员、中国计算机学会 CCF 杰出演讲者，担任或曾

担任 IEEE T-CSVT、IEEE T-BIOM、Pattern Recognition 等知名期刊编委，是 CVPR、ICCV、NeurIPS、AAAI、IJCAI、ACM MM 等知名国际会议的领域主席（Area Chair）。

## 夏勇 西北工业大学

**报告题目：**医学影像检测与分割



**讲者简介：**夏勇，西北工业大学计算机学院教授，研究方向为医学影像智能计算，近三年在 JAMA Network Open、IEEE-TPAMI/TMI/TIP、MedIA、NeurIPS、CVPR、MICCAI 发表论文 40 余篇，被引用近 9000 次(Google Scholar)，先后在 PROMISE12、ISBI 2019 C-NMC、BraTS2020/2021、MyoPS 2020、COVID-19 2020、KiTS21、COSMOS 2022、KiPA 22 等国际学科竞赛中名列前茅；担任中国体视学学会理事、中国图象图形学学会视觉大数据专委会常委、中国抗癌协会肿瘤影像专业委员会人工智能学组副组长等，先后担任 ISBI2017、MICCAI 2019/2020、ICASSP 2023 地区主席或分会主席等。个人主页：

<https://teacher.nwpu.edu.cn/yonxia.html>

## 侯淇彬 南开大学

**报告题目：**开放域目标检测/识别

**报告摘要：**开放域目标检测/识别已成为计算机视觉领域一大研究热点，其目的是借助于现有跨模态模型的泛化能力，实现对已定义类别物体的 few-shot 检测和对未定义类别的 zero-shot 检测。从功能角度说，开放域目标检测/识别是比弱监督和 zero-shot 方法更通用、更有效的目标检测任务。本次报告将概述过去一年开放域目标检测/识别领域的主要进展，并对未来研究方向做出展望。



**讲者简介：**侯淇彬，副教授，南开大学“百青计划”入选者，博士生导师。于 2019 年在南开大学获得博士学位，之后在新加坡国立大学从事博士后研究工作，主要研究方向为基础模型、分割及检测。目前已在 IEEE TPAMI / NeurIPS / CVPR / ICCV 等国际顶级期刊及会议上发表论文 30 余篇，Google Scholar 引用 7000+，两篇一作论文引用超过 1000，研究成果应用于华为旗舰手机智能拍照系统中并获得了华为公司 2017 年度优秀合作成果奖，入选第七届中国科协青年人才托举计划，曾获“吴文俊人工智能科学技术奖”自然科学奖二等奖等奖项。

## 王利民 南京大学

**报告题目：**视频动作与分析识别

**报告摘要：**视频动作分析与识别是视频理解领域的关键技术，其主要研究内容主要包括：视频动作识别、时序动作检测、时空动作检测等等。在过去两年中，首先构建通用的视频基础模型已经成为动作识别任务的重要性能增长点，本次报告将回顾视频基础模型架构（从 CNN 到 Transformer）和学习范式（从有监督到自监督），重点介绍视频 Transformer 架构和 VideoMAE 预训练范式；其次针对视频动作分析下游任务，简洁高效的分析框架已经成为复杂场景下动作理解关键技术，本次报告将介绍端到端的长时动作分析框架（PointTAD, BasicTAD）和一阶段时空动作检测框架（MOC, STMixer）。最后本报告还将简单回顾最新顶会关于视频动作分析与识别的前沿进展。



**讲者简介：**王利民，南京大学教授，博士生导师，国家海外高层次人才计划入选者，科技创新 2030-“新一代人工智能”重大项目青年科学家。2011 年在南京大学获得学士学位，2015 年香港中文获得博士学位，2015 年至 2018 年在苏黎世联邦理工学院（ETH Zurich）从事博士后研究工作。主要研究领域为计算机视觉和深度学习，专注视频理解和动作识别，在 IJCV、T-PAMI、CVPR、ICCV 等重要学术期刊和会议发表论文 70 余篇。根据 Google Scholar 统计，论文被引用 17000 余次，两篇一作论文取得了单篇引用超过 3000 的学术影响力。提出的 TSN 网络获得首届 ActivityNet 比赛冠军，已经成为动作识别领域基准方法。曾获得广东省技术发明一等奖，世界人工智能大会青年优秀论文奖。入选 2022-2023 年度 AI 2000 人工智能全球最具影响力学者榜单（计算机视觉方向），2022 年度全球华人 AI 青年学者榜单，2021-2022 年度爱思唯尔中国高被引学者榜单。

### 胡瀚 微软亚洲研究院

**报告题目：**视觉自监督学习

**报告摘要：**视觉自监督学习的主流范式在过去一年多的时间里经历了从对比学习方法到生成式方法的迁移。以 BEiT/MAE/SwinV2(SimMIM)为代表的生成式方法在预训练-微调范式下取得了优异的性能，更重要的是，它们被证明相比此前的方法具备更好的数据和模型可扩展性，也能很好的与多模态方法融合。本次 APR 将概述过去一年视觉自监督学习方面的主要进展，包括预训练方法本身及其相关性质的研究。



**讲者简介：**胡瀚，微软亚洲研究院首席研究员，西安交大兼职博士生导师。目前主要从事大模型相关的研究工作，其论文 Swin Transformer 荣获 ICCV2021 最佳论文（马尔奖），并被学界和业界广泛使用。他有 6 篇论文入选 PaperDigest Most Influential Papers (top-15) 榜单，代表性工作还包括可变形卷积系列、关系网络系列、SimMIM 等，它们促进了计算机视觉和自然语言处理在建模和学习上趋向统一。他在清华大学自动化系分别获得博士和本科学位，博士论文获得中国人工智能学会优博奖。曾多次担任计算机视觉领域顶级会议的领域主席。谷歌学术引用近 2 万次。

### 程焱 西北工业大学

**报告题目：**遥感目标检测

**报告摘要：**遥感目标检测是空天地海一体化观测系统的一项关键技术。与自然图像相比，遥感图像具有目标方向多变、目标类型及数量繁杂、特定领域样本稀缺、成像视角单一等特点，此外，不同平台、不同光照、天气条件、大气参数等都会对遥感图像获取产生影响。这些综合因素使得遥感目标检测面临着更大的挑战和更多的难点问题。本报告首先总结分析遥感目标检测面临的挑战，接下来重点概述近年来遥感目标检测的主要进展，主要包括有向目标检测、弱监督目标检测、小样本目标检测、目标型号识别、以及弱小目标检测等。



**讲者简介：**程建，西北工业大学长聘教授，博士生导师，信息融合技术教育部重点实验室副主任，入选国家“万人计划”青年拔尖人才，连续3年入选科睿唯安“全球高被引科学家”和爱思唯尔“中国高被引学者”。主要研究方向为光学遥感图像理解、计算机视觉等。以第一作者/通讯作者发表论文60余篇，包括PIEEE、TPAMI、CVPR、ICCV等，谷歌学术总引用1.3万余次，4篇第一作者论文单篇引用大于1000次，3篇论文入选年度中国百篇最具影响国际学术论文，获得2021年度IEEE TCSVT最佳论文奖、2021年度和2023年度IEEE地球科学与遥感学会最高影响力论文奖（IEEE GRSS Highest Impact Paper Award）等学术奖励，

获得吴文俊人工智能技术发明一等奖、陕西省科学技术一等奖等4项省部级科技奖励，担任IEEE GRSM、ISPRS JPRS、JRS等多个国际期刊编委。

### 邓伟洪 北京邮电大学

**报告题目：**人脸感知计算

**报告摘要：**古希腊哲学家西塞罗说：“世间的一切都写在脸上。”我们在脸上能够看到的信息，远远不止“TA是谁”。本次报告将概述近期人脸感知计算的主要进展，包括从监督学习到噪声标签学习和自监督预训练的理论研究进展，从身份属性的识别到生理指标估计、视线估计、唇语和微表情识别、疾病筛查的技术应用进展；从识别准确率的刷分到可解释性、鲁棒性、公平性和隐私保护的伦理研究进展等等。



**讲者简介：**邓伟洪，北京邮电大学人工智能学院教授，教育部青年长江学者，研究方向为计算机视觉与模式识别、可信人工智能、情感计算、多模态学习。近年来主持国家重点研发计划课题、国家自然科学基金等项目20多项，在IEEE TPAMI、IJCV、TIP、ICCV、CVPR等国际期刊和会议上发表论文100多篇，谷歌学术引用1.2万余次，曾担任ACM MM、ECCV、IJCAI、ICME、ICPR等会议的领域主席，曾入选北京市优秀博士学位论文、北京市科技新星、教育部新世纪优秀人才、Elsevier中国高被引学者等。

### 邓成 西安电子科技大学

**报告题目：**多模态融合感知

**报告摘要：**多模态融合能够结合不同的传感器来处理 and 关联物理世界中的数据，充分模拟人脑的信息感知机理。早期的多模态学习主要体现在视觉、文本模态在特征表示与感知结果等层面的融合。近年来，随着CLIP、ChatGPT、GPT4等大模型的飞速发展，以多模态融合技术为核心的感知、交互和智慧协同能力，不断支撑各类终端和应用的智能化水平提升。本次报告将概述过去一年多模态融合的主要进展，包括多模态融合表示相关理论、多模态预训练模型发展，以及相关的下游任务的最新成果，主要分为理解式任务(视觉问答、视觉推理、图文检索等)和生成式任务(图文生成、音视频生成等)两个部分。



**讲者简介:** 西安电子科技大学教授、博导, 国家级人才、国家“百千万人才工程”入选者, 中国计算机学会杰出会员。长期从事多模态智能、机器学习等领域的研究。主持国家自然科学基金重点项目等项目近 30 项。在国际 TOP 期刊和 CCFA 类国际会议上发表论文 150 余篇。担任《Pattern Recognition》等国际刊物副编辑; 担任国际顶级会议 CVPR 2021、CVPR2023、ICCV 2021、NeurIPS2023 领域主席。入围爱思唯尔中国高被引学者榜单和全球前 2% 顶尖科学家榜单。获 2016 年国家自然科学基金二等奖、2019 年陕西省自然科学一等奖。

### 魏哲巍 中国人民大学

**报告题目:** 图表示学习

**报告摘要:** 图表示学习是一类基于机器学习技术处理图数据的方法, 在众多领域展现出了卓越的性能, 其已经成为一种广泛应用的图分析方法。通常, 图表示学习包含了图神经网络、图嵌入表示、图自监督学习等多种方法, 这些方法已被广泛应用于社交网络、蛋白质结构预测、交通流量预测等许多领域。本次报告将概述过去一年图表示学习的主要进展, 包括图神经网络、图嵌入表示、图自监督和图 Transformer 等方向的最新方法, 以及对未来工作的一些展望。



**讲者简介:** 魏哲巍, 中国人民大学高瓴人工智能学院教授, 博导, 入选国家高层次青年人才。2008 年本科毕业于北京大学数学科学学院, 2012 年博士毕业于香港科技大学计算机系; 2012 年至 2014 年于奥胡斯大学海量数据算法研究中心担任博士后研究员, 2014 年 9 月加入中国人民大学担任副教授, 2019 年 8 月起任教授。数据库、理论计算机、数据挖掘、机器学习等领域的顶级会议及期刊上 (如 SIGMOD、VLDB、ICML、NeurIPS、KDD、SODA 等) 发表论文 60 余篇, 并获得数据库理论顶级会议 PODS2022 时间检验奖。主持国家自然科学基金重点项目、科技部 2030 重点专项课题。担任

PODS、ICDT 等大数据理论会议论文集主席以及 ICML、NeurIPS、WWW 领域主席; 担任鹏城实验室广州基地青年科学家。培养博士生获 2021 年百度奖学金 (全球 10 人)、2022 年微软学者 (亚太地区 12 人)。

## Tutorial 报告及讲者简介

时间	内容	地点
6月11日 (周日) 13:30-18:00	<b>Tutorial 1: 从 Transformer 到 GPT</b>	梅花厅
	报告题目: 从 Transformer 到 GPT 讲者: 邱锡鹏 (复旦大学) 报告题目: 从 Transformer 到 GPT 讲者: 刘鹏飞 (上海交通大学)	
6月12日 (周一) 13:30-18:00	<b>Tutorial 2: 扩散模型</b>	B5 馆
	报告题目: 扩散模型 讲者: 李崇轩 (中国人民大学)	
6月12日 (周一) 8:30-12:15	<b>Tutorial 3: Learning Environment Models for Reinforcement Learning</b>	B5 馆
	报告题目: Learning Environment Models for Reinforcement Learning 讲者: 俞扬 (南京大学)	
6月11日 (周日) 8:30-12:15	<b>Tutorial 4: 计算神经科学</b>	梅花厅
	报告题目: 计算神经科学 讲者: 吴思 (北京大学)	

## Tutorial 1: 从 Transformer 到 GPT

邱锡鹏 复旦大学

刘鹏飞 上海交通大学

**报告题目:** 从 Transformer 到 GPT

**报告摘要:** 目前 Transformer 在自然语言处理、计算机视觉领域取得了广泛的成功, 以其为基础构建的预训练模型 GPT 也带来了前所未有的计算能力, 不仅变革了艺术创造的形式, 也将赋能每一位人, 改变人与数据、工具的交互方式。本报告主要介绍 (1) Transformer 模型以及变体; (2) 预训练模型主要核心框架; (3) 提示学习的相关概念基础和应用。最后, 对大模型预训练及其未来发展趋势进行展望。



**讲者简介:** 邱锡鹏, 复旦大学计算机学院教授, 国家优青获得者, 于复旦大学获得理学学士和博士学位。主要从事自然语言处理、深度学习等方向的研究, 发表 CCF A/B 类论文 70 余篇, 获得 ACL 2017 杰出论文奖 (CCF A 类), CCL 2019 最佳论文奖, 有 4 篇论文入选 PaperDigest 发布的 IJCAI/ACL/EMNLP 的最有影响力论文 (各会议每年 10 篇)。出版开源专著《神经网络与深度学习》, Github 关注数 1.4 万, 豆瓣评分 9.4 分。主持开发了开源框架 FudanNLP 和 FastNLP, 已被国内外数百家单位使用。2015 年入选首届中国科协青年人才托举工程项目, 2018 年获钱伟长中文信息处理科学技术奖青年创新奖一等奖等。培养学生曾获中国中文信息学会优博、中国人工智能学会优博、上海市优博、微软学者、百度奖学金等。



**讲者简介:** 刘鹏飞, 上海交通大学清源研究院长聘教轨副教授, 生成式人工智能研究组负责人, 专注于自然语言的预训、生成和评估等研究方向; 在卡耐基梅隆大学语言技术研究所博士期间, 与 Graham Neubig 教授联合创立 Inspired Cognition 自然语言技术公司。在自然语言处理和人工智能领域发表学术论文 60 余篇, 谷歌学术引用 6000 余次。ACL 会议史上首次实现连续两年获得 System & Demo Paper Award; 提示工程 (Prompt Engineering) 概念最早提出者之一。曾获得中国人工智能学会优秀博士论文、上海市计算机学会优秀博士论文, 百度奖学金, 微软学者, IBM 全球博士英才计划奖等。领导开发 ExplainaBoard, DataLab 等开源工具, 相关工具月均总下载量超过 1 万次, 自 2021 年发布以来, 总下载量达 15 万余次。

## Tutorial 2: 扩散模型

李崇轩 中国人民大学

**报告题目：**扩散模型

**报告摘要：**扩散概率模型是一类新涌现的深度生成模型，这类方法逐步地对先验噪声分布去噪得以逼近数据分布。目前，扩散概率模型在数据合成质量、采样的多样性和数据密度估计等指标下取得了超越 VAE、GAN、FLOW 等经典深度生成模型的结果，也在诸多领先的视觉生成大模型中得到应用，有望成为未来通用生成式智能的重要组成部分。本次课程会介绍扩散概率模型的基本原理与代表性工作，介绍其在可控生成、多模态建模和三维场景建模等方面的前沿进展，并对扩散概率模型的下一步发展做简单的展望。



**讲者简介：**李崇轩 博士，中国人民大学准聘助理教授，博士生导师。2019 年博士毕业于清华大学，2021 年加入中国人民大学高瓴人工智能学院。研究方向为深度概率学习，相关工作发表于机器学习领域重要国际会议、期刊 40 余篇。代表性工作有：一致性理论下最优的半监督 GAN 方法 Triple-GAN；扩散概率模型在最大似然意义下的最优反向方差估计 Analytic-DPM；文图通用的多模态扩散概率大模型 Unidiffuser。李崇轩曾荣获机器学习领域重要国际会议 ICLR 杰出论文奖，吴文俊人工智能自然科学奖一等奖，吴文俊人工智能优秀青年奖，中国计算机学会优秀博士学位论文奖，北京市科技新星，中国博士后创新人才支持计划，主持国家自然科学基金面上项目，教育部产学结合协同育人项目等。

## Tutorial 3: Learning Environment Models for Reinforcement Learning

俞扬 南京大学

**报告题目：**Learning Environment Models for Reinforcement Learning

**报告摘要：**基于环境模型的强化学习方法利用与环境模型交互产生的虚拟样本帮助学习，从而避免大量现实样本采样造成的昂贵代价。因此基于环境模型的强化学习被认为是解决以往强化学习样本效率低下的一个重要途径。现有方法中环境模型的学习技术往往采取了最为简单的单步预测。然而，环境模型作为强化学习的训练环境，会被用于评估未知策略，这对泛化能力提出了更高的要求，用单步预测的技术难以学到高精度的环境模型。本次教程将介绍近期环境模型学习的方法和研究进展，以及基于环境模型的强化学习方法和应用。



**讲者简介:** 俞扬, 南京大学人工智能学院教授, 主要从事机器学习、强化学习的研究工作, 获 4 项国际论文奖励和 3 项国际算法竞赛冠军。入选国家级青年人才计划, 获 2020 CCF-IEEE 青年科学家奖, 入选 2018 IEEE AI's 10 to Watch, 获首届亚太数据挖掘“青年成就奖”, 并受邀在国际人工智能联合会 IJCAI'18 作 Early Career Spotlight。

## Tutorial 4: 计算神经科学

吴思 北京大学

**报告题目:** 计算神经科学

**报告摘要:** 计算神经科学的目的是用计算建模的方法阐明大脑信息加工的基本原理, 并在此基础上启发类脑智能的发展, 因此计算神经科学起到了脑科学与人工智能之间的桥梁作用。与深度学习网络不同, 生物神经网络通常具有不同架构, 信息加工更多依赖于网络的动力学过程, 了解这些不同点可以帮助我们思考如何实现类人的认知功能。本课程将重点介绍计算神经科学的基础知识, 包括神经元、突触和网络建模的基本知识, 一些神经建模在类脑计算中应用的例子, 以及一套方便、灵活且高效的神经建模编程工具。



**讲者简介:** 吴思, 北京大学心理与认知科学学院教授、麦戈文脑科学研究所研究员、北大-清华生命科学联合中心研究员。主要研究方向是计算神经科学和类脑计算, 尤其关注于神经系统信息加工的基本原理, 在过去二十多年里系统研究了大脑信息表征的正则化模型-连续吸引子神经网络。已发表论文上百篇, 包括大量神经科学顶级期刊和人工智能顶级会议。目前担任计算神经科学领域最大国际期刊 Frontiers in Computational Neuroscience 的共同主编, 中国神经科学学会《计算神经科学与神经工程》专业委员会主任等。

## Workshops 报告及讲者简介

### Workshop 1: 大模型对 CV/PR 的挑战与机会

主席：代季峰（清华大学） 张兆翔（中国科学院自动化研究所） 操晓春（中山大学）、  
王文海（上海人工智能实验室）

**时间：6月11日（周日）13:30-18:00 地点：B6馆**

时间	主持人	内容
13:30~14:00	操晓春	讲者：田永鸿（北京大学） 题目：国产 E 级智算平台上视觉大模型训练及应用
14:00~14:30	操晓春	讲者：田奇（华为） 题目：在大模型背景下，重新审视计算机视觉的发展方向
14:30~15:00	张兆翔	讲者：刘静（中国科学院自动化研究所） 题目：多模态预训练的进展回顾与展望
15:00~15:30	张兆翔	讲者：黄高（清华大学） 题目：视觉数据多样性带来的机遇与挑战
15:30~15:38	百度宣讲	
15:38~15:45	中场休息	
15:45~16:15	张兆翔	讲者：董胤蓬（清华大学） 题目：视觉基础模型的安全性及可信性
16:15~16:45	王文海	讲者：朱锡洲（商汤科技） 题目：面向多模态多任务的通用视觉感知大模型
16:45~17:15	王文海	讲者：张祥雨（旷视科技） 题目：基于解耦学习的可逆纵列神经网络
17:15~17:30	代季峰	Panel 嘉宾：田永鸿（北京大学），田奇（华为），刘静（自动化所），黄高（清华大学），董胤蓬（清华大学），朱锡洲（商汤科技），张祥雨（旷视科技）

## 田永鸿 北京大学

**报告题目：**国产 E 级智算平台上视觉大模型训练及应用

**报告摘要：**以 ChatGPT/GPT4 为代表的大模型正在影响 AI 领域的关键创新，但我们要思索我国科技和产业创新的机会在哪里？近年来，鹏城实验室采用国产芯片建设了“鹏城云脑 II”超级智算平台，填补了国产软硬件平台上千亿级 AI 大模型并行训练关键技术空白，打造了具有国际影响力的千亿参数大规模 AI 模型技术体系，高效支持了鹏程·盘古、鹏城·百度·文心、鹏程·神农等系列 AI 大模型的训练及行业应用。然而，视觉大模型训练是比自然语言大模型训练更具挑战性的任务。针对这一挑战，鹏城云脑 II 研发了规模视觉数据处理与模型训练平台，能支持近 PB 级别数据处理、千亿多模态模型的高效微调训练。在此基础上，构建了 24 亿图片高质量视觉数据集，训练了鹏城-大圣视觉基础模型，已开源 3 亿参数模型，在检测、重识别等 30+项下游任务上验证性能平均提升超 10%。从大模型到 AI4S，甚至是 AI4X（即 AI for Everything），这无疑是一场人工智能驱动的科技与产业革命。我们将与学术界和产业界探讨如何高效协作来应对这场科技与产业革命带来的巨大挑战并抓住发展机遇。



**讲者简介：**北京大学博雅特聘教授，博士生导师，IEEE Fellow，北京大学深圳研究生院信息工程学院院长，鹏城实验室网络智能部副主任兼云脑研究所所长，2018 年国家杰出青年基金获得者。主要研究方向为分布式机器学习、神经形态视觉和视频大数据。累计主持国家重点研发计划项目、国基金杰青/重点/重大仪器项目等国家、省部级与企业合作项目 40 余项，累计在 Nat. Sci. Data、IEEE Trans 等国际期刊和 ICML、NeurIPS 等国际会议发表学术论文 300 余篇，两获国际期刊和会议最佳论文奖；拥有美/中国发明专利 100 余项，获国家技术发明/进步二等奖各 1 次、教育部科技进步一等奖 1 次、中国电子学会技术发明/科技进步一等奖各 1 次、2022 年 IEEE 标准奖章和标准新兴技术奖、2022 年 ACM 戈登贝尔奖特别奖提名，国内外算法竞赛奖 10 余次，是首届高校计算机专业优秀教师奖励计划获奖者。曾任香港中文大学（深圳）和华中科技大学兼职教授，国际期刊 IEEE TCSVT/TMM/Multimedia 等编委，IEEE MIPR2020/ICME2021 大会主席，IEEE ICME2015/BigMM2015/ISM2015/MIPR2018/MIPR2019 程序主席，现任 IEEE 数据压缩标准委员会副主席兼 IEEE 2941 标准工作组组长、中国图象图形学会理事与交通视频专委会副主任等。他是科技部十四五重点专项“智能传感器”专家组成员、广东省十四五重点专项“新一代人工智能”专家组成员。

## 田奇 华为

**报告题目：**在大模型背景下，重新审视计算机视觉的发展方向

**报告摘要：**以 ChatGPT 为代表的大规模语言模型正在席卷 AI 领域，也将深刻地改变人们的生产 and 生活方式。相较于语言模型，视觉模型的参数量和通用性都落于下风，这与图像信号更复杂的结构和更大的体量相违背。在本次报告中，我将从 ChatGPT 带来的启发讲起，指出在大语言模型的预训练+微调机制背后，是“所学即所需”的交互环境及其所带来的“上下游一体”的任务体系。与之相对地，计算机视觉领域还无法构建相应的环境和体系，代理任务与实际需求的割裂比较严重，因而难以复刻自然语言领域的成功模式。基于上述观点，我将展望计算机视觉的发展方向，并对近期业界的代表性工作进行简要概述。



**讲者简介:** 田奇教授目前是华为云人工智能领域首席科学家, 2018年6月-2020年3月担任华为诺亚方舟实验室计算视觉首席科学家。田奇教授本科毕业于清华大学电子工程系, 后赴美国伊利诺伊大学香槟分校学习, 师从 Thomas S. Huang 教授获博士学位。后历任美国德克萨斯大学圣安东尼奥分校计算机系助理教授、副教授、和正教授 (2002-2019), 2010年获 Google Faculty Research Award, 2017年 UTSA 校长杰出研究奖、2016年获评多媒体领域 10 大最具影响力的学者, 并于 2018 年入选国家级领军人才创新项目。田教授也是教育部长江讲座教授、国家自然科学基金海外杰青, 中国科学院海

外评审专家。他是国际欧亚科学院院士(2021)、IEEE Fellow (2016)、CAAI Fellow (2022), 获得 2021 年中国人工智能学会吴文俊人工智能杰出贡献奖。田教授是 IEEE TMM, TCSVT, TNNLS, ACM TOMM, Multimedia Systems Journal 等多个期刊的 Associate Editor。他拥有多项美国专利, 在计算机视觉及多媒体方向顶级期刊及会议如 IEEE TPAMI, IJCV, TIP, TMM, CVPR, ICCV, ECCV, ACM MM 上发表文章约 650+ 余篇 (包括 200 篇 IEEE/ACM 期刊和 210 篇 CCF A 类会议文章), 谷歌学术引用次数 50500+, h 指数为 101, 有 8 篇论文获最佳论文奖或者最佳学生论文包括 ACM Multimedia 等。

### 刘静 中国科学院自动化研究所

**报告题目:** 多模态预训练的进展回顾与展望

**报告摘要:** 近年来, 从预训练模型到预训练大模型, 从文本、音频、视觉等单模态大模型, 到现在的图文、图文音等多模态预训练大模型, 无论在学术界还是企业界预训练模型都得到了广泛关注与爆发式发展。多模态预训练通过联合图文音等多模态内容进行模型学习, 其发展在多模态理解、搜索、推荐、问答, 语音识别与合成, 人机交互等应用领域中具有潜力巨大的市场价值。本报告主要包含三方面内容: 分析多模态预训练模型的重要性与必要性; 回顾当前多模态预训练的最新研究进展; 多模态预训练模型主要应用场景与未来展望。



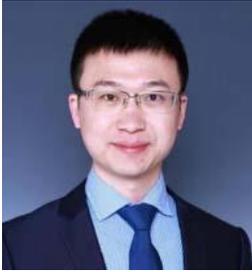
**讲者简介:** 中科院自动化所研究员/博导, 中国科学院大学岗位教授, 国家优青获得者。研究方向多模态分析与理解, 紫东太初大模型。曾获中国电子学会自然科学一等奖, 图像图形学会科学技术二等奖, 2022 年世界人工智能大会“卓越人工智能引领者奖 SAIL”。承担或参与多项国家自然科学基金项目、国家 973 课题、国家基金重大研究计划、国家重点研发等。已发表高水平学术论文 150 余篇, 谷歌学术引用 11000+ 次, SCI 他引次数 4000+ 次, 其中有三篇被 ESI 列为 Top1% 高被引论文。在视觉计算相关领域的多项国际学术竞赛中荣获冠军 10+ 项。

### 黄高 清华大学

**报告题目:** 视觉数据多样性带来的机遇与挑战

**报告摘要:** 在计算机视觉领域, 随着深度卷积网络和 Transformer 的发展与成熟, 骨干模

型创新带来的收益正在逐步递减。相比之下，数据的重要性愈发凸显，促成了以数据为中心（Data-Centric AI）的视觉技术的兴起。本报告将围绕视觉数据的多样性特点，从数据标注、数据扩增和动态推理等几个方面探讨数据多样化对视觉学习带来的机遇与挑战。



**讲者简介：**清华大学自动化系副教授，博士生导师。2015年获清华大学博士学位，2015年至2018年在美国康奈尔大学计算机系从事博士后科研工作。主要研究领域为深度学习和计算机视觉，提出了主流卷积网络 DenseNet。目前在 NeurIPS, ICML, CVPR 等国际顶级会议及 IEEE 多个汇刊共计发表学术论文 100 余篇，被引 40000 余次，最高单篇引用超过 3 万次。获国家优青、CVPR 最佳论文奖、达摩院青橙奖、世界人工智能大会 SAIL 先锋奖、中国自动化学会优秀博士学位论文、中国百篇最具影响国际学术论文、中国人工智能学会自然科学一等奖和吴文俊优秀青年奖等荣誉，入选北京智源学者、AI2000 人工智能最具影响力学者、《麻省理工科技评论》亚太区“35 岁以下科技创新 35 人”。

### 董胤蓬 清华大学

**报告题目：**视觉基础模型的安全性可信性

**报告摘要：**视觉基础模型近年来发展迅速，在图像感知、生成等诸多任务中取得了显著效果。然而，基础模型在发挥巨大作用的同时，其面临着新型的安全风险与可信性问题。本报告将介绍视觉基础模型的安全性可信性，包括在对抗攻击、后门攻击等恶意攻击下的鲁棒性以及其本身所引发的安全性问题。另一方面，我们将探讨如何更好地利用基础模型促进安全性可信性。



**讲者简介：**董胤蓬，清华大学计算机系博士后研究员，导师为朱军教授。于 2017 年和 2022 年在清华大学计算机系分别获得学士和博士学位。主要研究方向为机器学习与计算机视觉，聚焦深度学习鲁棒性的研究，先后发表 TPAMI、IJCV、CVPR、NeurIPS 等顶级国际期刊和会议论文四十余篇。担任 VALSE 执行 AC，担任 TPAMI、IJCV、ICML、NeurIPS、CVPR 等期刊和会议审稿人。曾在 ICML2021、AAAI2022 等国际会议上组织了对抗机器学习相关研讨会。曾获得 CCF 优秀博士学位论文激励计划、CCF-CV 学术新锐奖、微软奖学金、百度奖学金、字节奖学金、清华大学“水木学者”计划、博新计划等。

### 朱锡洲 商汤科技

**报告题目：**面向多模态多任务的通用视觉感知大模型

**报告摘要：**近年来，针对专用感知模型缺乏足够通用性与迁移能力的缺点，催生出了支持多模态输入和多任务输出的通用感知模型。为了满足这一通用性的需求，除了通用感知的模型构建外，还需要设计合适的训练范式以充分利用尽可能丰富的训练数据和监督信号。在本次报告中，我将介绍我们在多模态多任务通用视觉感知方面一系列的探索工作，包括

(1) 基于互信息最大化的通用多模态多任务预训练框架, (2) 首次成功应用于视觉和图文标杆任务的通用视觉感知模型, 并在各任务上取得与专用模型媲美的性能。

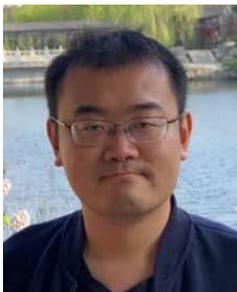


**讲者简介:** 商汤科技研究院资深算法研究员。2015 年本科毕业于中国科学技术大学。2020 年博士毕业于中国科学技术大学-微软亚洲研究院联合培养博士生项目。研究方向主要为图像和视频中的物体检测、多模态多任务通用感知模型、视觉自监督与弱监督预训练等。以第一作者在 CVPR、ECCV、ICLR、NeurIPS 等国际顶级会议上发表论文 10 余篇。根据 Google Scholar 统计, 论文被引用 6500 余次。2018 年获微软学者奖学金。所提出的 Deformable ConvNets v2 与 Deformable DETR 已成为目标检测领域的代表性工作。参与负责商汤多模态多任务大模型书生 2.0 及书生 2.5 项目, 在 20 余个不同场景的视觉和图文任务中取得世界领先水平, 其中 ImageNet 图像分类取得开源模型中最佳性能, COCO 物体检测取得世界第一水平。

### 张祥雨 旷视科技

**报告题目:** 基于解耦学习的可逆纵列神经网络

**报告摘要:** 视觉大模型架构在经历了百花齐放的时代后, 如今普遍采用 Transformer 作为基本的构造单元。然而相比之下, 神经网络的宏观架构设计 (macro design) 得到的关注却并不多。虽然传统的单路直筒状宏观架构具有简单高效的优点, 但是由于信息瓶颈效应, 模型在预训练的过程中不可避免地会丢失信息, 对模型的推广性、尤其是大模型这种追求通用性的应用带来了不利影响。为此, 本工作提出一种新的宏观架构设计思路, 通过将特征解耦学习 (disentangled feature learning) 的思想引入模型设计中, 提出可逆纵列神经网络 (Reversible Column Network) 来保证特征解耦的同时信息不丢失。总体架构则借鉴了 Hinton 的 GLOM, 通过将输入反复接入网络, 将静态图片统一到类似视频流处理的框架中。从结果上来看, 20 亿参数的模型仅使用  $3 \times 3$  的静态卷积便可以在 ImageNet 上达到 90% 的准确度, 下游任务也达到较高的水平。且框架兼容任何主流模型设计单元, 对于 CNN 和 ViT 均有明显的收益, 搭配掩码图像建模 (MIM) 等预训练框架之后相比 ViT+MAE 等经典算法亦有明显提升。此外由于本架构是可逆神经网络的特例, 因此继承了可逆网络节省显存的优势: 本工作的大部分实验均可在 2080ti 上完成, 这对于大模型训练无疑是重要的。模型与代码均已开源: <https://github.com/megvii-research/RevCol>。



**讲者简介:** 旷视科技研究院基础模型组负责人。2017 年博士毕业于西安交通大学。期间参加微软亚洲研究院联合培养博士生项目, 师从孙剑博士和何恺明博士。目前团队研究方向包括高性能卷积网络设计, AutoML 与自动化神经网络架构搜索, 深度模型的裁剪与加速等。曾在 CVPR/ICCV/ECCV/NIPS/TPAMI 等顶级会议/期刊上发表论文 20 余篇, 获 CVPR2016 最佳论文奖, 并多次获得顶级视觉竞赛如 ImageNet 2015、COCO 2015/2017/2018 冠军。代表作包括 ResNet/ShuffleNet v1/v2 等, 在业界得到广泛应用。Google Scholar 引用数 40000+。入选 2019 福布斯中国 30 岁以下精英榜

## Workshop 2: 遥感影像解译

主席：白翔（华中科技大学）程堪（西北工业大学）李军（中国地质大学（武汉））、  
孙显（中国科学院空天信息创新研究院）、方乐缘（湖南大学）

时间：6月12日（周一）13:30-18:00 地点：兰花厅 A

时间	主持人	内容
13:30~14:00	白翔	讲者：孟德宇（西安交通大学） 题目：基于噪声和图像先验建模的高光谱图像复原研究
14:00~14:30	李军	讲者：吴泽彬（南京理工大学） 题目：云边协同的多源遥感融合高效处理
14:30~15:00	方乐缘	讲者：张洪艳（中国地质大学（武汉）） 题目：基于弱监督深度学习框架绘制中国首幅1米分辨率土地覆盖一张图
15:00~15:30	方乐缘	讲者：陈思伟（国防科技大学） 题目：极化 SAR 图像精细解译与应用
15:30~15:45	中场休息	
15:45~16:15	孙显	讲者：邹征夏（北京航空航天大学） 题目：物理驱动的遥感图像合成与解译
16:15~16:45	孙显	讲者：姚西文（西北工业大学） 题目：高分遥感图像弱监督目标检测
16:45~17:15	程堪	讲者：杨学（上海人工智能实验室） 题目：视觉有向目标的高精度检测
17:15~18:00	程堪	Panel 嘉宾：孟德宇（西安交通大学）、吴泽彬（南京理工大学）、张洪艳（中国地质大学）、陈思伟（国防科技大学）、邹征夏（北京航空航天大学）、姚西文（西北工业大学）、杨学（上海人工智能实验室）

### 孟德宇 西安交通大学

**报告题目：**基于噪声和图像先验建模的高光谱图像复原研究（孟德宇/曹相湧）

**报告摘要：**从低信噪比的高光谱遥感图像中复原出清晰图像是遥感图像底层视觉研究的一项重要任务，其对于图像下游任务（比如分类、分割、检测等）有着一定的影响。高光谱图像复原质量很大程度上取决于所采用的噪声模型以及图像先验的准确性，本次报告将分别介绍高光谱图像复杂噪声的统计建模方法、高光谱图像先验的建模方法，以及两种提升现有高光谱图像复原方法泛化性的范式（即插即用赋权网络、深度参数化低秩正则子）。所提的方法论在多个高光谱图像复原数据集上都展示了优越的性能。



**讲者简介：**西安交通大学教授，博士生导师，国家级领军人才，卡内基梅隆大学访问学者，中国工业与应用数学学会副理事长，任大数据算法与分析技术国家工程实验室机器学习教研室负责人。发表论文百余篇，其中 IEEE 汇刊论文 60 余篇，计算机学会 A 类会议 40 篇，谷歌学术引用超过 21000 次。现任 IEEE Trans. PAMI, Science China: Information Sciences 等 7 个国内外期刊编委。目前主要研究聚焦于元学习、概率机器学习、可解释性神经网络等机器学习基础研究问题。

### 吴泽彬 南京理工大学

**报告题目：**云边协同的多源遥感融合高效处理

**报告摘要：**多源遥感融合能够通过各种遥感数据源的互补信息提升对地观测和目标探测的精度。但是，遥感观测分辨率的不断提升，使得遥感图像的数据量快速增长，机载/星载遥感探测任务每天产生的海量图像给多源遥感数据的有效处理带来了巨大挑战。针对海量遥感数据的有效管理和实时处理需求，提出了一种云边协同的多源遥感融合高效处理方法。一方面，构建高效可扩展的遥感大数据云平台，实现海量遥感数据的分布式高效存储，提出基于元启发式多任务优化调度的遥感大数据高效处理方法。另一方面，提出基于边缘端和云端协同交互多粒度并行优化的多源遥感融合快速处理方法，满足目标探测等遥感实时应用需求。



**讲者简介：**南京理工大学计算机科学与工程学院教授、博士研究生导师。IEEE GRSS Fellow and Senior Member Search Committee 主席，IEEE GRSS Nanjing Chapter 主席，IEEE JSTARS 编委。入选“长江学者奖励计划”青年学者、科技部中法杰出青年交流计划、江苏省杰青、江苏省六大人才高峰、江苏省 333 工程中青年科学技术带头人等，获得教育部霍英东青年教师奖、江苏省科学技术（自然科学类）二等奖（第 1）、上海市科技发明金奖（第 2）等。主要从事高维信号处理、多源遥感融合、大数据处理等领域的科研与教学工作。作为项目负责人承担了江苏省杰出青年基金、国家自然科学基金、科技部创新特区项目等 13 项省部级以上科研项目。在 IEEE

Transactions on Neural Networks and Learning Systems、IEEE Transactions on Signal Processing、IEEE Transactions on Image Processing 等高水平学术刊物上发表论文 109 篇，近 5 年以第 1 作者/通讯作者发表 SCI 收录论文 32 篇(包括 IEEE 汇刊论文 18 篇，SCI 收录 1 区 TOP 期刊论文 19 篇，ESI 高被引论文 5 篇)。获国家发明专利授权 19 项。

### 张洪艳 中国地质大学(武汉)

**报告题目：**基于弱监督深度学习框架绘制中国首幅 1 米分辨率土地覆盖一张图

**报告摘要：**随着城市化进程加快，我国地表覆盖变化频繁，大范围高分辨率土地覆盖制图需求迫切。但是，由于地物类型多样、手动标注费时、算力需求巨大等原因，全国尺度的高分辨率土地覆盖制图产品一直欠缺。为了填补这一空白，本报告介绍了一种基于弱监督深度学习的跨分辨率土地覆盖制图框架，采用现有 10 米分辨率土地覆盖制图产品、开源 OSM 数据作为训练标签，使用 1 米分辨率 Google 遥感影像进行全国尺度的高分辨率土地覆盖制图。针对影像数据与训练标签分辨率差异带来的标签噪声问题，通过构建多尺度特征分辨率保持卷积网络，利用弱监督模块与无监督损失函数筛选低分辨率标签中的可信部分作为监督源并抑制标签噪声，充分挖掘 10 米分辨率土地覆盖制图产品中的可靠标签信息，实现了大范围高分辨率土地覆盖制图任务。研究团队耗时 10 个月处理约 73.25 TB 高分辨率遥感数据，研制了我国首幅 1 米分辨率土地覆盖一张图(SinoLC-1)，填补了我国缺少国家尺度高分辨率土地覆盖产品的空白。使用数十万随机验证样本点与国家第三次土地调查对 SinoLC-1 产品精度进行了定量验证，总体准确率为 73.61%，Kappa 系数为 0.6595。



**讲者简介：**中国地质大学(武汉)教授、博导，国家级青年人才，湖北省杰青。主要从事智能信息处理、农业遥感监测等方向的研究工作。已发表 SCI 期刊论文 90 余篇，其中 ESI 热点论文 3 篇、高被引论文 16 篇，出版学术专著 1 部，授权/申请国家发明专利 20 余项。先后主持国家重点研发计划课题 1 项、国家自然科学基金 5 项和省部级科研项目 3 项。曾获测绘科学技术一等奖 2 项、湖北省自然科学二等奖 1 项、霍英东教育基金会青年教师奖二等奖 1 项和 IEEE 地球科学与遥感学会数据融合大赛冠军 2 项。IET Fellow, IEEE Senior Member, 连续 2 年入选斯坦福大学全球前 2% 顶尖科学家、

爱思唯尔(Elsevier)中国高被引学者和全球学者库前 10 万顶尖科学家等榜单，应邀担任 3 个 SCI 期刊副编辑、4 个国内外期刊编委。

### 陈思伟 国防科技大学

**报告题目：**极化 SAR 图像精细解译与应用

**报告摘要：**极化 SAR 能够获取目标丰富极化信息，有助于实现目标物理参数反演与识别，在对地观测、侦察监视等领域有着重要应用价值。对极化 SAR 目标散射机理进行精细解译具有重要的科学意义和应用价值。雷达目标通常具有散射多样性。特别是绕雷达视线旋

转方向的散射多样性，被认为是制约目标机理准确解译的瓶颈。然而，目标散射多样性中也蕴含着目标独特信息。相比于传统消除目标散射多样性的研究工作，报告人提出了利用目标散射多样性的研究思路，发展了极化旋转域解译理论方法，并在地物分类、目标检测、灾害评估等领域获得应用。本报告将介绍研究团队近年来在上述领域的研究进展。



**讲者简介：**国防科技大学博士生导师，获国家自然科学基金优秀资助。研究方向为极化雷达成像与识别、电子对抗和机器学习等。主持完成国家级项目 10 余项，发表论文 120 余篇，第一作者出版中文和英文专著 2 部，第一发明人授权专利 17 项，获国家科技进步二等奖、湖南省自然科学一等奖、军队科技进步一等奖、军队教学成果一等奖和 IEEE 地球科学与遥感学会青年成就奖。先后担任《IEEE Geoscience and Remote Sensing Letters》副编辑，《Geomatics, Natural Hazards and Risk》《Journal of Remote Sensing》《电子与信息学报》《雷达学报》和《信号处理》等编委、IGARSS 等国际国内会议技术委员会委员和分会主席。

### 邹征夏 北京航空航天大学

**报告题目：**物理驱动的遥感图像合成与解译

**报告摘要：**图像合成是近期人工智能领域前沿且基础的研究方向。遥感图像的合成对于数字化遥感场景构建、赋能下游图像解译等任务都起着关键的作用。有别于自然图像合成任务，遥感图像在俯拍成像视角、地物光谱先验等方面都具有特殊性。在本次报告中，讲者将介绍课题组在物理驱动的遥感图像合成方面的相关工作，包括光谱库驱动的遥感图像合成、基于成像模型的遥感图像合成，以及基于俯拍多平面表示的立体遥感场景合成。最后，还将介绍上述合成方法在遥感图像解译方面的应用。



**讲者简介：**北京航空航天大学宇航学院教授、博士生导师，国家级青年人才。主要研究方向包括遥感图像处理、计算机视觉、深度学习等，关研究成果以第一/通讯作者身份发表在 Proceedings of the IEEE、Nature Communications、TPAMI、TIP、TGRS、CVPR、ICCV、AAAI、ICRA 等重要期刊和会议，谷歌学术论文引用 3000 余次，单篇谷歌引用 1100 余次，开源项目 Github Star 4000 余次。入选 2022 全球前 2% 科学家名单，研究成果连续四年收录于斯坦福大学的旗舰课

程 CS231n (美国三院院士 Li Fei-Fei 主讲)，担任中国图象图形学学会宣传工委秘书长。相关研究被新华社、中央电视台、新科学人等多家高影响力媒体机构报道和采用，服务于航天国防、自动驾驶等重大应用需求，应用于福特汽车公司、全球首个自动驾驶测试场 MCity 以及国内外 6 颗遥感卫星。

## 姚西文 西北工业大学

**报告题目：**高分遥感图像弱监督目标检测

**报告摘要：**高分遥感图像目标检测与识别是遥感大数据智能解译的关键技术，在搜救、侦察、监视、预警等军民领域具有重要的应用价值。与自然图像相比，遥感影像幅面大、目标类型繁杂、数量众多，对其进行精细标注(如目标边框、关键部件等)更依赖专家知识，远超人类处理极限。本报告将重点汇报我们脑与人工智能团队利用弱监督深度学习实现遥感图像目标鲁棒检测与型号精细识别的相关研究进展。



**讲者简介：**西北工业大学副研究员，博士生导师，入选陕西省高校“青年杰出人才支持计划”，西北工业大学“翱翔新星”人才计划。主要研究方向为遥感影像智能解译。以第一作者/通讯作者在 IEEE TPAMI、IEEE TIP、IEEE TGRS、IEEE CVPR 等国际期刊和会议上发表论文 40 余篇，其中 10 篇论文入选 ESI 高被引和热点论文，1 篇论文入选 2018 年度中国百篇最具影响国际学术论文，1 篇论文获陕西省第十四届自然科学优秀学术论文二等奖。承担了国家自然科学基金委青年基金、面上项目、创新特区项目、博士后特别资助和省重点研发计划项目等 10 余项国家和省部级科研项目。荣获吴文俊人工智能技术发明奖一等奖，火箭军“智箭·火眼”人工智能挑战赛亚军。担任国际地球科学于遥感学会旗舰期刊 IEEE JSTARS 副编辑，《遥感学报》专刊主编，2022 年中国模式识别与计算机视觉大会领域主席，中国图象图形学学会遥感图像专委会委员。

## 杨学 上海人工智能实验室

**报告题目：**视觉有向目标的高精度检测

**报告摘要：**高精度的有向目标检测是遥感场景解译中必不可少的一项技术手段。我们总结了当前有向目标检测中存在的边界不连续、类正方形检测等问题，并为此设计了一种统一的解决方案：高斯建模和分布距离度量。首先，我们对高斯分布协方差总结了三个重要的性质，并发现了这些性质可以有效消除上述问题。然后结合梯度分析和尺度不变性证明，我们还发现使用基于 KL 散度的分布距离度量会让检测器在训练过程中能根据目标的形态动态地调整参数的优化权重，最终在高精度指标上取得了显著的提升。



**讲者简介：**上海交通大学计算机系博士，吴文俊人工智能博士班成员，研究方向是计算机视觉（目标检测），导师为严骏驰教授。在 TPAMI/IJCV/CVPR/ICCV/ECCV/ICML-/NeurIPS/ICLR 等 CCF-A/清华 A 类期刊和会议上发表一作论文 13 篇，2 篇入选 Paper Digest 最具影响力 AAAI21 论文列表（第 1 名和第 10 名），谷歌学术引用超过 3000。博士期间获得 CCF-CV 学术新锐学者，上海交通大学学术之星提名奖，并连续两年获得博士生国家奖学金。所开源的代码在社区累积获得 star 数超过 7000，包括两个最具影响力的旋转检测 AlphaRotate 和 MMRotate。

## Workshop 3: 围绕手机的计算影像学

主席：程明明（南开大学）王立君（大连理工大学）卢湖川（大连理工大学）、左旺孟（哈尔滨工业大学）、郭春乐（南开大学）

时间：6月11日（周日）8:30-12:15 地点：兰花厅 A

时间	主持人	内容
8:30~9:05	王立君	讲者：张磊（OPPO） 题目：当 A (academia) 碰到 I (industry)：图像复原和增强任务中 AI 的困境和机遇
9:05~9:40	王立君	讲者：戴玉超（西北工业大学） 题目：卷帘快门相机：模型、优化与学习
9:40~10:15	王立君	讲者：贾旭（大连理工大学） 题目：影像的时空分辨率增强
10:15~10:30	中场休息	
10:30~11:05	郭春乐	讲者：薛天帆（香港中文大学） 题目：如何设计一个真正意义上的智能相机
11:05~11:40	郭春乐	讲者：张健（北京大学） 题目：Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model
11:40~12:15	王立君、郭春乐	Panel 嘉宾：张磊（OPPO）、戴玉超（西北工业大学）、贾旭（大连理工大学）、薛天帆（香港中文大学）、张健（北京大学）、曹讯（南京大学）、杨文翰（鹏程实验室）、顾舒航（电子科技大学）（红色为官网没有的）

## 张磊 OPPO

**报告题目：**当 A (academia) 碰到 I (industry)：图像复原和增强任务中 AI 的困境和机遇

**报告摘要：**近十年人工智能已经取得突破性发展，在各行各业中取得巨大成功，通用人工智能也已经曙光在前。相较于 AI 在大模型、预训练、多模态、内容生成等方面的惊艳表现，AI 在底层视觉，尤其是在图像复原和增强领域，的实际工业应用远远落后，令人遗憾。在本报告中，我们从数据、评估准则、模型部署等几个层面探讨这种困境的原因，给出几个亟待解决的问题，只要 A (academia) 和 I (industry) 加强互动，增进理解，创造条件，AI 在图像复原和增强任务中必定也可以大放异彩。



**讲者简介：**张磊教授 (IEEE Fellow) 于 2006 年加入香港理工大学电子计算学系，2017 年起任职讲座教授。张磊教授长期致力于计算机视觉、图像处理、模式识别等方向的研究，是底层视觉方面的国际权威学者。张教授是 IEEE Trans. on Image Processing (TIP) 的高级编委，IEEE Trans. on Pattern Analysis and Machine Intelligence (TPAMI)、SIAM Journal of Imaging Sciences 等多个国际期刊的编委。从 2015 年至 2022 年，张教授连续被评为 Clarivate Analytics Highly Cited Researcher。张磊教授目前也于 OPPO 研究院任 AI 科学家，从事 AI 影像前沿技术的研发。

## 戴玉超 西北工业大学

**报告题目：**卷帘快门相机：模型、优化与学习

**报告摘要：**当前，绝大多数消费级相机都采用卷帘快门 (Rolling Shutter) 机制。与全局快门相机同时曝光的成像机制不同，卷帘快门相机以逐行曝光的方式进行数据采集，不同行的像素点具有不同的曝光时间，同时也对应着不同的相机位姿。针对全局快门相机模型设计的三维视觉模型无法直接应用于卷帘快门相机。本次报告内容涵盖针对卷帘快门相机从多视几何模型构建、几何优化求解和基于深度学习的卷帘快门图像处理方面的一系列工作。首先，针对卷帘快门相机相对位姿问题，建立了卷帘快门本质矩阵并设计了相应的位姿估计算法；其次，在卷帘快门图像校正方面，构建一个端到端对称去畸变网络，从连续帧卷帘快门图像中恢复出对应的全局快门图像；再次，介绍基于几何模型的卷帘快门图像时间超分辨率方法，能够从连续两帧卷帘快门图像提取出对应于这两帧中任意时刻的高质量全局快门图像视频；最后，对于卷帘快门相机相关研究进行展望和讨论。



**讲者简介：**戴玉超，男，西北工业大学电子信息学院教授、博士生导师，国家级青年人才。主要研究工作集中在机器视觉、智能感知、图像处理、人工智能等领域，聚焦复杂动态场景的三维重建与感知、深度学习和几何模型融合的稠密匹配、新型仿生视觉传感器和计算成像等问题。主持国家自然科学基金、科技部科技创新 2030 “新一代人工智能” 重大研究计划子课题、JKW 领域基金重点项目等科研项目。近年来在 TPAMI、IJCV、ICCV、CVPR、NeurIPS 等国际顶级期刊和会议上发表论文 70 余篇，谷歌学术引用超过 7700 次，H 因子 43。先后获得 CVPR 2012 最佳论文奖 (大陆高校 30 年来首次获得该奖项)、陕西省自然科学一等奖、中国图象图形学学会青年科学家奖、火箭军“智箭火眼”

人工智能挑战赛全国第一名、IEEE CVPR 2020 最佳论文奖提名、ECCV 2020 鲁棒计算机视觉挑战赛双目深度估计赛道冠军和光流估计赛道亚军、CVPR 2017 非刚性结构与运动恢

复挑战赛最佳算法奖、APSIPA 2017 年度峰会最佳深度学习/机器学习论文奖等奖项。担任 APSIPA 杰出讲者和 CVPR、ICCV、NeurIPS 等国际顶级会议领域主席，ACCV 2022 宣传主席，中国图象图形学报青年编委。

### 贾旭 大连理工大学

**报告题目：**影像的时空分辨率增强

**报告摘要：**随着智能手机的普及，人们对手机的影像能力提出了越来越高的要求，而深度学习算法的兴起为提升影像能力提供了强大的技术支撑。本次报告将聚焦于深度学习方法在影像的时间和空间分辨率增强方面的技术进展，从模型结构设计、真实场景退化和多传感器结合等方面展开介绍，并将讨论影像的时空分辨率增强方向的进一步研究思路和其在实际应用上的难点。



**讲者简介：**贾旭，大连理工大学人工智能学院院长聘副教授，辽宁省智能感知与理解人工智能重点实验室骨干成员，博士毕业于比利时鲁汶大学，是从 Tinne Tuytelaars 教授和 Luc Van Gool 教授。主要研究方向包括图像和视频的增强与生成、视觉目标检测跟踪以及类脑视觉。迄今在计算机视觉和机器学习领域顶级会议及期刊发表论文 30 余篇，Google Scholar 引用 6000 余次，申请国内外专利 10 余项。曾在 Google Research，商汤科技，华为诺亚方舟实验室等从事研究工作。主持并参与国家自然科学基金、科技部科技创新 2030 重大项目、科技委项目以及华为等企业合作项目若干项。

### 薛天帆 香港中文大学

**报告题目：**如何设计一个真正意义上的智能相机

**报告摘要：**AI 赋能的智能相机已广泛应用于我们的日常生活中，从手机相机上的图片美化到自动驾驶中的场景理解，都离不开智能相机。一个真正的智能相机系统，不仅仅是将 AI 算法应用于相机捕捉的图像，而是重新设计整个相机系统，包括捕捉、处理、编辑和分析。在本次演讲中，我将分享我们最近使用计算摄影和机器学习技术改进每个步骤的努力。这次报告将介绍如何在捕捉时使用多个摄像头和传感器来增强视觉信号或消除视觉障碍，如何使用合成数据训练高效的图像处理网络，包括降噪和消除光斑，以及讨论如何设计更高效的图像处理网络。



**讲者简介：**薛天帆教授是香港中文大学讯息工程系的 vice chancellor assistant professor。在此之前，他在谷歌研究院的计算摄影团队担任主任工程师，工作了五年以上，有着丰富的产学研结合经验。他毕业于麻省理工学院计算机科学与人工智能实验室，导师为 William T. Freeman 教授。2011 年在香港中文大学讯息工程系获得哲学硕士(M.Phil.)学位，2009 年在清华大学计算机科学与技术系获得本科学位。他的研究重点是计算摄影、计算机视觉和图形学以及机器学习。他的多项研究技术在学界和工业届都有很大反响：他研究的去反射技术被 Google Photoscan 使用，拥有超过 1000 万用户；他研究的快速双边学习技术被整合进了 Google Tensor 芯片；他研究的夜景算法活动 DPRreview 年度最佳创新奖，并被应用与谷歌 Pixel 手机中，也启发了很多其他夜景算法的开发。他也同时担任多个顶级会议和期刊的审稿人工作，并担任了 CVPR 2020 的网络主席，WACV 2023

和 CVPR 2023 领域主席。他的研究领域涵盖了计算摄影、机器学习、计算机视觉和计算机图形学。特别地，他致力于构建智能摄像头系统，并与视觉世界进行交互。这些技术可以惠及多个领域，包括人工智能视觉系统、手机摄像头、自动驾驶和医学成像等。

### 张健 北京大学

**报告题目：** Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model

**报告摘要：** Most existing Image Restoration (IR) models are task-specific, which can not be generalized to different degradation operators. In this work, we propose the Denoising Diffusion Null-Space Model (DDNM), a novel zero-shot framework for arbitrary linear IR problems, including but not limited to image super-resolution, colorization, inpainting, compressed sensing, and deblurring. DDNM only needs a pre-trained off-the-shelf diffusion model as the generative prior, without any extra training or network modifications. By refining only the null-space contents during the reverse diffusion process, we can yield diverse results satisfying both data consistency and realism. We further propose an enhanced and robust version, dubbed DDNM+, to support noisy restoration and improve restoration quality for hard tasks. Our experiments on several IR tasks reveal that DDNM outperforms other state-of-the-art zero-shot IR methods. We also demonstrate that DDNM+ can solve complex real-world applications, e.g., old photo restoration. Code: <https://github.com/wyhuai/DDNM>.



**讲者简介：** 张健，北京大学深圳研究生院信息工程学院助理教授/研究员、博导生导师，深圳市海外高层次人才。分别于 2007 年、2009 年、2014 年获得哈尔滨工业大学数学与应用数学学士学位、计算机科学与技术硕士学位及计算机应用博士学位，并先后在北京大学、香港科技大学和沙特国王科技大学做博士后访问研究。主要从事底层视觉、计算成像以及 AIGC 方面研究，共计发表包括 TPAMI、IJCV、TIP、CVPR、ECCV、ICCV、ICLR 等高水平权威国际期刊/会议论文 100 余篇。目前，负责北京大学视觉信息智能学习实验室(VILLA)。近两年，带领

VILLA 以第一或通讯作者在 CCF A 类会议/期刊上发表论文 20 余篇，Google Scholar 引用超过 4500 次，h-index 值为 35，i10-index 值为 63。多次获得国际/国内期刊会议最佳论文奖，连续三年入选人工智能与图像处理领域全球前 2% 顶尖科学家榜单。个人主页：<https://jianzhang.tech/>。

## Workshop 4: 多模态认知计算

主席：姬艳丽（电子科技大学）胡迪（中国人民大学）田亚鹏（德克萨斯大学达拉斯分校）、张正（哈尔滨工业大学（深圳））

时间：6月12日（周一）13:30-17:40 地点：B4馆

时间	主持人	内容
13:30~14:00	姬艳丽	讲者：汪萌（合肥工业大学） 题目：基于智能交互的心理与情绪计算分析
14:00~14:30	姬艳丽	讲者：李玺（浙江大学） 题目：多模态视觉结构建模和特征学习
14:30-15:00	姬艳丽	讲者：郑锋（南方科技大学） 题目：音视文融合感知与生成
15:00-15:30	姬艳丽	讲者：武宇（武汉大学） 题目：多模态感知与跨模态生成
15:30-15:40	中场休息	
15:40-16:10	胡迪	讲者：周航（百度） 题目：真人虚拟人的语音智能驱动
16:10-17:00	胡迪	Poster Spotlight（5分钟/篇）（官网没有） 讲者列表：王逸舟（182），马智祺（176），杜长德（97），杨丽鹤（115）
17:00-17:40	胡迪	Panel 嘉宾：汪萌（合肥工业大学）、李玺（浙江大学）、郑锋（南方科技大学）、李冠彬（中山大学）、武宇（武汉大学）、周航（百度）、张正（哈尔滨工业大学（深圳））、张晗（西北工业大学）

### 汪萌 合肥工业大学

**报告题目：**基于智能交互的心理与情绪计算分析

**报告摘要：**对基于视觉、对话等不同强度交互方式的生理信号和微反应检测技术进行介绍，探讨其在情绪、心理和人格分析中的应用。



**讲者简介：**汪萌，教授，博士生导师，合肥工业大学计算机与信息学院院长，合肥综合性国家科学中心人工智能研究院副院长。国际电气与电子工程师协会会员（IEEE Fellow）、国际模式识别协会会员（IAPR Fellow）。国家杰出青年科学基金获得者、国家重点研发计划“网络空间安全”专项项目负责人、国家重点研发计划“变革性技术关键科学问题”总体组专家、第八届教育部科技委信息学部委员。论文被引用三万余次，2016 至 2022 年度连续入选科睿唯安“全球高被引科学家”。

### 李玺 浙江大学

**报告题目：**多模态视觉结构建模和特征学习

**报告摘要：**互联网和物联网时代催生了海量多模态大数据，从这些海量数据中有效提取知识迫切需要各种人工智能的技术和手段。因此，如何进行人工智能驱动的视觉计算已经成为当今知识经济时代亟待解决的核心技术问题。本报告主要围绕数据驱动的人工智能学习方法，进行大规模多模态特征学习，从目标视觉感知特性、视觉特征表达、深度学习器构建机制、高层语义理解等多维度视角进行了深入剖析，并引入了大规模特征学习所涉及的主要研究问题和技术方法。然后系统地回顾了多模态特征表达和学习领域的不同发展阶段，介绍了近年来我们利用多模态特征学习进行视觉语义分析和理解所做的一系列代表性的研究工作及其实际应用。报告的最后将和大家一起探讨一下涉及多模态特征学习所面临的一些开放性问题 and 难题。



**讲者简介：**李玺教授，博导，IET Fellow，IEEE Senior Member，浙江大学求是特聘教授，国家杰出青年科学基金获得者，国家青年特聘专家，浙江大学上海高等研究院副院长，科技部科技创新 2030 新一代人工智能重大项目首席科学家，国家自然科学基金重点项目负责人，教育部重点规划研究项目负责人，第七届中国图象图形学会理事，浙江省杰出青年科学基金获得者，浙江省特聘专家，杭州市钱江特聘专家，浙江省 151 人才培养工程第二层次，中国信息与电子科技发展战略研究中心专家委员会特聘专家。主要从事计算机视觉、模式识别和机器学习等领域的研究和开发。

### 郑锋 南方科技大学

**报告题目：**音视文融合感知与生成

**报告摘要：**大型语言模型（ChatGPT）和多模态大型模型（GPT4）的快速发展，使人类首次实现通用人工智能成为可能。然而，这些现有的大型模型主要侧重于理解语言和弥合图像与语言之间的鸿沟。在面对更复杂的未裁剪长视频和集成更多模态信息（音频）时，他们仍然面临许多挑战。在本报告中，我们通过回答以下问题来探索“更长、更多”的任务。如何使用更多未配对的视频文本数据进行视觉语言模型的预训练？在视频描述中，如何实现类人生成更丰富多样的描述？如何在复杂的未裁剪长视频中有效地同时描述更多事件？如何对复杂的未裁剪长视频进行视听事件检测？如何利用视听融合的特征对花样滑冰进

行自动评分？



**讲者简介：**郑锋，南方科技大学副教授（研究员），国自然优秀青年基金获得者，于英国谢菲尔德大学获得博士学位。研究兴趣包括跨媒体计算、机器学习与计算机视觉。在国际顶级杂志和会议上包括 IEEE TPAMI/TIP/TNNLS, ICML, NeuIPS, AAAI, CVPR, ICCV, ECCV 发表 100 余篇学术论文，其中，高被引论文 2 篇，CCF 推荐 A 类论文 62 篇。指导学生在包括 CVPR ActivityNet, CVPR LOVEU, ACM MM PIC 等多个国际重要竞赛中多次取得了第一或第二名的好成绩。是 IET Image Processing 杂志副编辑，是 ACM MM 2020/2021 领域主席，ICME2021/IJCB2021 本地主席，PRCV2022 组委会主席，在多个国际顶级人工智能会议包括 CVPR, AAAI, IJCAI, NeuIPS, ICML, ICLR, KDD, ICCV 担任委员会成员。

### 武宇 武汉大学

**报告题目：**多模态感知与跨模态生成

**报告摘要：**多模态学习是深度学习领域当前火热的研究课题之一，其目标是对视觉、文本、音频等多种模态数据进行关联、感知、定位与生成。视频数据中天然涵盖多种模态的信息，如何利用好多种模态信息来提升对视频综合理解是值得探讨的研究课题。本次报告将首先介绍多模态学习包括音视频理解、视觉-语言等感知模型，以及基于扩散模型等一系列最新的多模态生成方法。



**讲者简介：**武宇，武汉大学计算机学院教授、博士生导师，入选国家级青年人才项目。本科毕业于上海交通大学，博士毕业于悉尼科技大学获得博士学位，后于普林斯顿大学从事博士后研究。主要从事视觉-语言理解、多模态理解、跨模态生成等方向研究。曾获 2020 年谷歌博士奖研金（Google PhD Fellowship，CV 方向全球范围仅 10 名获奖者），3 次获得 CVPR 国际学术竞赛的冠军（包括 ActivityNet、EPIC-Kitchens、YouTube-VOS 等）。担任人工智能顶会 NeurIPS 2023 领域主席，计算机视觉顶会 CVPR 2023 大会主要组织者、Workshop 主席、领域主席。

### 周航 百度

**报告题目：**真人虚拟人的语音智能驱动

**报告摘要：**随着 AIGC 和元宇宙概念的兴起，真人虚拟人作为数字世界人类的形象替身发挥着重要作用。通过语音等多模态输入对真人形象的虚拟人的唇形和动作进行驱动，是构建真人虚拟人的重要方式。本次报告将介绍基于最近基于语音对真人虚拟人唇形、头部、身体驱动方向的前沿方法，并展示百度数字人方面的最近技术成果。



**讲者简介：**周航，百度视觉技术部资深工程师，本科毕业于南京大学，博士毕业于香港中文大学。在 CVPR/ECCV/ICCV/AAAI/NeurIPS/SIGGRAPH 等会议发表论文 20 余篇，曾获得 CVPR, NeurIPS 最佳审稿人等称号。在多模态音视频联合学习和虚拟人的语音驱动，形象生成方面有着丰富的研究和落地经验，目前负责百度真人和 3D 虚拟人相关的技术研发。

## Workshop 5: 语言文字识别与理解

主席：白翔（华中科技大学）侯淇彬（南开大学）袁粒（北京大学）

时间：6月12日（周一）8:30-12:15 地点：兰花厅 B

时间	主持人	内容
8:30~8:55	白翔	讲者：金连文（华南理工大学） 题目：弱监督文字识别新思考
8:55~9:20	白翔	讲者：殷绪成（北京科技大学） 题目：开放集文字识别技术
9:20~9:45	白翔	讲者：姚聪（阿里巴巴达摩院） 题目：面向文档理解的视觉-语言联合建模
9:45~10:10	白翔	讲者：殷保才（讯飞） 题目：复杂场景下图文识别关键技术及产业化
10:10~10:18	白翔	“合合信息”金牌赞助商宣讲
10:18~10:43	白翔	讲者：苏劲松（厦门大学） 题目：基于图文语义融合的多模态神经机器翻译研究
10:43~11:08	白翔	讲者：刘禹良（华中科技大学） 题目：浅析多模态大型模型在文字图像智能领域的应用现状、挑战及未来展望
11:08-11:28	白翔	Poster Spotlight 讲者列表：王淦汀，陈喆，吴艳敏，黎向阳
11:28-12:15	白翔	Panel 嘉宾：金连文（华南理工大学）、殷绪成（北京科技大学）、姚聪（阿里巴巴达摩院）、殷保才（科大讯飞研究院）、苏劲松（厦门大学）、刘禹良（华中科技大学）

## 金连文 华南理工大学

**报告题目：**弱监督文字识别新思考

**报告摘要：**光学字符识别（OCR）是计算机视觉领域的一个重要研究问题，在文档数字化、图像理解、智能办公、智慧金融等领域具有非常广泛的应用前景。现有的方法和模型往往高度依赖高质量的标注数据，而获取这些数据通常耗时且费用高昂。本报告将重点探讨 OCR 领域中的三个问题：1) 如何减轻深度学习方法对高质量标注数据的依赖，2) 如何在数据稀缺或弱标注情况下解决 OCR 问题，3) 如何利用大规模未标注数据；针对这三个方面的问题，介绍一些代表性方法以及潜在的解决方案。最后，我将对 OCR 领域仍然存在的一些技术挑战和未来发展趋势进行探讨和展望。



**讲者简介：**金连文，男，华南理工大学二级教授，兼任中国图象图形学学会（CSIG）常务理事、广东省图象图形学会理事长、CSIG 文档图像分析与识别专委会主任、CSIG-MV、CAAI-PR 和 CAA-PRMI 专委会常务委员等职。主要研究领域为文字识别、文档图像理解等，在重要学术期刊及国际会议上发表论文 200 余篇，其中 SCI Q1 区+CCF A 类论文 100 余篇，Google Scholar 论文被引用数 10000 余次，H-Index 56。作为负责人先后主持国家科技支撑计划课题、国家重点研发计划课题、国家自然科学基金重点项目/面上项目、广东省自然科学基金团队项目/重点项目、企业合作项目等科研项目 40 余项；获省部级科技奖 5 项（其中一等奖 2 项，二等奖 3 项）；荣获中国人工智能学会、中国电子学会和中国图象图形学会科技进步二等奖各 1 项；指导学生参加 CVPR、ICDAR、ICPR、PRCV 等国际国内知名会议上的学术竞赛，荣获冠军 20 余次。

## 殷绪成 北京科技大学

**报告题目：**开放集文字识别技术

**报告摘要：**在诸如智能驾驶、网络信息安全等开放环境中，新数据、新模式、新类别不断涌现，模式识别与文字识别技术及应用出现新挑战。本报告首先简要回顾开放环境下模式识别的基本概念及主要技术，然后重点介绍其中的一个新方向——“开放集文字识别”（Open-Set Text Recognition），较详细地阐述其基本定义、通用框架和典型算法，最后对开放集文字识别技术应用进行了初步的展望。



**讲者简介：**殷绪成，教授、博士生导师，国家杰出青年科学基金获得者、科技创新 2030——“新一代人工智能”重大项目首席科学家，本科、硕士毕业于北京科技大学计算机系，博士毕业于中国科学院自动化研究所，现为北京科技大学计算机与通信工程学院院长、模式识别与人工智能技术创新实验室主任，中国图象图形学学会文档图像分析与识别专委会副主任/秘书长。他长期从事模式识别、文字识别、计算机视觉、人工智能芯片、工业智能技术及应用研究，在 IEEE T-PAMI、IEEE T-IP、CVPR、ICDAR 等学术期刊上发表近 100 篇论文；以第一完成人获北京市科技进步奖一等奖 1 项、教育部科技进步奖二等奖 1 项，带领团队连续四届（2013、2015、2017 和 2019 年）荣获国际文档分析与识别大会技术竞赛共 15 项冠军。

## 姚聪 阿里巴巴达摩院

**报告题目：**面向文档理解的视觉-语言联合建模

**报告摘要：**文档是一种内容丰富、形态多样的信息载体，在人们的日常工作、学习和生活中扮演着非常关键的角色。利用计算机从文档中以结构化的方式提取关键信息，不论是从研究还是应用的角度，都是一个值得深入探索的课题。近年来，视觉-语言联合建模技术的相关方法和应用发展十分迅速，本报告将梳理业界近期在利用视觉-语言联合建模技术执行文档理解任务方向的典型工作，并详细介绍我们最新的研究成果，包括结合几何信息的文档预训练模型 GeoLayoutLM 和端到端视觉信息抽取算法 ESP 等。



**讲者简介：**姚聪，达摩院自然语言实验室 OCR 团队负责人。在华中科技大学电信学院获得学士和博士学位，目前从事计算机视觉方向的研究，主要专业领域为自然场景 OCR、文档分析与理解。曾在旷视科技负责云服务业务相关视觉算法(自然场景 OCR、人脸识别等)的研发工作。在人工智能和计算机视觉领域的期刊 TPAMI、IJCV 以及会议 CVPR、ICCV、ECCV 等上发表学术论文 20 余篇，总引用量超过 1.2 万(根据 Google Scholar)。研究的 EAST、CRNN、DB 等方法在学术界和工业界均具有广泛的影响力。

### 殷保才 科大讯飞研究院

**报告题目：**复杂场景下图文识别关键技术及产业化

**报告摘要：**图文图像作为信息的重要载体，有着广泛的应用场景。但在实际应用过程中，面对复杂场景仍然存在一些挑战，包括复杂版面的理解、公式表格等复杂结构的识别等。本次报告将首先介绍科大讯飞在图文识别方向的布局，再结合实际应用过程中的复杂场景，介绍科大讯飞在公式识别、表格识别、图文理解方面的研究进展。在突破复杂场景中的技术难题后，图文识别技术将会迎来更多的应用场景，最后本报告对讯飞在图文识别方面的产业化应用进行介绍。



**讲者简介：**殷保才，博士，科大讯飞研究院副院长。负责科大讯飞计算机视觉方向的研究与落地工作，研究方向包括图文分析与识别、视觉交互与分析、多模态分析与理解等，带领团队在 MTWI 图文识别挑战赛、IDRID 眼底图分析、CVPR Object365 等国际竞赛获得 10 余项冠军，并在国际会议、期刊发表论文 20 余篇，申请专利 50 余项。获得 ISCSLP2022 最佳论文奖、吴文俊人工智能科技进步奖一等奖。作为子课题负责人、项目骨干参与多个国家、省部级科研项目。

### 苏劲松 厦门大学

**报告题目：**基于图文语义融合的多模态神经机器翻译研究

**报告摘要：**多模态机器翻译是近年来神经机器翻译研究的热点之一。在本次报告中，报告人将首先对多模态机器翻译的任务定义，应用场景和研究进展进行综述介绍，然后介绍报告人课题组在基于图文语义融合的多模态机器翻译模型构建方面的研究进展，最后对大模型时代的多模态神经机器翻译发展进行展望。



**讲者简介:** 苏劲松, 厦门大学信息学院教授, 博导, 国家特支计划青年拔尖人才, 福建省杰青, 厦门大学南强青年拔尖 A 类人才, 研究方向为自然语言处理, 文本生成, 机器翻译; 在 T-PAMI/AIJ/ACL 等权威期刊会议发表 CCF-A/B 论文 100 余篇; 担任中国中文信息学会理事、青工委委员、中国人工智能学会组织工作委员会委员、CCF 自然语言处理专业委员会委员; 担任 ACL2021/ACL2022/ACL2023/EMNLP 2019/EMNLP2020/EMNLP2022/COLING2022 领域主席, CCMT2021 程序委员会主席, 研究工作获得 2020 年中国中文信息学会钱伟长中文信息处理科学技术奖—汉王青年创新奖。

### 刘禹良 华中科技大学

**报告题目:** 浅析多模态大型模型在文字图像智能领域的应用现状、挑战及未来展望

**报告摘要:** 随着大型多模态模型的兴起, 自然语言处理和多模态学习领域取得了重大突破。本报告全面分析了现有开源多模态模型在文本识别、基于文本的视觉问题回答和关键信息提取等任务中的表现。定量评估显示, 多模态方法在文本识别等方面可以取得令人鼓舞的成果, 部分数据集甚至达到了业界领先水平。然而, 与领域特定方法相比, 多模态方法仍有明显的差距, 这表明为每个文字任务量身定制的专门技术仍具有关键意义。通过探讨这些模型在文本图像智能领域的优势和局限, 本报告旨在对文本领域应用多模态方法提供有价值的见解, 同时为改进零样本多模态技术的新想法的开发和评估提供基础。



**讲者简介:** 刘禹良, 华中科技大学人工智能与自动化学院, 研究员, 湖北省 A 类楚才。研究方向为人工智能, 主要聚焦在文字图像智能领域。主持青基 1 项, 2 项国家重点研发计划子课题负责人, 在 IEEE TPAMI, IJCV, NeurIPS, TIP, CVPR 等期刊会议发表论文 40 余篇, 一作及通讯 20 余篇。中国图象图形学报编委、CSIG 优博。担任 CCF-CV, CSIG 文档图像分析与识别专委会、CSIG-MV, CAAI-智能服务委员会委员。曾获 2 项互联网+金奖。研究成果应用于 Adobe 等知名企业。

## Workshop 6: ChatGPT 与计算机视觉

主席：谢凌曦（华为） 张东霖（江南大学） 李翔（南开大学）

时间：6月11日（周日）8:30-12:15 地点：B6 馆

时间	主持人	内容
8:30~9:30	李翔	本地政府环节
9:30~9:50	李翔	讲者：吴飞（浙江大学） 题目：ChatGPT：chat 的计算合成与涌现机理
9:50~10:10	李翔	讲者：白翔（华中科技大学） 题目：再谈 ChatGPT 那些事儿
10:10~10:30	李翔	讲者：车万翔（哈尔滨工业大学） 题目：ChatGPT——大语言模型向人类期望的对齐之路
10:30~10:50	李翔	讲者：鲁伟明（浙江大学） 题目：多模型协同的计算机视觉处理：以 HuggingGPT 为例
/	中场休息	
10:50~11:10	谢凌曦	讲者：郑志彤（OPPO） 题目：借鉴 ChatGPT 思想，小布助手在 AIGC 上的探索
10:50~11:30	谢凌曦	讲者：魏龙辉（华为） 题目：图文多模态预训练和生成技术进展
11:30~12:15	谢凌曦	Panel 嘉宾：吴飞（浙江大学）、白翔（华中科技大学）、车万翔（哈尔滨工业大学）、鲁伟明（浙江大学）、杨振宇（OPPO）、魏龙辉（华为）

## 吴飞 浙江大学

**报告题目：**ChatGPT: chat 的计算合成与涌现机理

**报告摘要：**人工智能于 1956 在达特茅斯启航后经历了两落三起曲折过程，包括评述人工智能“尚未产生当时承诺的重大影响、只能解决玩具类问题”的“莱特希尔 (Lighthill) 报告”、“爬上树梢不等于攀登月球”的早期神经网络受挫等故事。本报告回顾人工智能历史发展过程所带来的启示，同时介绍目前引起热议的人工智能现象级产品 ChatGPT 和 GPT-4 的技术特点、涌现机理和对人工智能未来提出的挑战。



工智能：模型与算法》。

**讲者简介：**浙江大学求是特聘教授，博士生导师。主要研究领域为人工智能、多媒体分析与检索。浙江大学上海高等研究院常务副院长、浙江大学人工智能研究所所长。国家杰出青年科学基金获得者（2016 年）、科技部科技创新 2030 “新一代人工智能”重大项目管理专家组成员和指南编制专家、教育部人工智能科技创新专家组工作组组长（2018.8-2020.12）、中国工程院院刊《Engineering》信息与电子工程学科执行主编、中国人工智能学会教育工作委员会主任委员，浙江大学第九届永平杰出教学贡献奖获得者。著有《人工智能导论：模型与算法》和《走进人工智能》（高等教育出版社），开设国家级首批线上一流课程《人

## 白翔 华中科技大学

**报告题目：**再谈 ChatGPT 那些事儿

**报告摘要：**ChatGPT 这一现象级人工智能产品已引起了全世界的强烈关注。本次报告在此前科普文章《浅析 ChatGPT: 历史沿革、应用现状及前景展望》的基础上，进一步总结和归纳了 ChatGPT 有关的技术进展和应用情况。希望回过头再谈 ChatGPT，能带给大家更深的认识和全新的视角。



**讲者简介：**华中科技大学人工智能与自动化学院教授，国家杰青获得者，IAPR Fellow，机器视觉与智能系统湖北省工程研究中心主任。主要研究方向向计算机视觉与模式识别、文档分析等。已在计算机视觉与模式识别领域一流国际期刊和会议如 PAMI、CVPR 等发表论文 80 余篇，谷歌总引用逾三万次，h 值 88。由于在场景文字检测与识别取得的显著成绩，2019 年在国际模式识别协会主办的国际文档分析与识别会议上获得 IAPR/ICDAR Young Investigator Award (青年研究者奖)。现任期刊 IEEE TPAMI, Int. J. of Document Analysis and Recognition, Pattern Recognition, Frontier of Computer Science, 中国科学-信息科学、自动化学报编委。现任 IAPR 混合会议特设委员会委员，IEEE SPS Wuhan Chapter 主席，曾任 CVPR、AAAI、ICPR 等人工智能主流国际会议的领域主席或竞赛主席十余次，并获得 AAAI-2019 Outstanding SPC Award，2014-2022 连续 9 年入选 Elsevier 中国高被引学者榜。

## 车万翔 哈尔滨工业大学

**报告题目：**ChatGPT——大语言模型向人类期望的对齐之路

**报告摘要：**2022 年 11 月 30 日，OpenAI 推出了全新一代的对话式人工智能工具——ChatGPT，因其表现出了惊艳的语言理解、生成以及知识推理等能力，获得了全网的热议和关注。ChatGPT 的成功表现，使人们看到了解决自然语言处理这一认知智能核心问题的一条可能的路径，并被认为是向通用人工智能目标迈出了坚实的一步，将对搜索引擎等应用

构成巨大的挑战，甚至将取代很多人的工作，更将颠覆很多领域和行业。那么，ChatGPT 究竟解决了什么科学问题，是如何解决该问题的，以及未来还会有哪些亟待解决的问题呢？希望通过本报告的粗浅分析，能部分回答以上的问题。



**讲者简介：**哈尔滨工业大学计算学部长聘教授/博士生导师，人工智能研究院副院长，国家级青年人才，龙江学者“青年学者”，斯坦福大学访问学者。现任中国中文信息学会理事、计算语言学专业委员会副主任兼秘书长；国际计算语言学学会亚太分会（AAACL）执委兼秘书长。目前承担国家自然科学基金重点项目、2030“新一代人工智能”重大项目课题等多项科研项目。著有《自然语言处理：基于预训练模型的方法》一书。曾获 AAAI 2013 最佳论文提名奖。负责研发的语言技术平台（LTP）已授权给百度、腾讯、华为等公司付费使用。2016 年获黑龙江省科技进步一等奖（排名第 2），2020 年获黑龙江省青年科技奖。

### 鲁伟明 浙江大学

**报告题目：**多模型协同的计算机视觉处理：以 HuggingGPT 为例

**报告摘要：**随着基于 Transformer 的预训练语言模型的发展，越来越多的工作将 Transformer 引入到计算机视觉任务中，并体现出了优异的性能。面对各类模态的数据以及众多繁杂的视觉任务，我们应该选择多模态统一大模型还是多模型协同工作方式？本报告将以 ChatGPT 为背景，以 HuggingGPT 为例，来探讨多模型协同的计算机视觉处理方式。



**讲者简介：**浙江大学计算机学院副教授，博士生导师，英国帝国理工学院访问学者，数字图书馆教育部工程研究中心骨干成员。主要研究领域包括人工智能、自然语言处理、跨媒体学习等，在 ACL、EMNLP、IJCAI、AAAI 等人工智能领域的著名国际会议和期刊上发表相关论文 60 余篇，曾获浙江省科技进步一等奖 2 项。

### 郑志彤 OPPO

**报告题目：**借鉴 ChatGPT 思想，小布助手在 AIGC 上的探索

**报告摘要：**随着大规模预训练技术的不断发展，在 NLP、CV 和跨模态领域有了比较好的发展，ChatGPT、SwintransformerV2、StableDiffusion、GPT4 是其中的明星代表，在工业界也产生了巨大的应用潜力。OPPO 在两年内先后启动了 CV 大规模预训练、NLP 大规模预训练、多模态预训练、AIGC 绘图和 AndesGPT 等项目，在 OPPO 小布助手和主题商店等应用场景研发了各种产品能力，产生了很大的业务价值。但是，也产生了不少 OPEN 的问题，比如“仅仅靠 ARLLM 是不是能产生 AGI？”、“ChatGPT 是不是适合整体搬到 CV 领域？”等。本报告将阐述上述问题，提出一些深层次的思考和探索方向，同时也展示 OPPO 多模态方向的一些成果。



**讲者简介：**OPPO 高级算法架构师，现任 OPPO 数智工程系统机器学习 TMG 主任、小布智能中心多模态学习负责人，促进小布从语音助手进化成多模态助手，打造 AndesGPT 多模态能力，落地了多个小布 AIGC 场景，包括小布绘画、锁屏壁纸生成、小布空间图片生成等。在机器学习领域有十几年经验，对 CV、NLP、音频、推荐系统等算法有深刻认知，对 AI 工程化有实操经验，十分关注通用智能的发展，特别是通用智能的关键一环多模态技术。硕士毕业于清华大学，本科毕业于中国人民大学。

### 魏龙辉 华为

**报告题目：**图文多模态预训练和生成技术进展

**报告摘要：**大规模预训练正在影响多个人工智能领域。近年来，多模态预训练模型逐渐兴起，并且展现了其在跨模态检索、AIGC、多模对话等方面的巨大应用潜力，吸引了很多工业界和学术界相关研究学者的关注。在本次报告中，讲者将简单介绍多模态理解、生成、对话三个维度预训练技术的发展脉络和相关前沿代表性工作。同时，讲者也将详细介绍华为云盘古大模型团队在多模态预训练方向上的几个代表性工作，以及对多模态预训练和相关领域未来发展的思考。



**讲者简介：**华为云盘古大模型团队高级研究员。2019 年毕业于北京大学信息科学技术学院数字媒体所。魏龙辉的主要研究领域为视觉自监督学习、多模态理解、生成、对话预训练技术、行人识别等相关方向，至今已在 CVPR、ICCV、ECCV、NeurIPS、AAAI、ACM MM、T-MM 等视觉或多媒体顶级会议与期刊发表相关论文二十余篇，谷歌学术统计引用量已达 2500 余次。其中，第一作者单篇最高引用据谷歌学术统计至今已达 1400 余次。担任 CVPR、ICCV、ECCV、NeurIPS、T-PAMI、T-IP、T-CSVT、T-MM 等多

个国际知名会议或期刊的审稿人。

## Workshop 7: 机器人具身智能

主席：弋力（清华大学） 王永才（中国人民大学） 杨旭（中国科学院自动化研究所）  
李超（江南大学）

时间：6月11日（周日）8:30-12:30 地点：B4馆

时间	主持人	内容
8:30-9:30	弋力	无锡国家高新技术产业开发区政产学研推介活动
9:30-9:55	弋力	讲者：孙富春（清华大学） 题目：物理特性驱动的机器人认知学习
9:55-10:20	弋力	讲者：蒋树强（中国科学院计算技术研究所） 题目：视觉导航：从模拟到现实
10:20-10:45	杨旭	讲者：卢策吾（上海交通大学） 题目：具身三维视觉PIE方案
10:45-11:10	杨旭	讲者：丛杨（中国科学院沈阳自动化研究所） 题目：3D视觉引导的机器人自主操作
11:10-11:35	王永才	讲者：徐婧林（北京科技大学） 题目：视频人体动作的细粒度分析与理解
11:35-12:00	王永才	讲者：付春江（优必选科技） 题目：人形机器人多模态应用
12:00-12:30	王永才	讲者：龚月（格灵深瞳） 题目：速度与“机”情——智能巡检机器人在列车运维中的应用
12:30-12:45	弋力	Panel 嘉宾：孙富春（清华大学）、蒋树强（中国科学院计算技术研究所）、卢策吾（上海交通大学）、丛杨（中国科学院沈阳自动化研究所）、徐婧林（北京科技大学）、龚月（格灵深瞳）

**孙富春 清华大学**

**报告题目：**物理特性驱动的机器人认知学习



**讲者简介：**清华大学计算机科学与技术系教授，博士生导师，IEEE/CAAI/CAA Fellow，国家杰出青年基金获得者；兼任清华大学校学术委员会委员，计算机科学与技术系长聘教授委员会副主任，清华大学人工智能研究院智能机器人中心主任。兼任担任国家重点研发计划机器人总体专家组成员，中国人工智能学会（CAAI）副理事长，中国自动化学会（CAA）和中国认知科学学会（IACS）常务理事，中国计算机学会（CCF）智能机器人专业委员会主任。兼任国际刊物《Cognitive Computation and Systems》主编，中国人工智能学会会刊《CAAI Artificial Intelligence Research》执行主编，国际刊物《IEEE Trans. on

Cognitive and Development Systems》，《IEEE Trans. on Fuzzy Systems》，《IEEE Trans. Systems, Man and Cybernetics》和《International Journal of Control, Automation, and Systems (IJCAS)》副主编或领域主编，刊物《Robots and Autonomous Systems》，《中国科学：F 辑》编委。

**蒋树强 中国科学院计算技术研究所**

**报告题目：**视觉导航：从模拟到现实

**报告摘要：**视觉导航是指具身智能体通过视觉实时感知环境，进行路径规划找到给定目标，这是具身交互任务中的一个重要步骤，是智能系统应用在现实世界中的一项重要能力，具有广泛应用潜力。受限于在现实场景中存在的训练数据不足、交互性不强、难以评测对比等问题，视觉导航等具身智能任务主要在模拟器环境下开展研究。本报告将首先介绍基于模拟器的视觉导航技术研究现状和进展，包括类别级/实例级物体导航、单物体/多物体导航等技术；其次将探讨视觉导航从模拟环境到现实环境的迁移与实施，介绍真实环境下视觉导航开展情况并给出演示。



**讲者简介：**中科院计算所研究员，博士生导师，国家杰出青年科学基金获得者，期刊《计算机研究与发展》、《JCST》、《ACM ToMM》编委，CCF 多媒体专委会秘书长、中国人工智能学会智能服务专委会副主任、ACM SIGMM 中国分会副主席、IEEE CASS 北京分会副主席。主要研究方向是图像/视频等多媒体内容分析、多模态智能技术和食品计算。主持承担科技创新 2030-“新一代人工智能”重大项目、国家自然科学基金等项目 20 余项。发表论文 200 余篇，获授权专利 18 项。先后获中国计算机学会科学技术奖、中国科学院青年科学家国际合作奖、CSIG 自然科学二等奖、吴文俊人工智能自然科学一等奖、北京市科技进步二等奖。

**卢策吾 上海交通大学**

**报告题目：**具身三维视觉 PIE 方案

**报告摘要：**该讲座介绍讲者具身三维视觉 PIE 方案。P (Perception) 介绍讲者全感知与交互感知工作。I (Imagination)，介绍讲者的物理世界概念驱动仿真推理框架。E (Execution) 介绍讲座通用元操作技能设想与工作（重点介绍：通用柔性物体操作与通用抓取工作）。

**讲者简介：**上海交通大学教授，博士生导师，2016 年或海外高层次青年引进人才，2018 年被《麻省理工科技评论》评为 35 位 35 岁以下中国科技精英（MIT TR35），2019 年获求是



杰出青年学者，2020 年获上海市科技进步特等奖（第三完成人），2021 获中国高被引学者，2022 年获 IROS 最佳论文之一（6/3579）。以通讯作者或第一作者在《自然》、《自然·机器学习》、TPAMI 等高水平期刊和会议发表论文 100 多篇；担任 Sciece 正刊，Nature 子刊，Cell 子刊等期刊审稿人，NeurIPS, CVPR, ICCV , ECCV , IROS , ICRA 领域主席。研究兴趣包括计算机视觉，机器人学习。

### 丛杨 中国科学院沈阳自动化研究所

#### 报告题目：3D 视觉引导的机器人自主操作

**报告摘要：**机器人自主操作能力不但是衡量机器人智能化水平的关键指标之一，对提高机器人在智能制造、家庭服务、医疗健康等领域的作业适应范围也具有重要意义。其中机器人感知和认知能力实现机器人自主操作的关键，而视觉特别是 3D 立体视觉和机器学习是机器人感知和认知的重要手段。虽然近年来涌现出许多令人兴奋的进展，但机器人感知和认知中的一些核心问题仍然没有得到很好解决，导致机器人还无法完成很多人类看似简单的工作，例如机器人视觉识别能力不足、自主操作泛化能力较弱等。报告结合机器人学国家重点实验室的背景和特点，主要阐述针对机器人感知和认知中的视觉识别和在线学习问题所开展的探索性研究工作。



**讲者简介：**丛杨，中国科学院沈阳自动化所研究员，国家杰青，博士生导师。担任机器人学国家重点实验室主任助理、辽宁省机器智能重点实验室主任。主要从事机器人视觉、机器学习、医学影像分析、大数据处理、机器人伺服等研究。担任中国图象图形学会理事，中国自动化学会青年工作委员会副主任等职务。2009 年起至 2011 年，先后在新加坡国立（NUS）和南洋理工大学（NTU）从事研究工作，2014 年在美国罗切斯特大学访问。先后主持国家重点研发计划项目、国家自然科学基金仪器专项、杰青、优青、重点、中科院课题多项。获得辽宁省自然科学一等奖、中国自动化学会自然科学奖一等奖、中国自动化学会青年科学家奖、辽宁省青年科技奖、辽宁省自然科学成果特等奖、中国科学院青年创新促进会优秀会员、辽宁省“百千万人才工程”、辽宁省“兴辽英才计划”青年拔尖人才等奖项。目前已在国内外期刊和会议上发表 80 余篇论文，参与出版 Springer 专著一部。目前，担任 IEEE Trans、《自动化学报》等国内外知名期刊的编委。

### 徐婧林 北京科技大学

#### 报告题目：视频人体动作的细粒度分析与理解

**报告摘要：**视频是人们观察和记录世界、进行内容分享和信息交互的重要载体，分析视频中人体动作对理解现实世界、服务人类社会具有重大意义，其作为基础性关键技术智能体育、智能安防、智能媒体、智能家居等领域具有广阔的应用前景。本报告首先介绍视频人体动作分析与理解的难点问题和方法；重点介绍视频体育运动定量评价与数据分析等问题，构建了首个细粒度竞技体育视频数据集 FincDiving，提出了基于过程感知的细粒度人体动作解析等方法，解决竞技体育运动定量评价的可回溯性问题；介绍视频人体动作意图难理解、人工标注不可靠等问题，提出了基于知识增强的意外动作定位等方法，突破现有模型稳定预测、泛化的局限性，提升了模型推理行为目标、认知他人计划的能力。



**讲者简介：**徐婧林，北京科技大学智能科学与技术学院副教授，西北工业大学博士、清华大学博士后。主要研究方向为视频动作理解、多视图分析与识别等，已发表 CCF A 类国际期刊和会议论文 14 篇，包括 IEEE TIP、TKDE、CVPR、AAAI、IJCAI 等。主持国家自然科学基金青年基金、中国博士后科学基金项目等多个项目，参与国家自然科学基金重点、面上、企业横向等多个项目。2022 年中国图象图形学学会优秀博士学位论文奖、2022 年西北工业大学优秀博士学位论文。担任《电子与信息学报》编委、中国自动化学会模式识别与机器智能专业委员会委员、中国图象图形学学会视觉大数据专委会委

员，担任国际权威期刊 (IEEE TPAMI、TIP、TNNLS、TCSVT、TMM 等) 和国际会议 (CVPR、ICCV、IJCAI、AAAI 等) 审稿人。

### 付春江 优必选科技

**报告题目：**人形机器人多模态应用

**报告摘要：**人形机器人的多模态应用是指在实际应用中结合多种感知和交互方式，以提高机器人的适应性和功能。这些模态包括视觉、听觉、触觉、运动和语言等，人形机器人需要具备高度的自主性来完成不同的任务，近年来的研究和实践表明语言大模型和多模态感知是实现这一目标的关键技术之一。报告结合人形机器人背景和特点，主要阐述针对人形机器人自主操作的初步探索，应用语言大模型与多模态感知，人形机器人实现人机自然语音交互，任务分解，路径规划，定位导航，动作和步态控制，物体检测，物体识别，灵巧操作，触觉传感与机械臂控制自主作业。为了人形机器人的高度适应性，以便在不同的环境和场景中完成任务，我们将来的工作会在机器人感知，认知和决策中的一些核心问题，如多语言多任务语音交互技术，增强机器人视觉识别能力，自主操作泛化能力，精确的物体操控和操作技术，运动控制技术，情感识别技术和情感表达技术在人形机器人上进行进一步探索。这些多模态应用使得人形机器人在日常生活、医疗、教育、娱乐，工厂，物流等领域都有着广泛的应用前景，可以为人类带来更加便利和高效的服务。



**讲者简介：**日本大阪大学博士，曾与美国斯坦福大学、日本京都大学、意大利 IIT 研究所、英国 Deepmind 公司等交流合作。师从野村泰伸教授，日本生体工学权威专家 (Global COE in silico IDE 项目总负责人)。付博士于 2017-2021 任职国际机器人顶级研究机构——本田公司先端技术研究所 (曾开发 ASIMO 机器人)，回国后任之江实验室 PI 研究院，带领视觉组、控制组及大双足本体组开展研究工作。现任深圳优必选科技有限公司副总裁、人形机器人事业部负责人，负责人形机器人从战略制定到业务经营的全面工作，包括产品定义、研发、生产、销售及带领团队研发下一代大型双足仿人机器人等。付博士十多年来一直致力于仿人机器人运动控制，双足行走建模，控制及稳定性复杂性分析。曾研究开发过的领域有人体手臂阻抗控制，手臂最优控制，人体行走控制及稳

定性分析，行走复杂性，行走辅助器械，人体全身与车辆座椅动力学交互，双足机器人行走控制，变阻抗关节控制等。

## 龚月 格灵深瞳

**报告题目：**速度与“机”情——智能巡检机器人在列车运维中的应用

**报告摘要：**中国目前有数千组动车和城轨列车，每年承载着几十亿人次的出行。为保证公共交通的安全与便捷，轨道交通的运维工作变得至关重要。列车运维场景具有环境结构化弱、通道狭窄、地面情况复杂、巡检采集精度要求高等特点。为此，格灵深瞳研发了新一代的智能巡检机器人，结合了自主导航、在线感知、手眼协同、虚拟示教等技术。使得机器人可以适应列车检修的复杂场景，做到快速部署，稳定运行。同时，搭配结合了深度学习与立体视觉的Deep3D算法引擎，解决列车故障诊断碎片化程度高，收敛速度慢等问题。本次报告将介绍列车运维面临的场景挑战，以及格灵深瞳利用以立体视觉技术为中心的解决方案。同时，也将针对列车的缺陷检测问题的难点做出分享。



**讲者简介：**龚月，北京格灵深瞳信息技术股份有限公司，智能机器人与工业检测开发部负责人，VP。从业8年，主要从事机器人和工业缺陷检测方面的产品研发工作，涉及技术方向：机器人自主导航，机械臂自主规划与控制，主动感知，3D重建，SLAM，缺陷检测，点云处理，深度学习等，目前聚焦在在轨交运维方向。领导开发的主要产品有：列车智能巡检机器人、线路智能巡检机器人、列车360动态检测系统、鹰眼列车故障分析引擎、小灵瞳学远程临场机器人等，目前产品已覆盖动车、地铁、货运机车等场景，在全国多个城市均有落地。

## Workshop 8: 物联网与工业应用

主席：吴帆（上海交通大学）陈旭（中山大学）黄倩怡（中山大学）朱书伟（江南大学）

时间：6月11日（周日）13:30-18:00 地点：B4 馆

时间	主持人	内容
13:30-13:35		论坛主席开场致辞
13:35-14:35	陈旭	本地安排活动
14:35-14:55	吴帆	讲者：杨旸（特斯联） 题目：面向用户算力需求特征的 6G 网络智能架构
14:55-15:15	吴帆	讲者：何源（清华大学） 题目：从应用视角看物联网技术发展——射频计算初探
15:15-15:35	陈旭	讲者：杨树森（西安交通大学） 题目：边云场景大数据计算：算法、平台与应用
15:35-15:55	陈旭	讲者：谢磊（南京大学） 题目：“工业视觉与边缘智能”助力新型电力系统
15:55-16:00	中场休息	
16:00-16:20	黄倩怡	讲者：董玮（浙江大学） 题目：大语言模型下的物联网低代码开发
16:20-16:40	黄倩怡	讲者：王晓飞（天津大学） 题目：云边端泛在协同人工智能
16:40-17:00	黄倩怡	讲者：丁二锐（百度） 题目：数字人/物/场生成的产业化实践
17:00-18:00	黄倩怡	Panel 嘉宾：杨旸（特斯联）、何源（清华大学）、杨树森（西安交通大学）、谢磊（南京大学）、董玮（浙江大学）、王晓飞（天津大学）、丁二锐（百度）、陈旭（中山大学）

## 杨旸 特斯联

**报告题目：**面向用户算力需求特征的 6G 网络智能架构

**报告摘要：**在未来的 6G 时代，用户千变万化的算力需求将成为 6G 网络架构和服务质量保障的重大挑战。本报告首先研究了当前云计算平台上海量任务的算力需求分布和典型特征，分析推导出大尺度和小尺度参数模型。其次，基于通算融合的网络智能架构，实现云网边协同互助，优化部署分布式的人工智能和算力资源。最后，大规模仿真结果显示智慧内生的网络智能架构可以大幅度提升多维度服务体验 and 用户满意度。



**讲者简介：**杨旸博士，IEEE Fellow，现任鹏城实验室兼职研究员、特斯联科技集团首席科学家、深圳市智慧城市科技发展集团有限公司高级顾问专家。研究领域包括 5G/6G 移动通信系统、智能物联网、多层次算力网络，开放无线测试验证平台等。已申请了 120 多项科技发明专利，发表了 300 多篇学术论文，出版了六部中英文专著。杨旸曾在香港中文大学、英国布鲁内尔大学、伦敦大学学院担任教职。目前，依托深圳市智慧城市科技发展集团有限公司，他牵头承担了国家重点研发计划“物联网与智慧城市关键技术及示范”重点专项项目《面向大湾区智慧城市群的 5G 泛在物

联基础设施建设及示范》

## 何源 清华大学

**报告题目：**从应用视角看物联网技术发展——射频频谱计算初探

**报告摘要：**应用是物联网技术进步的根本动力。经过二十多年发展，物联网技术发展水平离万物互联的愿景仍有明显的距离。有限的感知能力、过高的开销制约物联网向更广阔空间部署和应用。为了应对上述挑战，我们尝试从应用发展的角度思考物联网技术演进的趋势，探讨以射频频谱资源为中心的面向物联网的新型计算理论与关键技术。本报告将分享我们在物联网应用系统研发和前沿技术研究中取得的成果，讨论“射频频谱计算”的研究空间和方向。



**讲者简介：**何源，清华大学软件学院院长聘副教授、博导，CCF 高级会员，IEEE 高级会员。ACM SIGBED China 常委，中国计算机学会物联网专委会执委。分别于中国科学技术大学、中国科学院软件研究所、香港科技大学获得工学学士、工学硕士和博士学位。研究领域涉及物联网、无线网络、移动和普适计算等。担任《计算机研究与发展》、JCST、ACM Transactions on IoT、IEEE IoT Journal 等期刊编委，SECON 2022、EWSN 2019、DCOSS 2018、ICPADS 2017/2018 等国际会议程序委员会主席/副主席，CPS-IoT Week 研讨会主席，CPS-IoT Week 和 MobiSys 等会议出版主席，以及 SenSys、IPSN、INFOCOM、WWW 等国际会议的程序委员会委员，三次被评为 INFOCOM 杰出程序委员会委员。作为项目或课题负责人承担国家自然科学基金优秀青年基金、联合基金重点项目、面上项目以及国家重点研发计划重点专项等。发表论文 150 余篇，被引用 5000 余次。研究成果获得包括 SenSys 2022 和首届 IEEE 国际工业互联网大会在内的多个学术会议的最佳论文奖。

## 杨树森 西安交通大学

**报告题目：**边云场景大数据计算：算法、平台与应用

**报告摘要：**当前大数据技术面临广域网延迟、碎片化分布、跨主体合法合规流通困难等一系列挑战。边云协同环境计算技术是解决上述问题的有效途径之一。在实时性问题上，数据处理决策下沉到边缘侧，天然降低了广域网和城域网延迟；在样本稀缺和隐私

保护方面，在数据不出本地的前提下，进行跨主体联合建模，实现数据“可用但不可见”。报告将介绍团队近年在边云大数据分析领域的研究进展，包括算法设计、平台构建和在我国头部企业的应用案例。



**讲者简介：**杨树森，西安交通大学教授、博导，英国帝国理工大学博士，大数据算法与分析技术国家工程实验室常务副主任，智能网络与网络安全教育部重点实验室副主任，国家重点研发计划“高性能计算”重点专项总体组专家，华为公司重大项目负责人，国家青年千人。研究方向为边云协同大数据分析，提出了时空多尺度大数据聚类与在线决策，边云协同机器学习与视频分析，跨异质主体联邦学习等一系列创新算法与系统。发表了 CCF A/中科院 1 区/IEEE Trans 论文 50 余篇，包括 National Science Review, MOBICOM, SIGMOD, INFOCOM, IJCAI, TON, JSAC, TPDS, TMC, TIFS 等。研制了多个原型系统，部分系统在华为、中航发、南网、广汽等我国头部企业成功落地应用。近五年主持国家、企业科技项目 4000 余万元（千万级 1 项，百万级 7 项）。曾担任 Intel Europe 研究员，英国利物浦大学助理教授（终身教职），阿里巴巴访问学者。荣获首届阿里巴巴“达摩院青橙奖”（2018 年全国 9 人）和华为公司“2020 技术合作与突破卓越成就奖”（唯一获奖人）。

### 谢磊 南京大学

**报告题目：**“工业视觉与边缘智能”助力新型电力系统

**报告摘要：**本次报告将介绍新时代电力系统在数字化转型过程中所面临的新机遇和新挑战，以及近年来研究团队在工业视觉与边缘智能方面的研究进展，并通过“工业视觉赋能泛在感知”和“边缘智能助力数据计算”两方面的内容介绍，具体阐述在新型电力系统以及数字化建设中的关键技术及其具体应用案例。



**讲者简介：**谢磊，南京大学计算机科学与技术系教授，博士生导师，入选“教育部青年长江学者”、南京大学首届“仲英青年学者”，CCF 杰出会员，CCF 杰出演讲者。担任中国计算机学会普适计算专委会秘书长、南京大学-南方电网数字平台科技有限公司工业互联网联合实验室主任。曾获江苏省科学技术奖一等奖（2016 年、2019 年两次获得）、江苏省教学成果奖（高等教育类）特等奖、高校计算机专业优秀教师奖励计划、江苏省六大人才高峰创新人才团队、南京大学五四青年奖章、国际会议 MobiQuitous 最佳论文奖、国际期刊 TPDS 的 Spotlight Paper 等荣誉称号。主要研究领域为智能感知计算与边缘智能，目前共发表论文 100 余篇，包括国际一流学术会议 ACM MOBICOM、ACM UBICOMP、IEEE INFOCOM，国际一流学术期刊 ACM/IEEE TON、IEEE TMC 等。近年来，基于“无线感知”、“多模态感知”和“边缘智能”的研究成果为能源互联网和工业互联网领域“感传算一体化”的发展实现了基础理论突破，提供了核心技术支撑，并在多家大型制造和能源企业开展示范应用。相关成果荣获由“互联网+”和“挑战杯”等赛事的特等奖和一等奖，并受到中央电视台、新浪网、搜狐网报导。

### 董玮 浙江大学

**报告题目：**大语言模型下的物联网低代码开发

**报告摘要：**本报告从编程开发这一计算机科学与技术的核心研究话题展开，回顾无线传感网编程开发的重要历史性工作及启发。在此基础上，针对云边融合物联网低代码开发面临的协议异构、云-端执行环境异构等关键挑战，汇报本研究团队在云边融合一体开

发、跨技术统一协议架构、轻量化 WebAssembly 执行引擎等方向的科研探索。最后，结合最近快速发展的 ChatGPT 等 AI 技术，对未来基于大语言模型的智能物联系统以及开发方式进行了思考和展望。



**作者简介：**董玮，教授，博士生导师，中国计算机学会(CCF)高级会员，CCF 物联网专委会常务委员、CCF 互联网专委会委员、ACM SIGAPP China 专委会委员、ACM China 杭州分部执行委员、浙江省计算机学会嵌入式与物联网技术专委会秘书长。获全国优博论文提名，入选浙江大学求是青年学者、CCF-Intel 青年学者提升计划、浙江省 151 人才工程、国家及浙江省青年人才计划。主要研究方向包括物联网、智能边缘计算、计算机网络等。承担国家重点研发计划课题、国家自然科学基金、浙江省杰出青年基金等国家和省部级项目 10 多项。在 ACM MobiCom、MobiSys、UbiComp、USENIX NSDI、IEEE INFOCOM 等著名国际会议以及 IEEE/ACM Trans on Networking、IEEE Trans on Mobile Computing 等著名国际期刊上发表论文 160 余篇，其中 CCF A 类及 IEEE/ACM Trans 论文 70 余篇，3 次获 IEEE/ACM 国际会议最佳论文奖/最佳视频展示奖。授权发明专利 30 余项。主编《从创意到原型：物联网应用快速开发》，入选浙江省十三五新形态教材、浙江省十四五新工科重点教材、工信部十四五规划教材。指导的多位博士生获华为天才少年计划(2019)、ACM SIGAPP China 优秀博士论文(2022)、CCF 优秀博士论文(2022)等。近年来致力于云边缘融合的物联网软件与协议的研究，相关成果得到 IEEE Spectrum、ACM TechNews 等主流科技媒体的广泛报导，并在阿里云、华为、海康威视等头部企业得到深入应用，取得良好的经济和社会效益。获浙江省科技进步一等奖、浙江省计算机学会优秀成果奖、中国计算机学会科技进步杰出奖。

### 王晓飞 天津大学

**报告题目：**云边缘泛在协同人工智能

**报告摘要：**当前面向泛在协同人工智能的存算一体、云网一体技术正在快速发展，面对复杂多样的边缘和终端设备，面向泛在人工智能的云边缘算力资源类别复杂多样、算力需求庞大等挑战，探索泛在协同人工智能技术与应用。针对算力分布异构性强、计算协同效率低下、资源优化缺乏保障、产业生态维护性差四方面的问题，研究多级异构网络基础设施的算力感知与融合技术，实现泛在自适应分布式协同 AI 计算，保证分布式异构异质资源优化及公平调度，最终形成并维护灵活可追溯的 AI 数据计算模型生态体系，进而满足 AI 的泛在计算需求。有望为面向泛在协同人工智能技术提供新的理论和方案，引领新型云边缘 AI 算力管理理论的研究范式。



**作者简介：**王晓飞，天津大学智能与计算学部教授、博导，国家海外高层次人才引进计划优秀人才，获中国计算机学会分布式专委会青年创新奖。2017 年 IEEE 通讯协会 Fred W. Ellersick Prize 年度最佳杂志论文奖获得者，ACM Rising Star (Tianjin)，天津大学北洋青年学者。主要研究边缘智能理论、边缘计算系统架构、云边协同算法等，发表高水平科研论文 130 余篇(含 SCI 论文 70 余篇)，其中包括 ESI 高引 6 篇，热引 1 篇，谷歌学术引用 5200 次，申请发明专利 20 余项。担任多个一二区期刊副编委和专刊编委，主持了国家自然科学基金青年和面上项目、科技部重点研发计划课题和子课题等，还负责了中国电信、华为、圣戈班、中国电科院等企业项目二十余项，出版领域内第一本边缘智能书籍《Edge AI》被列为 Springer 的 2020 年中国高影响力作者图书 Top10，在国内外边缘计算相关学术和产业大会做主旨汇报和论坛发言五十余次。

## 丁二锐 百度

**报告题目：**数字人/物/场生成的产业化实践

**报告摘要：**AIGC 技术发展日新月异，正在快速改变内容生产和交互的方式。在视觉内容上，围绕核心的人/物/场内容，如何低成本生成是视觉 AIGC 产业规模化的关键。本报告以百度视觉技术在图像/视频生成、图形渲染、3D 生成等方向的积累为基础，以座舱、终端大屏、直播等为应用背景，重点介绍数字人建模(人像/发型等)、驱动(单目/智能驱动)及羽量化渲染，其次介绍数字物/场景的多视角、跨模态建模及编辑，最后介绍融合人/物/场的创新生成案例。



**讲者简介：**丁二锐，博士，正高级工程师，百度视觉技术部总监。曾获国家技术发明奖二等奖、中国专利银奖、中国电子学会科技进步一等奖、百度“最佳经理人”奖，相关论文曾获 ICDAR/FG 最佳(学生)论文奖。带领部门还获得北京市发明专利一等奖及 40 余项国际比赛冠军，在交通感知/智感超清/文字识别/数字人等领域打下坚实基础，由大模型等技术支撑的计算机视觉公有云连续 3 年被 IDC 评为市场第一。

## Workshop 9: 混合现实中的视觉与学习技术

主席：周晓巍（浙江大学） 胡瑞珍（深圳大学）

时间：6月11日（周日）13:30-17:00 地点：兰花厅A

时间	主持人	内容
13:30-13:55	周晓巍	讲者：许威威（浙江大学） 题目：神经网络渲染技术进展报告
13:55-14:20	周晓巍	讲者：王贝贝（南京理工大学） 题目：真实感材质的表示与恢复
14:20-14:45	周晓巍	讲者：刘利斌（北京大学） 题目：多模态和物理驱动的数字人动画
14:45-15:05	周晓巍	Spotlight
15:05-15:13	周晓巍	企业宣讲：蚂蚁技术研究院
15:13-15:20	茶歇	
15:15-15:45	周晓巍	讲者：潘焱（上海交通大学） 题目：混合现实环境中3D角色交互技术研发
15:45-16:05	周晓巍	讲者：许岚（上海科技大学） 题目：神经辐射场和虚拟数字人
16:05-17:00	周晓巍	Panel 嘉宾：许威威（浙江大学）、王贝贝（南京理工大学）、刘利斌（北京大学）、潘焱（上海交通大学）、许岚（上海科技大学）

## 许威威 浙江大学

**报告题目：**神经网络渲染技术进展报告

**报告摘要：**神经网络渲染技术采用深度学习端到端的方式生成图像，基于神经隐式表达简化了传统渲染技术对场景几何与材质精确表达的要求，近年来在基于图像的绘制、逆向重建等领域获得了快速的发展。本报告聚焦于介绍神经网络渲染技术在场景重建、材质计算、人体重建以及网络训练加速技术上的进展，并对大场景条件下分布式神经隐式表达构建进行探讨。



**讲者简介：**现任浙江大学计算机科学与技术学院 CAD&CG 国家重点实验室特聘教授，教育部长江学者，曾任日本立命馆大学博士后，微软亚洲研究院网络图形组研究员，杭州师范大学浙江省钱江学者特聘教授。主要研究方向为计算机图形学、数字孪生与虚拟现实。在国内外高水平学术会议和期刊发表论文 100 余篇，其中 ACM Transactions on Graphics、IEEE TVCG、IEEE CVPR、NIPS 等 CCF-A 类论文 50 余篇。获中国和美国授权专利 18 项。2014 年起受国家自然科学基金优秀青年基金资助，主持国家自然科学基金重点项目一项，获浙江省自然科学二等奖一项，2022 年入选教育部长江学者奖励计划。所研发的三维场景重建以及渲染技术应用于华为河图、百度阿波罗自动驾驶、杭州先临三维科技高精度扫描仪和深圳望尘科技等公司的产品。

## 王贝贝 南京理工大学

**报告题目：**真实感材质的表示与恢复

**报告摘要：**现实世界中的材质是丰富多彩的，在混合现实等应用中通常需要将现实世界中的材质外观具有真实感地呈现出来。这就一方面需要真实感的材质表示模型，另外一方面需要高质量的材质恢复方法。神经网络为材质的表达和恢复都带了新的思路。在本次报告中，我们将介绍近两年团队在材质表达和恢复方面的相关工作，包括多层材质的神经网络表示模型、BTF 的双表面表示以及轻量级获取以及面向布料材质的轻量级获取方法等。



**讲者简介：**南京理工大学，副教授。研究方向为计算机图形学渲染方向，研究主题聚焦于材质建模和高效渲染方法等。分别于 2009 年和 2014 年在山东大学获得学士和博士学位，曾在 INRIA 从事博士后研究。目前已在计算机图形学领域期刊和会议上发表学术论文 30 余篇。主持国家自然科学基金（青年和面上项目）、科技创新 2030—“新一代人工智能”重大项目子课题、省部级科研项目和公司的研发项目。作为骨干参与 Disney 游戏 Infinity 3 的研发，亮片材质实时渲染应用于迪士尼游戏。担任 SIGGRAPH 2023 程序委员会委员、GAMES 常务执行委员、中国图象图形学报青年编委。入选 2020 年中国图学会“青年托举”计划。个人主页：<https://wangningbei.github.io>

### 刘利斌 北京大学

**报告题目：**多模态和物理驱动的数字人动画

**报告摘要：**生成逼真的人类行为是计算机动画领域的一个基本问题，同时也是许多新兴领域（如数字人、元宇宙、AIGC 等）中最具挑战性的技术之一。在过去的一年中，我们看到了扩散模型、ChatGPT 等多模态模型在语言和图像生成领域的巨大成功，但如何将多模态信息引入角色动画并生成真实可信的动作依然是一项尚未解决的问题。本报告将简述计算机角色动画的基本知识和该领域近年来的主要进展，并介绍我们近期在语音文字驱动的动作生成、多模态动作编辑、以及基于物理仿真的生成式运动控制策略等方面的工作。



**讲者简介：**刘利斌，北京大学智能学院助理教授，2014 年博士毕业于清华大学，之后曾在加拿大英属哥伦比亚大学及美国迪士尼研究院进行博士后研究，后曾加入美国硅谷创业公司 DeepMotion Inc. 担任首席科学家。主要方向是计算机图形学，特别是数字人建模与动画、物理仿真、运动控制以及相关的优化控制、机器学习、增强学习等领域，曾获得 SIGGRAPH Asia 2022 最佳论文奖。他曾多次担任图形学领域重要国际会议如 SIGGRAPH (North America/Asia)、PacificGraphics、SCA 等的论文程序委员，以及图形学领域主要会议和期刊的审稿人。

### 潘焱 上海交通大学

**报告题目：**混合现实环境中 3D 角色交互技术研发

**报告摘要：**在迪士尼、皮克斯等主流动画工作室，动画师通常基于窗口、图标、菜单和指示设备（WIMP）交互界面，运用 Maya, Houdini 等商业软件制作动画过程复杂、成本高昂。尤其在 VR 动画制作中，360° 全景空间任何一点都可以成为观众观看的视点，传统动画采用镜头组接的制作方法已不再适合。申请人运用混合现实技术突破了传统基于 WIMP 动画制作交互界面不直观的问题，另外，针对 3D 智能角色呈现，从算法层面解决了 2D 屏幕固有的深度信息失真的难题，无需眼镜或其它复杂设备辅助，提高了用户与 3D 智能角色交互体验。

**讲者简介：**上海交通大学约翰霍普金斯计算机科学中心副教授。2015 年博士毕业于伦敦大学学院。曾任伦敦大学学院虚拟环境和计算机图形组高级研究助理。在迪士尼研究院参与主研华特迪士尼幻想工程，致力于 AI Character 核心项目研发，创新成果申请美国专利并已转化应用，且以此获得“Disney Inventor Award”创新奖。主要研究方向为虚拟现实，增强现实，动画制作，计算机图形，人机交互。曾担任 IEEE VR/TVCG 程序委员会委员，在 ACM SIGCHI, IEEEVR 等重要国际学术会议多次做口头报告，是顶级会议 ACM SIGCHI, IHCS 等 10 个学术会议和期刊的特约编辑和审稿人。



许岚 上海科技大学

**报告题目：**神经辐射场和虚拟数字人

**报告摘要：**随着神经表示技术和生成式人工智能的发展，图形图像渲染领域也迎来了重要的发展变革，在高质量的场景重建和真实渲染等方面取得了突破性进展，为虚拟数字人、元宇宙等等应用场景带来新机遇，需求也变得越发迫切。本次报告回顾过去一年课题组在数字人建模和渲染上的系列进展，重点探讨神经辐射场渲染和生成技术，展望神经网络建模和渲染技术对数字人领域的影响和未来发展方向。



**讲者简介：**上海科技大学信息科学与技术学院终身助理教授、研究员、博士生导师、MARS 实验室主任。2015 年从浙江大学毕业后于 2020 年在香港科技大学获得博士学位。研究方向聚焦于计算机视觉、计算机图形学和计算摄像学，目标是捕捉、感知并理解复杂现实世界中以人为中心的动态和静态场景，致力于光场智能重建理论与技术的研究，突破了动态神经辐射场和虚拟数字人的一批核心关键技术，率团队研制了系列光场装置，为新一代人工智能推动的轻量化超写实数字人提供了新范式，研究成果在多家明星初创公司得到应用，研制装备部署于北京冬奥。在 CVPR、SIGGRAPH、

IEEE TPAMI 等顶级刊物发表数十篇文章，并多次担任人工智能顶级会议 CVPR、ICCV、AAAI 等领域主席。

## Workshop 10: 目标检测与分割

主席：贾伟（合肥工业大学）李成龙（安徽大学）黄岩（中国科学院自动化研究所）丛润民（山东大学）

时间：6月11日（周日）08:30-12:10 地点：B5馆

时间	主持人	内容
8:30~9:00	李成龙	讲者：程明明（南开大学） 题目：粒度自适应的图像感知技术
9:00~9:30	李成龙	讲者：张兆翔（中国科学院自动化研究所） 题目：基于多传感器融合的视觉物体检测与分割
9:30~10:00	黄岩	讲者：张磊（重庆大学） 题目：开放环境视觉感知之开放域目标检测
10:00~10:30	黄岩	讲者：徐天阳（江南大学） 题目：信息牵引下的视觉目标定位与识别
10:30~10:40	中场休息	
10:40~10:48	黄岩	金牌赞助商“极视角”宣讲
10:48~11:18	黄岩	讲者：元玉慧（微软亚洲研究院） 题目：统一计算机视觉识别任务的 DETR 网络结构
11:18~11:48	丛润民	Poster Spotlight
11:48~12:15	丛润民	Panel 嘉宾：程明明（南开大学）、张兆翔（中国科学院自动化研究所）、张磊（重庆大学）、徐天阳（江南大学）、元玉慧（微软亚洲研究院）

## 程明明 南开大学

**报告题目：**粒度自适应的图像感知技术

**报告摘要：**图像中蕴含着丰富的多粒度信息。不论是简单的分类任务，还是更加复杂的目标检测、语义分割、物体跟踪、姿态估计等，几乎所有的视觉感知任务都需要丰富的多粒度信息的共同配合才能够进行鲁棒的判断。卷积神经网络中，小尺度信息可以通过卷积操作进行高效的建模，大尺度特征的建模通常计算量需求较大。本报告面向视觉感知任务需求，重点介绍如何高效地提取多粒度信息，并且通过神经网络架构高效搜索的方式对神经网络中的不同层自适应地选择特征提的粒度。本报告所介绍的粒度自适应视觉感知技术在多种计算机视觉任务中展现出了优越的性能，包括但不限于：图像分类、目标检测、语义分割、实例分割、关键点估计等。这些技术的应用范围也不局限于视觉信息处理，也被众多国际同行应用在语音合成、蛋白质结构预测、对抗样本鲁棒性研究等领域。



**讲者简介：**程明明，现任南开大学教授，计算机系主任。主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是计算机视觉和计算机图形学，在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇（含 IEEE TPAMI 论文 30 余篇），h-index 为 73，论文谷歌引用 3.8 万余次，单篇最高引用 4600 余次，多次入选中国高被引学者和全球高被引科学家。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得多项省部级科技奖励。现担任中国图象图形学学会副秘书长、天津市人工智能学会副理事长和 SCI 一区期刊 IEEE TPAMI, IEEE TIP 编委。

## 张兆翔 中国科学院自动化研究所

**报告题目：**基于多传感器融合的视觉物体检测与分割

**报告摘要：**视觉物体检测与分割是计算机视觉领域的核心问题，具有重要理论意义和应用价值。当前，在以自动驾驶、无人机、服务机器人等为代表的具身智能系统中，视觉场景感知涉及的传感器往往类型多样、数目多样。在多种传感器条件下，如何针对特定的传感器类型设计高效的视觉场景感知算法，如何对多种传感器的信息加以有机融合，具有重要的创新价值，也是推动实际应用性能的关键。本报告着重介绍我们课题组近期在摄像头、激光点云、时空信息融合、多传感器信息表征与融合等方面的一系列视觉物体检测与分割新方法，探索如何提升算法性能、降低模型标注、实现信息的高效表征。本报告会对该方向进行总结与展望。



**讲者简介：**张兆翔，中国科学院自动化研究所研究员、博士生导师，入选教育部长江学者，国家万人计划青年拔尖人才等。主要研究方向包括脑启发的神经网络建模、视觉认知学习、面向开放环境的场景感知与理解，在本领域 TPAMI、IJCV、JMLR 等顶刊和 CVPR、ICCV、ICLR、NeurIPS 等顶会发表论文 100 余篇，授权专利 20 余项，承担了国家自然科学基金重点项目、国家自然科学基金企业联合重点项目、国家重点研发项目等一系列国家级科研项目，是 IEEE 高级会员，中国计算机学会 CCF 杰出会员、中国人工智能学会 CAAI 杰出会员、中国计算机学会 CCF 杰出演讲者，担任或曾担任 IEEE T-CSVT、IEEE T-BIOM、Pattern Recognition 等知名期刊编委，是 CVPR、ICCV、NeurIPS、AAAI、IJCAI、ACM MM 等知名国际会议的领域主席（Area Chair）。

### 张磊 重庆大学

**报告题目：**开放环境视觉感知之开放域目标检测

**报告摘要：**目标检测作为计算机视觉中最基础的任务之一，已在深度学习引擎的推动下得到快速发展，检测精度及泛化能力显著提升。尽管如此，在实际应用中，由于数据不确定性、环境不可控性以及算法特异性，现有目标检测算法在开放域下的适应能力依然较差，难以满足开放、动态、复杂物理环境下的视觉感知与应用。本报告从迁移学习角度，包括基于域适配和域泛化的研究思路展开，汇报团队在开放域下的目标检测研究进展。



**讲者简介：**张磊，重庆大学微电子与通信工程学院教授，博导。主要研究兴趣为开放环境视觉感知中的域迁移泛化和模型鲁棒性问题。发表 IEEE Transactions (TPAMI、IJCV、TIP 等) 以及 CVPR/ICCV/ECCV/ICML/AAAI 等论文 100 余篇，专著 1 部，被引用 5 千余次。担任 IEEE Trans. Instrumentation and Measurement and Neural Networks 等期刊编委以及 ACM MM/CVPR/ICCV/ICLR/ICML/NeurIPS/AAAI/IJCAI 等会议领域主席或程序委员等，曾获吴文俊人工智能优秀青年奖、ACM SIGAI 中国新星奖、重庆市十佳科技青年奖。以第 1 完成人获得吴文俊人工智能自然科学奖和重庆市自然科学奖 2 项。

### 徐天阳 江南大学

**报告题目：**信息牵引下的视觉目标定位与识别

**报告摘要：**随着神经网络在计算机视觉领域中的广泛应用和大规模标注数据集的持续发布，视觉目标的定位和识别性能获得了显著提升，如何通过显示的建模设计将期望的语义观从输入中有效提取是关键问题。本报告将从信息牵引的角度出发，重点介绍多模态场景下的视频目标跟踪和视频识别模型，构建信息选择、信息融合、信息矫正的牵引模式，探索多模态交互、时序表观建模、非欧几何表示等机制，提升模型性能和可解释性，并阐述报告人对相关方向的理解和思考。



**讲者简介：**徐天阳，博士，江南大学人工智能与计算机学院副教授。2019 年毕业于江南大学模式识别与智能系统专业，2019 年在英国萨里大学 CVSSP 从事博士后研究工作，2021 年加入江南大学人工智能与计算机学院。研究方向为模式识别与计算机视觉，主要包括视频理解与流形学习。已发表期刊和会议论文 30 余篇，包括 IEEE T-PAMI、T-IP、T-IFS、T-KDE、IJCV、ICCV、AAAI 等，获中国图象图形学会 2021 年优秀博士论文奖，指导学生获 PRCV2022 最佳学生论文奖。在 CVPR、ICCV、ECCV、ACMMM 举办的相关学术竞赛中获冠军 5 次，亚军 3 次，季军 2 次，包括视觉目标跟踪竞赛 VOT、反无人机竞赛 Anti-UAV、多模态视频

分析竞赛 MMVRAC 等。

### 元玉慧 微软亚洲研究院

**报告题目：**统一计算机视觉识别任务的 DETR 网络结构

**报告摘要：**TDetection TRansformer (DETR) 已经被成功用于解决各种基础计算机视觉任务，包括但不限于二维视觉识别任务、三维视觉识别任务，以及视觉-语言理解任务等。DETR 成功包含三个要素，分别是 (1) 端到端的架构，消除了对复杂手工组件的依赖 (如 NMS 和 Anchor Generation); (ii) 一对一的匈牙利匹配，将复杂的视觉任务简化为一对一的搜索问题，使用一组可学习的查询特征表示与键值特征表示进行交互，并根据不同任务

的监督来收集有用的信息；(iii) Transformer 编码器-解码器，可以动态地对长距离的上下文信息进行建模。最近两年，许多研究工作提出了各种先进的改进设计，以提高 DETR 在不同计算机视觉任务中的性能，但是这些工作都依赖设计更加复杂的网络结构。在本次报告中，我们将详细介绍我们最近完成的更加简单且高效的三项研究工作（包括 H-DETR、Plain DETR 和 Vertex DETR），并讨论这些方法设计背后的机理，以及如何通过使用我们提出的方法来搭建基于 DETR 的 SOTA 2D 与 3D 物体检测系统。



**讲者简介：**元玉慧，微软亚洲研究院视觉计算组研究员，2014 年从南京大学计算机系获得学士学位，2017 年从北京大学智能科学技术学院获得硕士学位后，于同年加入微软亚洲研究院视觉计算组工作至今，期间 2022 初从中科院计算所获得博士学位，目前一直从事研究通用视觉理解，比如语义分割、物体检测等。代表性研究工作包括 OCRNet (ECCV 2020)与 H-DETR (CVPR 2023)，其中 OCRNet 被收录到 most influential papers in ECCV 2020 论文列表中 (位列第 8)。

## Workshop 11: 大模型背景下的自动驾驶与车路协同

主席：马超（上海交通大学） 李镇（香港中文大学(深圳)） 张瑞茂（香港中文大学(深圳)） 胡庆拥（牛津大学）

时间：6月12日（周一）08:30-12:15 地点：兰花厅 A

时间	主持人	内容
8:30-8:55	马超	讲者：李弘扬（上海人工智能实验室） 题目：道路结构认知 2.0-感知与高精地图之争
08:55-09:20	李镇	讲者：李鸿升（香港中文大学） 题目：Towards More Accurate and Robust 3D Scene Perception with LiDAR Point Clouds
09:20-09:45		讲者：王乃岩（图森未来） 题目：基于稀疏特征图的超远距离感知技术
09:45-10:10	张瑞茂	讲者：赵行（清华大学） 题目：视觉为中心的自动驾驶：3D 占用感知，运动预测与地图构建
10:10-10:25	金牌赞助商宣讲及休息 地平线企业宣讲	
10:25-10:50	张瑞茂	讲者：赵恒爽（香港大学） 题目：3D Scene Understanding: from Fundamental Design to Representation Learning and Future Frontiers
10:50-11:15	胡庆拥	讲者：张骞（地平线） 题目：BEV 新范式下的高阶自动驾驶技术
11:15-11:45		Poster Spotlight (4-6 个, 每个 5 分钟) 1. BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers 2. VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking 3. Benchmarking Robustness of 3D Object Detection to Common Corruptions in Autonomous Driving 4. Masked Images Are Counterfactual Samples for Robust Fine-tuning 5. MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction 6. CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving
11:45-12:15	马超	Panel 嘉宾：李弘扬（上海人工智能实验室）、李鸿升（香港中文大学）、王乃岩（图森未来）、赵行（清华大学）、赵恒爽（香港大学）

### 李弘扬 上海人工智能实验室

**报告题目：**道路结构认知 2.0-感知与高精地图之争

**报告摘要：**在自动驾驶方案中，感知和高精地图是两个对规划控制任务至关重要的模块。从车道线识别到道路结构认知、场景理解，是迈向高阶自动驾驶问题的必经之路。本次演讲将回顾去年 Valse 2022 留下的挑战问题，分享 OpenDriveLab 团队近一年在这方面的最新工作。更多信息详见：<https://opendrivelab.com/>



**讲者简介：**李弘扬，上海人工智能实验室青年科学家、高级工程师。2019 年香港中文大学博士毕业，在读期间荣获香港政府奖学金。以第一作者身份，在相关国际会议与期刊如 CVPR、NeurIPS、T-PAMI 等发表文章 10 余篇，累计谷歌学术引用 1600 余次，提交专利 20 余项。担任 CVPR、NeurIPS 2023 等会议领域主席。主持国家自然科学基金青年项目、上海市启明星项目。连续三年担任清华大学《计算机视觉》课程主讲人、上海交通大学行业博士生导师。2022 年获 Waymo 自动驾驶国际知名挑战赛第一名；团队提出的工作（BEVFormer、UniAD 等）在国际上取得领先地位，为多家自动驾驶公司提供了实际量产落地方案。提出的环视 3D 检测工作获 2022 年全球最有影响力的 AI 论文 Top 100。UniAD 工作获 CVPR 2023 最佳论文提名奖。

### 李鸿升 香港中文大学

**报告题目：**Towards More Accurate and Robust 3D Scene Perception with LiDAR Point Clouds

**报告摘要：**3D object detection and tracking is a fundamental module in autonomous driving vehicles. Although camera-based multi-view 3D detection methods (especially BEV-based methods) have show impressive performance, the LiDAR piont cloud-based methods still exhibit much better performance gap on achieving more accurate 3D scene perception. In this talk, I will introduce our group's latest methods on 3D object detection (ConQueR), 3D multi-object tracking (TrajectoryFormer), offline 3D detection (DetZero), etc., to improve 3D scene percpeition accuracy and robustness.



**讲者简介：**李鸿升博士现任香港中文大学多媒体实验室副教授、上海交通大学兼职博士生导师、西安电子科技大学“华山学者”讲座教授、上海人工智能实验室顾问科学家。他于 2006 年获华东理工大学自动化学士学位，2012 年于美国理海大学获得计算机科学博士学位。他的计算机视觉、深度学习、医学图像处理有着深厚的研究经验，他在计算机视觉、机器学习、医学图像处理顶级期刊和会议上（TPAMI、CVPR、ICCV、ECCV、NeurIPS、ICLR、MICCAI、IPMI 等）发表论文 110 余篇，获得了 2020 年 IEEE 电路与系统协会杰出青年作者奖、2022 年 AI 2000 人工智能最具影响力学者提名奖、2022 年全球前 2% 顶尖科学家、2021 年香港中文大学青年学者杰出研究成就奖等奖项。2016 年带领团队参加 ImageNet 2016 国际挑战赛，赢得了视频物体检测项目第一名。他担任国际顶级学术会议 NeurIPS 2021-2023、CVPR 2023、ICCV 2023、ICML 2023 领域主席，AAAI 2022 高级程序委员，国际期刊 IEEE Transactions on Circuits and Systems for Video Technology、Neurocomputing 等的副编辑。

### 王乃岩 图森未来

**报告题目：**基于稀疏特征图的超远距离感知技术

**报告摘要：**安全可靠的干线物流自动化运输依赖于可靠的超远距离感知。然而传统感知方法的运算代价随着感知距离的增大而快速增长，使得在实车部署这些算法不可行。在这个 talk 中，我会介绍图森未来在这方面的几个工作，包括纯视觉，纯 LiDAR 与多模态融合的方案。这些方法的核心观察在于随着感知距离的增大，需要计算的前景信息也快速减少，故我们可以设计多阶段的检测器，低成本地过滤掉背景信息，从而大幅节省计算开销。除了在公开数据集中取得了领先的性能，这套算法也被部署在图森未来的无人驾驶卡车中，实现了最远千米，可靠五百米的感知系统。



**讲者简介：**王乃岩，现为北京图森未来科技有限公司首席科学家，负责北京算法研发团队从事自动驾驶重卡的研发。在这之前，他于 2011 年本科毕业于浙江大学，2015 年博士毕业于香港科技大学计算机科学与工程系。他是 2014 Google PhD Fellow 计划入选者（全球 38 名），多次在国际数据挖掘和计算机视觉比赛中名列前茅。他发表的论文据 Google Scholar 统计总引用次数超过 15000 余次。

**赵行 清华大学**

**报告题目：**视觉为中心的自动驾驶：3D 占用感知，运动预测与地图构建

**报告摘要：**随着高级辅助驾驶的应用落地，视觉为主的自动驾驶感知技术受到了学术界和工业界的广泛关注。本次报告将讲述团队在视觉自动驾驶方向的三方面工作。第一个工作提出了基于视觉的 3D 占用感知任务，并开源了评测基准；第二个工作展示了视觉中丰富的语义信息能大幅提升轨迹预测的表现；第三个工作提出了一种全新的基于神经网络的地图构建范式，用于替代传统的高精度地图以及在线地图感知方法。



**讲者简介：**赵行，清华大学交叉信息研究院助理教授，博士生导师。博士毕业于麻省理工学院 MIT，后于谷歌无人车项目 Waymo 担任研究科学家。研究涵盖自动驾驶的整个算法栈，以及多模态和多传感器的机器学习。提出了自动驾驶感知和预测中一系列框架型的工作，为行业大多数公司所使用或借鉴。研究工作曾被 BBC, NBC, 麻省理工科技评论等报道。曾获 2015 年 ICCP 最佳论文奖，入选 2020 年福布斯中国 U30 科学精英榜。

**赵恒爽 香港大学**

**报告题目：**3D Scene Understanding: from Fundamental Design to Representation Learning and Future Frontiers

**报告摘要：**3D scene understanding has attracted broad interest in various potential applications such as autonomous driving, robot manipulation, augmented reality, etc. Different from 2D images, 3D point clouds are unordered and scattered in 3D space, resulting in difficulties in scene understanding under complex 3D scenarios. In this talk, I will talk about our recent explorations in the fundamental design of 3D raw operators and backbone structures, as well as representation learning algorithms. First, I will present Point Transformer, a self-attention-based architecture for

general point cloud processing. We design self-attention layers and use these to construct self-attention networks, which show superior performance on various 3D tasks. Second, I will introduce its enhanced version named Point Transformer V2, which develops powerful grouped vector attention and partition-based pooling mechanisms, and harvests more advanced performance than its predecessor. Third, I will demonstrate Masked Scene Contrast, a scalable framework for unsupervised 3D representation learning, which adopts both contrastive learning and reconstructive learning for efficient, effective, and scalable feature learning. Last, I will discuss several existing challenges and future frontiers for intelligent visual scene understanding.



**讲者简介:** Dr. Hengshuang Zhao is an Assistant Professor in the Department of Computer Science at The University of Hong Kong. Before that, he was a postdoctoral researcher at Massachusetts Institute of Technology and University of Oxford. He obtained his Ph.D. degree from The Chinese University of Hong Kong. His general research interests cover the broad area of computer vision, machine learning, and artificial intelligence, with a special emphasis on building intelligent visual systems. He and his team won several championships in competitive international challenges like the ImageNet Scene Parsing Challenge. He received the rising star award at the world artificial intelligence conference and was recognized as one of the most influential scholars in computer vision by AI 2000. He has served as an Area Chair for CVPR and NeurIPS. His works have been cited about 18,000+ times, with 10,000+ GitHub credits.

## 张骞 地平线

**报告题目:** BEV 新范式下的高阶自动驾驶技术

**报告摘要:** 2021 Tesla AI DAY, 特斯拉展示了基于 BEV 的自动驾驶感知新范式, 取得了惊艳的效果, 引起了工业界和学术界的广泛关注, 本次演讲主要介绍以 BEV 感知新范式的自动驾驶技术一些实践和探索, 主要包括三方面的工作, 端侧的以 BEV 为框架的端到端多任务感知技术、云侧的基于重建的自动标注技术以及感知到决策规划的端到端探索工作。



**讲者简介:** 张骞, 博士毕业于中国科学院自动化研究所模式识别与智能系统专业, 长期专注计算机视觉、机器学习、模式识别等技术方向的研究, 发表论文 40 余篇。现为地平线算法研发总监, 负责基础算法研发团队, 从事高阶自动驾驶技术在人工智能芯片上的探索创新与技术落地。获得了第十届吴文俊人工智能专项奖

芯片项目一等奖、2022 年中国汽车工程学会“科学技术奖科技进步一等奖”。

## Workshop 12: 多模态大模型与提示学习

主席：张利军（南京大学） 卢志武（中国人民大学） 牛玉磊（哥伦比亚大学）、丁明宇  
（加州大学伯克利分校）

时间：6月12日（周一）08:30-12:15 地点：B6 馆

时间	主持人	内容
8:30-8:40	胡迪	研讨会介绍
8:40-9:10		讲者：左旺孟（哈尔滨工业大学） 题目：预训练模型和语言增强的零样本视觉学习
09:10-09:40		讲者：余宙（杭州电子科技大学） 题目：知识增强的多模态预训练和提示学习
09:40-10:10		讲者：王云鹤（华为） 题目：多模态交织：高效模型架构和融合方法
10:10-10:40		讲者：邵婧（商汤科技） 题目：基于 LLM 的多模态提示学习：框架、提示数据及评测标准
10:40-10:52		中场休息
10:52~11:00		赞助商宣讲：真格基金
11:00-11:30		讲者：朱霖潮（浙江大学） 题目：模态鸿沟与交互式提示学习
11:30-12:15		Panel 嘉宾：左旺孟(哈尔滨工业大学)、余宙（杭州电子科技大学）、王云鹤（华为）、邵婧（商汤科技）、朱霖潮（浙江大学）

### 左旺孟 哈尔滨工业大学

**报告题目：**预训练模型和语言增强的零样本视觉学习

**报告摘要：**近年来，随着 CLIP、Stable Diffusion 等多模态预训练模型的出现，如何在各种下游任务中充分利用预训练大模型进行微调 and 提示学习，已成为近年来计算机和多模态学习的研究热点问题和重要发展趋势。针对上述挑战性问题，本报告主要包括三个方面：(1) 以 3D 点云分类为例，探讨如何将图像-语言预训练模型拓展至其他视觉模态如 3D 点云；(2) 以物体检测为例，探讨如何基于多模态预训练生成模型实现更复杂视觉任务的零样本学习。(3) 以多标签分类为例，探讨如何将语言数据作为视觉监督信息，进一步增强零样本视觉学习性能；通过上述分析介绍，期望更多种类的预训练模型（如：CLIP、Stable Diffusion）可以更为广泛地应用于各种视觉模态（如：图像、点云）和复杂视觉任务（如：分类、检测、分割、生成），推动多模态预训练模型在下游任务中的研究与实际应用。



**讲者简介：**左旺孟，哈尔滨工业大学计算机学院教授、博士生导师。主要从事图像增强与复原、图像编辑与生成、物体检测与目标跟踪、图像与视频分类等方面的研究。在 CVPR/ICCV/ECCV 等顶级会议和 T-PAMI、IJCV 及 IEEE Trans. 等期刊上发表论文 100 余篇。曾任 ICCV2019、CVPR2020/2021 等 CCF-A 类会议领域主席，现任 IEEE T-PAMI 和 T-IP 等期刊编委。

### 余宙 杭州电子科技大学

**报告题目：**知识增强的多模态预训练和提示学习

**报告摘要：**从海量数据中预训练得到的多模态模型在各种多模态任务上均取得令人惊叹的效果。如何在现有的“数据驱动”框架基础上进一步引入“知识指导”，从而进一步提升模型的表达能力和泛化能力是当前研究的热点和难点。在本报告中，我们介绍了两种融合知识的多模态学习框架：1) ROSITA：利用多模态数据内部隐含的模态内和模态间关联知识，实现内部知识增强多模态预训练学习；2) Prophet：利用多模态小模型提示语言大模型（GPT-3），实现外部知识增强的多模态提示学习。



**讲者简介：**余宙，杭州电子科技大学计算机学院教授，中国科协青托，浙江省计算机学会副秘书长。其长期致力于多模态分析与推理、多模态预训练、知识引导视觉问答等方面研究，主持国家自然科学基金面上、青年项目、国家重点研发计划子课题、浙江省自然科学基金杰青项目等多个国家级、省部级项目，在 CVPR、ICCV、ACMMM、TNNLS、TIP 等高水平期刊/会议论文 30 余篇。相关成果被国内外同行总引用 2700 余次，带领团队多次获得视觉问答领域挑战赛 VQA Challenge 的冠军。因在相关领域的突出贡献，获得 2020 年度浙江省自然科学一等奖（排名第二）、ACM 杭州新星奖等荣誉。

### 王云鹤 华为

**报告题目：**多模态交织：高效模型架构和融合方法

**报告摘要：**在多模态学习中，多种数据的融合以及 AI 模型架构本身对最终的效果提升都

是非常重要的。在这个报告中，首先会提出一种模态交织学习框架，动态地利用不同模态之间的优劣进行统筹优化。此外还会提出一种新型的骨干网络框 VanillaNet，区别于传统的深度卷积网络和视觉 Transformer，仅用 7 个卷积层，就可以逼近 ResNet-50 的性能，13 层就可以在 ImageNet 数据集上达到 83% 以上的 top1 精度，对视觉任务和多模态任务都有非常深远的意义。



**讲者简介：**王云鹤，华为算法应用部部长，在华为诺亚方舟实验室从事人工智能算法的研发和在实际业务中的应用落地。主要的研究领域包含计算机视觉、机器学习、模型压缩、高效 AI 计算等，多项算法落地工业界产品，主导的加法神经网络项目受到了业界的广泛关注。在相关领域发表 CCF A 类学术论文 80 余篇，包含 NeurIPS、ICML、CVPR、ICCV、IEEE TPAMI 等。担任 NeurIPS 和 ICML 的领域主席，VALSE 的高级领域主席。

### 邵婧 商汤科技

**报告题目：**基于 LLM 的多模态提示学习：框架、提示数据及评测标准

**报告摘要：**随着语言大模型的迅猛发展，文字成为连接各种模态信息的天然媒介，通过文字与其他模态信息的交互学习，我们可以在丰富的跨模态领域应用中获得更有泛化性的感知、理解和生成能力。由于跨模态数据和任务存在巨大鸿沟，如何做好关联对齐并有效融合仍面临巨大挑战。本次报告基于近期大模型进展，从支持从图像、视频及三维数据的理解、认知与生成等多种任务出发，介绍跨模态提示学习的数据和框架构建、评测标准设计等方面。并结合应用案例，深入探讨多模态大模型现状和挑战。



**讲者简介：**邵婧，商汤科技研究总监，2016 年于香港中文大学 MMLab 获得博士学位。长期从事计算机视觉和多媒体分析，重点研究城市级的大规模视觉感知、理解及通用视觉技术框架和开源生态。在 T-PAMI/IJCV/CVPR/NeurIPS/ACL 等发表论文 50 余篇，Google Scholar 被引 5800 余次。多次担任学术竞赛、论坛主席和论文审稿人。研究成果获 10 余篇国家和国际发明专利，并在多个产业领域落地。曾获“2022 年度深圳人工智能科技进步奖”。2021 年带领团队研发的新一代通用视觉体系“书生”是国内最早研发并开源的通用人工智能大模型之一，支撑业界大幅降低下游数据采集成本、快速验证新场景、实现长尾应用。

### 朱霖潮 浙江大学

**报告题目：**模态鸿沟与交互式提示学习

**报告摘要：**随着多模态数据的增加，多模态分析成为研究的热点。在多模态分析中，迁移与对齐技术能够将不同模态的信息对齐并进行多模态的迁移，提高任务的效果和性能。本次报告将介绍多模态领域常见的模态鸿沟问题，以及降低模态鸿沟的方法，包括基于提示词的迁移、多任务学习、零样本学习等。报告还将结合实验和应用案例，深入探讨多模态分析中的应用。



**讲者简介：**朱霖潮，浙江大学百人计划研究员、博士生导师。主要研究方向为时序建模、多模态分析及其应用、人工智能交叉领域研究。曾获得 THUMOS 动作识别竞赛冠军 (2015)、EPIC-KITCHENS 第一视角动作识别竞赛冠军 (2019, 2020)、CVPR MABe 多智能体行为建模竞赛冠军 (2022)等竞赛冠军。

## Workshop 13: 三维视觉技术前沿

主席：高盛华（上海科技大学）戴玉超（西北工业大学）王龙光（空军航空大学）、崔兆鹏（浙江大学）

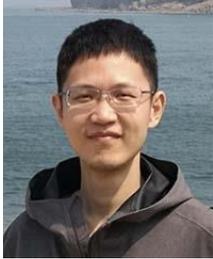
时间：6月12日（周一）08:30-12:15 地点：B4

时间	主持人	内容
8:30-9:00	戴玉超	讲者：任博（南开大学） 题目：含透明与反射物体的三维场景内容生成
9:00-09:30		讲者：杨佳琪（西北工业大学） 题目：点亮传统方法：几何三维点云配准方法新探索
09:30-10:00	崔兆鹏	讲者：申抒含（中国科学院自动化研究所） 题目：大规模复杂场景三维重建与理解
10:00-10:30		讲者：王鹤（北京大学） 题目：基于三维视觉的泛化物体操纵
10:30-10:38	崔兆鹏	赞助商宣讲（奥比中光）
10:38-10:55		中场休息
10:55-11:25	王龙光	讲者：刘玉身（清华大学） 题目：三维智能处理与重建
11:25-11:45		Poster Spotlight（4个，每个5分钟） 讲者：尹富坤（已确认）、陈思瑾（已确认）、丁润语（已确认）、张矩泽
11:45-12:15	高盛华	Panel 嘉宾：任博（南开大学）、杨佳琪（西北工业大学）、申抒含（中国科学院自动化研究所）、王鹤（北京大学）、刘玉身（清华大学）

## 任博 南开大学

**报告题目：**含透明与反射物体的三维场景内容生成

**报告摘要：**神经辐射场（Neural radiance field, NeRF）是近年间的重要三维场景内容生成方法，然而对于包含透明或反射材质的场景，NeRF 类算法的表现尚不尽如人意。在本次报告中，我将分享我们在增强 NeRF 在含透明与反射材质场景表现上的一些进展，涉及玻璃透射、平面焦散、镜面反射等多种常见现象的处理。相比已有方法，我们的方法在此类问题中能够获得更加显著更加准确的三维场景渲染结果或物体几何建模。



**讲者简介：**任博，于 2015 年于清华大学计算机科学与技术系获得工学博士学位。2015 年 7 月至今于南开大学计算机系任教。近期研究方向包括基于物理/基于机器学习的流体仿真与控制，三维场景重建与内容生成等。指导学生在国际顶级期刊会议 SIGGRAPH, TVCG, CVPR, ICCV 等发表文章十余篇。入选南开大学“百青”计划。主持或参与多项国家自然科学基金青年/面上项目，国家重点研发计划课题与国家重点实验室开放课题。

## 杨佳琪 西北工业大学

**报告题目：**点亮传统方法：几何三维点云配准方法新探索

**报告摘要：**三维点云配准是三维视觉领域的一个基础问题，近年来，随着深度学习热潮的到来，基于深度学习的三维点云配准方法涌现并取得令人瞩目的性能。相比之下，传统方法在该问题中逐渐被忽视。本报告将介绍团队针对三维配准重建传统方法的新探索。第一，我们提出了一种基于互投票的三维点云特征匹配方法，在“候选人-投票者”互投票置信度评分的框架下，为三维点云特征匹配进行离群点剔除，提高匹配集内点率。第二，我们提出了一种基于极大团的三维点云配准参数估计方法，为无序匹配构建图并提出一种全新的极大团图空间几何约束，在简单有效的迭代校验框架下取得了同期最优的配准性能，并能增强现有主流的深度学习方法。在训练样本缺乏、计算资源受限、适配全国产化平台的约束下，提出的传统方法再次取得了优异的三维配准效果。



**讲者简介：**杨佳琪，西北工业大学长聘副教授，空天地海一体化大数据应用技术国家工程实验室多域多维信息系统方向负责人，中国图象图形学学会三维视觉专委会（CSIG-3DV）秘书，陕西省电子信息类专业共同体电信类工作委员会秘书长。于 2014 年和 2019 年于华中科技大学获得学士和博士学位，2017 年至 2018 年在宾夕法尼亚大学进行公派联合培养，主要研究方向为计算资源、数据标签受限条件下的三维配准重建，致力于提出航天应用背景下的无需训练、适配全国产化平台、轻量化的三维重建算法。近年在 IEEE TPAMI、CVPR 等 CCF A/IEEE 汇刊/中科院一区的期刊会议发表论文 30 余篇，包括 CVPR 2023 best paper finalist(12/9155, 大陆高校第一单位唯一入选)。主持国家自然科学基金、国防纵向重点项目等国家级项目 3 项；陕西省秦岭生态保护课题等省部级项目 3 项；参与国防纵向重点项目等国家级项目 10 余项，成果应用于航天领域空间环境三维重建感知。

## 申抒含 中国科学院自动化研究所

**报告题目：**大规模复杂场景三维重建与理解

**报告摘要：**对大规模复杂场景进行高精高效的三维重建与理解，包括三维几何重建、三维

语义理解、三维结构表达等，是数字城市、无人系统、工业孪生等领域的共性需求。在这一研究中，计算机视觉、摄影测量、机器人等多个领域的研究不断交叉融合，几何视觉方法与深度学习方法也在不断发生碰撞。本次报告将介绍我们在大规模复杂场景三维重建与理解方面的最新研究进展，以及在城市实景三维建模、无人驾驶构图定位、工业三维数字孪生等多个领域的实际应用，并探讨大模型时代三维几何视觉研究的生存空间与核心价值。



**讲者简介：**申抒合，中国科学院自动化研究所研究员、博士生导师，中国科学院大学人工智能学院岗位教授，中科院工业视觉智能装备技术工程实验室副主任。研究领域为三维计算机视觉理论与应用，包括大规模场景三维重建、智能机器人三维环境感知、场景三维语义理解等。在计算机视觉、摄影测量、机器人等领域国际期刊和国际会议，如IJCV、IEEE TIP、ISPRS JPRS、CVPR、ECCV、ICRA、IROS等发表论文70余篇。作为项目负责人主持国家自然科学基金联合重点项目、北京市自然科学基金联合重点项目、中科院人才专项，以及各类企业研发项目20余项。入选中科院青年创新促进会、中科院自动化所特聘青年骨干、中科院特聘研究岗位，曾获2016年ACM北京新星奖、2018年中国图像图形学会科学技术二等奖等。

### 王鹤 北京大学

**报告题目：**基于三维视觉的泛化物体操纵

**报告摘要：**基于视觉的机器人物体操纵是机器人与物理世界交互的关键。目前物体操纵技术面临在泛化性和普适性上的制约。基于三维点云输入的方法减少了对物体外表颜色和纹理的依赖，拥有更好的泛化性，但是却受制于深度传感器，在透明和高光物体上无法工作，丧失了普适性。针对此问题，我们ICRA2023的工作GraspNeRF通过探究可泛化神经辐射场(NeRF)突破点云输入的限制，基于多视角RGB进行材质无关的抓取。在抓取之上，我们CVPR2023 highlight工作GAPartNet基于三维视觉输入探讨通过定义跨类别零件Generalizable and Actionable Parts对跨类别物体操纵提供了思路。基于这样的GAPart，我们CVPR2023的工作PartManip进一步基于三维视觉输入探索了跨类别物体操纵泛化的强化学习策略，展示了零件对奖励函数、状态策略学习和三维视觉策略学习的帮助。



**讲者简介：**王鹤博士是北京大学计算机学院前沿计算研究中心的助理教授和博士生导师。他创立并领导了北大具身感知与交互实验室(Embodied Perception and Interaction Lab, 主页:<https://hughw19.github.io>)。实验室立足三维视觉感知与机器人学，重点关注具身机器人在三维复杂环境中的感知和交互问题，研究目标是以可扩展地方式发展高泛化性的机器人视觉和控制系统。研究成果覆盖物体抓取、功能性操控、灵巧操作及寻物导航，于CVPR/ICCV/ECCV/TRO/RAL/ICRA/IROS/ICLR/NeurIPS/AAAI发表40余篇工作，获得世界人工智能大会优秀青年论文奖，Eurographics 2019最佳论文提名，ICLR 2021可泛化机器人物体操纵挑战赛ManiSkill无额外标注赛道冠军。他担任了CVPR2022和WACV2022的领域主席，Image and Vision Computing的副主编和诸多项会的审稿人、程序委员。在加入北京大学之前，他于2021年从斯坦福大学获得博士学位，师从美国三院院士Leonidas J Guibas教授，于2014年从清华大学获得学士学位。

**刘玉身 清华大学**

**报告题目：**三维智能处理与重建

**报告摘要：**三维智能处理与重建是计算机图形学、三维计算机视觉、数字孪生领域中的研究热点，本报告主要围绕激光扫描等获取的三维数据，介绍近两年三维智能识别、处理与重建的前沿技术，包括：三维智能识别技术，及其在对象识别、大规模模型检索中的应用；三维智能处理技术，及其在三维场景语义分割与补全中的应用，为后续场景重建与理解提供支撑；三维智能重建技术，及其在数字孪生领域中的应用。



**讲者简介：**刘玉身，清华大学软件学院院长聘副教授，博士生导师。研究方向是三维计算机视觉、几何智能处理与重建、BIM。主持国家自然科学基金项目 5 项，主持国家重点研发计划课题 2 项，参编国家标准 3 项。以第一或通讯作者，近年来在 TPAMI、TIP、CVPR、ICCV、NeurIPS、ICML 等 CCF A 类期刊和会议发表论文 40

多篇，谷歌学术总引用 3000 多次。担任多个 CCF A 类国际会议的高级程序委员会委员和国际重要期刊编委、中国图学会计算机图学会专委会委员、中国铁道学会智能铁路委员会委员等。研究成果获得中国图学会 2022 年度优秀博士学位论文、中国计算机学会 CAD&CG 专委会 2021 年度“优秀图形开源数据集”奖、工程信息化领域旗舰期刊 AEI 高被引研究奖、土木与建筑工程计算领域旗舰会议 (ICCCBE2016) 最佳学生报告奖、CAD/Graphics 2005 最佳学生论文奖等。教学成果获得清华大学精品课程、清华大学年度教学优秀奖、清华大学优秀班主任一等奖等，指导研究生多次荣获北京市优秀毕业生、清华大学优秀毕业生、校级优秀学位论文等。详见个人主页：<https://yushen-liu.github.io/>

## Workshop 14: 视觉内容生成

主席：李崇轩（中国人民大学）章国锋（浙江大学）于茜（北京航空航天大学）盛律（北京航空航天大学）

时间：6月11日（周日）13:30-17:15 地点：B5 馆

时间	主持人	内容
13:30-14:05	于茜	讲者：卢志武（中国人民大学） 题目：多模态通用生成模型的基本框架与最新进展
14:05-14:40		讲者：王杨（合肥工业大学） 题目：基于知识蒸馏场景下的样本生成
14:40-15:15		讲者：齐勇刚（北京邮电大学） 题目：跨模态生成视角下的抽象视觉理解
15:15~15:23		赞助商“阿里妈妈”宣讲
15:23-15:40		中场休息
15:30-16:15	盛律	讲者：古纾旸（微软亚洲研究院） 题目：From Paint by Example to Generalized Image Editing
16:15-16:50		讲者：易冉（上海交通大学） 题目：面向艺术肖像画的媒体艺术生成与评估
16:50-17:10		Poster Spotlight 讲者列表：张亚博，陶明，邓皓戈，郦洋
17:10-18:00	于茜	Panel 讨论 嘉宾列表：卢志武（中国人民大学）、王杨（合肥工业大学）、齐勇刚（北京邮电大学）、古纾旸（微软亚洲研究院）、易冉（上海交通大学）、盛律（北京航空航天大学）

**卢志武 中国人民大学**

**报告题目:** 多模态通用生成模型的基本框架与最新进展

**报告摘要:** 多模态预训练在过去三年得到飞速的发展,涌现出 OpenAI CLIP、Stable Diffusion 等优秀模型。受 ChatGPT 的启发,多模态预训练出现新的变化,开始更关注多模态通用生成。本报告首先给出多模态通用生成模型的定义与基本框架。然后从语言生成、图像生成和视频生成三个方面展开,介绍我们的最新研究进展。最后还会对多模态通用生成模型的发展趋势进行展望。



**讲者简介:** 卢志武博士,中国人民大学高瓴人工智能学院教授,博士生导师。2005年毕业于北京大学数学科学学院,获理学硕士学位;2011年毕业于香港城市大学计算机系,获 PhD 学位。研究方向为机器学习、计算机视觉等。设计首个公开的中文通用图文预训练模型文澜 BriVL,并发表于 Nature Communications。

**王杨 合肥工业大学**

**报告题目:** 基于知识蒸馏场景下的样本生成

**报告摘要:** 近年来,基于数据安全与隐私问题,无数据场景下的知识蒸馏方法获得了越来越多的关注。然而,现有方法侧重于恢复原始数据,忽略了生成样本对于学生模型的适应性,在涉及老师-学生网络模型存在较大差异的情况下,仍然面临着诸多挑战。在本次报告中,我们首先通过以下几个问题介绍关于知识蒸馏的差异化研究:老师学生模型差异性为什么以及什么时候会影响学生的性能?如何衡量老师学生间的差异性?然后,我们立足无数据量化任务,通过泛化误差分析,深入研究孤立学生模型所导致的欠拟合、过拟合问题。围绕生成样本的适应性,首次提出零和博弈观点进而分析无数据量化,解决了模型差异性,同时为生成样本构建上下边界,通过边缘优化,在无数据蒸馏过程中实现理想样本的生成。



**讲者简介:** 合肥工业大学多媒体实验室 教授,博士生导师。入选安徽省高层次人才计划,担任信息搜索领域顶级杂志 ACM Transactions on Information Systems (CCF Rank A) 副编 (Associate Editor, 2019 至今)。入选斯坦福大学 2022 年 9 月统计发布的人工智能与图像处理领域 前 2% 全球顶级科学家。至今在模式识别,多媒体计算相关领域的顶级杂志与会议上发表文章 80 篇,其中 ESI 高被引文章 7 篇,并且全部进入 top 1% 列表,发表源包括 Artificial Intelligence (Elsevier), International Journal of Computer Vision (IJCV), IEEE TIP, ACM TOIS, Machine Learning (Springer), IEEE TKDE, VLDB Journal, CVPR, ECCV, ACM SIGIR, ACM KDD, AAAI, IJCAI, ACM Multimedia, SCIENCE CHINA Information Sciences (中国科学:信息科学), 其中两篇论文入选 paperdigest2021/03 版本的 IJCAI 最有影响力文章之一,主持国家自然科学基金联合基金重点项目,面上项目等,同时担任国家自然科学基金 优秀青年基金(海外)项目,面上项目评审专家。谷歌学术引用 5000+, H-因子 34。

### 齐勇刚 北京邮电大学

**报告题目：**跨模态生成视角下的抽象视觉理解

**报告摘要：**对视觉信号的感知与理解是计算机视觉与机器学习的一项核心任务。相比于自然图像，抽象图像（如手绘草图、推理图）常常是由人创作的视觉信号，蕴含的信息更加抽象，机器视觉算法理解此类图像往往更加困难。齐勇刚博士将围绕抽象视觉理解任务，分享他在基于手绘的跨模态生成与检索、基于序列模型的生成式手绘分割、残缺手绘的生成式理解、基于扩散模型的手绘生成、生成式感知视觉问答等问题上的若干最新研究成果。



**讲者简介：**齐勇刚，博士，北京邮电大学人工智能学院助理教授，硕士生导师。2015年7月于北京邮电大学获得工学博士学位。2019年以国家基金委公派访问学者身份赴英国萨里大学 Centre for Vision, Speech and Signal Processing, CVSSP 交流访问。2013年4月至11月赴丹麦奥尔堡大学电子系以客座博士身份交流学习，2013年12月至2014年4月赴广州中山大学信科院访问交流。主要研究方向为草图识别与检索，感知编组(Perceptual Grouping)，少样本学习，多模态检索，生成式模型等。已发表包括国际顶级学术会议 CVPR、ECCV、BMVC 及期刊 TCSVT 等在内的 20 余篇学术论文。作为执行主席组织了 CCF-CV 视界无限研讨会第九期“手绘草图研究的前沿进展与未来趋势”，并曾在 ICCV2021 和 ECCV2022 作为主要组织者组织系列研讨会 Sketching for Human Expressivity。

### 古纾旸 微软亚洲研究院

**报告题目：** From Paint by Example to Generalized Image Editing

**报告摘要：** Recently, language-guided image editing techniques have become increasingly successful. However, these techniques often lack local control. In our study, for the first time, we investigate exemplar-guided image editing as a way to achieve more precise control. We achieve this by disentangling and re-organizing the source image and the exemplar using self-supervised training. To avoid the copy-paste trivial solution, we propose an information bottleneck and strong augmentations. Additionally, to ensure controllability, we design an arbitrary shape mask for the exemplar image. Our framework involves a single forward of the diffusion model without iterative optimization. This new technology allows for controllable editing of in-the-wild images with high fidelity. By combining it with language-guided editing methods, we propose a generalized image editing framework that can handle various image editing operations based on language instructions.



**讲者简介：**古纾旸，博士，在中国科学技术大学自动化系于 2017 年和 2022 年分别获得学士和博士学位，现为微软亚洲研究院研究员，主要研究方向为计算机视觉中的生成模型，特别是生成对抗网络和扩散模型的理论及其在 2D 和 3D 数据中的应用，以及对生成结果的质量评估等。目前已在 CVPR, ICCV, ECCV 等会议上发表多篇论文并担任多个会议与期刊的审稿人。

### 易冉 上海交通大学

**报告题目：**面向艺术肖像画的媒体艺术生成与评估

**报告摘要：**人工智能内容生成已经引发内容创作行业的产业革命，艺术肖像画是人工智能内容生成领域的重要分支。本次报告主要分享面向艺术肖像画的媒体艺术风格生成的系列研究工作。首先，针对成对数据、单风格生成，我们提出针对线条的距离变换损失和层次

化结构的生成模型，与基于流形映射的理论解释和精细艺术肖像生成模型；针对非成对数据、多风格生成，我们提出基于非对称循环结构的生成模型以解决源域、目标域信息不对称导致的信息隐藏问题，提出肖像线条画质量评估指标和基于风格空间搜索的未知新风格肖像画生成方法。其次，针对多模态信息融合问题，我们提出基于特征空间弹性匹配的语音驱动艺术肖像说话视频生成方法，基于语言模型指导的小样本扩散模型风格迁移方法，以及基于扩散模型的 3D 内容成方法。最后，针对内容生成质量难以评估的特点，我们提出了图像艺术美感评估方法，以助力后续的媒体艺术生成研究。



**讲者简介：**易冉，上海交通大学计算机科学与工程系助理教授。2016 年获得清华大学工学学士学位，2021 年获得清华大学工学博士学位。从事计算机图形学、计算机视觉等方面的研究。发表录用 30 余篇论文于 IEEE TPAMI、ACM TOG、SIGGRAPH、CVPR、ICCV、TVCG、AAAI 等国际期刊和会议，其中 CCF-A 国际期刊会议 21 篇。入选第八届中国科协“青年人才托举工程”，获 2021 年北京市科技进步二等奖(排名九)，2022 年中国图象图形学学会高等教育教学成果一等奖(排名五)，2021 年中国图象图形学学会石青女科学家奖(青英组)，2022 年瑞士 Chorafas 青年研究奖，2022 年中国人工智能学会优秀博士学位论文，2019 年中国计算机学会计算机视觉专委会学术新锐奖。担任中国图象图形学学会智能图形专委会、动画与数字娱乐专委会委员，TPAMI、IJCV、TIP、CVPR、ICCV、NeurIPS、ICLR、AAAI 等国际期刊会议审稿人。

## Workshop 15: 自监督视觉表征学习

主席：沈为（上海交通大学） 王兴刚（华中科技大学） 胡瀚（微软亚洲研究院）

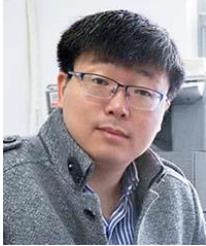
时间：6月12日（周一）13:30-18:00 地点：B6 馆

时间	主持人	内容
13:30-14:10	沈为	讲者：叶齐祥（中国科学院大学） 题目：视觉表征模型结构设计——从局部全局特征耦合的 Conformer 到全预训练模型 iTPN
14:10-14:50		讲者：纪荣嵘（厦门大学） 题目：从 ChatGPT 到国产多模态基础大模型
14:50-15:30	王兴刚	讲者：张拯（微软亚洲研究院） 题目：MIM 的理解，局限，提升与扩展
15:30-16:10		讲者：孔涛（字节跳动） 题目：自监督表示学习与机器人操作
16:10-16:25		中场休息
16:25-17:05	胡瀚	讲者：谢晋（南京理工大学） 题目：自监督点云场景流估计
17:05-17:30		Poster Spotlight 讲者列表：陈俊颖，陈恺，徐斌倩，周楠，申志强
17:30-18:00	沈为	Panel 嘉宾：叶齐祥（中国科学院大学）、纪荣嵘（厦门大学）、张拯（微软亚洲研究院）、孔涛（字节跳动）、谢晋（南京理工大学）

## 叶齐祥 中国科学院大学

**报告题目：**视觉表征模型结构设计——从局部全局特征耦合的 Conformer 到全预训练模型 iTPN

**报告摘要：**分析了局部卷积运算与全局注意力运算的互补性及辩证关系，将局部特征与全局特征耦合形成 Conformer 网络结构，小规模模型可以显著增强视觉表征能力，提升表征模型的性能下限。探讨了局部卷积运算造成的 Mask Image Modeling (MIM) 自监督学习信息泄露问题，提出了 Token Merging 操作，突破卷积或局部运算的局部约束，形成高效分层 Transformer 表征(HiViT)与全预训练的 Transformer 金字塔网络(iTPN)。在 ImageNet 分类任务上，不依赖于额外数据，iTPN-Base, iTPN-Large, iTPN-Huge 分别达到了 88.0%, 89.2%, 89.7% 的 Top-1 分类精度。系列研究从模型结构设计角度将视觉目标检测、分割等任务性能提升到了一个新高度。代码链接：[github.com/pengzhiliang/Conformer](https://github.com/pengzhiliang/Conformer) 与 [github.com/sunsmarterjie/iTPN](https://github.com/sunsmarterjie/iTPN)。



**讲者简介：**中国科学院大学特聘教授、国家杰出青年基金获得者、中科院卢嘉锡青年人才奖获得者、中国科学院优秀博士生导师、CVPR2023、NeurIPS2023 Area Chairs、国际期刊 IEEE TITS, IEEE TCSVT 编委。2013 至 2014 年在美国马里兰大学先进计算机技术研究所(UMIACS)任访问助理教授，2016 年 Duke 大学信息技术研究所(IID)访问学者。主要进行视觉表征学习与目标感知方向的研究，在 CVPR, ICCV, NeurIPS 等国际会议与 TPAMI, TNNLS, TIP 等期刊发表论文 100 余篇，Google 引用 10000 余次。承担了自然科学基金重点项目、开发的高精度目标感知方法支撑了华为、航天等单位的应用系统。曾获 ICCV2017、

CVPR2019、高分辨率对地观测重大专项遥感目标解译竞赛冠军、中国电子学会自然科学一等奖。培养多名博士生获中科院院长奖、中国科学院百篇优秀博士论文、博士后创新人才计划支持等。

## 纪荣嵘 厦门大学

**报告题目：**从 ChatGPT 到国产多模态基础大模型

**报告摘要：**随着深度学习技术的快速发展，ChatGPT 作为自然语言处理领域的重要突破，引起了广泛关注。本次探讨 ChatGPT 在自然语言处理领域中的重要性以及其所具有的优秀特性，继而引出国产多模态基础大模型的研发意义，以及如何通过结合多模态数据和深度学习技术来实现新型国产化基础模型的构建。具体来说，将讨论多模态数据的处理方法、模型构建的关键技术以及如何实现紧致化部署等相关技术点。本讲座的研究成果将为解决多模态大模型的部署难题和提高模型性能和效率提供借鉴和参考。



**讲者简介：**纪荣嵘，厦门大学南强特聘教授，科技处处长，人工智能研究院负责人。国家杰出青年科学基金获得者、国家优秀青年科学基金获得者、中组部万人计划青年拔尖人才。长期从事计算机视觉与机器学习等人工智能领域前沿技术研究，近年来发表 TPAMI, IJCV, ACM 汇刊、IEEE 汇刊、CVPR、NeurIPS 等会议长文百余篇。论文谷歌学术引用 1.9 万余次。曾获 2016 年教育部技术发明一等奖、2018 年省科技进步一等奖、2020 年省科技进步一等奖、2021 年教育部技术发明二等奖、2022 年霍英东青年科学奖。曾现主持科技部科技创新 2030 重大项目、军科委基础加强项目、国家自然科学基金联合重点、教育部装备预研创新团队项目等。任中国计算机学会 A 类国际会议 CVPR 和 ACM Multimedia 领域主席、中国图象图形学学会学术工委副主任、教育部电子信息类教指委委员，人工智能国家标准

工作组联合组长。

### 张拯 微软亚洲研究院

**报告题目：**MIM 的理解，局限，提升与扩展

**报告摘要：**Mask Image Modeling 已经成为视觉预训练的重要方法。然而，关于其机制的理解，有效性的边界，以及与其他预训练方法的关系等探索仍处于早期。本次报告中，我们会分享过去的一年中我们对 MIM 模型机制和有效性方面的探索和理解，并以此为起点，深入的探讨 MIM 与其他预训练模型（如 CLIP, DINO 等）的关系，以及 MIM 在性能边界和在小模型上应用 MIM 预训练的新进展。



**讲者简介：**微软亚洲研究院视觉计算组高级研究员，主要研究兴趣是构建通用的视觉感知系统，其研究方向涵盖了神经网络架构设计，预训练算法，以及物体检测、分割等。他是 Swin Transformer, SimMIM, Relation Network 和 Soft Teacher 等工作的主要作者之一。此外，他早期在自然场景文字检测的工作也具有开创性：Symmetry-based Text detection 首次提出将文字检测建模为分割问题，MFCN 则首次将全卷积网络（FCN）引入文字检测领域，并用于多方向文字检测。他曾于 2021 年获得 Marr Prize (ICCV 最佳论文奖)，其 Google Scholar 引用超过 14000。

### 孔涛 字节跳动

**报告题目：**自监督表示学习与机器人操作

**报告摘要：**近年来，自监督表示学习在计算机视觉、NLP 等领域取得了巨大的进步，一个自然的问题是自监督表示学习是否会对机器人领域产生大的影响呢？本次 talk，我将主要介绍两个与该话题密切相关的工作。首先我将介绍 iBOT，一种有机结合 Masked Image Model 和 self-distillation 的预训练模型，该模型不仅在表示学习的指标上获得 SoTA，并且可以有效捕捉输入图像的语义、部件和细节，表现出较大的泛化潜力。之后将介绍结合表示学习的机器人操作学习探索工作，包括数据、模型和部署，提出一种统一的机器人操作学习范式，使机器人具备多种操作任务的泛化性。



**讲者简介：**字节跳动人工智能实验室高级研究员，机器人研究负责人。在人工智能和机器人领域国际会议和期刊（CVPR/ICRA/NeurIPS/TPAMI 等）发表论文 30 余篇，获得超过 7000 次的谷歌学术引用。获得 2020 年中国人工智能学会优秀博士学位论文提名奖、2016 年 IROS 机器人抓取与操作竞赛冠军、2022 年 Habitat 国际主动语义导航挑战赛冠军等荣誉。担任国际会议 NeurIPS 2023 领域主席，以及多个顶会的审稿人、程序委员。

### 谢晋 南京理工大学

**报告题目：**自监督点云场景流估计

**报告摘要：**点云场景流估计是自动驾驶中的重要任务之一。本报告将介绍基于起点生成的端到端的自监督点云场景流估计框架。该框架利用源点云与目标点云之间的场景流一致性作为自监督信号，设计了一个起点生成网络，在点云坐标空间和特征空间迭代聚类生成场景流相似的起点。通过聚合起点的场景流信息，该框架还设计了起点引导的门控循环单元来更新预测的场景流。



**讲者简介：**南京理工大学教授，博士生导师，江苏特聘教授。博士毕业于香港理工大学计算机系，曾在纽约大学阿布扎比分校及纽约大学 Tandon 工程院任职博士后研究员、研究科学家，主要研究方向为模式识别和三维计算机视觉及在自动驾驶和机器人中的应用，已在 IEEE TPAMI、IEEE TIP、CVPR、ICCV、ECCV、NeurIPS、ICRA 等国际权威期刊和会议上发表论文多篇，曾获得亚洲模式识别大会 ACPR 2021 最佳论文奖。

## Workshop 16: 人工智能与科学计算

主席：王楠楠（西安电子科技大学）白磊（上海人工智能实验室）陈涛（复旦大学）

时间：6月12日（周一）8:30-12:15 地点：梅花厅

时间	主持人	内容
8:30-8:35	白磊	Workshop 简要介绍
08:35-09:05		讲者：杨跃东（中山大学） 题目：融合 HPC 和 AI 的药物分子设计
9:05~9:35		讲者：谢凌曦（华为） 题目：盘古气象大模型：3D 全球高分辨率气象预报方法
09:35-10:05		讲者：苏航（清华大学） 题目：非均匀物理数据分布的偏微分方程（PDEs）的通用求解算子
10:05-10:15		中场休息
10:15-10:45	陈涛	讲者：欧阳万里（上海人工智能实验室） 题目：从计算机视觉到 AI4Science-挑战与机遇
10:45-11:15		讲者：严骏驰（上海交通大学） 题目：量子图机器学习及其化学应用初探
11:15-12:15		Panel 嘉宾：谢凌曦（华为）、欧阳万里（上海人工智能实验室）、严骏驰（上海交通大学）、杨跃东（中山大学）、苏航（清华大学）

## 杨跃东 中山大学

**报告题目：**融合 HPC 和 AI 的药物分子设计

**报告摘要：**报告围绕药物智能设计中的蛋白质和药物分子的准确表征问题，在蛋白质和化合物的表示方面提出一系列新策略，并通过之间的多模态融合学习，准确预测蛋白质-药物的结合。该智能模型进一步与基于超算的分子模拟相结合，在七周内就成功发现全新 PROTAC 化合物，实验验证具有更好的药代性能。同时，介绍生物医药计算平台的最新进展，浅谈如何结合 HPC 和 AI 加速生物医药研究。



**讲者简介：**杨跃东,中山大学计算机学院教授、国家超算广州中心副总工程师,是国家级人才、科技部创新团队及广东省创新团队核心成员,主持国家重点研发课题、基金委面上项目及新冠专项、和广东省重点研发项目等。担任 Nature 子刊 Communication Biology 和 BMC Bioinformatics 杂志编委。在 Nature Machine Intelligence、Bioinformatics 等领域顶级杂志和人工智能顶会 (IJCAI、AAAI 等) 上已发表超 150 篇论文, 引用近 8000 次, 入选斯坦福大学发布的 2021 年度全球前 2% 最有影响力科学家。

获 2022 年广东省科技进步特等奖、中国电子学会科技进步一等奖、细胞出版社 2021 中国年度论文、2021 年世界人工智能大会青年优秀论文奖等。目前主要研究结合高性能计算及人工智能算法, 研发跨尺度多组学大数据挖掘及药物智能设计算法, 并结合“天河二号”开发“一站式”生物医药高性能计算服务平台。

## 谢凌曦 华为

**报告题目：**盘古气象大模型：3D 全球高分辨率气象预报方法

**报告摘要：**数值天气预报在每日天气预报、极端灾害预警、气候变化预测等领域发挥着重要作用,但是随着算力增长的趋缓和物理模型的逐渐复杂化,传统数值预报的瓶颈日益突出。研究者们开始挖掘新的气象预报范式,如使用深度学习方法预测未来天气。在数值方法应用最广泛的中长期预报任务中, AI 预报方法精度仍然显著低于数值预报方法。本次报告中,我将介绍盘古气象大模型,一种新的高分辨率全球 AI 气象预报系统。盘古气象大模型是首个超过传统数值预报精度的 AI 方法, 1 小时-7 天预测精度均高于传统数值方法 (欧洲气象中心的 operational IFS), 同时预测速度提升 10000 倍,能够在秒级时间内提供全球气象预报。古气象模型的水平空间分辨率达到 0.25 度,时间分辨率为 1 小时,覆盖 13 层垂直高度,可以精准地预测位势、湿度、风速、温度、海平面气压等气象特征。作为基础模型,盘古气象大模型能够直接应用于下游气象预报场景,例如在热带风暴轨迹预测中,盘古气象大模型的预测精度显著超过欧洲气象中心的细网格预报结果。日前,相关论文已经被 nature 正刊接收



**讲者简介：**谢凌曦博士目前是华为公司的高级研究员。他分别于 2010 年和 2015 年于清华大学获得本科和博士学位,并且于 2015 年至 2019 年期间在美国加州大学洛杉矶分校和约翰霍普金斯大学担任博士后研究员。谢凌曦博士的研究兴趣覆盖计算机视觉的各个方向,主要包括统计学习方法和深度学习模型的应用。他的研究工作覆盖图像分类、物体检测、语义分割和其他视觉任务,并积极推动自动机器学习算法在上述领域的应用。谢凌曦博士已经在国际顶级的学术会议和期刊上发表超过 90 篇论文,谷歌学术引用超过 10000 次。他于 2015 年获得清华大学优秀博士学位论文奖,并于 ICMR2015 会议上获得最佳论文奖。

### 苏航 清华大学

**报告题目：**非均匀物理数据分布的偏微分方程（PDEs）的通用求解算子

**报告摘要：**物理信息驱动的机器学习（PIML）作为一种新型机器学习范式，有效地将物理的先验知识和实验数据融合，赋能机器学习模型在面临高维度和不确定性问题时，得到更贴合物理规律的解决方案。本报告专注于非均匀物理数据分布的偏微分方程（PDEs）求解问题，探讨基于机器学习的神经算子求解方法。首先，对于几何边界条件复杂的 PDE 求解问题，我们引入了混合有限元方法的 Extra Field，构建了统一而高效的求解框架。随后，我们提出了通用神经运算符变换器（GNOT），这是一个基于变换器的学习运算符框架，能够处理多个输入函数和不规则网格。最后，我们引入了非均匀神经运算符（NUNO），这是一个针对非均匀数据设计的高效运算符学习的通用框架。同时，为了建立了全面、一致的基准测试，评估不同物理信息神经网络（PINNs）方法在解决偏微分方程中的效果，我们建立了一个 PINN 的评估测试基准 PINNacle，以期能为科学机器学习方法的发展提供有力工具。



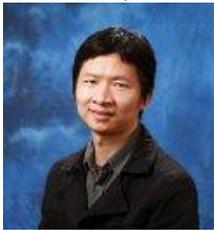
**讲者简介：**苏航，清华大学计算机系副研究员，入选国家“万人计划”青年拔尖人才，主要研究安全可信的机器学习算法，发表 CCF 推荐 A 类会议和期刊论文 50 余篇，谷歌学术论文引用 6000 余次，获得吴文俊人工智能自然科学一等奖，ICME 铂金最佳论文、MICCAI 青年学者奖和 AVSS 最佳论文等多个学术奖项，曾率队在 NeurIPS2017 对抗攻防等多个国际学术比赛中获得冠军。现任中国图像图形学会青工委执委、曾任 VALSE 执行 AC 委员会主席，担任 NeurIPS21 的领域主席（Area Chair）、AAAI22 Workshop Co-Chair，并在多次 ICML 等顶级国际会议上作为分论坛主席组

组织对抗学习专题研讨。

### 欧阳万里 上海人工智能实验室

**报告题目：**从计算机视觉到 AI4Science-挑战与机遇

**报告摘要：**以深度学习为代表的人工智能算法取得了飞速的发展，并大规模地应用到人类的生产生活实践中。Valsec 们大量学者关注计算机视觉问题，讲者的科研经历也是如此。另一方面，将人工智能技术应用到科学研究，利用人工智能算法解决当前科学的未解问题已经成为产学研关注的重点。本次报告将介绍这两个不同科研课题的共性、区别以及介绍 AI4Science 这一对于 Valsec 们比较新的课题中的挑战与机遇。同时将介绍上海人工智能实验室在 AI4Science 研究（包括材料、生物、气象、天文）的既有工作和未来探索。作为其中的一个工作，将介绍实验室最近的中期天气预报大模型“风乌”。



**讲者简介：**欧阳万里，上海人工智能实验室教授，香港中文大学博士毕业，负责人工智能驱动的交叉科学研究工作。长期从事计算机视觉与深度学习领域的研究，其科研成果在人工智能领域一流国际会议和期刊发表论文 160 余篇，论文总引用 32000 余次，H-Index 为 79。两篇文章入选 paperdigest CVPR/ICCV 最有影响力的文章。因其成果产业化价值，这些成果已转化为 112 项专利。担任 TPAMI、IJCV 和 Pattern Recognition 副主编。

**严骏驰 上海交通大学**

**报告题目：**量子图机器学习及其化学应用初探

**报告摘要：**随着量子计算与量子机器学习的发展，量子计算在各领域的优越性受到了广泛关注。由于量子计算在求解薛定谔方程方面的天然优势，量子计算或者更准确地说变分子算法被认为是探究分子动力学问题的有力工具。本次报告将围绕量子机器学习在化学计算中的应用，介绍量子计算以及量子机器学习的相关研究进展；同时对变分子算法，尤其是变分子算法在分子基态能预测等问题中的应用进行介绍；最后我们将介绍一些我们组近期利用量子机器学习方法求解量子化学问题的一些初步探索。



**讲者简介：**严骏驰，上海交通大学计算机系副教授。科技部 2030 新一代人工智能青年项目负责人、基金委优青、教育部资源建设深度学习首席专家。曾任 IBM 研究院首席研究员。研究方向为机器学习。发表 CCF-A 类第一/通讯作者论文过百篇，引用超万次。任 ICML、NeurIPS、CVPR 等领域主席、Pattern Recognition 等期刊编委。

## Workshop 17: 视觉知识和多重知识表达

主席：朱霖潮（浙江大学）丛润民（山东大学）张鼎文（西北工业大学）

时间：6月11日（周日）08:30-12:15 地点：兰花厅 B

时间	主持人	内容
08:30-08:40	朱霖潮	讲者：王树徽（中科院计算所） 题目：跨媒体多重知识获取与可信学习推理
08:40-09:10		讲者：付彦伟（复旦大学） 题目：先验信息引导的图片内容生成与编辑
09:10-09:40	丛润民	讲者：王文冠（浙江大学） 题目：知识与数据双驱动的视觉感知
09:40-10:10		讲者：谢伟迪（上海交通大学） 题目：基于知识驱动的多模态表征学习
10:10-10:20		中场休息
10:20-10:50	张鼎文	讲者：孙奕帆（百度） 题目：Exploring the Prototype and Scope of Visual Concept
10:50-11:10		Poster Spotlight（4位讲者）
11:10-12:15	朱霖潮	Panel 嘉宾：王树徽（中国科学院计算技术研究所）、付彦伟（复旦大学）、王文冠（浙江大学）、谢伟迪（上海交通大学）、孙奕帆（百度）

### 王树徽 中科院计算所

**报告题目：**跨媒体多重知识获取与可信学习推理

**报告摘要：**近年来，多源异构的跨媒体数据呈现爆炸性增长趋势，展现出了跨模态、跨数据源的复杂关联及动态演化特性，对人工智能的研究范式产生了深刻影响。跨媒体多重知识获取与可信学习推理是在对图像、视频、文本等多模态信息的概念识别和深度理解基础上，实现跨媒体内容的结构化解析和多粒度多模态知识获取，进而从学习机理、分析过程、推理结果等层面实现可解释、可回溯的跨媒体学习与推理，是发展可信跨媒体类人智能的关键技术。本报告围绕跨媒体知识结构的复杂特性与跨媒体分析推理难解释难泛化的固有问题，介绍研究组近期在跨媒体知识获取与学习推理方面的若干研究探索，包括人机协同的细粒度跨媒体事理知识构建、可信跨模态统一表征与知识学习、可解释、可回溯跨模态推理等。最后，基于已有研究结论，对知识驱动的跨媒体计算的发展前景进行展望。



**讲者简介：**王树徽，2006 年于清华大学获得工学学士学位，2012 年于中国科学院计算所获得工学博士学位，2014 年从中国科学院计算所博士后出站并留所工作、历任助理研究员、副研究员、研究员。从事图像视频理解、跨媒体分析推理与等方面的研究，发表及录用 IEEE/ACM 汇刊以及顶级学术会议论文 80 余篇。担任 ACM MM、IJCAI 等顶级国际会议领域主席，主持国基金面上项目等科研项目，参与科技创新 2030-新一代人工智能重大项目、973 课题、863 课题等项目研究，获国家自然科学基金委优秀资助。获得吴文俊人工智能自然科学一等奖、CCF 科学技术奖等奖励。

### 付彦伟 复旦大学

**报告题目：**先验信息引导的图片内容生成与编辑

**报告摘要：**多模态学习成为 AI 的热门技术趋势之一，尤其是在新内容生成方面。而得益于神经网络的发展，以图像修复、编辑及内容生产为代表的新一代图像内容智能技术，近年来取得了巨大的进步。以 Transformers、生成对抗网络、Diffusion model 等为代表的技术，成就了众多优秀的图像生成模型。先验信息指导，对图片内容生成和编辑尤为重要。相关工作在该领域取得了成功，促进了更多的模型/算法性能改进和有效的图像合成。本次报告通过回顾与讨论一系列图像合成任务及相关代表性方法，进而介绍本课题组先验信息引导的图像生成与编辑的相关工作，最后 jiang 探讨各种常见的图像合成方法（自动编码器、GAN、变换器），并对先验信息引导相关的“新内容”技术做系列展望。



**讲者简介：**付彦伟，博士，复旦大学大数据学院青年研究员，博士生导师，东方学者、国家青年千人计划学者。2014 年获得伦敦大学玛丽皇后学院博士学位，2015.01-2016.07，在美国匹兹堡迪士尼研究院任博士后研究员。付博士发表高水平论文 100 多篇：在 IEEE TPAMI 发表通讯作者/第一作者论文 11 篇，论文曾获得 IEEE ICME 2019 最佳论文，获得美国发明专利 6 项、中国专利 10 多项。研究方向侧重于基于迁移学习的多个任务，如 3D/4D 物体的建模；神经网络稀疏化学习、机械臂抓取；图像编辑及修复等。

### 王文冠 浙江大学

**报告题目：**知识与数据双驱动的视觉感知

**报告摘要：**大模型时代的到来标志着计算机视觉领域将迎来全新变革。在海量数据和超大

规模参数的加持下，神经网络展现出强大的表达和学习能力，而另一方面，人类智能建立在高度抽象的符号知识与逻辑先验之上。在计算机视觉领域新变革与大模型时代的背景之下，本次讲座围绕“知识与数据双驱动”的全新视觉感知范式，介绍如何将视觉知识、逻辑知识、专家知识为代表的多重知识与数据驱动的神经网络相结合，并以视觉识别、图像分割以及图像生成等经典视觉感知任务为例，探讨计算机视觉的未来发展方向。



**讲者简介：**王文冠，浙江大学计算机学院百人计划研究员，博士生导师，国家优秀青年基金（海外）获得者。2022~2023年，任悉尼科技大学（University of Technology Sydney）讲师。2020~2022年，任苏黎世联邦理工学院（ETH Zurich）博士后研究员。2018~2019年，先后任起源人工智能研究院（IIAI）研究员和资深研究员。2016~2018年在加州大学洛杉矶分校（UCLA）访学。2018年博士毕业于北京理工大学。主要研究方向为计算机视觉和人工智能。在顶级期刊和会议（如 TPAMI、IJCV、ICLR、NeurIPS、CVPR、ICCV、ECCV、AAAI、Siggraph Asia）发表学术论文 80 多篇。谷歌学术引用 1 万余次，H 指数 56。曾获澳大利亚研究理事会（Australian Research Council, ARC）优秀青年基金（Discovery Early Career Researcher Award, DECRA）（2022 年）、Elsevier 高被引中国学者（2020~2022 年）、世界人工智能大会优秀青年论文奖（2020 年）、中国人工智能学会优博奖（2019 年）、ACM 中国优博奖（2018 年）。

### 谢伟迪 上海交通大学

**报告题目：**基于知识驱动的多模态表征学习。

**报告摘要：**近年来，基于图像-文本学习的模型越来越受到关注。其中，大语言模型中编码了大量人类知识，使得模型具备更强的表征能力和泛化能力。在这种背景下，知识驱动的多模态表征学习逐渐成为了研究的热点。在本报告中，我将介绍我对知识驱动表征学习的理解，知识的编码方式，包括知识图谱和图像、文本等多模态数据的处理方法。接着，我们将以开放集视觉应用为例，阐述知识驱动表征学习在面向真实场景中的应用，例如，开放集物体检测、开放集动作识别、开放集物体分割、电影自动语音字幕等多个领域的应用效果，并探讨如何进一步提高其性能和可解释性。最后，我们将以医疗应用为另一个并列的应用场景，医疗图像分析、疾病预测和临床决策等方面阐述知识驱动表征学习在该知识密集领域中的应用价值，并探讨如何进一步提高其性能和可解释性。



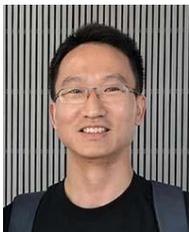
**讲者简介：**谢伟迪，上海交通大学副教授、博士生导师。2018 年于英国牛津大学视觉几何组（VGG）获得博士学位，先后担任牛津大学博士后研究员，高级研究学者，获 Oxford-Google DeepMind Scholarship, Magdalen Award (China-Oxford Scholarship Funds), Oxford Excellence Award, 上海市领军人才，科技部“新一代人工智能”重大项目青年项目负责人。发表论文 40 余篇，Google Scholar 引用超 6500 次，开源多个标准领域数据集，包括 VGGFace2, Voxceleb VGGSound, MoCA, 下载量超 25 万次。担任 CVPR2023, NeurIPS2023 领域主席。主要研究领域为大规模多模态表征学习。个人主页：<https://weidixie.github.io>。

### 孙奕帆 百度

**报告题目：**Exploring the Prototype and Scope of Visual Concept

**报告摘要：**Prototype and scope are two essentials of visual knowledge. This presentation discusses these two characteristics and their benefit on visual recognition tasks. Specifically, we

are interested in: 1) how to learn deep embeddings with hierarchical prototypes, 2) how to utilize the relationship between adjacent prototypes for cross-domain generalization and 3) how to learn an image classification transformer that benefits from embedding prototypes and scope simultaneously.



**讲者简介:** 孙奕帆, 百度资深研究员。主要研究方向为计算机视觉和光电成像系统。特别是在计算机视觉领域, 围绕深度度量学习、不良数据学习问题以及视觉大模型等方向开展了一系列研究与应用探索。在模式识别及计算机视觉国际顶级会议、期刊发表三十余篇论文, 目前 google scholar 被引用次数累计 5100 余次, 最高单篇论文引用次数 1800 余次。部分研究成果(如 PCB、Circle Loss) 获得产业应用, 并成为所在领域内重要基线方法。孙奕帆曾获得 2016 年某省部级科技进步二等奖(第一完成人)、2019 年吴文俊人工智能科学技术二等奖、中国图象图形学会优秀博士论文, 并曾承担 ICPR 领域主席、AAAI 高级程序主席等学术服务工作。

## Workshop 18: 女性研究者论坛

主席：雷柏英（深圳大学）张姗姗（南京理工大学）贲晔焯（山东大学）

时间：6月11日（周日）13:30-18:00 地点：兰花厅 B

时间	主持人	内容
13:30-13:55	雷柏英	讲者：邬霞（北京师范大学） 题目：时-空共变脑功能分析算法研究
13:55-14:20		讲者：景丽萍（北京交通大学） 题目：认知启发的目标跟踪探索
14:20-14:45		讲者：杨欣（华中科技大学） 题目：基于视觉的深度估计与定位
14:45-15:10		讲者：王珊珊（深圳先进技术研究院） 题目：知识驱动的机器学习智能医学成像与分析
15:10-15:25		中场休息
15:25-15:50	张姗姗	讲者：付莹（北京理工大学） 题目：暗光图像视频增强
15:50-16:15		讲者：霍静（南京大学） 题目：跨模态图像合成关键技术
16:40-17:05		讲者：张曼（北京邮电大学） 题目：步态识别技术和应用
17:05-18:00	贲晔焯	Panel 嘉宾：邬霞（北京师范大学）、景丽萍（北京交通大学）、杨欣（华中科技大学）、王珊珊（深圳先进技术研究院）、付莹（北京理工大学）、霍静（南京大学）、张曼（北京邮电大学）

**郭霞 北京师范大学**

**报告题目：**时-空共变脑功能分析算法研究

**报告摘要：**脑功能分析是开展脑科学研究的核心关键基础，其有效分析理论与应用方法对于脑认知研究和脑疾病辅助诊疗具有重要意义。研究表明，大脑功能是由多时空尺度上运行的众多神经回路的相互作用完成的，即不仅在空间上存在脑区协同，且在时间上存在着动态变化特性。本报告拟从大脑在外界刺激时产生的脑功能激活和脑功能网络变化出发，通过构建大脑时-空共变模式的算法模型，实现大脑功能活动过程的精确刻画，并进一步引入更具有泛化能力的模型框架，有望为个性化的大脑活动过程刻画提供更泛化的技术框架。



**讲者简介：**郭霞，北京师范大学人工智能学院教授、博士生导师，国家自然科学基金优秀青年基金、教育部新世纪优秀人才、吴文俊人工智能科学技术奖自然科学一等奖、教育部自然科学二等奖、茅以升北京青年科技奖获得者。主要研究方向为脑信号智能分析、类脑算法等。近年来，主持承担国家自然科学基金、国家重点研发计划等项目十余项、以第一/通讯作者在 IEEE TNNLS、NeuroImage、Medical Image Analysis、IPMI、MICCAI 等国内外重要学术期刊、会议发表论文 80 余篇。

**景丽萍 北京交通大学**

**报告题目：**认知启发的目标跟踪探索

**报告摘要：**新一代信息技术的快速发展，各领域已搜集到大规模视频数据，能否从中定位并跟踪指定目标已成为 AI+ 领域成功应用的关键要素。本报告针对复杂视频数据中目标跟踪任务，从心理学与认知科学角度出发，受人类认知过程中注意力机制、多模态增强和多线索校准的启发，探索视频中多重交互、概念自适应、渐进式关联的单/多目标跟踪理论和算法，部分研究成果在实际问题处理中产生较好的效果。



**讲者简介：**景丽萍，博士，北京交通大学计算机与信息技术学院教授、博士生导师，现任计算机与信息技术学院副院长。担任中国计算机学会人工智能与模式识别专委会委员、秘书，中国人工智能学会机器学习专委会常委。先后入选国家级青年人才计划、教育部创新团队负责人、北京市课程思政教学名师、北京市海淀区人大代表、北京交通大学卓越百人，获评北京交通大学“巾帼十杰”、“优秀教师”等。主持多项省部级、国家级项目，包括国家自然科学基金优秀项目、科技部新一代人工智能重大专项、国防科技创新重点项目、教育部人工智能算法战略研究项目、北京市自然科学基金重点研究专题等。近年来在国内外重要学术期刊和会议上发表多篇高质量论文（包括 CCF-A 类顶级国际学术会议 NeurIPS、AAAI、ACL、ICCV、ICLR、WWW、CVPR，顶级学术期刊 JMLR、IJCV、IEEE TPAMI 等），相关研究成果已成功应用于智能交通、智能芯片、智能教育、智能国防等领域。

**杨欣 华中科技大学**

**报告题目：**基于视觉的深度估计与定位

**报告摘要：**基于视觉的深度估计与定位技术被广泛应用于自动驾驶、增强现实、三维重建等多项任务中。本报告中我们将首先分享我们近期在深度估计方向上一系列研究工作，包

括：1) 基于注意力级联体的双目/多目立体匹配方法。通过在构建代价过程中充分结合局部几何特征与全局语义特征，能有效降低后续代价体聚合或多层级循环更新模块计算开销，提升聚合/更新效率；2) 基于视觉的深度与运动联合估计方法。通过结合光流估计和单视角深度估计网络的互补特征（即，光流网络学习匹配特征，单视角深度估计学习结构特征），能有效提升两项任务的精度和效率。进一步，利用光流、深度及相机位姿估计三者特征间的相互约束，能提升无监督深度、光流及位姿联合估计精度。在报告的第二部分，我们将介绍我们近期在多传感器里程计方面的研究进展，我们所提出并开源的视觉与雷达融合里程计 DSV-LOAM 在 KITTI Odometry 等多个数据集上获得 State-of-the-Art 的精度。



**讲者简介：**杨欣，女，华中科技大学电信学院教授，国家优秀青获得者；博士毕业于加州大学圣塔芭芭拉分校，主要从事计算机视觉和医学影像智能分析领域研究工作，已在国际权威期刊及学术会议上发表学术论文 90 余篇，授权美国/中国发明专利 20 余项，出版英文书籍章节 2 部；获得中国图象图形学会石青云女科学家奖；担任 SCI 期刊 IEEE-TMI 及 Multimedia System 编委及 ACM MM2018 和 MICCAI2019-2021 领域主席；现任 IEEE Bio Imaging Signal Processing (BISP) Technical Committee (TC) 技术委员会委员，CSIG 青工委副秘书长，CCF 多媒体专委会副秘书长等。

### 王珊珊 深圳先进技术研究院

**报告题目：**知识驱动的机器学习智能医学成像与分析

**报告摘要：**人工智能医学影像是近几年非常火爆的研究方向，然而，不同于自然图像，医学影像具有丰富的领域知识（如复杂的物理成像过程、多通道的硬件配置、细粒度的任务属性与多模态的数据特性等等），简单的将视觉任务网络搬迁至医学影像会带来各种挑战与问题。这里将给大家汇报我们利用医学影像领域知识进行机器学习建模的一些探索，并将提出一些开放性问题和大家进行探讨。



**讲者简介：**中科院深圳先进研究院研究员，博士生导师，国家优秀青年基金获得者，入选 2022 年度斯坦福发布“全球前 2% 顶尖科学家”榜单。研究方向机器学习、快速医学成像、放射组学等，迄今为止在 Nature 子刊、IEEE Trans 等发表学术论文 100 多篇，ESI 高被引 4 篇，PMB 年度亮点一篇，发明中国授权专利 9 项以及美国授权专利 2 项，4 项实现规模化产业应用。曾获国际华人医学磁共振协会杰出研究奖，吴文俊人工智能优秀青年奖，广东省科技进步一等奖等，IEEE Senior member, OCSMRM BoT/Life member, Gordon Plenary Lecturer, ISMRM NIBIB New Horizons Plenary Lecturer. 为多个 JCR 1 区国际 SCI 学术期刊的编委（如 IEEE Transactions on Medical Imaging,

Magnetic resonance in medicine, Pattern recognition, IEEE reviews in biomedical engineering and Biomedical signal processing and control)。

### 付莹 北京理工大学

**报告题目：**暗光图像视频增强

**报告摘要：**从低信噪比的暗光图像视频中重建清晰数据是底层视觉研究中的一项重要任务，

其在遥感、生物光学、诊断医学等领域有着广泛的应用。图像视频增强质量很大程度上取决于所采用的噪声模型以及图像先验的准确性，本次报告将从成像传感器的物理特性出发，构建图像视频时空先验，突破传统成像系统的限制，在暗光成像相关应用中展示其优越性能。



**讲者简介：**付莹，北京理工大学教授，博士生导师。2009年获西安电子科技大学电子信息工程专业学士学位，2012年获清华大学自动化专业硕士学位，2015年博士毕业于东京大学信息科学与技术专业。2016年加入北京理工大学计算机学院，2017年入选国家级高层次人才计划。主要从事计算机视觉、计算摄像、图像/视频处理等相关领域研究，已在IEEE TPAMI、IJCV、TIP等国际期刊和ICCV、CVPR、ICML等国际会议发表论文40余篇，曾获得ICML2020最佳论文奖。

### 霍静 南京大学

**报告题目：**跨模态图像合成关键技术

**报告摘要：**图像合成是指按照特定的对图像内容的编辑需求生成图像的过程，如艺术风格图像合成、人像属性编辑、文本指导的图像合成等。随着基于深度学习的风格迁移、生成对抗网络以及扩散模型等生成式模型的发展，图像合成领域取得了非常重要的突破，本次报告将主要介绍目前主流生成式模型的相关进展，分析不同复杂场景下图像合成问题的难点，同时着重介绍在跨模态图像合成领域，围绕模态解耦、可控编辑、语义细节保持等方面的技术方法以及我们取得的一些进展，最后，报告将对该领域的未来发展方向进行展望。



**讲者简介：**霍静，博士，南京大学计算机科学与技术系准聘副教授，分别于2017年，2011年在南京大学计算机科学与技术系获得博士学位以及在南京师范大学强化培养学院获得学士学位。曾分别在英国曼彻斯特大学，香港大学等高校进行学术访问交流。2022年入选国家高层次特殊人才青年拔尖人才计划，2018年获得江苏省计算机学会优博，2018年获得江苏省科学技术奖二等奖。研究方向为新型机器学习技术，包括生成式模型、小样本学习、强化学习及其智能感知与决策应用。主持国家自然科学基金面上基金、青年基金各1项，江苏省自然科学基金青年基金1项，参与国家自然科学基金重大项目1项，科技部2030新一代人工智能项目2项等。在相关研究领域的期刊会议发表论文40余篇，包括CVPR、ICCV、AAAI、IJCAI、ACM MM、TPAMI、TIP、TMM、TNNLS、TCYB、PR等。

### 张曼 北京邮电大学

**报告题目：**步态识别技术和应用

**报告摘要：**步态是远距离复杂场景下唯一可清晰成像的生物特征，难以伪装且成像时无需主动配合。步态识别通过人的身体体型和行走规律来识别身份或分析人体特征，在公共安全、智能家居、医疗体育等领域具有广泛的应用前景。非受控场景中常见的跨视角、遮挡等问题也为步态识别带来了极大的挑战。本次报告将介绍步态识别的概念、起源、发展，以及最新研究进展，以及在多个场景的应用现状。



**讲者简介：**张曼，北京邮电大学研究员、博士生导师，主要研究方向为计算机视觉、模式识别、生物特征识别，已主持国家自然科学基金面上项目、国家重点研发计划子课题、北京市科委人工智能专项等多项国家和省部级科研项目的研发工作，成果发表在人工智能本领域知名期刊和会议（IEEE TPAMI、IEEE TIP、AAAI 等）。曾获得北京市三八红旗奖章、北京市青年骨干、北京市科技新星、北京市科协青年人才托举、海淀青年榜样等奖项。现任 IEEE 女工程师协会北京分会主席、中国人工智能学会模式识别专委会副秘书长、中国图象图形学学会视觉检测专委会副秘书长、北京图象图

形学学会理事、北京市女科协理事、北京市青联十二届委员。

## Workshop 19: 优秀学生论坛

主席: 董胤蓬 (清华大学) 盛律 (北京航空航天大学) 邓志杰 (上海交通大学)

时间: 6月12日(周一) 13:30-18:00 地点: 兰花厅 B

时间	主持人	内容
13:30~13:40	董胤蓬	讲者: 张博航 (北京大学) 题目: Rethinking the Expressive Power of GNNs via Graph Biconnectivity
13:40~13:50		讲者: 许牧天 (香港中文大学-深圳) 题目: 基于遮蔽建模的三维室内场景自监督预训练方法及大规模多视角图片数据集
13:50~14:00		讲者: 陈涵晟 (同济大学) 题目: 基于单阶段扩散 NeRF 的 3D 生成与重建
14:00~14:10		讲者: 杨啸 (清华大学) 题目: Towards Effective Adversarial Textured 3D Meshes on Physical Face Recognition
14:10~14:20		讲者: 董博文 (哈尔滨工业大学) 题目: Self-Promoted Supervision for Few-Shot Transformer
14:20~14:30		讲者: 杨木李 (西安电子科技大学) 题目: Towards Distribution-Agnostic Novel Class Discovery
14:30~14:40		讲者: 黄振宇 (四川大学) 题目: Noisy Correspondence: A New Paradigm for Learning with Noisy Label
14:40~14:50		讲者: 方浩树 (上海交通大学) 题目: 人类级别的通用机器人抓取
14:50~15:00		讲者: 张曦予 (西北工业大学) 题目: 3D Registration with Maximal Cliques
15:00~15:30		邓志杰
15:30~15:40		中场休息
15:40~16:20	盛律	师生圆桌交流
16:20~17:00		导师总结与话题讨论
17:00~17:30		师生自由交流

## 张博航 北京大学

**报告题目:** Rethinking the Expressive Power of GNNs via Graph Biconnectivity

**报告摘要:** 如设计具有强大表达能力的图神经网络是图机器学习领域的一个核心话题。近年来, 虽然已经提出了大量的网络结构, 但人们尚不清楚这些网络结构能够系统地获得怎样的表达能力。这次报告将从图双连通性的新颖角度来系统探究这一领域的基本问题。图双连通性是图论中的核心概念, 它们在各类图学习中具有重要意义。然而, 我们的结果表明, 绝大多数主流的图神经网络不具有在图双连通性方面的任何表达能力。我们进而提出了一种系统性的方法, 通过引用距离信息来可证明地获得对各类图双连通性问题的表达能力。实践中, 该方法可以被基于 Transformer 的网络架构高效的实现, 并在各种任务中取得了比先前方法更好的实验结果。



**讲者简介:** 北京大学智能学院博士四年级学生, 导师是王立威教授。主要研究方向为机器学习, 聚焦图神经网络、鲁棒性、表达能力等领域的基础性问题。在 ICLR、ICML、NeurIPS 三大机器学习国际顶级会议上以第一作者身份发表 6 篇论文, 多篇论文获得口头报告。曾荣获 ICLR 2023 会议杰出论文奖。长期担任 TPAMI、ICLR、ICML、NeurIPS、CVPR 等期刊和会议审稿人。连续四年获得北京大学校长奖学金; 曾获北京大学石青院士优秀论文奖。

## 许牧天 香港中文大学-深圳

**报告题目:** 基于遮蔽建模的三维室内场景自监督预训练方法及大规模多视角图片数据集

**报告摘要:** 遮蔽建模 (Masked Modeling) 通过重建被遮蔽的视觉局部区域, 在各种视觉自监督预训练任务中显示了广泛的成功。然而, 由于数据的稀缺性和场景的复杂性, 将遮蔽建模应用于大规模三维场景仍然是一个开放的问题。我们提出了一种信息保留式重建方法, 该方法利用局部统计来发现和保留有代表性的结构点, 与渐进式重建方式相结合以专注于区域几何建模, 从而减少了预训练任务的 ambiguity, 以提取出有效的先验特征, 以服务于三维场景理解的一系列下游任务。2. 我们构建了一个多视角的图片数据集 MVImgNet。我们共收集约 22 万段 Object-Centric 多视角视频, 共 238 个常见物体类别, 每段视频我们均做了 Colmap 相机重建。从中, 我们也提取并标注了 8 万 7 千个物体 (150 物体类别) 的真实点云数据集 MVPNet。我们探索了数据集对 Nerf 重建、MVS 重建、点云分类以及 view-consistent 图像分类和多视角显著物体检测任务均有效。目前数据已开源。



**讲者简介:** 许牧天, 香港中文大学 (深圳) 二年级博士研究生, 导师为韩晓光教授。主要研究方向为三维物体或场景的理解与感知, 以及自监督学习、数据生成。在 CVPR, ECCV, AAAI 发表多篇论文, 其中一作五篇, 一篇一作文章被选为 ECCV 2022 口头汇报。截至目前, 一作论文引用量 250 余次。受邀担任 IJCV/TVCG/CVPR/ICCV/NeurIPS/AAAI 等顶级期刊/会议的审稿人。曾在香港大学 CVMI Lab 和深圳先进技术研究院多媒体中心访问。个人主页: <https://mutianxu.github.io/>。

### 陈涵晟 同济大学

**报告题目：**基于单阶段扩散 NeRF 的 3D 生成与重建

**报告摘要：**生成模型与神经渲染的快速发展使 3D 生成与重建成为了当下的研究热点之一。本次报告将分享我们在该领域的最新研究——单阶段扩散 NeRF (Single-Stage Diffusion NeRF, SSDNeRF)。SSDNeRF 提出了一种单阶段的 3D 扩散模型与 NeRF 联合端到端训练方法，通过大量物体的多视角图片同步进行 3D 重建与生成先验学习，无需任何多阶段训练或与预训练模型。在推理阶段，其生成先验可用于无条件采样新物体，亦可结合单张/多张图像的观测进行基于图像的 3D 重建，在多项任务中均可达到或超越以往 SOTA 的性能。



**讲者简介：**陈涵晟，同济大学 2023 届硕士生（上海市优秀毕业生），斯坦福大学计算机科学博士预录取生。他的研究兴趣为 3D 计算机视觉和图形学，主要方向包括 3D 生成与重建、神经渲染、几何视觉等。硕士期间在 CVPR 发表一作论文两篇，获 CVPR 2022 最佳学生论文奖，并担任 CVPR、ICCV、SIGGRAPH 等计算机视觉和图形学顶级会议的审稿人。先后在阿里巴巴达摩院，加州大学圣地亚哥分校（远程）实习。

### 杨啸 清华大学

**报告题目：**Towards Effective Adversarial Textured 3D Meshes on Physical Face Recognition

**报告摘要：**Face recognition is a prevailing authentication solution in numerous biometric applications. Physical adversarial attacks, as an important surrogate, can identify the weaknesses of face recognition systems and evaluate their robustness before deployed. However, most existing physical attacks are either detectable readily or ineffective against commercial recognition systems. The goal of this work is to develop a more reliable technique that can carry out an end-to-end evaluation of adversarial robustness for commercial systems. It requires that this technique can simultaneously deceive black-box recognition models and evade defensive mechanisms. To fulfill this, we design adversarial textured 3D meshes (AT3D) with an elaborate topology on a human face, which can be 3D-printed and pasted on the attacker's face to evade the defenses. However, the mesh-based optimization regime calculates gradients in high-dimensional mesh space, and can be trapped into local optima with unsatisfactory transferability. To deviate from the mesh-based space, we propose to perturb the low-dimensional coefficient space based on 3D Morphable Model, which significantly improves black-box transferability meanwhile enjoying faster search efficiency and better visual quality. Extensive experiments in digital and physical scenarios show that our method effectively explores the security vulnerabilities of multiple popular commercial services, including three recognition APIs, four anti-spoofing APIs, two prevailing mobile phones and two automated access control systems.



**讲者简介：**杨啸，清华大学计算机系博士四年级，导师为朱军教授，主要研究方向是机器学习中的对抗鲁棒性。目前在机器学习顶级会议 NeurIPS、CVPR、ICCV 等上发表十余篇论文。此外，多次在国际对抗攻防竞赛中均获得冠/亚军，包括国际安全极客大赛 GeekPwn 2020 DeepFake 赛道冠军，GeekPwn 2019 CAAD CTF 及 Adversarial Patch 两个赛道冠军，CVPR 2021 无限制扰动对抗攻击竞赛亚军等。多次担当 TPAMI、NeurIPS、CVPR、ECCV、ICCV 等审稿人，并参与共同举办 AAAI 2022 Workshop 等活动。

**董伯文 哈尔滨工业大学**

**报告题目:** Self-Promoted Supervision for Few-Shot Transformer

**报告摘要:** The few-shot learning ability of vision transformers (ViTs) is rarely investigated though heavily desired. In this work, we empirically find that with the same few-shot learning frameworks, replacing the widely used CNN feature extractor with a ViT model often severely impairs few-shot classification performance. Moreover, our empirical study shows that in the absence of inductive bias, ViTs often learn the low-qualified token dependencies under few-shot learning regime where only a few labeled training data are available, which largely contributes to the above performance degradation. To alleviate this issue, we propose a simple yet effective few-shot training framework for ViTs, namely Self-promoted sUpervision (SUN). Specifically, besides the conventional global supervision for global semantic learning, SUN further pretrains the ViT on the few-shot learning dataset and then uses it to generate individual location-specific supervision for guiding each patch token. This location-specific supervision tells the ViT which patch tokens are similar or dissimilar and thus accelerates token dependency learning. Moreover, it models the local semantics in each patch token to improve the object grounding and recognition capability which helps learn generalizable patterns. To improve the quality of location-specific supervision, we further propose: 1) background patch filtration to filtrate background patches out and assign them into an extra background class; and 2) spatial-consistent augmentation to introduce sufficient diversity for data augmentation while keeping the accuracy of the generated local supervisions. Experimental results show that SUN using ViTs significantly surpasses other few-shot learning frameworks with ViTs and is the first one that achieves higher performance than those CNN state-of-the-arts.



**讲者简介:** 哈尔滨工业大学与香港理工大学联合培养博士生，导师为左旺孟老师和张磊老师。他的主要研究方向为弱监督学习，少样本学习和迁移学习。他曾在 ICCV, ECCV, CVPR, ICLR 等会议发表多篇论文，并多次担任 CVPR, ICCV, ECCV, TIP 等会议或期刊的审稿人。

**杨木李 西安电子科技大学**

**报告题目:** Towards Distribution-Agnostic Novel Class Discovery

**报告摘要:** Novel Class Discovery (NCD) aims to discover unknown classes without any annotation, by exploiting the transferable knowledge already learned from a base set of known classes. Existing works hold an impractical assumption that the novel class distribution prior is uniform, yet neglect the imbalanced nature of real-world data. In this work, we relax this assumption by proposing a new challenging task: distribution-agnostic NCD, which allows data drawn from arbitrary unknown class distributions and thus renders existing methods useless or even harmful. We tackle this challenge by proposing a new method, dubbed "Bootstrapping Your Own Prior (BYOP)", which iteratively estimates the class prior based on the model prediction itself. At each iteration, we devise a dynamic temperature technique that better estimates the class prior by encouraging sharper predictions for less-confident samples. Thus, BYOP obtains more accurate pseudo-labels for the novel samples, which are beneficial for the next training iteration. Extensive experiments show that existing methods suffer from imbalanced class distributions, while BYOP outperforms them by clear margins, demonstrating its effectiveness across various

distribution scenarios.



**讲者简介:** 西安电子科技大学电子工程学院博士生, 导师为邓成教授。主要研究方向为迁移学习和分布外泛化。在 NeurIPS、CVPR、ICCV、TIP 等高水平会议和期刊上发表学术论文十余篇。曾获博士研究生国家奖学金等荣誉。多次担任 ICML、NeurIPS、ICLR、CVPR、ICCV、TCSVT、TNNLS 等顶级国际会议和期刊的审稿人。

### 黄振宇 四川大学

**报告题目:** Noisy Correspondence: A New Paradigm for Learning with Noisy Label

**报告摘要:** In this talk, I will introduce a new paradigm for label noise, namely, Learning with Noisy Correspondence. Specifically, the widely-studied label noise usually focuses on category-level annotation errors, and most existing studies take classification as the evaluation scenario. In contrast, noisy correspondence refers to the matching errors in paired data such as the mismatch pairs in cross-modal retrieval and wrongly-related question-answer pairs in machine reading comprehension. It is undeniable that such noise would provide unreliable/uncertain supervised information and significantly degrade the performance. In the presentation, I will introduce our recent works which study noisy correspondence in various applications with different causes of formation including the wrongly-matched image-text pairs in cross-modal retrieval, twin noisy labels in visible-infrared ReID, and self-generated noisy question-answer pairs in machine reading comprehension.



**讲者简介:** 黄振宇, 四川大学计算机学院博士研究生, 主要研究领域包含多模态学习和噪声学习; 在 TPAMI、TIP、NeurIPS、ICML 等顶级国际期刊和会议发表论文 10 篇, 包含 3 篇 Oral (ICML、NeurIPS) 论文; 获四川大学“十佳学术之星”、腾讯奖学金、VALSE 焦点论文奖等奖项或荣誉; 多次在 NeurIPS、AAAI、IJCAI 等学术会议和 VALSE、MLA 等国内学术交流论坛上作报告; 多次担任 JMLR、TPAMI、ICML、NeurIPS、ICLR 等顶级国际会议和期刊的审稿人。

### 方浩树 上海交通大学

**报告题目:** 人类级别的通用机器人抓取

**报告摘要:** 通用抓取能力是物体操作的基础, 其在智能体进化过程中具有重要地位。在人类进化史上, 高级的视觉引导抓取能力被认为是早期灵长类动物的特征之一。今天, 机器人社区为机器人手开发了不同的机械结构。无论是从智能进化的角度来看还是从行业角度来看, 有效地赋予不同机械结构的灵巧手由视觉引导的抓取能力是十分重要的。在今天的报告中, 我将分享如何仅从 40 个训练物体中高效学习机器人抓取, 对于不同的机

械手，均能达到人类级别的学习效率，抓取成功率与抓取速度。



<https://fang-haoshu.github.io>

**讲者简介：**方浩树，上海交通大学直博四年级博士生，入选致远荣誉博士计划与吴文俊班，本科毕业于上海交通大学。其主要研究方向为人类行为理解与机器人物体操作，研究愿景为赋予机器人灵巧的操作技能，代替人类在工业农业服务业中的繁琐工作。他的代表项目为 AlphaPose 与 AnyGrasp。在本科与博士期间，方浩树在 TPAMI, Nature, RAL, CVPR, ICCV, ECCV, AAAI, ICRA 等期刊和会议上发表二十余篇论文，其中十篇一作，总引用量 3500 余次，并担任 TPAMI, TRO, IJCV, TMM, TOC, CVPR, NIPS, ICCV, ECCV, ICLR 等期刊与会议审稿人。方浩树曾获 CF-CV 新锐奖，百度奖学金，微软学者奖学金，字节跳动奖学金，国家奖学金等。个人主页：

### 张曦予 西北工业大学

**报告题目：**3D Registration with Maximal Cliques

**报告摘要：**三维点云配准是计算机视觉中的一个基本问题，旨在寻找最佳姿态以对齐点云对。本文提出了一种基于极大团（MAC）的三维配准方法。目的是放松先前的最大团约束，并在图中挖掘更多的局部一致性信息以进行准确的姿态假设生成：首先，构建兼容性图以呈现初始对应关系之间的亲和关系；其次，在图中搜索极大团，每个极大团代表一个共识集，然后执行节点引导的团选择，其中每个节点对应于具有最大图权重的极大团。最后，通过 SVD 算法计算团的变换假设，然后使用最佳假设进行配准。在 U3M、3DMatch、3DLoMatch 和 KITTI 数据集上进行的广泛实验，证明 MAC 有效地提高了配准精度，优于各种最先进的方法，并提高了深度学习方法的性能。我们的工作证明了传统几何方法针对三维点云配准重建仍有巨大的潜力。



**讲者简介：**西北工业大学计算机学院在读硕士，主要研究领域为三维点云特征匹配与点云配准，导师为杨佳琪副教授。已在 IEEE TPAMI 2023 上发表论文《Mutual Voting for Ranking 3D Correspondences》，在 CVPR 2023 投稿的论文《3D Registration with Maximal Cliques》入围最佳论文候选（录用率 0.13%，中国大陆高校第一作者单位唯一入选）。

## Workshop 20: 开放域的学习方法与视觉任务

主席：苏航（清华大学） 徐航（华为诺亚方舟实验室） 张磊（重庆大学） 宫辰（南京理工大学）

时间：6月12日（周一）13:30-18:00 地点：梅花厅

时间	主持人	内容
13:30~14:00	苏航	讲者：代季峰（清华大学） 题目：开放域物体识别算法
14:00~14:30		讲者：段楠（微软亚洲研究院） 题目：从视觉内容生成到视觉任务自动化
14:30~15:00		讲者：张伟楠（上海交通大学） 题目：面向开放域的虚实迭代强化学习
15:00~15:30		讲者：罗平（香港大学） 题目：混合大模型
15:30~15:45		Poster Spotlight
15:45~16:15	徐航	讲者：张幸幸（启元实验室） 题目：动态开放环境的具身持续学习方法
16:15~16:45		讲者：王文海（上海人工智能实验室） 题目：大规模基础模型在开放世界任务中的研究与应用
16:45~17:15		讲者：梁小丹（中山大学） 题目：多模态开放域检测大模型及应用
17:15~18:00		Panel 嘉宾：宫辰(南京理工大学)、苏航(清华大学)、代季峰(清华大学)、段楠(微软亚洲研究院)、张伟楠(上海交通大学)、梁小丹(中山大学)、罗平(香港大学)、张幸幸(启元实验室)、王文海(上海人工智能实验室)

## 代季峰 清华大学

**报告题目：**开放域物体识别算法

**报告摘要：**开放世界物体识别已成为提高人工智能系统能力的关键组成部分。目前的物体识别模型受限于对封闭世界假设和预定义物体类别的依赖，导致在现实场景中的性能表现不佳。本次演讲将介绍针对开放世界物体识别的新方法，旨在克服这些局限。开放世界物体识别任务为机器人、自动驾驶汽车以及其他 AI 驱动系统的应用开辟了新的发展方向。



**讲者简介：**代季峰，清华大学电子工程系副教授，博士生导师。在 2009 年和 2014 年于清华大学自动化系分别获得工学学士和博士学位，博士生导师周杰教授。2014 年至 2019 年在微软亚洲研究院视觉组工作，担任首席研究员、研究经理。2019 年至 2022 年在商汤科技研究院工作，担任执行研究总监，二级部门长。2022 年 7 月全职加入清华大学电子工程系。他的研究兴趣包括计算机视觉、深度学习等。他在相关领域发表国际期刊、会议文章 50 余篇，论文总引用 2 万 6 千余次。多篇论文成为物体识别领域里程碑式的成果，被编入世界一流大学视觉课程讲义，被选入深度学习权威框架 PyTorch 成为标准算子。他连续两年获得物体识别领域权威的 COCO 比赛冠军，之后历届冠军系统也使用了他提出的算法。他提出的算法获得自动驾驶感知领域权威的 Waymo 2022 竞赛冠军。他是顶刊 IJCV 的编委，和顶会 NeurIPS 2023, ICCV 2023, CVPR 2023, CVPR 2021, ECCV 2020 的领域主席，ICCV 2019 的宣传主席。

## 段楠 微软亚洲研究院

**报告题目：**从视觉内容生成到视觉任务自动化

**报告摘要：**近年来，生成式基础模型 (generative foundation model) 让人们看到构建通用人工智能的曙光。以 DALL E 和 Stable Diffusion 为代表的视觉生成模型在视觉内容生成领域同样发展迅速，并快速迭代到高清图片生成和超长视频生成等任务。于此同时，受益于基础模型的强大自然语言理解、推理和代码生成能力，该类模型已经开始被用于自动完成更多样化的视觉理解和生成任务。在这一大背景下，本报告第一部分将简要介绍 MSRA 在视觉内容生成领域的典型工作(女蜗系列模型)，第二部分将重点介绍 MSRA 在视觉任务自动化领域的最新进展(TaskMatrix.AI)。报告的结尾将对未来研究方向做一个简单的讨论和展望。



**讲者简介：**段楠博士，微软亚洲研究院首席研究员，自然语言计算团队研究经理，中国科学技术大学兼职博导，天津大学兼职教授，中国计算机协会杰出会员，主要从事自然语言处理、代码智能、多模态智能、机器推理等研究，多次担任 NLP/AI 学术会议程序主席、评测主席、资深领域主席和领域主席，发表学术论文 100 余篇，持有专利 20 余项。

## 张伟楠 上海交通大学

**报告题目：**面向开放域的虚实迭代强化学习

**报告摘要：**强化学习正在成为可泛化机器人操作和运动控制的新范式。但强化学习依赖大量交互样本进行训练，因此几乎只能在模拟环境中进行，而模拟环境与真实环境之间存在着真实性差异，因此需要虚实迭代(Sim-to-Real)技术来完成策略从虚拟环境到真实环境的迁

移。在本次报告中，我将概述虚实迭代的基本技术，并以四足机器人运动控制为案例，讨论虚实迭代技术在机器人强化学习中的研究工作。具体地，我将讨论面向不同地形和不同具身形态条件下的虚实迭代泛化技术，这些技术已经在真机上验证，并有望在真实场景中取得有价值的落地。



**讲者简介：**张伟楠博士现任上海交通大学计算机系长聘副教授、博士生导师，科研领域包括强化学习、信息检索和数据科学，相关的研究成果在国际会议和期刊上发表 150 余篇学术论文，谷歌学术引用 1 万余次，获得 4 个最佳论文奖项，入选爱思唯尔中国高被引学者，出版教材《动手学强化学习》，相关成果被应用于华为鸿蒙系统，产生了显著的经济效益。张伟楠长期 TPAMI 编委、FCS 期刊的青年编委以及 NeurIPS、ICML、IJCAI、AAAI 等会议的领域主席。作为负责人承担国家自然科学基金项目和科技部重大项目课题，入选上海市科委英才扬帆计划、中国科协青年人才托举工程，获得吴文俊人工智能优秀青年奖和达摩院青橙奖。张伟楠于 2011 年在上海交通大学计算机系 ACM 班获得学士学位，于 2016 年在伦敦大学学院计算机系获得博士学位。

### 梁小丹 中山大学

**报告题目：**多模态开放域检测大模型及应用

**报告摘要：**近期多模态预训练基础大模型如 GPT4 等在多个视觉语言跨模态任务下得到突出的性能，但是其无法做到细粒度的语义对齐和定位，如利用大量图文对进行开放域物体检测学习。本次报告介绍研究组利用大规模图文数据图进行细粒度物体和文本对齐模型，利用千万数据来训练大模型的模型设计。首先讲介绍一种新的针对开放域检测的端到端预训练模型 DetCLIPv2。其次介绍如何利用 caption 数据来增强开放域检测预训练模型。最后依托于开放域物体检测能力，设计了针对多模态机器人视觉语言导航的多模态决策模型，结合多模态预训练大模型和大规模语言模型如 Chatgpt，在多个 VLN 任务得到明显的性能提升，并支持 Zero-shot 场景的泛化性。



**讲者简介：**梁小丹，中山大学副教授，逸仙学者，国家万人计划青年拔尖人才，IEEE Senior Member。研究领域为多模态视觉语言理解，数字人生成和驱动，可解释 AI 和因果推断机器学习模型。Google 学术引用超 18000 次。现担任 Image and Vision Computing 和 Neural Networks 期刊的 Associate Editor，曾担任 CVPR/ICML/ICCV/Neurips/ACMMM 等会议领域主席和 Ombud chair 等，荣获 ACM 中国新星提名奖，阿里巴巴达摩院青橙奖，CSIG 石青青青年女科学家奖，吴文俊人工智能优秀青年奖，中国科协青年人才托举计划获得者，中国图像图形学会科技技术一等奖，CCF

优秀博士论文奖和 ACM China 优秀博士论文奖,ACM 中国新星奖提名，入选 2022Elsevier 中国高被引学者榜单等。

### 罗平 香港大学

**报告题目：**混合大模型

**报告摘要：**基础模型在多任务学习方面取得了重大进展。然而，在迁移到开放域过程中，多任务学习的潜力尚未充分发挥。我们将介绍一种混合通用基础模型参数迁移学习方法，同时适用于视觉、语言和视觉-语言混合任务。不同任务的相似性可以在一个统一的、模态无关的空间中进行预测，生成一个可扩展的任务关系图，可以深入解释不同模态之间任务

的可迁移性，有理论保证。另外，将介绍混合 Diffusion Paths 用于极大的提高文本-图像生成模型的生成能力。



**讲者简介：**香港大学计算机系副教授。2014 年获香港中文大学信息工程系博士，师从汤晓鸥教授和王晓刚教授。近三年在 TPAMI、ICML、ICLR、NeurIPS、CVPR 等发表 100+ 文章（其中 Oral 25+，Top-10 Most Influential Paper 4），谷歌学术引用 32000+。曾获多个奖项如 2011 香港政府博士奖学金，2013 微软亚洲学者奖，2014 ImageNet 图像识别挑战赛亚军、2015 AAAI Easily Accessible Paper (0.5%)，2020《麻省理工科技评论》亚太区 35 岁以下创新者(MIT TR35)，2022 ACL 杰出论文(0.2%)，2022 Computational Visual Media 期刊最佳论文提名(1/2)。入选 2021~2023 AI2000 计算机视觉最具影响力学者，斯坦福大学全

球前 2% 科学家 (single-year + career long lists)。长期担任 ICML, ICLR, NeurIPS, CVPR 等领域主席。近三年指导 20+ 博士生，多人获重要奖项如 Nvidia Fellowship, Baidu Fellowship, WAIC 云帆奖等。

### 张幸幸 启元实验室

**报告题目：**动态开放环境的具身持续学习方法

**报告摘要：**目前主流的机器学习范式是在静态、封闭的数据集上学习到一个模型，并假设其应用环境和之前训练数据的属性相同，因而无法像生物智能一样快速适应动态和开放环境的挑战。如何通过持续学习这一过程，模拟生物的学习过程和学习能力，发展新型机器学习理论和方法，对于提升算法/智能体对开放、动态环境的适应能力具有十分重要的意义。本次报告将首先定义具身持续学习的概念——智能体在与动态开放环境的交互演化中实现终身的学习记忆；进而分析其中的灾难性遗忘这一理论问题——稳定性可塑性权衡与泛化性分析；然后介绍两类前沿的具身持续学习方法，包括具身持续学习中的主动遗忘、归一化等机制，以及预训练范式下基于 Prompt 的持续学习策略；最后会对未来具身持续学习的发展，尤其是提升智能体对于开放动态环境中新任务/新环境/新形态的适应能力进行展望。



**讲者简介：**张幸幸，启元实验室研究员，清华大学博士后，北京交通大学博士。研究方向为数据时空高效的学习、推理与应用，特别是零、小样本学习和持续学习。相关工作发表于机器学习领域重要的国际期刊和会议，包括 TPAMI、TNNLS、TMM、TCSVT、ICLR、IJCAI、ECCV、CVPR 等 30 余篇，牵头撰写持续学习领域的综述，系统地梳理了持续学习领域的最新进展，提出了未来的发展方向，arXiv 版本一经上线即受到人工智能社区的广泛关注（新智元、专知等人工智能自媒体专题报道，Twitter 关注量 50k+）。

### 王文海 上海人工智能实验室

**报告题目：**大规模基础模型在开放世界任务中的研究与应用

**报告摘要：**近年来，以 ChatGPT 为代表的大规模语言模型已在自然语言处理开放世界任务中获得了极大的成功。然而，由于其受到模态的限制，这些模型在计算机视觉和视觉语言任务中的应用仍然受到限制。在本次报告中，我们将从大规模视觉基础模型开始，探讨视

觉基础模型与大规模语言模型在以视觉为中心的开放世界任务中的应用。本次报告将分为三个部分：(1) 大规模视觉基础模型的构建和训练；(2) 视觉基础模型和大规模语言模型在开放世界视觉交互系统中协同应用；(3) 更灵活的大规模语言模型在开放世界视觉和视觉语言任务中的使用方法。



**讲者简介：**上人工智能实验室青年科学家。研究方向为视觉基础模型研究，上海人工智能实验室“书生”系列视觉基础模型核心开发者。主要成果发表在顶级期刊和会议：TPAMI, CVPR, ICCV, ECCV, ICLR, NeurIPS 等共 31 篇论文，其中 17 篇为一作/共一/通信。研究成果获得了总共超 7000 次引用，单篇最高引用超 1500 次。其中 2 项成果分别入选 CVPR, CVMJ 最佳论文候选，一次入选 ESI 高被引论文（前 1%），4 分别入选 CVPR 2020, ICCV 2021, NeurIPS 2021, ECCV 2022 十大最具影响力论文，一次入选 2022 年百篇最高引 AI 论文。担任 TPAMI, IJCV, CVPR, ICCV, ECCV, IJCAI 等多个国际会议/期刊的高级程序

委员会委员/程序委员会委员/审稿人。

## Poster 交流论文一览表

1	Shiming Chen, Ziming Hong, Wenjin Hou, Guo-Sen Xie, Yibing Song, Jian Zhao, Xinge You, Shuicheng Yan, Ling Shao	卡耐基梅隆大学&阿联 酋人工智能大学	TPAMI 2022
	TransZero++: Cross Attribute-Guided Transformer for Zero-Shot Learning		
2	Mingjun Xu, Lingyun Qin, Weijie Chen, Shiliang Pu, Lei Zhang	重庆大学	CVPR 2023
	Multi-view Adversarial Discriminator: Mine the Non-causal Factors for Object Detection in Unseen Domains		
3	Congyu Qiao, Ning Xu, Xin Geng	东南大学	ICLR 2023
	Decompositional Generation Process for Instance-Dependent Partial Label Learning		
4	Mingwu Zheng, Haiyu Zhang, Hongyu Yang, Di Huang	北京航空航天大学	CVPR 2023
	NeuFace: Realistic 3D Neural Face Rendering from Multi-view Images		
5	Haojie Zhao, Junsong Chen, Lijun Wang, Huchuan Lu	大连理工大学	CVPR 2023
	ARKitTrack: A New Diverse Dataset for Tracking Using Mobile RGB-D Data		
6	Qianyu Zhou, Xiangtai Li, Lu He, Yibo Yang, Guangliang Cheng, Yunhai Tong, Lizhuang Ma, Dacheng Tao	上海交通大学	TPAMI 2022
	TransVOD: End-to-End Video Object Detection with Spatial-Temporal Transformers		
7	Mingjia Li, Binhui Xie, Shuang Li, Chi Harold Liu, Xinjing Cheng	北京理工大学	AAAI 2023
	VBLC: Visibility Boosting and Logit-Constraint Learning for Domain Adaptive Semantic Segmentation under Adverse Conditions		
8	Bencheng Liao, Shaoyu Chen, Xinggang Wang, Tianheng Cheng, Qian Zhang, Wenyu Liu, Chang Huang	华中科技大学	ICLR 2023
	MapTR: Structured Modeling and Learning for Online Vectorized HD Map Construction		
9	Guangwei Gao, Zixiang Xu, Juncheng Li, Jian Yang, Tiejong Zeng, Guo-Jun Qi	南京邮电大学	TIP 2023
	CTCNet: A CNN-Transformer Cooperation Network for Face Image Super-Resolution		
10	Siwei Wang, Xinwang Liu, Suyuan Liu, Jiaqi Jin, Wenxuan Tu, Xinzhong Zhu, En Zhu	国防科技大学	NeurIPS 2022
	Align then Fusion: Generalized Large-scale Multi-view Clustering with Anchor		

	Matching Correspondences		
11	Nian Liu, Xiao Wang, Deyu Bo, Chuan Shi, Jian Pei	北京邮电大学	NeurIPS 2022
	Revisiting Graph Contrastive Learning from the Perspective of Graph Spectrum		
12	Yang Shen, Xuhao Sun, Xiushen Wei	南京理工大学	CVPR 2023
	Equiangular Basis Vectors		
13	Yuxiang Wei, Zhilong Ji, Xiaohe Wu, Jinfeng Bai, Lei Zhang, Wangmeng Zuo	哈尔滨工业大学	CVPR 2023
	Inferring and Leveraging Parts from Object Shape for Improving Semantic Image Synthesis		
14	Kaixin Xiong, Shi Gong, Xiaoqing Ye, Xiao Tan, Ji Wan, Errui Ding, Jingdong Wang, Xiang Bai	华中科技大学	CVPR 2023
	CAPE: Camera View Position Embedding for Multi-View 3D Object Detection.		
15	Yabo Zhang, Mingshuai Yao, Yuxiang Wei, Zhilong Ji, Jinfeng Bai, Wangmeng Zuo	哈尔滨工业大学	NeurIPS 2022
	Towards Diverse and Faithful One-shot Adaption of Generative Adversarial Networks		
16	Yifei Wang, Qi Zhang, Tianqi Du, Jiansheng Yang, Zhouchen Lin, Yisen Wang	北京大学	ICLR 2023
	A Message Passing Perspective on Learning Dynamics of Contrastive Learning		
17	XueFeng Zhu, Tianyang Xu, Zangyong Tang, Zucheng Wu, Haodong Liu, Xiao Yang, XiaoJun Wu, Josef Kittler	江南大学	AAAI 2023
	RGBDIK: A Large-scale Dataset and Benchmark for RGB-D Object Tracking		
18	Chang Jiang, Fei Gao, Biao Ma, Yuhao Lin, Nannan Wang, Gang Xu	杭州电子科技大学	CVPR 2023
	Masked and Adaptive Transformer for Exemplar Based Image Translation.		
19	Hu Yu, Naishan Zheng, Man Zhou, Jie Huang, Zeyu Xiao, and Feng Zhao	中国科学技术大学	ECCV 2022
	Frequency and Spatial Dual Guidance for Image Dehazing		
20	Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, Luc Van Gool	香港中文大学	ECCV 2022
	OSFormer: One-Stage Camouflaged Instance Segmentation with Transformers		
21	Yuxuan Cai, Yizhuang Zhou, Qi Han, Jianjian Sun, Xiangwen Kong, Jun Li, Xiangyu Zhang	旷视科技	ICLR 2023

	Reversible Column Networks		
22	Wenke Huang, Mang Ye, Zekun Shi, He Li, Bo Du	武汉大学	CVPR 2023
	Rethinking Federated Learning with Domain Shift: A Prototype View		
23	Hongbo Bi, Ranwan Wu, Ziqi Liu, Huihui Zhu, Cong Zhang, Tianzhu Xiang	东北石油大学	PR 2023
	Cross-modal Hierarchical Interaction Network for RGB-D Salient Object Detection		
24	Fukun Yin, Wen Liu, Zilong Huang, Pei Cheng, Tao Chen, Gang YU	复旦大学	NeurIPS 2022
	Coordinates Are NOT Lonely -- Codebook Prior Helps Implicit Neural 3D Representations		
25	Jingcheng Ni, Nan Zhou, Jie Qin, Qian Wu, Junqi Liu, Boxun Li, Di Huang	北京航空航天大学	ECCV 2022
	Motion Sensitive Contrastive Learning for Self-supervised Video Representation		
26	Yuecong Xu, Jianfei Yang, Haozhi Cao, Keyu Wu, Wu Min, Zhenghua Chen	新加坡科技研究局&新加坡南洋理工大学	ECCV 2022
	Source-free Video Domain Adaptation by Learning Temporal Consistency for Action Recognition		
27	Bowen Dong, Pan Zhou, Shuicheng Yan, Wangmeng Zuo	哈尔滨工业大学	ICLR 2023
	LPT: Long-tailed Prompt Tuning for Image Classification		
28	Ziheng Chen, Tianyang Xu, Xiao-Jun Wu, Rui Wang, Zhiwu Huang, Josef Kittler	江南大学	AAAI 2023
	Riemannian Local Mechanism for SPD Neural Networks		
29	Donghao Zhou, Chunbin Gu, Junde Xu, Furui Liu, Qiong Wang, Guangyong Chen, Pheng-Ann Heng	中国科学院深圳先进技术研究院	CVPR 2023
	RepMode: Learning to Re-parameterize Diverse Experts for Subcellular Structure Prediction		
30	Zhilu Zhang, Ruohao Wang, Hongzhi Zhang, Yunjin Chen, Wangmeng Zuo	哈尔滨工业大学	ECCV 2022
	Self-Supervised Learning for Real-World Super-Resolution from Dual Zoomed Observations		
31	Zhilu Zhang, Rongjian Xu, Ming Liu, Zifei Yan, Wangmeng Zuo	哈尔滨工业大学	NeurIPS 2022
	Self-Supervised Image Restoration with Blurry and Noisy Pairs		
32	Mengxue Qu, Yu Wu, Wu Liu, Qiqi Gong, Xiaodan Liang,	北京交通大学	ECCV 2022

	Olga Russakovsky, Yao Zhao, Yunchao Wei		
	SiRi: A Simple Selective Retraining Mechanism for Transformer-Based Visual Grounding		
33	Jiakang Yuan, Bo Zhang, Xiangchao Yan, Tao Chen, Botian Shi, Yikang Li, Yu Qiao	复旦大学	CVPR 2023
	Bi3D: Bi-domain Active Learning for Cross-domain 3D Object Detection		
34	Sijin Chen, Hongyuan Zhu, Xin Chen, Yinjie Lei, Gang Yu, Tao Chen	复旦大学	CVPR 2023
	End to End 3D Dense Captioning with Vote2Cap-DETR		
35	Xin Chen, Biao Jiang, Wen Liu, Zilong Huang, Bin Fu, Tao Chen, Jingyi Yu, Gang YU	腾讯科技	CVPR 2023
	Executing your Commands via Motion Diffusion in Latent Space		
36	Rui Zhao, Ruiqin Xiong, Jing Zhao, Zhaofei Yu, Xiaopeng Fan, Tiejun Huang	北京大学	NeurIPS 2022
	Learning Optical Flow From Continuous Spike Streams		
37	Xiang Chen, Hao Li, Mingqiang Li, Jinshan Pan	南京理工大学	CVPR 2023
	Learning A Sparse Transformer Network for Effective Image Deraining		
38	Ren Wang, Haoliang Sun, Yuling Ma, Xiaoming Xi, Yilong Yin	山东大学	CVPR 2023
	MetaViewer: Towards A Unified Multi-View Representation		
39	Sen Yang, Wen Heng, Gang Liu, Guozhong Luo, Wankou Yang, Gang Yu	东南大学	ICLR 2023
	Capturing the Motion of Every Joint: 3D Human Pose and Shape Estimation with Independent Tokens		
40	Lingyu Gu, Yongqi Du, Yuan Zhang, Di Xie, Shiliang Pu, Robert C Qiu, Zhenyu Liao	华中科技大学	NeurIPS 2022
	Lossless: Compression of Deep Neural Networks: A High-dimensional Neural Tangent Kernel Approach		
41	Jiaxu Zhang, Junwu Weng, Di Kang, Fang Zhao, Shaoli Huang, Xuefei Zhe, Linchao Bao, Ying Shan, Jue Wang, Zhigang Tu	武汉大学	CVPR 2023
	Skinned Motion Retargeting with Residual Perception of Motion Semantics & Geometry		
42	Yichuan Mo, Dongxian Wu, Yifei Wang, Yiwen Guo, Yisen Wang	北京大学	NeurIPS 2022

	When Adversarial Training Meets Vision Transformers: Recipes from Training to Architecture		
43	Xin Zou, Weiwei Liu	武汉大学	JMLR 2023
	Generalization Bounds for Adversarial Contrastive Learning		
44	Xiyuan Li, Xin Zou, Weiwei Liu	武汉大学	NeurIPS 2022
	Defending Against Adversarial Attacks via Neural Dynamic System		
45	Huajun Zhou, Bo Qiao, Lingxiao Yang, Jianhuang Lai, Xiaohua Xie	中山大学	CVPR 2023
	Texture-guided Saliency Distilling for Unsupervised Salient Object Detection		
46	Luting Wang, Yi Liu, Penghui Du, Zihan Ding, Yue Liao, Qiaosong Qi, Biaocong Chen, Si Liu	北京航空航天大学	CVPR 2023
	Object-Aware Distillation Pyramid for Open-Vocabulary Object Detection		
47	Ke Li, Kaiyue Feng, Yi-zhe Song	北京邮电大学	CVPR 2023
	Photo Pre-Training, But for Sketch.		
48	Zongbo Han, Zhipeng Liang, Fan Yang, Liu Liu, Lanqing Li, Yatao Bian, Peilin Zhao, Bingzhe Wu, Changqing Zhang, Jianhua Yao	天津大学	NeurIPS 2022
	Umix: Improving importance weighting for subpopulation shift via uncertainty-aware mixup		
49	Hong Chen, Xin Wang, Chaoyu Guan, Yue Liu, Wenwu Zhu	清华大学	ICML 2022
	Auxiliary Learning with Joint Task and Data Scheduling		
50	Zekai Wang, Weiwei Liu	武汉大学	ICML 2022
	Robustness Verification for Contrastive Learning		
51	Yuxin Dong, Tieliang Gong, Shujian Yu, Chen Li	西安交通大学	IEEE Trans. Inf. Theory 2023
	Optimal Randomized Approximations for Matrix-based Rényi's Entropy		
52	Xiao Wang, Zhe Chen, Bo Jiang, Jin Tang, Bin Luo, and Dacheng Tao	安徽大学	TIP 2022
	Beyond Greedy Search: Tracking by Multi-Agent Reinforcement Learning-Based Beam Search		
53	Pengwei Liang, Junjun Jiang, Xianming Liu, Jiayi Ma	哈尔滨工业大学	ECCV 2022
	Fusion from Decomposition: A Self-Supervised Decomposition Approach for Image Fusion		

54	Jianfeng Dong, Shengkai Sun, Zhonglin Liu, Shujie Chen, Baolong Liu, Xun Wang	浙江工商大学	AAAI 2023
	Hierarchical Contrast for Unsupervised Skeleton-based Action Representation Learning		
55	Zhihong Fu, Zehua Fu, Qingjie Liu, Wenrui Cai, Yunhong Wang	北京航空航天大学	IJCAI 2022
	SparseTT: Visual Tracking with Sparse Transformers		
56	Teng Wang, Yixiao Ge, Feng Zheng, Ran Cheng, Ying Shan, Xiaohu Qie, Ping Luo	南方科技大学&香港大 学	CVPR 2023
	Accelerating Vision-Language Pretraining with Free Language Modeling		
57	Yingdong Hu, Renhao Wang, Kaifeng Zhang, Yang Gao	清华大学	ECCV 2022
	Semantic-Aware Fine-Grained Correspondence		
58	Ming Tao, Bing-Kun Bao, Hao Tang, Changsheng Xu	南京邮电大学	CVPR 2023
	GALIP: Generative Adversarial CLIPs for Text-to-Image Synthesis		
59	Weicong Liang, Yuhui Yuan, Henghui Ding, Xiao Luo, Weihong Lin, Ding Jia, Zheng Zhang, Chao Zhang, Han Hu	北京大学	NeurIPS 2022
	Expediting Large-Scale Vision Transformer for Dense Prediction without Fine-tuning		
60	Jie Hu, Linyan Huang, Tianhe Ren, Shengchuan Zhang, Rongrong Ji, Liujuan Cao	厦门大学	CVPR 2023
	You Only Segment Once: Towards Real-Time Panoptic Segmentation		
61	Wanxing Chang, Ye Shi, Hoang Duong Tuan, Jingya Wang	上海科技大学	NeurIPS 2022
	Unified Optimal Transport Framework for Universal Domain Adaptation		
62	Tianyang Xu, Zhen-Hua Feng, Xiao-Jun Wu, Josef Kittler	江南大学	TIP 2023
	Toward Robust Visual Object Tracking With Independent Target-Agnostic Detection and Effective Siamese Cross-Task Interaction		
63	Jian Zhang, Lei Qi, Yinghuan Shi, Yang Gao	南京大学	ECCV 2022
	MVDG: A Unified Multi-view Framework for Domain Generalization		
64	Chunlin Yu, Ye Shi, Zimo Liu, Shenghua Gao, Jingya Wang	上海科技大学	AAAI 2023
	Lifelong Person Re-Identification via Knowledge Refreshing and Consolidation		
65	Jingting Li, Zichao Dong, Shaoyuan Lu, Sujing Wang, Wenjing Yan, Yinhuan Ma, Ye	中国科学院心理研究所	TPAMI 2022

	Liu, Changbing Huang, Xiaolan Fu		
	CAS(ME)3: A Third Generation Facial Spontaneous Micro-Expression Database With Depth Information and High Ecological Validity		
66	Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, Zhongxuan Luo	大连理工大学	CVPR 2022
	Toward Fast, Flexible, and Robust Low-Light Image Enhancement		
67	Hao Lu, Wenzhe Liu, Zixuan Ye, Hongtao Fu, Yuliang Liu, Zhiguo Cao	华中科技大学	NeurIPS 2022
	SAPA: Similarity-Aware Point Affiliation for Feature Upsampling		
68	Shaofei Huang, Zhenwei Shen, Zehao Huang, Zi-han Ding, Jiao Dai, Jizhong Han, Naiyan Wang, Si Liu	北京航空航天大学	CVPR 2023
	Anchor3DLane: Learning to Regress 3D Anchors for Monocular 3D Lane Detection		
69	Juncheng Wang, Junyu Gao, Yuan Yuan, Qi Wang	西北工业大学	TIP 2023
	Crowd localization from gaussian mixture scoped knowledge and scoped teacher		
70	Qi Zhang, Yifei Wang, Yisen Wang	北京大学	NeurIPS 2022
	How Mask Matters: Towards Theoretical Understandings of Masked Autoencoders		
71	Caixia Yan and Qinghua Zheng, Xiaojun Chang, Zhihui Li, Weili Guan, Zongyuan Ge, Lei Zhu	西安交通大学	TPAMI 2022
	ZeroNAS: Differentiable Generative Adversarial Networks Search for Zero-Shot Learning		
72	Zhiqi Li, Wenhai Wang, Hongyang Li, Enze Xie, Chonghao Sima, Tong Lu, Qiao Yu, Jifeng Dai	南京大学	ECCV 2022
	BEVFormer: Learning Bird's-Eye-View Representation from Multi-Camera Images via Spatiotemporal Transformers		
73	Xiangbo Shu, Binqian Xu, Jinhui Tang, Liyan Zhang	南京理工大学	TPAMI 2022
	Multi-Granularity Anchor-Contrastive Representation Learning for Semi-Supervised Skeleton-Based Action Recognition		
74	Chuangang Yang, Zhulin An, Helong Zhou, Fuzhen Zhuang, Yongjun Xu, Qian Zhan	中国科学院计算技术研究所	TPAMI 2023
	Online Knowledge Distillation via Mutual Contrastive Learning for Visual Recognition		
75	Chao Xia, Jie Yang, Jiyue Wang, Bing Chen	上海交通大学	MICCAI 2022 & TMI 2023

	An End-to-End Combinatorial Optimization Method for R-band Chromosome Recognition with Grouping Guided Attention		
76	Yanpeng sun, Qiang Chen, Xiangyu He, Jian Wang, Haocheng Feng, Junyu Han, Errui Ding, Jian Cheng, Zechao Li, Jingdong Wang	南京理工大学	NeurIPS 2022
	Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning		
77	Zixiang Zhao, Haowen Bai, Jiangshe Zhang, Yulun Zhang, Shuang Xu, Zudi Lin, Radu Timofte, Luc Van Gool	西安交通大学	CVPR 2023
	CDDFuse: Correlation-Driven Dual-Branch Feature Decomposition for Multi-Modality Image Fusion		
78	Enyan Dai, Jie Chen	宾夕法尼亚州立大学	ICLR 2022
	Graph-Augmented Normalizing Flows for Anomaly Detection of Multiple Time Series		
79	Ziqi Zhou, Lei Qi, Xin Yang, Dong Ni, Yinghuan Shi	南京大学	CVPR 2022
	Generalizable Cross-modality Medical Image Segmentation via Style Augmentation and Dual Normalization		
80	Wei Shang, Dongwei Ren, Yi Yang, Hongzhi Zhang, Kede Ma, Wangmeng Zuo	哈尔滨工业大学	CVPR 2023
	Joint Video Multi-Frame Interpolation and Deblurring under Unknown Exposure Time		
81	Zongheng Tang, Yifan Sun, Si Liu, Yi Yang	北京航空航天大学	CVPR 2023
	DETR with Additional Global Aggregation for Cross-domain Weakly Supervised Object Detection		
82	Peng Cui, Yang Yue, Zhijie Deng, Jun Zhu	清华大学	NeurIPS 2022
	Confidence-based Reliable Learning under Dual Noises		
83	Yukang Chen, Jianhui Liu, Xiangyu Zhang, Xiaojuan Qi, Jiaya Jia	香港中文大学	CVPR 2023
	VoxelNeXt: Fully Sparse VoxelNet for 3D Object Detection and Tracking		
84	Yuanbiao Gou, Peng Hu, Jiancheng Lv, Joey Tianyi Zhou, Xi Peng	四川大学	NeurIPS 2022
	Multi-Scale Adaptive Network for Single Image Denoising		
85	Xiang An, Jiankang Deng, Kaicheng Yang, Jaiwei Li, Ziyong Feng, Jia Guo, Jing Yang, Tongliang Liu	格灵深瞳	ICLR 2023
	Unicom: Universal and Compact Representation Learning for Image Retrieval		

86	Yurui Zhu, Tianyu Wang, Xueyang Fu, Xuanyu Yang, Xin Guo, Jifeng Dai, Yu Qiao, Xiaowei Hu	中国科学技术大学	CVPR 2023
	Learning Weather-General and Weather-Specific Features for Image Restoration Under Multiple Adverse Weather Conditions		
87	Ye Zhu, Jie Yang, Si-Qi Liu, Ruimao Zhang	香港中文大学(深圳) & 深圳市大数据研究院	MIDL 2023 & PMLR
	Inherent Consistent Learning for Accurate Semi-supervised Medical Image Segmentation		
88	Juze Zhang, Ye Shi, Yuexin Ma, Lan Xu, Jingyi Yu, Jingya Wang	上海科技大学	AAAI 2023
	IKOL: Inverse kinematics optimization layer for 3D human pose and shape estimation via Gauss-Newton differentiation		
89	Man Zhou, Hu Yu, Jie Huang, Feng Zhao, Jinwei Gu, Chen Change Loy, Deyu Meng, Chongyi Li	南洋理工大学	NeurIPS 2022
	Deep Fourier Up-sampling		
90	Chun-Mei Feng, Bangjun Li, Xinxing Xu, Yong Liu, Huazhu Fu, Wangmeng Zuo	新加坡科技研究局	CVPR 2023
	Learning Federated Visual Prompt in Null Space for MRI Reconstruction		
91	Qingjie Zeng, Yutong Xie, Zilin Lu, Yong Xia	西北工业大学	CVPR 2023
	PEFAT: Boosting Semi-supervised Medical Image Classification via Pseudo-loss Estimation and Feature Adversarial Training		
92	Hao Huang, Ziyang Chen, Huanran Chen, Yongtao Wang, Kevin Zhang	北京大学	CVPR 2023
	T-SEA: Transfer-based Self-Ensemble Attack on Object Detection		
93	Tong Wei, Kai Gan	东南大学	CVPR 2023
	Towards Realistic Long-Tailed Semi-Supervised Learning: Consistency Is All You Need		
94	Yingqian Wang, Longguang Wang, Gaochang Wu, Jungang Yang, Wei An, Jingyi Yu, Yulan Guo	国防科技大学	TPAMI 2022
	Disentangling Light Fields for Super-Resolution and Disparity Estimation		
95	Jiacheng Li, Chang Chen, Wei Huang, Zhiqiang Lang, Fenglong Song, Youliang Yan, Zhiwei Xiong	中国科学技术大学	CVPR 2023
	Learning Steerable Function for Efficient Image Resampling		
96	Bingnan Yang, Mi Zhang, Zhan Zhang, Zhili Zhang,	武汉大学	CVPR 2023

	Xiangyun Hu		
	TopDiG: Class-agnostic Topological Directional Graph Extraction from Remote Sensing Images		
97	Changde Du, Kaicheng Fu, Jinpeng Li, Huiguang He	中国科学院自动化研究所	TPAMI 2023
	Decoding Visual Neural Representations by Multimodal Learning of Brain-Visual-Linguistic Features		
98	Yinpeng Dong, Shouwei Ruan, Hang Su, Caixin Kang, Xingxing Wei, Jun Zhu	北京航空航天大学	NeurIPS 2022
	《Viewfool: Evaluating the robustness of visual recognition to adversarial viewpoints》		
99	Jintao Guo, Na Wang, Lei Qi, Yinghuan Shi	南京大学	CVPR 2023
	ALOFT: A Lightweight MLP-like Architecture with Dynamic Low-frequency Transform for Domain Generalization		
100	Yinpeng Dong, Caixin Kang, Jinlai Zhang, Zijian Zhu, Yikai Wang, Xiao Yang, Hang Su, Xingxing Wei, Jun Zhu	北京航空航天大学	CVPR 2023
	Benchmarking Robustness of 3D Object Detection to Common Corruptions in Autonomous Driving		
101	Jianfeng Dong, Xianke Chen, Minsong Zhang, Xun Yang, Shujie Chen, Xirong Li, Xun Wang	浙江工商大学	ACM MM 2022
	Partially Relevant Video Retrieval		
102	Xueyang Fu, Jie Xiao, Yurui Zhu, Aiping Liu, Feng Wu, Zheng-Jun Zha	中国科学技术大学	TPAMI 2023
	Continual Image Deraining with Hypergraph Convolutional Networks		
103	Xiaojun Jia, Yong Zhang, Baoyuan Wu, Ke Ma, Jue Wang, Xiaochun Cao	中国科学院信息工程研究所	CVPR 2022
	LAS-AT: Adversarial Training with Learnable Attack Strategy		
104	Kangkan Wang, Guofeng Zhang, Suxu Cong, Jian Yang	南京理工大学	CVPR 2023
	Clothed Human Performance Capture with a Double-layer Neural Radiance Fields		
105	Jie Xiao, Xueyang Fu, Feng Wu, Zheng-Jun Zha	中国科学技术大学	NeurIPS 2022
	Stochastic Window Transformer for Image Restoration		
106	Zeyang Zhang, Xin Wang, Ziwei Zhang, Haoyang Li, Zhou Qin, Wenwu Zhu	清华大学	NeurIPS 2022
	Dynamic Graph Neural Networks Under Spatio-Temporal Distribution Shift		

107	Xing Wei, Yifan Bai, Yongchao Zheng, Dahu Shi, Yihong Gong	西安交通大学	CVPR 2023
	Autoregressive Visual Tracking		
108	Zhe Chen, Yuchen Duan, Wenhai Wang, Junjun He, Tong Lu, Jifeng Dai, Yu Qiao	南京大学	ICLR 2023
	Vision Transformer Adapter for Dense Predictions		
109	Yijian Qin, Xin Wang, Ziwei Zhang, Pengtao Xie, Wenwu Zhu	清华大学	ICML 2022
	Graph Neural Architecture Search Under Distribution Shifts		
110	Jiangjun Peng, Yao Wang, Hongying Zhang, Jianjun Wang, Deyu Meng	西安交通大学	TPAMI 2022
	Exact Decomposition of Joint Low Rankness and Local Smoothness Plus Sparse Matrices		
111	Yucheng Shi, Yahong Han, Yu-an Tan, Xiaohui Kuang	天津大学	NeurIPS 2022
	Decision-based Black-box Attack Against Vision Transformers via Patch-wise Adversarial Removal		
112	Bowen Zhao, Chen Chen, Shu-Tao Xia	清华大学	ICLR 2023
	DELTA: degradation-free fully test-time adaptation		
113	Lihe Yang, Lei Qi, Litong Feng, Wayne Zhang, Yinghuan Shi	南京大学	CVPR 2023
	Revisiting Weak-to-Strong Consistency in Semi-Supervised Semantic Segmentation		
114	Ren Yang, Radu Timofte, Luc Van Gool	商汤科技	IJCAI 2022
	Perceptual Learned Video Compression with Recurrent Conditional GAN		
115	Xiyao Liu, Ziping Ma, Junxing Ma, Jian Zhang, Gerald Schaefer, Hui Fang	中南大学	CVPR 2022
	Image disentanglement autoencoder for steganography without embedding		
116	Jie Huang, Yajing Liu, Feng Zhao, Keyu Yan, Jinghao Zhang, Yukun Huang, Man Zhou, Zhiwei Xiong	中国科学技术大学	ECCV 2022
	Deep Fourier-based Exposure Correction Network with Spatial-Frequency Interaction		
117	Qizhou Wang, Feng Liu, Yonggang Zhang, Jing Zhang, Chen Gong, Tongliang Liu, Bo Han	香港浸会大学	NeurIPS 2022
	Watermarking for Out-of-distribution Detection		

118	Minting Pan, Xiangming Zhu, Yunbo Wang, Xiaokang Yang	上海交通大学	NeurIPS 2022
	Iso-Dream: Isolating and Leveraging Noncontrollable Visual Dynamics in World Models		
119	Aihua Zheng, Peng Pan, Hongchao Li, Chenglong Li, Bin Luo, Chang Tan, Ruoran Jia	安徽大学	ACM MM 2022
	Progressive Attribute Embedding for Accurate Cross-modality Person Re-ID		
120	Zeyu Gao; Anyu Mao; Kefei Wu; Yang Li; Liebin Zhao; Xianli Zhang; Jialun Wu; Lisha Yu; Chao Xing, Tieliang Gong, Yefeng Zheng, Deyu Meng, Min Zhou, Chen Li	西安交通大学	TMI 2023
	Childhood Leukemia Classification via Information Bottleneck Enhanced Hierarchical Multi-Instance Learning		
121	Shenyuan Gao, Chunluan Zhou, Jun Zhang	香港科技大学	CVPR 2023
	Generalized Relation Modeling for Transformer Tracking		
122	Ye Du, Yujun Shen, Haochen Wang, Jingjing Fei, Wei Li, Liwei Wu, Rui Zhao, Zehua Fu, Qingjie Liu	北京航空航天大学	NeurIPS 2022
	Learning from Future: A Novel Self-Training Framework for Semantic Segmentation		
123	ZhongCheng Wu, TingZhu Huang, LiangJian Deng, HongXia Dou, Deyu Meng	电子科技大学	NeurIPS 2022
	Tensor Wheel Decomposition and Its Tensor Completion Application		
124	Yinhuai Wang, Jiwen Yu, Jian Zhang	北京大学	ICLR 2023
	Zero-Shot Image Restoration Using Denoising Diffusion Null-Space Model		
125	Zhi Gao, Yuwei Wu, Xiaomeng Fan, Mehrtash Harandi, Yunde Jia	北京理工大学	TPAMI 2022
	Learning to Optimize on Riemannian Manifolds		
126	Yanmin Wu, Xinhua Cheng, Renrui Zhang, Zesen Cheng, Jian Zhang	北京大学	CVPR 2023
	EDA: Explicit Text-Decoupling and Dense Alignment for 3D Visual Grounding		
127	Runyu Ding, Jihan Yang, Chuhui Xue, Wenqing Zhang, Song Bai, Xiaojuan Qi	香港大学	CVPR 2023
	PLA: Language-Driven Open-Vocabulary 3D Scene Understanding		
128	Fengyin Lin, Mingkang Li, Da Li, Timothy Hospedales, Yizhe Song, Yonggang Qi	北京邮电大学	CVPR 2023

	Zero-Shot Everything Sketch-Based Image Retrieval, and in Explainable Style		
129	Qiang Wang, Haoge Deng, Yonggang Qi, Da Li, Yi-Zhe Song	北京邮电大学	ICLR 2023
	SketchKnitter: Vectorized Sketch Generation with Diffusion Models		
130	Zhiqiang Shen, Xiaoxiao Sheng, Longguang Wang, Yulan Guo, Qiong Liu, Xi Zhou	上海交通大学	CVPR 2023
	PointCMP: Contrastive Mask Prediction for Self-supervised Learning on Point Cloud Videos		
131	Yang Li, Yichuan Mo, Liangliang Shi, Junchi Yan, Xiaolu Zhang, Jun Zhou	上海交通大学	NeurIPS 2022
	Improving Generative Adversarial Networks via Adversarial Learning in Latent Space		
132	Weiqing Min, Zhiling Wang, Yuxin Liu, Mengjiang Luo, Liping Kang, Xiaoming Wei, Xiaolin Wei, Shuqiang Jiang	中国科学院计算技术研 究所	TPAMI 2023
	Large Scale Visual Food Recognition		
133	Yue Wang, Jinlong Peng, Jiangning Zhang, Ran Yi, Yabiao Wang, Chengjie Wang	上海交通大学	CVPR 2023
	Multimodal Industrial Anomaly Detection via Hybrid Fusion		
134	Xingxing Wei, Ying Guo, Jie Yu, Bo Zhang	美团	TPAMI 2022
	Simultaneously Optimizing Perturbations and Positions for Black-box Adversarial Patch Attacks		
135	Xiangyang Li, Zihan Wang, Jiahao Yang, Yaowei Wang, Shuqiang Jiang	中国科学院计算技术研 究所	CVPR 2023
	KERM: Knowledge Enhanced Reasoning for Vision-and-Language Navigation		
136	Kai Chen, Zhili Liu, Lanqing Hong, Hang Xu, Zhenguo Li, Dit-Yan Yeung	香港科技大学	CVPR 2023
	Mixed Autoencoder for Self-supervised Visual Representation Learning		
137	Kaican Li, Kai Chen, Haoyu Wang, Lanqing Hong, Chaoqiang Ye, Jianhua Han, Yukuai Chen, Wei Zhang, Chunjing Xu, Dit-Yan Yeung, Xiaodan Liang, Zhenguo Li, Hang Xu	华为诺亚方舟实验室	ECCV 2022
	CODA: A Real-World Road Corner Case Dataset for Object Detection in Autonomous Driving		
138	Zhe Chen, Xiao-Jun Wu, Tianyang Xu, Josef Kittler	江南大学	TIP 2023

	Fast Self-guided Multi-view Subspace Clustering		
139	Sanyi Zhang, Xiaochun Cao, Guo-Jun Qi, Zhanjie Song, Jie Zhou	中国科学院信息工程研 究所	TIP 2022
	AIParsing: Anchor-Free Instance-Level Human Parsing		
140	Chenchen Jing, Yunde Jia, Yuwei Wu, Xinyu Liu, Qi Wu	北京理工大学	CVPR 2022
	Maintaining Reasoning Consistency in Compositional Visual Question Answering		
141	Qixin Hu, Yixiong Chen, Junfei Xiao, Shuwen Sun, Jieneng Chen, Alan Yuille, Zongwei Zhou	华中科技大学	CVPR 2023
	Label-Free Liver Tumor Segmentation		
142	Jinkun Cao, Ruiqian Nai, Qing Yang, Jialei Huang, Yang Gao	清华大学	NeurIPS 2022
	An Empirical Study on Disentanglement of Negative-free Contrastive Learning		
143	Zhiling Wang, Weiqing Min, Zhuo Li, Liping Kang, Xiaoming Wei, Xiaolin Wei, Shuqiang Jiang	美团	TIP 2022
	Ingredient-Guided Region Discovery and Relationship Modeling for Food Category-Ingredient Prediction		
144	Sheng Xu, Yanjing Li, Mingbao Lin, Peng Gao, Guodong Guo, Jinhu Lu, Baochang Zhang	北京航空航天大学	CVPR 2023
	Q-DETR: An Efficient Low-Bit Quantized Detection Transformer		
145	Dengsheng Chen, Jie Hu, Wenwen Qiang, Xiaoming Wei, Enhua Wu	美团	ICLR 2023
	Rethinking skip connection model as a learnable Markov chain		
146	Qing Ma, Junjun Jiang, Xianming Liu, Jiayi Ma	哈尔滨工业大学	TIP 2022
	Multi-Task Interaction Learning for Spatsiospectral Image Super-Resolution		
147	Jingyuan Xu, Weiwei Liu	武汉大学	NeurIPS 2022
	On Robust Multiclass Learnability		
148	Sixian Zhang, Xinhang Song, Weijie Li, Yubing Bai, Xinyao Yu, Shuqiang Jiang	中国科学院计算技术研 究所	CVPR 2023
	Layout-based Causal Inference for Object Navigation		
149	Lue Fan, Feng Wang, Naiyan Wang, Zhaoxiang Zhang	中国科学院自动化研究 所	NeurIPS 2022
	Fully Sparse 3D Object Detection		

150	Xin Jin, Ling-Hao Han, Zhen Li, Chun-Le Guo, Zhi Chai, Chongyi Li	南开大学	CVPR 2023
	DNF: Decouple and Feedback Network for Seeing in the Dark		
151	Yue Li, Jiayong Peng, Juntian Ye, Yueyi Zhang, Feihu Xu, Zhiwei Xiong	中国科学技术大学	CVPR 2023
	NLOST: Non-Line-of-Sight Imaging with Transformer		
152	Zhili Liu, Kai Chen, Jianhua Han, Lanqing HONG, Hang Xu, Zhenguo Li, James Kwok	华为诺亚方舟实验室	ICLR 2023
	Task-customized Masked Autoencoder via Mixture of Cluster-conditional Experts		
153	Yuanyuan Liu, Wei Dai, Chuanxu Feng, Wenbin Wang, Guanghao Yin, Jiabei Zeng, Shiguang Shan	中国地质大学	ACM MM 2022
	MAFW: A Large-scale, Multi-modal, Compound Affective Database for Dynamic Facial Expression Recognition in the Wild		
154	Guoli Jia, Jufeng Yang	南开大学	ECCV 2022
	S2-VER: Semi-Supervised Visual Emotion Recognition		
155	Andong Lu, Zhang Zhang, Yan Huang, Yifan Zhang, Chenglong Li, Jin Tang, Liang Wang	安徽大学	ECCV 2023
	Illumination Distillation Framework for Nighttime Person Re-Identification and A New Benchmark		
156	Zijian Zhu, Yichi Zhang, Hai Chen, Yinpeng Dong, Shu Zhao, Wenbo Ding, Jiachen Zhong, Shibao Zheng	上海交通大学	CVPR 2023
	Understanding the Robustness of 3D Object Detection with Bird's-Eye-View Representations in Autonomous Driving		
157	Kunyu Wang, Xueyang Fu, Yukun Huang, Chengzhi Cao, Gege Shi, Zheng-Jun Zha	中国科学技术大学	CVPR 2023
	Generalized UAV Object Detection via Frequency Domain Disentanglement		
158	Shengzhe Liu, Xin Zhang, Jufeng Yang	南开大学	ACM MM 2022
	SER30K: A Large-Scale Dataset for Sticker Emotion Recognition		
159	Yao Xiao, Ziyi Tang, Pengxu Wei, Cong Liu, Liang Lin	中山大学	CVPR 2023
	Masked Images Are Counterfactual Samples for Robust Fine-tuning		
160	Man Liu, Feng Li, Chunjie Zhang, Yunchao Wei, Huihui Bai, Yao Zhao	北京交通大学	CVPR 2023

	Progressive Semantic-Visual Mutual Adaption for Generalized Zero-Shot Learning		
161	ChuanXian Ren, YouWei Luo, DaoQing Dai	中山大学	TPAMI 2023
	BuresNet: Conditional Bures Metric for Transferable Representation Learning		
162	Guodong Wang, Yunhong Wang, Jie Qin, Dongming Zhang, Xiuguo Bao, Di Huang	北京航空航天大学	ECCV 2022
	Video Anomaly Detection by Solving Decoupled Spatio-Temporal Jigsaw Puzzles		
163	Hui Li, Tianyang Xu, Xiao-Jun Wu, Jiwen Lu, Josef Kittler	江南大学	TPAMI 2023
	LRRNet: A novel representation learning guided fusion framework for infrared and visible images		
164	Huanyu Zhou, Qingjie Liu, Yunhong Wang	北京航空航天大学	CVPR 2023
	Learning Discriminative Representations for Skeleton Based Action Recognition		
165	Tinglei Feng, Yingjie Zhai, Jufeng Yang, Jie Liang, Deng-Ping Fan, Jing Zhang, Ling Shao, Dacheng Tao	南开大学	TPAMI 2022
	IC9600: A Benchmark Dataset for Automatic Image Complexity Assessment		
166	Xiaorong Qin, Xinhang Song, Shuqiang Jiang	中国科学院计算技术研究所	CVPR 2023
	Bi-level Meta-learning for few-shot domain generalization		
167	RuiQi Wu, ZhengPeng Duan, ChunLe Guo, Zhi Chai, ChongYi Li	南开大学	CVPR 2023
	RIDCP: Revitalizing Real Image Dehazing via High-Quality Codebook Priors		
168	Wenzheng Zeng, Yang Xiao, Sicheng Wei, Jinfang Gan, Xintao Zhang, Zhiguo Cao, Zhiwen Fang, Joey Tianyi Zhou	华中科技大学	CVPR 2023
	Real-time Multi-person Eyeblick Detection in the Wild for Untrimmed Video		
169	Lianyu Hu, Liqing Gao, Zekang liu, Wei Feng	天津大学	AAAI 2023
	Self-Emphasizing Network for Continuous Sign Language Recognition		
170	Zhijie Shen, Zishuo Zheng, Chunyu Lin, Lang Nie, Kang Liao, Shuai Zheng, Yao Zhao	北京交通大学	CVPR 2023
	Disentangling Orthogonal Planes for Indoor Panoramic Room Layout Estimation with Cross-Scale Distortion Awareness		
171	Jiechong Song, Chong Mou, Shiqi Wang, Siwei Ma, Jian Zhang	北京大学	CVPR 2023

	Optimization-Inspired Cross-Attention Transformer for Compressive Sensing		
172	Chong Mou, Yanze Wu, Xintao Wang, Chao Dong, Jian Zhang, Ying Shan	北京大学	ECCV 2022
	Metric Learning Based Interactive Modulation for Real-World Super-Resolution		
173	Zhengzheng Tu, Zhun Li, Chenglong Li, Jin Tang	安徽大学	TIP 2022
	Weakly Alignment-Free RGBT Salient Object Detection With Deep Correlation Network		
174	Zi Wang, Chenglong Li, Aihua Zheng, Ran He, Jin Tang	安徽大学	AAAI 2022
	Interact, Embed and Enlarge: Boosting Modality-specific Representations for Multi-Modal Person Re-identification		
175	Yixin Yang, Jin Han, Jinxiu Liang, Imari Sato, Boxin Shi	北京大学	CVPR 2023
	Learning Event Guided High Dynamic Range Video Reconstruction		
176	Tinglei Feng, Jiaxuan Liu, Jufeng Yang	南开大学	CVPR 2023
	Probing Sentiment-Oriented Pre-Training Inspired by Human Sentiment Perception Mechanism		
177	Hanyue Lou, Minggui Teng, Yixin Yang, Boxin Shi	北京大学	CVPR 2023
	All-in-focus Imaging from Event Focal Stack		
178	Zhigang Dai, Bolun Ca, Yugeng Lin, Junying Chen	华南理工大学	TPAMI 2022
	Unsupervised Pre-Training for Detection Transformers		
179	Ding Jia, Yuhui Yuan, Haodi He, Xiaopei Wu, Haojun Yu, Weihong Lin, Lei Sun, Chao Zhang, Han Hu4	北京大学	CVPR 2023
	DETRs with Hybrid Matching		
180	Yanbin Hao, Hao Zhang, Chong-Wah Ngo, Xiangnan He	中国科学技术大学	CVPR 2022
	Group Contextualization for Video Recognition		
181	Shixiang Tang, Cheng Chen, Qingsong Xie, Meilin Chen, Yizhou Wang, Yuanzheng Ci, Lei Bai, Feng Zhu, Haiyang Yang, Li Yi, Rui Zhao, Wanli Ouyang	浙江大学	CVPR 2023
	HumanBench: Towards General Human-centric Perception with Projector Assisted Pretraining		
182	Yuanzheng Ci, Yizhou Wang, Meilin Chen, Shixiang Tang, Lei Bai, Feng Zhu, Rui Zhao, Fengwei Yu, Donglian Qi, Wanli	上海人工智能实验室	CVPR 2023

	Ouyang		
	UniHCP: A Unified Model for Human-Centric Perceptions		
183	Haiyang Yang, Xiaotong Li, Shixiang Tang, Feng Zhu, Yizhou Wang, Meilin Chen, Lei Bai, Rui Zhao, Wanli Ouyang	北京大学	ICLR 2023
	Cycle-consistent Masked AutoEncoder for Unsupervised Domain Generalization		
184	Xingxuan Zhang, Yue He, Renzhe Xu, Han Yu, Zheyang Shen, Peng Cui	清华大学	CVPR 2023
	NICO++: Towards Better Bechmarks for Out-of-Distribution Generalization		
185	Xinyi Ying, Li Liu, Yingqian Wang, Ruoqing Li, Nuo Chen, Zaiping Lin, Weidong Sheng, Shilin Zhou	国防科技大学	CVPR 2023
	Mapping Degeneration Meets Label Evolution: Learning Infrared Small Target Detection with Single Point Supervision		

## 合作单位简介

# oppo

铂金合作单位

OPPO

OPPO 是全球五大智能手机品牌之一，也是全球领先的智能设备制造商和创新者。作为技术驱动型的科技公司，OPPO 在全球建立了 9 大智能制造中心、6 大研究所、5 大研发中心，以人工智能、云服务、大数据等前沿技术驱动、软件产品和互联网服务的开发，引领 5G、AI、影像处理、新材料新工艺等技术在智能终端上的发展应用。万物互融时代，OPPO 坚持“3+N+X”的科技跃迁战略。其中“3”指的是三项基础技术，包括硬件基础技术、软件基础技术和服务基础技术；“N”指的是能力中心，包括互联互通、多媒体、人工智能和安全隐私等；“X”指的是差异化技术，其中包含了 OPPO 的强项快充、影像、新形态和 AR 增强现实技术等。

截至 2021 年 3 月，OPPO 全球专利申请数量超过 61,000 件，专利授权数量超过 26,000 件，并仍在持续增长。据世界知识产权组织(WIPO)发布 2020 年国际专利条约(PCT)申请数量排行榜，OPPO 全球排名前十。

专利搭载在产品上，就变成了功能与用户体验。除了持续引领 SUPERVOOC 闪充生态，推进 5G 普及之外，OPPO 近期迭代了一系列凝结科技的功能：凭借更准确的语义分析与用户实现情感交互冰冷的小布助手，带动机器学习与语音语义突破；借助全链路色彩管理系统、感知人像与画质增强引擎、视频防抖，背后是计算机视觉与多媒体等技术在硬件及软件上的创新。

OPPO 在智能终端布局日趋完善，以高端旗舰 Find X3 系列手机、Reno5 系列新品手机等智能手机为中心，持续完善 IoT 生态场景，推出了智能电视、OPPO Watch、OPPO 手环、Enco X 真无线降噪耳机等产品，为用户提供智能美好的完整体验。在前沿科技层面，OPPO 秉承对美的一贯追求。OPPO X 2021 卷轴屏概念机是 OPPO 手机形态探索的最新成果，屏幕如画卷般伸展，呈现几乎“零折痕”的屏幕效果。OPPO AR Glass 2021 则结合了全时空间计划 AR 应用，探索“虚实融合”数字世界进一步升级。

# tu 图森未来

铂金合作单位

图森未来

图森未来于 2015 年 9 月成立，在中国、美国设有研发中心，员工超过 1000 人。2021 年 2 月，图森未来完成超过 4.15 亿美元的 E 轮融资。4 月，图森未来正式登陆纳斯达克挂牌上市，成为全球无人驾驶第一股。

图森未来自主研发的以摄像头为主要传感器，融合激光雷达、毫米波雷达的 L4 级无人驾驶卡车解决方案，具备感知、定位、决策、控制等无人驾驶核心功能，能够实现货运卡车在干线物流场景和半封闭枢纽场景下的全无人驾驶。

目前，图森未来已率先在中国、美国和欧洲多地展开测试和商业化试运营，持续获得主营收入。在中国，获得首张无人驾驶卡车公开道路测试牌照和示范应用资格，并在东海大桥开展应

用于洋山港和芦潮港中心站等物流场景的多式联运。在美国，图森未来已经开始向包括 UPS、McLane 在内的 18 家客户提供无人驾驶物流服务。在欧洲，于瑞典南泰利耶（Södertälje）和延雪平（Jönköping）之间的 E4 高速公路上进行载货无人驾驶卡车测试。



## 铂金合作单位

AutoDL

视拓云团队来自于山世光老师创建的中科视拓，专注云计算和 AI 开发者社区两个细分领域，面向“大 AI 圈”内的科研工作者运营 AutoDL.com 和 CodeWithGPU.com 两个产品。AutoDL 提供弹性、好用、省钱的 GPU 算力；CodeWithGPU 提供算法复现服务和内容交流社区。有算法就有复现，能复现才是好算法，复现算法就上 CodeWithGPU.com。



## 铂金合作单位

华为

华为是全球领先的 ICT 基础设施和智能终端提供商，致力于把数字世界带入每个人、每个家庭、每个组织，构建万物互联的智能世界。我们在通信网络、IT、智能终端和云服务等领域为客户提供有竞争力、安全可信的产品、解决方案与服务，持续为客户创造价值。

华为云 EI 是企业智能的使能者，通过云服务的方式（公有云、专属云等模式），提供一个开放、可信、智能的平台，结合产业场景，使能企业应用系统能看、能听、能说，让更多的企业便捷地使用 AI 和大数据服务，加速业务发展，造福社会。

华为消费者 BG AI 技术应用部是华为面向全场景智慧硬件的 AI 应用技术研发和能力中心。聚焦基于 1+8 硬件的计算机视觉、听觉、多传感器融合感知和识别技术、算力提升和数据分析预测技术，致力于面向消费者全场景打造 1+8 产品的硬件智慧体验。

中央媒体技术院是华为公司媒体技术创新与工程能力中心，肩负着公司手机拍照、ARVR、视频、音频及音视频标准、媒体体验与测评等领域的技术研究、创新和突破任务，确保华为公司媒体产品技术竞争力业界持续领先。

诺亚方舟实验室是华为的人工智能研究中心，立足于人工智能基础算法研究，聚焦打造数据高效和能耗高效的 AI 引擎，推动计算机视觉、语音和自然语言处理、决策推理等 AI 领域发展，助力华为公司主航道业务 AI 使能。



## 金牌合作单位

百度

百度是拥有强大互联网基础的领先 AI 公司。是全球为数不多的提供 AI 芯片、软件架构和应用程序等全栈 AI 技术的公司之一，被国际机构评为全球四大 AI 公司之一。百度以“用科技让复杂的世界更简单”为使命，坚持技术创新，致力于“成为最懂用户，并能帮助人们成长的全球顶级高科技公司”。

百度公司 2000 年 1 月 1 日创立于中关村，创始人李彦宏拥有“超链分析”技术专利，也使中国成为美国、俄罗斯、和韩国之外，全球仅有的 4 个拥有搜索引擎核心技术的国家之一。百度每天响应来自 100 余个国家和地区的数十亿次搜索请求，是网民获取中文信息和服务的最主要入口，服务 10 亿互联网用户。

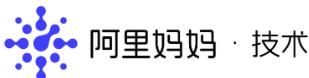
基于搜索引擎，百度演化出语音、图像、知识图谱、自然语言处理等人工智能技术；最近 10 年，百度在深度学习、对话式人工智能操作系统、自动驾驶、AI 芯片等前沿领域投资，使得百度成为一个拥有强大互联网基础的领先 AI 公司。



金牌合作单位

字节跳动

字节跳动成立于 2012 年 3 月，公司使命为 “Inspire Creativity, Enrich Life（激发创造，丰富生活）”。公司业务覆盖 150 个国家和地区，拥有 11 万名员工。字节跳动在全球推出了多款有影响力的产品，包括今日头条、抖音、西瓜视频、飞书、TikTok、Lark、Helo 等。截至 2021 年 6 月，字节跳动旗下产品全球月活跃用户数超过 19 亿。



金牌合作单位

阿里妈妈

阿里妈妈成立于 2007 年，是淘天集团商业数智营销中台，管理并运营淘天集团搜索、展示、信息流等多种消费场景下的营销产品及技术解决方案、覆盖互联网主流 APP 全域营销资源。

阿里妈妈拥有淘天集团独家核心商业数智能力，致力于深研 AI 前沿技术，引领了 AI 在互联网营销领域的探索和大规模应用，并通过技术创新驱动淘天集团数智营销业务高速增长。

2021 年阿里妈妈发布全新的品牌主张“让每一份经营都算数”及全新的经营方法论-“深链经营”，全面实现从全域营销到全域经营的升级。



金牌合作单位

合合信息

合合信息是一家人工智能及大数据科技企业，基于自主研发的领先的智能文字识别及商业大数据核心技术，为全球 C 端用户和多元行业 B 端客户提供数字化、智能化的产品及服务。

公司 C 端业务主要为面向全球个人用户的 APP 产品，包括扫描全能王（智能文字扫描及识别 APP）、名片全能王（智能名片及人脉管理 APP）、启信宝（企业商业信息查询 APP）3 款核心产品；公司 B 端业务为面向企业客户提供以智能文字识别、商业大数据为核心的服务，形成了包括基础技术服务、标准化服务和场景化解决方案的业务矩阵，满足客户降本增

效、风险管理、智能营销等多元需求，助力客户实现数字化与智能化的转型升级。凭借领先的自主研发技术、成熟的产品落地能力、优质的用户体验及服务质量，公司的C端产品覆盖了全球百余个国家和地区的亿级用户，B端服务覆盖了近30个行业的企业客户。在B端业务方面，公司智能文字识别与商业大数据服务已覆盖了银行、证券、保险、政府、物流、制造、地产、零售等近30个行业的众多头部客户，《财富》杂志2020年发布的世界500强公司名单中，公司客户已覆盖超过80家。



## 金牌合作单位 蚂蚁技术研究院

随着数字化转型的深入和科技引领创新成为趋势，蚂蚁集团持续投入前沿技术，布局隐私计算、区块链、图计算、分布式数据库和绿色计算五大数字化“根技术”。为进一步强化和完善蚂蚁技术探索和创新机制，提升技术布局的前瞻性，体现蚂蚁在前沿硬核技术领域全力以赴、攻坚克难的决心，于2021年成立蚂蚁技术研究院。根据公开信息，蚂蚁集团连续3年科研投入增长率超39%，目前在职工工技术人员占比达63%，在全国工商联发布的“2022民营企业研发投入”榜单中排名全国第六。

蚂蚁技术研究院致力做有用、有想象力的科研，开展前瞻科技探索，支撑蚂蚁科技发展，为数字化和智能化社会提供科技底座。下设六大实验室，分别为交互智能实验室、图计算实验室、密码学实验室、计算系统实验室、程序设计和编译器实验室和数据库实验室。



## 金牌合作单位 极视角科技

极市(Extreme Mar)是极视角科技旗下的AI开发者生态，为开发者提供一站式线上便捷算法开发平台，同时提供大咖技术分享及直播、社区交流与线下沙龙、以及一系列的算法竞赛等丰富的内容与服务。

极市开发者生态自2015年起，迄今已经积累超240,000名海内外专业算法开发者，影响力覆盖300,000+AI从业者/学生群体，极市希望与开发者们一起打造计算机视觉行业的生态圈，携手用算法改变世界。



## 金牌合作单位 奥比中光

奥比中光是行业领先的3D视觉感知整体技术方案提供商。公司已构建起“全栈式技术研发能力+全领域技术路线布局”的3D视觉感知技术体系，在技术纵向上对包括深度引擎芯片、感光芯片、专用光学系统等在内的核心底层技术以及SDK、行业应用算法等全链路技术进行全栈式自主研发，在技术横向上对结构光、iToF、双目、dToF、Lidar、工业三维测量进行全领域布局。

以“让所有终端都能看懂世界”为使命，奥比中光坚持“3D 视觉传感器+自主底层核心技术+完整应用方案”的市场定位，致力于将 3D 视觉感知技术应用于“衣、食、住、行、工、娱、医”等领域；在智能手机、线下零售等市场已率先突破百万级出货量，在生物识别、消费电子、AIoT、工业三维测量等市场实现多项商业应用，服务全球超过 1000 家客户及众多开发者。

奥比中光总部位于深圳，在上海、西安、美国密歇根设有分支机构。截至 2021 年，奥比中光累计申请专利超 1100 件，承担科技部国家重点研发计划项目“面向服务机器人的三维视觉传感器研发及产业化应用”、“3D 视觉感知广东省新一代人工智能开放创新平台”等多项国家级、省级重大科研项目，荣获“吴文俊人工智能科技进步奖”、“中国专利优秀奖”、“中国芯”优秀技术创新产品、“广东省科学技术奖科技进步奖一等奖”、“深圳市科技进步奖技术开发类一等奖”等重要奖励。

ZhenFund  
真格基金

金牌合作单位  
真格基金

真格基金是由徐小平、王强先生于 2011 年联合红杉资本中国基金创立的早期投资机构，累计管理资金总规模超 10 亿美元。真格基金自创立伊始，一直积极在互联网、移动互联网、未来科技、人工智能、企业服务、医疗健康、大消费、教育等领域寻找最优秀的创业团队和引领时代的投资机会。

真格基金陆续投资了 700 余家创业公司，从早期陪伴了兴盛优选、美菜、小红书、Nuro、VIPKID、依图科技、英雄互娱、地平线、Momenta、找钢网、罗辑思维、禾赛科技、出门问问、蜜芽、晶泰科技等公司成长为独角兽企业。自 2011 年起，真格基金被投公司世纪佳缘(NASDAQ: DATE)、聚美优品(NYSE: JMEI)、兰亭集势(NYSE: LITB)、51Talk(NYSE: COE)、牛电科技(NASDAQ: NIU)、老虎证券(NASDAQ: TIGR)、亿航智能(NASDAQ: EH)、逸仙电商(NYSE: YSG)、优客工场(NASDAQ: UK)、水滴(NYSE: WDH)等陆续上市，超过 50 个项目通过多种方式实现退出，获得投资回报。自 2014 年清科“中国股权投资年度排名”设立早期投资机构排名以来，真格基金已连续 6 年获得“中国早期投资机构 30 强”第 1 名；真格基金创始人徐小平从 2016 年起连续五年入选福布斯“全球最佳创投人榜单( the Midas List )”，在 2019 年榜单中排名第 11 位。

真格基金总部位于北京，并陆续布局上海和深圳。“寻找真正优秀的创业者，缔造引领科技创新并改变世界的伟大公司”是徐小平和王强先生创立真格的初衷。真格基金具有的诚信正直、友善谦卑、亲和温暖、专业效率与理想主义并重的文化价值观得到了海内外创业者的认可和赞赏。

做创业者的垫脚石，真格正在不懈努力。

 地平线  
Horizon Robotics

金牌合作单位  
地平线

地平线是边缘人工智能计算平台的全球领导者。得益于前瞻性的软硬结合理念，地平线自主研发兼具极致效能与开放易用性的边缘人工智能芯片及解决方案，可面向智能驾驶以及更广泛的通用 AI 应用领域提供全面开放的赋能服务。目前，地平线是国内唯一一家实现汽车智能芯片前装量产的企业。



## 银牌合作单位 金山办公

北京金山办公软件股份有限公司（688111.SH）（以下简称“金山办公”），是中国领先的办公软件产品和服务提供商。作为一家源自中国的科技公司，秉持“让智慧绽放”的品牌理念，金山办公在过去 34 年持续深耕办公赛道，不断打磨技术和产品服务，始终致力于把最简单高效的办公体验带给众多个人、家庭和组织，帮助个人用户更轻松快乐的创作和生活，帮助组织客户更高效安全的运行与发展。

凭借以 WPS、金山文档、稻壳儿等为代表的办公产品和服务，金山办公为来自全球 220 多个国家和地区的用户提供办公服务，截至 2021 年 12 月，公司主要产品月度活跃设备数为 5.44 亿，其中 WPS office PC 版月度活跃设备数 2.19 亿，移动版月度活跃设备数 3.21 亿，持续领先其他国产办公软件。



## 银牌合作单位 融科联创

融科联创（天津）信息技术有限公司，总部位于天津市武清区京津科技谷，是一家集服务器研发、生产、营销为一体的高新技术企业。公司建有专业的服务器定制化生产基地，可实现年产量 20 万台，并在北京、成都、杭州、长沙、深圳、武汉、西安、新加坡等地设立分支业务机构。

目前，融科联创业务已覆盖人工智能、云计算、互联网、教育、科研院所、广电、政府、工业企业、物联网等众多行业和应用领域。先后在云计算数据中心、高校 AI 教育、冷冻电镜生物分子研究、基因测序、广播电视录播系统、铁路智能安检、人脸识别、高性能计算平台、自动驾驶、无人配送车等行业应用方向成功落地。

多年来，融科联创获得了众多自主知识产权和荣誉资质。包括国家高新技术企业、天津市雏鹰企业、天津市瞪羚企业、天津市创新型中小企业、ISO 三体系认证、中国强制性产品 3C 认证，以及 80 余项专利和软件著作权认证等。

同时，融科联创还与超微（Supermicro STAP 成员）、英特尔（Intel 钛金级合作伙伴）、英伟达（NVIDIA 合作伙伴）等国际行业巨头紧密合作，整合产业优势资源，为客户提供优质的产品和解决方案，致力于成为人工智能、云计算、大数据、物联网领域卓越的解决方案提供商！



## 银牌合作单位 思腾合力

思腾合力（SITONHOLY）成立于 2009 年，总部位于天津滨海新区逸仙科学工业园，是 AI 服务器与 HPC 基础架构解决方案商，作为 NVIDIA 精英级别合作伙伴，专注于人工智能领域，提供云计算、AI 服务器、AI 工作站、系统集成、产品定制、软件开发、边缘计算等产品和整体解决方案，拥有自主品牌 AI 服务器及通用 X86 服务器，适用于深度学习训练及推理

等场景，覆盖服务器、静音工作站等多种产品形态，已经打造出了一套完全自主的软硬件结合的产品生态，致力于成为行业领先的人工智能基础架构解决方案商。

思腾合力拥有完善的研发、生产、制造基地，2021年收购包头市易慧信息科技有限公司开启云计算业务，形成以天津为生产及研发基地，北京为营销中心，南京、深圳、成都、武汉、西安、内蒙古覆盖全国主要区域的营销和售后服务机构，为更全面地服务客户提供了有力保障。

思腾合力成立十多年来深耕教育及科研行业，从业 AI 领域研究、高性能计算的重点高校百分之八十都采用了思腾产品及解决方案，为各专业的科学实验研究提供了完备的 AI 加速解决方案。目前合作客户包括清华大学、北京大学、北京理工大学、中科院计算所、中科院自动化所、中科院半导体所、中科院信息工程所，以及国内知名人工智能公司等各企事业单位。



银牌合作单位

美团

美团的使命是“帮大家吃得更好，生活更好”。作为一家生活服务电子商务平台，公司聚焦“Food + Platform”战略，以“吃”为核心，通过科技创新，和广大商户与各类合作伙伴一起，努力为消费者提供品质生活，推动生活服务业需求侧和供给侧数字化升级。

2018年9月20日，美团正式在港交所挂牌上市。美团将始终坚持以客户为中心，不断加大在科技研发方面的投入，更好承担社会责任，更多创造社会价值，与广大合作伙伴一起发展共赢。



银牌合作单位

格灵深瞳

格灵深瞳是一家行业领先的人工智能公司，专注于将先进的计算机视觉、大数据分析、机器人和人机交互技术与应用场景深度融合，提供面向智慧金融、城市管理、商业零售、轨交运维、体育健康、元宇宙等领域的人工智能产品及解决方案。格灵深瞳成立于2013年，是国内计算机视觉行业和算法技术的早期探索者和实践者，并于2022年3月在上海证券交易所科创板挂牌上市，成为A股第一家AI计算机视觉上市企业。

格灵深瞳创始人、董事长、CEO 赵勇博士毕业于美国布朗大学计算机工程系，专攻计算机视觉和运算影像学，曾任谷歌总部研究院资深研究员。赵勇博士被评为中关村“高聚工程”高端领军人才，被聘为“北京市特聘专家”；荣获“北京市抗击新冠肺炎疫情先进个人”、“北京市劳动模范”等荣誉称号；荣获北京市华侨华人“京华奖”；赵勇博士曾担任北京市第十五届、十六届人大代表。

2023年4月，格灵深瞳发布了上市后的首份年报，报告期内格灵深瞳全年各项业绩表现亮眼。报告显示，格灵深瞳2022年全年逆势增长，并首次实现了盈利，成为A股首家盈利的AI计算机视觉公司。



银牌合作单位

美图影像

美图公司成立于 2008 年 10 月，以人工智能为驱动，是中国领先的影像处理及社交平台。美图公司的使命是“让每个人都能简单变美”，围绕着“美”创造了一系列软硬件产品如美图秀秀、美颜相机、美拍、美图宜肤以及美图魔镜，改变了用户创造与分享美的方式。2016 年 12 月 15 日，美图公司在香港联合交易所（“联交所”）主板挂牌上市，股票代码 1357。

2010 年，美图成立了美图影像实验室（MT Lab, Meitu Imaging & Vision Lab），致力于研究计算机视觉及深度学习相关的前沿影像技术，如人脸技术（该项技术可精准检测及识别人脸和面部特征）、极智美颜技术（通过分析使用者年龄、性别等人脸属性并据此定制个性化美颜方案）、图像分割技术等。目前，美图影像实验室（MT Lab）已开发一系列专有技术，注册超过 451 项专利（其中发明专利 256 项），并持有 192 项软件产品的著作权。2019 年 4 月，美图影像实验室（MT Lab）正式上线美图 AI 开放平台（ai.meitu.com），专注于人脸技术、人体技术、图像识别、图像处理、图像生成等核心领域，为客户提供经市场验证的专业 AI 算法服务和解决方案。



银牌合作单位

超集信息

苏州超集信息科技有限公司以“以卓越的计算机技术，创造智慧、互联、绿色的解决方案，帮助客户应对各类计算性能和智能制造的复杂挑战”为公司使命，注入二十余年深耕我国高性能计算行业的专注情怀，已成为数据中心、高性能计算、人工智能、液冷计算、智能制造以及 OEM 解决方案值得信赖的优秀企业，致力于为广大用户提供更强大、更稳定、更降耗的高性能计算产品，更专业的智能制造技术服务支持。

公司研产销一体，在苏州设立有年产量超 12 万台的生产基地，并设立有多个专属产品研发空间、多条全功能生产线，可实现高性能计算产品从研发到销售的快速迭代，灵活满足不同规模客户的定制化生产需求。

公司以苏州为核心，立足东南沿海，辐射全国市场，荣获国家高新技术企业、AAA 级信用企业、江苏省专精特新企业、苏州乃至苏南国家自主创新示范区瞪羚企业等荣誉。

近年来，公司也联合英特尔、英伟达、AMD、希捷、深信服等业内知名合作伙伴，相继推出了液冷计算、PlatforMax 智算融合平台等计算解决方案，加速了数字经济时代的应用落地。作为计算解决方案的优选品牌，公司为清华大学、北京大学、江南大学、文远知行、卫宁健康、暗物智能、网易、微软等单位，提供了性能优越的计算解决方案。

超集信息愿与您携手，以极致计算成就极致未来。



银牌合作单位

爱诗科技

爱诗科技是一家专注于通用视觉多模态大模型的初创公司，旨在通过 AIGC 技术赋能不同行业和场景的内容创意工作者与公司。我们希望通过视觉多模态大模型服务内容生产全价

价值链，提升从内容的生成、安全、分析、分发到商业化各环节的效率。我们将 AIGC 的能力覆盖视觉各模态，包括图像、3D、视频等。

视觉是人类最重要的信息交互模态，我们希望通过自主研发的视觉多模态大模型丰富、深刻内容的创作与管理，最终实现拓展类人视觉境界的愿景，创造更多更好的视觉体验。

公司创始人/CEO 王长虎博士深耕计算机视觉、人工智能领域 20 年。曾任字节跳动视觉技术负责人，参与了抖音和 TikTok 等产品从 0 到 1 的建设和发展，搭建了字节跳动视觉算法平台和业务中台，带领团队将相关技术广泛应用到字节跳动全线产品中。曾任微软亚洲研究院主管研究员，专注前沿研究，主导了多个全球首创的前沿 CV 研究。曾任龙湖数科智能引擎部总经理，负责集团各航道的信息化、线上化、自动化和智能化升级，打造业务数智化闭环。

王长虎的主要研究方向包括计算机视觉、机器学习、多媒体分析等，是 IEEE 和 ACM 高级会员，发表了近百篇论文，是 20 多个国际会议与期刊的编委或审稿人，拥有和申请了数百个中国和美国发明专利。



银牌合作单位

趋动云

趋动科技作为软件定义 AI 算力技术的领导厂商，专注于为全球用户提供国际领先的数据中心级 AI 算力虚拟化和资源池化软件及解决方案，已完成中关村高新、国高新、“专精特新”等企业认证。趋动科技的 OrionX 猎户座 AI 算力资源池化软件能够帮助用户提高资源利用率和降低 TCO，提高算法工程师的工作效率。趋动科技的双子座 GEMINI AI 训练平台，为客户提供强大的 AI 算力管理服务以及高效的算法开发和训练支持，能够化繁为简，帮助企业建好 AI 平台、管好 GPU、用好 AI 服务。

依托全球领先的 AI 算力池化技术，趋动科技重磅推出趋动云 VirtAI Cloud，为万千企业和 AI 开发者带来又便宜、又好用的 AI 算力池化云服务。凭借标准化、可复制的产品架构，趋动科技得到了包括互联网、金融、电信运营商、自动驾驶、能源、科研机构 and 高校等大量行业头部客户的认可。

资本市场对于趋动科技的发展充满信心——趋动科技成立两年多已经完成近亿美元的融资，顶级的投资机构持续支持趋动科技的发展，包括国开装备基金、沙特阿美旗下多元化风投基金 Prosperity7 Ventures、元禾重元、招银国际、顺为、高瓴、嘉御、戈壁、讯飞和涌铨在内的国内外顶级 VC 正在见证趋动科技锐意进取的脚步。



银牌合作单位

商汤

商汤科技以“坚持原创，让 AI 引领人类进步”为使命，“以人工智能实现物理世界和数字世界的连接，促进社会生产力可持续发展，并为人们带来更好的虚实结合生活体验”为愿景。作为亚洲最大的 AI 软件公司，商业化赋能多个行业。商汤基于 SenseCore 为客户开发并提供具有模块化灵活性的标准软件平台，业务已涵盖智慧商业、智慧城市、智慧生活、智能汽车四大板块。“大装置+大模型”迎时代机遇，SenseCore AI 大装置实现高效率、低成本、规模化 AI，商汤超大模型进入加速成长期。面向未来，商汤将通用人工智能（AGI）作为核心发展

战略，围绕关键领域，构建 AGI 核心能力。推出包含语言大模型“商量 SenseChat”等一系列生成式 AI 模型及应用在内的“日日新 SenseNova”大模型体系，前瞻布局 AGI 基础设施，人机共智，创造新型研究范式。

## 部分组织单位简介



中国人工智能学会

Chinese Association for Artificial Intelligence

主办单位：中国人工智能学会

中国人工智能学会（Chinese Association for Artificial Intelligence, CAAI）成立于1981年，是经国家民政部正式注册的我国智能科学技术领域唯一的国家级学会，是全国性4A级社会组织，挂靠单位为北京邮电大学；是中国科学技术协会的正式团体会员，具有推荐“两院院士”的资格。

目前拥有51个分支机构，包括43个专业委员会和8个工作委员会，覆盖了智能科学与技术领域。学会活动的学术领域是智能科学技术，活动地域是中华人民共和国全境，基本任务是团结全国智能科学技术工作者和积极分子通过学术研究、国内外学术交流、科学普及、学术教育、科技会展、学术出版、人才推荐、学术评价、学术咨询、技术评审与奖励等活动促进我国智能科学技术的发展，为国家的经济发展、社会进步、文明提升、安全保障提供智能化的科学技术服务。

学会自主创办全球人工智能技术大会、中国人工智能大会、中国智能产业高峰论坛、中国AI+创新创业大赛、“华为杯”全国大学生智能设计竞赛、国际人工智能会议、IEEE云计算与智能系统国际会议等规模化、系列化学术活动。拥有“吴文俊人工智能科学技术奖”、“中国人工智能学会优秀博士学位论文评选”、“学会会士评选”、“学会先进个人”等多个奖项与评选。

欢迎广大科技工作者踊跃加入中国人工智能学会！



中国图象图形学学会

CHINA SOCIETY OF IMAGE AND GRAPHICS

主办单位：中国图象图形学学会

中国图象图形学学会（China Society of Image and Graphics，缩写CSIG）成立于1990年，是经国家民政部批准成立的国家一级学会，是中国科学技术协会的正式团体成员。由从事图像图形基础理论与应用研究、软硬件技术开发及应用推广的专家学者和相关科技工作者组成。

中国图象图形学学会的宗旨是团结广大图像图形领域的科技工作者，积极开展图像图形基础理论和高新技术的研究，促进该学科技术的发展和在国民经济各个领域的推广应用。本学会专业领域涵盖了数字图像处理、图像理解、计算机视觉、图像压缩与传输、电视技术、科学计算可视化、虚拟现实、多媒体技术、模式识别、计算机图像图形学、医学影像处理、计算机动画、空间信息系统等。

中国图象图形学学会的主要任务是：开展科学研究与学术交流，活跃学术思想，促进学科发展，推广先进技术，发现、培养和举荐人才，提供技术咨询与服务，普及图像图形科技知识，传播科学思想和方法，编辑出版学术和科普书刊，加强同国内外学术团体和科技工作者的友好交往。

## VALSE——学术华尔兹

VALSE发源于2011年，是Vision And Learning SEminar的简写，取“华尔兹舞”之意。旨在为全球计算机视觉、模式识别、机器学习、多媒体技术等相关领域的华人青年学者提供一个平等、自由的学术交流舞台。发起VALSE的主要动机是我们深感中国计算机视觉与机器学习领域缺少一个以华人青年学者为主体的常态化学术交流舞台。有鉴于此，2011年初，山世光、潘纲、刘青山和颜水成共同讨论了发起一个视觉与学习领域华人青年学者研讨会的想法，之后该想法得到了李学龙、徐东、周志华、马毅等青年学者的大力支持。为此，首届视觉与学习青年学者研讨会于2011年4月8日-9日在杭州成功举行。此后，主要发起人山世光、潘纲、刘青山、颜水成、李学龙等共同讨论确定了VALSE这一名称。之后由山世光牵头起草、逐步完善了VALSE作为一个学术社区的组织原则和发展规划，特别是大会讲者由指导委员会按照“诺贝尔奖”模式推荐和选举产生的原则。

截至目前，VALSE已成功举办12届，分别为VALSE2011（杭州），VALSE2012（西安），VALSE2013（南京），VALSE2014（青岛），VALSE2015（成都），VALSE2016（武汉），VALSE2017（厦门），VALSE2018（大连），VALSE2019（合肥），VALSE2020（线上），VALSE2021（杭州），VALSE2022（天津）。VALSE2023将于2023年6月10-12日在无锡举行，由中国人工智能学会、中国图象图形学学会主办，江南大学、无锡国家高新技术产业开发区管理委员会承办，江苏省人工智能学会、无锡市计算机学会、中国图象图形学学会青年工作委员会协办。除大会演讲之外，VALSE年度大会还不断推陈出新，逐渐增加了Poster/Spotlight/Demo/Tutorial、年度进展评述、Workshops和工业界技术分享等环节，参会人数也逐渐增长到了5000人以上。在此过程中，逐渐形成了由山世光、潘纲、刘青山、颜水成、李学龙、周志华、徐东、马毅、周昆、高新波、何晓飞、余凯、杨健、黄华、白翔等15名青年学者组成的指导委员会。2018年，VALSE成立了指导委员会委员45岁退休的原则，故高新波、马毅、杨健、周志华四位老师进入顾问委员会，同时吸纳了华刚、汪萌、虞晶怡和张敏灵四位老师进入指导委员会。此外，VALSE大会也吸引了越来越多的企业参加，已成为计算机视觉与机器学习领域“产-学-研”合作交流的重要平台。

为配合VALSE系列年度研讨会，VALSE主要发起人之一山世光于2014年6月18日创建了VALSE专业学术交流QQ群，即VALSE-A群（2000人满）。此后，逐渐开通了VALSE-B-R群（群号：137634472）。从而形成了一个近两万两千人的视觉与学习青年学者在线社区。自2014年9月开始VALSE每周或隔周定期举办VALSE Webinar学术报告会，在已有品牌活动的基础上，我们最新推出了《VALSE短教程》、《VALSE论文速览》以经济、便捷的在线形式，将众多青年学者的最新工作和学术思想呈现给散落在世界各地的青年学者和研究生。自2020年4月以来，活动迁移至B站直播。VALSE Online 迄今已组织288期的在线学术报告。特别是众多知名青年学者的亲临报告（如：颜水成、王晓刚、屠卓文、凌海滨、沈春华、张磊、朱军、李纯明、印卧涛、熊红凯、刘利刚、齐国君、刘烨斌、毕彦超、Philip Torr、华刚、刘小明等），更大大激励了VALSE Online的发展，目前VALSE B站有3.9万粉丝，B站所存放的视频都是VALSE每周的Webinar录制的视频，视频的累计播放量71.6万，单个视频的最高播放量在4.9万次。逐渐形成了一个独具特色、经济高效、便捷实用的在线学术交流舞台。VALSE历史视频都会更新在B站空间，欢迎在B站搜索VALSE\_Webinar关注我们！也可以通过链接直接观看：<https://space.bilibili.com/562085182>。

VALSE Online 是青年学者自组织、自管理的舞台。其兴起不仅得益于VALSE指导委员会成员的大力支持，更离不开逐渐形成的VALSE Online组织团队。除发起人山世光之外，越来越多的青年学者参与了进来。特别是白翔（华中科技大学）、程明明（南开大学）、孟德宇（西安交通大学）、彭莹（四川大学）、贾伟（合肥工业大学）、郑海永（中国海洋

大学)、纪荣嵘(厦门大学)、姬艳丽(电子科技大学)、张利军(南京大学)、章国锋(浙江大学)、左旺孟(哈尔滨工业大学)、张兆翔(自动化所)、何晖光(自动化所)、禹之鼎(CMU)、王乃岩(图森未来)、苏航(清华大学)、欧阳万里(悉尼大学)等,都为VALSE Online的发展付出了大量心血。为了更好地组织VALSE年度及在线活动,VALSE成立了常务AC委员会(LACC),资深AC委员会(SACC),执行AC委员会(EACC)(名单参见后面的委员会名单),220名青年学者积极参与到了相关活动的组织中。

关于VALSE的更多信息,请访问VALSE总主页:<http://valser.org>(特别鸣谢中国海洋大学郑海永教授搭建该平台)。欢迎大家扫码关注之后页面的VALSE微信公众号,查看VALSE的最新消息。

借此机会,我们要诚挚感谢本届VALSE大会的合作单位,包括:OPPO、图森未来、视拓云、华为、百度、字节跳动、阿里妈妈、合合信息、蚂蚁技术研究院、极市平台、奥比中光、真格基金、地平线、金山办公、融科联创(天津)、思腾合力(天津)、美团、格灵深瞳、美图、超集信息、爱诗科技、趋动科技、商汤科技。感谢这些公司的负责人和联系人对赞助VALSE而做出的努力,谢谢你们!

上述成绩的取得更离不开众多 VALSER 们的支持和鼓励,尤其是众多常态化参与VALSE Webinar 报告会的老师和同学们,我们深表谢意!今后,我们将继续集思广益、创新学术交流和合作模式,更好地搭建视觉与学习领域华人学术交流大舞台,为本领域的产研学发展起到更好的促进作用。

## VALSE 在线活动参与方法介绍

1、VALSE 每周举行的 Webinar 活动依托 B 站直播平台进行，欢迎在 B 站搜索 **VALSE\_Webinar** 关注我们！

直播地址：

<https://live.bilibili.com/22300737>；

历史视频观看地址：

<https://space.bilibili.com/562085182/>

2、VALSE Webinar 活动通常**每周三**晚上 20:00 进行，但偶尔会因为讲者时区问题略有调整，为方便您参加活动，请关注 VALSE 微信公众号：**valse\_wechat** 或加入 VALSE QQ **R 群**，群号：**137634472**）；

**\*注：**申请加入 VALSE QQ 群时需验证**姓名、单位和身份**，缺一不可。入群后，请实名，姓名身份单位。身份：学校及科研单位人员 T；企业研发 I；博士 D；硕士 M。

3、VALSE 微信公众号一般会在**每周四**发布下一周 Webinar 报告的通知。

4、您也可以通过访问 VALSE 主页：<http://valser.org>/直接查看 Webinar 活动信息。Webinar 报告的 PPT（经讲者允许后），会在 VALSE 官网每期报告通知的最下方更新。



## VALSE 2022-2023 年度 Webinar 报告一览表

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
2022.1.19	赵国英 (芬兰 Oulu 大学)	Emotional Interfaces with Face Analysis (基于脸部分析的情感接口)		洪晓鹏	洪晓鹏 (哈尔滨工业大学)		
2022.3.9	杨宗鑫 (浙江大学)	视频理解中的多目标联合分割	杨宗鑫 (浙江大学)、王立君 (大连理工大学)、王利民 (南京大学), 林巍岷 (上海交通大学)、杨帆 (快手)	丛润民	丛润民 (北京交通大学)		
	王立君 (大连理工大学)	基于外观模型的视频目标跟踪与分割					
2022.3.16	王连生 (厦门大学)	一种放射影像预训练方法	王连生 (厦门大学)、崔智铭 (香港大学)、秦璟 (香港理工大学)、周郁音 (UCSC)、于乐全 (香港大学)	于乐全、窦琪	窦琪 (香港中文大学)、于乐全 (香港大学)		
	崔智铭 (香港大学)	Shape-aware tooth segmentation in digital dentistry					
	秦璟 (香港理工大学)	Rethink Deep Learning Models for Medical Image Analysis from an Epistemological Perspective					
2022.3.23	贺通 (上海人工智能实验室)	3D instance segmentation with dynamic convolution	贺通 (上海人工智能实验室)、陈挺 (Google Brain)、代季峰 (商汤科技)、王鹤 (Peking University)、张兆翔 (中国科学院自动化研究所)	欧阳万里、周晓巍	欧阳万里 (悉尼大学)	周晓巍 (浙江大学)	
	陈挺 (Google Brain)	Pix2seq: A Language Modeling Framework for Object Detection					
2022.3.30	俞扬 (南京大学)	数据驱动的强化学习及其工业应用		魏秀参	魏秀参 (南京理工大学)		
2022.4.6	方聪 (北京大学)	Convex Formulation of Overparameterized Deep Neural Networks	方聪 (北京大学)、袁晓彤 (南京信息工程大学)、刘日升 (大连理工大学)、左旺孟 (哈尔滨工业大学)、黄雷 (北京航空航天大学)	王旗龙	王旗龙 (天津大学)	黄雷 (北京航空航天大学)	
	袁晓彤 (南京信息工程大学)	基于随机模型的深度学习优化					
	刘日升 (大连理工大学)	基于梯度的双层优化方法初探					
2022.4.13	张铁林 (中国科学院自动化研究所)	基于类脑脉冲神经网络的视听觉信息处理	吴小俊 (江南大学)、唐华锦 (浙江大学)、张兆翔 (中国科学院自动化研究	李玺	李玺 (浙江大学)		

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
	余肇飞 (北京大学)	面向类脑视觉的生物视觉编解码机制和模型	所)、张铁林 (中国科学院自动化研究所)、余肇飞 (北京大学)、于云龙 (浙江大学)				
2022.4.20	弋力 (清华大学)	面向交互的四维动态场景理解	弋力 (清华大学)、赵恒爽 (香港大学)、韩晓光 (香港中文大学 (深圳))、史少帅 (德国马克思普朗克研究所)、王兴刚 (华中科技大学)、徐凯 (国防科技大学)	王鹤	王鹤 (北京大学)	盛律 (北京航空航天大学)	
	赵恒爽 (香港大学)	Scene Understanding in 3D and 2D-3D					
2022.5.25	林绍辉 (华东师范大学)	深度模型压缩与加速技术及其应用	陈涛 (复旦大学)、林绍辉 (华东师范大学)、周天异 (新加坡科技研究局 A*STAR)、俞刚 (腾讯)、王云鹤 (华为诺亚方舟实验室)	陈涛	胡鹏 (四川大学)		
	陈涛 (复旦大学)	Efficient Neural Architecture Design for Visual Analysis					
2022.6.1	沈春华 (浙江大学)	从单目深度估计到单目三维场景重建		刘昊	夏勇 (西北工业大学)	刘昊 (宁夏大学)	
2022.6.8	田渊栋 (Meta AI Research (FAIR))	Towards Better Understanding of Contrastive Learning		刘日升	苏航 (清华大学)	刘日升 (大连理工大学)	
2022.6.15	陆路 (UPenn)	Learning operators using deep neural networks for diverse applications	陆路 (Upenn)、王建勋 (University of Notre Dame)、许志钦 (上海交通大学)、孙浩 (中国人民大学)	黄高	黄高 (清华大学)		
	王建勋 (University of Notre Dame)	Leveraging physics-induced bias in scientific machine learning for computational mechanics					
2022.6.29	宫明明 (墨尔本大学)	CRIS: CLIP-Driven Referring Image Segmentation	宫明明 (墨尔本大学)、苏冰 (中国人民大学)、曹越 (微软亚洲研究院)、刘同亮 (悉尼大学)、胡鹏 (四川大学)	刘宇	刘宇 (大连理工大学)	刘同亮 (悉尼大学)	
	苏冰 (中国人民大学)	What to contrast?					
2022.7.6	李霄霄 (The University of British Columbia)	Federated learning for healthcare: from theory to practice	李霄霄 (The University of British Columbia)、徐子乐 (Nvidia)、李响 (哈佛大	雷柏英	雷柏英 (深圳大学)	陈浩 (香港科技大学)	

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
	徐子乐 (Nvidia)	Techniques and Tools for Collaborative Development of AI Models across Institutes	学、麻省总医院)、陈浩 (香港科技大学)、周郁音 (UC Santa Cruz)、秦璟 (香港理工大学)				
	李响 (哈佛大学、麻省总医院)	联邦学习在医学图像处理的应用					
2022.7.13	张敬 (The University of Sydney)	ViTAE Transformer 超级视觉深度学习模型	张敬 (The University of Sydney)、刘同亮 (The University of Sydney)、韩波 (香港浸会大学)、宫辰 (南京理工大学)、庄博涵 (Monash University)、王超岳 (京东探索研究院)	程明明	程明明 (南开大学)	刘同亮 (The University of Sydney)、韩波 (香港浸会大学)、宫辰 (南京理工大学)	
	刘同亮 (The University of Sydney)	标签学习简介——助力超级深度学习					
2022.8.3	周博磊 (加州大学洛杉矶分校)	生成模型的解释性, 交互性, 小样本学习, 以及我对 Dalle2 的看法	周博磊 (加州大学洛杉矶分校)、李崇轩 (中国人民大学)、左旺孟 (哈尔滨工业大学)、胡瀚 (微软亚洲研究院)、宋颀 (斯坦福大学)	谢凌曦	谢凌曦 (华为云计算公司)	左旺孟 (哈尔滨工业大学)	
	李崇轩 (中国人民大学)	扩散概率模型与其加速推断算法					
2022.8.10	高弘扬 (Iowa State University)	Molecular Representation Learning via Heterogeneous Motif Graph Neural Networks	高弘扬 (Iowa State University)、白璐 (中央财经大学)、严骏驰 (上海交通大学)、杨旭 (西安电子科技大学)	江波	江波 (安徽大学)	杨旭 (西安电子科技大学)	
	白璐 (中央财经大学)	Transitive-Aligned Graph Neural Networks for Graph Classification					
2022.8.17	王瑞平 (中国科学院计算技术研究所)	视觉场景图—表示、生成与应用	王瑞平 (中国科学院计算技术研究所)、李永露 (香港科技大学)、卢策吾 (上海交通大学)、郑伟诗 (中山大学)、王利民 (南京大学)	丁长兴	丁长兴 (华南理工大学)	王利民 (南京大学)	
	李永露 (香港科技大学)	Three-Stages in Human-Object Interaction Detection					
	卢策吾 (上海交通大学)	Bridging Isolated Islands in Human Activity Understanding					
2022.8.31	张林 (同济大学)	图像雾霾分析: 模拟、评估和消除	张林 (同济大学)、任文琦 (中山大学)、刘家瑛 (北京大学)、潘金山 (南京理工大学)	吴庆波	吴庆波 (电子科技大学)	潘金山 (南京理工大学)	
	任文琦 (中山大学)	融合先验知识的图像视频复原方法研究					
2022.9.7	李鸿升 (香港中文大学)	MPPNet: Multi-Frame Feature Intertwining with Proxy Points for 3D Temporal Object Detection	李鸿升 (香港中文大学)、Fisher Yu (苏黎世联邦理	马超			

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
	Fisher Yu (苏黎世联邦理工)	High-Quality 4D Scene Understanding in Autonomous Driving	工)、沈春华 (浙江大学)、王乃岩 (图森未来)、许春景 (华为技术有限公司)		马超 (上海交通大学)		
2022.9.14	吴祖焯 (复旦大学)	基于 Transformer 的视频内容理解	吴祖焯 (复旦大学), 寿政 (新加坡国立大学), 赵行 (清华大学), 刘洋 (北京大学)	朱霖潮	朱霖潮 (浙江大学)	刘洋 (北京大学)	
	寿政 (新加坡国立大学)	大规模视频-语言预训练					
2022.9.21	万军 (中国科学院自动化研究所)	多模态人脸防伪研究: 方法、数据、竞赛	万军 (中国科学院自动化研究所)、邓伟洪 (北京邮电大学)、周文柏 (中国科学技术大学)、易冉 (上海交通大学)	彭春蕾	彭春蕾 (西安电子科技大学)	丁长兴 (华南理工大学)	
	邓伟洪 (北京邮电大学)	安全可信的人脸识别					
2022.10.12	陈小军 (深圳大学)	Self-supervised Image Clustering	陈小军 (深圳大学)、康昭 (电子科技大学)、胡迪 (中国人民大学)、胡鹏 (四川大学)	杨旭	杨旭 (西安电子科技大学)		
	康昭 (电子科技大学)	自监督思想在聚类中的应用					
2022.10.19	代季峰 (清华大学)	通用感知模型研究	代季峰 (清华大学)	夏勇	夏勇 (西北工业大学)		
2022.11.23	王杨 (合肥工业大学)	Switchable Online Knowledge Distillation	王杨 (合肥工业大学)、徐行 (电子科技大学)、黄增丰 (复旦大学)、刘新旺 (国防科技大学)	姬艳丽	姬艳丽 (电子科技大学)		
	徐行 (电子科技大学)	Cross-modal Correlation Learning via Knowledge Transfer and Distillation					
2022.12.1	徐东 (香港大学)	迈向有影响力的研究: 漫谈视觉域适应和深度视频压缩	徐东 (香港大学)	刘偲	刘偲 (北京航空航天大学)		
2022.12.14	周建涛 (澳门大学)	面向社交网络的多媒体信息安全和取证	周建涛 (澳门大学)、余梓彤 (南洋理工大学)、李雷达 (西安电子科技大学)、万人杰 (香港浸会大学)	李皓亮	李皓亮 (香港城市大学)	万人杰 (香港浸会大学)	
	余梓彤 (南洋理工大学)	Towards Trustworthy Face Physiological Sensing and Security systems					
2022.12.21	孙显 (中国科学院空天信息创新研究院)	大规模遥感图像多任务学习方法探索与应用	孙显 (中国科学院空天信息创新研究院)、张鼎文 (西北工业大学)、程臻 (西北工业大学)、冯婕 (西安电子科技大学)	杨曦	杨曦 (西安电子科技大学)		
	张鼎文 (西北工业大学)	浅析遥感图像智能理解					
2022.12.28	古纾旸 (微软亚洲研究院)	向量量化的扩散概率模型 (Vector quantized diffusion models)	古纾旸 (微软亚洲研究院)、宋佳铭 (NVIDIA)	李崇轩			

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
	宋佳铭 (NVIDIA Research)	用扩散模型高效求解逆问题	Research)、卢志武 (中国人民大学)、赵洲 (浙江大学)		李崇轩 (中国人民大学)		
2023.1.6	黄泽昊 (图森未来)	基于 Predictor 的神经网络结构搜索方法及其在自动驾驶中的应用		姬艳丽	姬艳丽 (电子科技大学)		
	王峰 (图森未来)	全稀疏 3d 检测器					
	谢凌曦 (华为)	华为云盘古视觉基础模型的挑战难题发布					
	秦征 (OPPO)	用户体验驱动的数字之美					
	冯展鹏 (深圳云天励飞技术股份有限公司)	Towards Automated Artificial Intelligence					
	张号遼 (深圳云天励飞技术股份有限公司)	ParC-Net: Position Aware Circular Convolution with Merits from ConvNets and Transformer					
2023.2.15	黄民烈 (清华大学)	深度揭秘 chatGPT: 原理, 技术和启示		苏航	苏航 (清华大学)		
2023.2.22	翟广涛 (上海交通大学)	媒体质量评价	翟广涛 (上海交通大学)、李弘扬 (上海人工智能实验室)、徐畅 (悉尼大学)、陈涛 (复旦大学)、苗振伟 (阿里巴巴)	陈涛	陈涛 (复旦大学)		
	李弘扬 (上海人工智能实验室)	自动驾驶场景道路结构认知与关系推理					
2023.3.15	吴金建 (西安电子科技大学)	面向事件相机的感光芯片设计及异步信号智能处理技术研究		戴玉超	吴金建 (西安电子科技大学)	戴玉超 (西北工业大学)	
2323.3.22	庞天宇 (新加坡 Sea AI Lab)	Trustworthy Diffusion Models & Diffusion Models for Trustworthy ML	庞天宇 (新加坡 Sea AI Lab)、Chaowei Xiao (ASU)、崔鹏 (清华大学)、徐畅 (悉尼大学)、苏航 (清华大学)	张弘扬	张弘扬 (滑铁卢大学)	徐畅 (悉尼大学)	
	Chaowei Xiao (ASU)	Towards Socially Responsible machine learning: Security and Ethics					
2023.3.29	宋睿华 (中国人民大学)	多模态预训练模型及在智能创作领域的应用	宋睿华 (中国人民大学)、朱霖潮 (浙江大学)、段楠 (微软亚洲研究院)、刘洋 (北京大学)	刘洋	刘洋 (北京大学)	朱霖潮 (浙江大学)	
	朱霖潮 (浙江大学)	多模态分析中的迁移与对齐技术					
2323.4.12	于力军 (Carnegie Mellon University)	Masked Generative Video Transformer	于力军 (Carnegie Mellon University)、Kihyuk Sohn	孟德宇		蒋路 (Google)	

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
	Kihyuk Sohn (Google Research)	Video Probabilistic Diffusion Models in Projected Latent Space	(Google Research)、陈启峰 (香港科技大学)、寿政 (National University of Singapore)、付彦伟 (复旦大学)		孟德宇 (西安交通大学)	Research & Carnegie Mellon University)	
2023.4.19	王文冠 (浙江大学)	Knowledge-Driven Perception	彭宇新 (北京大学)、俞俊 (杭州电子科技大学)、王瑞平 (中科院计算所)、吴琦 (阿德莱德大学)、王文冠 (浙江大学)、朱磊 (山东师范大学)	李冠彬	李冠彬 (中山大学计算机学院)	王文冠 (浙江大学)、吴琦 (阿德莱德大学)	
	朱磊 (山东师范大学)	高效能跨模态检索					
2023.4.26	郑锋 (南方科技大学)	“Segment Anything”浅谈及应用介绍	程明明 (南开大学)、张磊 (粤港澳大湾区数字经济研究院(IDEA))、王兴刚 (华中科技大学)、谢凌曦 (华为)、郑锋 (南方科技大学)、张力 (复旦大学)	李冠彬、山世光	李冠彬 (中山大学)	张正 (哈尔滨工业大学(深圳))、胡庆拥 (天津大学)	
	张力 (复旦大学)	Vision foundation model and its application on autonomous driving					

# VALSE 2022-2023 年度 Student Webinar 报告一览表

日期	讲者	报告题目	嘉宾	主持人	主办 AC	协办 AC	责任 AC
2022.1.13	梁思源 (中国科学院信息工程研究所)	浅谈顶会论文投稿	梁思源 (中国科学院信息工程研究所)、庞树茂 (南方医科大学)、王玫 (北京邮电大学)、白翔 (华中科技大学)、孟德宇 (西安交通大学)	闫杰熹	刘偲 (北京航空航天大学)	中国图象图形学报	
	庞树茂 (南方医科大学)	科技论文写作经验分享					
	王玫 (北京邮电大学)	科研路上的心得与体会					
2022.1.23	陶仁帅 (北京航空航天大学)	如何写好一个 Rebuttal		秦浩桐	刘偲 (北京航空航天大学)		
	李一鸣 (清华大学)	从审稿人的视角浅析如何撰写 Rebuttal					
	张松阳 (上海科技大学)	从 Submission 到 Rebuttal, 再到 Re-submission					
2022.5.29	都一凡 (中国人民大学)	视觉-语言预训练模型综述	张含望 (新加坡南洋理工大学)、段楠 (微软亚洲研究院)、吴琦 (澳大利亚阿德莱德大学)、魏忠钰 (复旦大学)	刘乾 秦浩桐 刘偲	刘偲 (北京航空航天大学)		
	鲍航波 (哈尔滨工业大学)	BEiT: BERT Pre-Training of Image Transformers					
	张宋扬 (罗切斯特大学)	视频辅助的句法分析					
	卢雨洁 (加州大学)	视觉想象辅助语言理解					
2022.12.10— 2022.12.11	谢琦 (西安交通大学)	参数化卷积方法及其应用		盛律	盛律 (北京航空航天大学)	刘偲 (北京航空航天大学)	
	雍宏巍 (香港理工大学)	深度神经网络优化技术的相关探索					
	郑玉棒 (电子科技大学)	张量分解方法及其在高维数据复原中的应用		周满			
	刘晋源、刘铸 (大连理工大学)	面向底层视觉的双层优化方法研究					
	郭春乐 (南开大学)	The Beauty of Curve in Image Enhancement					
	岳宗胜 (南洋理工大学)	深度贝叶斯建模技术在底层视觉任务中的应用		马龙			
	芮翔宇 (西安交通大学)	去噪模型中的加权可以在多个模型中迁移: 理论初探					

胡启明 (天津大学)	Trash or Treasure? An Interactive Dual-Stream Strategy for Single Image Reflection Separation				
肖杰 (中国科学技术大学)	Stochastic Window Transformer for Image Restoration				
赵子祥 (西安交通大学)	Discrete Cosine Transform Network for Guided Depth Map Super-Resolution				
黄杰 (中国科学技术大学)	Exposure-Consistency Representation Learning for Exposure Correction				
王荫槐 (北京大学)	GAN Prior based Null-Space Learning for Consistent Super-Resolution				
马龙 (大连理工大学)	高效低光照视觉增强方法研究				
周满 (南洋理工大学)	漫谈底层视觉中的傅里叶先验到定制化底层视觉理论			李重仪	
王红 (腾讯)	领域知识驱动的深度 CT 图像复原方法				
徐涵、唐霖峰 (武汉大学)	面向实际应用需求的图像融合方法				
牟冲 (北京大学)	面向真实场景的交互式图像重建			芮翔宇	
张凯 (苏黎世联邦理工大学)	Deep Plug-and-play/ Unfolding/ Blind Image Restoration				

## VALSE 各委员会

### 顾问委员会

高新波	西安电子科技大学	马毅	UC Berkeley
周志华	南京大学	杨健	南京理工大学

### 指导委员会

白翔	华中科技大学	何晓飞	飞步科技
华刚	Wormpex AI Research	黄华	北京理工大学
李学龙	西北工业大学	刘青山	南京信息工程大学
潘纲	浙江大学	山世光	中国科学院计算技术研究所
汪萌	合肥工业大学	徐东	悉尼大学
颜水成	冬海集团	虞晶怡	上海科技大学
余凯	地平线机器人	张敏灵	东南大学
周昆	浙江大学		

### 常务 AC 委员会 (LACC)

<b>主席:</b>	白翔	华中科技大学		
<b>副主席:</b>	程明明	南开大学	纪荣嵘	厦门大学
<b>常务 AC:</b>	程明明	南开大学	白翔	华中科技大学
	纪荣嵘	厦门大学	姬艳丽	电子科技大学
	韩琥	中国科学院计算技术研究所	刘日升	大连理工大学
	贾伟	合肥工业大学	欧阳万里	悉尼大学
	孟德宇	西安交通大学	王楠楠	西安电子科技大学
	彭玺	四川大学	章国锋	浙江大学
	王琦	西北工业大学	张兆翔	中国科学院自动化所
	张利军	南京大学	左旺孟	哈尔滨工业大学
	郑海永	中国海洋大学	郭裕兰	国防科技大学
	戴玉超	西北工业大学	王兴刚	华中科技大学
	卢策吾	上海交通大学	魏秀参	南京理工大学
	苏航	清华大学	周晓巍	浙江大学
	张姗姗	南京理工大学		

### 资深 AC 委员会 (SACC)

<b>主席:</b>	姬艳丽	电子科技大学		
<b>副主席:</b>	严骏驰	上海交通大学	谢凌曦	华为技术有限公司
	张姗姗	南京理工大学		
<b>资深 AC:</b>	陈涛	复旦大学	樊彬	北京科技大学
	丛润民	北京交通大学	高常鑫	华中科技大学
	洪晓鹏	哈尔滨工业大学	高陈强	重庆邮电大学
	胡鹏	四川大学	连宙辉	北京大学
	林巍峤	上海交通大学	刘德	北京航空航天大学

李冠彬	中山大学	明悦	北京邮电大学
潘金山	南京理工大学	任传贤	中山大学
马超	上海交通大学	王利民	南京大学
王文冠	苏黎世联邦理工	王兴刚	华中科技大学
王云鹤	华为诺亚方舟实验室	夏勇	西北工业大学
魏云超	悉尼科技大学	徐畅	悉尼大学
许永超	武汉大学	严骏驰	上海交通大学
杨猛	中山大学	张磊	重庆大学
张林	同济大学	张姗姗	南京理工大学
赵健	军事科学院	郑伟诗	中山大学
谢凌曦	华为技术有限公司	江波	安徽大学

### 执行 AC 委员会 (EACC)

**主席:** 郭裕兰 国防科技大学

**副主席:** 胡鹏 四川大学

李冠彬 中山大学

**执行 AC:** 白亚龙 京东 AI 研究院

曹越 微软亚洲研究院

陈涛 复旦大学

代登信 ETH Zurich

丁长兴 华南理工大学

董伟生 西安电子科技大学

窦琪 香港中文大学

冯尊磊 浙江大学

宫明明 墨尔本大学

郭晓杰 天津大学

韩晓光 香港中文大学 (深圳)

胡迪 中国人民大学

胡鹏 四川大学

黄高 清华大学

贾旭 大连理工大学

焦剑波 牛津大学

雷柏英 深圳大学

李冠彬 中山大学

李雷达 西安电子科技大学

李文 电子科技大学

林迪 天津大学

刘峰 密歇根州立大学

刘俊 新加坡科技设计大学

刘洋 北京大学

马超 上海交通大学

彭春雷 西安电子科技大学

雷柏英 深圳大学

郑乾 浙江大学

贲晔焯 山东大学

常晓军 蒙纳士大学

丛润民 北京交通大学

邓欣 北京航空航天大学

董超 中国科学院深圳先进技术研究院

董宣毅 悉尼科技大学

段立新 电子科技大学

宫辰 南京理工大学

顾舒航 悉尼大学

韩波 香港浸会大学

洪晓鹏 西安交通大学

胡建方 中山大学

胡玮 北京大学

黄雷 北京航空航天大学

江波 安徽大学

柯秋红 墨尔本大学

李策 兰州理工大学

李皓亮 香港城市大学

李爽 北京理工大学

李镇 香港中文大学 (深圳)

林绍辉 华东师范大学

刘昊 宁夏大学

刘同亮 悉尼大学

刘宇 大连理工大学

牛玉磊 新加坡南洋理工大学

秦杰 阿联酋起源人工智能研究院

任文琦	中国科学院信息工程研究所	沈 为	上海交通大学
盛 律	北京航空航天大学	舒祥波	南京理工大学
宋 杰	浙江大学	眭亚楠	清华大学
隋 尧	哈佛大学	谭明奎	华南理工大学
涂志刚	武汉大学	万人杰	新加坡南洋理工大学
汪婧雅	上海科技大学	王 栋	大连理工大学
王 鹤	北京大学	王立君	大连理工大学
王旗龙	天津大学	王 鑫	清华大学
王奕森	北京大学	王 正	东京大学
王智慧	大连理工大学	韦星星	北京航空航天大学
文碧汉	新加坡南洋理工大学	吴金建	西安电子科技大学
吴 琦	阿德莱德大学	吴庆波	电子科技大学
夏 勇	西北工业大学	谢凌曦	华为数字技术有限公司
徐 迈	北京航空航天大学	徐 易	阿里巴巴(美国)集团
杨二昆	北卡罗来纳大学教堂山分校	杨 恒	深圳爱莫科技有限公司
杨文瀚	南洋理工大学	杨 曦	西安电子科技大学
杨 欣	华中科技大学	杨 旭	中国科学院自动化所
杨 旭	西安电子科技大学	姚权铭	第四范式
叶 茫	武汉大学	于乐全	香港大学
张鼎文	西北工业大学	张弘扬	滑铁卢大学
张 健	北京大学	张平平	大连理工大学
张瑞茂	香港中文大学(深圳)	张长青	天津大学
张 正	哈尔滨工业大学(深圳)	赵恒爽	牛津大学
赵文达	大连理工大学	赵 洋	合肥工业大学
郑 乾	新加坡南洋理工大学	周天飞	ETH Zurich
周 毅	东南大学	朱 磊	University of Cambridge
朱霖潮	浙江大学	朱鹏飞	天津大学
朱盈盈	华中科技大学	邹常青	华为加拿大研究院
董胤蓬	清华大学	曹相湧	西安交通大学
陈冠英	香港中文大学(深圳)	陈 浩	香港科技大学
陈使明	美国卡耐基梅隆大学、 阿联酋人工智能大学	陈威华	阿里巴巴达摩院
程光亮	英国利物浦大学	崔兆鹏	浙江大学
代季峰	清华大学	丁明宇	加州大学伯克利分校
冯 婕	西安电子科技大学	傅雪阳	中国科学技术大学
高广谓	南京邮电大学	顾 实	电子科技大学
郭 青	新加坡科技研究局	郭宗辉	中国科学院计算技术研究所
胡庆拥	University of Oxford	黄文炳	中国人民大学高瓴人工智能学院
霍 静	南京大学	况 琨	浙江大学
李崇轩	中国人民大学	李 响	西安交通大学
刘 晗	大连理工大学	刘夏雷	南开大学计算机学院

刘瑶瑶	约翰霍普金斯大学	刘 勇	中国人民大学高瓴人工智能学院
陆 昊	华中科技大学	倪张凯	同济大学
屈靛琼	香港大学	任冬伟	哈尔滨工业大学
孙奕帆	百度	唐彦嵩	清华大学
田亚鹏	德克萨斯大学达拉斯分校	王 啸	北京航空航天大学
王一帆	大连理工大学	武 宇	武汉大学计算机学院
谢雨彤	澳大利亚阿德莱德大学	杨 帅	MMLab@南洋理工大学
杨 旭	东南大学计算机学院	弋 力	清华大学
易 冉	上海交通大学	于 茜	北京航空航天大学
元玉慧	微软亚洲研究院	张 力	复旦大学

**秘书处:**

朱盈盈	华中科技大学	程 一	中国科学院计算所
班瀚文	中国科学院计算所		

# 会场及酒店

## 会场位置

会议地点：无锡太湖国际博览中心  
交通地址：无锡滨湖区太湖新城清舒道 88 号



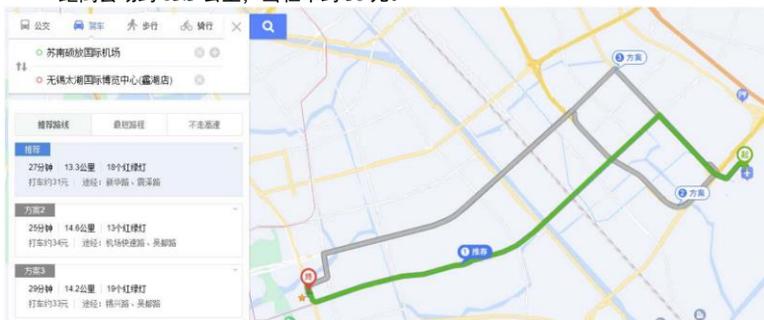
## 交通信息

### 交通情况概览



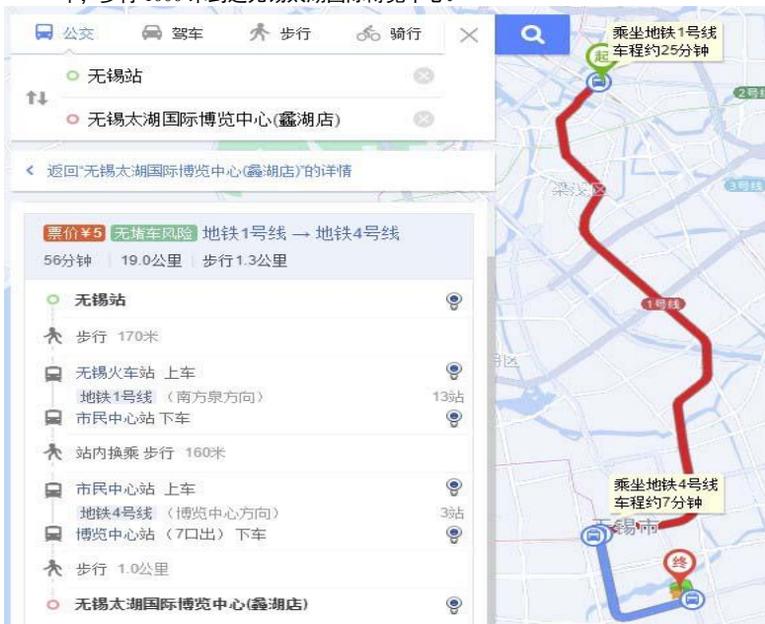
## 1. 苏南硕放国际机场

- **打车线路【全程估计 27 分钟】**  
距离会场约 13.3 公里，出租车约 36 元。



## 2. 无锡站

- **地铁线路【全程估计 56 分钟】**  
路线：无锡火车站 → 博览中心，5 元/人。  
起点：无锡火车站上车，乘坐地铁 1 号线（南方泉方向）到市民中心站换乘；  
换乘：市民中心站内换乘地铁 4 号线（博览中心方向），到博览中心站（7 口出）下车，步行 1000 米到达无锡太湖国际博览中心。

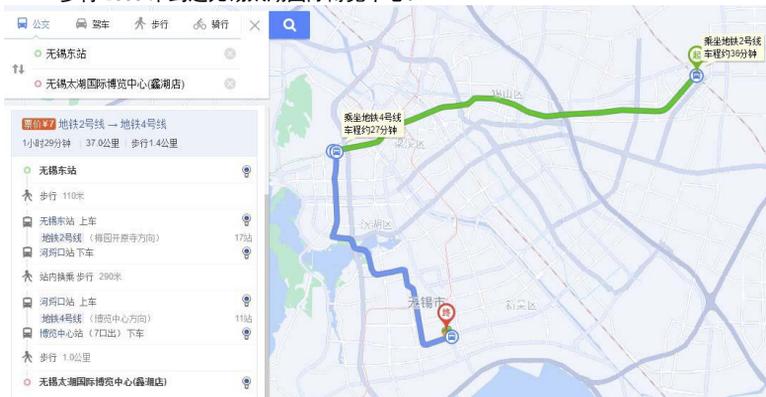


- ▶ **打车线路【全程估计 33 分钟】**  
距离会场约 16.5 公里，出租车约 50 元。



### 3. 无锡东站

- ▶ **地铁线路【全程估计 89 分钟】**  
路线: 无锡东站 → 博览中心, 7元/人。  
起点: 无锡东站上车, 乘坐地铁 3 号线 (梅园开原寺方向) 到河埭口站换乘;  
换乘: 河埭口站内换乘地铁 4 号线 (博览中心方向), 到博览中心站 (7 口出) 下车,  
步行 1000 米到达无锡太湖国际博览中心。



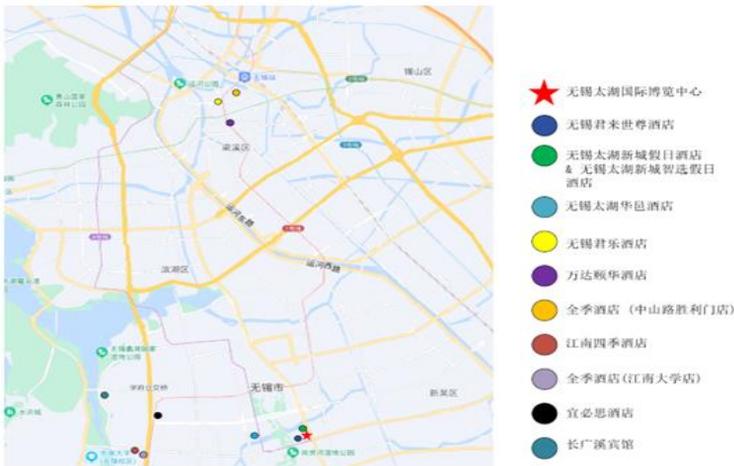
- ▶ **打车线路【全程估计 33 分钟】**  
距离会场约 22.7 公里，出租车约 65 元



## 酒店信息

### 大会协议酒店信息

为方便各位老师和同学参会，组委会前期调研了会场周边的酒店，步行范围内可达会场（无锡太湖国际博览中心）的有“无锡君来世尊酒店”、“无锡太湖新城假日酒店”、“无锡太湖新城智选假日酒店”三家酒店。另有豪华型酒店“无锡太湖华邑酒店”距会场 2000 米，需乘坐地铁 4 号线 2 站。其余酒店均位于地铁沿线和交通较为便利的地方，适宜选择。同时也非常建议参会者自行通过携程等公共平台自行搜索会场附近酒店。



组委会将在大会网站上公开电子版会议通知，方便各位参会者报销。酒店特惠订房协议价如下：

酒店名称	房型	酒店价格	酒店地址	酒店联系人	联系方式	备注
无锡君来世尊酒店	高级房（大床）	480，含单早	无锡市滨湖区和风路111号	王经理	18921122882	预订时间：周一到周五：早上8:30-下午16:30；预定需要说出优惠口令“Valse 2023”；协议优惠价有效期2023年6月9日到2023年6月12日；酒店步行300米即可到达会场
	高级房（双床）	480，含双早				
	园景房（大床）	550，含单早				
	园景房（双床）	550，含双早				
	湿地房（大床）	600，含单早				
	湿地房（双床）	600，含双早				
无锡太湖新城假日酒店 <i>（已订满）</i>	高级房（大床）	380，含单早	无锡市滨湖区清舒道77号	石琴	18896718872 0510-8052 6121	预定需要说出优惠口令“Valse 2023”；协议优惠价有效期2023年6月9日到2023年6月12日；酒店步行300米即可到达会场
	高级房（双床）	380，含双早				
	豪华房（大床）	430，含单早				
	豪华房（双床）	430，含双早				
	超豪华房（大床）	480，含单早				
	超豪华房（双床）	480，含双早				
无锡太湖新城智选假日酒店 <i>（已订满）</i>	标准房（大床）	280，含单早				
	标准房（双床）	280，含双早				
	高级房（大床）	330，含单早				
	高级房（双床）	330，含双早				
无锡太湖华邑酒店	高级房（大床）	490，含单早	无锡市经开区金融六街15号	张杰杰	13921658416	预定需要提供成功报名截图，以协议优惠价预定须在2023年5月20日之前；酒店乘坐两站地铁即可到达会场
	高级房（双床）	490，含单早				
	豪华房（大床）	590，含单早				
	豪华房（双床）	590，含单早				
	行政房（大床）	800，含单早				
	行政房（双床）	800，含双早				
无锡君乐酒店	高级大床房（大床）	280，含单早	江苏省无锡市梁溪区中山路401号	龚子懿	13151932901	早餐另加一份需要50元；预定需要说出优惠口令“Valse 2023”；协议优惠价有效期2023年6月9日到2023年6月12日；打车约28分钟。
	高级双床房（双床）	280，含单早				
	豪华大床房（大床）	320，含单早				
	豪华双床房（双床）	320，含单早				

万达颐华酒店	高级大床房 (大床)	380, 含单早	江苏省无锡市梁溪区中山路218号	顾经理	15061879360	预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约26分钟。
	高级双床房 (双床)	380, 含双早				
	豪华大床房 (大床)	430, 含单早				
	豪华双床房 (双床)	430, 含双早				
	颐华大床房 (大床)	480, 含单早				
全季酒店 (中山路胜利门店)	标准间 (双床)	360, 含双早	江苏省无锡市梁溪区解放东路1008号	吴文鹏	13222535250	6月10日当天价格需要额外加20元; 预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约29分钟。
	高级大床房 (大床)	360, 含单早				
江南四季酒店	公寓双床 (双床)	240, 含双早	江苏省无锡市滨湖区锦溪路100号1号楼	朱峰	18168878120	预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约13分钟。
	公寓套房 (双床)	328, 含双早				
	高级大床 (大床)	328, 含单早				
	高级双床 (双床)	328, 含双早				
	豪华大床 (大床)	388, 含单早				
	豪华双床 (双床)	388, 含双早				
全季酒店 (江南大学店)	大床房 (大床)	324.72, 含单早	江苏省无锡市滨湖区锦溪路100号	冯经理	13961802596	预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约13分钟。
	高级大床房 (大床)	351.12, 含单早				
	高级双床房 (双床)	377.52, 含双早				
宜必思酒店	高级大床房 (大床)	240, 含单早	江苏省无锡市滨湖区观山路100号星光广场173单元	戴俊彦	13861863201	预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约13分钟。
	商务大床房 (大床)	260, 含单早				
	双床房 (双床)	260, 含双早				
	家庭房 (双床)	300, 含双早				
长广溪宾馆	高级标准间 (双床)	300, 不含早	江苏省无锡市滨湖区缘溪道1号	王经理	13861850830	早餐每份30元; 预定需要说出优惠口令“VALSE 2023”; 协议优惠价有效期2023年6月9日到2023年6月12日; 打车约16分钟。
	高级大床房 (大床)	300, 不含早				

**请凭借 VALSE 2023 订房优惠口令(或报名成功截图)自行与酒店联系人预定, 费用自理, 先到先得。**



主办单位：中国人工智能学会 中国图象图形学学会

承办单位：江南大学 无锡国家高新技术产业开发区管理委员会

协办单位：江苏省人工智能学会 无锡市计算机学会 中国图象图形学学会青年工作委员会