

2024 YINCHUAN
CHINA MM

中国多媒体大会 China Multimedia 2024

宁夏·银川市 2024年7月24-26日

会议手册



2024 中国多媒体大会

由中国图象图形学学会（CSIG）、中国计算机学会（CCF）主办，CSIG 多媒体专业委员会、CCF 多媒体技术专业委员会承办，北方民族大学联合承办，宁夏大学、宁夏师范大学协办的中国多媒体大会（ChinaMM 2024），定于 2024 年 7 月 24-26 日在宁夏回族自治区银川市召开。本次大会将为多媒体领域的学术界和产业界提供前沿的技术发展动态、丰富多样的学术交流活动、全面前瞻的产业技术宣传平台，以广泛促进学术分享与交流、产业合作与互动，推动产学研联动发展，提升我国多媒体技术的研究与应用水平。

本次会议将邀请本领域国内外学术界和企业界的顶尖专家与学者做主题报告，分享最新的前沿技术与产业热点。会议期间，还将组织学术高峰论坛、前沿技术专题论坛、企业论坛等多角度、多层次的丰富活动，为参会者提供丰富技术盛宴，全方位地扩大交流范畴，创新交流形式，增强交流实效。

主办单位

中国图象图形学学会（CSIG）

中国计算机学会（CCF）

承办单位

CSIG 多媒体专业委员会

CCF 多媒体技术专业委员会

联办单位

北方民族大学

协办单位

宁夏大学

宁夏师范大学

▶ 目 录 ◀

中国多媒体大会组委会	1
会议须知	2
会议地点	3
会议日程概览	4
会议详细日程	7
特邀报告	17
企业特邀报告	25
专题论坛 1: 智能艺术与设计 AI Art & Design	27
专题论坛 2: 教育多模态大模型	30
专题论坛 3: 软硬协同的高能效视觉计算	33
专题论坛 4: 多模态大模型自适应学习	37
专题论坛 5: 三维点云压缩与表征学习	41
专题论坛 6: 面向自然科学与工业应用的多媒体智能分析技术	44
专题论坛 7: 面向开放复杂环境的跨时空行人关联	48
专题论坛 8: AIGC 与媒体内容安全	51
专题论坛 9: 沉浸式媒体技术赋能文旅产业	55
专题论坛 10: 面向具身智能的多模态感知与交互	59
专题论坛 11: 三维视觉理论、技术与应用	62
专题论坛 12: 面向机器的多媒体编解码技术: 从标准框架到评价指标	66
专题论坛 13: 面向视频图像计算的专用硬件处理器	70
专题论坛 14: 多模态大模型和可控生成	73
专题论坛 15: 未来视觉信息编码与智能技术	76
专题论坛 16: 面向跨媒体内容的智能可视化计算	80
专题论坛 17: 多媒体智能传播	84
专题论坛 18: 面向开放世界的可信多模态通用大模型研究	87
专题论坛 19: 多模态感知与行为协同的具身智能	91
专题论坛 20: AI4Food: 食品和健康计算	95
专题论坛 21: 三维生成大模型	99
专题论坛 22: 高保真数字人建模与驱动研究及其应用	103
专题论坛 23: 多模态情感脑机接口与应用	107
专题论坛 24: 面向开放环境的多媒体内容理解	110
钻石赞助专题论坛: 面向视觉智算的关键技术与应用	114
最佳论文评选论坛	118
ChinaMM2024 论文墙报展示 (英文)	119
ChinaMM2024 论文墙报展示 (中文)	121
会场平面图	124
赞助企业展宣	128
赞助企业展宣	129

中国多媒体大会组委会

大会荣誉主席

高文
北京大学

杨士强
清华大学

李波
北京航空航天大学

陈熙霖
中国科学院计算所

大会主席

黄华
北京师范大学

黄庆明
中国科学院大学

保文星
北方民族大学

程序委员会主席

姜育刚
复旦大学

徐迈
北京航空航天大学

吕科
中国科学院大学

组织委员会主席

刘家琰
北京大学

冯峰
宁夏大学

贺钧
北方民族大学

本地主席

马金林
北方民族大学

李卫军
北方民族大学

刘昊
宁夏大学

赞助主席

马思伟
北京大学

王涛
航天宏图

李泽超
南京理工大学

张新峰
中国科学院大学

论坛主席

孙立峰
清华大学

聂礼强
哈尔滨工业大学（深圳）

鲍秉坤
南京邮电大学

宣传主席

刘东
中国科学技术大学

许倩倩
中国科学院计算所

张飞飞
天津理工大学

评奖主席

刘青山
南京邮电大学

蒋树强
中国科学院计算所

洪日昌
合肥工业大学

出版主席

胡海苗
北京航空航天大学

陈智能
复旦大学

刘武
中国科学技术大学

财务主席

贾川民
北京大学

甘甜
山东大学

张越一
中国科学技术大学

张小伍
北方民族大学

► 会议须知 ◀

会议签到

- 7月23日 | 签到时间: 14:00-22:00 | 地点: 银川国际交流中心酒店大堂
- 7月24日 | 签到时间: 08:00-22:00 | 地点: 银川国际交流中心酒店大堂
- 7月25日 | 签到时间: 08:00-18:00 | 地点: 银川国际交流中心酒店大堂

注意事项

- 注意安全防范, 妥善保管好个人财物、资料, 休息或离开房间时务必锁好房门。
- 请各位嘉宾根据《会议手册》日程安排按顺序报告, 并关注临时通知。如遇到特殊情况不能按时到会, 请提前告知会务组, 以便于协调。
- 会场多媒体设备由会务组统一提供, 报告者可通过拷入 U 盘的方式, 提前备好演讲文稿电子版 (PPT/PDF) 用于试场、报告等; 每位演讲者报告时长包括演讲和提问交流时间。
- 本次会议凭借胸卡进入会场, 凭餐券用餐, 请随身携带。
- 遵守会场秩序, 会议开始前请将手机关闭或静音, 保持会场安静。
- 外出离开时请拿一张酒店名片, 便于返回。
- 身体不适者请及时与会务组人员联系, 以确保安全。

► 会议地点 ◀

银川国际交流中心

具体地址：银川市金凤区亲水北大街 222 号



交通信息

银川河东机场

打车 全程 41.5 公里，约 40 分钟。

公交 河东机场乘坐**机场巴士 1 号线**，在西港航空酒店下车后，步行至胜利街南塘巷口公交站，乘坐**公交 102 路**至国际会展中心公交站，换乘**公交 13 路**至亲水大街景观水道下车，步行 606 米到达银川国际交流中心。全程约 2 小时。

银川站

打车 全程 6.1 公里，约 13 分钟。

公交 火车西站乘坐**公交 52 路**，15 站后在银川国际交流中心东公交站下车，步行 9 分钟到达银川国际交流中心。全程约 52 分钟。

► 会议日程概览 ◀

7月23日 星期二

时间	日程	地点
14:00-22:00	会议注册	大堂

7月24日 星期三

时间	日程	地点
08:00-22:00	会议注册	大堂
09:00-17:00	论文墙报展示	黄河厅 前厅
时间	日程	地点
09:00-12:00	专题论坛：智能艺术与设计 AI Art & Design	多功能厅 1
	专题论坛：教育多模态大模型	多功能厅 2
	专题论坛：软硬协同的高能效视觉计算	多功能厅 3
	专题论坛：多模态大模型自适应学习	多功能厅 4
	专题论坛：三维点云压缩与表征学习	多功能厅 5
	专题论坛：面向自然科学与工业应用的多媒体智能分析技术	多功能厅 6
	钻石赞助专题论坛：面向视觉智算的关键技术与应用	贵宾厅 1
09:30-12:00	最佳论文评选	多功能厅 7
13:30-17:00	专题论坛：面向开放复杂环境的跨时空行人关联	多功能厅 1
	专题论坛：AIGC 与媒体内容安全	多功能厅 2
	专题论坛：沉浸式媒体技术赋能文旅产业	多功能厅 3
	专题论坛：面向具身智能的多模态感知与交互	多功能厅 4
	专题论坛：三维视觉理论、技术与应用	多功能厅 5
	专题论坛：面向机器的多媒体编解码技术：从标准框架到评价指标	多功能厅 6
19:00-20:00	CSIG 多媒体专业委员会 工作会议	贵宾厅 1
20:00-21:00	CCF 多媒体技术专业委员会 工作会议	贵宾厅 1

7月25日 星期四

时 间	日 程	地 点
08:00-18:00	会议注册	大堂
09:00-17:00	论文墙报展示	黄河厅 前厅
时 间	日 程	地 点
08:30-09:00	开幕式	黄河厅
09:00-09:50	大会特邀报告 报告人：郑庆华（同济大学）	
09:50-10:20	合影留念	
10:20-11:10	大会特邀报告 报告人：王耀南（湖南大学）	
11:10-12:00	大会特邀报告 报告人：高 文（北京大学、鹏城实验室）	
14:00-14:40	大会特邀报告 报告人：蔡达成（新加坡国立大学）	黄河厅
14:40-15:20	大会特邀报告 报告人：陈熙霖（中国科学院计算技术研究所）	
15:20-15:55	企业特邀报告 报告人：王苕社（博雅睿视）	
15:55-16:10	茶 歇	
16:10-16:50	大会特邀报告 报告人：李 波（北京航空航天大学）	
16:50-17:30	大会特邀报告 报告人：马华东（北京邮电大学）	
17:30-18:00	企业特邀报告 报告人：张 迪（快手科技）	
18:30-21:00	晚 宴	银川厅

7月26日 星期五

时 间	日 程	地 点
09:00-12:00	专题论坛：面向视频图像计算的专用硬件处理器	多功能厅 1
	专题论坛：多模态大模型和可控生成	多功能厅 2
	专题论坛：未来视觉信息编码与智能技术	多功能厅 3
	专题论坛：面向跨媒体内容的智能可视计算	多功能厅 4
	专题论坛：多媒体智能传播	多功能厅 5
	专题论坛：面向开放世界的可信多模态通用大模型研究	多功能厅 6
13:30-17:00	专题论坛：多模态感知与行为协同的具身智能	多功能厅 1
	专题论坛：AI4Food：食品和健康计算	多功能厅 2
	专题论坛：三维生成大模型	多功能厅 3
	专题论坛：高保真数字人建模与驱动研究及其应用	多功能厅 4
	专题论坛：多模态情感脑机接口与应用	多功能厅 5
	专题论坛：面向开放环境的多媒体内容理解	多功能厅 6

会议详细日程

7月24日 星期三

时间	日程	地点
08:00-22:00	会议注册	大堂
09:00-17:00	论文墙报展示	黄河厅 前厅
时间	日程	地点
09:00-12:00	专题论坛：智能艺术与设计 AI Art & Design 组织者： 牟伦田 副教授 北京工业大学 高峰 研究员 北京大学 王泽宇 助理教授 香港科技大学（广州） - 人工智能对艺术创作的影响：现状与未来（张军平 教授 复旦大学） - 生成式大模型与艺术创作的差距（杨熙 智能评测负责人 北京智源人工智能研究院） - 基于脑状态解码的多媒体用户满意度评估（邬霞 教授 北京理工大学） - 面向未来的智能叙事设计研究：具身交互叙事（郭子淳 教授 北京化工大学） - 智能穿搭：大模型时代的创新与赋能（赵健 研究员 中国电信人工智能研究院&西北工业大学） • 圆桌讨论	多功能厅 1
09:00-12:00	专题论坛：教育多模态大模型 组织者： 黄 华 教授 北京师范大学 刘 静 研究员 中国科学院自动化研究所 姜文斌 教授 北京师范大学 - 教育场景中以语言为核心的多模态大模型（文继荣 教授 中国人民大学） - 新一代人工智能工具对未来教育的影响和应对（贺樑 教授 华东师范大学） - 多模态教学大模型面临的挑战与应对措施（刘均 教授 西安交通大学） - 通用人工智能技术进展和教育典型应用（王士进 正高级工程师 科大讯飞） - 多模态大模型在数学领域的前沿应用与挑战（田密 CTO 好未来） • 圆桌讨论	多功能厅 2
09:00-12:00	专题论坛：软硬协同的高能效视觉计算（铂金赞助-AutoDL 冠名） 组织者： 鲁继文 教授 清华大学 陈 涛 研究员 复旦大学	多功能厅 3

	• 端侧软硬件协同高效计算 (程健 研究员 中国科学院自动化研究所) • 低比特深度神经网络量化 (刘祥龙 教授 北京航空航天大学) • 高效深度模型优化理论与技术 (谭明奎 教授 华南理工大学) • 面向边缘多模态计算的软硬协同优化探索 (李超 教授 上海交通大学) • 高效扩散模型及其应用 (唐彦嵩 特聘研究员 清华大学) • 破解用卡难复现难解决方案 (余佳 首席运营官 AutoDL) • 圆桌讨论	
09:00-12:00	专题论坛：多模态大模型自适应学习 组织者：常晓军 讲席教授 中国科学技术大学 刘 慧 教授 山东财经大学 朱 磊 研究员 同济大学 • 黑盒大模型与优化算法的协同方法 (左旺孟 教授 哈尔滨工业大学) • AI4Science-机遇与挑战 (欧阳万里 领军科学家 上海人工智能实验室) • 多模态大模型感知 (赫然 研究员 中国科学院自动化所) • 优化驱动学习及多模态感知应用 (刘日升 教授 大连理工大学) • 大模型驱动的智能文档处理:探索与应用 (丁凯 事业部副总经理 合合信息) • 天权遥感大模型关键技术与实践 (袁梦珂 算法研究员 航天宏图)	多功能厅 4
09:00-12:00	专题论坛：三维点云压缩与表征学习 组织者：李 革 教授 北京大学 高 伟 研究员 北京大学 • 点云数据获取：高分辨光场计算采集与重建 (金欣 教授 清华大学) • 点云压缩：从技术到标准 (许晓中 高级首席研究员 腾讯多媒体实验室) • 面向点云场景理解的跨域跨模态表征学习 (雷印杰 教授 四川大学) • 迈向高效且灵活的点云智能编解码 (高伟 研究员 北京大学) • 圆桌讨论	多功能厅 5
09:00-12:00	专题论坛：面向自然科学与工业应用的多媒体智能分析技术 组织者：聂 婕 特聘教授 中国海洋大学 王 鑫 副研究员 清华大学 费 豪 研究员 新加坡国立大学 纪家沂 博士后研究员 厦门大学 • 从 Science for AI 到 AI for Science：人工智能与基础科学的“双重奏” (田永鸿 博雅特聘教授 北京大学) • 水下光学高分辨率三维成像 (董军宇 教授 中国海洋大学) • 推荐大模型初探：面向用户行为的理解、对齐与模拟 (王翔 特任教授 中国科学技术大学) • 多模态语义分割及其应用 (丁恒辉 青年研究员 复旦大学)	多功能厅 6

	- 面向垂直领域的多模态事件预测 (马云山 博士后研究员 新加坡国立大学) - 图文大模型高效感知与计算 (罗根 博士 厦门大学)	
09:00-12:00	钻石赞助专题论坛：面向视觉智算的关键技术与应用 组织者：王苕社 副研究员 北京大学 韩巍 总经理 北京博雅睿视科技有限公司 - 智算中心建设经验分享及视频领域解决方案介绍 (田原 院长 中通服咨询设计研究院通信院) - 开放的生态-国产算力发展之路 (梅迪 创始人、CEO 广州希姆半导体科技有限公司) "道合声远 菁彩未来", 超高清产业生态发展分享 (王斌 产业研究总监 世界超高清视频产业联盟) - xkISP: 开源 ISP 处理器及应用 (范益波 教授 复旦大学) "视觉智能 2.0" 时代对算力挑战的思考和实践 (余世兵 解决方案负责人 北京凌川科技有限公司) - 用"芯"构建可信进化的视觉智算解决方案 (李俊 解决方案架构师 北京博雅睿视科技有限公司) - 圆桌讨论	贵宾厅 1
09:30-12:00	最佳论文评选	多功能厅 7
13:30-17:00	专题论坛：面向开放复杂环境的跨时空行人关联 组织者：李 波 特聘教授 北京航空航天大学 张永飞 教授 北京航空航天大学 - 域泛化行人再识别：挑战、算法和数据 (廖胜才 主任科学家 起源人工智能研究院) - 行人再识别新思考：统一模型和新评估标准 (郑伟诗 教授 中山大学) - 多模态行人检索与生成 (叶茫 教授 武汉大学) - 基于大模型的开放场景行人感知 (张史梁 长聘副教授 北京大学) - 以芯赋能边缘 AI, 打通 AI 应用最后一公里 (郑文先 副总裁 深圳云天励飞技术股份有限公司) - 圆桌讨论	多功能厅 1
13:30-17:00	专题论坛：AIGC 与媒体内容安全 组织者：卢 伟 教授 中山大学 钱振兴 教授 复旦大学 高 赞 教授 山东省人工智能研究院 程志勇 教授 合肥工业大学 程皓楠 副研究员 中国传媒大学 圆桌讨论主席：查正军 教授 中国科学技术大学 从智能计算到 AIGC 安全 (罗向阳 教授 解放军信息工程大学)	多功能厅 2

	• 大语言模型安全与隐私保护（王骞 特聘教授 武汉大学） • AI 大模型的安全与隐私风险（沈超 特聘教授 西安交通大学） • AIGC 生成内容检测与溯源技术（张卫明 教授 中国科学技术大学） • AI 视觉内容生成与安全对抗研究（董晶 研究员 中国科学院自动化研究所） • 圆桌讨论	
13:30-17:00	专题论坛：沉浸式媒体技术赋能文旅产业 组织者：马思伟 博雅特聘教授 北京大学 叶 龙 教授 中国传媒大学 张新峰 副教授 中国科学院大学 • 沉浸式赋能文旅新质生产力：逻辑、现状与前瞻（邓宁 教授 北京第二外国语学院） • 汉字字体与人工智能（郭宗明 教授 北京大学） • 虚实交互和 AIGC 的融合及在文旅行业的应用（穆凯辉 副总裁 蓝海彤翔集团） • 数字科技美学 打造地球的红飘带（柳启阳 副总经理 贵州省文化旅游科技有限公司） • 数字文旅的无限可能：探索元宇宙中应用的新场景（汤启鹏 总经理 宽堂（北京）文化发展有限公司） • 全息音频---手机对象音频服务框架探索与实践（林松 音频系统规划设计专家 OPPO） • 圆桌讨论	多功能厅 3
13:30-17:00	专题论坛：面向具身智能的多模态感知与交互 组织者：赵铁松 教授 福州大学 金 枝 副教授 中山大学 陈鸣锴 副教授 南京邮电大学 • 视觉与触觉的交融——面向复杂物体感知与抓取的柔性触觉技术（丁文伯 副教授 清华大学深圳国际研究生院） • 脑启发的触觉感知和学习（陈立翰 副教授 北京大学） • Taxel 触觉图像超分辨率重建方法（刘倩 教授 大连理工大学） • 面向多模态业务的视觉-触觉信号恢复技术（魏昕 教授 南京邮电大学） • 视触媒体信息融合与交互（韩腾 研究员 中国科学院软件研究所） • 圆桌讨论	多功能厅 4
13:30-17:00	专题论坛：三维视觉理论、技术与应用 组织者：马 展 教授 南京大学 马凯光 教授 南京航空航天大学 丁丹丹 副教授 杭州师范大学 • 三维场景的机器人自主探测与重建（刘利刚 教授 中国科学技术大学）	多功能厅 5

	• 三维场景精确重建：从图形学研究到 AI 赋能（郭延文 教授 南京大学） • 多源融合 SLAM-现状与挑战（刘勇 教授 浙江大学） • 局部自适应的三维信号处理（范晓鹏 特聘教授 哈尔滨工业大学） • 面向沉浸式媒体的实时渲染与轻量表征（廖依伊 特聘研究员 浙江大学） • 全息通信——未来的沟通方式（周席龙 研究员 联想研究院） • 圆桌讨论	
13:30-17:00	专题论坛：面向机器的多媒体编解码技术：从标准框架到评价指标 组织者：虞 露 求是特聘教授 浙江大学 陈志波 教授 中国科学技术大学 • 面向机器视觉的可控视频压缩编码（鲁国 助理教授 上海交通大学） • VCM 与 DCM 核心技术与标准进展（于化龙 讲师 浙江大学宁波科创中心） • 面向机器视觉的语义解耦与智能编码（金鑫 助理教授 宁波东方理工大学） • 面向机器视觉任务编码的评价指标（高长生 博士后研究员 新加坡南洋理工大学） • 智能任务驱动的媒体质量评价（胡姚姒 博士后研究员 香港理工大学）	多功能厅 6
19:00-20:00	CSIG 多媒体专业委员会 工作会议	贵宾厅 1
20:00-21:00	CCF 多媒体技术专业委员会 工作会议	贵宾厅 1

7月25日 星期四

时间	日程	地点
08:00-18:00	会议注册	大堂
09:00-17:00	论文墙报展示	黄河厅 前厅
时间	日程	地点
08:30-09:00	会议开幕式	黄河厅
09:00-09:50	大会特邀报告：从计算智能迈向认知智能 报告人：郑庆华（同济大学）	
09:50-10:20	合影留念	黄河厅
10:20-11:10	大会特邀报告：高光谱机器视觉感知技术及发展趋势 报告人：王耀南（湖南大学）	
11:10-12:00	大会特邀报告：鹏城系列大模型与数字视网膜应用实践 报告人：高 文（北京大学、鹏城实验室）	
14:00-14:40	大会特邀报告：From Large Foundation Models to Network of Experts 报告人：蔡达成（新加坡国立大学）	
14:40-15:20	大会特邀报告：面向行人辅助的场景理解与交互 报告人：陈熙霖（中国科学院计算技术研究所）	
15:20-15:55	企业特邀报告：视觉智算--过去，现在和将来 报告人：王苕社（博雅睿视）	
15:55-16:10	茶 歇	黄河厅
16:10-16:50	大会特邀报告：复杂场景视频图像处理 报告人：李 波（北京航空航天大学）	
16:50-17:30	大会特邀报告：从多媒体传感网到视频物联网：探索与展望 报告人：马华东（北京邮电大学）	
17:30-18:00	企业特邀报告：可灵 AI 视觉生成大模型及应用 报告人：张 迪（快手科技）	银川厅
18:30-21:00	晚 宴	

7月26日 星期五

时间	日程	地点
09:00-12:00	专题论坛：面向视频图像计算的专用硬件处理器 组织者：范益波 教授 复旦大学 - 超高速视频采集和编解码系统的开发（周金佳 副教授 日本法政大学） - 人工智能硬件处理器设计（徐科 首席科学家 中兴微电子） - 面向智能计算的多层次优化存储系统与应用（殷海兵 教授 杭州电子科技大学） - 大规模图像处理架构演进及硬件加速业务分享（王文强 阿里云服务器研发资深技术专家 阿里云） - 面向新一代标准的视频编码处理器硬件架构（范益波 教授 复旦大学）	多功能厅 1
09:00-12:00	专题论坛：多模态大模型和可控生成 组织者：李 玺 求是特聘教授 浙江大学 李 辉 讲师 江南大学 - 三维生成：从类别限定到开放世界（韩晓光 助理教授 香港中文大学（深圳）） - 多模态引导的可控三维数字人生成（晏轶超 助理教授 上海交通大学） - 多模态大模型在有限样本场景下的泛化（于云龙 研究员 浙江大学） - 基于多模态大模型的长视频理解（王高昂 研究员 浙江大学）	多功能厅 2
09:00-12:00	专题论坛：未来视觉信息编码与智能技术（铂金赞助-快手冠名） 组织者：翟广涛 特聘教授 上海交通大学 朱树元 教授 电子科技大学 贾川民 助理教授 北京大学 - 沉浸式媒体智能技术与应用（马思伟 博雅特聘教授 北京大学） 3D 视频智能编码与处理方法（雷建军 讲席教授 天津大学） - 面向端到端视频编码的帧间预测技术（李礼 特任教授 中国科学技术大学） - 未来视觉应用与技术发展（郭勐 技术经理 中国移动研究院） - 机器视觉编码国际国内标准进展（张园 副所长 中国电信研究院） - 智能编解码前沿探索与形态（谷嘉文 快手音视频算法专家 快手）	多功能厅 3
09:00-12:00	专题论坛：面向跨媒体内容的智能可视计算 组织者：赵一凡 副教授 北京航空航天大学 蒋 铎 副教授 北京航空航天大学 朱 林 助理教授 北京理工大学 - 细粒度多模态运动分析（彭宇新 博雅特聘教授 北京大学） - 多模态视觉结构学习（李玺 求是特聘教授 浙江大学） - 生成式 AI 的缘起、发展与挑战（邓成 教授 西安电子科技大学）	多功能厅 4

	• 人的生物特征生成与可视化 (贾伟 教授 合肥工业大学) • 数据视角下的智能图像降噪 (王立志 准聘教授 北京理工大学) • 圆桌讨论	
09:00-12:00	专题论坛: 多媒体智能传播 组织者: 刘安安 教授 天津大学 靳国庆 研究员 人民日报社传播内容认知全国重点实验室 • 大模型驱动的引导式推荐初探 (冯福利 特任教授 中国科学技术大学) • 通用人工智能的安全、对齐与治理 (田天 创始人及首席执行官 瑞莱智慧 RealAI 公司) • 基于社交机器人的多媒体传播安全可控 (王岚君 研究员 天津大学) • 多模态对话与生成大模型技术前沿介绍与思考 (魏龙辉 华为终端云高级研究员 华为技术有限公司)	多功能厅 5
09:00-12:00	专题论坛: 面向开放世界的可信多模态通用大模型研究 组织者: 吴建龙 副教授 哈尔滨工业大学 (深圳) 张 正 副教授 哈尔滨工业大学 (深圳) • 多模态大模型: 技术挑战与行业应用实践 (田永鸿 博雅特聘教授 北京大学) • 机器学习的数学技术 (孟德宇 教授 西安交通大学) • 视觉表征模型的结构设计与物理启发 (叶齐祥 教授 中国科学院大学) • 多模态视频理解基础模型 InternVideo (王利民 教授 南京大学) • Multimodal Fusion on Low-quality Data (张长青 教授 天津大学) • 圆桌讨论	多功能厅 6
13:30-17:00	专题论坛: 多模态感知与行为协同的具身智能 组织者: 蒋树强 研究员 中国科学院计算技术研究所 宋新航 副研究员 中国科学院计算技术研究所 • 具身感知场景中的反思学习 (郑锋 副教授 南方科技大学) • 基于多模态大模型的智能感知与决策 (范鹤鹤 百人计划研究员 浙江大学) • 不确定性下的人机交互策略规划 (蔡盼盼 副教授 上海交通大学) • 多模态大模型驱动的开放世界具身感知与交互 (弋力 助理教授 清华大学) • 面向通用机器人的具身多模态大模型系统 (王鹤 助理教授 北京大学) • 圆桌讨论	多功能厅 1
13:30-17:00	专题论坛: AI4Food: 食品和健康计算 组织者: 闵巍庆 副研究员 中国科学院计算技术研究所 陈静静 副教授 复旦大学 石吉勇 特聘教授 江苏大学 圆桌讨论主席: 明朝燕 副教授 浙大城市学院	多功能厅 2

	卢 剑 教授 宁夏医科大学 - 基于图像分析的奶粉品质在线检测技术研究 (崔晓晖 特聘教授 江南大学) - 农产品无损检测与智能计算技术 (邹小波 教授 江苏大学) - 面向营养健康的食品计算和多模态大模型 (蒋树强 研究员 中国科学院计算技术研究所) - 基于组学数据的精准营养配方设计 (张烽 教授 江南大学) - 结合垂域大模型和多智能体框架的营养健康场景赋能 (魏忠钰 副教授 复旦大学) - 圆桌讨论	
13:30-17:00	专题论坛：三维生成大模型 组织者：白 静 教授 北方民族大学 聂为之 教授 天津大学 - 基于神经辐射场与 3D 高斯的三维生成与重建 (易冉 助理教授 上海交通大学) - 基于超图计算的 3D 内容理解 (高跃 长聘副教授 清华大学) - 基于大模型的多物体三维场景生成 (李曼祎 副研究员 山东大学) - 基于流模型的三维点云高质量生成 (毛爱华 教授 华南理工大学) - 视频内容生成与理解 (陈磊 助理研究员 清华大学) - 面向数据高效的 3D 场景理解 (赵娜 助理教授 新加坡科技设计大学) - 圆桌讨论	多功能厅 3
13:30-17:00	专题论坛：高保真数字人建模与驱动研究及其应用 组织者：吕 科 特聘教授 中国科学院大学 杨 易 求是讲席教授 浙江大学 张举勇 教授 中国科学技术大学 3D 高斯数字人 (刘烨斌 长聘教授 清华大学) - 多模态融合趋势下数字人的机遇与挑战 (林惊 教授 中山大学) - 单目多人人体 3D 模型重建关键技术 (刘武 特任教授 中国科学技术大学) - 知识驱动的数字人重建和生成技术 (朱霖潮 百人计划研究员 浙江大学) - 可重光照的 4D 数字人重建 (黄石生 副教授 北京师范大学) - 基于扩散模型先验的三维数字人生成与编辑 (郭玉东 特任副研究员 中国科学技术大学) - 数字人在真实场景的具体应用 (王金宝 助理教授 深圳大学)	多功能厅 4
13:30-17:00	专题论坛：多模态情感脑机接口与应用 组织者：郑文明 教授 东南大学 黄晓华 教授 南京工程学院	多功能厅 5

	• 基于信息论学习的类脑计算与脑机接口 (陈霸东 教授 西安交通大学) • 基于脑成像技术的大脑功能激活状态研究 (吴奇 教授 上海交通大学) • 多媒体情感计算 (赵思成 副研究员 清华大学) • 多模态表征学习及其在大模型中的应用 (陈杰 副教授 北京大学) • 多源信息融合的情感识别与抑郁症评估 (蒋冬梅 研究员 鹏城实验室) • 圆桌讨论	
13:30-17:00	专题论坛：面向开放环境的多媒体内容理解 组织者：叶齐祥 特聘教授 中国科学院大学 杨小汕 副研究员 中国科学院自动化研究所 谢国森 教授 南京理工大学 圆桌讨论主席：谭明奎 教授 华南理工大学 • 开放环境下无人驾驶预测与决策技术研究 (李长升 教授 北京理工大学) • 开放场景下的视频行为理解 (舒祥波 教授 南京理工大学) • 领域自适应开放识别 (李晶晶 百人计划研究员 电子科技大学) • 区域级多模态模型研究 (万方 长聘教轨助理教授 中国科学院大学) • 通义 mPLUG 多模态对话大模型技术 (严明 NLP 高级算法专家 阿里巴巴通义实验室) • 圆桌讨论	多功能厅 6

▶ 特邀报告 ◀

主持人：黄华 单位：北京师范大学

时 间

2024 年 7 月 25 日 上午 09:00-09:50

地 点

黄河厅



郑庆华

中国工程院院士、同济大学校长

从计算智能迈向认知智能

嘉宾介绍

郑庆华，中国工程院院士，同济大学校长（副部长级）、党委副书记。智能网络与网络安全教育部重点实验室主任，教育部科技委学部委员，教育部大学计算机教学指导委员会主任，国家杰出青年基金获得者，国家自然科学基金创新群体负责人。主要从事大数据知识工程、网络舆情监测等方面的研究，获国家科技进步二等奖 3 项，国家教学成果一等奖 1 项、二等奖 2 项，省部级科技进步一等奖 6 项，获 2022 年度何梁何利基金“科学与技术进步奖”，中国自动化学会科技进步特等奖，并先后获得中国科协“求是”杰出青年奖、中国青年科技奖、国务院政府特殊津贴专家、宝钢优秀教师特等奖、全国高等学校优秀骨干教师、全国信息产业科技创新先进工作者等荣誉。

▶ 特邀报告 ◀

主持人：黄庆明 单位：中国科学院大学

时间

2024年7月25日 上午 10:20-11:10

地点

黄河厅



王耀南

中国工程院院士、湖南大学教授
中国图象图形学学会理事长

高光谱机器视觉感知技术及发展趋势

嘉宾介绍

王耀南，中国工程院院士，机器人技术与智能控制专家，湖南大学教授，机器人视觉感知与控制技术国家工程研究中心主任。任中国科协委员、中国图象图形学学会理事长、中国自动化学会会士、中国计算机学会会士、中国人工智能学会会士、全国智能机器人创新联盟副理事长、国家自然科学基金委员会专家咨询委员、中国自动化学会常务理事、中国人工智能学会监事、教育部科技委人工智能与区块链技术委员会委员等。曾任国家 863 计划智能机器人领域专家、欧盟第五框架国际合作重大项目首席科学家。

长期从事机器人感知与控制技术教学科研工作，成果获国家技术发明二等奖 1 项、国家科技进步二等奖 4 项、何梁何利基金科学与技术进步奖、国际 IEEE 机器人与自动化领域“工业应用最高奖”，省部级一等奖 12 项。发表国际 IEEE 等 SCI 论文 200 余篇，出版机器人感知与智能控制等著作 15 部，获国家发明专利 90 余项，培养博士 80 余名。荣获国家百千万工程人才、德国杰出洪堡学者、全国高等学校优秀教师、全国五一劳动奖章、全国先进工作者、全国创新争先奖、全国教材建设先进个人等荣誉称号。

报告摘要

机器视觉作为机器人感知系统的“高精度眼睛”，其发展对机器人起着重要的作用。高光谱机器视觉具备多模态成像系统与智能分析处理算法，是当前机器视觉领域的研究前沿。本报告将从研究背景与意义、研究现状与面临挑战、关键技术等方面介绍高光谱机器视觉的现实发展与未来趋势。首先，以高端制造过程中面临的机器人感知手段有限、测量检测精度低、缺陷样本数量少等挑战为牵引，介绍了高光谱机器视觉研究中高速高精快照式光谱成像、无监督异常检测方法、跨场景模型边缘端部署等方面；最后介绍了高光谱机器视觉未来的发展趋势与展望。

▶ 特邀报告 ◀

主持人：黄庆明 单位：中国科学院大学

时间

2024年7月25日 上午 11:10-12:00

地点

黄河厅



高 文

中国工程院院士、北京大学博雅讲席教授
信息与工程科学部主任、鹏城实验室主任

鹏城系列大模型与数字视网膜应用实践

嘉宾介绍

高文，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。

报告摘要

GPT 为代表的 AI 通用大模型已经对人类经济生活产生了巨大的影响。基于鹏城云脑先进自主 E 级智能算力平台，鹏城实验室研发了参数量达 20 亿的鹏城·大基视觉与多模态系列基础模型，训练完成了鹏城·脑海自然语言大模型底座。3 亿参数视觉基础模型与 2.9 亿参数多模态基础模型已开源。此外，还构建了首个判别生成一体多模态基础模型，同时支持判别与生成任务；研发了增量式多模态基础模型，首次具备模型对新模态的持续扩展能力。视觉与多模态基础模型结合实验室牵头制定的全球首个端边云协同技术国际标准，支撑了“数字视网膜”计算架构，在全国多个城市的智能交通、综合治理等重点领域应用，取得了显著社会效益。鹏城·脑海是国内首个完全自主可控、安全可控、开源开放的自然语言预训练大模型底座，参数级别达到 2000 亿，并且输出内容符合中文核心价值观；鹏城实验室已经开放脑海模型合作，与外部伙伴一起建设大模型的垂直领域应用。

► 特邀报告 ◀

主持人: 姜育刚 单位: 复旦大学

时间 2024年7月25日 下午 14:00-14:40

地点 黄河厅



蔡达成 Chua Tat-Seng

新加坡国立大学讲席教授

From Large Foundation Models to Network of Experts

嘉宾介绍

Dr. Chua is the KITHCT Chair Professor at the School of Computing, National University of Singapore (NUS). He is also the Distinguished Visiting Professor of Tsinghua University, the Visiting Pao Yue-Kong Chair Professor of Zhejiang University, and the Distinguished Visiting Professor of Sichuan University. Dr. Chua was the Founding Dean of the School of Computing from 1998-2000. His main research interests include unstructured data analytics, video analytics, conversational search and recommendation, and robust and trustable AI. He is the co-Director of NExT, a joint research Center between NUS and Tsinghua University.

Dr. Chua is the recipient of the 2015 ACM SIGMM Achievements Award, and the winner of the 2022 NUS Research Recognition Award. He is the Chair of steering committee of Multimedia Modeling (MMM) conference series, and ACM International Conference on Multimedia Retrieval (ICMR) (2015-2018). He was the General Co-Chair of ACM Multimedia 2005, ACM SIGIR 2008, ACM Web Science 2015, ACM MM-Asia 2020, WSDM 2023, and TheWebConf (or WWW) 2024. He serves in the editorial boards of several international journals. Dr. Chua is the co-Founder of two technology startup companies in Singapore.

报告摘要

The emergence of Large Foundation Models (LFM's) that offer significant capabilities in content comprehension, generation, and flexible human-level dialogues, has revolutionized the way we seek and consume information. At sufficiently large scale, universal LFMs possess the strengths of domain versatility, output diversity, as well as semantic coherency and alignment at human-level. But being too versatile, diverse, large and open-domain come at the price of requiring high cost for operation and hallucination with the accompanying problems of trust and safety. There is a need to develop efficient solutions that can retain most of the strengths of LFMs while being safe, trustable and accurate in solving task-specific problems. This talk presents recent research and trends of LFMs at

two levels. The first is at the media level, by analyzing recent advancements in aligning text and multimodal contents to achieve higher quality multimodal understanding and generation, as well as research towards multimodal alignment, instruction tuning and reliable & faithful LFM. At the next level, we discuss current trends towards agent-based systems, and present our proposal to advance the agent-based models towards the network of experts (NoE) framework. The NoE will be task specific, for tasks such as recommendation, event forecasting and Fintech etc., while domain agnostic. It will support a network of small-sized LFMs or experts with diverse capabilities; some with high performance and with high level of trust and safety on domain specific tasks; while others are domain agnostic and with high versatility and diversity. These experts can be configured dynamically to collaboratively tackle more complex problems beyond what each is capable of doing. As NoE requires only smaller sized LFMs tuned to solving various tasks, they can be undertaken by smaller academic and industry labs, and thus offering the possibility of democratizing the core of LFM research to all.

► 特邀报告 ◀

主持人：姜育刚 单位：复旦大学

时间

2024年7月25日 下午 14:40-15:20

地点

黄河厅



陈熙霖

中国科学院计算技术研究所研究员

面向行人辅助的场景理解与交互

嘉宾介绍

陈熙霖，中国科学院计算技术研究所研究员，ACM / IAPR / IEEE Fellow，中国计算机学会会士，其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交互，特别是在生物启发的特征表示与学习、以人为中心的感知与分析、场景理解等方面开展了系统的研究。目前担任 JVCIR 的 Senior AE，计算机学报、模式识别与人工智能和 VRIH 的副主编，以及多个刊物的(领域)编委。担任(过)多个国内外大会的主席和程序委员会主席，并十多次担任 CVPR / ICCV / ECCV / NeurIPS 等会议的领域主席。研究工作在国内外重要刊物和会议上发表论文 400 多篇，先后获得过国家自然科学二等奖、北京市高等学校教学名师奖。

报告摘要

视障者的出行受交通基础设施等多方面的影响。面向特定人群在半结构化道路上的出行需求，发展相应的理解和交互技术，改善出行条件具有重要的作用。报告分析了行人辅助的半结构化道路环境理解与交互问题的特点与挑战，并将介绍在相应方面开展的研究工作，包括对场景感知与理解、综合多传感信息的定位与分析、高效的人机交互等。相关的技术可以拓展至支持非结构化下的移动装置，具有很好的泛化性。报告最后对未来的工作进行了展望。

▶ 特邀报告 ◀

主持人：吕科 单位：中国科学院大学

时间

2024年7月25日 下午 16:10-16:50

地点

黄河厅



李 波

北京航空航天大学长江学者特聘教授

复杂场景视频图像处理

嘉宾介绍

李波，北京航空航天大学计算机学院长江学者特聘教授、杰青，北航人工智能研究院常务副院长。兼任国务院学位委员会软件工程学科评议组成员、教育部人工智能科技创新专家组工作组副组长、军委装备发展部人工智能装备应用基础技术专家组成员。

当前主要研究方向为计算机视觉、机器学习、知识推理、嵌入式智能系统。已主持国家、省部级课题 40 余项，是国家重点研发计划项目“公共安全监控视频安全共享与特征分析关键技术研究”项目负责人、国家 973 计划项目“数字媒体理解的理论与方法研究”首席科学家，原总装“十五”、“十二五”某边海防视频系统型号总设计师。在本领域重要期刊和国际学术会议发表学术论文 100 余篇，有国内外发明专利 100 余件，获国家技术发明二等奖 1 项、国家科技进步二等奖 1 项。

报告摘要

人类获取的感知信息 80%来自视觉，视频图像处理在经济社会发展、国家安全保障中发挥着极其重要的作用。在复杂场景中，受天气复杂、环境复杂、目标变化复杂等因素的影响，需要解决恶劣天气获取视频“看不清”、复杂场景目标识别“认不准”、跨时空目标“关联难”等挑战问题。

在本报告中，建立了基于非均匀光照模型的降质图像清晰还原的理论方法，提出了融合时空信息的复杂场景目标准确识别技术，发明了基于多视角融合的跨时空目标关联技术，相关成果已在边海防侦察监视、公安视频侦查、弹载红外制导、卫星图像判读等方面规模化应用，解决了重大工程应用难题。

► 特邀报告 ◀

主持人：吕科 单位：中国科学院大学

时间

2024年7月25日 下午 16:50-17:30

地点

黄河厅



马华东

北京邮电大学教授、校学术委员会副主任

从多媒体传感网到视频物联网：探索与展望

嘉宾介绍

马华东，教授，北京邮电大学学术委员会副主任。2010-2021 年历任计算机学院执行院长、网络技术研究院执行院长等职务。2009 年获得国家杰出青年科学基金，2010 年被聘为国家 973 计划项目“物联网体系结构基础研究”首席科学家，2011 年度长江学者特聘教授，2019 年入选国家基金委创新群体“物联网基础理论与关键技术”项目负责人，2023 年度中国计算机学会王选奖获得者。兼任国务院学位委员会学科评议组成员，教育部科技委委员；中国计算机学会常务理事，曾任物联网专委会主任、多媒体专委会副主任；中国图象图形学学会常务理事，曾任多媒体专委会副主任；中国人工智能学会常务理事、副秘书长，IEEE/CCF/CAAI 会士。长期从事物联网、多媒体、人工智能领域的研究，在多媒体传感网、新型感知技术、视频物联网、媒体智能处理算法等方面取得一批有国际影响力的原创或创新性成果，获得授权国家发明专利 80 余项，发表论文 400 余篇，论著被同行学者引用 1.5 万次。排名第 1 获得国家级教学成果二等奖（2 次）、教育部自然科学一等奖、中国电子学会科学技术奖一等奖等奖励；获 IEEE Transactions on Multimedia 年度最佳论文奖等国际学术奖励。

报告摘要

本报告首先回顾了物联网发展长期面临的挑战，然后介绍了我们在多媒体与物联网交叉方向研究进行的若干探索，特别介绍了最近在视频物联网网络架构、移动视频智能传输、云边端协同视频计算等相关领域研究进展。人工智能的发展将驱动物联网的变革，本报告展望了人工智能时代的物联网发展一些开放的问题，如多模态类人感知、内生智能网络、智能服务生态，上述问题的突破将推动物联网技术的深层次发展。

► 企业特邀报告 ◀

主持人：姜育刚 单位：复旦大学

时间

2024年7月25日 下午 15:20-15:55

地点

黄河厅



王苦社

北京大学计算机学院副研究员
北京博雅睿视科技有限公司创始人

视觉智算--过去，现在和将来

嘉宾介绍

研究兴趣主要包括视频编码与质量评价、智能视频处理等。深度参与了视频编码国际标准和国家标准的制定工作，提交 200 余项技术提案；主持多项国家级重点项目和自然科学基金项目；作为主要完成人，获得 2019 年中国电子学会技术发明特等奖，2020 年度国家技术发明一等奖，2022 年度中国电影电视学会科技进步一等奖，2022 年度中国电子学会创新团队，2022 年度北京市科技进步一等奖。

报告摘要

随着人工智能技术的快速发展，全视频时代随之到来，视频与视觉技术越来越多的在行业发展中扮演核心角色。本报告着重回顾了视觉智算的内涵，发展现状，以及未来的发展趋势。同时随着技术本身的不断成熟，越来越多的硬件设备都支持了 AI 算力，充分发挥 AI 算力在行业应用中的价值，对于行业的可持续发展具有至关重要的作用。本报告结合首颗支持 AVS3 编码芯片的架构设计和研发过程，阐述了视觉智算的关键要素和未来发展趋势。

► 企业特邀报告 ◀

主持人：吕科 单位：中国科学院大学

时间

2024年7月25日 下午 17:30-18:00

地点

黄河厅



张 迪

快手科技副总裁

大模型与多媒体技术团队负责人

可灵 AI 视觉生成大模型及应用

嘉宾介绍

张迪，2003~2010年，在上海交通大学取得计算机系本科和硕士学位。曾任阿里巴巴集团资深技术专家，在阿里集团搜推广业务进入深度学习时代的技术演变中发挥了核心作用。现任快手科技副总裁，大模型与多媒体技术团队负责人，带领团队负责生成式 AI 大模型的研发和应用，构建了超大规模 AI 基础设施，研发了快意大语言模型，可图文生图、可灵视频生成大模型等，并将大模型与多媒体技术用于快手的核心业务中。

报告摘要

随着人工智能技术的飞速发展，大模型已成为人工智能领域的一个重要研究方向和推动行业发展的关键力量。本报告全面探讨了快手公司在这一领域的最新进展，包括可图文生图模型、可灵文生视频模型、语音大模型等关键技术，以及在 AIGC 等关键应用方向。通过深度分析和案例研究，报告揭示了这些技术如何有效地增强平台内容的丰富性、互动性和个性化，同时显著提高运营效率和用户体验。此外，报告也讨论了这些技术所带来的挑战，以及未来发展趋势，旨在为业界提供深入见解和指导，推动行业的持续进步和发展。

► 专题论坛 ◀

专题论坛 1：智能艺术与设计 AI Art & Design

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 多功能厅 1

组织人

- 牟伦田 (北京工业大学 副教授)
- 高峰 (北京大学 研究员)
- 王泽宇 (香港科技大学 (广州) 助理教授)

论坛介绍

随着 Sora、Suno 的兴起，人工智能大模型引发了艺术界、科技界和普通大众的极大震动与广泛关注。人类艺术创作将被 AI 取代吗？人机如何高效协同创新？人类的常识、情感、审美、价值如何与 AI 创作、设计结合？如何让 AI 在艺术创作或产品设计中迸发灵感？从脑电解码角度如何评估用户对视频服务的满意度？人工智能与具身交互技术的融合将如何改变设计范式？大模型与生成式人工智能如何赋能智能穿搭？这些问题需要学术界、产业界专家与人工智能、艺术、设计领域的从业者一起自由思辨和研讨。

报告嘉宾

嘉宾介绍



张军平 复旦大学

张军平，复旦大学计算机科学技术学院教授、博士生导师，中国自动化学会普及工作委员会主任。主要研究方向包括人工智能、机器学习、图像处理、生物认证、智能交通及气象预测。至今发表论文 100 余篇，连续两年（2022、2023）入选全球前 2% 顶尖科学家榜单终身科学影响力排行榜。著有《人工智能极简史》《爱犯错的智能体》《高质量读研》，主编《人机混合增强智能》，译著《统计学习要素》（第二版）。其中《爱犯错的智能体》曾获中国科普创作领域最高奖，即中国科普作家协会第六届优秀科普图书金奖；《人工智能极简史》获第 19 届文津奖提名图书，并入选“中国好书”六一专榜。

报告题目

人工智能对艺术创作的影响：现状与未来

报告摘要：近年来人工智能发展迅猛，尤其在艺术创作领域，出现了 Midjourney、Pika、Sora、Suno 等一系列多模态的应用，从其效果来看大有替代艺术创作者的趋势。那么，在此形势下，人工智能研究者 and 艺术创作者应该如何应对、适应和突破呢？我将在此报告里，介绍人工智能在艺术创作方面的现状、影响及对未来的思考。

嘉宾介绍



杨 熙 北京智源人工智能研究院

杨熙，北京智源人工智能研究院智能评测负责人，北京大学博士，中国科学院计算所博士后。主要从事大模型评测方面的研究，主持建设大模型评测平台“天秤”，已经对国内外百余个主流大模型进行了评测；在主客观自动批改、口语能力评测及诊断等方面有深入的研究和成功的产品化经验。多篇研究成果发表于 IJCAI/ACM 等领域顶会顶刊，承担或参与多项国家级重大项目。

报告题目

生成式大模型与艺术创作的差距

报告摘要：随着生成式大模型如 GPT4、Sora 等的兴起，创意产业见证了技术的强大潜力与初步影响。这些模型虽然在艺术创作如绘画、文学中展示出惊人的能力，却仍存在与真正艺术创作的差距。在本次报告中，将基于大模型发展的最新突破，通过对生成式模型的能力分析与评价，分析相关问题和挑战，并展望未来。人机协作的艺术形式将成为探索的重点，通过结合人类的创造力与大模型的能力，共同推动艺术的发展。

嘉宾介绍



邬 霞 北京理工大学

邬霞，北京理工大学计算机学院教授，主要研究方向为人工智能与脑科学，致力于运用人工智能技术理论和方法，挖掘人脑特征、探索认知功能的脑机制，并受脑机制启发，研究新型类脑人工智能算法。以第一/通讯作者在高水平期刊及顶级国际会议发表论文 100 余篇。2023 年获国家杰出青年科学基金；2021 年获茅以升科学技术奖-北京青年科技奖；2020 年获吴文俊人工智能科学技术奖自然科学一等奖、教育部高等学校科学研究优秀成果

奖一等奖。

报告题目

基于脑状态解码的多媒体用户满意度评估

报告摘要：多媒体服务不仅可以传播信息，更能为用户提供情感价值和社交功能，其呈现形式直接影响到人们的认知和情感状态，进而影响用户满意度。用户的满意程度高度依赖于人的主观感知，对于优化多媒体服务非常重要。脑电（EEG）信号包含丰富的脑活动信息，因其在脑状态解码方面的出色表现越来越受到研究者的关注。此次报告将介绍利用 EEG 对多媒体服务进行用户满意程度评价的脑状态解码相关工作，以及脑科学为多媒体业务优化和发展提供的神经生理学科学依据。在此基础上，进一步讨论脑状态解码研究与多媒体艺术创作领域相结合的发展潜力，以及未来可能面临的挑战。

嘉宾介绍



郭子淳 北京化工大学

郭子淳，北京化工大学教授，主要研究方向为未来影像与智能媒体设计、具身交互叙事设计等。入选北京化工大学百人青年英才计划。清华大学未来实验室客座研究员，北京交叉科学学会设计创新专业委员会委员，北京市智慧广电重点实验室副主任，CCF 人机交互专委会、虚拟现实专委会执行委员。主持国家社会科学面上、教育部青年等基金。多篇论文入选教育部高等学校文科学术文摘、人大复印资料，相关成果被中宣部及广电总局采纳。电影作品获第 13 届澳门国际电影节最佳导演、第 11 届“学院奖”二等奖、伦敦猎鹰国际电影节最佳惊悚片等。

报告题目

面向未来的智能叙事设计研究：具身交互叙事

报告摘要：当今，人工智能与具身交互技术的融合正处于设计学范式的临界点。研究认为，具身智能技术正在改变未来的叙事形态，以具身交互技术为核心的具身叙事设计开始浮现，围绕“身体观”的交互感知已经成为未来智能叙事设计的重要表征，传统的叙事设计研究开始从传统的“离身叙事”转向“具身设计”，这为未来智能设计的研究路径和学科交叉提供了全新的思路——具身交互叙事设计。

嘉宾介绍



赵 健 中国电信人工智能研究院、西北工业大学

赵健，中国电信人工智能研究院多媒体认知学习实验室（EVOL Lab）负责人、青年科学家，西北工业大学光电与智能研究院研究员，博士毕业于新加坡国立大学。发表 CCF-A 类论文 40 余篇，一作代表作包括 2 篇 T-PAMI、3 篇 IJCV 等，技术成果应用于百度、蚂蚁金服等 6 个科技行业领军企业。入选中国科协及北京市科协“青年人才托举工程”，主持 JKW 某特区项目、国家自然青年科学基金项目 6 项。曾获 2023 中国人工智能学会吴文俊人工智能优秀青年奖、2022 中国人工智能学会吴文俊人工智能自然科学奖一等奖（2/5）、CCF-A 类会议 ACM MM'18 唯一最佳学生论文奖（一作 1/208），7 次在国际重要科技赛事夺冠。

报告题目

智能穿搭：大模型时代的创新与赋能

报告摘要：近年来人工智能技术快速发展，从早期 AI 的衍生，到深度学习井喷式发展、AI 的新一轮爆发，到大模型推动 AI 技术革新、多模态大模型问世，以及具身智能现实应用，在大模型“百家争鸣”、“百花齐放”的时代 AI 的刚需到底在哪里？本报告面向民生经济重大需求，介绍多模态大模型与生成式人工智能在智能穿搭方向的前沿探索与应用创新，为推动“智能+”场景赋能科技创新和转型升级提供支撑。

► 专题论坛 ◀

专题论坛 2：教育多模态大模型

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 多功能厅 2

组织人

- 黄 华 (北京师范大学 教授)
- 姜文斌 (北京师范大学 教授)
- 刘 静 (中国科学院自动化所 研究员)

论坛介绍

论坛邀请国内教育人工智能领域的多位知名学者和专家，共同探讨教育多模态大模型技术的最新研究成果、发展趋势以及在教育领域的应用经验。生成式大语言模型技术发展迅速，已经成为人工智能的新基石。近年来，大模型技术快速演进，所处理的信息从单一文本模态逐渐扩展至更多模态，所应用的领域也已经成功拓展至多个行业。多模态大模型技术为教育数字化带来了新的契机。教育部近期启动了人工智能赋能教育行动，其中就包括打造生成式人工智能教育专用大模型。因此，非常有必要举办教育多模态大模型技术论坛，促进我国教育多模态大模型的研发与应用。我们期待通过本次论坛，促进教育多模态大模型领域的技术交流，激发更多创新思维和合作火花，共同推动教育多模态大模型技术的蓬勃发展，为引领中国新一代智能教育技术革命贡献力量。

报告嘉宾

嘉宾介绍



文继荣 中国人民大学

文继荣，教授，中国人民大学信息学院院长、高瓴人工智能学院执行院长，曾任微软亚洲研究院高级研究员和互联网搜索与挖掘组主任。长期从事人工智能和大数据领域的研究工作，近年来尤其专注于大模型相关的研究和应用。入选国家海外高层次人才计划、北京市卓越青年科学家计划等。担任北京市第十四届政协常委、中央统战部党外知识分子建言献策专家组专家、第八届教育部科技委委员、中国计算机学会常务理事、SIGIR 2020 程序委员会主席等。

报告题目

教育场景中以语言为核心的多模态大模型

报告摘要：近年来，以大语言模型为代表的新一代人工智能技术改变了传统的机器学习范式，初步展现出一些通用人工智能特征。然而，现有的大模型在理解复杂语义和逻辑关系的准确性、图文多模态场景的扩展性、生成内容的可信性等方面还存在很多问题，无法适用以中高考为代表的教育场景问答任务。这次讲座将介绍我们改进大语言模型的细粒度对齐和长距离依赖的多种尝试，探讨训练 GPT4 风格的多模态大模型的关键要素和图理解的难点与解决方案，并讨论如何增强大模型的可信度。

嘉宾介绍



贺 樑 华东师范大学

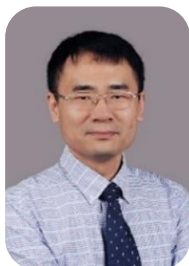
贺樑，华东师范大学计算机科学与技术学院副院长、教授、博士生导师，市政协十四届委员，九三学社上海市委员会委员、常委，九三学社华东师范大学委员会主委。主要研究方向为自然语言处理、知识计算和人机混合增强智能。目前担任国家科技创新 2030 新一代人工智能重大项目管理专家组成员，国家智能社会治理实验特色基地（教育）华东师范大学主任，中国工程科技发展战略上海研究院人工智能创新发展研究室主任，华东师范大学中国智慧研究院副院长，上海市高新技术领域人工智能方向“十四五”和中长期科技发展规划编制组组长，主持多项国家科技支撑、上海市科委重大重点科技攻关项目。

报告题目

新一代人工智能工具对未来教育的影响和应对

报告摘要：人工智能技术的迅猛发展使教育变革变得尤为迫切，AI 技术不仅会渗透“教学管评”各个环节，使人机协同的教学成为新常态，也对职业教育如何调整、教育中的伦理如何监管等问题的解决提出了更多挑战。从长远看，个性化教育时代如何演进，未来我们如何学习，需要培养什么人才，都是当下这个时代赋予我们的问题，而我们则需面向这个 AI 时代，在技术、产品、测评、理念等各个维度上加以布局，做出正确的预判和应对。

嘉宾介绍



刘 均 西安交通大学

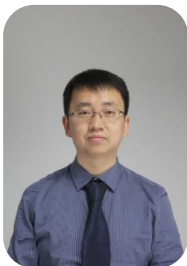
刘均，教授，博士生导师，国家“万人计划”科技创新领军人才，斯坦福大学高级访问学者，IEEE 高级会员。担任 IEEE TNNLS、Information Fusion 编委以及多个国际期刊的客座编辑。近年来，承担了国家重点研发计划项目、国家 863 课题、国家自然科学基金重点项目。在 IJCV、IEEE TPAMI、IEEE TKDE、ICDE、KDD 等重要期刊与国际会议上发表论文百余篇，出版学术专著 2 部，获得 ISSRE 2016、ICBK 2018 等最佳论文奖。授权发明专利 20 项。获国家科技进步二等奖、国家教学成果二等奖，中国自动化学会科技进步特等奖以及多项省部级科技奖励。获陕西省优秀博士论文指导教师、王宽诚育人奖等奖励与荣誉。

报告题目

多模态教学大模型面临的挑战与应对措施

报告摘要：以 ChatGPT 为代表的大模型大大提升了通用智能与内容生成能力，有望成为支撑教育数字化的人工智能基础设施。然而，现在大模型应用于教学场景还面临一系列挑战：一是生成内容存在幻觉、意识形态、价值观等问题，可能对学生产生严重误导；二是跨媒体教学内容理解能力弱，特别是对于包含示意图的教学资源，难以实现深度解析；三是缺少可解释高阶推理能力，难以应对教学中涉及逻辑推理、常识推理与数值推理的复杂问题。该报告围绕上述挑战，汇报我们开展的初步工作。

嘉宾介绍



王士进 科大讯飞

王士进，博士，正高级工程师，现任科大讯飞 AI 研究院执行院长、认知智能全国重点实验室副主任。主要从事自然语言处理、认知大模型等方向研究，承担多项国家重点研发计划，曾获安徽省科学技术进步奖一等奖、吴文俊人工智能科技进步奖一等奖、中国科协求是杰出青年成果转化奖。他主导科大讯飞“认知大模型技术及应用”专项工作，发布的讯飞星火大模型达到国内领先水平。

报告题目

通用人工智能技术进展和教育典型应用

报告摘要：本报告首先分析了人工智能的发展阶段，并提出当前以认知大模型为代表的通用人工智能技术引发全球广泛关注。然后还分析了从认知大模型到多模态大模型的技术特性、发展趋势及应用价值。其次，报告汇报了科大讯飞研发星火大模型的成果和研发经历。最后重点介绍了大模型服务多个行业，特别是教育方向的探索经验。

嘉宾介绍



田 密 好未来

田密，好未来集团 CTO、好未来美校业务负责人、清华大学-好未来智能教育信息技术联合研究中心主任、四川大学-好未来“汇智未来”研究中心主任，高级工程师。毕业于北京航空航天大学计算机软件与理论专业，师从李德毅院士，研究方向不确定性人工智能。长期从事搜索引擎、地图应用、LBS 大数据挖掘、在线教育 AI 技术等研究工作，先后担任过腾讯垂直搜索总监、腾讯地图技术&产品总监、高德地图技术副总裁、阿里巴巴集团资深总监、好未来学而思培优产研负责人、好未来集团副 CTO 等职位。

报告题目

多模态大模型在数学领域的前沿应用与挑战

报告摘要：多模态大模型通过整合来自不同模态的数据，展示出在各种任务上的卓越性能，为解决复杂问题提供了新的视角和方法。然而，尽管多模态大模型的应用前景广阔，它们在数学领域也面临着一系列挑战，包括数据的整合效率、模型的可解释性、以及模型在特定数学问题上的精确性等。本报告将讨论和展示多模态大模型如何在数学领域中识别模式、预测结果，并与传统的数学解决方案相结合，以提高解决问题的效率和准确性。同时，还将讨论在实际应用中如何克服数据融合的困难、提高模型的泛化能力以及确保计算过程的透明度和可追溯性。探讨这一主题不仅能推动数学与 AI 的交叉融合，也对理解和优化多模态大模型的工作机制具有重要意义。

► 专题论坛 ◀

专题论坛 3：软硬协同的高能效视觉计算

(铂金赞助-AutoDL 冠名)

时间

2024 年 7 月 24 日 上午 09:00-12:00

地点

多功能厅 3

组织人

• 鲁继文 (清华大学 教授)

• 陈 涛 (复旦大学 研究员)

论坛介绍

2023 年引爆全球的大模型成为了学术界和产业界的关注焦点。虽然大模型在各个任务上取得了显著效果，但与此同时其巨大的参数量、高昂的训练和推理成本，导致了对各类硬件和数据资源的过度消耗，同时对其在边缘设备上低延迟和高能效的端侧部署也提出了极大挑战。

因此针对上述问题，我们在本论坛中围绕如何有限算力资源下进行高能效的深度模型设计以及高效率端侧视觉部署展开深入讨论和研究分享。论坛的几位讲者将分别从：

- 1) 端侧软硬件协同高效计算
- 2) 低比特深度神经网络量化
- 3) 高能效深度模型优化理论与技术
- 4) 面向边缘多模态计算的软硬协同优化探索
- 5) 高能效扩散模型及其应用

上述几个专题进行报告与圆桌讨论，通过本论坛为学术界相关研究人员提供解决算力受限下视觉和多媒体内容分析研究的新思路，促进大模型在各类终端普适设备上的应用，推动数据高效、硬件高效的绿色人工智能产业更好发展。

报告嘉宾

嘉宾介绍



程 健 中国科学院自动化研究所

程健，中国科学院自动化所研究员、中国科学院大学岗位教授、中科南京人工智能创新研究院副院长，国家级领军人才。主要从事深度学习、图像与视频内容分析、芯片架构设计等方面研究，在 IEEE TPAMI、TCAD、TNNLS、NeurIPS、ICML、ICLR、HPCA、MICRO、CVPR、ICCV、AAAI 等相关领域发表学术论文 100 余篇；先后主持科技部创新 2030-“新一代人工智能”重大项目、中国科学院战略先导课题、国家自然科学基金等项目。相关成果曾先后获得中国科学院卢嘉锡青年人才奖、中国电子学会自然科学一等奖、中国图象图形学会科技二等奖、山东省科技一等奖、江苏省科技一等奖等。

报告题目

端侧软硬件协同高效计算

报告摘要：以 ChatGPT 为代表的大模型是当前最热的人工智能话题之一，并正在改变人工智能的应用范式。由于模型参数量庞大，及对算力和数据的巨大需求，导致大模型的训练和推理成为瓶颈。模型压缩可大幅减少网络模型的计算复杂度和资源消耗，使其更好地应用和部署在手机、机器人、物联网等计算资源受限设备上。本报告首先分析当前大模型计算遇到的挑战问题；然后从模型量化压缩、稀疏剪枝、加速器架构设计等方面介绍团队的最新研究成果；最后分享该领域的未来发展趋势。

嘉宾介绍



刘祥龙 北京航空航天大学

北京航空航天大学教授，博士生导师，国家优秀青年科学基金获得者。现任北京航空航天大学计算机学院副院长、复杂关键软件环境全国重点实验室副主任。主要研究模型压缩、智能安全、开放认知。近年来，主持国家重点研发计划、国家自然科学基金、国防科技创新重点项目、科技创新 2030 重大项目等多项国家课题；发表 TPAMI、IJCV、ICML、NeurIPS、CVPR、USENIX Security 等国际顶级期刊和会议 100 余篇。研究成果受到新华社、人民网等权威媒体和杂志关注。曾获省部级/学会科技奖一等奖 2 项、北京市科技新星、中国计算机学会优秀博士学位论文等。

报告题目

低比特深度神经网络量化

报告摘要：大数据、大模型的学习范式极大提升了深度神经网络的性能，促进了计算机视觉、语音识别、自然语言处理等领域 的快速发展及广泛应用。然而，巨量的数据和参数引发了深度神经网络模型存储和计算等瓶颈问题，尤其在存储和算力等计算资源有限的端侧场景，深度神经网络的部署应用面临严重制约。本报告将围绕神经网络常用压缩技术——低比特量化，介绍线性量化与二值量化两种典型的量化方式，重点介绍团队近年来针对 MLP、CNN、Transformer 等常见神经网络架构以及分类、检测、分割、超分辨等不同任务提出的系列量化方法。

嘉宾介绍



谭明奎 华南理工大学

谭明奎，华南理工大学教授、博士生导师，担任华南理工大学“大数据与智能机器人”教育部重点实验室副主任，华南理工大学计算中心主任，入选中组部高层次人才培养青年项目、广东省“珠江人才团队”，2022 年入选美国斯坦福大学全球前 2% 顶尖科学家榜单。主持国家自然科学基金、广东省“新一代人工智能”重大专项等多个国家和省部级项目。近五年以第一作者或者通讯作者发表学术论文 60 余篇，包括 IEEE TPAMI 等 IEEE 汇刊论文 24 篇，以及 NeurIPS、ICML、ICLR、CVPR 等 CCF-A 人工智能顶级会议论文 30 余篇，担任 NeurIPS、ICML、AAAI、CVPR 等多个人工智能会议的领域主席。获得 2019 年“世界华人数学家联盟最佳论文奖(ICCM Best Paper)”等多项奖励。

报告题目

高效能深度模型优化理论与技术

报告摘要：基于深度学习的人工智能技术在智慧城市、智慧医疗等诸多领域取得了重要突破，但深度学习存在模型学习效率低、深度模型构建难、模型轻量化和低功耗部署难等问题，成为人工智能技术规模化应用的桎梏。如何实现深度模型紧致训练，构建高效能模型结构，推动深度模型在低算力场景的规模化应用，是人工智能领域的国际研究前沿。我们针对对高效能深度模型优化理论与计算加速技术的关键难题进行了攻关，取得了一系列成果：1) 针对紧致模型训练难题，研究高效紧致模型训练新方法 和闭环一致性紧致学习范式，突破现实场景深度模型训练瓶颈；2) 针对高效能模型结构设计难题，提出渐进式模型结构敏捷构建框架和高效模型结构自适应优化方法，推动模型 便捷低成本自动化应用；3) 针对深度模型计算加速难题，研发细粒度模型冗余感知与剪枝方法及高效能模型量化与敏捷部署平台，支撑深度模型在低算力场景的规模化应用。

嘉宾介绍



李 超 上海交通大学

李超，上海交通大学教授，博士生导师。致力于研发高性能可扩展的计算机系统，主持国家自然科学基金委优秀青年科学基金，科技创新 2030 “新一代人工智能”重大项目课题，入选上海市青年科技启明星计划；担任 A 类国际期刊 IEEE Transactions on Computers 的副主编 (AEIC) 和领域首席编委。中国计算机学会杰出会员，体系结构专委会副主任；发表论文 110 余篇，获最佳/优秀/焦点论文奖 6 次；授权国内外发明专利 20 余项。入选全国高校计算机专业优秀教师奖励计划，主编/参编教材 2 部。博士毕业于美国佛罗里达大学，先后获得中国计算机学会青年人才托举计划、TCSC 可扩展计算技术委员会早期职业成就奖、上海交通大学孙永强青年学者奖、CCF 分布式计算与系统青年创新先锋等荣誉。

报告题目

面向边缘多模态计算的软硬协同优化探索

报告摘要：多模态计算 (multi-modal computing) 是边缘微小型数据中心或者嵌入式设备上的一类日益重要的任务。这种计算任务依赖一系列模态传感器来收集包括图像、音频、文本等数据，并借助多模态深度神经网络来生成更富洞见的智能分析结果。相比单一模态分析，多模态计算能够提供更高的精确度，因此在无人驾驶、自主系统、机器人、人工智能物联网等领域有重要应用。作为一种更加复杂的边缘计算负载，多模态计算任务在算力和存力资源消耗、功耗能耗等方面的问题不容忽视，存在许多研究机遇与挑战。本次报告汇报我们近期在多模态计算负载特征分析、多模态计算基准测试集开发、以及边缘智算应用“能源-计算”一体化协同设计方面的进展。

嘉宾介绍



唐彦嵩 清华大学

唐彦嵩，清华大学深圳国际研究生院特聘研究员、教研系列助理教授、博士生导师。分别在清华大学自动化系获得工学学士和博士学位，在英国牛津大学从事博士后工作，并先后在美国加州大学洛杉矶分校和微软亚洲研究院进行访问研究。主要从事人工智能与计算机视觉等领域的相关工作，在国际权威期刊和会议上发表论文 40 余篇，主持国家重点研发计划课题、国家自然科学基金青年科学基金等国家级项目，入选第八届中国科协青年人才托举

工程和微软亚洲研究院“铸星计划”，担任中国人工智能学会模式识别专业委员会（CAAI-PR）副秘书长等学术职务。

报告题目

高效扩散模型及其应用

报告摘要：随着扩散模型的出现，视觉生成模型领域迎来了前所未有的发展机遇。扩散模型通过迭代式的去噪过程及其对文本等多模态信息的交互理解能力，实现了在生成质量和用户可控性方面的新突破。然而，扩散模型庞大的模型体量及长时间的迭代效率带来了高昂的推理成本，限制了其在低能耗设备上的应用场景。本次报告将围绕上述问题，重点介绍团队近期针对扩散模型结构设计的模型压缩及高效推理方案，以及其在图象增强、人体姿态估计以及动作生成等方面的应用。

嘉宾介绍



余佳 AutoDL

余佳，AutoDL 的首席运营官（COO）。

报告题目

破解用卡难复现难解决方案

报告摘要：针对实验室用卡困难、机器管理混乱、算法复现难等痛点问题，AutoDL 将在报告中介绍给出了弹性、好用、省钱的解决方案。AutoDL.com 是目前全国最大的 C 端 AI 算力租用平台之一。平台具有从服务器生产制造、AIDC 建设到算力云服务的全产业链能力。其运营超过 1.4w 张、20 多个型号的 GPU 和国产 AI 加速芯片，具备“万卡”和“千 P”算力的调度能力；面向“大 AI 圈”内的科研工作者和科技企业提供算力分时租赁服务、云计算服务以及私有云部署方案。CodeWithGPU.com 是旗下的算法复现社区，拥有 40w+ 开发者。

► 专题论坛 ◀

专题论坛 4：多模态大模型自适应学习

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 多功能厅 4

组织人

- 常晓军（中国科学技术大学 讲席教授）
- 刘 慧（山东财经大学 教授）
- 朱 磊（同济大学 研究员）

论坛介绍

近年来，以 CLIP 模型为代表的大规模视觉-语言预训练模型（Vision-Language Models, VLMs）在零样本识别、小样本识别以及多模态内容理解与生成等任务上取得了显著的性能突破。然而，VLMs 在实际应用中可能会面临数据分布的动态变化，新兴的数据类别和模式的不断涌现，这些挑战可能会影响到模型在实际应用中的泛化能力，使其无法达到预期的性能表现。为此，越来越多的研究人员开始集中关注如何有效地将 VLMs 自适应地迁移到开放场景下的各种具体任务中，并提出了一系列创新的技术方法，如提示学习、测试阶段自适应和持续学习等，旨在提升 VLMs 的鲁棒性、自适应性和可信性。本论坛以“多模态大模型自适应学习”为核心议题，旨在汇聚该领域的专家学者，就最新的理论发展、技术创新及应用实践进行深入交流与探讨。期望通过此次论坛的举办，能够进一步促进 VLMs 在多模态领域的研究进展，推动相关技术的创新与应用发展。

报告嘉宾

嘉宾介绍



左旺孟 哈尔滨工业大学

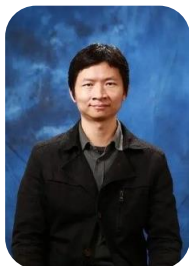
左旺孟，哈尔滨工业大学计算机学院教授、博士生导师。主要从事图像增强与复原、图像编辑与生成、物体检测与目标跟踪、图像与视频分类等方面的研究。在 CVPR/ICCV/ECCV 等顶级会议和 T-PAMI、IJCV 及 IEEE Trans. 等期刊上发表论文 100 余篇。曾任 ICCV2019、CVPR2020/2021、ECCV 2022 等顶级会议领域主席，现任 IEEE T-PAMI 和 T-IP 等期刊编委。

报告题目

黑盒大模型与优化算法的协同方法

报告摘要：随着语言大模型和多模态大模型的持续发展，当前的基础模型如 ChatGPT、Gemini、GPT4o 等虽然日益强大，但逐渐不再开源模型结构与代码。为此，报告将关注两个问题：(1) 如何对黑盒大模型进行微调，使之更适用于特定的下游任务；(2) 如何利用黑盒大模型强大的推理与求解能力，更好地求解优化问题。针对上述问题，报告分别介绍综合 Adapter 和提示学习的大模型黑盒调优方法 CBBT，以及梯度下降与黑盒大模型优化器（Model as an Optimizer, MaaO）相结合的提示学习算法，期待未来几年黑盒大模型的利用和优化可以获得更多的关注与发展。

嘉宾介绍



欧阳万里 上海人工智能实验室

欧阳万里，上海人工智能实验室领军科学家，实验室科学智能部门主任，曾任悉尼大学电子信息工程学院研究主任。其团队在 ImageNet 和 COCO 竞赛多次获得第一。获 IEEE TCSVT 期刊最佳论文，两篇文章入选 paperdigest CVPR/ICCV 最有影响力的文章。担任人工智能领域顶级期刊 TPAMI 和 IJCV 副编辑，CVPR2023 资深领域主席，AAAI24、CVPR2021、ICCV2021 领域主席。

报告题目

AI4Science-机遇与挑战

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。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。本次报告将介绍上海人工智能实验室在 AI4Science 研究的探索。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”。

嘉宾介绍



赫 然 中国科学院自动化所

赫然，中国科学院自动化所模式识别国家重点实验室研究员，中国科学院大学岗位教授，国际模式识别学会会士(IAPR Fellow)。从事模式识别应用基础理论研究，并应用到计算机视觉、生物特征识别和人工智能安全；近期主要聚焦在非监督机器学习以及深度合成中遇到的瓶颈问题。出版信息理论学习专著 1 部，在本领域国际主流期刊 IEEE T-PAMI 和 IJCV 上发表论文 23 篇，获 IEEE 信号处理协会最佳青年论文奖、ICPR 最佳科学论文奖、CSIG 自然科学一等奖、北京市科技进步二等奖等，承担了国家自然科学基金优秀青年科学基金、北京自然科学基金杰出青年科学基金和中国科学院青年促进会优秀会员等项目。他是 IEEE T-IP、IEEE T-BIOM、Pattern Recognition 和自动化学报等多个国内外期刊编委，以及 NeurIPS、ICML、CVPR、ECCV 等领域主席。

报告题目

多模态大模型感知

报告摘要：多模态大模型 (MLLM) 通过整合预训练的视觉编码器和语言模型，在跨模态理解和跨模态推理等复杂任务上取得了显著进展。然而，MLLM 中视觉模块和语言模块之间的能力差异，以及连接层的薄弱，导致模型产生严重的幻觉现象，限制了其在实际应用中的表现。本报告介绍团队在通用视觉基础模型、跨模态对齐方法以及高效微调技术的研究工作，以提升零样本和小样本下游任务的性能，实现跨模态理解和推理。

嘉宾介绍



刘日升 大连理工大学

刘日升，大连理工大学软件学院教授（破格），教育部重点实验室副主任，日本立命馆大学兼职教授。近年来在计算机视觉、深度学习、优化方法等领域发表 IEEE 汇刊及 CCF 推荐 A 类会议论文 100 余篇（含第一作者 TPAMI 论文 10 篇）。获各级自然科学成果奖 4 项，CCF 推荐学术会议论文奖 7 篇。主持国家自然科学基金优青项目、企业联合重点项目、军科委创新特区项目、科技部重点研发计划课题等，入选辽宁省青年拔尖人才、百千万人才工程。

报告题目

优化驱动学习及多模态感知应用

报告摘要：近年来，视觉与学习领域经历了从“小模型、小数据”到“大模型、大数据”的跨域式发展，但在开放场景下如何构建模型可解释、过程可控制、结果可预期的高效鲁棒学习方法仍然面临挑战。本报告将重点汇报团队在开放场景下优化驱动学习理论方法方面的系列工作，并展示面向无人系统全天候自主作业的多模态感知关键技术与智能装备研究进展。

嘉宾介绍



丁 凯 合合信息

丁凯博士，毕业于华南理工大学，合合信息智能技术平台事业部副总经理，高级工程师职称，CSIG 文档图像分析与识别专委会委员，CSIG 机器视觉专委会委员，CSIG 青年工作委员会委员，上海科技大学企业导师，华南理工大学校外研究生导师，获得上海市人才发展基金资助。研究方向为人工智能，模式识别，文档分析与理解，OCR，知识图谱等。近三年在 IEEE TMM, IEEE TAI, CVPR, ACL, ECCV, ACM MM 等国际期刊和会议上发表论文十余篇，申请和授权的发明专利超过 50 项，带领团队获得 ICDAR, ICPR 等十多项国际学术竞赛冠军，作为主要完成人获得 CSIG 科技进步奖二等奖。近年来主持研发的名片识别理解，电子文档还原，多模态文档理解，知识图谱构建与挖掘等项目支撑了合合信息多项核心业务，获得了显著的经济效益和学术评价。

报告题目

大模型驱动的智能文档处理:探索与应用

报告摘要：大模型的发展给智能文档处理领域带来了非常大的机会于挑战，一方面大模型发展使得传统的文档处理技术受到了挑战，同时智能文档技术也可以成为大模型技术很好的补充，让大模型可以更好的分析和理解文档。具体来说，一方面智能文档处理技术可以将海量的各种格式的文档数据转化为可以供大模型训练的高质量的训练数据，另外一方面也可以使得在大模型在推理过程中，大幅度提升文档理解的准确率。本次报告主要聚焦于智能文档处理助力大模型训练与推理这个方向，介绍了合合信息在大模型驱动的智能文档处理方面的一些探索和应用实践。

嘉宾介绍



袁梦轲 航天宏图

袁梦轲，航天宏图算法研究员。2019年毕业于中国科学院自动化研究所，获工学博士学位。目前主要从事遥感图像处理、智能解译技术研究和工程应用，具体包括遥感图像变化检测、地物分类、遥感图像增强、快速正射影像生成等。在 TIP、TVCG、TMM、JSTAR、PR、CVPR、ICCV、IROS、IGARSS 等国际期刊和会议上已发表论文 20 余篇，申请专利 6 项。主持和作为核心成员参与自然科学基金、重点研究项目和重要技术应用项目 11 项。

在多项国际和国内知名遥感领域比赛获得荣誉，如 2022 年人工智能·昇腾 AI 创新大赛 2022（湖北赛区）行业创新赛道金奖，2023 年 IEEE GRSS DFC 大赛屋顶细粒度分类和建筑高度估计双赛道冠军，2024 年“SAR 水体提取”赛道亚军，2023 年全国大数据与智能计算挑战赛“基于亚米级影像的精细化目标检测”赛道二等奖等。

报告题目

天权遥感大模型关键技术与实践

报告摘要：遥感大模型作为一种新兴的遥感数据处理技术，具有准确率高和泛化能力强等优势。本次报告将介绍航天宏图 PIE-Engine 天权大模型遥感云服务产品的关键技术和应用实践。该产品依托航天宏图国际领先的女媧星座等多模态卫星影像资源，结合视觉大模型、人工智能内容生成等前沿技术，提供地物全要素解译、特定场景目标检测、道路提取、水体提取、变化检测、影像超分、去雾以及三维快速构建等服务，旨在解决复杂、多样遥感数据的处理分析挑战，为用户提供高效、精准以及高度泛化的智能遥感云平台服务。产品关键技术多个国际、国内知名遥感领域大赛（IEEE 数据融合大赛、天智杯、全国人工智能应用场景创新挑战赛等）中获得冠亚军，并广泛应用于自然资源、城市治理、交通导航、数字农村以及灾害应急响应等领域。

► 专题论坛 ◀

专题论坛 5：三维点云压缩与表征学习

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 多功能厅 5

组织人

• 李 革 (北京大学 教授)

• 高 伟 (北京大学 研究员)

论坛介绍

基于点云实现的三维视觉技术和应用的快速发展引起了学术界和工业界的广泛关注和研究热潮。三维点云广泛应用于自动驾驶、数字博物馆、沉浸式媒体、元宇宙和具身智能等场景，以高精度的三维建模能力改善人类视觉体验和机器分析性能。点云具有海量数据规模，因此急需研究高效压缩编码方法及标准，实现高效数据存储和传输；同时点云自身的无序性、不规则性和复杂性造成其智能处理方法面临新的挑战，需要研究更加有效的表征学习理论与技术方法，提升点云智能处理效果。本论坛拟针对三维点云压缩与表征学习的前沿问题展开学术研讨，涉及理论、方法、标准及应用。

报告嘉宾

嘉宾介绍



金 欣 清华大学

金欣，清华大学教授/博士生导师，深圳市“鹏城学者”特聘教授，深圳市优秀教师，2018 年中国电子信息科技创新团队带头人。致力于计算光学成像领域研究，承担了 10 余项重点重大项目，在 Nature 子刊和 IEEE Transactions 等上发表论文 170 余篇。荣获 2022 年日内瓦国际发明特别展金奖、2016 年国家科技进步二等奖和 2015 年广东省科学技术一等奖等荣誉。兼任 ISO/IEC JTC 1/SC 29/WG 04 AHG on Lenslet Video Coding 联合主席，IEEE VSPC 技术委员会委员，IEEE Trans. Broadcasting 和 Scientific Reports 副主编。深圳市女科技工作者协会副会长。

报告题目

点云数据获取：高分辨光场计算采集与重建

报告摘要：同时捕获空间光场的强度、方向信息，能为真实场景的三维重建、沉浸式漫游提供丰富有效的真实数据，是沉浸式三维应用走向实际的关键。但是，同步采集时空间分辨率和角度分辨率间的矛盾，对光场采集提出的巨大挑战。本报告从理论和方法创新两个层面探讨突破空间-角度分辨率极限的光场计算采集与重建方法，为实际场景的三维点云重建奠定高质量数据基础。

嘉宾介绍



许晓中 腾讯多媒体实验室

Xiaozhong Xu has been a Senior Principal Researcher and Senior Manager of Multimedia Standards at Tencent Media Lab, since 2017. He was with MediaTek as a Senior Staff Engineer and Department Manager of Multimedia Technology Development, from 2013 to 2017. Prior to that, he worked for Zenverge (acquired by NXP in 2014), a semiconductor company focusing on multi-channel video transcoding ASIC design, from 2011 to 2013. He also held technical positions at Thomson Corporate Research (now Technicolor) and Mitsubishi Electric Research Laboratories. His research interest lies in the general area of multimedia, including video, image and volumetric data coding, processing and transmission. He has been an active participant in various multimedia standardization activities for over fifteen years. Xiaozhong Xu received the B.S. and Ph.D. degrees from Tsinghua University, Beijing China in electronics engineering, and the MS degree from Polytechnic school of engineering, New York University, NY, USA, in electrical and computer engineering. He was a recipient of the Science and Technology Award from China Association for Standardization in 2020.

报告题目

点云压缩：从技术到标准

报告摘要：点云可以灵活方便地表达三维物体或场景的空间结构及表面属性，因而应用广泛。伴随着大规模的点云数据不断积累，点云的海量特性给点云的使用带来了巨大挑战。点云数据的高效存储、传输、发布、共享和标准化，成为点云应用的关键。本报告从压缩，质量评价和系统传输几个层面，探讨支撑实际场景点云应用的若干关键技术和标准化进展。

嘉宾介绍



雷印杰 四川大学

雷印杰，四川大学教授、博士生导师，四川大学电子信息学院副院长。主要研究领域包括人工智能、计算机视觉。近年来，主持了包括国家自然科学基金重点项目、国家重点研发计划课题、国防基础科研重点项目、四川省杰出青年科学基金等三十余项国家级、省部级及企业委托科研项目。在人工智能等领域的国内外顶级学术会议和期刊上发表学术论文 80 余篇，包括 CVPR、ICCV、ECCV（均为 CCF-A 类人工智能顶级国际会议），EMNLP（自然语言处理顶级国际会议），IEEE TPAMI、TIP、TNNLS、TMM（均为中科院一区）等，并申请了三十余项国家发明专利。先后担任 CVPR、ICCV、ACM MM、ECCV、AAAI、WACV 等人工智能领域国际知名学术会议的领域主席（AC）、高级程序委员会委员（SPC）以及程序委员会委员（PC）。

报告题目

面向点云场景理解的跨域跨模态表征学习

报告摘要：三维扫描技术的日益演进催生点云逐步成为数字化时代中场景感知的重要信息载体，其已经在自然灾害态势感知、道路交通安全服务、机器人巡线自主导航等国家重大战略需求中发挥了不可替代的作用。然而，

如何实现跨传感器异构信息融合以形成完备的复杂场景观测依旧亟待突破的挑战性难题。因此，以跨域跨模态表征学习为核心理论基础，协同融合分布复杂、特征异构的多源点云数据，以准确理解复杂多样的三维视觉场景，具有广泛的研究价值与应用前景。本报告将分别探讨跨域表征学习、跨模态表征学习在点云场景理解中各关键任务的最新前沿进展，并介绍课题组的相关研究工作。此外，还将从一体化建模、交互式推理等方面进行展望。

嘉宾介绍

高 伟 北京大学



高伟，北京大学信息工程学院（深圳）研究员/博士生导师，IEEE 高级会员，广东省青年拔尖人才。研究领域为三维点云编码与处理。发表高水平论文 120 余篇，申请专利 90 余项，提交标准提案 40 余项，出版英文专著 1 部。荣获 2021 年 IEEE 多媒体学术新星奖（Runner Up Award）。担任国际期刊 Elsevier Signal Processing 编委。担任国际 IEEE CASS VSPC、APSIPA IVM 技术委员会委员。曾在 ACM MM、IEEE ICME 等会议上组织研讨会和专题会议、担任讲习班讲者。搭建 OpenPointCloud、OpenAICoding 等多个开源算法库。

报告题目

迈向高效且灵活的点云智能编解码

报告摘要：伴随着深度学习在点云压缩领域的应用，压缩效率不断提升，为编码冗余移除提供了更多的潜力空间。然而智能编解码方案具有很多灵活性方面的缺陷，包括在多码率支持、复杂度控制等方面，严重限制了深度学习点云编码技术的发展和應用。在本报告中，将重点汇报报告人围绕深度学习点云编码领域开展的率失真优化、灵活性优化等方面所提出的新方法，以及中国自主深度学习点云编码标准制定工作的最新进展。

► 专题论坛 ◀

专题论坛 6：面向自然科学与工业应用的多媒体智能分析技术

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 多功能厅 6

组织人

- 聂 婕 (中国海洋大学 特聘教授)
- 王 鑫 (清华大学 副研究员)
- 费 豪 (新加坡国立大学 研究员)
- 纪家沂 (厦门大学 博士后研究员)

论坛介绍

随着人工智能技术的迅猛发展，自然科学领域和工业界都积累了大量的图像和视音频数据等多媒体数据资源。这些海量的多媒体数据形成了规模庞大的资源库，为科学研究和工业应用提供了丰富的知识和研究对象。然而，从海量科学多媒体数据中获取有价值的信息和规律是一项巨大的挑战，迫切需要智能感知和理解技术支持。利用多媒体和人工智能等技术实现对自然科学和工业场景中多媒体大数据的智能感知和理解，可以给人们提供更快速数据理解和认知能力，深入挖掘其潜在规律和发展趋势。然而，面对复杂的自然规律和繁多的工业需求，深度模型如何有效迁移其强大的感知和推理能力至关重要。

本论坛专注于媒体智能分析技术与应用这一前沿议题，邀请了来自学术界和工业界的资深专家和优秀青年学者进行报告并参与圆桌讨论，包括教师，博士后以及优秀青年博士。讨论的主旨包括多媒体智能分析技术如何更好的赋能自然科学基础研究和工业应用，并就相关科学问题展开讨论，以梳理该领域的研究现状，识别存在的问题和挑战，促成全面的视角和多样化的解决方案，推动自然科学研究和工业应用的进步和创新，并进一步完善多媒体智能分析的理论范式和技术框架。论坛旨在为跨媒体分析领域的学者、从业人员及研究生搭建交流平台，共同探讨该领域的核心科学问题、未来发展趋势，及从业者的职业和学术发展路径，分享个人成长经历，以促进多媒体领域在教育、科技和人才培养方面的全面进步。

报告嘉宾

嘉宾介绍



田永鸿 北京大学

田永鸿，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，2024 年首批国家杰出青年基金延续资助计划获得者。主要研究方向为分布式机器学习、脉冲神经网络和神经形态视觉。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 Nature/Science 子刊、IEEE Trans 等国际期刊和 ICML、NeurIPS 等国际会议发表学术论文 350 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 110 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、2023 年广东省科技进步特等奖、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，多个国际期刊编委和国际会议

大会主席/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。

报告题目

从 Science for AI 到 AI for Science: 人工智能与基础科学的“双重奏”

报告摘要：在大模型时代，企业成为科技创新的绝对主力，因此我们必须面对的问题是“高校应该在人工智能创新体系中发挥什么作用？”一方面，克服当前人工智能局限性的关键是需要搭建包括脑科学、神经科学和数理科学在内的基础科学与人工智能间的桥梁，利用科学基本原理来启发新的 AI 模型研发，探索更高效的通用人工智能之路；另一方面，AI 大模型正在改变基础科学领域的研究范式，可用于解决科学问题、引导科学假设生成并推进科学发现。本报告将从 Science for AI 到 AI for Science 两个角度来描述最近的一些研究探索和进展，以期达到抛砖引玉之效。

嘉宾介绍

董军宇 中国海洋大学



董军宇，教授，博士生导师。分别于 1993、1999 年在中国海洋大学获得学士和硕士学位，2003 年在英国 Heriot-Watt 大学获得博士学位。国家高层次人才计划科技创新领军人才。主要研究方向为计算机视觉和海洋人工智能。现任中国海洋大学信息科学与工程学部部长，国际计算机学会(ACM)青岛分会共同主席，法国 Interdisciplinary Graduate School for the Blue Planet (ISBlue)国际学术委员会委员。先后主持承担国家重点研发计划、科技部国际合作专项、国家自然科学基金重大仪器（自由申请）专项等 10 个国家级项目，是国际期刊 Intelligent Marine Technology and Systems 的创刊主编。

报告题目

水下光学高分辨率三维成像

报告摘要：由于海底工作环境复杂恶劣、能见度低，传统水下摄像无法进行三维精准测量，而声学成像无法获取高分辨率数据及表面颜色纹理信息。水下高精度光学三维成像产业发展于二十一世纪初，得益于水下机器人、水下成像技术、计算机视觉技术的进步而不断提升。水下高精度三维成像技术及装备产业需求广泛，可以覆盖从海洋工程、水利枢纽等各类基础建设的检测、监测，到海底结构物精细化三维扫描测量、海洋物种中包括鱼类、珊瑚、牡蛎礁、贝壳堤等生物的高分辨率三维形态分析，再到水下考古对海底沉船、遗址等高价文物的三维测绘与挖掘研究等领域。本报告主要汇报项目组研发的三维测量仪器，其精度可达毫米级，多项性能指标（包括探测距离、距离分辨率、水平分辨率、耐压深度、采样速率等）有望超过国外同类产品的水平，同时可以灵活搭载国产潜器。

嘉宾介绍



王翔 中国科学技术大学

王翔，中国科学技术大学特任教授、博导，2021 年国家优青（海外）获得者。研究方向包括可解释与可信人工智能、图深度学习、多模态大模型、信息检索与推荐，2022、2023、2024 连续三年入选 AI 2000 全球人工智能最具影响力学者、“爱思唯尔 2022 中国高被引学者”，承担国家基金委重大研究计划培育项目、科技部新一代人工智能国家科技重大专项大模型课题。相关的研究成果在 CCF A 类国际顶级会议和期刊发表论文 90 余篇，谷歌学术引用一万七千余次，其中 10 篇论文是相关会议的最具影响力论文和最高引论文，3 篇论文入选相关会议的 Best Paper Final List，1 篇论文获得国际基础科学大会前沿科学奖。

报告题目

推荐大模型初探：面向用户行为的理解、对齐与模拟

报告摘要：将大语言模型（LLMs）与推荐系统相结合，是追求更具互动性的个性化信息过滤技术，其中的关键在于多模态数据（自然语言与用户行为）的统一表征与对齐。本报告将分享我们近期的工作，探讨大模型作为推荐系统中的核心组件。这些工作包括解释传统推荐系统的表示、创建与推荐系统互动的用户代理，以及探索推荐系统作为帮助用户的交互工具。其中，我们研究了如何利用大模型解析和理解来自传统推荐系统的表示，以便更好地对齐推荐模型与用户的需求和偏好；此外，我们开发了用户智能体，模拟真实用户与推荐系统进行互动，从而提高用户体验的便利性和自然性。我们还探索了将推荐系统作为一种交互工具，辅助用户在各种应用场景中做出更明智的决策。

这些发展的核心在于强调 LLMs 与推荐模型集成方式的转变——不仅在于提升预测的准确性，还在于彻底革新这些模型的交互能力。通过这种创新的整合方式，我们不仅能够提供更精确的推荐，还能实现更为智能和人性化的用户交互体验。

嘉宾介绍



丁恒辉 复旦大学

丁恒辉，复旦大学青年研究员，国家优青（海外）、上海市海外高层次青年人才，上海市计算机学会副秘书长。2016 年于西安交通大学获学士学位，2020 年于新加坡南洋理工大学获博士学位。曾在南洋理工大学和苏黎世联邦理工学院任博士后研究员。主要从事计算机视觉、多模态大模型、场景理解、图像和视频分割等研究。过去 6 年内共发表论文 70 多篇，包括 50 多篇 CCF-A 类论文和 10 多篇 CCF-B 类论文。担任 IET Computer Vision 期刊编委、Visual Intelligence 期刊编委，担任多个 CCF-A 类国际顶级会议的领域主席，如 CVPR、NeurIPS、ACM MM 等，担任多个 CCF-A 类国际顶级会议的资深程序委员会委员，如 AAAI、IJCAI 等。

报告题目

多模态语义分割及其应用

报告摘要：多模态语义分割是场景理解的关键技术，现有方法和数据集通常研究受限场景和强约束条件下的图像和视频分割。本报告从面向开放复杂环境的多模态图像视频分割实际应用需求和挑战出发，介绍广义指向性目标分割 GRES、复杂场景视频分割 MOSE、基于动作描述的视频目标分割 MeVis 等新任务、新数据集和

关键技术，推动多模态语义分割在更真实场景下的应用研究。

嘉宾介绍

马云山 新加坡国立大学



马云山，新加坡国立大学计算机学院 NExT++ 研究中心博士后研究员，于 2022 年 3 月从新加坡国立大学计算机学院获得博士学位，师从 Prof. Chua Tat-Seng。研究兴趣包括多模态事件预测，计算时尚，推荐系统等。他已在 ACM MM、SIGIR、KDD、WWW、ACL、WSDM、CIKM、TOIS、TKDE 和 TMM 等顶级会议期刊上发表多篇论文。其论文曾入选 ACM MM 2018 最佳论文奖候选，荣获 ACM ICMR 2021 最佳学生论文奖。此外，他担任多个国际会议（ACM MM, KDD, SIGIR, WWW, WSDM, AAAI, IJCAI 等）和期刊（如 TKDE, TMM, TNNLS, TWEB 等）的程序委员和审稿人。

报告题目

面向垂直领域的多模态事件预测

报告摘要：多模态事件预测作为一个新兴的研究问题，旨在捕捉多模态数据中的时间演化模式并进行预测。特别是在某些垂直领域，如股票市场、国际政治、网络安全等，时序事件预测需要多种模态的数据，包括图、文本、图像等。传统的多模态事件预测方法在时序性、多模态和通用的理解能力上存在局限性。为弥补这些不足，我们率先采用强大且具有通用性的多模态大型基础模型（LFM）进行多模态事件预测。具体来说，我们提出了多种基于 LFM 的方法，包括基于自反馈机制生成可解释的股票市场预测，引入图片信息的多模态国际政治事件预测，以及基于网络威胁情报的复杂安全事件抽取及分析。大量的实证研究表明，LFM 展现了优越的性能以及巨大的应用潜力，为多模态事件预测提供了多个有价值的研究方向。

嘉宾介绍

罗根 厦门大学



罗根，博士毕业于厦门大学，师从纪荣嵘教授。研究方向为跨媒体计算，围绕该方向在 IJCV, IEEE TIP, CVPR, ECCV, NeurIPS 及 ACM MM 等 CCF 推荐期刊和会议发表论文 20 余篇，谷歌学术引用 800 余次，其中 5 篇论文被 CVPR, ECCV 和 ACM MM 等领域顶会评为口头报告（Oral Paper），主持首届国家自然科学基金青年学生基础研究项目（博士研究生项目）。

报告题目

图文大模型高效感知与计算

报告摘要：多模态图文大模型被认为是人工智能领域的新顶峰，是机器逼近人类智慧的重要体现。多模态图文大模型的进步不仅对人工智能领域的学术研究构成了重大推动，也为工业界的实际生产带来了颠覆性的变革。最近，开源图文大模型在感知和计算层面进行了诸多尝试，并取得了长足进步。报告回顾了图文大模型在感知和计算层面的重要进展，并围绕报告人近年的研究工作开展进一步的讨论和思考。

► 专题论坛 ◀

专题论坛 7：面向开放复杂环境的跨时空行人关联

时间 2024 年 7 月 24 日 下午 13:30-17:00

地点 多功能厅 1

组织人

- 李 波 (北京航空航天大学 特聘教授)
- 张永飞 (北京航空航天大学 教授)

论坛介绍

行人关联旨在通过跨摄像机下行人图像等信息的检索与关联，实现目标行人定位与轨迹发现，是开放场景非配合应用中人脸识别等身份识别技术的重要补充手段，已成为智能视频监控系统的关键环节和支撑技术。近年来，在学术界和工业界的广泛关注下，行人关联技术取得了突飞猛进的发展。然而，在面向实际应用中长期大范围跨时空开放复杂环境下，由于场景复杂多变、行人目标众多且变化大等原因，开放复杂环境下的行人目标关联面临着一系列新的挑战难题，导致行人关联技术仍然未能像人脸识别、车牌识别等技术一样大规模应用。在此背景下，本论坛邀请国内外学术界和工业界研究部门的专家，深入交流和探讨开放复杂环境下行人目标关联的挑战问题、最新研究、应用进展和发展趋势，希望推动行人关联技术研究的发展和技术的落地应用。

报告嘉宾

嘉宾介绍



廖胜才 起源人工智能研究院

廖胜才，博士，起源人工智能研究院（IIAI）主任科学家，IEEE 高级会员。2005 年获中山大学数学与应用数学学士学位，2010 年获中国科学院自动化所模式识别与智能系统博士学位，2010 年至 2012 年任美国密歇根州立大学计算机系博士后研究员，2012-2018 年间在中国科学院自动化所历任助理研究员、副研究员。廖博士主要从事模式识别和计算机视觉方面的研究工作，特别是人脸和行人检测与识别、分析与生成，和智能视频分析。在国际主流期刊和会议上发表论文 100 余篇，论文被引用 20000 余次，H-Index 52。其代表性工作、发表在国际顶会 CVPR2015 上的 LOMO+XQDA 是行人再识别和度量学习的代表性算法，被国内外学者引用 2400 余次，并在 2015 年最具影响力的 CVPR 论文中排名第 11。曾任 IJCB 2022 程序主席，ICPR、ICB、CVPR、ICCV、ECCV、NeurIPS、WACV 等主流国际会议领域主席，IEEE TIP 和 TMM 副主编，IEEE T-BIOM 客座副主编，和 Springer《生物特征识别百科全书》助理编辑。曾获 ICB 2006 优秀学生论文奖，ICB 2007 最佳论文奖，北京 2008 奥运会突出贡献奖（人脸识别电子票证系统），并荣获国家科技进步二等奖一项（排名第 5）。此外，还获得 IJCB 2014 最佳审稿人奖、CVPR 2019 和 CVPR 2021 杰出审稿人奖；指导学生获得 ICB 2015 和 CCB 2016 最佳学生论文奖；荣获 CVPR2017 行人检测竞赛冠军和 ICCV2019 黑夜行人检测竞赛冠军。

报告题目

域泛化行人再识别：挑战、算法和数据

报告摘要：行人再识别是近年来的热门研究领域，随着深度学习的发展取得了很大的进步。但是已有模型在不同场景下的泛化能力依然较差。虽然迁移学习被大量地研究用于增强模型在新场景下的适应性，但其代价是为了应用到处需要深度学习训练。为此，面向实际应用迫切需要研究开箱即用的行人再识别——即可泛化的行人再识别。本报告将从该问题所面临的挑战、我们提出的一些方法、及大规模虚拟数据对模型泛化性的提升等方面，全面阐述可泛化的行人再识别这一前沿研究课题。

嘉宾介绍



郑伟诗 中山大学

郑伟诗博士，中山大学计算机学院教授/副院长、教育部“长江学者奖励计划”特聘教授、英国皇家学会牛顿高级学者，现任教育部机器智能与先进计算重点实验室主任。长期研究协同与交互分析理论与方法，解决人体建模和机器人行为的视觉计算问题。担任 IEEE T-PAMI 等期刊的编委。主持承担国家级重点类项目和人才项目 5 项、以及广东省自然科学基金委卓越青年团队(负责人)项目等。获国家教学和省部级科技奖励 6 项。

报告题目

行人再识别新思考：统一模型和新评估标准

报告摘要：当前行人再识别研究主要着力于提高两两摄像区域同一人的图像匹配准确率，针对各种困难（遮挡、跨模态、换装等）设计出高效的专用模型。然而，专用模型存在较大局限性。为此，我们提出了一个多功能模型，能够自适应地克服不同困难。此外，由于行人再识别的一个重要目的是在摄像头网络中追踪行人，我们提出了连续一致行人再识别和相应的评估标准。不同于以往针对模型结构的改进，我们通过连续一致性行人再识别分析了多摄像区域下摄像头质量对行人再识别的影响，有效地检测出有缺陷的摄像机设置，有助于对视频监控环境做出有益的调整。

嘉宾介绍



叶 茫 武汉大学

叶茫，武汉大学教授，国家优青（海外）、中国科协青年托举人才、湖北省百人计划创新人才。主要研究方向人工智能、多媒体检索、多模态理解、联邦学习等，发表国际期刊会议论文 90 余篇，其中第一/通讯作者发表 CCF-A 类论文 50 余篇，谷歌学术引用 7000 余次，10 篇论文入选 ESI 高被引论文，入选 2023 湖北省优秀科技论文。担任 CVPR、ACM MM、ECCV 等顶级学术会议领域主席。主持国家自然科学基金面上项目、国家自然科学基金国际合作项目、湖北省重点研发计划等 10 余项科研项目。获谷歌优秀奖学金、计算机视觉顶会 ICCV2021 无人机特定行人检索赛道冠军、2021 年斯坦福排行榜“全球前 2% 顶尖科学家”、2022 年度百度 AI 华人青年学者（计算机视觉领域）等荣誉。

报告题目

多模态行人检索与生成

报告摘要：多模态行人检索在智慧城市等领域有重要应用，现有方法通常独立的去研究单一的跨模态检索任务，极大的限制了实际不确定场景的灵活性，本次报告将介绍如何利用多模态预训练模型，构建不确定多模态场景

下的行人检索基础模型等相关工作，分享文本引导的人像生成相关研究进展及展望多模态行人理解未来趋势。

嘉宾介绍



张史梁 北京大学

张史梁，北京大学长聘副教授，担任 CVIU、IET Computer Vision 编委、ACM TOMM 客座编委，CVPR、AAAI、ICPR 等会议领域主席，以第一作者和通讯作者在 T-PAMI、IJCV、CVPR、NeurIPS 等权威期刊与会议发表论文 100 余篇，引用 10000 余次。是国家青年特聘专家，首批北京市杰出青年科学基金获得者，主持重点研发计划国际合作专项、联合基金重点项目等项目，获电子学会科技进步一等奖、教育部科技发明一等奖、大川研究助成奖、CCF 优博、英伟达先锋研究奖、NEC 美国实验室突出技术奖等。

报告题目

基于大模型的开放场景行人感知

报告摘要：大模型打开通用人工智能大门，揭开了智力时代的大幕。构建能够处理视觉数据的多模态大模型已经成为计算机视觉领域的研究热点。本次报告将介绍我们在构建多模态大模型、提升多模态大模型指代理解能力等方面的近期工作，同时将介绍如何将多模态大模型应用于开放场景行人姿态感知、换衣行人重识别、开集目标识别等视觉任务。

嘉宾介绍



郑文先 深圳云天励飞技术股份有限公司

郑文先，深圳云天励飞技术股份有限公司（股票代码：688343）副总裁、党委书记。本科毕业于北京航空航天大学，清华大学创新领军工程博士在读，北京大学光华管理学院 EMBA，是深圳市地方领军人才、青岛市政协委员、深圳市青年科技人才协会副会长、深圳青年联合会委员、广东省人工智能学会副秘书长、深圳市科创委创赛产业专家、深圳市计算机学会副会长、湖南师范大学硕士生导师。目前，在深圳云天励飞技术股份有限公司担任副总裁、党委书记，全面负责云天励飞的战略规划制定、运营管理、区域拓展、公共事务、品牌策划和知识产权等工作。

报告题目

以芯赋能边缘 AI，打通 AI 应用最后一公里

报告摘要：大模型出现对 AI 行业带来巨大变革。大模型具备的泛化能力，可有效解决开放复杂环境下跨时空行人关联面临的难题。但大模型研发需要投入大量算力、人才和资金，一般企业和个人难以承担。此外，当前大模型以云端部署为主，难以满足各行业和场景对大模型使用实时性和隐私保障的需求。在此背景下，端侧 AI 应运而生。端侧 AI 可让大模型在使用场景本地运行甚至自训练迭代，让 AI 深入赋能到各场景毛细血管。本报告将从当前人工智能发展现状、边缘 AI 的发展趋势、AI 芯片如何赋能边缘 AI 发展、视觉大模型的边缘自训练迭代等方面，探讨边缘 AI 技术如何与大规模视频目标识别应用的产业化方向。

► 专题论坛 ◀

专题论坛 8: AIGC 与媒体内容安全

时间 2024 年 7 月 24 日 下午 13:30-17:00

地点 多功能厅 2

组织人

- 卢伟 (中山大学 教授)
- 高赞 (山东省人工智能研究院 教授)
- 程皓楠 (中国传媒大学 副研究员)
- 钱振兴 (复旦大学 教授)
- 程志勇 (合肥工业大学 教授)

圆桌讨论主席

- 查正军 (中国科学技术大学 教授)

论坛介绍

近年来随着生成式人工智能领域取得关键技术突破, 愈发逼真的深度伪造媒体内容加剧了国际社会“信任赤字”, 带来国防安全、政治安全 and 经济安全等领域诸多问题。各国加速研发应对虚假媒体信息的识别技术, 我国中央网信办也在 2023 年底明确提出“集中整治技术生成虚假短视频”专项行动。深入开展高准确度、高可靠性的媒体内容安全理论与方法研究已成为保卫国家舆论安全的迫切需求。本论坛围绕 AIGC 生成内容检测与溯源、大语言模型安全与隐私保护等方面展开研讨, 旨在探讨 AIGC 与媒体内容安全的理论基础、技术方法、实践案例和未来发展趋势, 邀请了国内相关领域的专家学者共同参与交流分享。

报告嘉宾

嘉宾介绍



罗向阳 解放军信息工程大学

罗向阳, 男, 1978 年生, 信息工程大学教授、博导, 河南省网络空间态势感知重点实验室主任, 国防科技卓越青年基金获得者, 国家重点研发计划项目首席科学家, 先后入选河南省科技创新杰出青年和杰出人才、中原领军人才、军队学科拔尖人才。主要从事网络与信息安全、人工智能安全领域研究。先后主持国家自然科学基金 6 项 (其中重点 3 项), 主持国家重点研发计划、军队和省部级科研项目 30 余项; 在 IEEE TPAMI、TIFS、TDSC、TCSVT、TMM、ACM TOMM 等 IEEE/ACM 系列汇刊、《中国科学》、《计算机学报》、ACM CCS、IJCAI、WWW 等国内外重要期刊/会议发表论文 300 余篇。获全国百篇优博提名和全军优博, 获河南省自然科学一等奖、河南省科技进步一等奖、教育部技术发明一等奖、中国电子学会技术发明一等奖各 1 项, 军队和河南省科技进步二等奖 4 项。获国家教学成果一等奖、河南省教学成果一等奖和军队教学成果一等奖各 1 项。

报告题目

从智能计算到 AIGC 安全

报告摘要: 生成式人工智能已成为世界各国战略资源和软实力竞争的焦点。然而, 这一技术的应用也带来了新的安全风险。尽管已有多种 AIGC 安全治理方法被提出, 但现有研究仍然存在一些不足, 例如伪造鉴别方法过

度依赖数据标签、来源取证方法安全性不足、主动防御方法在实际应用场景中效果不佳、未形成完整的取证链等。解决 AIGC 安全治理关键难题，形成媒体智能鉴别与防护体系显得至关重要。报告重点汇报团队在无监督范式下的深度伪造人脸检测方面的研究进展，提出的一些检查方法，以及构建的智能取证平台。

嘉宾介绍



王 骞 武汉大学

王骞，IEEE Fellow，武汉大学二级教授、国家网络安全学院执行院长，ACM SIGSAC China 主席，国家重大人才工程特聘教授，基础加强计划重点基础研究项目“技术首席专家”，国家海外高层次人才引进人才，获国家优秀青年科学基金资助。长期从事网络空间安全领域研究，致力于推动人工智能系统安全、移动智能终端安全与隐私保护、应用密码学创新发展与应用落地，获浙江省自然科学奖一等奖、中国电子学会自然科学奖一等奖。入选 CCF-Intel Young Faculty Researcher Program，获 IEEE TCSC Award for Excellence in Scalable Computing (Early Career Researcher)、IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award 等多项国际学术荣誉，及 10 余次国际会议最佳论文奖。担任 IEEE TDSC、IEEE TIFS、IEEE TETC 等国际刊物编委，以及网络安全领域四大会议 ACM CCS、USENIX Security、NDSS 等国际会议程序委员会委员。

报告题目

大语言模型安全与隐私保护

报告摘要：自 2022 年底 ChatGPT 出现以来，大语言模型凭借其对于人类语言优秀的理解和生成能力，在智能助手、代码生成、文本翻译等场景中得到广泛应用。但在过去一年中，随着全世界研究人员对于大语言模型的深入研究，其中存在的安全与隐私问题也逐渐凸显。例如，攻击者可以通过精心设计提示词，使模型生成恶意内容，甚至输出隐私训练数据。针对这些问题，研究者们也提出了一些潜在的解决方案，如生成内容检测、数字水印、基于人类反馈的强化学习等。本报告将分别针对大语言模型系统中的三种主要数据类型：文本数据、模型和提示词，分别探讨其中存在的安全与隐私问题、应对方法与局限。最后，本报告还将展望大语言模型未来发展中可能面临的新的安全与隐私问题。

嘉宾介绍



沈 超 西安交通大学

沈超，西安交通大学人才办副处长、二级教授，教育部长江学者特聘教授，教育部创新团队负责人，国家重点研发计划首席科学家，国防基础加强计划首席科学家，重点研发计划“先进计算与新兴软件”重点专项指南专家组成员。主要从事智能系统可信、安全、控制与测试的研究工作，发表学术刊物 180 余篇，获最佳论文奖 9 次。牵头获陕西省科学技术一等奖、中国自动化学会科学技术一等奖、达摩院青橙奖、霍英东教师一等奖、MIT TR35 China、国家优秀青年科学基金、IEEE SMC Early Career Award、陕西省五四青年奖章等。主持国家重大、重点、国际(地区)合作等项目 30 余项，制定国内外标准 5 项，多份建言被中央办公厅等采纳。担任 IEEE TDSC、TCYB 汇刊等 10 余个国际期刊编委、IEEE Xi'an SMC&CS 主席、ACM SIGSAC China 副主席、中国人工智能学会组织工委副主任等。

报告题目

AI 大模型的安全与隐私风险

报告摘要：近些年，以 GPT 为首的大模型技术开启了 AI 研究的新纪元，然而在美好的大模型蓝图下，攻击者也可能利用大模型的脆弱性对个人安全、社会安全、乃至国家安全带来严重的威胁。本报告将从大模型的保密性、完整性、隐私性三个角度出发，分析 AI 模型安全与隐私在大模型时代的各种挑战，探讨大模型的隐私泄漏、模型窃取、对抗攻击、后门威胁、输出安全以及公平性与偏见等热门研究问题。本报告旨在发现并分析大模型安全和隐私风险，推动大模型安全应用与可持续发展。

嘉宾介绍



张卫明 中国科学技术大学

张卫明，中国科学技术大学 教授、博导，网络空间安全学院副院长。主要研究兴趣包括信息隐藏和人工智能安全。已在国际著名学术期刊和会议 IEEE TIT、TPAMI、TIFS、TIP、CVPR、S&P、NDSS、ICCV、NeurIPS、AAAI 等发表论文 200 多篇。主持基础加强重点项目、国家自然科学基金重点、国家重点研发课题、国家 863 等项目 20 余项。获得军队科技进步一等奖、安徽省自然科学奖一等奖、安徽省教学成果特等奖、ACM SIGSOFT 杰出论文奖。入选 2021 年长三角人工智能十大杰出人物。

报告题目

AIGC 生成内容检测与溯源技术

报告摘要：生成式人工智能的快速发展为人们的学习、工作与生活带来便利的同时，也给网络空间安全带来了全新的挑战，诸如利用 AI 撰写虚假新闻、学术造假和网络诈骗等。因此 AI 生成内容的检测与溯源技术成为迫切需求。此报告将介绍 AI 生成文本、图像的被动检测方法和主动溯源方法。

嘉宾介绍



董晶 中国科学院自动化研究所

董晶，女，博士，中国科学院自动化研究所研究员、博士生导师，CSIG/CCF/IEEE 高级会员，CAAI 杰出会员，中国科学院青年创新促进会会员，国家高层次青年人才计划入选者，目前担任中国图象图形学学会 (CSIG) 理事、副秘书长、女工委秘书长，IEEE 亚太区执委、奖励委员会主席、IEEE 信号处理协会全球成员发展主席。主要从事人工智能安全与对抗、计算机视觉、多媒体内容取证等前沿方向的技术研究，她先后以课题（子课题）负责人承担了 4 项国家自然科学基金、3 项国家重点研发课题、1 项工信部国家重大专项子项及 20 余项省部级科研项目（课题）。已在国际权威期刊及学术会议上发表学术论文 90 余篇，已授权 26 项中国专利含 3 项美国专利。她曾获 4 次最佳（优秀）论文奖、2016 年度 IBM 学院奖、2019 年度中国人工智能学会杰出贡献奖、2020 年度 CSIG 石青云女科学家奖、2021 年度北京青年优秀科技论文奖、2021 年度 CSIG 科技奖二等奖（排名第一）、2021 年度吴文俊人工智能科学技术奖（技术进步科普类）、2022 年度中国发明协会创业创新奖一等奖（排名第一）、2022 年度国家广电总局 MediaAIAC 大赛深度合成技术应用类一等奖（排名第一）以及 2023 年度吴文俊人工智能科学技术奖（技术发明一等奖、排名第二）。

报告题目

AI 视觉内容生成与安全对抗研究

报告摘要：由 ChatGPT 驱动的技术范式变革的话题近期频发热议，随着多模态大模型其相关应用的纷纷落地并被广泛应用，越来越多的研究热点已围绕由 AIGC 等前沿智能技术引发的隐私保护和数据安全隐忧开展。本报告将重点面向构建智慧社会安全治理体系，围绕人工智能伪造音视频的鉴定和治理关键技术，重点关注数字人像伪造与鉴伪的 AI 对抗研究，介绍 AI 视觉内容“伪造”与“鉴别”、“攻击”与“防御”相互博弈和相互促进的独特及可持续发展机制，通过研究深度伪造与鉴伪的机理、机制和线索梳理，展开 AI 视觉深度合成与鉴别应用的相关前沿探讨，并介绍相关研究的前沿技术进展与发展趋势。

► 专题论坛 ◀

专题论坛 9：沉浸式媒体技术赋能文旅产业

时间 2024 年 7 月 24 日 下午 13:30-17:00

地点 多功能厅 3

组织人

- 马思伟（北京大学 博雅特聘教授）
- 叶 龙（中国传媒大学 教授）
- 张新峰（中国科学院大学 副教授）

论坛介绍

以全景视频、3D 视频、VR/AR 等为代表的沉浸式交互化媒体新业态，是未来信息技术融合创新的主战场，深刻影响着文化、教育、旅游等行业的转型升级。但是，现阶段沉浸式视音频技术体系尚未定型，业态发展面临多重问题与挑战：业务系统方面，沉浸式视频直播、VR 社交等多层次与多场景交互应用受制于网络带宽与算力负载，云边端协同机制尚不能根据应用的交互特性进行智能化设计，因此无法满足个性化、社交化的新型业务需求；应用示范方面，虽然沉浸式媒体技术产业发展条件基本成熟，但是应用模式单一，缺乏新业态和新模式的创新，拓宽内容消费场景、助力区域文化传播等方面仍有很大提升空间。

文旅产业是沉浸式媒体的重要应用场景，国家十四五规划和 2035 远景目标也提出加快发展新型文化企业、文化业态、文化消费模式，推动沉浸式体验等新型文旅服务发展。发展新一代沉浸式体验型文化和旅游消费内容，开发沉浸化体验技术应用系统平台与产品，培育数字文化产业新业态和新消费模式，成为贯彻落实国家文化产业数字化战略的重要任务。本论坛旨在汇集国内外相关领域的专家学者，共同探讨沉浸式媒体技术在文旅产业中的应用现状、挑战与未来发展趋势，汇聚专家观点，以期形成科技赋能文旅产业的新思路。

报告嘉宾

嘉宾介绍



邓 宁 北京第二外国语学院

邓宁，教授，北京第二外国语学院旅游科学学院副院长，数字文旅研究中心（文化和旅游部研究基地）主任，文化和旅游部信息科技领域专家，中国旅游景区协会信息化领域专家，国家自然科学基金管理学部通讯评审专家。博士毕业于北京理工大学计算机专业，曾任职于联想研究院和华为北研所。目前主要研究方向：数字文旅、目的地数字营销、旅游大数据、UGC 可视化素材挖掘。近年来在 TM、JTR、JDMM、CIIT、《旅游学刊》等国内外顶级旅游管理 SSCI 和 CSSCI 期刊上发表学术论文；主持国家自然科学基金面上项目、省部级项目、中国首个旅游 APEC 国际项目；建设国内首门“旅游大数据”本科课程，获教育部一流本科课程（在线）；承担文化和旅游部及各省市区数字文旅相关课题 30 余项；开发运营具有自主知识产权的“众誉旅游大数据网络口碑分析平台”，拥有相关专利 6 项。

报告题目

沉浸式赋能文旅新质生产力：逻辑、现状与前瞻

报告摘要：沉浸式文旅是目前旅游业新质生产力发展的重要组成部分，也是目前旅游产品创新的具体体现形式，随着智慧旅游的发展数字技术已经不再局限于旅游管理，服务，营销的“老三样”，更结合内容作为提升游客在地体验和目的地形象投射的有效手段。本话题将涉及沉浸式文旅产品的分类，具体案例，特点及产品发展最新趋势等，促进沉浸式技术与文旅实际场景相结合。更好协调各方力量和资源为赋能文旅高质量发展提供技术和理论支撑。

嘉宾介绍



郭宗明 北京大学

郭宗明，北京大学二级教授、博士生导师、享受国务院政府特殊津贴专家，1987/90/94 于北京大学分别获得学士/硕士/博士学位。目前担任北京大学王选计算机研究所学术委员会主任、科研副所长，电子出版新技术国家工程中心主任，国家语委中国文字字体设计与研究中心主任；同时兼任文化旅游部重点实验室“闽台非遗文化数字化保护与智能处理”学术委员会主任（厦门大学），CSIG 数字文化遗产专委会副主任。主要研究方向为智能媒体处理技术。曾主持、领导研发团队攻克多项技术与工程难关，研制成功多项工程应用系统并获得广泛应用。获中国高等学校十大科技进展、国家科技进步二等奖、教育部科技进步一等奖和二等奖、北京市技术发明奖各一次，还获得王选新闻科技奖—杰出人才奖和北京市优秀师德先进个人等荣誉，6 次获得北京大学优秀博士论文指导导师称号，截至到目前累计发表学术论文 200 多篇，获得授权专利 100 多项。

报告题目

汉字字体与人工智能

报告摘要：汉字是迄今为止连续使用时间最长的文字，也是上古时期各大文字体系中唯一传承至今的文字，更是目前世界上唯一仍被广泛使用的高度发展的表意文字。本报告介绍了汉字的起源，和发展历史，以及最近的二次汉字信息化技术革命—汉字激光照排技术和智能化汉字技术，并展示了最新各种酷炫的汉字字体以及人工智能技术在字体识别、生成、风格化等方面的应用。

嘉宾介绍



穆凯辉 蓝海彤翔集团

穆凯辉，蓝海彤翔集团副总裁，CTO。参与了多项国家科技重大专项和国家重点研发计划项目，主持开发了蓝海创意云平台，以云计算、人工智能、区块链、超高清低延时视频传输、5G 实时渲染等为技术底座，打造了多项优质产品服务和解决方案，建立了完善的生态技术体系，为全球 40 多个国家和地区提供影视动漫、展览展示、VR/AR、数字文旅、智慧城市、数字孪生、电商直播、新零售、在线教育等专业服务。

报告题目

虚实交互和 AIGC 的融合及在文旅行业的应用

报告摘要：虚实融合和智能交互技术迅速发展，不断突破，应用市场广泛。AIGC 技术如火如荼，对虚拟融合和智能交互技术产生了极大的影响，逐渐在其中扮演成核心角色，不断加速内容的生成速度和质量，推动智能交互的创新应用，成为构建未来数字世界的关键驱动力之一。在文旅行业，虚实融合和智能交互应用广泛，正在

改变文旅行业的面貌，不仅提升了游客的沉浸式体验，也为行业带了新的商业模式和增长点。随着技术的不断进步，未来的文旅体验将更加个性化、智能化和沉浸化。

嘉宾介绍



柳启阳 贵州省文化旅游科技有限公司

柳启阳,贵州省文化旅游科技有限公司副总经理，中级经济师。

报告题目

数字科技美学 打造地球的红飘带

报告摘要：充分利用数字创新优势，推进数字科技与文化旅游深度融合，在保护传承和创新发展中做亮红色文化，打造出首个以长征为主题的全域行浸式数字体验馆——红飘带，助推多彩贵州文化旅游强省建设，更好满足人民群众对美好生活的新期待。

嘉宾介绍



汤启鹏 宽堂（北京）文化发展有限公司

汤启鹏，宽堂（北京）文化发展有限公司总经理，原阿里巴巴集团传统文化产业带业务负责人；文化和旅游部非遗研培专项专家、文化和旅游部文化产业“双创”人才专家，现重点专注于文旅、文博、传统文化 IP 数字化场景创新应用。

报告题目

数字文旅的无限可能：探索元宇宙中应用的新场景

报告摘要：随着数字技术的飞速发展，元宇宙作为虚拟世界与现实世界融合的新平台，为文旅行业带来了前所未有的机遇。同时，元宇宙将为文旅行业带来更多创新的应用场景，元宇宙艺术与创作应用、元宇宙博物馆、元宇宙非遗传承体验等不仅能够丰富人们的文化生活、文化科普教育，还能够推动文旅产业的数字化转型，实现文旅与科技的深度融合。本报告重点以潭柘寺元宇宙祈福馆等数字文旅项目为例，旨在探讨元宇宙与数字技术在文旅领域的创新应用，展现数字文旅的无限可能。潭柘寺元宇宙祈福馆是北京千年古刹潭柘寺的线上祈福空间，依托全新的区块链技术，用户可以不受地理限制随时随地进入元宇宙祈福馆祈福、禅修、游览、收藏数字藏品等，祈福馆独特的“摸鱼”文化已成为用户每日打卡任务。元宇宙祈福馆以数字化的形式，让古老传统文化得以在元宇宙中焕发新生，同时为年轻群体用户提供了一种全新的文旅体验和价值认同。

嘉宾介绍



林 松 OPPO

林松，2000 年本科毕业于南京大学声学专业。历任智多微电子多媒体研发经理、ForteMedia 电声调测系统技术负责人、华为终端音频主任工程师等。在 OPPO 负责软件音频相关的产品技术规划和设计，并在沉浸式音频、AI 音频等热点技术方向做了较多的前瞻性工作。全息音频即为期间落地的重要成果之一，为 OPPO 产品在音频领域提供差异化的竞争力。

报告题目

全息音频---手机对象音频服务框架探索与实践

报告摘要：全息音频是 OPPO 基于手机端推出的新一代空间音频技术。基于空间音频中的对象音频(Object-Based Audio)技术，对 Android 系统的音频服务框架做了关键性的改造，实现了传统手机音频体验的三维化，并提升了多个声音一起播放时的声音辨识度。更为有意义的是，全息音频将手机音频信号的操控维度，从传统的时频域转化到了更为直观的空间域上，从而在未来的 XR 时代，带来更多交互和体验上的可能性。

► 专题论坛 ◀

专题论坛 10：面向具身智能的多模态感知与交互

时间 2024 年 7 月 24 日 下午 13:30-17:00

地点 多功能厅 4

组织人

- 赵铁松（福州大学 教授）
- 金 枝（中山大学 副教授）
- 陈鸣锴（南京邮电大学 副教授）

论坛介绍

具身智能技术将 AI 与机器人结合，能够像人一样能够与环境交互感知、自助规划、决策、行动、执行任务，其中融合机器视觉、自然语言理解、认知和推理、机器人学、博弈伦理、机器学习等，横跨多个学科方向。本论坛将关注具身智能中的多种媒体形式，如图像、视频、触觉等，以及其跨模态交互研究，将针对具身智能中的多模态媒体感知与交互，邀请了领域知名学者专家，阐明相关理论与方法的最新进展，探讨相关技术的未来发展趋势与热门课题，引导不同领域的专家关注新兴媒体课题，引发新讨论，激发新灵感，拓展新思路，形成新成果。

报告嘉宾

嘉宾介绍



丁文伯 清华大学深圳国际研究生院

丁文伯，清华大学深圳国际研究生院副教授、博士生导师，国家青年特聘专家。主要研究领域为机器人感知及智能人机交互系统。曾获 IROS 2023 最佳应用论文提名奖；2019 年和 2022 年中国电子学会自然科学二等奖；第 47 届日内瓦国际发明展金奖；2015 年中国光学工程学会创新技术奖二等奖；2015 年度 IEEE Scott Helt Memorial Award (1/5)；2015 年清华大学特等奖学金等荣誉。担任信号处理期刊 DSP 副编辑、IEEE JSTSP 机器人特刊首席客座编辑。

报告题目

视觉与触觉的交融——面向复杂物体感知与抓取的柔性触觉技术

报告摘要：具有触觉感知功能的电子皮肤使得智能机器人能够进行灵巧的操作并与人类和周围环境进行自然交互。然而，由于信号解耦的瓶颈，使用单一触觉传感机制同时感知几何特征和材料属性仍然是一个挑战。在本次报告中，我将介绍团队在基于柔性电子皮肤的多模态传感器研制以及多光谱视触融合感知架构方面的工作，并探讨触觉传感在具身智能、人形机器人等领域的技术难题和应用前景。

嘉宾介绍



陈立翰 北京大学

陈立翰，北京大学心理与认知科学学院副教授，脑与认知科学系副系主任，哲学博士（2010年，慕尼黑大学）。中国心理学会工程心理学专业委员会委员（2016-至今），中国仪器仪表学会力触觉感知与交互专业委员会委员（2023-2028），《Perception》编委（2021-至今）。主持科技创新 2030“脑科学与类脑研究”等课题项目研究。研究方向为多感觉注意、触觉感知等。

报告题目

脑启发的触觉感知和学习

报告摘要：人类通过手指的精细触觉感知和学习，与外界进行交互获得具身认知。本报告汇报课题组在指端的触觉客体识别、触觉数量感知和朝向知觉学习等领域的研究进展，揭示触觉感知和学习的关键脑认知过程与神经机制，以及探讨触觉与视觉客体加工的“超通道”属性、不同感觉通道的客体信息表征和加工的普遍知觉组织规律。相关研究有望为基于人工智能的多模态目标检测和识别等应用场景，提供一定的理论依据和启发价值。

嘉宾介绍



刘 倩 大连理工大学

刘倩，教授，博士生导师，大连理工大学计算机科学与技术学院副院长。2016年获德国洪堡科研基金，主要研究方向为触觉信号处理与通信技术、多模态人机交互、无线多媒体通信等。近5年以第一/通讯作者发表高水平期刊和会议论文30余篇。2021年获辽宁省技术发明一等奖（灵巧机械臂本体设计与智能控制技术）。曾任亚洲触觉大会（AsiaHaptics）程序委员会主席（2022）、IEEE 视听触觉环境与游戏会议程序委员会主席（2017-2018），现担任IEEE 触觉编码标准工作组秘书长。

报告题目

Taxel 触觉图像超分辨率重建方法

报告摘要：触觉感知是人类的重要感知之一，是人类与周围环境互动、获取有关物体及其表面性质的重要途径。触觉信号通常分为三类：力信号、振动信号和触觉图像（tactile pattern）。触觉图像可以通过分布式触觉传感阵列捕捉表面纹理和接触物体形状等复杂触觉细节。但是由于传感器硬件的物理低分辨率限制，商用触觉传感阵列通常难以直接获得高分辨率触觉图像。本报告重点介绍如何从传感器采集的低分辨率数据获得高分辨率触觉图像，从有监督学习框架到无监督主动触觉探索，逐步深入介绍触觉图像超分辨率重建的几种最新方法。

嘉宾介绍



魏 昕 南京邮电大学

魏昕，南京邮电大学教授、博士生导师，江苏高校青蓝工程中青年学术带头人。目前的研究方向为多媒体通信与信息处理。近年来，主持国家自然科学基金、江苏省自然科学基金、江苏省教学改革重中之重课题等。在IEEE TCOM/JSAC/TMM/TCSVT/WCM、C&E、PR等期刊发表SCI、SSCI学术论文30余篇，出版Springer英文学术专著2本；担任中国通信学会通信理论与信号处理专委会委员，担任多个通信与信号处理领域国际学术会议

分论坛主席；获得中国通信学会自然科学一等奖、国际会议最佳论文奖等多项学术奖励；获得中国国家发明专利 20 余项，其中 8 项已实现转化。

报告题目

面向多模态业务的视觉-触觉信号恢复技术

报告摘要：随着多媒体及无线通信技术的飞速发展，人们在视听需求得到极大满足的同时，开始追求更多维度、更高层次的感官体验。研究表明，将触觉等感官信息融入传统音视频而形成的多模态业务，可望带来更为丰富的用户沉浸式体验。然而，面对视觉和触觉信号在自身特性、传输需求等方面上的巨大差异，想要真正意义上实现面向多模态业务的视觉—触觉高效的信息处理，却并非易事。鉴于此，本报告将针对三类典型的多模态业务，分别介绍与之相适配的视觉-触觉信号恢复技术，具体包括触觉辅助下的视觉恢复技术、基于云边协同的跨模态触觉恢复技术、基于语义的视触觉相互补偿技术等，并介绍相关实际系统的研发以及应用情况。

嘉宾介绍



韩 腾 中国科学院软件研究所

韩腾，中国科学院软件研究所研究员，博导，获中国科学院人才项目支持。本科毕业于西安交通大学工业设计系，博士毕业于加拿大曼尼托巴大学人机交互实验室。其主要研究方向为自然人机交互、触觉感知和柔性触觉界面。发表 ACM CHI、UIST、IMWUT/UbiComp 等人机交互国际顶级会议/期刊论文三十余篇，两次获最佳论文提名奖，并发表国际高水平 SCI 期刊 NPJ Flex. Elect.、Nature Comm.、Adv. Intel. Sys.等。目前主持、参与国家自然科学基金重点项目、面上项目等国家课题，并主导了多项企业合作项目。

报告题目

视触媒体信息融合与交互

报告摘要：人工智能、物联网、虚拟现实等信息技术的发展加速了信息空间与物理世界的深度融合，人与智能信息空间的交互、混合将是未来社会的发展形态。智能信息空间的交互脱离了桌面式图形用户界面的范式，从二维（视觉、听觉）拓展到三维（视觉、听觉、触觉与本体觉）。触觉和本体觉参与的视听触融合呈现对交互的优化调控和认知干预是发展交互生理心理学模型和交互技术的重要突破口。这次报告我将介绍研究组近期在围绕触觉反馈和本体觉感知的自然交互技术方面取得的成果，包括触觉、本体觉对人在信息空间中的交互行为、效能和体验的优化机制，多元触力觉渲染和柔性触觉致动技术，以及结合多场景研究的视听触融合呈现技术和应用。

► 专题论坛 ◀

专题论坛 11：三维视觉理论、技术与应用

时间 2024 年 7 月 24 日 下午 13:30-17:00

地点 多功能厅 5

组织人

- 马 展 (南京大学 教授)
- 马凯光 (南京航空航天大学 教授)
- 丁丹丹 (杭州师范大学 副教授)

论坛介绍

近年来,自动驾驶、机器人、虚拟/增强现实等应用崛起,相应的核心技术——三维视觉也得到了广泛关注。三维视觉突破了传统的二维图像空间,实现了三维空间的分析、理解和交互,涉及深度图像获取、视觉定位、三维建模、三维数据传输与重建、三维理解等多个研究领域。本专题论坛将针对三维视觉的热点问题展开,介绍最前沿的三维建模、动态点云获取与压缩、基于 NeRF、3D Gaussian Splatting 等技术的三维重建、场景理解、SLAM 技术及多传感器融合技术等等,促进领域内专家的交流与合作,并探索未来的发展方向。

报告嘉宾

嘉宾介绍



刘利刚 中国科学技术大学

刘利刚,中国科学技术大学教授、博导,国家自然科学基金委“杰出青年”获得者,从事计算机图形学研究。于 2001 年在浙江大学获得应用数学博士学位。曾获得陆增镛 CAD&CG 高科技奖一等奖、国家自然科学基金二等奖(排名 3)等奖项。任中国工业与应用数学学会几何设计与计算专业委员会(CSIAM GDC)主任、国际几何建模与处理(GMP)协会指导委员会委员、亚洲图形学协会(Asiagraphics)秘书长。

报告题目

三维场景的机器人自主探测与重建

报告摘要: 利用移动机器人对未知室内场景进行自主三维扫描与重建是机器人与计算机图形学领域共同关注的重要研究方向。相比于人手持扫描,机器人自动扫描具有不易疲劳、稳定性好、控制精度高等优点。我们分别提出了基于单机器人和多机器人的三维场景的自主探索与重建方法,对未知场景的探索、重建和理解能在一次规划中完成。虚拟与真实场景的实验结果展现了所提出方法的可行性和高效性。

嘉宾介绍



郭延文 南京大学

郭延文，南京大学教授、博导。浙江大学博士、美国伊利诺伊大学香槟分校访问学者。主要研究方向为计算机图形学和三维计算机视觉，研究成果发表在 ACM TOG、IEEE T-PAMI/TIP/TVCG、等领域顶级期刊和 Siggraph, NeurIPS/ICCV/CVPR/ECCV 等顶级会议，主持国家自然科学基金重点项目、十三五装发预研项目以及江苏省杰出青年科学基金。担任中国图像图形学会理事，获得华为公司“难题揭榜”火花奖以及陆增镛 CAD&CG 高科技奖等奖励。

报告题目

三维场景精确重建：从图形学研究到 AI 赋能

报告摘要：三维场景重建在视景仿真以及影视特效等领域具有广泛应用。然而现实世界场景的精确重建一直是一个有挑战性的难题，传感器直接采集的数据往往数据量大且存在缺陷，难以直接应用，而传统手工建模方法工作量巨大且效率低下。随着以深度学习为代表的人工智能技术的发展，三维场景分析以及重建的研究取得了很大进步。本报告将介绍我们从三维场景理解暨三维视觉到三维物体和场景的重建等方面的研究进展以及在工业场景数字孪生方面的典型应用，并介绍我们发布的第一个真实激光雷达扫描室内场景数据集 LiDAR-NET。

嘉宾介绍



刘 勇 浙江大学

刘勇，浙江大学智能系统与控制研究所教授，浙江大学控制学院智能驾驶与未来交通中心主任，浙江大学先进智能系统研究中心副主任，浙江大学控制科学与工程学院党委委员，浙江省机器人专家。获浙江省科学技术进步一等奖、浙江省自然科学一等奖、浙江省科学技术一等奖、浙江省自然科学学术二等奖、浙江省杰出青年科学基金项目，入选中组部万人计划青年拔尖人才、浙江省有突出贡献青年科技人才和浙江省 151 人才项目，以第一作者或通讯作者在 IEEE Transactions、ICRA、CVPR、IJCAI、ICCV、IROS、ICLR、AAAI/IAAI 等知名期刊和机器人顶级会议发表论文百余篇。主要研究方向：自主机器人与智能系统、机器人自主规划与导航控制、视觉识别与模式识别、SLAM 技术及多传感器融合技术。

报告题目

多源融合 SLAM-现状与挑战

报告摘要：SLAM 技术是自主移动平台中的核心技术之一。近年来，随着无人驾驶和自主移动机器人的飞速发展，SLAM 技术也获得了飞速的进展，逐渐从基础研究走向了深度的实践应用。然而在实际机器人和无人驾驶平台应用中，SLAM 技术还面临着环境复杂多变、大规模结构化与非结构化场景交替、机器人平台剧烈运动以及光照、季节变化显著等挑战问题，因而采用多种传感器、多种信息源以及多种特征手段的多源融合 SLAM 成为了解决目前挑战问题的必然选择，本报告将针对目前 SLAM 在实践应用中面临的多种挑战问题开展介绍，分享目前的技术解决方案和最新进展。

嘉宾介绍



范晓鹏 哈尔滨工业大学

范晓鹏，哈尔滨工业大学计算学部二级教授/博导、长江学者、智能接口与人机交互研究中心主任、认知智能与内容安全教育部重点实验室副主任，哈工大苏研院《数字孪生与具身智能》团队负责人。入选国家级高层次人才（2021年）、新世纪优秀人才（2011年）、微软铸星计划（2010年）等。2009年于香港科技大学电子工程系获博士学位。2013年获IEEE标准杰出贡献奖。2023年获电子学会创新团队奖。主要研究兴趣包括图形图像技术、人机交互技术等，发表IEEE TIP、IEEE TVCG、IEEE TCSVT、CVPR、ACM MM、AAAI、ICML、NeurIPS等国际期刊和会议论文近200篇，获发明专利20余项，十余项技术被国际/国内标准采纳。主持与承担国家级项目10余项，包括担任科技部重点研发计划项目负责人、国家自然科学基金委重点项目负责人等。研发的多项技术在华为公司、神华集团、字节跳动等落地应用。2017年作为程序主席主办CCF推荐会议PCM2017。担任中国人工智能学会（CAAI）教育工作委员会副主任、中国计算机产业协会元宇宙专委会副会长、黑龙江省计算机学会学术工作委员会主任等。

报告题目

局部自适应的三维信号处理

报告摘要：随着数字城市、元宇宙等应用需求的增长，以及计算机处理能力的提升，以点云、Mesh等为代表的三维信号处理技术获得了更多的关注。点云、Mesh等三维数据在采集和重建过程中会受到精度的限制以及噪声的污染，并且信号分布以及噪声分布均具有空间不一致性，如何深入挖掘空间不一致性，提升三维信号的精度并去除噪声是目前仍需要关注的问题。本报告将首先回顾传统以及基于深度学习的点云和Mesh去噪增强技术，然后介绍基于深度学习的具有一定局部自适应能力的三维信号处理技术，包括滤波与网络结合的方法、基于元学习的方法、基于师生框架的方法、基于图割的方法等，并探讨这些技术应用于其他任务的新思路和新途径。

嘉宾介绍



廖依依 浙江大学

廖依依，2013年获西安交通大学学士学位，2018年获浙江大学博士学位。2018至2021年，在德国马克斯普朗克智能系统研究所（MPI-IS）及德国图宾根大学从事三年博士后研究。研究兴趣为三维视觉，包括场景重建、场景语义理解、可控图像生成。累计发表文章二十余篇，代表作有KITTI-360, GRAF, Deep Marching Cubes等。担任CVPR2023、3DV 2022、BMVC 2021-2022的领域主席。

报告题目

面向沉浸式媒体的实时渲染与轻量表征

报告摘要：近年来，以NeRF为代表的神经渲染技术的飞速发展让照片级真实度的沉浸式媒体成为可能。面向沉浸式媒体实际应用时，一个理想的场景表征应该同时支持实时渲染和高效传输，而现有许多方法往往难以实现时间和空间的权衡。本次报告将分享我们在渲染加速、表征压缩方面作出的一些探索，旨在向同时支持实时渲染与高效传输的沉浸式媒体迈进。

嘉宾介绍



周席龙 联想研究院

周席龙，大连理工大学自动化学士，北京大学软件工程硕士，现任联想研究院资深研究员，负责实时全息通信系统的研发与落地，基于联想实时全息通信技术平台 LHRTC (Lenovo Holographic Real-time Communication) 研发的业界首个 27 寸 单人自由视点裸眼 3D 全息通信会议系统，在中关村论坛等重量级场合进行展示，获得北京日报，中国教育电视台等媒体的推荐报道，拥有国内外专利 20 余项，5G 超高清传输虚实融合互动技术获中国多媒体企业创新奖。

报告题目

全息通信——未来的沟通方式

报告摘要：从看的清到看的真，数字世界正在从 2D 向 3D 进化，科幻片中如同面对面的沉浸临场感沟通方式正在逐步成为现实，报告将从整个全息通信系统设计出发，讨论全息数据采集、数据压缩、传输协议及光场显示设备等多个领域当前面临的技术挑战与解决方案，最后，展望全息通信的未来发展趋势，预测其在光场显示技术，5G、AI 的驱动下，将如何进一步革新人们的沟通方式，推动多媒体领域的创新与进步。

► 专题论坛 ◀

专题论坛 12：面向机器的多媒体编解码技术：从标准框架到评价指标

时间

2024 年 7 月 24 日 下午 13:30-17:00

地点

多功能厅 6

组织人

• 虞露 (浙江大学 求是特聘教授)

• 陈志波 (中国科学技术大学 教授)

论坛介绍

大模型时代，视频编解码技术作为超高清以及泛音视频产业发展的基础，是未来产业竞争的制高点。然而，由于智能化应用视频等领域的快速发展以及面向机器视觉的视频编码和处理需求的爆发增长，传统编码工具的性能已趋于极限，难以满足通用化、智能化的视频应用需求。因此，基于神经网络技术的面向机器任务编码，将成为下一代视频编码技术发展的重要突破口，其针对机器视觉进行编码，提取关键特征信息、传输特征流而不再是单纯视频信号流，在降低传输压力的同时提升机器视觉分析任务的性能和效率，满足对视频感知、分析、理解等智能应用方面的需求。2019 年 MPEG 成立编码专家组推进机器视觉编码标准（简称 VCM, Video Coding for Machines）的发展，2020 年国内也成立了面向机器智能的数据编码工作组（简称 DCM, Data Coding for Machines），在上述背景下，本次论坛将围绕国内外编码标准制定及科研进展为核心展开讨论，介绍一线学者在相关方面的研究成果及未来思考。

报告嘉宾

嘉宾介绍

鲁 国 上海交通大学



鲁国，上海交通大学电子工程系助理教授、博士生导师。2020 年博士毕业于上海交通大学，主要研究方向是视频编码。在 IEEE T-PAMI/T-IP 及 CVPR/ECCV/ICCV 等上发表论文 30 余篇。作为项目负责人先后承担自然科学基金委青年项目、华为、平头哥等项目，参与自然科学基金重点项目等。担任权威期刊 IJCV, IEEE T-CSVT 客座编辑，在 CVPR/ACMMM 上组织多次智能编码 tutorial，担任 AAAI SPC。入选中国科协青年人才托举工程，获首届 IEEE 视觉信号处理与通信新星奖，中国图象图形学学会优秀博士论文奖，上海交通大学优秀博士论文奖。

报告题目

面向机器视觉的可控视频压缩编码

报告摘要：面向机器视觉的视频编码主要是研究在给定下游任务的情况下如何提升压缩编码效率。本报告将分析当前面向机器视觉视频编码的主要技术路径，重点探讨两个方面。一是如何提升现有传统视频编码器如 H.265 在当前场景下的效率，增强其在不同视觉任务下的能力。另外一个方面是探讨在基于端到端智能视频编

码时，如何让其面向下游任务时实现更为可控的压缩编码，灵活在不同任务间进行切换。报告将介绍我们团队面向这些任务提出的基于 Mask 的语义特征提取与分层编码架构，以及基于编码器可控的端到端视频编码。报告还将探讨当前面向视觉的视频编码面临的挑战与机遇。

嘉宾介绍

于化龙 浙江大学宁波科创中心



于化龙，浙江大学（虞露教授课题组）获得博士学位，目前在浙江大学宁波科创中心任讲师。主要研究方向为视频编码优化及其标准化、智能媒体传输和面向机器智能的数据编码。主持参与 4 项国家省部级科研项目，发表相关学术论文 8 篇，被 AVS 标准采纳近十篇技术提案，为 ISO/IEC JTC 1/SC 29 WG2/4 国际标准工作组贡献数十项技术提案，授权 5 项标准核心专利，参与发布 4 项国际和行业标准。现任 WG4 面向机器智能的视频/特征编码标准的核心实验负责人，中国面向机器智能的数据编码（DCM）标准工作组参考软件编辑人、需求文档编辑人之一，中国数字音视频编解码技术标准（AVS）的智能媒体格式标准的编辑人。

报告题目

VCM 与 DCM 核心技术与标准进展

报告摘要：计算机视觉、人工智能的蓬勃发展对驱动智能的海量数据提出了更高的压缩编码要求。主流的图像视频数据编码以视觉观看为目标，面向机器智能的数据编码（Data Coding for Machines, DCM）为智能分析的语义保真提供了高效压缩的可能。近几年，国内和国际上的标准组织都开展了面向机器智能的数据压缩方向的标准化工作：国内 DCM 标准工作组即将发布国际上首个面向机器智能的图像压缩标准，以低算力的代价相较于 VVC 能够在获得近 70% 的编码效率提示，同时，该标准支持多任务并满足特征/像素等不同层次的后端数据需求；国际 ISO/IEC JTC 1/SC 29 WG4 (MPEG Video) 标准工作组立项了面向机器的视频编码（VCM）标准，从前处理、后处理和内核编解码等方面开展工作，相较于 VVC 获得了显著的编码效率提升。本报告将主要介绍 DCM 标准和 VCM 标准的发展历程和关键技术。

嘉宾介绍

金鑫 宁波东方理工大学



金鑫，中国科学技术大学博士（陈志波教授课题组），新加坡国立大学 LV Lab 访问学者，微软亚洲研究院 MSRA 明日之星，曾获 ACM SIGAI China（国际计算机学会中国人工智能分会）优博奖、中国科学院院长特别奖。现任宁波东方理工大学（暂名）助理教授、博士生导师。在计算机视觉及深度学习领域发表论文 40 余篇，第一作者成果涵盖业内顶级会议及期刊，如 CVPR、ICCV、ECCV、ACMMM、NeurIPS、AAAI、IEEE TIP、IEEE TMM 等，谷歌学术引用超过 3200 余次，且担任这些会议及期刊论文的审稿人/PC。积极参与图像及视频压缩编码相关的国际、国内标准化（面向机器智能编码）工作。主持国家自然科学基金青年科学基金、浙江省自然科学基金、宁波市高新区重点研发项目等课题 5 项。在 CVPR 2024、ECCV 2024 以及 IEEE VCIP 2024 上组织表征解耦、组合泛化以及生成式 AI 编码技术相关的 tutorial 及 special session。

报告题目

面向机器视觉的语义解耦与智能编码

报告摘要：近年来，大数据及人工智能技术催生出面向机器的海量应用，如车联网、无人驾驶、工业互联网、智慧与平安城市、视频监控等机器视觉内容和数据，相比日趋饱和的面向人类视觉的压缩，面向机器智能的数据编码（Video Coding for Machine, VCM）应用场景更为广泛。本报告首先回顾了近年来 VCM 技术的发展历史及现状，其次总结介绍近年来 VCM 技术从任务驱动编码 特征编码 结构化语义编码的演变趋势，最后以本课题组在 VCM 领域的基于语义解耦学习的结构化编码框架技术积累和研究成果为实例，探索和展望如何高效快速地利用现有预训练大模型等 AGI/AIGC 技术，推动 VCM 的实际落地推广与应用。

嘉宾介绍



高长生 新加坡南洋理工大学

高长生，新加坡南洋理工大学博士后研究员（Weisi Lin 教授组），于 2023 年获中国科学技术大学博士学位（吴枫、刘东教授课题组）。2022-2023 年在瑞士洛桑联邦理工学院（Touradj Ebrahimi 教授组）进行访问交流，并在 2018-2019 年在微软亚洲研究院任实习研究员。主要研究方向包括图像视频编码、特征编码以及质量评价等，相关研究成果发表于 TMM、TCSVT、TIP 等相关期刊和国际视频编码领域会议。

报告题目

面向机器视觉任务编码的评价指标

报告摘要：由于图像视频数据的爆发式增长，人力已无法完成对海量数据的分析。结合计算机视觉的发展，图像视频数据的分析转而交由机器进行分析。考虑到图像视频数据的存储和传输需求，面向机器视觉的编码的研究成为近年来数据压缩的热点。面向机器视觉的编码目前主要可以分为图像编码、视频编码以及特征编码。对于三种不同的编码对象，都存在一个共有的问题：如何评价解码后图像视频或特征的质量。评价指标的关键问题在于如何使得失真度量能够真正反映语义失真，一个好的评价指标能够应用于编码的优化过程中并提升编码器的性能。本报告将以面向机器视觉的编码为基础，分别对现有的图像、视频以及特征编码工作中的评价指标展开论述，并在此基础上探讨未来具有潜力的研究方向。

嘉宾介绍



胡姚姒 香港理工大学

胡姚姒，香港理工大学博士后研究员（陈长汶教授组），于 2023 年获得武汉大学博士学位（陈震中教授课题组），2021-2022 年于微软亚洲研究院任实习研究员并获得明日之星。主要从事图像视频生成与质量评价方向的研究，在 IEEE TIP、IEEE TMM、CVPR 等国际期刊及会议上发表论文十余篇，参与多项国际标准化工作，担任 IJCV、IEEE TIP、CVPR、ICCV 等期刊会议审稿人。

报告题目

智能任务驱动的媒体质量评价

报告摘要：在机器视觉应用不断拓展的趋势下，探索任务驱动的质量评价方法不仅能助力于面向机器的编解码技术的优化，同时也能为大量数据的使用提供筛选依据。由于不同的驱动任务对质量评价的特征设计提出了不

同的要求，本报告以遥感影像智能分析为例，分析了压缩失真对于遥感影像分类的影响，并分享了在遥感场景下分类精度预测的相关工作。此外，本报告也通过介绍我们在虚拟现实、音视频通话等场景下的质量评价国际标准的制定工作，对从面向人到面向机器的质量评价场景进行了展望。

► 专题论坛 ◀

专题论坛 13：面向视频图像计算的专用硬件处理器

时间 2024 年 7 月 26 日 上午 09:00-12:00

地点 多功能厅 1

组织人

- 范益波（复旦大学 教授）

论坛介绍

视频图像处理需要强大的算力来支撑，当前视频与图像的计算主要基于 CPU、GPU 等通用处理器，在面临视频压缩、图像增强、8K 分辨率等计算上难以满足高算力、高带宽、低延迟的实时计算需求，面临严峻的技术挑战。尤其是视频技术朝向多模态、大模型、生成式方向发展，带来更为严峻的算力与内存墙难题。因此，如何通过算法与硬件的联合优化，进而实现领域专用处理器，最终完成高性能、低延迟的视频图像计算，是本论坛的主题。

报告嘉宾

嘉宾介绍



周金佳 日本法政大学

周金佳，现任日本法政大学理工学院应用信息系的副教授。本科毕业于上海交通大学，硕士和博士毕业于日本早稻田大学。曾任日本学术振兴会特别研究员，日本学术振兴机构 JST 项目负责人，日本大阪大学特聘副教授。研究方向包括基于深度学习的图像与视频压缩技术，图像生成技术，超低功耗图像传感器技术等。主持多项国家科研项目，其中包括日本学术振兴机构 JST 的大型科研项目 Presto，致力于研究开发基于压缩感知技术的视频采集及视频编码系统。曾参与设计研发了一系列视频编解码芯。在 IEEE TIP, TMM, JSSC, ISSCC 等权威期刊与国际会议上发表学术论文 100 余篇。

报告题目

超高速视频采集和编解码系统的开发

报告摘要：超高速的视频采集和编解码系统可以显著提升视频的实时处理能力，减少视频传输和播放中的延迟。高效的视频编解码算法能够在保证视频质量的前提下，极大地压缩视频数据量，从而降低存储和传输的成本。本次报告将介绍我们提出的一种高速视频采集和编解码系统。我们把视频的采集，编码和解码作为一个整体进行设计开发。该系统结合了压缩感知技术，模拟域和数字域协作编码技术，和基于深度学习的编解码协同训练技术。

嘉宾介绍



徐科 中兴微电子

徐科博士于 2000 年及 2003 年在复旦大学分别获得学士及硕士学位，2007 香港中文大学获博士学位，2009 年加拿大多伦多大学进行博士后研究，2021 年被香港中文大学授予"Distinguished Alumni Award"(全球仅 8 位)。曾在 IBM, Qualcomm 加拿大多伦多及美国圣地亚哥总部就职，先后担任过资深工程师，芯片架构师，研发科学家等职位，并获“高通之星”等荣誉称号。回国后加入中兴微电子，担任人工智能首席科学家及芯片首席架构师，移动网络和移动通讯多媒体技术国家重点实验室学术带头人、研究员，电子科技大学客座教授。主要研究方向包括计算机视觉及人工智能，大算力人工智能芯片设计，高性能图形图像处理芯片等。共发表国际学术论文 30 余篇，申请专利 160 余件。

报告题目

人工智能硬件处理器设计

报告摘要：报告首先介绍了一个端到端基于神经网络的图像处理算法，包括图像处理、低照度增强、以及超解析，可以在高通 8gen1 处理器上达到 2K60fps 的实时性能；报告随后介绍了各种人工智能处理器架构，包括中兴微电子前后三代基于 DSA 的神经网络加速器。

嘉宾介绍



殷海兵 杭州电子科技大学

殷海兵，男，杭州电子科技大学教授，博士生导师，入选浙江省钱江特聘专家、浙江省中青年学科带头人、151 人才培养计划。主持国家科技部重点项目 1 项、研发课题 1 项、NSFC 基金项目 3 项及其他省部级项目 10 余项，参与国家及省部项目 10 余项。在 IEEE TCSVT TIP 等期刊发表论文 50 余篇，DCC、ISCAS 等会议论文 40 余篇，出版著作 2 部。获得包括浙江省科技进步一等奖、浙江省高校科研成果二等奖等奖励 4 项。主要研究方向为视频编解码、视觉智能算法和芯片架构设计与验证等。

报告题目

面向智能计算的多层次优化存储系统与应用

报告摘要：视频、点云等数据智能处理应用日益广泛，边缘设备急剧增加，边缘数据呈现爆炸式增长，面临大量非结构化数据汇聚难、存储难、处理难挑战。基于深度学习的高效检索、紧凑编码、智能分析等智能算法计算需求极其巨大，传统边缘计算存储架构面临存储墙、功耗墙和延时访问等挑战，难以满足应用场景数据高速访问和实时处理需求，急需解决云端边缘存储计算架构设计优化与低延时高通量数据访问处理问题。本报告介绍面向智能计算的分布式边缘存储应用方面前期探索，包括存算一体架构、分布式边缘存储、编程模型框架、高效检索和端边云协同计算等，分析提升海量数据高效计算存储和智能处理效率的可能路径。

嘉宾介绍



王文强 阿里云

王文强，阿里云服务器研发资深技术专家，2008年毕业于中国科学院研究生院，从事多年图像处理的FPGA硬件方案设计。先后参与了QQ相册、微信朋友圈、淘宝、支付宝等多款业界知名产品的图像处理后台硬件加速方案设计。

报告题目

大规模图像处理架构演进及硬件加速业务分享

报告摘要：主要介绍阿里云的图像编解码硬件加速技术与业务分享。涵盖以下几个方面：架构方面，介绍阿里图像处理系统硬件架构的演进路线，介绍存储计算分离方案的落地，结合业务特性寻求存储和计算的平衡。吞吐方面，介绍如何通过软硬件一体的优化达到单机上万 QPS 的极致吞吐；图像压缩性能方面，重点介绍 JPEG/Webp/Heic/AVIF 等图像处理产学研合作的相关成果。最后，介绍数据中心适合硬件加速的业务场景及解决的关键问题，同时讨论在多媒体处理方面的硬件架构演进的几个方向。

嘉宾介绍



范益波 复旦大学

范益波，复旦大学微电子学院教授。本科毕业于浙江大学，硕士毕业于复旦大学，博士毕业于日本早稻田大学。主要从事视频图像、人工智能、硬件架构与 SoC 芯片设计研究。出版专著 2 部，发表学术论文 140 多篇，取得发明专利 50 多项。获上海市“东方学者”特聘教授、浙江省“钱江特聘专家”等。创建 OpenASIC 开源芯片论坛，发布 xk264、xk265 开源视频编码处理器 IP 核、xkISP 开源图像处理器 IP 核、xkDLA 开源神经网络处理器 IP 核。从事图像 ISP 处理器、视频 VPU 处理器、人工智能 NPU 处理器的前沿技术与处理器 IP 核开源。

报告题目

面向新一代标准的视频编码处理器硬件架构

报告摘要：视频编解码（Codec）处理器是图像、视觉芯片中最核心的处理模块，其中涉及非常多的算法研究与硬件架构优化问题，特别是视频编解码还是各大国际标准组织推进行业发展的重点领域，潜在科技壁垒和商业价值巨大。本次演讲主要涉及如下几个方面：1. 视频编解码专用硬件处理器研究介绍；2、新一代视频编码标准—VVC、AVS3、SVAC3. 3、新一代视频编解码处理器硬件架构--xkVVC；4、面向下一代安防的视频编解码处理器--xkSVAC；5. 处理器开源与未来研究方向。

► 专题论坛 ◀

专题论坛 14：多模态大模型和可控生成

时间 2024 年 7 月 26 日 上午 09:00-12:00

地点 多功能厅 2

组织人

• 李 玺（浙江大学 求是特聘教授）

• 李 辉（江南大学 讲师）

论坛介绍

随着人工智能领域的不断发展，多模态大模型和可控生成技术日益成为关注的焦点。多模态基础模型、多模态感知理解、大模型智能体、图像视频以及 3D 生成技术等，在过去的几年里，已经取得了显著的进展，为图像、语言、音频等多种模态数据的处理和生成提供了强大的工具和方法。然而，由于不同模态空间信息的表示不同、训练数据的固有偏见以及上下游任务的领域差异等，使得多模态大模型与可控生成还存在诸多挑战。本次论坛将围绕多模态大模型和可控生成的发展，探讨当下先进的技术方案以及未来的发展趋势。本论坛旨在为学术界和产业界的研究人员、工程师和从业者提供一个交流和分享的平台，共同探讨多模态大模型和可控生成技术的最新进展、挑战和应用。通过此次交流，有望推动这些技术的发展，加速其在各个领域的应用和落地。

报告嘉宾

嘉宾介绍



韩晓光 香港中文大学（深圳）

韩晓光博士，现任香港中文大学（深圳）理工学院助理教授，校长青年学者。他于 2017 年获得香港大学计算机专业博士学位。其研究方向包括计算机视觉和计算机图形学等，在该方向著名国际期刊和会议已发表论文近 100 篇，包括顶级会议和期刊 SIGGRAPH(Asia), CVPR, ICCV, ECCV, NeurIPS, ACM TOG, IEEE TPAMI 等。他曾获得吴文俊人工智能优秀青年奖，广东省杰出青年基金资助，香港中文大学（深圳）青年科研奖。

目前也担任 CVPR2023/2024, NeurIPS 2023 以及 ECCV2024 领域主席，同时也是 IEEE TVCG 以及 Computer&Graphics 的编委。他的工作曾两次获得 CCF 图形开源数据集奖（DeepFashion3D 和 MVMingNet），曾两次入选 CVPR 最佳论文列表。

报告题目

三维生成：从类别限定到开放世界

报告摘要：本报告将包括以下部分：首先，我将快速回顾一下我们组过去几年所做的一些工作，这些工作关注特定类别的 3D 内容重建或生成。接下来，我将介绍我们最近在开放世界三维生成方面的工作。最后，我将分享一些关于如何构建 3D 基础模型的想法。

嘉宾介绍



晏轶超 上海交通大学

晏轶超，上海交通大学人工智能研究院助理教授，博士生导师。获上海交通大学电子工程学士、博士学位，法国里昂中央理工学院硕士学位，曾担任阿联酋起源人工智能研究院研究科学家。主要研究方向为计算机视觉、图形学技术及其在虚拟现实、数字多媒体中的应用，发表包括 TPAMI、CVPR、NeurIPS 在内的国际高水平论文 30 余篇，担任十余个国际顶级会议与期刊审稿专家。主持国家自然科学基金青年项目、CCF-阿里巴巴青年科学家基金等项目 5 项。入选上海市海外高层次人才计划，获 2020 年度中国图象图形学学会优秀博士论文奖。

报告题目

多模态引导的可控三维数字人生成

报告摘要：面对大规模三维数字人的生产需求，结合生成式模型与传统图形学建模产生高拟真、规模化的三维数字人正逐渐成为研究热点。本次报告主要介绍如何结合大模型进行可控三维数字人的生成，包括三维人脸重建、三维人脸编辑以及人体动作生成。三维人脸重建使用大模型引导人脸材质重建，三维人脸编辑使用拖拽方式实现三维人脸几何结构的快速编辑，人体驱动使用文本生成三维数字人的行为交互。

嘉宾介绍



于云龙 浙江大学

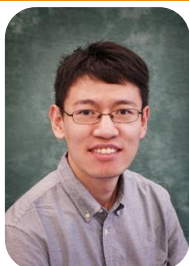
于云龙，浙江大学信电学院研究员，博士生导师。从前从事计算机视觉和机器学习在小样本场景下的理论研究工作，在国际知名期刊和国际会议上发表论文 40 余篇，包括 NeurIPS, CVPR, IJCAI, ECCV, TIP, TNNLS, CSVT 等，申请专利 10 余项。曾作为项目负责人主持国家自然科学基金青年项目，并参与国家自然科学基金重点项目课题等。在多模态学习、机器学习理论等领域积累了丰富的研究经验。

报告题目

多模态大模型在有限样本场景下的泛化

报告摘要：基础模型因其在大量感知任务上的可扩展性和出色的效果而受到计算机视觉领域的重视。然而，如何将预训练模型在少样本和零样本情况下的迁移到下游任务存在挑战。在本报告中，我们将介绍几种迁移学习和提示微调技术，专门解决在有限数据条件下微调视觉基础模型上的挑战，以用于图像识别、遥感分割和视频理解等下游任务。

嘉宾介绍



王高昂 浙江大学

王高昂，浙江大学国际联合学院研究员，伊利诺伊大学厄巴纳香槟校区兼聘助理教授，博士生导师。在高质量国际期刊及国际会议上共发表论文 68 篇，包括 IJCV、TIP、CVPR、ICCV 等中国科学院 Top 期刊和 CCF A 类会议论文 21 篇。在 2017-2019 年度的英伟达智慧城市挑战赛的 5 个赛道中分别取得 4 次冠军、1 次亚军成绩。主持和参与多项国家及省

部级课题，包括国家自然科学基金、科技创新 2030-“新一代人工智能”、浙江省自然科学基金重大及重点项目等。

报告题目

基于多模态大模型的长视频理解

报告摘要：随着大模型技术的发展，目前的视频理解系统通常会将视频基础模型和大语言模型进行结合，可有效克服特定预定义视觉任务的限制。然而，现有视频理解系统大多数只能处理较短视频。对于长视频，计算复杂度和内存成本方面会带来极大挑战。除此之外，将预训练的单帧视觉模型迁移到长视频任务，往往需要大量训练数据进行时序建模，这会带来灾难性遗忘以及额外的计算代价。在本次报告中，我们将汇报如何借助长短时记忆机制，通过无训练（training-free）的方式，将基础视觉模型有效迁移到万帧级别的长视频理解任务中。

► 专题论坛 ◀

专题论坛 15：未来视觉信息编码与智能技术

(铂金赞助-快手冠名)

时间

2024 年 7 月 26 日 上午 09:00-12:00

地点

多功能厅 3

组织人

- 翟广涛 (上海交通大学 特聘教授)
- 朱树元 (电子科技大学 教授)
- 贾川民 (北京大学 助理教授)

论坛介绍

视觉信息是人类与现实世界交互的主要承载方式。近年来，视觉信息的内容、形式和数据总量呈现多维度增长，全息数据、三维视觉、8K 超高清等未来视觉信息内容存储和传输代价巨大，迫切需要进一步提升编码压缩效率。未来视觉信息编码技术作为核心研究主题正面临新一轮技术挑战，全真视觉体验也在各种智能技术的赋能下迎来了新机遇。本论坛聚焦未来视觉信息编码以及智能技术发展这一前沿交叉主题，邀请学术界和工业界资深专家进行报告并参与圆桌讨论。内容涵盖三维全息高效视频技术、语义智能编码、面向机器视觉的压缩编码及标准化等主题。计划为视觉智能技术领域的专家学者、从业人员和研究生提供交流平台，探讨领域核心科学问题，推动未来视觉信息编码进程，促进多媒体领域教育、科技、人才三位一体发展。

报告嘉宾

嘉宾介绍



马思伟 北京大学

马思伟，IEEE Fellow，北京大学博雅特聘教授，视频与视觉技术国家工程研究中心副主任，国家杰出青年科学基金、科学探索奖获得者。长期从事高效视频编码研究，聚焦于资源约束条件下的率失真优化编码理论与方法。2020 年入选国家杰出青年科学基金资助项目，已发表国际期刊论文 100 余篇，已获授权发明专利 80 多项。自 2002 年起，陆续参与组织 AVS 系列国家标准制定，担任 AVS 工作组视频组组长。他承担过国家重点研发计划、国家自然科学基金重点、863 计划、国家科技支撑计划等多项国家级项目；曾获国家技术发明奖一等奖、国家技术发明奖二等奖、国家科学技术进步奖二等奖、中国电子学会科学技术奖技术发明特等奖、北京市科学技术进步一等奖等奖励；担任中国图象图形学学会理事、会员发展与服务工委主任、多媒体专委会副主任。在超高清视频高效压缩、视觉失真度量和率失真优化编码方面，他取得了一系列重要创新成果，主导制定了超高清视频编码国家标准，研制的 AVS2 4K/AVS3 8K 超高清编解码器，支撑了央视 CCTV-4K/CCTV-8K 超高清频道的开播。

报告题目

沉浸式媒体智能技术与应用

报告摘要：全景超高清、三维自由视角视频等沉浸式视觉媒体在赛事直播、文艺演出、文物展陈、全息视讯等应用场景创新得到广泛关注，相比于传统视频应用，沉浸视频编解码、传输和交互都需要新的技术、标准支撑，近来发展迅速，预示着更广阔的发展前景和市场潜力。本报告主要介绍团队在沉浸式媒体技术与应用方面的最新工作进展情况，包括沉浸式视频编解码技术及系统应用、沉浸式媒体内容生成技术及沉浸式文旅应用等内容。

嘉宾介绍



雷建军 天津大学

雷建军，讲席教授，博士生导师/硕士生导师，天津大学科学技术发展研究院副院长。2007年毕业于北京邮电大学，获信号与信息处理专业博士学位。现工作于天津大学电气自动化与信息工程学院信息与通信工程学科。2012年8月至2013年8月，在美国华盛顿大学做访问学者。在国内外学术期刊及会议上发表SCI/EI论文100余篇，包括IEEE Trans长文50余篇；作为负责人及主要参加人参加了国家自然科学基金、国家重点研发计划、国家863计划、“新一代宽带无线移动通信网”国家科技重大专项、国家国际科技合作项目、天津市自然科学基金、天津市人工智能重大专项等多项科研项目；获国家杰出青年科学基金、优秀青年科学基金、天津市杰出青年科学基金、天津市科技进步一等奖（排名第1）、天津市技术发明一等奖（排名第2）、天津市优秀硕士论文指导教师、天津市工程专业学位优秀指导教师，入选了天津市中青年科技创新领军人才、天津市131人才工程、北洋学者计划。

报告题目

3D 视频智能编码与处理方法

报告摘要：人类生活的世界是3D立体的，由于受技术条件的限制，传统的显示技术采用2D平面的影像表达和传递对世界的认识。继黑白、彩色、高分辨率信息在屏幕上完美再现之后，3D视频技术成为了新一代信息技术的发展方向 and 趋势。3D视频能够再现真实景物的立体信息，提供沉浸式的立体视觉感受和灵活的交互式体验，在多个领域都具有广阔应用前景。报告将介绍3D视频基本原理、3D视频系统面临的主要挑战阐述3D智能视频编码与处理方法，并展望可能的研究方向。

嘉宾介绍



李 礼 中国科学技术大学

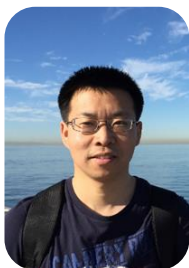
李礼，博士，中国科学技术大学电子工程与信息科学系特任教授、博士生导师。2011年、2016年在中国科大分别获得学士、博士学位。他的主要研究领域是多媒体压缩，在国际期刊和会议上发表论文80余篇，谷歌学术引用超过3000次。获授权专利20余项，被国际、国内标准化组织采纳的技术方案10余项。获2019年国家技术发明二等奖（序5），获高等教育（本科）国家级教学成果奖二等奖（序12），获2023多媒体新星奖。获VCIP 2016最佳10%论文奖和ICIP 2019最佳10%论文奖。获2023年基于神经网络的视频编码挑战赛冠军（端到端赛道）和2016年ICME光场图像压缩挑战赛冠军。担任视频编码领域顶级期刊T-CSVT副编辑以及客座编辑，担任ICME 2021、MMSP 2024、VCIP 2023、VCIP 2022和VCIP 2016组委会成员。

报告题目

面向端到端视频编码的帧间预测技术

报告摘要：混合视频编码框架经过数十年的发展已经进入瓶颈，基于统计信号处理进一步提升视频编码效率举步维艰。近年来，端到端视频编码技术展现出了很大的潜力，为视频编码领域带来了新的生机。本次报告将分享课题组在端到端视频编码领域的最新研究进展，特别是在帧间预测技术方面的一些成果，主要包括两个方面：一方面，我们致力于进一步挖掘视频帧之间的相关性，以提升视频编码效率；另一方面，我们致力于进一步提升视频帧间预测的人机适配性，以同时服务于人类和机器视觉。

嘉宾介绍



郭 勐 中国移动研究院

郭勐，博士，现任中国移动研究院业务研究所技术经理，中国通信标准化协会 TC1 WG3 工作组副组长。研究领域包括移动视频核心能力及产品研发、新型多媒体业务设计以及标准制定。

报告题目

未来视觉应用与技术发展

报告摘要：沉浸媒体技术正推动着娱乐、教育、医疗、军事、旅游以及零售等众多行业的变革。本报告将围绕视频语义编码、3D 数据编码、质量评测及标准化等方面进行分享。

嘉宾介绍



张园 中国电信研究院

张园，中国电信股份有限公司研究院大数据与人工智能研究所副所长。担任 ISO/IEC JTC1 SC29 机器视觉编码专家组召集人、DCM 面向机器智能的数据编码国家标准工作组副组长、ITU-T SG16 WP3 音视频技术与沉浸式应用工作组主席、Q12/16 智慧视频课题组报告人、FG-MV WG2 元宇宙应用与系统工作组主席等 5 项国际任职和 5 项国内任职。曾获中国标准创新贡献二等奖、中国电子学会科技进步一等奖、中国通信标准化协会科学技术一等奖等二十余项省部/学会/企业/个人奖项。牵头制定国际标准二十余项，发表论文 10 余篇。

报告题目

机器视觉编码国际国内标准进展

报告摘要：报告将介绍机器视觉的概念起源，机器视觉的系统组成和技术发展、机器视觉相关标准组织及工作进展、典型应用场景及产业化，包括 ISO/IEC JTC1、ITU-T、SAC 等机器视觉编码、应用、系统测试等相关标准组织的工作及最新进展，智慧工业、智慧安防、智慧交通等产业应用，结合机器视觉研究中心工作探讨未来技术路径和标准化工作方向，以及对垂直行业与通信产业的价值和意义。

嘉宾介绍



谷嘉文 快手

谷嘉文，博士，2020年毕业于清华大学计算机系，2014年美国数学建模比赛特等奖，2020年北京市优秀博士毕业生、清华大学计算机系优秀博士毕业生。博士主攻视频编解码算法研发，累计发表论文12篇。在博士阶段深度参与编解码相关业务和算法的落地，完成4K HEVC全景VR拼接+直播系统、8K HEVC 60FPS实时编码系统。全景拼接系统荣获日内瓦发明展金奖，应用于瑞士蒙特勒爵士音乐节展厅；8K HEVC编码器是当年国内首个实现实时的HEVC编码直播系统，最终应用于2019年男篮世界杯8K+5G转播活动和2019国庆70周年“8K+5G”慢直播活动。毕业后入职快手，从事智能编解码算法的研发工作，工作期间相关上线档位合计为公司带来数亿带宽节省，发表专利30余篇。

报告题目

智能编解码前沿探索与形态

报告摘要：随着深度学习和大模型的火热，其在视频编解码的领域上的应用也受到业界的持续关注。在当前，深度学习算法已经在视频编解码压缩上发展迅速。然而，受限于昂贵的端上算力，智能编解码方向的应用和普及还存在一些距离。本次报告将结合快手研发成果和线上业务场景，介绍了智能算法在视频编解码方向的应用，并会分享当前一些可行性方向的探索成果和对外来智能编解码形态的展望。

► 专题论坛 ◀

专题论坛 16：面向跨媒体内容的智能可视化计算

时间 2024 年 7 月 26 日 上午 09:00-12:00

地点 多功能厅 4

组织人

- 赵一凡（北京航空航天大学 副教授）
- 蒋 铨（北京航空航天大学 副教授）
- 朱 林（北京理工大学 助理教授）

论坛介绍

随着近年来跨媒体信息内容的快速涌现，传统的单一视觉媒体已经难以适应现实应用对物理环境的精确感知和计算需求。图像、视频、文本、音频以及海量的结构化数据为可视内容的高效计算提供了重要的数据原料支撑。以跨模态、跨视图计算为核心的跨媒体内容融合互联，已逐步成为实现交通安防、卫星通信、数字孪生等重要应用的关键技术底座。本论坛以“如何利用跨媒体间的知识结构内在关联，实现智能化、精细化的可视内容计算”为核心，探讨多源异构、多视角跨媒体互补学习以及图像、视频、文本等跨模态媒体关联融合的关键技术。旨在探究跨媒体内容间的内在知识结构关联，为视觉内容的跨媒体可视计算提供新颖的研究思路和启发。本论坛聚焦跨媒体内容智能可视计算的关键技术和应用，邀请多位领域专家和知名学者介绍相关研究进展和前沿发展趋势，围绕细粒度多模态分析与生成、多模态视觉结构学习、生成式 AI 技术等主题开展深入探讨。

报告嘉宾

嘉宾介绍



彭宇新 北京大学

彭宇新，北京大学二级教授、博雅特聘教授、国家杰出青年科学基金获得者、国家万人计划科技创新领军人才、科技部中青年科技创新领军人才、863 项目首席专家、中国人工智能产业创新联盟专家委员会主任、中国工程院“人工智能 2.0”规划专家委员会专家、中国电子学会会士、中国人工智能学会会士、中国图象图形学会会士、副秘书长、提名与奖励委员会副主任、北京图象图形学会副理事长。主要研究方向为跨媒体分析、计算机视觉、机器学习、人工智能。以第一完成人获 2016 年北京市科学技术奖一等奖和 2020 年中国电子学会科技进步奖一等奖，2008 年获北京大学宝钢奖教金优秀奖，2017 年获北京大学教学优秀奖。主持了 863、国家自然科学基金重点等 30 多个项目，发表论文 200 多篇，包括 ACM/IEEE Trans 和 CCF A 类论文 100 多篇。多次参加由美国国家标准技术局 NIST 举办的国际评测 TRECVID 视频样例搜索比赛，均获第一名。主持研发的跨媒体互联网内容分析与识别系统已经应用于公安部、工信部、国家广播电视总局等单位。担任 IEEE TMM、TCSVT 等期刊编委。

报告题目

细粒度多模态运动分析

报告摘要：运动是生命的基本要素，是现实世界的客观描述。与粗粒度、单模态的运动分析不同，细粒度、多

模态的运动分析描述真实世界更准确，同时更具挑战性。细粒度多模态运动分析旨在利用文本、视频、3D 姿态等不同模态信息的推理能力对目标动作进行细粒度识别、定位、评价、重建等，关键科学问题是如何精准辨识时空边界模糊的细粒度子动作，如何利用文本信息对人体动作进行精细重建等。可广泛应用于智能安防、智慧医疗、智能体育、智慧传媒等领域。本报告将介绍我们最新的 6 篇 CVPR 2024 论文，涵盖了我们在细粒度多模态运动分析的近期工作进展，包括细粒度动作质量评价、文本驱动细粒度时空动作定位、文本驱动的三维人体姿态估计等，并对未来发展方向与趋势进行讨论与展望。

嘉宾介绍



李 玺 浙江大学

李玺，浙江大学求是特聘教授，IET Fellow，IEEE Senior Member，国家杰青，国家青年特聘专家，主要从事人工智能领域研究，提出了多因子耦合模型学习新理论，探索了知识引导的模型结构设计与搜索新方法，开拓了复杂异构图像结构语义理解的新技术。担任国际权威期刊的编委和顶级会议的 Area Chair，获得 2021 年中国图象图形学会自然科学奖二等奖，2021 年中国电子学会科技进步一等奖，2021 年中国产学研合作促进会产学研合作创新与促进奖，2022 年世界互联网领先科技成果，2022 年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，2023 年中国发明协会发明创业奖创新奖一等奖，2023 年陆增镛 CAD&CG 高科技奖一等奖，两项中国北京市自然科学技术奖（包括一等奖和二等奖），以及一项中国专利优秀奖。荣获四项最佳学术论文奖、腾讯好专利、华为优秀合作成果奖和火花价值奖。成果应用于华为、阿里、海康等企业。

报告题目

多模态视觉结构学习

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多模态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，进行大规模图像/视频数据的视觉特征学习，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，并引入了大规模多模态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。

嘉宾介绍



邓 成 西安电子科技大学

邓成，男，西安电子科技大学二级教授、博士生导师。国家级高层次人才，国家百千万人才工程入选者，国家有突出贡献中青年专家，陕西省重点科技创新团队负责人，陕西省教学名师，陕西省师德标兵，教育部电子信息类教学指导委员会秘书长。长期从事人工智能领域的研究工作，主持国家自然科学基金重点项目、科技部重点研发计划等项目近 30 项。在本领域国际一流期刊和 CCF A 类会议上发表论文 200 余篇，连续多年入选爱思唯尔中国高被引学者榜单。研究成果获 2019 年、2023 年陕西省自然科学一等奖（第 1）、2016 年国家自然科学二等奖（第 3）。

报告题目

生成式 AI 的缘起、发展与挑战

报告摘要：生成式 AI 代表了机器学习与计算机视觉领域的一次革命性跃迁，其核心价值在于赋予机器创造高质量、连贯数据的能力，无论是文本、图像、或是视频。本报告首先将追溯生成模型的起源，分析其从基础模型到大模型的演变过程，并聚焦于这一历史进程中的关键技术突破与理论创新。我们还将探讨生成式大模型的最新发展，并讨论其在图像、文本、音频、视频生成等多个领域的显著成就。最后，本报告将对生成式 AI 的未来前景进行前瞻性思考，探讨如何通过提升模型的生成质量、稳定性、鲁棒性以及可解释性，推动生成式 AI 的进一步发展，以及其对人类文化、社会价值和道德伦理的深远影响。

嘉宾介绍



贾 伟 合肥工业大学

贾伟，博士，合肥工业大学计算机与信息学院教授，博士生导师，智能科学与技术系系主任。中国图象图形学学会青年工作委员会副主任兼秘书长。中国自动化学会模式识别与机器智能专业委员会常务委员。20 余次担任国际国内知名会议的程序主席、组织主席等。已发表 SCI 论文近百篇，其中在 CVPR、ICCV、IEEE Transactions 会刊等国际顶级会议及权威期刊上发表论文 CCF A 类及中国科学院 1 区论文 50 多篇。主要研究兴趣为人工智能、生物特征识别、计算机视觉、模式识别等。

报告题目

人的生物特征生成与可视化

报告摘要：身份认证已成为数字化和智能化社会的基础服务之一。生物特征识别是进行身份认证最为有效的技术手段之一。在过去的 10 年中，深度学习成为生物特征识别最为主流的处理技术。深度学习的模型训练需要大量的样本，然而，由于隐私保护等问题，采集真实的生物特征数据并公开发布愈发困难，在这种情况下，使用生成方法合成高真实感的生物特征数据已经为生物特征识别技术中数据集构建的重要手段。本报告将对主流的生物特征模态的图像生成方法进行介绍，包括指纹图像生成、人脸图像生成和虹膜图像生成等。掌纹识别和掌静脉识别是非常有前景的生物特征识别技术，本报告也将介绍研究团队在掌纹和掌静脉图像生成方面的最新工作。

嘉宾介绍



王立志 北京理工大学

王立志，北京理工大学准聘教授、博士生导师。主要研究领域为计算摄像与图像处理，主持多项国家级项目，包括国家优秀青年科学基金项目、“应用创新”装备预研。发表论文 50 余篇，担任 IEEE TIP 编委。荣获 CCF A 类会议 ACM MM 2022 最佳论文提名奖、中国电子学会 2018 年度优秀博士学位论文奖、IEEE VCIP 2016 最佳论文奖。

报告题目



数据视角下的智能图像降噪

报告摘要：图像降噪是计算摄像的经典任务，在多个领域具有实际应用需求。依赖于深度学习技术，智能图像降噪已经成为图像降噪的主流解决方案，其关键问题是构建带噪图像与清晰图像之间的数据映射，其中数据扮演着决定性的作用。本次报告将从数据视角下介绍智能图像降噪的最新进展，包括数据采集、数据增强、以及数据生成，并展望未来发展趋势。

► 专题论坛 ◀

专题论坛 17：多媒体智能传播

时间

2024 年 7 月 26 日 上午 09:00-12:00

地点

多功能厅 5

组织人

• 刘安安（天津大学 教授）

• 靳国庆（人民网股份有限公司 研究员）

论坛介绍

随着生成式人工智能时代的到来，海量的图像、视频、文本等跨媒体数据呈现爆炸式增长。在这种情况下，多模态媒体的智能传播显得尤为重要，如何实现面向多源异构、相互关联的多模态媒体数据的跨媒体认知能力学习，联合实现面向视觉、语言等跨媒体信息动态抽取和场景理解，塑造基于大模型引导的生成式个性化推荐系统，以及防范内容和传播风险，对于提升用户体验保障传播安全都至关重要。本论坛将聚集来自工业界和学术界的专家，讨论议题涵盖智能传播中多模态信息的生成、分析、推荐与传播，以及具体应用实践。多媒体智能传播是一个充满挑战和机遇的研究领域，也是数字化时代媒体传播发展的必然趋势。

报告嘉宾

嘉宾介绍



冯福利 中国科学技术大学

冯福利，中国科学技术大学特任教授。研究领域：信息检索、数据挖掘、机器学习等，发表国内外顶级会议和期刊论文 100 篇，谷歌学术引用近 10000 次，承担科技部重点研发计划项目课题、基金委面上项目等国家级项目，研究成果在多家公司的商业系统应用。曾获 SIGIR 2021 最佳论文提名奖、WWW 2018 最佳演示论文奖。担任 Frontiers in Big Data 编委 (AE)，众多顶级期刊审稿人，会议 SPC/PC，包括 SIGIR、WWW、SIGKDD、NeurIPS、ICML、ICLR、ACL、TOIS、TKDE、TPAMI、TNNLS、Nature Sustainability。

报告题目

大模型驱动的引导式推荐初探

报告摘要：随着推荐系统在内容分发传播过程中的作用日益增长，面向信息茧房破除、有益兴趣培养、认知能力提升等目标的引导式推荐成为信息检索领域的前沿问题。本报告介绍引导式推荐的核心问题与最新进展，包括兴趣驱动的引导序列优化、基于溢出效应的社交引导策略优化。同时，大语言模型以及其驱动的智能体因丰富的世界知识和突出的规划能力在复杂任务中取得了惊人的效果。本报告介绍大模型智能体在引导式推荐场景下的前沿探索，包括面向长期推荐效能的智能体学习、大模型驱动的引导路径规划。最后，本报告将展望引导式推荐与大模型推荐智能体未来的发展方向。

嘉宾介绍



田 天 瑞莱智慧 RealAI 公司

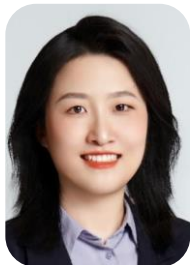
田天，瑞莱智慧 RealAI 公司联合创始人及首席执行官。2018 年博士毕业于清华大学计算机系，高级工程师，他致力于建设安全可控的人工智能生态，发表人工智能顶级国际会议和期刊论文十余篇，参与了数十项国家/行业/团体标准制定工作，2021 年公司研发成果“新一代安全可靠可控的人工智能平台”入选乌镇峰会世界互联网领先科技成果。田天还担任上海人工智能战略咨询委员会人工智能治理专委会委员，曾荣获“吴文俊人工智能优秀青年奖”，入选北京市科技新星计划。2019 年曾入选福布斯中国“30 岁以下精英榜”。

报告题目

通用人工智能的安全、对齐与治理

报告摘要：随着通用人工智能技术在各类场景落地应用，其自身技术不足引起的新型安全隐患带来越来越多实际事故和损失，与此同时垂直领域应用，包括智能传播中，需要更好的对齐，此外隐私泄露、深度伪造等滥用问题也亟需治理手段。RealAI 提出了一系列技术和平台产品，让 AI 应用更加安全、可靠、可控。

嘉宾介绍



王岚君 天津大学

王岚君，天津大学新媒体与传播学院研究员，博士生导师。加入天津大学前，她曾任 IBM 中国研究院高级主任研究员、华为云技术专家。她曾作为华为云模型可信负责人，参与建立了业界首个可实施的人工智能云服务安全标准，该成果载入 2021 年华为公司企业年报。此外，她已发表高水平论文 40 余篇，拥有授权专利 20 余项。她的主要研究方向为：可信人工智能。

报告题目

基于社交机器人的多媒体传播安全可控

报告摘要：现今在智能传播领域，对于安全可控问题的讨论往往局限于对多媒体内容的静态审核与处理，而忽略了在传播过程中所面临的安全挑战。然而，随着社交机器人的兴起，我们有了一种新的载体来应对这些挑战。本报告将深入探讨社交机器人在多媒体传播中的作用，重点关注其在虚假信息检测、虚假用户识别以及传播效果方面的影响。通过分析社交机器人的可控性，我们将探讨如何利用这一技术手段来确保多媒体传播链路的安全可控性，为智能传播的发展开辟新的可能性。

嘉宾介绍



魏龙辉 华为技术有限公司

魏龙辉，华为终端云高级研究员，2023 年博士毕业于中国科学技术大学。当前主要研究领域为视觉预训练技术、多模态理解\生成\对话大模型以及其他开放场景学习领域研究。目前已累计发表 30+ 领域顶级会议\期刊论文，谷歌学术引用 3500 余次，单篇一作最高引用 1800 余次。担任顶会 ACM MM' 24 领域主席，以及 CVPR/ICCV/ICML 等领域顶会及顶刊审稿人，连续获得 AMiner 评选的 2022 年以及 2023 年多媒体领域最具影响力学者提名。

报告题目

多模态对话与生成大模型技术前沿介绍与思考

报告摘要：多模态大模型已经成为当前多媒体领域乃至整个人工智能领域最为火热的研究方向之一，国内外相关大型研究机构在上述领域也不断提出多个变革性算法来推动领域的快速发展。本次报告将聚焦于回顾近年来多模态对话与生成大模型技术的整体发展脉络和代表性算法、并简单介绍本人在上述领域的研究工作，最后将重点分析多模态对话与生成大模型领域所面临的挑战并展望未来的发展方向。

► 专题论坛 ◀

专题论坛 18：面向开放世界的可信多模态通用大模型研究

时间 2024 年 7 月 26 日 上午 09:00-12:00

地点 多功能厅 6

组织人

• 吴建龙（哈尔滨工业大学（深圳）副教授）

• 张 正（哈尔滨工业大学（深圳）副教授）

论坛介绍

随着人工智能的飞速发展，我们正见证着一个划时代的转变——各种模态，如视觉、语言和音频等，原本彼此独立，现如同人类的五感一般，逐渐交织融合，极大地推动人工智能朝着一个全新的“通感”时代迈进，其中多模态大型模型发挥了非常重要的角色。当前，多模态机器学习正从局限于特定数据集的封闭型学习，向能够适应动态变化的数据和实时环境的开放型大模型研究和应用迈进。这些开放型多模态大模型展现了在不确定语义、变化环境、甚至对抗性干扰下的强大适应和推理能力，能够充分利用不同模态间的互补信息，提升整体性能。本论坛旨在邀请多模态领域的顶尖专家和学者，分享他们在可信多模态内容感知与理解、开放场景下大模型的增量学习、大模型的高性能压缩及高效训练等方面的前沿技术和研究成果。我们将深入探讨在开放动态环境中研究和应用多模态大模型的最新技术趋势和挑战，为参与者提供一个关于多模态学习未来方向的全面视角。

报告嘉宾

嘉宾介绍

田永鸿 北京大学



田永鸿，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，2024 年首批国家杰出青年基金延续资助计划获得者。主要研究方向为分布式机器学习、脉冲神经网络和神经形态视觉。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 Nature/Science 子刊、IEEE Trans 等国际期刊和 ICML、NeurIPS 等国际会议发表学术论文 350 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、2023 年广东省科技进步特等奖、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，多个国际期刊编委和国际会议大会主席/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。

报告题目

多模态大模型：技术挑战与行业应用实践

报告摘要：在新一轮人工智能浪潮中，国内大模型产业迅速发展，极大地促进了科技创新和产业升级。多模态大模型是当前大模型技术的最重要发展趋势，它通过融合多种数据类型来显著提高了对复杂现实问题的处理能力。本报告将详细分析多模态大模型在数据收集、视觉编码器训练、多模态对齐以及应用部署等方面所遇到的技术挑战，并以面向空间治理的大模型为例，探讨其在视觉编码处理和小目标异常检测等关键技术的应用，以及该模型在林草、国土、水利等多个行业中的实际部署和应用流程，展示多模态大模型在实现广泛业务覆盖和高效资源利用方面的能力。

嘉宾介绍



孟德宇 西安交通大学

孟德宇，西安交通大学教授，博导，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，谷歌学术引用超过 29000 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。

报告题目

机器学习的数学技术

报告摘要：以深度学习/大模型为代表的机器学习方法与技术为当今科技领域的研究焦点。然而，相比技术水平的迅猛提升，机器学习基础理论研究进展远远滞后，大量技术经验发现无法找到理论支撑，以科学性为前提的学科大厦面临危局。重建机器学习理论体系已成为当今重大科技前沿问题。针对这一挑战，本报告将以深度学习三大技术实验现象（任务泛化能力、智能涌现现象、鲁棒-精确悖论）理论内涵作为分析对象，介绍未来机器学习理论可能发展的新型学习、统计与物理理论新框架，以及本研究组分别针对通信、生命、医疗技术领域的场景动态适应、生命智能涌现、智能可靠诊疗三大技术问题的机器学习数学技术方面的初步探索。

嘉宾介绍



叶齐祥 中国科学院大学

叶齐祥，中国科学院大学特聘教授、国家杰出青年基金获得者、中国科学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、ICLR2024 Area Chairs、国际期刊 IEEE TITS, IEEE TCSVT 编委。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 150 余篇。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。曾获中国电子学会自然科学一等奖。培养多名博士生获中国科学院院长奖、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。

报告题目

视觉表征模型的结构设计与物理启发

报告摘要：本报告分析了局部卷积运算与全局注意力运算的互补性，将局部特征与全局特征耦合形成 Conformer 网络结构，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，iTPN-Base, iTPN-Large, iTPN-Huge 分别达到了 88.0%, 89.2%, 89.7% 的 Top-1 分类精度。在模型结构设计的基础上，探索了物理原理启发的基础模型，从热传导视角探索表征模型，取得了初步效果，代码：github.com/pengzhiliang/Conformer、github.com/sunsmarterjie/iTPN、github.com/MzeroMiko/VMamba。

嘉宾介绍



王利民 南京大学

王利民，南京大学教授，博士生导师，国家海外高层次青年人才计划入选者，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 年在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、NeurIPS 等学术期刊和会议发表论文 100 余篇。根据 Google Scholar 统计，论文被引用 23000 余次，两篇一作论文取得了单篇引用接近或超过 4000 的学术影响力。在视频分析领域做出了一系列有重要影响力的研究工作，例如：TSN 网络架构，VideoMAE 预训练方法，MixFormer 跟踪器等。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，ACM MM 2023 最佳论文荣誉提名奖。入选 2022 年度 AI 2000 人工智能全球最具影响力学者榜单，2022 年度全球华人 AI 青年学者榜单，2021-2023 年度爱思唯尔中国高被引学者榜单。担任 CVPR/ICCV/NeurIPS 等重要国际会议的领域主席和计算机视觉领域旗舰期刊 IJCV 的编委。

报告题目

多模态视频理解基础模型 InternVideo

报告摘要：构建多模态基础模型已经成为计算机视觉领域的研究热点。视频理解面临着数据维度高、信息容量大、场景变化多等核心挑战，如何构建通用视频理解基础模型已经成为现阶段一项极具挑战的任务。本次报告将主要介绍多模态视频理解基础模型 InternVideo 及其背后的关键技术，包括单模态视频自监督预训练方法 VideoMAE，多模态视频弱监督预训练方法 UMT，和多模态视频交互对话模型 VideoChat。同时还将介绍多模态视频数据集 InternVid 和多模态视频评测基准 MVBench。最后将展望多模态视频基础模型发展趋势。

嘉宾介绍



张长青 天津大学

张长青，天津大学智能与计算学部教授/博士生导师、人工智能学院副院长，国家“万人计划”青年拔尖人才，其主要研究方向为机器学习/计算机视觉/智能医疗。2017-2018 年在北卡罗莱纳大学教堂山分校 (UNC) 担任研究员。在 Nature Communications/TPAMI/ICML 等顶级期刊和国际会议上发表论文 50 余篇，多篇论文入选 ICML/CVPR/NeurIPS 口头报告或亮点论文。研究成果获得中国图象图形学学会自然科学奖一等奖、ICME 最佳论文等奖励，入选百度发布的全球高潜力 AI 华人青年学者榜单、斯坦福大学发布的全

球 Top 2% 顶尖科学家榜单、爱思唯尔“中国高被引学者”。受邀为 IEEE TPAMI、CVPR、ICCV、NeurIPS、ICML、ICLR 等国际期刊及会议审稿。

报告题目

Multimodal Fusion on Low-quality Data

报告摘要：多模态信息融合在科学发现、医疗诊断、机器人等领域具有广泛和重要应用。不同模态数据蕴含的信息具有互补性、冗余性、动态性、不平衡、不完整等诸多复杂关联和不确定性，其对多模态数据融合的效果产生了深刻的影响。本报告将从统一的视角对低质量多模态数据进行刻画，并介绍面向低质量多模态数据的融合理论、方法和应用。

► 专题论坛 ◀

专题论坛 19：多模态感知与行为协同的具身智能

时间 2024 年 7 月 26 日 下午 13:30-17:00

地点 多功能厅 1

组织人

- 宋新航（中国科学院计算技术研究所 副研究员）
- 蒋树强（中国科学院计算技术研究所 研究员）

论坛介绍

具身智能的目标是实现智能系统在真实世界的感知、决策和行为交互，从而能够更好地适应复杂、动态、开放的环境，以及和人类和其他具身智能体进行有效的沟通和协作。这种全面的、多模态的智能系统，更贴近人类的认知和行为模式，代表着人工智能未来发展的一个重要方向。然而，该领域面临的挑战包括多模态数据的稀缺性和多样性、模型的复杂性和可扩展性、验证环境的规范性和统一性，以及虚拟到现实的泛化问题等。在这个背景下，本论坛旨在总结当前具身智能领域的前沿进展，并通过主题演讲和圆桌讨论的形式深入分析和探讨当前现状、研究内涵和未来路线。我们期望通过这个论坛，能够总结更多共识、提供更多启迪、推动领域发展。同时，多媒体内容在具身智能的研究和应用中发挥重要作用，如图像、视频、音频等提供丰富的感知信息，为具身智能提供数据基础。通过分析和理解多媒体内容，具身智能可以更好地理解和适应环境，完成复杂任务。因此，多媒体内容的分析和利用，也是具身智能研究的重要组成部分，我们也期待在论坛上，与专家和学者共同探讨多媒体内容与技术具身智能中的应用和挑战，推动具身智能的进一步发展。

报告嘉宾

嘉宾介绍



郑 锋 南方科技大学

郑锋，鹏城研究员，南方科技大学副教授，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。研究方向为视觉语言大模型，多模态融合具身智能。在相关领域发表 CCF 推荐 A 类论文 90 余篇。指导学生在包括 CVPR ActivityNet, LOVEU, MeViS, ACM MM PIC 等多个国际重要竞赛中多次获得冠军。开源项目中，提出的 Pyramid 再识别算法收录 OpenCV，推出的 Track Anything 获得 GitHub 6000+ Stars，热度周/月排名分别进入前二。是 IET IPR 杂志副编辑，多次担任 ACM MM, NeurIPS 领域主席，ICME/IJCB 本地主席。主持包括科技部重点研发课题等相关科研项目 10 余项。研究的相关技术成果成功向中兴通讯、中国科学院国家空间中心，华为，腾讯数平、腾讯优图等机构实现应用转化。

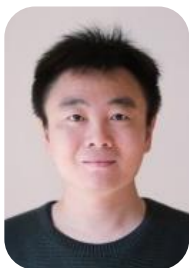
报告题目

具身感知场景中的反思学习

报告摘要：在人工智能的快速演进中，语言和多模态大模型通过数据与模型规模的扩展，实现了知识涌现的突破。尽管如此，具身智能领域，作为通向通用人工智能的关键，仍面临空间信息数据的不足。基于此，我们依托鹏城实验室，首次确立了具身数据标准，并成立了具身数据联盟，赢得了学术界和工业界的广泛关注。针对

具身感知中基于多模态大模型的感知技术所遭遇的幻觉问题，我们提出了一种新颖的幻觉抑制技术，该技术采用反思学习机制，显著降低了模型幻觉，优化了性能。

嘉宾介绍



范鹤鹤 浙江大学

范鹤鹤，浙江大学计算机科学与技术学院百人计划研究员，获国家级青年人才项目。主要研究方向包括多模态大模型、生成式人工智能、三维视觉感知、和微观结构理解与生成。曾担任 ACM MM、IEEE ICIP 领域主席。曾在 TPAMI、ICLR、CVPR、ICCV 等期刊和会议上发表论文 40 余篇。

报告题目

数据知识双轮驱动的具身智能

报告摘要：具身智能技术旨在赋予机器人或虚拟智能体在物理或虚拟环境中进行感知、行动和决策的能力。最近，在具身智能领域，多模态大模型正在引领智能感知与决策的革新，并深刻影响多智能体系统的决策过程。多模态大模型能够处理并解析来自多种感官的数据，从而实现更加精准的环境理解、推理和决策。不仅显著增强了机器的环境感知能力，同时也大幅提升了决策的质量与多智能体间的协调效率。在多智能体系统中，多模态大模型能够帮助各个智能体学习如何更好地与其他智能体协作或进行竞争。通过理解并预测其他智能体的行为，这些模型能够优化整体群体策略，以高效实现各项任务目标。本报告将从多个角度介绍多模态大模型在智能感知与决策中的应用，展示其在现代智能系统中的关键作用和潜力。

嘉宾介绍



蔡盼盼 上海交通大学

蔡盼盼，上海交通大学副教授，博士生导师，2023 年获海外高层次青年引进人才。2022 年 8 月加入上海交通大学清源研究院，任长聘教轨副教授，研究领域为人工智能、机器人策略规划、机器人学习、自动驾驶，在机器人期刊 IJRR、T-RO、工业自动化期刊 TII、机器人会议 RSS 等发表大量高水平论文。担任机器人期刊 IEEE T-RO 编委，曾担任机器人会议 (ICRA、ISER) 副编辑，RSS、CoRL、ICAPS 等机器人与人工智能会议的程序委员会成员，以及 IJRR、TRO、RAL、AURO、RSS、ICRA、IROS、ACC、IJCAI 等机器人、人工智能期刊与会议的审稿人。曾在机器人会议 RSS 主办融合规划与学习研讨会。代表工作曾被纳入柏林工业大学研究生课程。

报告题目

不确定性下的人机交互策略规划

报告摘要：机器人与现实世界的交互充满了不确定性。机器人往往需要在这些不确定性下进行实时决策规划。例如，自动驾驶汽车在城市环境中必须与大量行人、车辆实时交互，并与他们紧密协作以实现安全、舒适的驾驶并快速到达目的地。在拥挤的交通中，人类往往表现出复杂的、不规范的交通行为，造成充满不确定性的动态交互环境和高度困难的机器人决策问题。然而，不确定性下的最优决策规划具有极高的计算成本，在实时约束下很难实现。

本次分享将围绕机器人在高不确定性动态环境中的实时决策问题，从问题建模、实时规划、融合规划与学习三

方面展开讨论。演讲者将首先介绍为混乱场景自动驾驶构建的部分可观察马尔科夫决策过程 (POMDP) ; 随后, 探讨如何利用大规模并行化加速 POMDP 规划算法, 实现大规模不确定性下的实时决策; 最后, 讨论规划算法与机器人学习 (如强化学习、自监督学习等) 的融合。通过为实时规划学习搜索启发、抽象行为空间、采样分布等子模块, 攻克决策规划的高计算复杂度问题和学习的数据饥渴问题。

嘉宾介绍



弋 力 清华大学

弋力博士, 现任清华大学交叉信息研究院助理教授, 国家优青 (海外)。他在斯坦福大学取得博士学位, 导师为 Leonidas J. Guibas 教授, 毕业后在谷歌研究院任研究科学家。在此之前, 他在清华大学电子工程系取得了学士学位。他近期的研究兴趣涵盖三维视觉和具身人工智能, 他的研究目标是使智能机器人具备理解三维世界并与之互动的能力。他在计算机视觉、计算机图形学以及机器学习领域的顶级会议发表论文六十余篇, 并担任 CVPR 2022-2024、IJCAI 2023、NeurIPS 2023 领域主席。他的工作在领域内得到广泛关注, 引用数 20000+, 代表作品包括 ShapeNet Part, 光谱图 CNN, PointNet++ 等。

报告题目

多模态大模型驱动的开放世界具身感知与交互

报告摘要: 当前的具身智能研究大多聚焦于在特定环境下针对特定技能的学习, 距离实现在开放环境中解决各种任务的通用智能目标还有较大差距。近年来, 多模态大模型的快速发展使得在开放环境中的感知理解任务表现出色。因此, 如何利用和发展多模态大模型以驱动具身通用智能的发展, 成为了一个值得探索的自然想法。本次报告将从两个角度展开讨论: 一是如何从现有的多模态大模型中继承知识, 以驱动通用泛化的感知与交互; 二是如何发展更加适配于具身智能体的多模态大模型。一方面, 我们可以利用现有多模态大模型关于海量知识的表征, 以规划智能体更加多样化的交互行为, 大幅拓展其交互技能。另一方面, 我们从想象力和交互力的角度出发, 打造面向具身规划的多模态大模型。我们相信具身大模型会为通用具身智能的发展提供强有力的范式。

嘉宾介绍



王 鹤 北京大学

王鹤博士是北京大学前沿计算研究中心 (CFCS) 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 (EPIC Lab), 致力于通过发展具身技能及具身多模态大模型推进通用具身智能。他同时担任北大-银河通用具身智能联合实验室主任和北京智源人工智能研究院具身智能研究中心主任。王鹤博士在计算机视觉、机器人学和人工智能领域的顶级会议和期刊上发表了五十余篇论文, 并获得包括 ICCV、ICRA、Eurographics 在内的多次最佳论文提名和候选。他担任 CVPR 和 WACV 的领域主席, 同时担任多项顶会的审稿人和程序委员。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。

报告题目

面向通用机器人的具身多模态大模型系统

报告摘要：通用机器人是具有任务和环境通用属性和革命性意义的下一代机器人。本报告将讨论通过具身多模态大模型系统融合视觉语言等模态可以直接对高度泛化通用的物理任务输出动作的解决方案。报告将介绍由本体和数据构成基石层、大脑和小脑构成能力层的具身多模态大模型系统方案。对于本体和数据，我们将对人形机器人的形态进行探讨，提供硬件的发展思路和基于合成大数据的泛化训练数据获取途径。对于能力层，我们完全通过合成数据和 Sim2Real 实现了多个泛化的移动和操作技能，包括二指和灵巧抓取、铰接类物体操作、柔性物体操作、端到端视觉语言导航大模型等等，这些构成了小脑。而对于大脑，我们将展示 GPT-4V 为代表的非具身多模态大模型进行视觉感知、任务规划和调用中层的三维视觉技能，实现从家用电器泛化操作到开放指令物体摆放的能力。最后，报告将展望通用机器人的未来，讨论其中的机会和挑战。

► 专题论坛 ◀

专题论坛 20: AI4Food: 食品和健康计算

时间 2024 年 7 月 26 日 下午 13:30-17:00

地点 多功能厅 2

组织人

- 闵巍庆 (中国科学院计算技术研究所 副研究员)
- 陈静静 (复旦大学 副教授)
- 石吉勇 (江苏大学 特聘教授)

圆桌讨论主席

- 明朝燕 (浙大城市学院 副教授)
- 卢 剑 (宁夏医科大学 教授)

论坛介绍

面向膳食领域的多媒体内容分析自 2000 年在多媒体领域开始兴起, 由于食物对人类生存、生活与健康的基础性作用和关键性影响使得该领域近来受到越来越广泛的关注。人类生活行为记录的数字化使得饮食图像和菜谱文本等膳食多媒体数据爆发增长, 加之包括 ChatGPT 等各种大模型在内的人工智能技术的迅速崛起, 为该传统领域的进一步发展带来新的机遇, 并逐渐催生了食品计算这一新兴领域。由于膳食仅次于遗传, 是影响人类健康的第二大因素, 因而面向膳食领域的食品和健康计算得到了尤为广泛的关注, 其主要采用计算方法对膳食图像、菜谱文本和各种膳食属性 (如烹饪方式、口味、营养素等) 等膳食多媒体数据进行分析处理, 推进膳食领域的智能化和精准化转型升级, 并满足人们日益增长的膳食分析和健康管理等各种需求, 最终辅助解决膳食相关的健康问题。为此, 本论坛将围绕膳食领域的食品和健康计算, 从食品图像的多粒度分析、基于高光谱视觉的食品品质检测、多模态膳食营养摄入分析、膳食领域的基础模型以及在食品工业和临床营养等方面的应用介绍最新研究成果, 以期促进该领域的发展, 丰富和扩展多媒体分析的新方法、新应用和新服务。

报告嘉宾

嘉宾介绍



崔晓晖 江南大学

崔晓晖, 武汉大学国家网络安全学院二级教授, 江南大学特聘教授, 博士生导师。美国路易斯维尔大学讲习教授、食品安全区块链联盟负责人、国家“食品安全大数据关键技术”重点研发计划首席科学家、教育部软件工程教学指导委员会委员、国家留学基金委项目评审委员会委员。长期从事大数据、集群智能、区块链、食品计算、社会计算研究。主持了国家重点研发计划、国家自然科学基金、军委装发重点基金等 10 余项研发项目。研究成果被美国 MSNBC 新闻、美国时代自由杂志、英国新科学家杂志以及人民网、中国青年报、央广网等国内外媒体报道。

报告题目

基于图像分析的奶粉品质在线检测技术研究

报告摘要: 大数据技术的日趋成熟和机器学习方法的日渐丰富, 为建立更精准、高效的奶粉品质软测量模型提

供了思路。奶粉品质测量是奶粉重要的冲调性指标之一，主要包括分散性、润湿性、溶解性和下沉性等，是消费者评价奶粉品质优劣的直观依据。现有的测量方法多为周期长、效率低、损耗高的实验室方法。针对这些问题，讲者提出结合了机器学习和大数据技术的奶粉品质软测量模型旨在使用通过小样本数据扩增技术（扩散模型及生成对抗网络）扩充的奶粉颗粒微观形状因子数据集实现实时、高效、精准的奶粉品质在线软测量。

嘉宾介绍



邹小波 江苏大学

邹小波，博士生导师，教授，现任江苏大学副校长。邹小波教授长期从事食品、农产品质量与安全快速无损检测技术与智能装备研究。主要包括计算机视觉检测技术与装备、电子嗅觉检测技术与装备、近红外光谱检测技术与装备；光谱成像检测技术与装备；生物传感器检测技术与装备；食品品质快速无损与在线检测装备；中央厨房成套装备等。先后主持国家重点研发计划、国家科技支撑计划、国家自然科学基金等项目 10 余项。近年来在食品品质的多维仿生感知、特色食品加工过程参量的多维分布成像化检测等方面取得了创新性成果。出版专著 6 部，其中英文 4 部。发表 SCI/EI 论文 300 余篇，授权专利 50 余件。研究成果曾获全国百篇优秀博士论文、国家技术发明二等奖、中国青年科技奖、教育部育技术发明一等奖等奖励和荣誉。

报告题目

农产品无损检测与智能计算技术

报告摘要：以“加工工艺变化”应对“加工过程变化”的“以变应变”机制是实现农产品智能加工的核心。与生理生化反应弱、组分变化慢的工业产品相比，农产品为自身生理生化反应强、反应速率/位点不均一、组分状态变化复杂的非稳态生物体；现有检测方法对生产线上农产品加工过程变化的感知效率不高，易导致“以变应变”的农产品智能化计算机制失效。本报告将从无损检测的主要内涵和关键特征出发，介绍嗅觉可视化等农产品无损检测新技术，以及农产品原料属性评价、加工过程参量感知、成品品质分级过程中的无损检测及智能计算应用案例。

嘉宾介绍



蒋树强 中国科学院计算技术研究所

蒋树强，中国科学院计算技术研究所研究员，中国科学院智能信息处理重点实验室副主任，国家杰出青年科学基金获得者；先后主持承担科技创新 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，任 IEEE 与 CCF 高级会员，国际期刊 ACM TOMM 编委，CCF 多媒体专委会秘书长、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、食品计算，在 IEEE/ACM Transactions 和 CCF-A 类会议上发表论文 60 余篇，研发成果应用或转移到多个企事业单位；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。

报告题目

面向营养健康的食品计算和多模态大模型

报告摘要：当前，人工智能（AI）技术的快速发展得到各行各业的高度关注，特别是近期大模型技术的出现正在深刻影响人工智能技术的发展与演进，如何利用行业特点、深入行业痛点、聚焦行业需求开展 AI 技术研究是人工智能技术赋能千行百业的重要方面。食品关乎人类生存发展、关乎人民生命健康、关乎个人日常生活，开展食品领域的人工智能技术研究意义重大、需求广泛。食品计算旨在用计算的方法对食品数据进行分析 and 建模，以期对食品科学、营养科学和农学等食品相关领域的研究和应用提供智能化与数字化支持。本报告面向食品营养健康，从食品识别、推荐、食品大数据挖掘等几个方面介绍相关研究与应用成果，并介绍团队在食品多模态大模型方面的最新研究进展与系统演示，并对未来食品计算的发展进行了展望。

嘉宾介绍



张 烽 江南大学

张烽，江南大学生物工程学院、医学院、环境与生态学院教授，博士生导师，上海交通大学生命科学技术学院博士，江南大学附属医院营养科（功能食品临床评价中心）副主任、GCP 机构办公室副主任。中国抗癌协会肿瘤营养专业委员会副秘书长和常务委员、中国营养学会营养与保健食品分会五届委员会委员暨特殊医学用途配方食品应用专委会委员、中国食品科学技术学会医学食品分会委员、江苏省中医药学会慢病管理专委会常务委员、江苏省免疫学会转化医学专业委员会委员等学会职务。自 2012 年起，从事肠道微生物代谢过程及与疾病的关系研究，通过肠道菌群鸟枪法宏基因组测序结合代谢组、转录组等多组学技术，探讨肠道菌群代谢物参与疾病发生发展的机制。相继在 Science、Nature Commun、Microbiol Mol Biol Rev、Crit Rev Food Sci Nutr 等高水平杂志发表论文 30 余篇，累计影响因子 242。主持国家自然科学基金面上项目 2 项、“十四五”重点研发计划子课题、江苏省自然科学基金面上项目，江苏省卫健委面上项目等国家级及省部级项目各 1 项，市级课题及横向课题 20 余项。获江苏省新技术引进二等奖和江苏省中医药科技奖二等奖各一项。参编著作 3 本，1 项专利获授权。曾获江苏政府留学奖学金资助，2016 年在加拿大 ISME 年会上作口头报告，2019 年在哈佛医学院波士顿儿童医院作学术报告。“太湖人才计划”高端人才和顶尖医学专家团队成员。担任《Frontiers in Immunology》编委、《Molecular Metabolism》、《中国慢性病预防与控制杂志》、《上海交通大学学报》审稿专家。

报告题目

基于组学数据的精准营养配方设计

报告摘要：针对特定疾病不同阶段患者营养缺陷和代谢特征不明，疾病特定全营养食品配方设计不精准等问题，本研究采用基础全营养配方食品对患者进行干预。借助肠道菌群宏基因组、血液代谢组和血液全谱氨基酸分析等手段，分析干预前后多组学数据与正常对照组之间的差距，采用深度学习的方法计算出更适合的配方，并在新的人群中进行验证，最终获得针对特定疾病特定阶段的精准全营养配方食品。

嘉宾介绍



魏忠钰 复旦大学

魏忠钰，复旦大学大数据学院副教授，智能复杂体系实验室双聘研究员，博士生导师，复旦大学数据智能与社会计算实验室（Fudan DISC）负责人，复旦大学自然语言处理（Fudan NLP）团队成员。于哈尔滨工业大学获得学士和硕士学位，香港中文大学博士，美国德州大学达拉斯分校博士后。担任多个重要国际会议包括 EMNLP、ACL 的高级领域主席和组织委员会成员。主要研究领域包括大模型技术和社会计算，获得 2019 年度 CIPS 社交媒体处理专委会新锐奖，2021 年上海市启明星计划，2022 年 CCF 自然语言处理专委会新锐学者奖。

报告题目

结合垂域大模型和多智能体框架的营养健康场景赋能

报告摘要：近来大语言模型（LLM）展示出强大的自然语言交互能力，为营养健康助手走进人们的生活带来了希望。健康咨询场景通常较为复杂，个人助手需要有丰富的营养健康知识，具备通过多个轮次对话了解用户意图，并给出专业、详实回复的能力。在交互能力之外，复杂的营养健康场景往往还需要用户建模、文档分析、健康产品推荐等多个功能支持。针对这一问题，讲者所在团队提出了将垂域大模型和多智能体框架相结合的方法赋能营养健康场景的实际需求。

► 专题论坛 ◀

专题论坛 21：三维生成大模型

时间 2024 年 7 月 26 日 下午 13:30-17:00

地点 多功能厅 3

组织人

• 白 静 (北方民族大学 教授)

• 聂为之 (天津大学 教授)

论坛介绍

三维生成技术和大型模型 (如 GPT 和 DALL-E) 的快速发展引起了学术界和工业界的广泛关注。如今, 三维内容生成正逐渐向人工智能和深度学习转变, 为视频游戏、电影、虚拟现实等领域提供强大支持, 并拓展至机器人视觉、自动驾驶等应用, 开辟了计算机视觉、图形学、人工智能等新研究方向。尽管该领域备受关注, 仍面临诸多挑战, 如算法效率、数据需求、生成质量控制、伦理和隐私等问题。为此, 我们拟举办主题为“三维生成大模型”的论坛, 邀请 3D 生成大模型领域的优秀青年学者嘉宾做学术报告介绍相关方向的前沿知识以及嘉宾的科研成果, 且与参会学者共同讨论该领域的热点问题以及未来发展趋势。学术报告和讨论所涉及的具体研究方向包括基于神经辐射场与 3D 高斯的三维生成与重建、基于大模型的多物体三维场景生成、基于流模型的三维点云高质量生成、基于超图计算的三维模型数据分析等。

报告嘉宾

嘉宾介绍



易 冉 上海交通大学

易冉, 博士, 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 60 余篇论文于 IEEE TPAMI、ACM TOG、SIGGRAPH、CVPR 等期刊会议, 其中 CCF-A 类论文 40 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, 2022 年中国图象图形学学会高等教育教学成果一等奖, 2022 年 CCF-腾讯犀牛鸟基金卓越项目, 2021 年中国图象图形学学会石青云女科学家奖(青英组), 2022 年中国人工智能学会优秀博士学位论文, 2022 年瑞士 Chorafas 青年研究奖, 2023 年 ACM SIGAI China 新星奖, 2019 年中国计算机学会计算机视觉专委会学术新锐奖。TPAMI、IJCV、TIP、CVPR、ICCV、NeurIPS、ICLR、AAAI 等国际期刊会议审稿人。

报告题目

基于神经辐射场与 3D 高斯的三维生成与重建

报告摘要: 随着 3D AIGC 技术的发展, 高效可控的三维内容智能生成成为领域内亟待解决的重要研究问题。本次报告中将围绕神经辐射场与 3D 高斯驱动的三维内容生成, 介绍基于单张图像高保真三维模型生成、实时神经辐射场加速、稀疏视角三维重建等方面的研究工作。

嘉宾介绍



高 跃 清华大学

高跃，清华大学长聘副教授、博士生导师，国家青年特聘专家、DeepTech 2022 年中国智能计算科技创新人物。主要研究领域为人工智能、计算机视觉及医学图像处理，提出并发展了超图计算理论及方法，形成了面向有效数据不足、信息关联复杂等环境的新方法。研究成果发表论文百余篇，论文引用万余次，出版《Hypergraph Computation》等英文专著，多次入选科睿唯安高被引科学家和爱思唯尔中国高被引学者，担任 International Journal of Computer Vision 和 Medical Image Analysis 等国际重要期刊编委。曾获得中国电子学会自然科学一等奖、福建省科技进步一等奖等。

报告题目

基于超图计算的 3D 内容理解

报告摘要：许多生物、社会和技术系统的复杂性源于系统单元之间丰富性的相互作用。社会交流、化学反应、生命系统到生态系统中相互作用通常发生在三个或更多对象的组中，并不能简单地用二元关系来描述。有效地利用这些高阶关联结构能够增强对各类系统的建模能力，帮助理解和预测其动态行为。超图是建模此类高阶交互、即高阶关联的天然结构。实际应用中通常也面临数据不足和关联复杂难题。本报告主要介绍面向高阶关联的超图计算理论、方法及应用。具体而言，从由数据到高阶关联的映射、由高阶关联到语义表示的映射及高阶关联语义领域泛化三个维度介绍高阶关联的超图结构建模、数据和高阶关联协同的超图语义计算。报告最后介绍基于超图计算的 3D 内容理解方法与应用。

嘉宾介绍



李曼玮 山东大学

李曼玮，山东大学，软件学院，副研究员。2013-2018 年在山东大学计算机学院交叉研究中心攻读博士学位，在此期间曾访问香港大学、中国科学技术大学、以色列特拉维夫大学并进行项目合作，并于 2017-2018 年在加拿大西蒙弗雷泽大学进行联合培养。2019-2021 年继续在加拿大西蒙弗雷泽大学 GrUVi 实验室进行博士后研究，同时在 Intelligent Project Solutions Inc. 兼职担任 AI 科学家。2022 年加入山东大学软件学院人机交互与虚拟现实研究中心担任副研究员。研究领域涉及计算机图形学、三维视觉、人工智能等，主要关注基于数据驱动的三维模型理解与内容生成，如三维模型的语义理解和表征学习、室内场景的设计与建模、基于图像或点云的三维重建与生成等。目前已在计算机图形学、计算机视觉等领域 CCF-A 类及权威期刊和会议发表论文多篇，参与项目包括加拿大 NRC-IRAP、Mitacs Acceleration、国家自然科学基金面上项目等，担任 Siggraph、TOG、TVCG、CVPR、ECCV 等国际顶级会议或期刊审稿人。获得 2022 年 ACM 济南分会新星奖，以及山东大学未来计划、山东省优秀青年科学基金（海外）、国家自然科学基金青年项目等资助。

报告题目

基于大模型的多物体三维场景生成

报告摘要：近年来，随着预训练大模型的快速发展，其在三维内容生成领域展现出巨大潜力，能够快速生成各种形象的三维模型。基于预训练大模型的三维场景生成技术能够突破固定类别数据集的限制，自动从文本描述

和图像条件中生成大规模和高质量的三维场景内容，具有自动化程度高、扩展性强、交互简单等优点。本次报告将介绍基于预训练大模型的三维场景生成领域的最新方法，重点关注具有大量物体和复杂空间关系的三维室内场景生成方向的问题与挑战，包括基于文本描述的场景生成、基于图像属性迁移的场景编辑等。

嘉宾介绍



毛爱华 华南理工大学

毛爱华，教授，博士生导师，华南理工大学，广东省高等学校优秀青年教师，广东高校优秀青年创新人才入选者，获 2015-2016 年于美国 The Ohio State University 计算机科学与工程学院进行访问研究，在人工智能、机器学习、图形图像、智能计算等方面进行了系统深入的研究探索，同时取得一系列的研究产出。近年来已发表三大检索学术论文 60 余篇，其中发表在 SIGGRAPH Asia、IEEE TVCG、CAD、AMM 等高水平 SCI 期刊论文 30 余篇，授权/申请发明专利 15 项，并获得学术奖励 6 项。前期研究成果已成功应用于 2012 年伦敦奥运会和 2008 年北京奥运会香港代表队智能比赛服研发。主持完成多项国家自然科学基金项目、广东省科技计划项目、广州市科技计划项目，授权/申请发明专利 10 项，并获得国际、国内学术奖励 5 项。主持 2019 年教育部首批国家级本科线下一流课程，获得 2019 年全国高校创新教学大赛一等奖，2013 首届全国高校微课教学比赛广东省赛区一等奖、2013 学年华南理工大学青年教师本科课堂教学竞赛一等奖，获 2017 年广东省教学成果奖二等奖，2021、2019 和 2017 年华南理工大学教学成果奖二等奖（排名第一），华南理工大学本科教学优秀教师“南光奖”（教学最高奖），华南理工大学“三育人”之“教书育人”先进个人，2021 年和 2011 年华南理工大学优秀共产党员。

报告题目

基于流模型的三维点云高质量生成

报告摘要：近年来，随着深度数据传感器的发展，点云已被广泛用于各种应用中，例如文化遗产重建，地球物理信息系统，自动驾驶，实时定位与地图构建和虚拟/增强现实。但由于硬件的限制及现实世界光照和材料的复杂性，LiDAR 等传感器获得的点云通常是稀疏的和不规则的，这会进一步影响到下游任务的性能。因此，将稀疏和不规则的点云上采样为稠密和均匀的数据，在计算机图形学领域中引起了广泛关注。本研究针对点云上采样问题构建了一个基于可逆神经网络的方法，通过对残差块引入 Lipschitz 约束实现了无约束的 Jacobian 架构。大量量化和可视化的实验结果表明，该方法的生成点云质量在分布均匀性、细节保真度、表面重建质量和计算效率等方面都优于目前最先进的方法。

嘉宾介绍



陈磊 清华大学

陈磊，清华大学自动化系助理研究员。长期从事人工智能、计算机视觉、视频分析等方面研究，在 CVPR、ECCV、IJCAI、AAAI、TCSVT、TMM 等期刊上会议上发表论文 10 余篇，获授权国家发明专利 3 项，主持博士后创新人才支持计划，国家自然科学基金委青年项目，担任国际期刊 JVCI 客座编委，TIP、TCSVT、TMM、CVPR、ICCV、ECCV 等期刊会议的审稿人，中国计算机学会多媒体技术专委会和中国自动化学会模式识别与机器智能专委会委员。

报告题目

视频内容生成与理解

报告摘要：视频人体动作分析与理解已成为计算机视觉领域的重要研究方向，在视频监控、人机交互等众多领域中展现出巨大的应用潜力。报告将深入探讨视频人体动作分析中的几个关键任务，包括动作的生成与预测、长程动作的分割以及基于文本的动作定位。主要包括如何通过生成实现人体动作预测、长程动作的分割方法、基于文本的动作定位研究等，最后对视频动作理解的未来发展趋势进行展望。

嘉宾介绍



赵娜 新加坡科技设计大学

赵娜，新加坡科技设计大学(SUTD)信息系统技术与设计学院助理教授，于新加坡国立大学(NUS)计算机学院(SoC)获得博士学位。她的主要研究方向为三维计算机视觉与机器学习。研究兴趣主要聚焦于增强现实、机器人及自动驾驶等实际应用中的场景理解问题，用于帮助智能设备（如机器人及智能车）在现实环境中进行深入的场景感知与理解。赵娜博士的研究成果主要发表在计算机视觉与机器学习领域的顶级会议及期刊，如 CVPR, ICCV, ECCV, AAAI, IJCV，同时担任相关顶会顶刊的审稿人。她曾于 2020 年获得 NUS SoC 研究成果奖并获得 NUS SoC 2021 年度 IMDA Excellence Prize（最佳博士论文奖）。

报告题目

面向数据高效的 3D 场景理解

报告摘要：我们生活在一个三维立体的世界中，自然而然地以三维方式感知周围环境。三维场景理解，即通过各种传感器分析和解释物理世界的过程，在许多新兴的工业应用中至关重要，如自动驾驶、增强现实和机器人感知。这一领域涉及多项任务，包括三维语义分割、三维目标检测和三维实例分割，每项任务都需要不同程度的语义细节。尽管现代深度学习方法在这些任务中取得了显著成功，但往往依赖于大规模标注数据集，这些数据集获取成本高昂且耗时。为了解决这一挑战并提高数据利用率，我们探索了三种实用的机器学习策略：1) 半监督学习，2) 小样本学习，以及 3) 自监督学习。这些策略利用现有数据减少对广泛标注的依赖。本次报告将介绍我们在这三个领域的最新进展，展示它们如何应用于各种三维场景理解任务，以在最小标注数据的情况下实现稳健的性能。

► 专题论坛 ◀

专题论坛 22：高保真数字人建模与驱动研究及其应用

时间 2024 年 7 月 26 日 下午 13:30-17:00

地点 多功能厅 4

组织人

- 吕 科 (中国科学院大学 特聘教授)
- 杨 易 (浙江大学 求是讲席教授)
- 张举勇 (中国科学技术大学 教授)

论坛介绍

随着人工智能技术的不断发展,虚拟现实、增强现实、数字孪生和元宇宙正在逐渐从概念走向现实,数字人是其中不可或缺的角色,是近年来产业聚焦的热点和研究的前沿。数字人是存在于非物理世界中多种计算机手段合成的具有多重人类特征的综合体,数字人在泛娱乐、金融、文旅、教育以及医疗等领域都有应用,论坛主题主要介绍数字人建模与驱动中的关键技术,包括基于情感感知的数字人表情生成方法,多模态协同的人体姿态估计与动作捕捉、面向多场景的数字人三维重建与驱动过程,以及数字人仿真测试与演示验证。希望基于此次论坛达到学术交流,推动数字人技术的研究与发展,体现其在突破多领域交叉融合方面具有显著的实用价值。

报告嘉宾

嘉宾介绍



刘焯斌 清华大学

刘焯斌,清华大学长聘教授,国家杰青基金获得者。研究方向为三维视觉、数字人重建、3D 生成与交互。发表 CVPR/ICCV/ECCV 论文 60 篇,其中 Oral 论文 14 篇。TPAMI/TOG 论文 18 篇。多次担任 CVPR、ICCV、ECCV 领域主席,担任 IEEE TVCG、CGF 编委,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)。

报告题目

3D 高斯数字人

报告摘要: 近期,3D 高斯点云渲染(Guassian Splatting)作为一种新的可微渲染技术横空出世。相比与 NeRF,以其更高的渲染质量、更快的渲染速度、兼容传统渲染管线等特点,在近半年来取得极高的关注。借助 Guassian Splatting 在 3D 场景表达和渲染的优势,介绍报告人在 3D 数字人生成方面的研究工作,包括高保真的头部数字化身生成、高保真的人体数字化身生成、人体数字化身的换装生成、人体人头数字化身重光照、动态三维编辑等方面的研究工作。

嘉宾介绍



林 惊 中山大学

林惊, 教授, 国家杰出青年基金获得者, Fellow of IAPR / IET, 曾任商汤科技首席研发总监/研究院执行院长。长期从事多模态人工智能、机器学习等领域的应用基础研究, 作为首席科学家项目负责人, 承担国家 2030 科技创新重大项目, 入选国家万人计划青拔人才; 曾带领商汤科技研发团队搭建大规模 A 基础设施, 开拓新兴行业。在国际顶级学术期刊和会议发表论文 300 余篇, 论文被引用近 3 万次(谷歌学术统计), 多次入选全球高被引学者榜单; 获权威期刊 Pattern Recognition 年度最佳论文奖, 多媒体计算旗舰会议 ICME 最佳论文钻石奖, 计算机视觉旗舰会议 ICCV 最佳论文奖提名; 获中国图像图形学会科学技术一等奖、吴文俊人工智能自然科学奖, 省级自然科学一等奖; 指导博士生(梁小丹、王可泽等)获得 CCF 优秀博士论文奖、ACM China 优秀博士论文奖及 CAAI 优秀博士论文奖。(曾) 担任知名期刊 IEEE Trans.Human-Machine Systems, IEEE Trans.Multimedia, IEEE Trans. Neural Networks and Learning Systems(Associate Editor), 十余次担任 IEEE CVPR、ICCV、NeurIPS、KDD、ACM Multimedia 等国际会议的领域主席。

报告题目

多模态融合趋势下数字人的机遇与挑战

报告摘要: 在更新随着生成式人工智能模型、多模态融合框架以及各种体积渲染方法的快速发展, 数字人相关的技术和应用也有了显著的进展, 也涌现出各类型基于数字人的产品和解决方案。本报告将讨论当前数字人技术的最新进展, 包括面向数字人生成及驱动的 AIGC 基础模型、真人复刻的数字人演讲视频生成、数字人多语言能力 & 情绪变换、数字人服饰生成和迁移等, 并且探讨数字人技术的发展趋势。

嘉宾介绍



刘 武 中国科学技术大学

刘武, 中国科学技术大学, 特任教授, IET fellow。在多媒体内容分析与搜索领域开展了较为深入的研究, 在重要学术国际会议和期刊上发表论文 100 余篇, 曾获得了 IEEE 多媒体领域三大国际旗舰期刊和会议 IEEE TMM 2019 最佳论文奖, IEEE MM 2018 最佳论文奖和 IEEE ICME 2016 最佳学生论文奖。入选了北京市科技新星计划, 《麻省理工科技评论》亚太区“35 岁以下科技创新 35 人”, 获得了天津市科技进步特等奖、ACM 中国新星奖、中国科学院优秀博士论文奖、中国多媒体企业创新技术奖等奖项, 担任了 IEEE T-MM Associate Editor, IEEE ICME 2022 和 ACM MM Asia 2021 技术委员会主席, 以及 ACM MM、AAAI、ACL、CIKM 等领域主席, 作为主席组织过 14 个国际 Workshop、Tutorial、Special Session 和竞赛。

报告题目

单目多人人体 3D 模型重建关键技术

报告摘要: 随着数字人和 AIGC 浪潮席卷全球, 人体 3D 模型重建、渲染和动作生成技术也成为了学术界和工业界关注的焦点。与已有依赖复杂设备和模型的工作不同, 本报告将介绍如何借助单个相机, 高效率、高鲁棒性地重建多人拥挤情况下的人体三维模型, 精确捕捉人体动作信息, 快速生成任意姿态的纹理外观信息, 以及如何基于文本大模型快速生成新的人体动作序列, 从而为数字人提供更多样化的动作内容, 并展示如何使用以

上技术来服务具体的应用。

嘉宾介绍



朱霖潮 浙江大学

朱霖潮，浙江大学计算机科学与技术学院百人计划研究员、博士生导师，入选国家级青年人才项目，获首届谷歌学术研究奖（2021 年）等荣誉。曾在澳大利亚悉尼科技大学担任讲师（助理教授）。主要研究方向为人工智能、跨媒体智能及其应用、人工智能通用基础模型等。曾获得 CVPR MABe 多智能体行为建模竞赛冠军（2022）等 8 项国际竞赛冠军。曾担任 ECCV、ICIP、PRCV 等会议的领域主席，并多次在 CVPR、ICME 等会议上组织专题研讨会。

报告题目

知识驱动的数字人重建和生成技术

报告摘要：本报告首先介绍数字人应用中遇到的诸多挑战与问题，并对知识驱动策略的优势进行深入分析。本报告将进一步探讨了在数字人重建和多媒体生成等领域的研究进展。在数字人重建领域，本报告将讨论几何与其他类型先验知识整合应用的技术。通过分析具体案例，本报告讨论专用知识嵌入和结构化表达技术在数字人领域的应用。最后，本报告将讨论结合数据和知识驱动的研究发展趋势。

嘉宾介绍



黄石生 北京师范大学

黄石生，北京师范大学副教授，硕士生导师，曾于清华大学计算机系进行博士学习和博士后研究工作，2021 年博士后出站进入北京师范大学人工智能学院工作至今。主要围绕计算机图形学中智能三维建模、动态视图合成、虚拟数字人建模等领域从事教学与科研工作，先后在 ACM SIGGRAPH、ACM TOG、IEEE TVCG、IEEE CVPR、IEEE ICRA 等国际会议或期刊正发表论文约 30 篇，主持国家自然科学基金青年基金、中国博士后科学基金面上项目等多项，参与科技部 973 项目、重点研究计划、国家自然科学基金重点研究项目等，获 2023 年度 CCF-百度松果基金学者等荣誉。长期担任 ACM SIGGRAPH、ACM SIGGRAPH Asia、IEEE TVCG 等期刊会议审稿人，获 CVMJ 2022 年度优秀审稿人。

报告题目

可重光照的 4D 数字人重建

报告摘要：近年来，从单目视频中重建可驱动的 4D 数字人技术已取得了长足的进步，但是如何进一步支持对 4D 数字人的自由编辑功能仍然是有待解决的问题。本报告以 4D 数字人的可重光照重建为出发点，结合实验室团队在该领域的最新研究，介绍可重光照的 4D 数字人重建工作进展。该工作将当前先进的 3D Gaussian Splatting 渲染技术，拓展到 4D 数字人重建问题上，并通过引入 4D Gaussian 层面的物理渲染技术，得到高效率、高质量的可重光照 4D 数字人建模效果，在可重光照效率和质量方面均超过已有最新的方法。此外，本报告还结合当前火热的大模型技术，就大模型支持下的可重光照 4D 数字人研究方向，提出一些新的研究思路 and 展望。

嘉宾介绍



郭玉东 中国科学技术大学

郭玉东，中国科学技术大学特任副研究员。2021 年博士毕业于中国科学技术大学，后曾于杭州像衍科技公司担任研发总监，期间研发的数字人相关产品已在杭州文广电视台和中央广播电视总台多个栏目得到应用。研究方向为三维视觉与数字人，近年来在相关领域的顶级期刊与会议上发表论文 10 余篇，曾获得华人数学家联盟最佳论文奖(若琳奖)，并多次担任 CVPR、ICCV、TVCG、SIGGRAPH ASIA、TMM 等期刊会议审稿人。

报告题目

基于扩散模型先验的三维数字人生成与编辑

报告摘要：如何高效、高质量地建模超写实三维数字人一直是数字人领域的关键技术问题。近年来，扩散模型在人像生成与编辑领域已取得突破性成果。然而，由于缺少高质量数据，目前三维数字人生成模型在渲染质量上距离人像生成模型有较大差距。本次报告将介绍我们在利用扩散模型先验，从简单输入构建可自然驱动、编辑的三维数字人方面的研究进展。

嘉宾介绍



王金宝 深圳大学

王金宝，深圳大学（大数据系统计算技术国家工程实验室）助理教授。主要研究领域涉及计算机视觉和机器学习，长期专注于数字人建模与驱动、图像异常检测、图像多样化生成，致力于将视觉技术应用于真实工业场景和虚拟现实的人机交互环境。主持国家自然科学基金项目和参与国自然重点国际（地区）合作与交流项目，获得腾讯“犀牛鸟”-深圳大学青年教师科研基金项目奖。在 IEEE TIP、IEEE TCSVT、ICCV、ACM MM、NeurIPS 和 ICLR 等重要国际期刊和会议上发表论文 30+，高引用论文 1 篇；担任国际期刊和会议的审稿人，如 IEEE TIP、IEEE TCSVT、ACM MM、NeurIPS、AAAI、IJCAI、ICML。

报告题目

数字人在真实场景的具体应用

报告摘要：随着人工智能技术的飞速发展，尤其是以生成式大模型为代表带来的技术突破，数字人在多个行业领域的应用开始获得实质性进展。这些基于大模型的数字人，不仅在交互性和智能化方面有了显著提升，也为行业应用开辟了新的可能性。本报告通过综合分析当前的应用现状、面临的挑战以及未来的发展趋势。可以看到，基于大模型的数字人正逐步成为科技进步和行业革新的重要驱动力。未来，随着技术的不断优化，数字人的应用将更为广泛和深入，极大地推动社会生产力和生活方式的变革。

► 专题论坛 ◀

专题论坛 23：多模态情感脑机接口与应用

时间 2024 年 7 月 26 日 下午 13:30-17:00

地点 多功能厅 5

组织人

• 郑文明（东南大学 教授）

• 黄晓华（南京工程学院 教授）

论坛介绍

情感智能研究的目的是通过赋予计算机识别、理解、表达和适应人的情感的能力来建立和谐人机环境，并使计算机具有更高的智能。近年来，随着脑科学、神经科学和人工智能等学科的迅速发展，使得我们对大脑功能和疾病有了更深入的理解；情感脑机接口开辟了非传统的大脑信息导出通道，实现大脑与外围设备的直接互动，可以实现更快速、精准的医疗诊断和治疗，帮助患者进行功能重塑。

本论坛旨在通过研究和探讨多模态情感脑机接口与医疗健康的交叉研究进展，将汇聚国内相关领域专家学者，为大家带来一系列在多模态情感脑机接口的前沿研究与先进技术，将通过阐述国内多模态情感脑机接口以及医疗健康的最新研究动态，对这一技术所面临的挑战性问题 and 可能的解决思路进行较为深入的分析和探讨，提高多模态情感脑机接口技术在国家和国际学术的影响力，促进产学研转化。

报告嘉宾

嘉宾介绍



陈霸东 西安交通大学

陈霸东，西安交通大学人工智能与机器人研究所教授，国家级领军人才。2008 年毕业于清华大学计算机专业获博士学位。研究领域包含机器学习、人工智能、脑机接口、机器人。在国际知名期刊及会议发表学术论文 300 余篇，论文被引用 1.3 万余次。获授权国家发明专利 20 余件，出版学术专著 6 部。入选世界排名前 2% 科学家名单和爱思唯尔中国高被引学者榜单。获教育部自然科学一等奖、中国自动化学会自然科学一等奖、中国自动化学会青年科学家奖等。担任中国认知科学学会理事、IEEE 汇刊 TNNLS/TCDS/TCSVT 编委。主持了国家自然科学基金重大研究计划重点支持项目、联合基金重点项目、973 计划课题、国家重点研发计划课题等科研项目。

报告题目

基于信息论学习的类脑计算与脑机接口

报告摘要：信息论在机器学习领域获得广泛应用并引起越来越多学者的关注。研究者针对不同的机器学习问题提出了各种信息论学习方法，如监督学习中的最小误差熵准则和表示学习中的信息瓶颈原则。本报告介绍信息论学习（Information Theoretic Learning）基本概念，阐述新的学习范式与方法，并探讨信息论学习在类脑计算与脑机接口等领域中的应用。

嘉宾介绍



吴 奇 上海交通大学

吴奇, 国家杰出青年基金获得者、上海交通大学教授。研究方向聚焦类脑计算、空中以人为中心的平台协同, 正在解决面向重大型号空中自主加油的技术攻关。任 IEEE T-NNLS, IEEE T SMCA, IEEE T-ITS, IEEE T-CDS 等 6 个 Trans 编委。

报告题目

基于脑成像技术的大脑功能激活状态研究

报告摘要: 脑功能激活状态反映了人类在面对特殊刺激, 或者完成指定任务时大脑的工作状态。研究脑功能激活状态对理解脑的工作原理、揭示脑疾病的致病机理等有着重大意义。本报告面向 fNIRS 和 EEG 两种典型的脑成像信号, 分享我们团队大脑功能激活状态的工作: 1) 针对 fNIRS 信号特征提取手段较初级、对信号深层次信息挖掘能力弱的问题, 提出了一种将 fNIRS 信号映射成二维图像的线性映射场 LMF 算法, 显著增强了特征提取深度, 为脑成像信号的特征提取提供了全新的思路; 2) 针对 EEG 信号丰富的节律特征较少地被利用、时域特征提取不够充分的问题, 提出了一种多节律并行、基于多尺度时域/空间域卷积层的时-空-频三模态脑神经网络 TSFNN 模型, 融合了 EEG 信号多种模态的多尺度特征; 3) 针对构建静态 EEG 脑图网络的局限性、多尺度空间卷积核在脑电通道较多的情况下信息丢失严重的问题, 提出了基于图的时-空-频三模态脑图神经网络 GTSFNN 模型, 通过定义不同的脑图网络模板和可学习的掩膜构建动态的脑图网络解决了上述问题。

嘉宾介绍



赵思成 清华大学

赵思成, 清华大学副研究员, 国家级青年人才, ACM/IEEE/CCF/CSIG 高级会员, 哈尔滨工业大学博士, 加州大学伯克利分校和哥伦比亚大学博士后。研究方向为情感计算、多媒体、机器学习等。发表 IEEE/ACM 汇刊、CCF 推荐 A 类论文 60 余篇, 谷歌学术引用 10000 余次。担任中国图象图形学学会 CSIG 情感计算与理解专委会秘书长、IEEE TAFCC/CSIG 通讯编委、IEEE TFS/ACM TOMM/IEEE MM 等国际期刊的首席客座编辑、NeurIPS/ACM MM/ECCV/AAAI/IJCAI/ICASSP 等国际会议的领域主席或资深程序委员。主持或参与国家自然科学基金、国家重点研发计划、CCF-滴滴盖亚学者科研基金等项目。获得 CSIG 优秀博士论文、ACM SIGMM 中国新星奖、媒体建模旗舰会议 MMM 2015 最佳论文 Runner-up 奖等奖励, 入选 AI 2000 人工智能全球最具影响力学者榜单和全球前 2% 顶尖科学家榜单。

报告题目

多媒体情感计算

报告摘要: 情感计算在人机交互、行为决策等系列智能活动中起到重要作用。该报告将围绕多媒体情感计算展开, 系统介绍课题组的相关工作。该报告首先总结多媒体情感计算的研究背景、应用场景和主要挑战, 接着从共性语义理解、个性化感知预测与标签分布学习、有限标注机器学习等三方面研究内容介绍部分代表性工作, 涵盖图像、文本、脑电、社交媒体与多模态数据, 最后讨论未来可能的研究方向。

嘉宾介绍



陈 杰 北京大学

陈杰，北京大学信息工程学院副教授，主要研究方向是计算机视觉、AI4Science，致力于表征学习及其在大模型中的应用。他在 Nature communication, TPAMI、IJCV、CVPR/ICCV/NeurIPS 等重要刊物上发表论文 200 余篇，入选“全球前 2% 顶尖科学家”榜单，最高单篇引用达到 3000 余次，总引用 10000 余次。担任国际顶级期刊 TPAMI 和 IJCV 的客座编辑，国际顶级会议的研讨会主席，例如 (ICCV, CVPR, ECCV, ACM MM 等)。累计主持或作为核心骨干参与国家发改委重大科技基础设施建设项目、科技创新 2030 — “新一代人工智能”重大项目、国家工信部 5G 项目建设、国家自然科学基金、广东重点项目和应急项目等 10 余项。获国家科技进步二等奖，入围 ACM 2022 戈登贝尔新冠特别奖（国际上高性能计算应用领域的最高学术奖项）。国家科协科技服务团专家，深圳市鹏城孔雀特聘岗位人才。

报告题目

多模态表征学习及其在大模型中的应用

报告摘要：报告首先介绍多模态的表征学习及其在计算机视觉领域的应用，例如物体分类、物体检测、目标分割和图文生成等。接着介绍了如何训练大模型，并将其用于生命科学领域，实现了对未来高风险病毒变异株的预测，该工作入围了 ACM2022 年戈登贝尔新冠特别奖，也是本次入围唯一来自中国的团队。该奖是国际高性能计算应用领域最高奖。

嘉宾介绍



蒋冬梅 鹏城实验室

蒋冬梅，鹏城实验室研究员、博士生导师，网络智能部视觉智能研究所副所长。研究方向为多模态基础模型与多模态情感计算，主持国家自然科学基金重点项目和面上项目、科技部政府间国际合作项目、陕西省重点研发项目等十余项，作为课题负责人或技术骨干参与多项国家重大重点项目和鹏城实验室重大项目。在国内外学术期刊与国际会议上发表论文 100 余篇，获陕西省自然科学优秀学术论文二等奖、陕西高等学校科学技术奖一等奖、陕西高等学校科学技术研究优秀成果奖二等奖。曾担任 ACII2015、BigMM2018 国际会议的本地组织主席，ACII2017 至 ACII2023 国际会议的出版主席，以及 ICMi2021 的宣传主席。

报告题目

多源信息融合的情感识别与抑郁症评估

报告摘要：现代社会人们生活节奏快，精神压力大，抑郁症呈高发趋势，迫切需要智能评估与筛查手段。本研究组突破了基于音视频、脑电、日常信息的情感识别与精神状态评估关键技术，构建了多模态抑郁症智能辅助评估系统。本报告将分两部分介绍近期工作：多模态情感识别与多模态精神状态评估。在多模态情感识别部分，重点介绍高泛化性可解释面部动作单元分析、面部区域与情感动态的空时注意力建模、多模态情感交互增强，以及多模态情感流 Benchmark 数据集和情感流检测；在多模态精神状态评估部分，重点介绍基于音视频、脑电或日常信息的抑郁状态评估，以及基于脑电的精神压力评估。最后，对多模态情感分析及抑郁症评估的未来工作进行展望。

► 专题论坛 ◀

专题论坛 24：面向开放环境的多媒体内容理解

时间

2024 年 7 月 26 日 下午 13:30-17:00

地点

多功能厅 6

组织人

- 叶齐祥（中国科学院大学 特聘教授）
- 杨小汕（中国科学院自动化研究所 副研究员）
- 谢国森（南京理工大学 教授）

圆桌讨论主席

- 谭明奎（华南理工大学 教授）

论坛介绍

随着移动互联网、物联网、5G 等新一代信息技术的快速发展，智能手机、智能手环等电子设备在日常生活得到了广泛普及和使用，随之引发文本、音频、图像、视频等多媒体数据的爆发式增长。可以说当今社会已经全面步入多媒体大数据时代。传统多媒体内容理解模型一般是在封闭环境下进行学习、训练和测试的；同时，所用数据集来自人工划分并且符合独立同分布假设。但在实际应用中的开放环境下，训练和测试数据往往不再满足独立同分布条件，导致已有模型性能急剧下降，难以推广应用。本次论坛主要围绕近年来开放环境多媒体内容理解领域的热点问题和前沿技术进行探讨，旨在为国内学术同行搭建平台，促进各位学者同仁之间的交流与合作。

报告嘉宾

嘉宾介绍



李长升 北京理工大学

李长升，北京理工大学计算机学院教授，博士生导师，国家优秀青年基金获得者。2013 年于中国科学院自动化所取得工学博士学位。在加入北京理工大学之前，先后在 IBM 研究院，阿里巴巴达摩院，以及电子科技大学计算机科学与工程学院工作。主要研究方向包括机器学习、数据挖掘、计算机视觉等。在 CCF A/IEEE 汇刊等国际顶级会议及期刊上发表学术论文 90 余篇。其中，以第一作者发表 CCF A/IEEE 汇刊 17 篇。先后主持国家自然科学基金优秀青年科学基金、国家重点研发计划课题等项目 10 余项；参与国家自然科学基金重点项目、173 国防基础加强重点项目、装发装备技术基础等。授权中国、美国、日本等国内外发明专利 50 余件。现担任多个国际顶级期刊和会议的审稿人、程序委员会委员、高级委员、领域主席等。

报告题目

开放环境下无人驾驶预测与决策技术研究

报告摘要：随着时代的进步和科技的快速发展，自动驾驶技术正深刻地塑造未来的交通系统和出行方式。在自

自动驾驶研究中，动态障碍物多模态轨迹预测和自车决策规划成为关键的技术挑战，直接影响了自动驾驶系统的性能、安全性和可靠性。面对复杂的多模态环境，包括静态要素（例如车道线、交通标识和信号灯）以及动态障碍物（如其他车辆、行人和自行车），自动驾驶汽车必须综合考虑各个动静态因素，考虑它们之间的相互影响，以准确预测周围动态障碍物的未来运动轨迹。通过对预测结果进行综合分析，自动驾驶系统可以做出可靠且安全的决策和规划，这对于实现高效、安全的自动驾驶系统至关重要。本报告主要介绍我们在多模态轨迹预测和自车规划的研究进展，主要包括瞬时轨迹预测、任意轨迹预测、噪声轨迹预测、规划模型压缩等。

嘉宾介绍



舒祥波 南京理工大学

舒祥波，南京理工大学计算机科学与工程学院/人工智能学院教授、院长助理、博士生导师、国家优秀青年基金获得者、江苏省杰出青年基金获得者、CCF/IEEE 高级会员。近年主要研究兴趣为视频内容分析、视频行为理解，在 TPAMI、TIP、TNNLS、CVPR、ICCV、ACM MM 等国际期刊/会议上发表学术论文近 100 篇，其中 ESI 高被引论文 7 篇；获中国电子学会自然科学一等奖、ACM MM 2015 最佳论文提名、MMM 2016 最佳学生论文奖、江苏省优秀博士论文奖、中国人工智能学会优秀博士论文奖；承担国家自然科学基金重点/面上/青年项目、国家重点研发课题、国防基础科研项目等国家级项目。担任 CSIG 青工委副秘书长，以及 IEEE TNNLS、IEEE TCSVT 等期刊编委。

报告题目

开放场景下的视频行为理解

报告摘要：视频数据在网络与物理空间持续呈现爆发式增长，人体行为理解是视频内容理解的主要研究任务，是人机交互、安防监控、虚拟现实等领域的核心技术需求之一。在许多开放场景中，由于数据规模受限、训练数据缺乏、数据易受攻击等因素，给视频行为理解带来新的挑战。基于此，本报告将探讨开放场景下的各类视频行为理解研究任务，重点介绍课题组近年利用大模型、基础模型、联邦学习等技术解决弱/半/无监督场景、多视角场景、可信场景下的行为预测与识别问题。

嘉宾介绍



李晶晶 电子科技大学

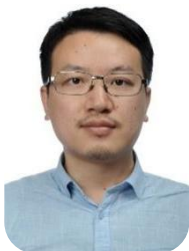
李晶晶，电子科技大学计算机学院百人计划研究员，博导，人社部“博新计划”博士后。博士学位论文获得 2018 年中国电子学会优秀博士论文奖。入选 2019 年电子科技大学“学术新人奖”，2020 年电子科技大学“人才托举计划”青年项目，2020 年电子科技大学“百人计划”。主要研究方向为机器学习，计算机视觉和多媒体，特别是欠标注场景下的机器学习。目前已在 TPAMI, TIP, TKDE, MM 和 CVPR 等 JCR 一区期刊及 CCF A 类会议上发表长文七十余篇，获得授权专利十项。担任 TPAMI, TIP, TCYB, TNNLS, TKDE, CVPR, AAAI, MM 等期刊和会议审稿人/领域主席/高级程序委员/程序委员。研究成果入选 ESI 高被引，热点，以及中国百篇最具影响国际学术论文。以第一作者获得 2023 年度吴文俊人工智能科学技术奖自然科学奖（二等）。

报告题目

领域自适应开放识别

报告摘要：领域自适应学习大幅减轻了开放识别模型数据获取和标注的成本，成倍提升了识别模型的泛化能力和适用性。为了不断攻克更复杂的问题、提升已有算法的表现，深度学习模型正变得越来越庞大复杂，其动辄数十上百亿的参数量对现有的模型自适应技术提出了极高的要求。为了让更多前沿的预训练模型能被自适应迁移应用于可穿戴、通信、医疗等计算和数据资源受限的特种环境中，高效自适应迁移学习算法成为了新的研究热点和必须解决的技术难题。本报告将分享课题组近年来在高效自适应迁移学习方面的探索和其在开放式别场景的应用。

嘉宾介绍



万 方 中国科学院大学

万方，中国科学院大学长聘教轨助理教授，博士生导师，博士后创新人才支持计划获得者、中国科学院百篇优秀博士学位论文。研究方向为计算机视觉和多模态模型，以监督信息不完备条件下的目标感知方法为主，如弱监督目标检测、主动学习目标检测等。发表包括 IEEE CVPR, ICCV, NeurIPS, AAAI, ECCV 等国际会议与 TPAMI, TNNLS, TIP, TCSVT, PR, TGRS 等国际期刊 30 余篇。获得中国电子学会自然科学奖一等奖，中国科学院院长奖，主持国家自然科学基金青年基金项目，主持华为诺亚方舟实验室科研项目，参与包括国自然重点、面上项目等项目 5 项。

报告题目

区域级多模态模型研究

报告摘要：报告针对多模态模型的区域指代任务，介绍一种名为 DynRefer 的动态分辨率方法，通过模拟人类视觉认知的分辨率适应性，实现高精度的多模态区域指代。DynRefer 首先追求充分的信息编码。它构建了一组围绕被指代区域的视图，模拟了人眼的注视和扫视。然后，DynRefer 实施动态视图选择。这使得用于指代的视觉信息能够更好地拟合人类偏好，从而提高了区域级多模态模型的表征适应性。大量实验证明，DynRefer 在多个区域级多模态任务上实现互相提升，包括区域级字幕生成、开放词汇区域识别和属性检测，在多个任务上使用一个模型实现了新的最先进水平。

嘉宾介绍



严 明 阿里巴巴通义实验室

严明，阿里巴巴通义实验室 NLP 高级算法专家，NLP 团队多模态大模型基础技术负责人。主要研究方向为多模态内容理解、AI Agent 技术和大模型等。在 ICML、ACL、EMNLP、MM、ICLR、CVPR、ICCV、PAMI 等国际顶级会议及期刊上发表论文五十余篇，在 SQuAD/GLUE/MSMARCO/VQA 等 10 多个国际赛事中取得第一，并 4 次取得超越人类基准的结果。参与打造的 AliceMind 语言模型体系入围 2021 世界人工智能大会最高奖 SAIL 奖 Top30，并获得达摩院最高荣誉奖项达摩院项目团队奖，目前负责 X-PLUG 开源体系构建。

报告题目

通义 mPLUG 多模态对话大模型技术

报告摘要：纯文本大模型方兴未艾，多模态领域也开始涌现出多模态对话大模型工作，OpenAI 的 GPT-4、GPT-4o、Google 的 Gemini 都具备强大的多模态理解能力，但是迟迟未向公众开放模型和具体实现细节，因此学术界和企业界开始在这个方向上发力研究并开源。在阿里巴巴通义实验室 mPLUG 多模态系列工作中，之前的 mPLUG、mPLUG-2 和 mPLUG-Owl2 分别被 EMNLP2022、ICML2023 和 CVPR2024 Highlight 录用，其中 mPLUG 工作在 VQA 榜单首超人类的成绩，mPLUG-Owl 作为最早的一批 GPT4 平替模型，是目前多模态大模型评测的一个常用基准 baseline。本报告主要介绍阿里巴巴通义 mPLUG 系列的多模态对话大模型 mPLUG-Owl 和 mPLUG-Owl2，通过延续 mPLUG 系列的模块化训练思想，把 LLM 升级为了多模态对话大模型。

► 专题论坛 ◀

钻石赞助专题论坛：面向视觉智算的关键技术与应用

时间 2024 年 7 月 24 日 上午 09:00-12:00

地点 贵宾厅 1

组织人

- 王苦社 (北京大学 副研究员)
- 韩巍 (北京博雅睿视科技有限公司 总经理)

论坛介绍

视觉分为人眼视觉 HV 及机器视觉 CV，计算从最初的数值计算逐渐演变为科学计算、关键计算和智慧计算，数字中国背景下，我们已迈入视觉信息占主导地位的时代，AI 计算正呈指数级增长，成为主要的计算需求，数据正逐渐成为国家基础性、战略性资源，是驱动数字经济发展的“新能源”，视觉智算主要是围绕“算据、算法、算力”进行，算法链接应用，决定发展的数字经济广度和质量，算力则决定着发展的速度和高度，算据依托于大数据，决定着数字经济发展的深度。

以算力基建化为主体，将通用算力与多元异构芯片集成，融合多种算力，进行端边云协同；以算法服务化为引领，通过提供预置行业算法、构建预训练大模型、推进算法模型持续升级、提供专业化数据和算法服务；以服务智件化为依托，将人工智能算法开发和模型训练从专业化、高门槛向泛在化、易用型转变。

本论坛旨在汇集相关领域的学术专家、行业大咖、企业管理者，共同探讨上述各方面的问题及解决措施，并汇聚多方观点，以期形成新的解决思路。

报告嘉宾

嘉宾介绍

田 原 中通服咨询设计研究院通信院



田原，中通服咨询设计研究院通信院院长，中共党员，工学硕士，中国节能协会碳中和专业委员会委员，江苏省土木建筑学会工程管理专业委员会委员，一级注册结构师。主要从事通信工程、结构工程、数据中心双碳等多类型项目的设计、项目管理及课题研究，拥有十余项专利成果。带领团队在算力、双碳、5G-A、低空经济、视联网、工业 pon 等方向均取得多项突破性成果。主持编写中国电信工业 pon 设计、施工、运维规范；参与编写国内首个 5G 增强低空网络白皮书。荣获绽放杯、光华杯、“华彩杯”算力应用创新大赛等诸多奖项。

报告题目

智算中心建设经验分享及视频领域解决方案介绍

报告摘要：首先介绍中通服咨询设计研究院的发展史、愿景及定位和资质荣誉，其次就算力发展的政策环境、近年全国算力发展规模、算力中心整体架构进行分析，并展示我院算力能力体系、算力产品体系、算力服务体系，然后就“东数西算”工程等优秀案例进行分享；在视频监控领域，主要围绕电信运营商的视频解决方案和案例进行报告。

嘉宾介绍



梅 迪 广州希姆半导体科技有限公司

梅迪，毕业于华中科技大学及中国科学院大学电子通信工程专业。有近二十年顶尖外企电子器件、半导体产品线管理经验。2019 年创立希姆计算，致力于开展数据中心级别高端芯片设计研发，熟悉 RISC-V 技术路线，对专用计算和通用计算发展及 AI 芯片的商业化理解有前瞻思考。

报告题目

开放的生态-国产算力发展之路

报告摘要：近年来，随着全球对芯片自主可控需求的增长以及物联网、边缘计算等领域的需求不断扩大，RISC-V 得到了广泛关注和应用，逐渐成为第三大指令集架构。在 AI 时代，RISC-V 更是迎来爆发的新机遇，从物联网设备、边缘计算逐渐迈向 AI 计算、高性能计算等领域。

今年以来，我国企业在 RISC-V+AI 开源技术体系上更多地成为主力贡献者，成为全球化开放共享的主导力量。这已经变成了一个国家战略。用开源打破生态垄断，降低企业拥有核心技术的门槛，让每个企业都能低成本地做自己的芯片，形成智能芯片的汪洋大海，满足无处不在的智能需求。

嘉宾介绍



王斌 世界超高清视频产业联盟

王斌，现任世界超高清视频产业联盟（UWA）产业研究总监。负责产业洞察与研究、产业与技术规划、专家组运营工作。超过 25 年 ICT 领域研发和市场经验，曾任华为首席专家、产品研发 SPDT 经理，在网络、软件平台、云服务领域具有多年产品规划研发与上市经验。目前专注视听产业领域规划研究与创新开源产业工作。

报告题目

"道合声远 菁彩未来"，超高清产业生态发展分享

报告摘要：超高清技术已经在个人、家庭、车载、体育赛事以及包括医疗，工业等各领域广泛应用。超高清根技术从六维度扩展到 3D，XR，元宇宙。在最近两年，随着超高清在公众娱乐，入户，车载等场景逐渐深入，以及 AI 技术的快速发展，超高清和 AI 在内容制作，终端，传输放映各领域不断深入。本报告将分享超高清最新热点，超高清根技术标准的制定及生态进展以及对未来发展的展望。

嘉宾介绍



范益波 复旦大学

范益波，本科毕业于浙江大学，硕士毕业于复旦大学，博士毕业于日本早稻田大学。主要从事视频图像、人工智能、硬件架构与 SoC 芯片设计研究。出版专著 2 部，发表学术论文 140 多篇，取得发明专利 50 多项。获得上海市“东方学者”特聘教授、浙江省“钱江特聘专家”等。创建 OpenASIC 开源芯片论坛，发布 xk264、xk265 开源视频编码处理器 IP 核、xkiSP 开源图像处理器 IP 核、xkDLA 开源神经网络处理器 IP 核。致力于从事图像 ISP 处理器、视频 VPU 处理器、人工智能 NPU 处理器的前沿技术研究与处理器 IP 核开源。

报告题目

xkiSP：开源 ISP 处理器及应用

报告摘要：图像信号处理器（ISP）是视频采集、机器视觉的最前端核心模块，主要负责对传感器输出的原始 RAW 信号做计算成像，实现传感器输出高画质的 RGB 或 YUV 信息。ISP 的性能直接影响到图像和视频的质量。一个优秀的 ISP 能够让设备在各种复杂的环境和条件下拍摄出令人满意的图像，为用户带来更好的视觉体验。同时，也为图像识别、计算机视觉等技术的应用提供了高质量的数据基础。本次报告将介绍我们开发的开源 ISP 处理器—xkiSP，具体包括如下几个部分：1) xkiSP 的全栈算法与开源处理器架构；2) 基于 AI 的 ISP Tuning 框架；3) AI-ISP 的架构探索与挑战；4) xkiSP 的改进与应用展望。

嘉宾介绍



余世兵 北京凌川科技有限公司

余世兵，拥有 10+ 年音视频相关领域的产品经理工作经验。2017 年创立在线视频生产平台 Onvideo，2020 年被快手收购加入快手音视频团队，目前为北京凌川科技有限公司解决方案负责人。

报告题目

“视觉智能 2.0”时代对算力挑战的思考和实践

报告摘要：近期快手可灵爆火，普通用户实实在在体验到了文生视频大模型带来的震撼，视觉智能正在从“人脸识别 1.0”到“智能可用 2.0”加速升级和落地。视觉 AI 2.0 时代，视频内容生产、理解、分发和互动的技术链路；视觉智能对算力、技术底座的需求都将提出全新挑战。孵化于快手的凌川科技团队，多年来围绕智能视频处理、AI 算力两个方向持续开展核心技术攻关和应用落地探索，对视觉 AI2.0 时代算力挑战积累了一定的思考和实践经验。

嘉宾介绍

**李 俊 北京博雅睿视科技有限公司**

李俊，拥音视频领域核心厂商服务 15 年以上，拥有丰富的产品规划、解决方案与项目管理经验。在超高清视频逐渐成为主导媒介的大趋势下，专注利用编解码算法优化、AI 模型构建、大数据分析、以及云计算等技术的组合应用，探索如何采用信创国产化的技术、产品与解决方案，构建全链路自主可控视觉智算产业生态。

报告题目**用“芯”构建可信可进化的视觉智算解决方案**

报告摘要：视觉主导的信息时代，信息的处理由计算迈入智算，AVS 视觉智算芯片的问世、数字视网膜的演进、大模型的出圈及落地，为算据积累、算法优化及算力协同提供了新的解决思路，使得基于高质量窄带智算、高并发可信互联、分布式端测推理的视觉智算解决方案触手可及，博雅睿视致力于视觉前沿技术的研究，探索构建看得全、识得准、自进化、有依据的可信视觉智算解决方案。

▶ 最佳论文评选论坛 ◀

主持人：胡海苗 单位：北京航空航天大学

时间 2024年7月24日 上午 09:30-12:00

地点 多功能厅 7

候选论文清单

ID	Paper ID	题 目
1	45	基于域适应增强和多尺度特征融合细化的跨域小样本目标检测方法
2	54	基于组合上下文提示的大型语言模型领域知识问答研究
3	78	基于语义分离和特征融合的人脸编辑技术
4	114	基于差异化特征提取的交叉半监督语义分割网络
5	129	基于时域注意力机制的动态点云几何压缩方法
6	199	细粒度情感可控的视频生成算法研究
7	213	基于通道变换和 Transformer 的高光谱图像变化检测方法
8	14	KN-VLM: Knowledge-guided Vision-and-Language Model for Visual Abductive Reasoning
9	25	BGFlow: Brightness-guided Normalizing Flow for Low-light Image Enhancement
10	74	Exploring the Privacy Protection Capabilities of Chinese Large Language Models
11	82	Point Cloud Upsampling via Implicit Shape Priors Discovery and Refinement
12	158	Dark Knowledge Associative Guided Hashing for Unsupervised Cross-Modal Retrieval
13	171	Stereoscopic Visual Attention Model Based on Bioinspiration
14	190	Inter-image Token Relation Learning for Weakly Supervised Semantic Segmentation
15	224	3D Human Model Guided Pose Transfer via Progressive Flow Prediction Network

ChinaMM2024 论文墙报展示（英文）

主持人：刘武 单位：中国科学技术大学

时间 2024年7月24、25日 09:00-17:00

地点 黄河厅 前厅

论文展示列表

ID	题 目
1	Mutual-weighted Feature Disentanglement for Unsupervised Domain Adaptation
2	Multi-Modality Boundary-Guided Network for Generalizable Image Manipulation Localization
3	Superpixels Rectify CAMs: Contrastive Activation Maps with Superpixel Rectification for Weakly Supervised Semantic Segmentation
4	KN-VLM: Knowledge-guided Vision-and-Language Model for Visual Abductive Reasoning
5	UnSP: Improving Event-to-Image Reconstruction with Uncertainty Guided Self-Paced Learning
6	Exploring Coherence from Heterogeneous Representations for OCR Image Captioning
7	Micro-expression Recognition Based on Optical Flow Features and Multi-attention Mechanism
8	Facial Action Unit Detection with Emotion Consistency: A Cross-Modal Learning Approach
9	BGFlow: Brightness-guided Normalizing Flow for Low-light Image Enhancement
10	Leveraging Occupancy Map to Accelerate Video-based Point Cloud Compression
11	Parentheses Insertion Based Sentence-level Text Adversarial Attack
12	Dual-guided Multi-modal Bias Removal Strategy for Temporal Sentence Grounding in Video
13	Manifold Structure Preservation of Latent Domain and Category Diversity of Target Domain for Unsupervised Domain Adaptation
14	AdapSyn: Anomaly Detection Based on Triplet Training With Adaptive Anomaly Synthesis
15	Self-Supervised Tone Mapping for Extreme High Dynamic Range Images
16	LightNet: A Lightweight Monocular Depth Estimation for High-level Guidance and Channel Re-alignment Optimization
17	A Blockchain-based Verifiable Image Editing System
18	Exploring the Privacy Protection Capabilities of Chinese Large Language Models
19	Point Cloud Upsampling via Implicit Shape Priors Discovery and Refinement
20	Exploring Multi-Dimensional Interests for Session-based Recommendation
21	Quality Evaluation for Text-to-Image Generation: A Survey

ID	题 目
22	GES: Generalization Enhancement Strategy Based on Ensemble Learning for Open Domain Image Manipulation Detection
23	Uniform-reference Threshold-dynamic skipping for Video Compressive Sensing
24	Interactive Image Segmentation with Non-uniform Noise Diffusion Model
25	Chatting with Interactive Memory for Text-based Person Retrieval
26	Prioritized Experience Replay in Dual Experience Pools Based on State Similarity via KL Divergence
27	GACFNet: A Global Attention Cross-Level Feature Fusion Network for Aerial Image Object Detection
28	Pro-MA: Progressively Margin-based Attribution in VLMs
29	Dark Knowledge Associative Guided Hashing for Unsupervised Cross-Modal Retrieval
30	Exploring Cross-modal Representation Learning via Hierarchical Bi-direction Concept Interaction for Text-Video Retrieval
31	Degradation Compensation and Multi-dimensional Reconstruction Based Pre-processing for Video Coding
32	A Multi-scale Discriminator Style Transfer Model Integrating Attention Mechanism
33	Stereoscopic Visual Attention Model Based on Bioinspiration
34	Knowledge-Guided Quantization-Aware Training for EEG-Based Emotion Recognition
35	Multi-View Anomaly Detection via Hybrid Instance-Neighborhood Aligning and Cross-View Reasoning
36	Question-augmented Knowledge Retrieval for Visual Question Answering
37	Scene-Text Aware Cross-Modal Retrieval Based on Semantic Matching
38	Inter-image Token Relation Learning for Weakly Supervised Semantic Segmentation
39	Constraint Embedding for Prompt Tuning in Vision-Language Pre-Trained Model
40	Thick Cloud Removal for Aerial Video Based on Semi-supervised Segmentation and Optical Flow Estimation
41	Temporal Refinement and Multi-grained Matching for Moment Retrieval and Highlight Detection
42	SPFont: Stroke Potential Features Embedded GAN for Chinese Calligraphy Font Generation
43	Lossless Medical Ultrasound Image Compression Based on Frequency Domain Decomposition
44	3D Human Model Guided Pose Transfer via Progressive Flow Prediction Network
45	Integrated Gaussian CLIP for Continual Learning
46	ROI-weighted Dual Attention Spatial Convolution Fusion Algorithm Based on Deep-board Learning for Identifying Benign and Malignant Thyroid Nodule Tumors
47	Global-local Co-regularization Network for Facial Action Unit Detection
48	CCA-Net: A Cascade Connecting Attention Network for LiDAR 3D Object Detection
49	AMAN:Attention-Modulated Adversarial Network for Blind Mural Image Completion

ChinaMM2024 论文墙报展示 (中文)

主持人: 陈智能 单位: 复旦大学

时间 2024 年 7 月 24、25 日 09:00-17:00

地点 黄河厅 前厅

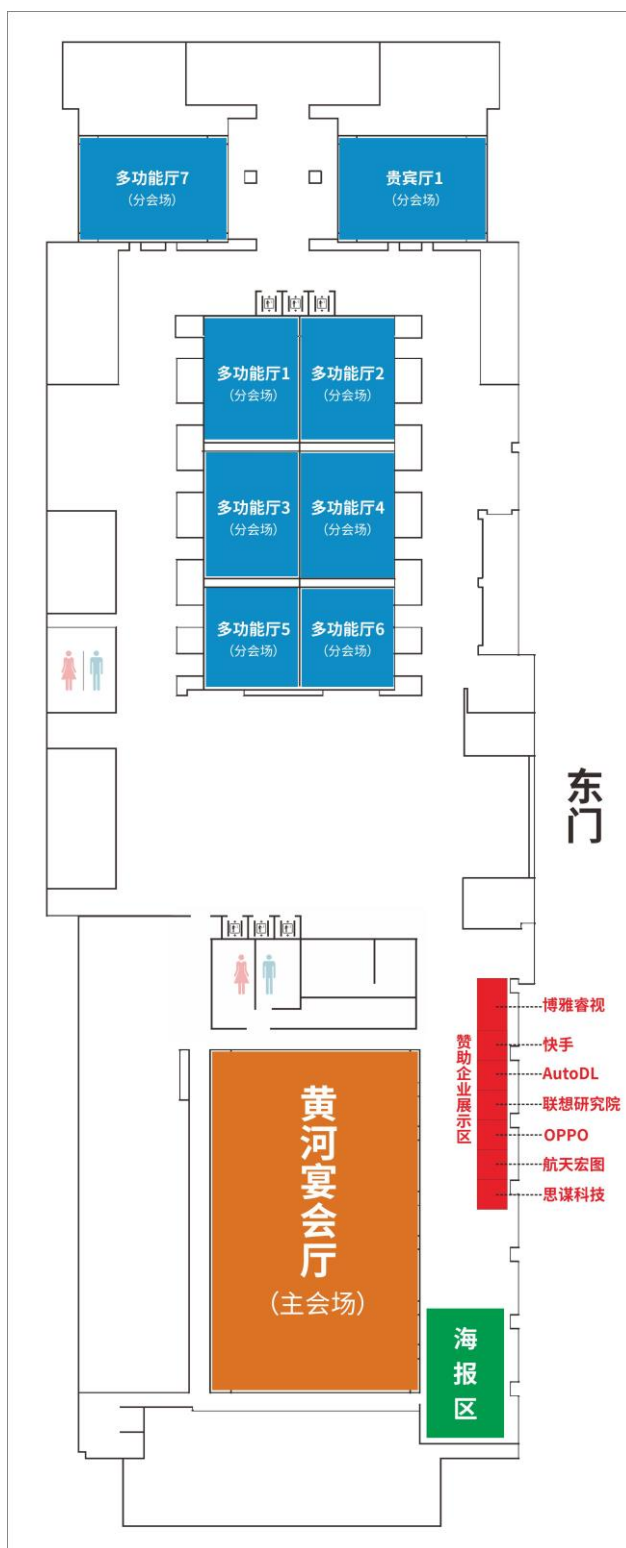
论文展示列表

ID	题 目
1	基于特征空间增强重放和偏差校正的类增量学习
2	LLM 引导的有解释性分数描述的盲图像质量评价
3	跨模态信息融合的视频文本检索
4	息肉边界线索深度融合网络用于息肉图像分割
5	基于域适应增强和多尺度特征融合细化的跨域小样本目标检测方法
6	时空分离变形记忆网络用于视频异常检测
7	大语言模型开放域多标签动作识别
8	基于组合上下文提示的大型语言模型领域知识问答研究
9	时空融合的车载环视摄像头外参联合标定算法
10	X 射线安检图像目标检测: 高精度实时检测模型及大规模基准数据集
11	基于最大反射率先验和变分正则化的夜间图像去雾算法
12	基于变分注意力知识选择和预训练模型的对话生成方法
13	大语言模型作为知识库的查询风格自适应转换
14	基于动态 KNN 图构建与图注意力卷积神经网络的多生物特征识别
15	基于多角度拓扑细化的超图卷积网络
16	基于多层次感知的 RGB-D 显著性目标检测网络
17	基于语义分离和特征融合的人脸编辑技术
18	基于全局和局部对比学习特征生成的小样本图像分类
19	基于个性化联邦学习的行人重识别方法
20	高动态链路下基于主机状态感知的视频码率自适应算法
21	基于深度学习的多目标跟踪综述
22	基于混合注意力的双分支 U-Net 高光谱全色锐化

ID	题 目
23	噪声标注条件下非对称互学习引导的跨域行人重识别
24	基于差异化特征提取的交叉半监督语义分割网络
25	基于显著性掩模混合的小样本图像分类
26	基于知识感知的图优化推荐模型
27	GPU 资源自感知的模型动态部署方法
28	基于时域注意力机制的动态点云几何压缩方法
29	面向端到端图像压缩的可分级码率控制方法
30	基于先验分布远离的模型逆向攻击方法
31	多尺度一致性对抗学习的半监督医学图像分割方法
32	基于多模态特征融合的异常行为检测方法
33	基于动态多粒度图卷积网络的人体骨架行为识别
34	注意力重调制和局部知识转移的弱监督语义分割
35	基于空间特征增强图卷积网络人体骨架行为识别
36	基于特征增强和模态交互的视频异常行为检测
37	基于多模态图学习的自闭症检测方法
38	基于置信度引导提示学习的多模态方面级情感分析
39	基于多类型提示互补的弱监督时序动作定位
40	感知约束和引导下的特征点增强局部水印算法
41	基于混合注意力和双向门控网络的高光谱遥感图像变化检测方法
42	基于稀疏注意力和可变形特征交叉融合的多源遥感图像分类方法
43	细粒度情感可控的视频生成算法研究
44	基于视觉语言双向感知的图像指代分割方法
45	音效伪造数据集的构建与基准检测方法研究
46	基于跨网络特征补偿的声呐图像质量评价
47	CLIP 驱动的基于对比学习和图神经网络的指称图像分割
48	基于通道变换和 Transformer 的高光谱图像变化检测方法
49	基于门控注意力聚合网络的多源遥感图像分类方法

ID	题 目
50	基于双分支 Transformer 的多源遥感图像分类方法
51	一种基于骨架关键点的遮挡场景行为识别方法
52	基于全局和局部动态注意力机制的高光谱图像去噪方法
53	基于双分支网络架构的安检 X 光图像违禁品分割
54	融合领域知识与自适应时空片段 Transformer 的人体骨架行为识别方法
55	基于跨模态超图优化学习的多模态情感分析
56	基于多模态无人机视频的分布式光伏故障检测算法及其云平台研究
57	基于新型势场模型的二维骨架计算
58	结合特征增强与双分支特征学习的视频行人重识别
59	基于球谐函数局部特征投影的三维人脸生成方法
60	VVC 低延迟分层编码的 CTU 级优化算法
61	基于大语言模型知识增强和多特征融合的中文命名实体识别方法

会场平面图



▶ 联办单位展宣 ◀

北方民族大学

北方民族大学直属于国家民族事务委员会，是一所综合性普通高等院校。学校坐落在宁夏回族自治区银川市，是我国唯一一所建立在民族地区的中央部委高校。



学校坚持以习近平新时代中国特色社会主义思想为指导，深入学习贯彻习近平总书记关于加强和改进民族工作的重要思想、关于教育和人才工作的重要论述，全面贯彻党的民族政策和教育方针，坚持党的全面领导，坚持立德树人根本任务，坚持铸牢中华民族共同体意识主线要求，坚持激发改革创新根本动力，坚持走内涵式高质量发展之路，成为研究宣传党的民族理论政策、培养造就各民族优秀干部人才、服务支撑民族地区经济社会发展的重要基地。被国家民委授予全国民族团结进步教育基地。

学校现有文学、理学、工学、法学、历史学、管理学、经济学、艺术学、教育学 9 个学科门类的 77 个本科专业。学校现有博士生、硕士生、本（预）科生 2.3 万余人在读，是中华民族大家庭的缩影。现有教职工近 1500 人，聘任院士 4 名，入选国家级人才项目或获得荣誉称号 10 人次。“图像处理与智能计算教师团队”获批为全国高校黄大年式教师团队。学校现有 1 个国家地方联合工程研究中心、1 个示范型国家国际科技合作基地；2 个国家级“全国化工及机械基础科普基地”“电子科学与物理示范中心科普基地”。学校坚持开放办学，积极开展国内外合作交流。合肥工业大学为教育部指定的对口支援高校。学校与北京大学、银川市人民政府、宁夏共享集团等深入开展校校、校地、校企合作，互促共赢。学校设有 3 个教育部/国家民委国别和区域研究中心，获批教育部中外合作办学项目 1 项。学校积极开展“双创”教育，实施“专创融合‘5+3’教育打造‘六个一’创新人才培养体系”。

站在新的历史起点上，学校坚持铸牢中华民族共同体意识工作主线，坚守民族院校初心使命，传承民族院校红色基因，全面深化综合改革，深入推进转型升级，扎实推进内涵建设，实现高质量发展，努力将学校建设成为特色鲜明的高水平现代化综合大学！

► 协办单位展宣 ◀

宁夏大学信息工程学院

宁夏大学信息工程学院是宁夏大学为适应高等教育转型发展、满足社会对IT人才的需求,在整合学校相关专业和学科优质资源的基础上于2016年5月13日正式挂牌成立的新学院,是宁夏回族自治区应用型转型试点学院,2017年与NIIT(印度国家信息学院)合作办学。建院以来,全院抢抓网络强国、数字中国发展机遇,将大数据人工智能与信息安全领域作为重点培育。学院已建有“宁夏大数据与人工智能省部共建协同创新中心”、“宁夏‘东数西算’人工智能与信息安全重点实验室”、“网络安全等级保护测评中心”、并行分布处理与大数据挖掘分析、高性能计算、嵌入式系统、图形图像与信号处理、物联网、计算机网络与下一代互联网、数据库与软件工程、计算机组成原理等13个研究实验室和计算机技术、软件工程2个校内实训基地。



学院现有专任教师67人,其中教授13人,副教授35人,具有博士学位者44人,硕士生导师47人。入选中国科协青年托举人才1人、宁夏科技创新领军人才1人、宁夏科协青年科技人才托举工程5人、宁夏青年拔尖人才1人、中科院“西部之光”青年学者5人、中国计算机学会高级会员5人,特聘宁夏大学“贺兰山学者”5人。学院现有计算机科学与技术、网络工程、软件工程3个本科专业。其中计算机科学与技术专业入选国家首批一流本科专业建设“双万计划”,同时也是西部一流建设学科。拥有1个一级学科硕士点(计算机科学与技术),2个学术型硕士学位授权点(计算机软件与理论、计算机应用技术)和1个专业型硕士学位授权点(计算机技术)。

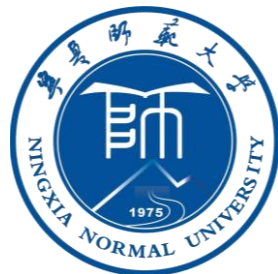
学院核心科研创新团队围绕智慧农业、健康信息、生态计算、网络攻防等方向,协同东部先进网络与人工智能优势技术,形成面向西部地域特色智慧产业与数字经济的行业应用服务平台,初步探索了宁夏“东数西算”协同创新模式,落地国家算力网络枢纽节点时机成熟、基础牢固、优势显著。学院秉承“坚持协同创新,注重资源整合;坚持需求导向,侧重应用研究;坚持前沿跟踪,着眼西部特色”的建设原则,以“学为学者,工为工匠”为院训精神,将学院建设成为西部一流应用研究型信息学院。

▶ 协办单位展宣 ◀

宁夏师范大学

宁夏师范大学坐落于宁夏南部六盘山下历史名城——固原市原州区，隶属宁夏回族自治区人民政府管理，是宁夏唯一的师范类本科院校。

学校前身为 1975 年建立的六盘山大学。1978 年经国务院批准在六盘山大学基础上建立固原师范专科学校，1994 年更名为固原师范高等专科学校，2006 年经教育部批准更名为宁夏师范学院，升格为本科院校。2012 年获批“服务国家特殊需求人才培养项目”——学士学位授予单位开展培养硕士专业学位研究生试点工作单位，并于 2013 年开始招生。2013 年通过教育部本科教学工作合格评估，同年，获得地方公费师范生培养资格。2016 年成为宁夏首批整体应用转型试点高校、自治区“西部一流”学科建设单位，获批外国留学生招生资质。2018 年获批硕士学位授予单位、自治区高水平本科教育试点学校、自治区人工智能助推教师队伍建设行动试点学校。2019 年顺利通过教育部本科教学工作审核评估。2021 年 11 月，获批“自治区开展本科层次职业教育试点建设高校”。2022 年入选“教育部师范教育协同提质计划”，2023 年自治区人民政府学位委员会办公室将宁夏师范大学列入 2024-2026 新增博士学位授予立项建设单位。



学校现有古雁、文苑两个校区，总占地面积 92.84 万平方米，校舍建筑总面积 32.25 万平方米。学校教学科研仪器设备总值 6133.61 万元。中外文纸质图书 114.49 万册，数字资源 109 万种。《宁夏师范学院学报》面向全国公开发刊，是首届全国“百强”社科学报。学校现有教学单位 15 个、教辅单位 8 个、科研单位 2 个、党群行政单位 19 个。在人文学科、社会学科、理学、工学等 4 个学科门类共有 18 个硕士学位授权点。有法学、教育学、文学、历史学、理学、工学、管理学、艺术学 8 个学科门类 41 个普通本科专业，初步形成研究生教育、本科教育不同层次，全日制高等教育、高等职业教育、成人继续教育、留学生教育等不同类型的人才培养体系。

“十四五”时期，学校确立了“升大创博”新目标，将以习近平新时代中国特色社会主义思想为指导，坚持社会主义办学方向，落实立德树人根本任务，坚持师范教育办学主体，突出教师教育优势特色，推进文法、理工、艺术、体育多学科协调发展，以培养造就适应经济社会发展需要的高素质专业化创新型人才为方向，以建设高水平教学研究型大学为目标，以服务基础教育和经济社会发展为己任，到 2025 年，建成教师教育特色优势鲜明、服务地方发展能力强的西部地区高水平师范大学和国家师范教育基地。

► 赞助企业展宣 ◀

钻石赞助



北京博雅智视科技有限公司
BEIJING BOYA INTELLIGENT TECHNOLOGY CO., LTD.

铂金赞助



金牌赞助



银牌赞助



▶ 赞助企业展宣 ◀

博雅睿视

北京博雅睿视科技有限公司（简称：博雅睿视），作为超高清视频与智能视觉解决方案提供商，国家高新技术企业，北京市专精特新企业，依托北京大学视频与视觉技术国家工程研究中心的科研力量，以及在视频编码与处理、AI 机器视觉、沉浸式视频交互等领域深厚的技术积累，为客户提供超高清视频编解码 IP/芯片与智能视觉系统解决方案。公司核心团队多次获得国家、省部及相关行业奖项荣誉，包括 2020 年国家技术发明一等奖、2022 年北京市科学技术进步一等奖。

► 赞助企业展宣 ◀

AutoDL

视拓云主要从事 AI 云计算、算力 C 端零售、算力资产运营、开发者社区服务和 AI 服务器硬件的设计、生产和销售等业务。视拓云运营的 AutoDL.com 是目前全国最大的 C 端 AI 算力租用平台之一。平台具有从服务器生产制造、AIDC 建设到算力云服务的全产业链能力。其运营超过 1.4w 张、20 多个型号的 GPU 和国产 AI 加速芯片，具备“万卡”和“千 P”算力的调度能力；面向“大 AI 圈”内的科研工作者和科技企业提供算力分时租赁服务和云计算服务。公测 2 年多来，根据开票数据统计，AutoDL 累计为超过 35w 开发者，142 所 985、211、双一流高校，540 多所其他高校，以及 3000 多家企业提供了弹性、省钱、好用的普惠 AI 云算力服务。

► 赞助企业展宣 ◀

快手

快手是全球领先的内容社区和社交平台，致力于成为全球最痴迷于为客户创造价值的公司。快手目前是中国最广泛使用的应用之一，2024 年第一季度快手应用国内平均日活跃用户达 3.94 亿，国内平均月活跃用户达 6.97 亿。自 2011 年创立以来，快手始终秉承“用有温度的科技提升每个人独特的幸福感”这一初心，用极致的技术为亿万用户提供极致的用户体验。目前，由快手自主研发的大语言模型“快意”、图片生成大模型“可图”和视频生成大模型“可灵”，已与搜索、评论、商业化等多场景结合，并已进行多个短视频直播产品的内测，打造出“AI 智能对话”、“AI 小快”、智能客服等智能应用。此外，快手 StreamLake 自研的 SL200 芯片在 MSU2022 世界编码器大赛中夺得 16 项第一，该芯片集成诸多领先关键技术，可以最大限度地提升视频音画质量，同等质量下对比业界主流可节省约 30% 带宽成本。2023 年 7 月快手正式投入运营快手智能云乌兰察布数据中心，作为中国最大的「大数据+人工智能」数据中心之一，预计年节电达 5 亿度左右，IDC 总成本和总能耗均下降 30% 以上。

同时，快手通过构建以“短视频 + 直播”为基础的内容社区，建立起真实、美好、多元、有益的内容生态，让每位用户都能发现所需、发挥所长。快手坚持构建普惠的数字社区，并逐步由人工智能驱动的内容创作平台发展为涵盖内容消费、知识传播、电子商务、本地生活服务、蓝领招聘等功能于一体的数字平台，将技术和产品赋能应用于各个产业，为社会创新发展与持续进步注入动力，为数实融合与数字中国建设贡献力量。

► 赞助企业展宣 ◀

OPPO

OPPO 是一家全球领先的智能设备制造商，成立于 2004 年，产品遍布 60 多个国家和地区。公司在全球拥有 260000 多个零售店和 3100 多个客户服务门店，拥有八大智能制造中心和伦敦设计中心，员工人数超过 48000 人。OPPO 以人性化科技为基因，2008 年推出“笑脸手机”A103，2011 年进入智能手机市场，以 Find 系列探索未来。2018 年推出 Find X，2019 年推出 Reno 系列，2021 年推出 Find N 系列，2024 年进入 AI 手机时代，Find X7 系列 AI 赋能多场景。

OPPO 不仅深耕手机业务，还构建了多智能终端生态，包括穿戴、声学、平板等 IoT 产品，提供跨场景高频使用的入门级产品。公司优化以 ColorOS 为核心的软件平台，为全球 6.57 亿多用户打造人性化、智能的移动操作系统，并提供软件商店、云服务、智能助手等增值服务。OPPO 自 2009 年开启海外市场业务，覆盖多个地区，成为世界第四大智能手机品牌。公司相信年轻的力量，推出 Renovators 全球青年创享计划，为艺术高校、青年艺术家和爱好者提供舞台。OPPO 研究院发起“微笑提案”科创赋能平台，征集创新技术解决方案，共创智慧生活。

► 赞助企业展宣 ◀

航天宏图

航天宏图信息技术股份有限公司（股票代码：688066），成立于 2008 年，是国内领先的卫星互联网企业、科创板首批上市企业。公司研发了具有完全自主知识产权的遥感与地理信息一体化软件 PIE（Pixel Information Expert），拥有国内首个遥感与地理信息云服务平台 PIE-Engine，实现遥感基础软件的国产化替代；规划了我国规模最大的多层次、多模式混合遥感卫星星座——“女娲星座”，一期工程计划发射 54 颗业务卫星，其中首批 4 颗雷达遥感卫星已于 2023 年 3 月 30 日成功发射，2024 年计划发射 12 颗 X 波段 SAR 成像卫星，业务能力向全卫星产业链拓展；构建面向全国的无人机生产与服务体系，打造“天空地”一体化服务能力。为政府、企业、高校以及其他有关部门提供基础软件产品、系统设计开发、遥感云服务等空间信息应用整体解决方案。

公司总部位于北京，在全国设有 200 余个分支机构，在西安、成都、武汉、南京、长沙等地建立研发中心。公司现有员工 3300 余人，其中博士 130 余人、硕士 1000 余人，海外留学归国人才及行业专家 200 余人，工程技术人员占 80%以上。设立博士后工作站，与岳清瑞院士团队共建城市应急管理卫星应用院士工作站，建设 1 个北京市工程实验室、2 个自然资源部重点实验室、4 个自然资源部工程技术创新中心。拥有国家重点高新技术企业、双软企业、CMMI L5、测绘甲级、系统建设和服务能力评估 4 级、工程咨询资信资格、通用航空企业经营许可、民用无人机驾驶员训练机构等资质。拥有 200 余项发明专利和 1000 余项软件著作权。

► 赞助企业展宣 ◀

联想研究院

联想研究院创立于 1999 年，是联想集团的公司级技术研发机构，聚焦未来 2-5 年的前沿技术创新，是联想核心技术的提供者和创新先锋。

目前，联想集团首席技术官、高级副总裁、联想技术委员会副主席芮勇博士，领导联想研究院的工作。加入联想之前，芮博士在微软任职 18 年，曾任微软亚洲研究院(MSRA)常务副院长，领导关键的人工智能团队和研究院的所有工程团队。芮博士还是中国计算机学会会士，中国人工智能学会会士，欧洲科学院 (Academia Europaea) 外籍院士，国际计算机学会会士(ACM Fellow)，国际电气电子工程学会会士(IEEE Fellow)，国际模式识别学会会士(IAPR Fellow) 以及国际光学工程学会会士(SPIE Fellow)。

► 赞助企业展宣 ◀

思谋科技

思谋科技 (SmartMore), 智能制造的持续创新者, 以“致力持续创新, 创造卓越价值”为使命, 是一家具备“光”“机”“电”“算”“软”全栈领先能力的标准软硬一体化产品及解决方案供应商。经过二十多年的技术积累与沉淀, 结合独特的商业思考, 思谋科技持续打造更具拓展性和普惠价值的智能工业和数智创新平台、工业大模型, 推动探索工业制造业的智能化升级和数字化转型。

目前, 思谋已通过自研的智能工业平台、智能传感器产品以及智能一体化设备, 服务了全球近 300 家行业头部企业, 以技术促进更高效、更灵活、更先进智造的发展; 此外, 思谋还不断拓宽智造外延, 基于“智造+”平台与数智化解决方案, 自主研发了数字化制造管理系统, 覆盖了从产线到工厂的应用场景, 为客户在全球范围内提供全面而优质的产品与方案服务。

思谋自成立以来一直坚持国际化发展路线, 致力于成为一家服务全球的全场景智能公司, 吸引了近千名来自世界知名学府和行业领先企业的国际化人才。公司已在香港、深圳、上海、北京、苏州、杭州、重庆、新加坡和日本东京等多地设有前沿技术研发与商务中心, 在东南亚和欧洲等地建立了代表处及合作伙伴网络, 商业布局遍布全球。



▶ 赞助企业展宣 ◀

抖音集团

抖音集团作为字节跳动国内信息和服务业务板块，秉承“激发创造，丰富生活”的使命，致力于为用户提供更优质的内容及服务。公司拥有抖音、今日头条、西瓜视频、番茄小说、懂车帝等产品。

官方网站：<https://www.bytedance.com/zh/>

人才招聘：<https://jobs.bytedance.com/>

主办单位：中国图象图形学学会(CSIG) 中国计算机学会(CCF)

承办单位：CSIG多媒体专业委员会 CCF多媒体技术专业委员会

联办单位：北方民族大学

协办单位：宁夏大学 宁夏师范大学

钻石赞助



北京博雅睿视科技有限公司
BEIJING BOYA REALSCENE TECHNOLOGIES CO., LTD

铂金赞助



AutoDL



快手

金牌赞助

oppo



航天宏图
PIESAT

联想研究院 Lenovo Research

银牌赞助



SmartMore 思谋



抖音集团
Douyin Group



扫码查看电子版手册

