VALSE 2025 珠海

视觉与学习青年学者研讨会

会议手册

主办单位：中国图象图形学学会

承办单位：中山大学

协办单位：中国图象图形学学会青年工作委员会
AutoDL

2025年6月6-8日

中国·珠海

目 录

VALSE 2025 欢迎您

欢迎大家来到广东珠海，享受VALSE给大家带来的学术盛宴。

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

围绕上述目标，过去十四年来，VALSE逐渐形成了自己的特色社区文化、找准了自己的使命，包括：

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

第十五届VALSE大会于2025年6月6-8日在珠海举行。届时大会将呈上6个大会主旨报告、12个领域年度进展(APR)报告、4场Tutorial、20场专题研讨会(Workshop)、390余篇顶会顶刊Poster，合计共有百余位知名青年学者共同带来视觉与学习等人工智能领域的一次学术盛会。同时，在会场合作单位展台区还将展示合作单位精彩的演示。

鉴于VALSE坚持不向参会者收取任何费用，故特别感谢华为、AutoDL、百度、OPPO、腾讯优图、美团、合合信息、阿里妈妈、无问芯穹、银河通用、阿里云、小米、TCCI、九章云极、极视角、真格基金、美图、金山办公、思腾合力、思谋科技、爱诗科技、趋动科技、首都在线（GpuGeek）、将门、易方达、摩尔线程、数据堂、英博数科、元戎启行、云天励飞、杭州宇树科技等合作单位对本次会议提供的鼎力支持。

VALSE 2025 组委会

大会组委会

大会主席 (General Chairs)

郑　跃	中山大学
赖剑煌	中山大学
刘青山	南京邮电大学
潘　纲	浙江大学
郑伟诗	中山大学

程序委员会主席 (Program Chairs)

姬艳丽	中山大学
郭裕兰	中山大学
苏　航	清华大学
左旺孟	哈尔滨工业大学
李　玺	浙江大学

Tutorial Chairs

王楠楠	西安电子科技大学
夏　勇	西北工业大学
刘日升	大连理工大学
戴玉超	西北工业大学
谢雨彤	阿德莱德大学

Poster Chairs

郑　乾	浙江大学
刘　波	重庆邮电大学
明　悦	北京邮电大学
张　青	中山大学
胡　迪	中国人民大学
郭晓杰	天津大学

展览主席

王福田	安徽大学
冷佳旭	重庆邮电大学
杨　猛	中山大学
余建兴	中山大学

Website Chair

郑海永	中国海洋大学

Finance Chairs

程明明	南开大学
吴岸聪	中山大学

Registration Chairs

郭宗辉	中国海洋大学
袁　霖	重庆邮电大学
高陈强	中山大学
郭春乐	南开大学
刘晋源	大连理工大学
陈建国	中山大学

宣传主席

贾　伟	合肥工业大学
刘　昊	宁夏大学
张瑞茂	中山大学
涂志刚	武汉大学
李　镇	香港中文大学（深圳）

Sponsorship Chairs

山世光	中科院计算所
姬艳丽	中山大学

Workshop Chairs

徐天阳	江南大学
胡　鹏	四川大学
刘夏雷	南开大学
张姗姗	南京理工大学
江　波	安徽大学
李　策	兰州理工大学
任传贤	中山大学
舒祥波	南京理工大学
沈　为	上海交通大学
赵　健	中国电信集团、西北工业大学

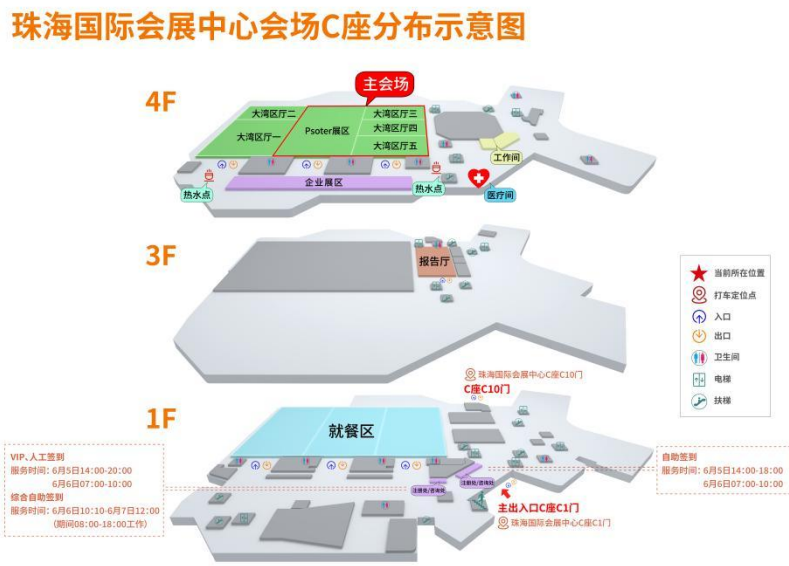
APR Chairs

谢凌曦	华为技术有限公司
张利军	南京大学
李冠彬	中山大学
崔兆鹏	浙江大学
冯　婕	西安电子科技大学
魏秀参	东南大学

本地主席

胡建芳　中山大学
常晓斌　中山大学
金　枝　中山大学
王瑞轩　中山大学
唐彦嵩　清华大学深圳国际研究生院

VALSE 2025 会场分布



VALSE2025 会议总体日程

6 月 6 日		
时间	**内容**	**地点**
08:40-09:10	开幕式 颁发赞助商感谢牌	四楼大湾区厅
09:10-09:40	大会主旨报告-1 报告人：白翔（华中科技大学） 报告题目：迈向通用文字识别：文档智能模型的进展与趋势	
09:40-09:50	铂金企业宣讲（华为技术有限公司） 宣讲人：徐晓刚（华为技术有限公司） 报告题目：华为在 CV 及媒体领域的进展及挑战	
09:50-10:00	铂金企业宣讲（AutoDL） 宣讲人：余佳（COO） 报告题目：破解用卡难复现难解决方案	
10:00-10:30	大会主旨报告-2 报告人：程明明（南开大学） 报告题目：高效视觉感知与个性化生成	
10:30-10:45	休息	
10:45-11:15	大会主旨报告-3 报告人：俞扬（南京大学） 报告题目：大模型背景下的强化学习	
11:15-11:30	2024-2025 年度 CV 与 ML 领域重要学术进展	
11:30-12:00	年度进展评述（一）	
	讲者：郝建业（天津大学） 题目：AI 智能体	
	讲者：仉尚航（北京大学） 题目：大模型复杂推理	
12:00-14:00	午餐	
14:00-14:30	大会主旨报告-4 报告人：孟德宇（西安交通大学） 报告题目：机器学习的“变”与“不变”	四楼大湾区厅
14:30-15:00	大会主旨报告-5 报告人：鲁继文（清华大学） 报告题目：视觉感知与自动驾驶	
15:00-15:20	休息	
15:20-18:00	年度进展评述（二）	

	讲者：傅朝友（南京大学） 题目：多模态基础模型	四楼大湾区厅
	讲者：袁粒（北京大学） 题目：视频生成基础模型	
	讲者：弋力（清华大学） 题目：具身智能	
	讲者：张兆翔（中国科学院自动化研究所） 题目：世界模型理论与框架	
	讲者：吕凯风（清华大学） 题目：大模型基础理论	
	讲者：虞晶怡（上海科技大学） 题目：3D 视觉基础模型	
	讲者：谢伟迪（上海交通大学） 题目：医疗通用大模型	
	讲者：邹征夏（北京航空航天大学） 题目：遥感通用大模型	
	讲者：顾实（浙江大学） 题目：脑启发与认知计算	
	讲者：叶茫（武汉大学） 题目：大模型安全与伦理	
18:40-21:00	VIP 交流活动	四楼大湾区厅 1
	6 月 7 日	
08:30-12:00	**Workshop 3：开放环境视觉理解与生成** **组织者**：陈使明（阿联酋人工智能大学(MBZUAI)）、谢国森（南京理工大学）、杨小汕（中国科学院自动化研究所） **讲者**：鲍秉坤（南京邮电大学）、彭玺（四川大学）、舒祥波（南京理工大学）、张正（哈尔滨工业大学（深圳））、谢伟迪（上海交通大学）、葛铁铮（阿里妈妈）、代季峰（集智进化）	四楼大湾区厅 1
	Workshop 4：低空智能感知与理解 **组织者**：丛润民（山东大学）、万人杰（香港浸会大学）、李锋（合肥工业大学） **讲者**：齐俊桐（上海大学）、白慧慧（北京交通大学）、朱鹏飞（天津大学）、钟平（国防科技大学）、邹征夏（北京航空航天大学）、赵健（中国电信集团）	四楼大湾区厅 2
	Workshop 10：基于生成式 AI 驱动的具身智能 **组织者**：汪婧雅（上海科技大学）、贾奎（香港中文大学（深圳））、胡瑞珍（深圳大学） **讲者**：盛律（北京航空航天大学），黄思远（北京通用人工智能研究院），张直政（银河通用），石野（上海科技大学），穆尧（上海交通大学）	四楼大湾区厅 3
	Workshop 13：大模型基础理论-从理论视角看待大模型技术发展 **组织者**：刘勇（中国人民大学）、李健（北京师范大学） **讲者**：车万翔（哈尔滨工业大学）、张奇（复旦大学）、龚铁梁	四楼大湾区厅 4

	（西安交通大学）、黄维然（上海交通大学/上海创智学院）、方聪（北京大学）、蒋宁（九章云极）	
	Workshop 12：AI4Science，诺贝尔奖后的思考 **组织者**：欧阳万里（上海人工智能实验室）、周浩（清华大学智能产业研究院）、周东展（上海人工智能实验室） **讲者**：邵斌（中关村人工智能研究院）、王笑楠（清华大学）、符天凡（南京大学）、朱霖潮（浙江大学）、岳翔宇（香港中文大学）、肖文之（字节跳动）、吴保东（无问芯穹）	四楼大湾区厅 5
	Tutorial 1：：大模型基础理论结构与微调 **讲者**：王立威（北京大学）、陈键飞（清华大学）	三楼报告厅
12:00-13:30	Poster/午餐	四楼大湾区厅 Poster 展区
13:30-17:00	**Workshop 7：多模态认知计算** **组织者**：胡迪（中国人民大学）、刘武（中国科学技术大学）、赵健（中国电信集团） **讲者**：谢洪涛（中国科学技术大学）、魏云超（北京交通大学）、梁小丹（中山大学）、沈为（上海交通大学）、王正（武汉大学）、徐文超（香港科技大学）、张勇（美团）	四楼大湾区厅 1
	Workshop 15：人体动作理解与生成 **组织者**：徐婧林（北京科技大学）、曾润浩（深圳北理莫斯科大学）、涂志刚（武汉大学） **讲者**：彭宇新（北京大学）、林巍峣（上海交通大学）、张姗姗（南京理工大学）、秦杰（南京航空航天大学）、涂志刚（武汉大学）	四楼大湾区厅 2
	Tutorial 2：科学智能（AI for Science） 讲者：欧阳万里（上海人工智能实验室）、孙浩（中国人民大学）	四楼大湾区厅 3
	Workshop 16：深度连续学习研讨会 **组织者**：刘夏雷（南开大学）、洪晓鹏（哈尔滨工业大学）、张幸幸（清华大学） **讲者**：邓成（西安电子科技大学）、周宇（南开大学）、叶翰嘉（南京大学）、余璐（天津理工大学）、朱飞（中国科学院香港创新研究院人工智能与机器人创新中心）、张幸幸（清华大学）	四楼大湾区厅 4
	Workshop 18：面向移动终端的 AI 图像增强 **组织者**：程明明（南开大学）、高广谓（南京邮电大学）、郭鑫（华为技术有限公司） **讲者**：郭鑫（华为技术有限公司）、任文琦（中山大学）、张宇伦（上海交通大学）、左旺孟（哈尔滨工业大学）、王诗淇（香港城市大学）、郭春乐（南开大学）、周尚辰（南洋理工大学）、郭丰俊（上海合合信息）	四楼大湾区厅 5
	Workshop 2：多模态学习助力智慧医疗：从科研到落地 **组织者**：谢雨彤（穆罕默德·本·扎耶德人工智能大学）、雷柏英（深圳大学）、夏勇（西北工业大学） **讲者**：沈定刚（上海科技大学生物医学工程学院、上海联影智能医疗科技有限公司）、王珊珊（中科院深圳先进院）、陈浩（香	三楼报告厅

	港科技大学）、隋尧（北京大学）、张建鹏（阿里巴巴达摩院、浙江大学）	
17:00-18:30	Poster/晚餐	四楼大湾区厅 Poster 展区
18:30-22:00	**Workshop 1：多模态遥感通用模型** **组织者**：赵文达（大连理工大学）、冯婕（西安电子科技大学）、洪丹枫（中国科学院空天信息创新研究院） **讲者**：夏桂松(武汉大学)、方乐缘(湖南大学)、张洪艳(中国地质大学)、郑钰辉(青海师范大学)、聂婕(中国海洋大学)、赵文达(大连理工大学)、仇明(阿里云)	四楼大湾区厅 1
	Workshop 9：生物特征识别遇上 AI+Science **组织者**：朱翔昱（中国科学院自动化研究所）、雷震（中国科学院自动化研究所）、何晖光（中国科学院自动化研究所） **讲者**：刘宁（中科院生物物理研究所）、常乐（中科院脑科学与智能技术卓越创新中心，神经科学研究所）、李晓白（浙江大学）、李婧婷（中国科学院心理研究所）、朱翔昱（中国科学院自动化所）	四楼大湾区厅 2
	Workshop 11：类脑成像：当事件相机遇到 AI **组织者**：吴金建（西安电子科技大学）、余肇飞（北京大学）、杨文（西安电子科技大学） **讲者**：戴玉超（西北工业大学）、侯军辉（香港城市大学）、郑乾（浙江大学）、周易（湖南大学）、田永鸿（北京大学）	四楼大湾区厅 3
	Workshop 17：人工智能大模型安全 **组织者**：彭春蕾（西安电子科技大学）、刘晗（大连理工大学）、刘艾杉（北京航空航天大学） **讲者**：韦星星（北京航空航天大学）、蔺琛皓（西安交通大学）、郭青（南开大学）、吴海威（电子科技大学）、周文柏（中国科学技术大学）、董胤蓬（清华大学）、王珂尧（百度）	四楼大湾区厅 4
	Workshop 19：空间智能的感知与决策 **组织者**：苏航（清华大学）、马月昕（上海科技大学）、赵恒爽（香港大学） **讲者**：蒋树强（中国科学院大学）、杨睿刚（上海交通大学）、Jianlan Luo（UC Berkeley）、王越（浙江大学）、刘希慧（香港大学）、苏航（清华大学）	四楼大湾区厅 5
	Tutorial 3：深度学习网络架构设计 讲者：黄高（清华大学）、谢琦（西安交通大学）	三楼报告厅
6 月 8 日		
08:30-12:00	**Workshop 6：视觉通用模型** **组织者**：王立君（大连理工大学）、李冠彬（中山大学）、黄岩（中科院自动化所） **讲者**：王兴刚（华中科技大学）、董超（中国科学院深圳先进技	四楼大湾区厅 1

	术研究院）、代季峰（清华大学）、王栋（大连理工大学）、陈隆（香港科技大学）、张严浩（OPPO）	
	Workshop 5：大场景多对象的重建与生成 **组织者**：李坤（天津大学）、季梦奇（北京航空航天大学）、马健（天津大学） **讲者**：王程（厦门大学）、刘子纬（南洋理工大学）、马楠（北京工业大学）、李坤（天津大学）、马月昕（上海科技大学）、温建伟（北京拙河科技有限公司）	四楼大湾区厅 2
	Tutorial 4：可信机器学习 **讲者**：韩波（香港浸会大学）、姚江超（上海交通大学）、张杰（中国科学院计算技术研究所）	四楼大湾区厅 3
	Workshop 14：自动驾驶多模态世界模型 **组织者**：李镇（香港中文大学（深圳））、贾旭（大连理工大学）、贾伟（合肥工业大学） **讲者**：李弘扬（香港大学）、马超（上海交通大学）、王新宇（华为技术有限公司）、王乃岩（小米）、赵昊（清华大学）、高晋（中国科学院大学）、范略（中国科学院自动化研究所）	四楼大湾区厅 4
	Workshop 20：优秀学生论坛 **组织者**：韩晓光（香港中文大学（深圳））、李永露（上海交通大学）、于茜（北京航空航天大学）、胡迪（人民大学） **讲者**：陈科研（北京航空航天大学）、刘世隆（清华大学）、卫雅珂（中国人民大学）、黄文柯（武汉大学）、刘欣鹏（上海交通大学）、王健伊（新加坡南洋理工大学）、倪旻恒（香港理工大学）、窦志扬（香港大学）	四楼大湾区厅 5
	Workshop 8：语言-视觉协同生成 **组织者**：杨易（浙江大学）、武宇（武汉大学）、刘希慧（香港大学） **讲者**：王亚星（南开大学）、饶安逸（香港科技大学）、王文海（上海人工智能实验室）、杨思蓓（中山大学）、李崇轩（中国人民大学）、曹浩宇（腾讯优图）	三楼报告厅

Tutorial 与 Workshop 日程一览

时间地点	报告与讲者信息
Tutorial 1: **大模型基础理论结构与微调** 6 月 7 日 08:30-12:00 三楼报告厅	报告题目：大模型时代机器学习理论 讲者：王立威（北京大学）
	报告题目：基于量化稀疏的高效训练推理：理论及算法 讲者：陈键飞（清华大学）
Tutorial 2: **科学智能（AI for Science）** 6 月 7 日 13:30-17:00 四楼大湾区厅 3	报告题目：AI for Science-机遇与挑战 讲者：欧阳万里（上海人工智能实验室）
	报告题目：复杂时空动力系统智能科学计算 讲者：孙浩（中国人民大学）
Tutorial 3: **深度学习网络架构设计** 6 月 7 日 18:30-21:30 三楼报告厅	报告题目：高效视觉基础模型架构设计 讲者：黄高（清华大学）
	报告题目：深度网络的基础模块设计：数据先验与架构融合 讲者：谢琦（西安交通大学）
Tutorial 4: **可信机器学习** 6 月 8 日 08:30-12:00 四楼大湾区厅 3	报告题目：基于噪声数据的可信机器学习 讲者：韩波（香港浸会大学）
	报告题目：基于不均衡数据的可信机器学习 讲者：姚江超（上海交通大学）
	报告题目：对抗场景下的可信机器学习 讲者：张杰（中国科学院计算技术研究所）
Workshop 3: **开放视觉环境理解与生成** 6 月 7 日 08:30-12:00 四楼大湾区厅 1	组织者：陈使明（阿联酋人工智能大学(MBZUAI)）、谢国森（南京理工大学）、杨小汕（中国科学院自动化研究所）
	讲者：鲍秉坤（南京邮电大学） 题目：记忆机制启发的视频行为理解
	讲者：彭玺（四川大学） 题目：噪声关联学习
	企业宣讲：阿里妈妈（葛铁铮） 题目：计算机视觉 in 阿里妈妈
	讲者：舒祥波（南京理工大学） 题目：开放场景下人体行为智能计算

时间地点	报告与讲者信息
	讲者：张正（哈尔滨工业大学） 题目：高效能跨模态关联分析
	企业宣讲：集智进化（代季峰） 题目：集智进化宣讲
	讲者：谢伟迪（上海交通大学） 题目：生成式视觉-语言模型研究
Workshop 4: 低空智能感知与理解 6 月 7 日 08:30-12:00 四楼大湾区厅 2	组织者：丛润民（山东大学）、万人杰（香港浸会大学）、李锋（合肥工业大学）
	讲者：齐俊桐（上海大学） 题目：无人机集群感知控制技术与应用
	讲者：白慧慧（北京交通大学） 题目：无人机图像智能感知方法研究
	讲者：朱鹏飞（天津大学） 题目：低空智能感知关键技术及应用
	讲者：钟平（国防科技大学） 题目：无人机对地遥感目标识别与对抗
	讲者：邹征夏（北京航空航天大学） 题目：三维遥感场景感知与新视角合成研究进展
	讲者：赵健（中国电信集团） 题目：Anti-UAV：复杂场景低慢小目标智能感知
Workshop 10：基于生成式 AI 驱动的具身智能 6 月 7 日 08:30-12:00 四楼大湾区厅 3	组织者：汪婧雅（上海科技大学）、贾奎（香港中文大学）、胡瑞珍（深圳大学）
	讲者：盛律（北京航空航天大学） 题目：面向具身智能的单视图三维内容生成：从物体到场景的生成路径
	讲者：黄思远（北京通用人工智能研究院） 题目：Empower Generalist Robot with Human-like Interactions: From Humans to Humanoids
	讲者：张直政(银河通用) 题目：合成数据开启端到端具身大模型训练新范式
	企业宣讲：银河通用（张直政） 题目：银河通用具身大模型机器人的跨行业应用
	讲者：石野（上海科技大学） 题目：扩散模型驱动的具身智能：理论与算法前沿突破
	讲者：穆尧（上海交通大学） 题目：从多模态认知到具身执行：大规模具身数据自动生成与具身大模型训练

时间地点	报告与讲者信息
Workshop 13: 大模型基础理论-从理论视角看待大模型技术发展 6 月 7 日 08:30-12:00 四楼大湾区厅 4	组织者：刘勇（中国人民大学）、李健（北京师范大学）
	讲者：车万翔（哈尔滨工业大学） 题目：迈向推理时代——长思维链机理及应用
	讲者：张奇（复旦大学） 题目：大语言模型能力来源与边界
	企业宣讲：九章云极（蒋宁） 题目：高性能高弹性普惠算力加速大模型科研
	讲者：龚铁梁（西安交通大学） 题目：领域泛化的信息论协同机制
	讲者：黄维然（上海交通大学/上海创智学院） 题目：矩阵信息论视角下的表征学习
	讲者：方聪（北京大学） 题目：随机梯度下降算法在高维回归问题中正则效应与泛化性能分析
Workshop 12：AI4Science，诺贝尔奖后的思考 6 月 7 日 08:30-12:00 四楼大湾区厅 5	组织者：欧阳万里（上海人工智能实验室）、周浩（清华大学智能产业研究院）、周东展（上海人工智能实验室）
	讲者：邵斌（中关村人工智能研究院） 题目：Toward Protein Dynamics Simulations with Ab Initio Accuracy
	讲者：王笑楠（清华大学） 题目：AI+能源化工新材料:大小通专模型并进
	企业宣讲：无问芯穹（吴保东） 题目：面向大模型科学研究的算力平权技术探索
	讲者：肖文之（字节跳动） 题目：Protenix—生物分子复合物结构预测与设计
	讲者：符天凡（南京大学） 题目：深度学习赋能的药物发现与开发
	讲者：朱霖潮（浙江大学） 题目：AI 赋能科学智能仿真和设计
	讲者：岳翔宇（香港中文大学） 题目：从 AI 到 AI4Science：我们的实践与探索
	组织者： 胡迪（中国人民大学）、刘武（中国科学技术大学）、赵健（中国电信集团）
	讲者：谢洪涛（中国科学技术大学） 题目：大模型时代下的场景文本检测、识别与编辑
	讲者：魏云超（北京交通大学） 题目：视觉智能推理

时间地点	报告与讲者信息
Workshop 7：多模态认知计算 6 月 7 日 13:30-17:00 四楼大湾区厅 1	企业宣讲：美团（张勇） 题目：本地生活场景中视觉生成的研究与应用
	讲者：梁小丹（中山大学） 题目：“慢思考”在多模态基础模型与具身智能中的探索
	讲者：沈为（上海交通大学） 题目：视觉语言的细粒度对齐
	讲者：王正（武汉大学） 题目：视觉隐匿感知与对抗
	讲者：徐文超（香港科技大学） 题目：多模态学习中的模态竞争及解决方案
Workshop 15：人体动作理解与生成 6 月 7 日 13:30-17:00 四楼大湾区厅 2	组织者：徐婧林（北京科技大学）、曾润浩（深圳北理莫斯科大学）、涂志刚（武汉大学）
	讲者：彭宇新（北京大学） 题目：基于多模态大模型的视觉内容理解与生成
	讲者：林巍峣（上海交通大学） 题目：语义驱动的视频事件内容感知推理与编码
	讲者：张姗姗（南京理工大学） 题目：知识驱动的人-物体交互理解与生成
	讲者：秦杰（南京航空航天大学） 题目：“分久必合”：视频动作检测-分割-预测一体化方法研究
	讲者：涂志刚（武汉大学） 题目：“以人为中心”视频行为识别与生成
Workshop 16: 深度连续学习研讨会 6 月 7 日 13:30-17:00 四楼大湾区厅 4	组织者：刘夏雷（南开大学）洪晓鹏（哈尔滨工业大学）张幸幸（清华大学）
	讲者：邓成（西安电子科技大学） 题目：基于脑启发的深度增量学习
	讲者：周宇（南开大学） 题目：增量目标检测技术研究
	讲者：叶翰嘉（南京大学） 题目：基于模型兼容的增量学习方法探究
	讲者：余璐（天津理工大学） 题目：持续学习中的少遗忘、强鲁棒和可解释研究
	讲者：朱飞（中国科学院香港创新研究院人工智能与机器人创新中心） 题目：多模态持续学习
	讲者：张幸幸（清华大学） 题目：预训练下的持续学习理论、方法与应用

时间地点	报告与讲者信息
Workshop 18：面向移动终端的 AI 图像增强 6 月 7 日 13:30-17:00 四楼大湾区厅 5	组织者：程明明（南开大学）、高广谓（南京邮电大学）、郭鑫（华为技术有限公司）
	讲者：郭鑫（华为技术有限公司） 题目：移动终端影像算法中的业务难题
	讲者：任文琦（中山大学） 题目：扩散模型驱动图像超分辨技术探索
	讲者：张宇伦（上海交通大学） 题目：图像重建轻量化
	企业宣讲：上海合合信息（郭丰俊） 题目：AI 浪潮下的图像内容安全
	讲者：左旺孟（哈尔滨工业大学） 题目：人脸和文本图像复原方法研究
	讲者：王诗淇（香港城市大学） 题目：提升低光图像质量：从客观质量到主观感知的转变
	讲者：郭春乐（南开大学） 题目：面向移动终端的影像 AI 娱乐化应用
	讲者：周尚辰（南洋理工大学） 题目：视觉内容补全与提取
Workshop 2：多模态学习助力智慧医疗：从科研到落地 6 月 7 日 13:30-17:00 三楼报告厅	组织者：谢雨彤（穆罕默德·本·扎耶德人工智能大学）、雷柏英（深圳大学）、夏勇（西北工业大学）
	讲者：沈定刚（上海科技大学） 题目：生成式人工智能与疾病诊疗
	讲者：王珊珊（中国科学院深圳先进技术研究院） 题目：多模态可泛化医学影像人工智能基础模型研究
	讲者：陈浩（香港科技大学） 题目：大模型赋能智慧医疗：挑战和未来
	讲者：隋尧（北京大学） 题目：神经成像质量与效率：面向临床的优化
	讲者：张亚琴（中山大学附属第五医院） 题目：人工智能助力乳腺疾病影像解读
	讲者：张建鹏（阿里巴巴达摩院、浙江大学） 题目：基于视觉语言预训练的开放场景医学影像理解
	组织者：赵文达（大连理工大学）、冯婕（西安电子科技大学）、洪丹枫（中国科学院空天信息创新研究院）

时间地点	报告与讲者信息
Workshop 1：多模态遥感通用模型 6 月 7 日 18:30-22:00 四楼大湾区厅 1	讲者：夏桂松（武汉大学） 题目：面向开放场景的多源遥感图像变化检测
	讲者：方乐缘（湖南大学） 题目：多模态遥感通用大模型研究初探
	企业宣讲：阿里云（仇明） 题目：通义大模型在科研领域的应用与实践
	讲者：张洪艳（中国地质大学） 题目：大范围高分辨率土地覆盖遥感制图
	讲者：郑钰辉（青海师范大学） 题目：高光谱遥感图像智能解译与多模态通用大模型的思考
	讲者：聂婕（中国海洋大学） 题目：面向海洋科学遥感的多模态智能计算模型
	讲者：赵文达（大连理工大学） 题目：非完备图像信息目标智能感知
Workshop 9：生物特征识别遇上 AI+Science 6 月 7 日 18:30-22:00 四楼大湾区厅 2	**组织者**：朱翔昱（中科院自动化研究所）、雷震（中科院自动化研究所）、何晖光（中科院自动化研究所）
	讲者：刘宁（中科院生物物理所） 题目：利用深度卷积神经网络理解熟悉面孔识别的神经机制
	讲者：常乐（中科院神经科学研究所） 题目：灵长类大脑对面孔信息的神经编码机制
	讲者：李晓白（浙江大学） 题目：远程脉搏波测量及其在生物识别领域的应用
	讲者：李婧婷（中科院心理所） 题目：心理学实验驱动的微表情分析与情感理解
	讲者：朱翔昱（中科院自动化所） 题目： 马尔视觉理论研究：2D-2.5D-3D 表征的涌现
Workshop 11：类脑成像：当事件相机遇到 AI 6 月 7 日 18:30-22:00 四楼大湾区厅 3	组织者：吴金建（西安电子科技大学）、余肇飞（北京大学）、杨文（西安电子科技大学）
	讲者：戴玉超（西北工业大学） 题目：事件相机视觉：从运动感知与影像生成
	讲者：侯军辉（香港城市大学） 题目：Event Tracker：Shifting Object Tracking into Temporal-Continous Domain
	讲者：郑乾（浙江大学） 题目：见微知著：基于事件信号的物理过程建模与环境感知方法

时间地点	报告与讲者信息
	讲者：周易（湖南大学） 题目：Enabling Faster and Safer Robots with Neuromorphic Event-based Vision
	讲者：田永鸿（北京大学） 题目：神经形态智能感知计算理论与方法
Workshop 17：人工智能大模型安全 6 月 7 日 18:30-22:00 四楼大湾区厅 4	组织者：彭春蕾（西安电子科技大学）、刘晗（大连理工大学）、刘艾杉（北京航空航天大学）
	讲者：韦星星（北京航空航天大学） 题目：多模态大模型可信度评估及增强
	讲者：蔺琛皓（西安交通大学） 题目：人工智能大模型安全的完整与可用
	企业宣讲：百度（王珂尧） 题目：面向可解释性的多模态人脸活体检测方法研究
	讲者：郭青（南开大学） 题目：面向多模态模型的攻击与防御方法研究
	讲者：吴海威（电子科技大学） 题目：信息泡沫时代如何练就火眼金睛
	讲者：周文柏（中国科学技术大学） 题目：大模型生成内容安全：从文本到多模态生成内容的安全防护体系
	讲者：董胤蓬（清华大学） 题目：面向多模态大模型的安全性与可信性
Workshop 19: 空间智能的感知与决策 6 月 7 日 18:30-22:00 四楼大湾区厅 5	组织者：苏航（清华大学）、马月昕（上海科技大学）、赵恒爽（香港大学）
	讲者：蒋树强（中国科学院大学） 题目：基于具身场景记忆的视觉导航
	讲者：杨睿刚（上海交通大学） 题目：Simulating the World from the World
	讲者：Jianlan Luo （UC Berkeley） 题目：Intelligent Physical Agents: High-Performance Learning for Generalist Robots
	讲者：王越（浙江大学） 题目：从视觉反馈控制到 VLA
	讲者：刘希慧（香港大学） 题目：3D Object Parsing: From Surface Segmentation to Generative Amodal Segmentation

时间地点	报告与讲者信息
	讲者：苏航（清华大学） 题目：跨越虚实鸿沟：基座模型驱动的具身智能泛化之路
Workshop 6：视觉通用模型 6 月 8 日 08:30-12:00 四楼大湾区厅 1	组织者： 王立君（大连理工大学）、李冠彬（中山大学）、黄岩（中科院自动化所）
	讲者：王兴刚（华中科技大学） 题目：重建 vs 生成 - 解决潜在扩散模型中的优化困境
	讲者：董超（中国科学院深圳先进技术研究院） 题目：通用底层视觉前沿探索
	企业宣讲：OPPO 题目：面对 AI 手机的多模态生成与问答探索实践
	讲者：代季峰（清华大学） 题目：多模态基础模型研究
	讲者：王栋（大连理工大学） 题目：通用视觉感知模型初探
	讲者：陈隆（香港科技大学） 题目：Narrowing the Gaps: Towards Real-World Multimodal Reasoning & Generation Models
Workshop 5：大场景多对象的重建与生成 6 月 8 日 08:30-12:00 四楼大湾区厅 2	组织者：李坤（天津大学）、季梦奇（北京航空航天大学）、马健（天津大学）
	讲者：王程（厦门大学） 题目：城市空间中激光雷达感知定位与人身份识别
	讲者：刘子纬（南洋理工大学） 题目：From Multimodal Generative Models to Dynamic World Modeling
	讲者：马楠（北京工业大学） 题目：大场景下的移动智能体多对象目标检测与配准定位技术
	讲者：李坤（天津大学） 题目：大场景下的群体三维重建与生成
	讲者：马月昕（上海科技大学） 题目：面向人-机-环境协同共生的感知与生成
	讲者：温建伟（北京拙河科技有限公司） 题目：亿像素光场重建与智能分析在产业界的应用探索
	组织者：李镇（香港中文大学 (深圳)）、贾伟（合肥工业大学）、贾旭（大连理工大学）
	讲者：李弘扬（香港大学） 题目：迈向具备高泛化性的通用自动驾驶世界模型

时间地点	报告与讲者信息
Workshop 14：自动驾驶多模态世界模型 6 月 8 日 08:30-12:00 四楼大湾区厅 4	讲者：马超（上海交通大学） 题目：自动驾驶多模态场景理解与生成
	讲者：王新宇（华为技术有限公司） 题目：面向未来智能辅助驾驶的 WEWA 架构
	讲者：王乃岩（小米） 题目：什么是适合物理世界 AI 的表征学习？
	企业宣讲：小米（王乃岩） 题目：走进小米汽车
	讲者：赵昊（清华大学） 题目：可控高效的时空世界模型在自动驾驶中的应用
	讲者：高晋（中国科学院大学） 题目：基于扩散模型的时空一致 4D 内容生成初探
	讲者：范略（中国科学院自动化研究所） 题目：生成和重建一体化的自动驾驶世界模型
	Panel 嘉宾：金鑫（宁波东方理工大学）、李弘扬（香港大学）、马超（上海交通大学）、王新宇（华为）、王乃岩（小米）、赵昊（清华大学）、高晋（中国科学院大学）、范略（中国科学院自动化研究所）
Workshop 20：优秀学生论坛 6 月 8 日 08:30-12:00 四楼大湾区厅 5	组织者：韩晓光（香港中文大学（深圳））、李永露（上海交通大学）、于茜（北京航空航天大学）、胡迪（中国人民大学）
	讲者：陈科研（北京航空航天大学） 题目：遥感图像高效视觉基础模型研究
	讲者：刘世隆（清华大学） 题目：LLaVA-Plus：学习使用工具创建多模态智能体
	讲者：卫雅珂（中国人民大学） 题目：平衡多模态学习
	讲者：黄文柯（武汉大学） 题目：多模态大模型的领域微调
	讲者：刘欣鹏（上海交通大学） 题目：Something out of Nothing for Human Motion Understanding
	讲者：王健伊（新加坡南洋理工大学） 题目：视频超分大模型
	讲者：倪旻恒（香港理工大学） 题目：基于多模态认知的视觉推理与生成
	讲者：窦志扬（香港大学） 题目：From Static 3D Geometry to Dynamic 4D Content: Analysis, Recovery, and Generation

时间地点	报告与讲者信息
Workshop 8：语言-视觉协同生成 6 月 8 日 08:30-12:00 三楼报告厅	组织者：杨易（浙江大学）、武宇（武汉大学）、刘希慧（香港大学）
	讲者：王亚星（南开大学） 题目：文生图模型中文本和图像表征的思考
	讲者：饶安逸（香港科技大学） 题目：Bridging the Representation Gap between Humans and Computers for Video Production
	企业宣讲：腾讯优图（曹浩宇） 题目：迈向 GPT-4O 级实时音视频交互：全模态大模型的进展与趋势
	讲者：王文海（上海人工智能实验室） 题目：书生·万象多模态大模型的技术演进与应用探索
	讲者：杨思蓓（中山大学） 题目：面向语言-视觉智能的基础表征、跨模态生成与主动交互
	讲者：李崇轩（中国人民大学） 题目：LLaDA：大语言模型新范式

大会报告及讲者简介

白翔　华中科技大学

报告题目：迈向通用文字识别：文档智能模型的进展与趋势

报告摘要：在大模型时代，文字识别技术已经取得了显著的进步，展示了实现通用 OCR 的潜力。在本次报告中，首先我将全面分析大模型在 OCR 识别方面的表现；接着，我将介绍团队在多任务统一的文字识别方法，面向文档智能理解的多模态大模型，大模型智能文档推理等技术进展；最后，我将对文档智能的发展趋势进行展望。

讲者简介：华中科技大学教授、博导，国家杰出青年基金获得者，IAPR Fellow，国际期刊 Pattern Recognition 副主编（A-EIC）。主要从事计算机视觉与模式识别、多模态大模型等方面研究，在 Nature Machine Intell.、IEEE TPAMI、CVPR 等国际一流期刊和国际会议发表论文 150 余篇。担任国际顶级期刊 IEEE TPAMI 编委，顶级会议 CVPR、ICCV、ECCV、AAAI、IJCAI、NeurIPS 的领域主席，国际文档分析与识别会议 ICDAR 2025 大会主席。曾获 ACL 2024 最佳论文奖（Best Paper Award）、2024 年湖北省青年科技创新奖、2023 年湖北省自然科学一等奖（排 1）、2021 年全国科技系统抗击新冠疫情先进个人、2021 年中国图象图形学学会自然科学一等奖（排 1）、2019 年国际模式识别协会青年学者奖（IAPR/ICDAR Young Investigator Award）。他是视觉与学习青年研讨会（VALSE）的指导委员会成员，VALSE 在线学术报告会（VALSE Webinar）活动的共同发起人。

程明明　南开大学

报告题目：高效视觉感知与个性化生成

报告摘要：高效的视觉感知与内容生成算法能够显著提升目标检测、场景理解与动态变化的识别精度与处理速度，为智能系统提供更加可靠的实时决策支持。本报告将从多粒度特征提取、伪装目标检测、高效生成模型训练、个性化生成等角度介绍高效视觉感知与个性化内容生成的进展。报告中所涉及的技术都立足于国产芯片与深度学习框架进行了开源。此外，报告还将针对当前感知与生成模型面临的关键技术挑战展开分析，并对未来技术发展趋势与应用前景进行展望。

讲者简介：南开大学二级教授，媒体计算团队学术带头人。主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是人工智能、计算机视觉和计算机图形学，在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇（含 IEEE TPAMI 论文 40 余篇），h-index 为 100，论文谷歌引用 6 万余次，单篇最高引用 5 千余次，多次入选全球高被引科学家和中国高被引学者。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得教育部自然科学一等奖 2 项、其他省部级科技奖 2 项。培养的 4 名博士生获得省部级优秀博士论文奖。现担任天津市视觉计算与智能感知重点实验室主任、中国图象图形学学会副秘书长、天津市人工智能学会副理事长和顶级期刊 IEEE TPAMI,

IEEE TIP 和《中国科学：信息科学》编委。

俞扬　南京大学

报告题目：大模型背景下的强化学习

报告摘要：2024 年图灵奖授予研究强化学习的先驱。强化学习已从早期游戏任务扩展到机器人控制等复杂物理环境中的应用。本次报告将回顾强化学习技术发展历史，并汇报在大模型和具身智能受到高度关注的背景下，强化学习技术的发展与变化，包括面向大语言模型的强化学习、借助大模型增强强化学习通用性、面向具身智能的决策等方面的发展趋势。

讲者简介：俞扬，南京大学人工智能学院教授。主要从事人工智能、机器学习、强化学习方向的研究，工作获 5 项国际论文奖、3 项国际算法竞赛冠军。入选国家青年人才计划、IEEE Intelligent Systems “AI’s 10 to Watch”，获 CCF-IEEE 青年科学家奖，首届亚太数据挖掘“青年成就奖”，并受邀在国际人工智能联合大会 IJCAI 2018 上作“青年亮点报告”。

孟德宇　西安交通大学

报告题目：机器学习的“变”与“不变”

报告摘要：在深度学习快速迭代的浪潮中，前沿研究聚焦于从“变”的角度构建机器学习方法，如增加数据/标记规模、设计创新网络架构、构建多样学习模式等。然而，从机器学习的基础研究视角，我们可发现机器学习的各个环节中存在更为本质的”不变性“规律与内涵，如数据高维标记空间的低维特征模式、网络基础模块的不变/等变性结构本质、学习模式设计的内在统一性规律等。把握这些不变性内涵，有利于我们更深刻理解提升机器学习泛化性、鲁棒性、可解释性机理的理论方法途径，更合理利用这些基础原理设计更加简洁合理的学习模式，构建更加具有深刻内涵的机理-数据双驱动、知识-网络相融合的有效机器学习方法。基于此，本报告将介绍针对高维标记空间低维不变性隐空间提炼的“标记分布建模”理论与方法、针对网络基础卷积模块旋转-尺度-仿射等变性结构刻画的“参数化卷积”理论与方法、针对机器学习方法超参设置不变性规律提炼的“模拟学习方法论”理论与方法，从而尝试探讨对机器学习方法如何从“变”中提炼其“不变”内涵的方法论思想，为机器学习的基础研究与工程应用提供一种可参考的视角。

讲者简介：孟德宇，西安交通大学数学与统计学院教授，任大数据分析与计算分析工程实验室统计与大数据中心常务副主任。长期致力于机器学习基础理论与方法的研究，在机器学习相关领域期刊会议发表论文百余篇，谷歌学术引用超过 35000 次。现任 IEEE Trans. PAMI，National Science Review 等 7 个国内外期刊编委。

鲁继文　清华大学

报告题目：视觉感知与自动驾驶

报告摘要：自动驾驶是人工智能与机器人领域的研究热点，在工业制造、农业生产、交通运输、现代服务等领域有着重要应用前景。报告将介绍自动驾驶视觉感知近年来的主要研究进展，包括视觉场景生成、三维占据预测、端到端自动驾驶、自动驾驶大模型等方法与技术，同时深入分析其优缺点与应用潜能，最后将对自动驾驶视觉感知的未来发展趋势进行展望。

讲者简介：清华大学长聘教授，自动化系副主任，全国重点实验室副主任，国家杰出青年科学基金获得者，IEEE/IAPR Fellow。主要研究计算机视觉、模式识别、具身智能、人工智能安全，发表 PAMI、IJCV、CVPR、ICCV、ECCV 论文 200 余篇，获授权国家发明专利 60 余项，主持国家自然科学基金重点基金 3 项，国家重点研发计划项目 1 项，北京市重点项目 2 项，获国家级教学成果奖二等奖 1 项，公安部科学技术奖一等奖 1 项，中国电子学会自然科学奖一等奖 2 项。担任中国仿真学会理事、视觉计算与仿真专业委员会主任，中国自动化学会专家咨询工作委员会副主任，国际期刊 Pattern Recognition Letters 主编，IEEE T-IP/T-MM/T-CSVT/T-BIOM 编委，培养 7 名博士生获北京市和全国一级学会优秀博士学位论文。

年度进展评述（APR）及讲者简介

郝建业　天津大学

报告题目：AI 智能体

报告摘要：本报告首先会介绍传统强化学习背景和基础，然后介绍在大模型时代下新的决策模型学习范式，以及强化学习如何助力决策模型及其所面临的挑战和最新技术进展，同时介绍在具身智能和 agent 等方向最新进展。

讲者简介：天津大学智算学部教授，华为诺亚决策推理实验室主任。主要研究方向为强化学习、具身智能和多智能体系统。发表人工智能领域 CCF-A 类国际会议和期刊论文 100 余篇，专著 2 部。获得国家自然科学基金委优青、国家科技部 2030 人工智能重大课题、基金委人工智能重大培育等项目资助 10 余项，研究成果获国际会议最佳论文奖 3 次，NeurIPS 20-22 大会竞赛冠军 4 次。相关成果在国产工业基础软件智能化、自动驾驶、游戏 AI、互联网广告及推荐、5G 网络优化、工业物流调度等领域落地应用。

仉尚航　北京大学

报告题目：多模态大模型复杂推理

报告摘要：本报告将概述大模型复杂性推理的重要性及其研究背景。总结近一年大模型复杂推理的基础进展与主要技术路线，并介绍多模态融合与复杂推理能力提升的关键探索与最新成果，同时列举多模态大模型应用于复杂推理任务的具体案例。最后，分析多模态大模型在复杂性推理的未来趋势、潜力与挑战。

讲者简介：仉尚航，北京大学计算机学院研究员、助理教授、博士生导师、博雅青年学者。致力于具身智能多模态大模型方向的研究，在人工智能顶级期刊和会议上发表论文 120 余篇，Google Scholar 引用数 1.5 万余次，荣获世界人工智能顶级会议 AAAI 最佳论文奖。由 Springer Nature 出版《Deep Reinforcement Learning》，至今电子版全球下载量近三十万次，入选中国作者年度高影响力研究精选。入选美国“EECS Rising Star”、“全球 AI 华人女性青年学者榜”、“中国科协青年百人会”、“AI100 青年先锋”（麻省理工科技评论）。曾获多项国际竞赛前三名，中关村仿生机器人大赛优胜奖，入选 2024 具身智能科技前沿热点工作。曾多次在国际顶级会议 NeurIPS、ICML 上组织 Workshop，担任 AAAI 2022&2023&2024 高级程序委员。博士毕业于美国卡内基梅隆大学，并于加州大学伯克利分校从事博士后研究。

傅朝友　南京大学

报告题目：多模态大模型

报告摘要：多模态大模型是一类能够同时处理文本、图像、音频等多种模态信息，并进行统一理解与生成的人工智能模型。本报告将简要回顾过去一年该领域的技术进展与代表性成果，涵盖模型架构演进、跨模态能力提升、开源生态动向及典型应用实践等，并结合当前发展趋势，探讨未来面临的挑战与潜在机遇。

讲者简介：南京大学智能科学与技术学院研究员、助理教授、博导，入选中国科协青年人才托举工程。2022 年博士毕业于中国科学院自动化研究所；2022-2024 年通过腾讯“天才少年”计划加入优图实验室担任高级研究员，作为 Technology&Project Leader 从事学术研究和工程落地工作；2024 年 8 月加入南京大学。研究方向为多模态智能，已发表国际期刊、会议论文 20 余篇，谷歌学术引用 3000 余次，作为 Owner 的 GitHub 开源项目累计获得 1.8 万余次 Stars。代表性工作包括 VITA 多模态大模型系列（一作 VITA-1.0&-1.5，通讯 Long-VITA，2.4 千 GitHub Stars），MME 多模态评测基准系列（一作 MME&Video-MME，引用千余次）和 Awesome-MLLM 多模态社区（Owner，1 万余次 GitHub Stars）等。曾获中科院院长特别奖、IEEE Biometrics Council 最佳博士学位论文、北京市优秀博士学位论文、中科院优秀博士学位论文、小米青年学者-科技创新奖、CVPR 杰出审稿人等。

袁粒　北京大学

报告题目：视频生成基础模型

报告摘要：视频生成技术自 2024 年初进入爆发期，以 Sora 为代表的视频生成模型以 DiT 路线为核心首先破圈引发关注。随后闭源模型如 Pika、可灵、Vidu、PixelVerse、Runway 等模型陆续在商业化上取得进展，开源模型包括混元、CogVideo、Open-Sora Plan 等模型陆续推出。但是视频生成技术仍然面临可控性、内容一致性与资源消耗等问题，因此其应用爆发仍未到达高峰。该年度进展报告将阐述这些挑战，并详细介绍视觉生成领域的新模型和新技术，展望该领域未来的发展方向。

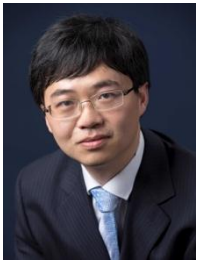
讲者简介：北京大学深圳研究生院助理教授、博士生导师、入选国家高层次青年人才计划、国家优秀留学生奖(归国类)、2023 年福布斯亚洲 30U30 名单等，主持国家科技创新 2030 重大项目课题和国自然基金等。 研究方向为以视觉为中心的多模态学习，以第一/通讯作者在国际期刊和顶会上发表论文 40 余篇，包括 Nature Computational Science、IEEE TPAMI/CVPR 等，代表性一作论文单篇被引数千余次，论文总引一万余次，代表性应用工作包括 ChatExcel、Open-Sora Plan 等。

弋力　清华大学

报告题目：具身智能

报告摘要：2024 年度具身智能快速发展，大模型与机器人融合深入，视觉-语言-动作模型突破。人形机器人运动控制技术显著提升。自主泛化、自我纠正能力增强，基础模型研发加速。具身智能仿真平台和数据集建设加快推进。本次报告讲究以上进展进行梳理讨论。

讲者简介：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，导师为美国三院院士 Leonidas J. Guibas 教授，毕业后在谷歌研究院任研究科学家。他近期的研究聚焦于三维视觉与具身智能，他的研究目标是赋予机器人理解并与三维世界交互的能力。他在计算机顶级会议期刊上已发表论文七十余篇，引用数两万余次，代表作品包括 ShapeNet Part，SyncSpecCNN，PointNet++等，大大影响了三维深度学习这一领域的出现与发展。此外他还曾担任 CVPR、IJCAI、NeurIPS 等顶会的领域主席与 SIGGRAPH TPC 等。



张兆翔　中国科学院自动化研究所

报告题目：世界模型理论与框架

报告摘要：世界模型作为理解和预测环境动态变化的核心技术，被视为是实现通用人工智能的关键。本报告系统回顾了世界模型的最新研究进展及其在多个领域的应用。近年来，世界模型在通用视频、自动驾驶和机器人领域取得了显著进展，展现出广阔的应用前景。在通用视频领域，世界模型开启了视频生成在交互可控性和物理真实性的新纪元。在自动驾驶领域，世界模型能够根据当前行驶场景预测未来可能出现的情况，帮助车辆提前做出决策，提高行驶安全性和效率。在机器人领域，世界模型为机器人提供安全的虚拟训练环境，使其能在部署前进行大量模拟训练和测试。展望未来，世界模型有望加速通用人工智能走向物理世界，推动具身智能体从“观察世界”走向“理解世界、推演世界、影响世界”的智能时代。

讲者简介：张兆翔，中国科学院自动化研究所多模态人工智能系统重点实验室研究员、博士生导师，模式识别实验室常务副主任，中国科学院大学岗位教授，IAPR Fellow，入选“教育部长江学者”和“国家万人计划青年拔尖人才”。研究方向是模式识别、具身智能、智能体学习。先后主持了国家自然科学基金重点项目、联合基金重点、重点国际(地区)合作研究、北京市重点研发计划、中科院先导科技专项、启元国家实验室重点项目、装备部重点项目等多项国家级重点项目，在 IEEE T-PAMI、CVPR 等本领域国际主流期刊与会议发表论文 200 余篇，授权发明专利 35 项。获北京市科技进步奖一等奖（排名第一）、北京市科技奖中关村杰出青年人物奖、中国电子学会科技进步一等奖等。他是或曾是 IJCV、IEEE T-CSVT、IEEE T-BIOM、PR 等知名期刊编委。

吕凯风　清华大学

报告题目：大模型基础理论

报告摘要：近年来，大模型技术突飞猛进，在诸多关键的下游任务上展现出卓越的性能。然而，大模型的发展仍高度依赖于实验经验，缺乏系统的理论指导。本报告将围绕与大模型相关的表示能力理论、优化理论、训练动力学等核心主题，系统回顾近期基础理论的研究进展，并探讨其在预训练加速、新型架构设计、思维链推理、长文本推理、AI 对齐等关键问题上的影响与启示。

讲者简介：清华大学交叉信息研究院助理教授，主要从事机器学习理论研究，尤其关注大模型的优化与泛化、神经网络的训练动力学分析等方向。在加入清华大学任教前，他曾在加州大学伯克利分校西蒙斯计算理论研究所担任博士后研究员。他 2024 年博士毕业于普林斯顿大学（师从 Sanjeev Arora 教授），2019 年本科毕业于清华大学。其研究成果发表在 NeurIPS、ICML、ICLR 等机器学习顶级会议上。

虞晶怡　上海科技大学

报告题目：3D 视觉基础模型

报告摘要：随着对三维世界理解需求的日益增长与传感器技术的持续突破，面向物理世界的统一化 3D 视觉基础模型已成为近期的研究热点。本报告聚焦三维表征革新、跨模态对齐及数据高效学习三大核心技术挑战，系统梳理了 3D 视觉基础模型在场景重建、动态交互感知和空间认知推理方面的最新研究进展，并探讨其在具身智能、元宇宙和工业数字孪生等领域的应用潜力。

讲者简介：虞晶怡教授，OSA Fellow，IEEE Fellow，ACM 杰出科学家，智能感知与人机协同教育部重点实验室主任。他于 2000 年获美国加州理工学院（Caltech）双学士学位，2005 年获美国麻省理工学院（MIT）博士学位。现任上海科技大学讲席教授，校副教务长兼信息科学与技术学院院长。虞教授长期从事计算机视觉、计算成像、计算机图形学、生物信息学等领域的研究工作，在他带领下，上科大团队在 2024 年拿到了包括 CVPR、SIGGRAPH、DAC、VIS、MICCAI 等 7 个计算机学科顶会的最佳论文或最佳论文提名奖。他先后获得美国国家科学基金杰出青年奖（NSF CAREER Award），美国空军研究院杰出青年奖 (AFOSR YIP Award)，白玉兰纪念奖。在智能光场研究上，他拥有十余项国际 PCT 专利，已广泛应用于智慧城市、数字人、人机交互等场景。他曾经担任 IEEE TPAMI、IEEE TIP 等多个顶级期刊编委，并担任国际人工智能顶会 CVPR 2021 和 ICCV 2027 的程序主席、ICCV 2025 的大会主席。他是达沃斯世界经济论坛（WEF）“全球议程理事会”理事。

谢伟迪　上海交通大学

报告题目：医疗通用大模型

报告摘要：近年来，基于大模型的医疗 AI 进展迅速，在这次的报告中，我将尝试介绍领域内的一些重要进展，包括以下几个方面：（1）医学领域的语言大模型；（2）面向医学场景的多模态大模型，包括放射学影像，皮肤，眼科，EHR，基因等；（3）针对癌症诊断金标准的病理学大模型；（4）针对罕见病筛查，诊断的多模态大模型；（5）面向疾病诊断与治疗用药推荐的多智能体系统；（6）针对致病机理理解的基因-蛋白质大模型。

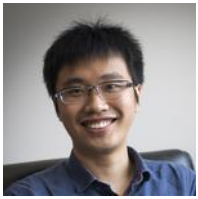
讲者简介：上海交通大学长聘轨副教授，教育部 U40 获得者，国家级青年人才(海外)，上海市海外高层次人才计划获得者，上海市启明星计划获得者，科技部科技创新 2030 —“新一代人工智能”重大项目青年项目负责人，国家基金委面上项目负责人。 博士毕业于牛津大学视觉几何组（Visual Geometry Group，VGG），首批 Google-DeepMind 全额奖学金获得者，China-Oxford Scholarship 获得者，牛津大学工程系杰出奖获得者。主要研究领域为计算机视觉，医学人工智能，共发表论文超 80 篇，包括 Nature Communications，NPJ Digital Medicine，CVPR，ICCV, NeurIPS, ICML, IJCV 等，Google Scholar 累计引用超 14500 余次，多次获得国际顶级会议研讨会的最佳论文奖和最佳海报奖，最佳期刊论文奖，MICCAI Young Scientist Publication Impact Award Finalist； Nature Medicine， Nature Communications 特邀审稿人，计算机视觉和人工智能领域的旗舰会议 CVPR，NeurIPS，ECCV 的领域主席。

邹征夏　北京航空航天大学

报告题目：遥感通用大模型

报告摘要：结合特定领域知识，研究具有领域特色的通用大模型技术，已成为各垂直领域、尤其是航空航天遥感领域的重要发展趋势。本次报告将围绕解译大模型、生成大模型和多模态大模型三个方向，回顾遥感通用大模型的最新进展，分析其中的关键方法与面临的挑战，并对未来的发展方向进行展望。

讲者简介：北京航空航天大学教授、博士生导师，国家级青年人才。主要研究方向为遥感图像处理、深度学习，以第一/通讯作者身份在 Proceedings of the IEEE、Nature 子刊、TPAMI、TIP、CVPR、ICCV 等重要期刊和会议发表论文 30 余篇，合作发表包括 Nature 正刊封面论文在内的学术论文共计 80 余篇；论文引用近一万次，单篇引用 3800 余次，GitHub 订阅 4000 余次；担任 IEEE 高级会员、TIP 期刊编委（Associate Editor）、Nature 旗下首个工程类期刊 Communications Engineering 特刊编辑，入选 AI100 青年先锋计划（首批，全国 65 人）。研究成果被国家自然科学基金委、新华社、新科学人、麻省理工科技评论等重要媒体报道，连续 4 年收录在斯坦福大学的著名公开课“CS231n 深度学习与计算机视觉”，成果应用于高分一号、巴遥一号、委遥一号、高景一号 03 星等遥感卫星，服务海上交通监测、城市规划、民生保障等重要应用。



顾实　浙江大学

报告题目：脑启发与认知计算

报告摘要：“脑启发与认知计算”融合神经科学与人工智能，旨在模拟与借鉴人脑认知机制以提升计算系统智能水平。本年度进展报告将系统梳理神经启发算法、认知建模、类脑架构等方面的最新研究成果，展示该领域的多学科融合趋势与应用前景。

讲者简介：顾实，现任浙江大学计算机科学与技术学院长聘副教授，曾入选 2017 年国家青年特聘专家， 同年入选福布斯中国“30 岁以下 30 人”榜单，主要研究方向为计算神经科学与类脑人工智能。相关研究发表在 Nature Communications、Science Advances、PNAS 等国际期刊和 ICLR、ICML，NeurIPS 等人工智能国际会议上。

叶茫　武汉大学

报告题目：大模型安全与伦理

报告摘要：大模型从单模态到多模态的快速演进，重构了安全与伦理风险挑战边界。本报告首先总结数据隐私保护和隐私计算进展，然后聚焦大模型安全前沿，介绍攻击方式的演化、防御机制的构建与安全评估体系，最后探讨大模型伦理对齐难题，分享可解释性最新成果，展望大模型安全与伦理的未来方向与机遇。

讲者简介：武汉大学计算机学院教授、博士生导师、智能科学系主任、国家高层次青年人才。长期从事联邦学习、多模态大模型高效微调与安全等领域研究，以第一/通讯作者发表 CCF-A 类论文 80 余篇，谷歌学术引用 10000 余次。担任 CCF-A 类 IEEE TIP 和 IEEE TIFS 等 SCI 期刊编委，CVPR、ICLR、NeurIPS、ICML 等会议领域主席，VALSE 执行 AC 等学术职务。主持国自科-香港联合基金、科技部重点研发计划课题等 10 余项科研项目。

Tutorial 报告及讲者简介

时间	内容	地点
6 月 7 日 08:30-12:00	**Tutorial 1：大模型基础理论结构与微调** 报告题目：大模型时代机器学习理论 讲者：王立威（北京大学） 报告题目：基于量化稀疏的高效训练推理：理论及算法 讲者：陈键飞（清华大学）	三楼报告厅
6 月 7 日 13:30-17:00	**Tutorial 2：科学智能（AI for Science）** 报告题目：AI for Science-机遇与挑战 讲者：欧阳万里（上海人工智能实验室） 报告题目：复杂时空动力系统智能科学计算 讲者：孙浩（中国人民大学）	四楼大湾区厅 3
6 月 7 日 18:30-22:00	**Tutorial 3：深度学习网络架构设计** 报告题目：高效视觉基础模型架构设计 讲者：黄高（清华大学） 报告题目：深度网络的基础模块设计：数据先验与架构融合 讲者：谢琦（西安交通大学）	三楼报告厅
6 月 8 日 08:30-12:00	**Tutorial 4：可信机器学习** 报告题目：基于噪声数据的可信机器学习 讲者：韩波（香港浸会大学） 报告题目：基于不均衡数据的可信机器学习 讲者：姚江超（上海交通大学） 报告题目：对抗场景下的可信机器学习 讲者：张杰（中国科学院计算技术研究所）	四楼大湾区厅 3

Tutorial 1: 大模型基础理论结构与微调

王立威 北京大学

报告题目：大模型时代机器学习理论

报告摘要：经典机器学习理论关于表示、优化、泛化三方面的研究在大模型时代遇到严峻挑战。由于训练中数据的单次（或少次）使用方式，传统优化与泛化问题在大模型中几乎不再存在。另一方面，由于思维链的广泛应用，大模型的表示问题也已不再限于神经网络本身。本报告将探讨大模型中机器学习的新问题新理论以及相应方法。分别探讨表示、优化与泛化理论基础与前沿。

讲者简介：王立威 北京大学智能学院教授。长期从事机器学习研究。在机器学习理论方面取得一系列成果。在机器学习国际权威期刊会议发表高水平论文 200 余篇。曾获机器学习顶级会议 NeurIPS 2024 最佳论文奖，ICLR 2023 杰出论文奖、ICLR 2024 杰出论文提名奖。担任人工智能权威期刊 TPAMI 编委。曾入选 AI’s 10 to Watch，是首位入选的中国学者。

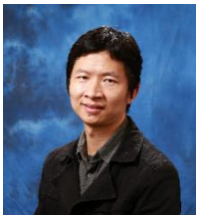
陈键飞 清华大学

报告题目：基于量化稀疏的高效训练推理：理论及算法

报告摘要：大模型所需计算成本高昂，而低精度、稀疏等高效训练推理方法均在原有计算基础上引入了近似，可能会引起精度损失。本报告将介绍近似梯度下降理论，该理论可以为高效的近似训练方法的收敛性、收敛速度提供理论保证。基于该理论，将分别介绍通过量化和稀疏两条技术路线设计的前馈神经网络计算加速、注意力计算加速、激活压缩、优化器压缩、通信压缩等高效训练推理算法。将从机器学习的角度出发，介绍高效训练的过程中遇到的训练不稳定等问题及克服方法。

讲者简介：陈键飞，清华大学计算机系准聘副教授。2010-2019 年获清华大学学士和博士学位。从事高效机器学习研究，谷歌学术引用 5000 余次。担任 IEEE TPAMI 的编委，担任 ICLR 等会议领域主席。获得 CCF 青年人才发展计划、清华大学学术新人奖等。

Tutorial 2：科学智能（AI for Science）



欧阳万里 上海人工智能实验室

报告题目：AI for Science-机遇与挑战

报告摘要：近十年，以深度学习为代表的人工智能算法取得了突飞猛进的发展，并大规模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 Tutorial 探索重大科学问题研究的范式，研究从微观到宏观自然科学的共性 AI 算法，通过人工智能与物理、化学、生物等自然学科的结合，加速人工智能在气象、新材料研发等领域的探索，赋能各行业发展。

讲者简介：欧阳万里，上海人工智能实验室 领军科学家。曾任悉尼大学电子信息工程学院研究主任。研究领域：模式识别、深度学习、计算机视觉、AI for Science。谷歌学术引用 5,4000+, H-index 指数 103。澳大利亚未来学者杰出青年人才计划、悉尼大学杰出科研校长奖、澳大利亚计算机科学领导者奖。领导研发“风乌”气象大模型。ICCV 最佳审稿人，担任人工智能领域顶级期刊 TPAMI 和 IJCV 副编，CVPR2023 资深领域主席，CVPR2021、ICCV2021 领域主席。获得 ImageNet 和 COCO 物体检测第一名。

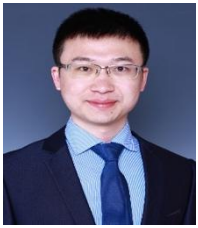
孙浩 中国人民大学

报告题目：复杂时空动力系统智能科学计算

报告摘要：对复杂动力/动态系统演化行为开展有效且高效的科学计算，一直是各学科领域的前沿挑战。以深度学习为代表的人工智能技术虽在科学计算领域有着巨大潜力，但具有可解释性差、鲁棒性差、泛化性差、误差不可控、对数据依赖性强等问题。因此，把数据驱动的机器学习与知识驱动的符号计算融合，发展新的智能科学计算理论和方法势在必行。本报告面向复杂时空动力系统（如湍流、气象/大气、反应扩散过程等）建模与仿真场景，介绍一系列数据-机理驱动智能科学计算方法，解决上述挑战难题，为复杂系统建模、加速仿真、知识发现搭建新的范式。

讲者简介：孙浩，中国人民大学高瓴人工智能学院“长聘副教授、博导”，国家高层次青年人才，哥伦比亚大学博士、麻省理工学院博士后，曾任美国匹兹堡大学、美国东北大学终身序列助理教授、博导。从事智能科学计算理论方法与前沿交叉研究。在 Nature Machine Intelligence、Nature Communications、National Science Review 等国际一流期刊和 NeurIPS、KDD、ICLR、IJCAI 等人工智能领域顶会发表论文 80 余篇。主持国家高层次人才计划青年项目、国家自然科学基金（重大培育、面上）、美国科学基金（重点、面上、专题）等科研项目十余项。先后获得福布斯北美 30 位 30 岁以下精英榜（科学类）、美国十大华人杰出青年、中国智能计算科技创新人物等荣誉。

Tutorial 3: 深度学习网络架构设计



黄高 清华大学

报告题目：高效视觉基础模型架构设计

报告摘要：进入大模型时代，神经网络的训练与推理效率变得日益重要。本报告将从神经网络架构设计的角度，针对影响模型计算效率的三个主要因素，即网络深度、参数规模和数据序列长度，探讨提升效率的方法与技术。报告将首先简要回顾神经网络深度的演进，分析网络深度对计算效率的影响；其次，介绍如何利用样本自适应和时空自适应动态神经网络解决大参数与效率的矛盾；最后，将重点介绍如何利用线性注意力和差分注意力实现高效的长序列建模，为设计与训练高效的大模型提供新的思路与方法。

讲者简介：黄高，清华大学自动化系副教授，博士生导师。主要研究领域为深度学习与智能系统，提出了 DenseNet、CondenseNet、动态神经网络等代表性深度学习模型。共计发表学术论文 100 余篇，被引 8 万余次，最高单篇引用超过 5 万次。获国家优青、CVPR 最佳论文奖、达摩院青橙奖、世界人工智能大会 SAIL 奖、亚洲青年科学家奖、AI 2000 人工智能最具影响力学者、MIT TR35 等，担任 IEEE T-PAMI、IEEE T-BD、Pattern Recognition 等国际重要期刊编委和 CVPR、ICCV、NeurIPS、ICML 等人工智能顶级会议领域主席。

谢琦 西安交通大学

报告题目：深度网络的基础模块设计：数据先验与架构融合

报告摘要：本报告回顾深度网络基础模块的设计演进历程，从早期全连接层的通用近似性、CNN 的局部平移等变机制，到 Transformer 的自注意力机制，系统探讨数据先验与架构设计的紧密关联。重点剖析自注意力类模块的建模机制，揭示其与图像、视频中非局部自相似性先验的数学同构性，并深入分析旋转、尺度等几何对称先验如何引导新型模块设计（如群等变卷积与 Steerable 滤波器）。通过实例阐释几何先验的显式编码对网络泛化性的理论增益，进一步探讨大模型时代下数据先验与架构融合的前景与挑战，为构建高效鲁棒的基础模块提供新思路与技术路径。

讲者简介：谢琦，西安交通大学数学与统计学院副教授，博导。于 2013 年 7 月和 2020 年 12 月分别获西安交通大学理学学士与理学博士学位。2017 年 8 月至 2018 年 9 月曾赴普林斯顿大学访学。目前主要从事机器学习与计算机视觉的基础问题研究。在 CCF A 类期刊与会议发表论文 21 篇，IEEE Trans.论文 15 篇，其中以第一或通讯作者在领域顶刊 TPAMI 发表论文 4 篇；4 篇论文入选 ESI 高被引论文。2015 年至今，谷歌学术被引 5000 余次，H 指数为 23。曾获 2022 年 CCF 优秀博士学位论文奖”、“2021 年 ACM 中国优博提名奖”、2024 年华为“火花奖”等奖项。

Tutorial 4: 可信机器学习

韩波 香港浸会大学

报告题目：基于噪声数据的可信机器学习

报告摘要：标签噪声学习是机器学习领域的核心研究课题之一，其聚焦于解决训练数据中误标注（即噪声标签）样本对模型性能造成的负面干扰，从而提升机器学习系统整体的稳健性与泛化能力。该领域的发展历经了传统统计机器学习时代的理论基础，到深度学习时代的算法变革，再到当前大模型范式下的持续探索，标签噪声学习始终具有关键的价值与现实意义。本报告将简要回顾标签噪声学习领域的发展历程与最新研究现状，从训练策略、数据建模与正则化方法三个层面，分析现有方法的核心思想及其所面临的主要挑战，并对未来可能的研究方向进行展望。相关成果发表在会议 NeurIPS、ICML、ICLR 以及期刊 IEEE Transactions on Pattern Analysis and Machine Intelligence、Journal of Machine Learning Research 等。

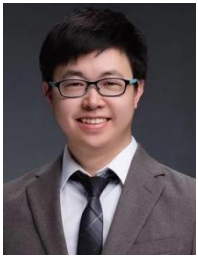
讲者简介：韩波，香港浸会大学机器学习方向助理教授、可信机器学习与推理组主任，日本理化学研究所先进智能中心项目梅峰访问科学家，其研究重点是机器学习、深度学习、基础模型及其应用。 曾是 MBZUAI 机器学习系的访问学者，微软亚洲研究院和阿里巴巴达摩院的访问教职研究员，也是日本理化学研究所先进智能中心项目的博士后研究员。 在悉尼科技大学获得计算机科学博士学位。 目前担任 NeurIPS 高级领域主席，以及 NeurIPS、ICML 和 ICLR 领域主席。 还担任 IEEE TPAMI、MLJ 和 JAIR 副主编，以及 JMLR 和 MLJ 编委会成员。

姚江超 上海交通大学

报告题目：基于不均衡数据的可信机器学习

报告摘要：不均衡学习是一个广泛存在且研究已久的领域，随着深度学习的发展，近些年来有了一些新的设计和改善。然而，随着整个人工智能领域趋势的演进，原有仅仅围绕类别不均衡的研究设定未必符合或者考虑了各种情形下的领域需求，如何把不均衡学习拓展到泛在情形就变得很重要。本报告将简要回顾不均衡学习领域的发展历史和现状，分析现有不均衡学习领域需要考虑的关键点，重点围绕将经典的不均衡学习的思想推广到表征学习、偏标签学习和增量学习的场景中，来克服样本不均衡影响的问题进行展开。我们将分别介绍我们在应对上述情景中提出的记忆增强对比学习方法、全局特征校正方法和惯性增强方法，来解决不均衡学习方法的泛在适配的难点，通过大量实验论证我们方法的有效性，相关成果发表在 ICML、ICLR 和 Transactions on Image Processing 上。

讲者简介：姚江超，上海交通大学助理教授、博导，上海人工智能实验室双聘青年科学家。上海交通大学和悉尼科技大学博士，曾任阿里巴巴达摩院算法专家。长期致力于鲁棒机器学习、大模型预训练相关技术和医疗人工智能应用研究，发表高水平论文 80 余篇，担任 ICML、NeurIPS、ICLR 领域主席，Transactions on Machine Learning Research 和 Neural Networks 执行主编，曾获教育部科技进步一等奖。



张杰 中国科学院计算技术研究所

报告题目：对抗场景下的可信机器学习

报告摘要：近年来，深度学习模型在计算机视觉、自然语言处理等领域取得了显著进展，随着深度学习在高安全需求领域（如自动驾驶、医疗诊断）的广泛应用，模型在对抗场景下的可信性成为关键挑战。本次报告主要围绕对抗场景下的可信机器学习展开，系统性地介绍对抗攻击的基本原理、防御策略及面向多模态大模型对抗攻防、安全评估的最新研究进展，探讨如何构建对抗场景下安全可信的人工智能系统。

讲者简介：张杰，中国科学院计算技术研究所副研究员、博士生导师，安全可信的智能算法研究组负责人，主要从事可信视觉计算、算法攻防、多模态大模型安全评估与防御等基础和应用研究，在知名期刊会议 TPAMI/IJCV/CVPR/ICCV/NeurIPS 等上发表论文 50 余篇，获 2024 年中国图象图形学学会自然科学奖一等奖、2024 年中国铁道学会科技进步奖二等奖。先后入选北京市科技新星、姑苏领军人才计划、中科院青促会和微软青年学者铸星计划。担任 ACL 领域主席、智能安全期刊青年编委。

Workshops 报告及讲者简介

Workshop 1: 多模态遥感通用模型

组织者：赵文达（大连理工大学）、冯婕（西安电子科技大学）、洪丹枫（中国科学院空天信息创新研究院）

时间：6 月 7 日（周六）18:30-22:00　地点：四楼大湾区厅 1

时间	主持人	内容
18:30-19:00	冯婕	讲者：夏桂松（武汉大学） 题目：面向开放场景的多源遥感图像变化检测
19:00-19:30	冯婕	讲者：方乐缘（湖南大学） 题目：多模态遥感通用大模型研究初探
19:30-19:40	冯婕	企业宣讲：阿里云（仇明） 题目：通义大模型在科研领域的应用与实践
19:40-20:10	赵文达	讲者：张洪艳（中国地质大学） 题目：大范围高分辨率土地覆盖遥感制图
20:10-20:40	赵文达	讲者：郑钰辉（青海师范大学） 题目：高光谱遥感图像智能解译与多模态通用大模型的思考
20:40~20:50	中场休息	
20:50-21:20	洪丹枫	讲者：聂婕（中国海洋大学） 题目：面向海洋科学遥感的多模态智能计算模型
21:20-21:50	洪丹枫	讲者：赵文达（大连理工大学） 题目：非完备图像信息目标智能感知

组织者：赵文达 大连理工大学

个人简介：赵文达，博士生导师。研究方向为多模态图像分析，在包括 IEEE TPAMI，IEEE TIP，CVPR，AAAI 等本领域顶级期刊和会议上发表学术论文 50 余篇。获得山东省科技进步一等奖，中国指控学会技术发明二等奖，IEEE MMTC Best Conference Paper Award。大连市高端人才，大连市科技之星。担任中国人工智能学会智能融合专业委员会副秘书长，视觉与学习青年学者研讨会议（VALSE）执行 AC 等。

组织者：冯婕 西安电子科技大学

个人简介：冯婕，西安电子科技大学教授，博士生导师，研究方向空天遥感图像智能处理，发表学术论文 80 余篇，其中包括 CCFA 类和中科院 I 区论文 40 余篇，ESI 高被引/热点论文 9 篇，出版专著 2 部，连续多年入选“全球前 2%顶尖科学家榜单”。主持军科委基础加强领域基金、装备预研教育部联合基金等。入选中国科协青年托举人才计划、陕西省特支计划青年拔尖人才等。担任 Frontiers in Imaging 副主编、Remote Sensing 编委。获得中国自动化学会自然科学二等奖、中国航天集团技术进步奖二等奖。担任国际期刊 Frontiers in Imaging 副主编、Remote Sensing 期刊编委、中国电子学会青年科学家俱乐部理事等。

组织者：洪丹枫 中国科学院空天信息创新研究院

个人简介：洪丹枫，中国科学院空天信息创新研究院研究员，博士生导师，中国科学院大学岗位教授，科睿唯安全球“高被引科学家”，日本 JSPS 学者。德国慕尼黑工业大学（TUM）博士，曾为德国宇航中心（DLR）研究员兼“光谱视觉”课题组组长、法国格勒诺布尔阿尔卑斯大学-傅立叶实验室（GIPSA Lab）客座研究员、德国亥姆霍兹联合会人工智能研究院（HACUI）AI 科研顾问。主要研究方向为人工智能、多模态大数据、大模型、地球科学等。获得《麻省理工科技评论》-- 中国智能计算创新人物奖、中国指挥与控制学会青年科技奖、IEEE GRSS 杰出青年奖（Early Career Award）、Remote Sensing 青年科学家奖、国际高光谱顶级会议 WHISPERS 杰出论文奖（Jose Bioucas Dias 奖、Paul Gader 奖）等。担任 IEEE TIP、IEEE TGRS、ISPRS JP&RS、Information Fusion 等期刊副主编/编委。

夏桂松 武汉大学

报告题目：面向开放场景的多源遥感图像变化检测

报告摘要：遥感影像变化检测是实现地表动态监测与地理信息更新的关键技术，在城市管理、环境评估和灾害响应等领域具有重要价值。当前方法多依赖双时相遥感影像，并预设有限的变化类别，难以适应开放场景下多样化的变化需求。本次报告聚焦于大规模预训练模型在变化检测中的构建与应用。通过引入大语言模型解析检测目标表达，并结合多源遥感特征，提升模型对任意语义变化的理解与识别能力。同时，借助地图等高层语义信息实现图像内容的语义对齐，有效增强模型在单图条件下的变化感知能力。该研究旨在推动变化检测从封闭类别识别向开放语义建模转变，为遥感影像智能解译提供更具适应性和扩展性的解决方案。

讲者简介：夏桂松，武汉大学弘毅特聘教授，现任武汉大学人工智能学院副院长（主持工作）、国家多媒体软件工程技术研究中心副主任；主持国家自然科学基金杰青、优青、联合重点、重大研究计划等纵向研究项目 20 余项，系列成果发表业内知名期刊/会议论文 150 余篇，谷歌学术引用 3 万余次，并在国家重要工程和知名企业相关业务场景中成功应用；获得湖北省自然科学一等奖 1 项、中国测绘科技进步一等奖 3 项、IEEE GRSS 最有影响力论文奖、中国图象图形学会优秀博士论文指导教师等荣誉；应邀兼任 2 个 SCI 一区期刊编委、中国图象图形学学会遥感图像专业委员会副主任、中国自动化学会模式识别和机器智能委员会常务委员。

方乐缘 湖南大学

报告题目：多模态遥感通用大模型研究初探

报告摘要：随着遥感观测手段的持续丰富与遥感数据体量的指数级增长，构建具备跨模态感知、任务泛化与知识迁移能力的遥感大模型，已成为推动遥感智能化演进的关键路径之一。本报告系统梳理了遥感基础模型的发展脉络，聚焦于模型架构设计、预训练范式、多模态融合机制及其跨任务适应能力的演进趋势。在此基础上，报告进一步探讨多模态遥感数据集构建、评测基准体系与伦理规范等关键支撑要素，分析从单一遥感模态模型向多模态融合模型的技术演化路径，及其在遥感-语言模型等新兴跨模态任务中的应用潜力。最后，报告深入剖析当前遥感大模型在通用性、可解释性与部署效率等方面所面临的主要挑战与未来发展方向，为构建高效、可信与泛化能力强的遥感智能体系提供理论支撑与技术参考。

讲者简介：方乐缘，湖南大学教授，国家杰青、国家优青，科睿唯安（Clarivate Analytics）全球“高被引科学家”，爱思唯尔中国高被引学者。获得国家自然科学二等奖 2 项、IEEE GRSS 最高影响力论文奖、湖南省自然科学一等奖 2 项等。担任 SCI 期刊 IEEE Transactions on Image Processing、IEEE Transactions on Neural Networks and Learning System、IEEE Transactions on Geoscience and Remote Sensing 等期刊编委。现主要从事深度学习、弱监督学习以及在遥感图像处理与分析等方面的研究。研究成果在国际权威期刊和会议发表论文

200 余篇，其中 IEEE TPAMI、IJCV、TIP 等本领域权威期刊论文 100 余篇，Google scholar 引用 17000 余次，ESI 高被引 22 篇，ESI 热点论文 4 篇。主持国家杰青、基金委联合重点、国家重点研发课题等项目。

张洪艳 中国地质大学

报告题目：大范围高分辨率土地覆盖遥感制图

报告摘要：作为遥感对地观测研究的核心任务之一，土地覆盖制图可为　生态、环境、农业和双碳等多学科研究提供共性基础底图。针对大范围高分辨率土地覆盖制图中高分辨率训练样本缺乏的关键难题，本报告介绍了一种基于弱监督深度学习的跨分辨率土地覆盖制图框架 L2HNet，利用弱监督策略挖掘带噪粗标签中的可靠监督信息，并在此基础上绘制了中国首幅国家尺度 1 米分辨率土地覆盖图。为进一步解决地物复杂多样、区域迁移性要求高等问题，本报告进一步提出利用历史低分辨率标签进行大范围高分辨率土地覆盖制图的网络 Paraformer，结合并行的 CNN 和 Transformer 分支，提取遥感影像中的高分辨率细节纹理与全局多样地貌特征，实现高效的大范围高分辨率土地覆盖制图。

讲者简介：张洪艳，男，教授、博导，国家级青年人才，湖北省杰青获得者。主要从事智能信息处理、农业遥感监测等方向的研究工作。在相关方向发表 SCI 期刊论文 130 篇，其中 ESI 热点论文 4 篇、ESI 高被引论文 23 篇和爱思唯尔年度热门论文 1 篇，出版学术专著 1 部，授权国家发明专利 12 项。主持国家重点研发计划课题 1 项、国家自然科学基金 5 项和省部级项目 3 项。研究成果先后荣获测绘科学技术一等奖 3 项、湖北省自然科学二等奖 1 项和 IEEE 地球科学与遥感学会数据融合大赛冠军 3 项。IET Fellow，IEEE Senior Member，连续 4 年入选斯坦福大学全球前 2%顶尖科学家、爱思唯尔中国高被引学者和全球学者库前 10 万顶尖科学家等榜单，应邀担任 IEEE JSTARS、PE & RS、Computers & Geosciences 等 SCI 期刊副主编。

郑钰辉 青海师范大学

报告题目：高光谱遥感图像智能解译与多模态通用大模型的思考

报告摘要：高光谱智能解译是地球观测领域的前沿方向。报告聚焦高光谱解译的关键技术挑战，重点探讨光谱-空间特征耦合建模、多模态数据协同分析等问题，并分享三项研究成果：提出多尺度交叉融合 Transformer 高光谱与多光谱融合方法，通过多尺度交叉注意力与多方向级联条带卷积，解决异构数据尺度差异，提升融合质量；构建局部与全局信息感知的遥感图像显著性目标检测方法，采用边缘-语义联合挖掘及上下文感知建模，实现遥感图像小目标检测；设计空-谱聚类注意力 Transformer 分类方法，创新性地将空间和光谱聚类簇嵌入自注意力计算过程，降低冗余并实现空谱特征的深度聚焦。最后，探索将上述方法拓展至多模态遥感通用模型框架，讨论统一特征表达范式与跨模态关联机制。

讲者简介：郑钰辉，教授、博导，国家级青年人才，江苏省杰出青年基金获得者。全重实

验室副主任、教育部重点实验室主任。现为中国人工智能学会智能服务专委会常务委员、中国认知科学学会认知与类脑计算专委会委员、中国图象图形学学会青工委委员。围绕社会与空间视觉感知计算，主持国家自然科学基金委重点项目、国际（地区）交流与合作项目等课题。已在 IEEE-TPAMI、IEEE-TIP、IJCAI、AAAI、ACM-MM 等人工智能、计算机视觉领域国际权威期刊与会议上发表学术论文百篇以上。曾获 2023 年度江苏省高等学校科学技术研究成果奖一等奖、2022 年度江苏省科学技术奖二等奖，2024 年度 CCF 科技进步二等奖。

聂婕 中国海洋大学

报告题目：面向海洋科学遥感的多模态智能计算模型

报告摘要：海洋科学遥感在全球海洋监测与环境预报中发挥着关键作用，相较于陆地光学遥感，由于受云层遮蔽、观测手段有限和观测环境动态变化等因素影响，普遍存在时空分布稀疏、缺测严重及多源异构等问题，对传统以单模态和高完整性为基础的遥感处理方法提出了严峻挑战。本报告系统分析了海洋科学遥感与传统光学遥感在数据形态、物理约束和任务目标等方面的关键差异，融合物理先验与数据驱动的方法以弥合数据空白、提升推理能力，提出一种面向海洋场景的多模态智能计算模型架构，通过引入海洋动力学知识约束、利用多源遥感与数值模式数据互补协同建模，并结合深度学习对缺失信息进行合理估补和泛化推理，有效提升在稀疏观测条件下的海洋信息恢复和预测能力。

个人简介：聂婕，中国海洋大学，“英才工程第一层次”教授，博士生导师，围绕海洋人工智能交叉研究方向发表百余篇高水平论文，提出“海洋多模态智能计算”等面向海洋科学的人工智能关键技术，担任科技部重点研发计划项目“数据驱动的海洋生态系统健康评价”青年科学家项目首席，围绕该方向主持国家自然科学基金优秀青年项目“海洋多模态智能计算”、区域联合（重点）项目“海洋大数据知识演化与复杂问题协同推理”等国家级项目 6 项，以第一发明人授权国家发明专利 45 余项，获得山东省科技进步一等奖 2 项，山东计算机学会科技进步一等奖 1 项，担任 IEEE TCSVT、IJDE，IMTS 等国际顶尖期刊编委。

赵文达 大连理工大学

报告题目：非完备图像信息目标智能感知

报告摘要：利用多模态遥感图像间互补信息，通过多模态图像融合提升图像质量。针对图像融合中缺少真值监督、目标语义难挖掘等问题，从自监督学习、多任务联合学习角度报告多模态图像融合解决方案。进一步，针对遥感样本标注困难问题，从样本生成、文本语义生成等角度报告小样本遥感目标识别解决方案。

讲者简介：赵文达，博士生导师。研究方向为多模态图像分析，在包括 IEEE TPAMI，IEEE TIP，CVPR，AAAI 等本领

域顶级期刊和会议上发表学术论文 50 余篇。获得山东省科技进步一等奖，中国指控学会技术发明二等奖，IEEE MMTC Best Conference Paper Award。大连市高端人才，大连市科技之星。担任中国人工智能学会智能融合专业委员会副秘书长，视觉与学习青年学者研讨会议（VALSE）执行 AC 等。

Workshop 2：多模态学习助力智慧医疗：从科研到落地

组织者：谢雨彤（穆罕默德·本·扎耶德人工智能大学）、雷柏英（深圳大学）、夏勇（西北工业大学）

时间：6 月 7 日（周六）13:30-17:00　　地点：三楼报告厅

时间	主持人	内容
13:30~14:10	夏勇	讲者：沈定刚（上海科技大学生物医学工程学院，上海联影智能医疗科技有限公司） 题目：生成式人工智能与疾病诊疗
14:10~14:40	夏勇	讲者：王珊珊（中国科学院深圳先进技术研究院） 题目：多模态可泛化医学影像人工智能基础模型研究
14:40~15:10	谢雨彤	讲者：陈浩（香港科技大学） 题目：大模型赋能智慧医疗：挑战和未来
15:10~15:40	谢雨彤	讲者：隋尧（北京大学） 题目：神经成像质量与效率：面向临床的优化
15:40~16:10	雷柏英	讲者：张亚琴（中山大学附属第五医院） 题目：人工智能助力乳腺疾病影像解读
16:10~16:40	雷柏英	讲者：张建鹏（阿里巴巴达摩院、浙江大学） 题目：基于视觉语言预训练的开放场景医学影像理解
16:40~17:00	谢雨彤	Panel 嘉宾：沈定刚（上海科技大学生物医学工程学院，上海联影智能医疗科技有限公司）、王珊珊（中科院深圳先进院）、陈浩（香港科技大学）、隋尧（北京大学）、张亚琴（中山大学附属第五医院）、张建鹏（阿里巴巴达摩院，浙江大学）、夏勇（西北工业大学）

组织者：谢雨彤 穆罕默德·本·扎耶德人工智能大学

个人简介：谢雨彤，穆罕默德·本·扎耶德人工智能大学(MBZUAI)助理教授，2022-2024 年在澳大利亚阿德莱德大学机器学习研究所任博士后研究员。研究方向为医疗人工智能，重点关注于有限标注下医学数据的高效分析和解读，多模态医学数据分析。迄今为止，在 IEEE-TPAMI、IJCV、CVPR、MICCAI 等中科院一区顶级期刊和领域顶级会议发表 50 余篇论文，4 篇入选 ESI 高被引，谷歌学术总引用 5900 余次。曾获中国图象图形学学会优秀博士学位论文奖、陕西省自然科学优秀学术论文奖等，2022-2024 连续三年入选斯坦福大学全球前 2%顶尖科学家榜单。2023-2025 连续三年担任领域顶级会议 MICCAI 的领域主席，并担任 VALSE 2025 Tutorial 主席，VALSE 2024 & 2025 执行领域主席和医学影像青年论坛（MICS）委员会委员。

组织者：雷柏英 深圳大学

个人简介：雷柏英，国家级青年人才，深圳大学特聘教授，西安电子科技大学客座教授，博士生导师，获新加坡南洋理工大学博士学位。主要研究方向为医学图像处理和人工智能。在 IEEE TPAMI、TMI 等以第一/通讯作者 (含共同)发表 SCI 论文 100 余篇(9 篇 ESI 高被引，1 篇热点论文)。谷歌学术总引用超万次，H 指数 58。主持国家自然科学基金联合基金重点 1 项，面上 2 项等项目 20 余项（含国家级 7 项）。现任 IEEE TNNLS、TCYB、TMI、JBHI、Medical Image Analysis 等 10 种 SCI 期刊编委。IEEE BISP、BIIP、BSP 等技术委员会委员，医学图像顶级学术会议 MICCAI 领域主席（2021-2023）。IEEE 广州分部 WIE 主席，人工智能 A 类会议 AAAI、IJCAI 程序委员会委员，IEEE EMBS 杰出讲师，入选 2024 年 ScholarGPS 的近 5 年全球 0.05%高引学者榜单和“全球前 2%顶尖科学家”（2020-2023）和终身影响力榜单（2023），获 “强国青年科学家”提名（2022，全国共 40 人），CSIG 石青云女科学家奖（2022）。

组织者：夏勇 西北工业大学

个人简介：夏勇，西北工业大学长聘教授、博士生导师、空天地海一体化大数据应用技术国家工程实验室成员，研究方向为医学影像智能计算，近 5 年在本领域顶级期刊 JAMA Network Open、Radiology、IEEE-TPAMI/TMI/TIP/JBHI、IJCV、MedIA 和顶级会议 NeurIPS、CVPR、ECCV、AAAI、IJCAI、MICCAI 发表论文 80 余篇，谷歌引用 1.6 万余次（H-Index = 60），现为中国体视学学会理事、中国计算机学会数字医学分会常委、中国抗癌协会肿瘤影像专业委员会人工智能学组副组长。个人主页：https://teacher.nwpu.edu.cn/yongxia.html

沈定刚 上海科技大学生物医学工程学院 上海联影智能医疗科技有限公司

报告题目：生成式人工智能与疾病诊疗

报告摘要：本报告探讨了多模态医疗大模型的研究与应用，以及对医疗领域的影响。首先介绍传统医疗影像 AI 的核心应用，包括成像、全身器官及靶区分割、疾病诊断等。接着，介绍医疗大模型的研发策略，提出通过借鉴通用大模型的方法，构建具备医疗专业性的大模型；例如，通过影像、文本和视频数据的融合，为复杂医疗场景提升诊疗和决策（包括智能勾画、精准诊断和报告生成）。此外，还介绍在手术室全景建模、实时术中导航、术后报告生成等场景中的应用，实现操作流程的智能化。

讲者简介：沈定刚，上海科技大学教授、生物医学工程学院创始院长，联影智能联席 CEO，IEEE/AIMBE/IAPR/MICCAI/ISMRM/IAMBE Fellow，美国 The Academy for Radiology & Biomedical Imaging Research 杰出研究者奖，2024 IEEE EMBS 技术成就奖。曾任美国 UNC-Chapel Hill 终身教授、冠名杰出教授，宾夕法尼亚大学 (UPenn) 助理教授，约翰霍普金斯大学 (Johns Hopkins University) 讲师。世界上最早开展医学影像人工智能研究的科学家之一，并最先将深度学习应用于医学影像。发表论文 1500 余篇，H-index 156，引用 10 万余次。三个国际期刊高级编辑（Senior Editor），六个国际期刊主编/副主编/编委。

王珊珊 中科院深圳先进院

报告题目：多模态可泛化医学影像人工智能基础模型研究

报告摘要：当前，医学多模态智能系统面临数据异质性高、标注数据稀缺以及模型泛化能力不足等关键挑战。尽管现有多模态学习方法在一定程度上实现了不同类型医学数据的融合，但在实际应用中仍存在融合效果不理想、模型对新场景适应性差等问题。本报告聚焦于医学多模态联合表征学习的泛化性研究，重点探索结合自监督学习与对比学习的预训练策略，充分挖掘大规模无标注医学数据的潜力，增强模型对不同模态间关联的理解能力及其跨领域迁移能力。通过减少对人工标注数据的依赖，有望显著提升模型在不同医院、设备和人群中的泛化表现。本研究将应用于多个实际场景，包括鲁棒性与稳定性兼具的快速磁共振成像、肺部文本-图像跨模态理解，以及面向开放环境的多部位、多模态、多尺度对齐与 Zero-shot 自适应学习等。

专家简介：王珊珊，中国科学院深圳先进技术研究院研究员、博士生导师，国家优青，2021-2024 入选 Elsevier 和斯坦福 “全球前 2%顶尖科学家” 榜单。长期从事人工智能、快速医学成像、放射组学与多模态分析等研究，在 Nature 子刊、IEEE Trans 等发表高质量论文。在人工智能快速医学成像与成像物理引导的小样本学习技术做出开创性工作，曾受邀在第 31 届国际医学磁共振年会给大会主题冠名报告（入选率约 1/6000，英国伦敦）及美国第 10 届 GRC 活体磁共振给大会主题报告（美国安多弗）。核心技术转让两家医疗公司，装机超千台，含多款磁共振与低剂量 CT 设备。曾荣获吴文俊人工智能技术发明一等奖(1/6)、OCSMRM 杰出研究奖(1/1)、广东省技术发明与科技进步一等奖(2/10)、广东省青年科技奖(1/1)等；先后主持科技部 2030 新一代人工智能重大课题、NSFC 联合基金重等国家级项目 6 项；担任 SCI 学术期刊的副主编/编委（如 IEEE Transactions on Medical Imaging, IEEE

Transactions on Computational Imaging, Magnetic Resonance in Medicine, Pattern Recognition、Biomedical Signal Processing and Control 等)。

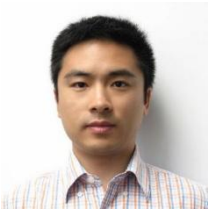
陈浩 香港科技大学

报告题目：大模型赋能智慧医疗：挑战和未来

报告摘要：人工智能大模型极大地提高了视觉计算和自然语言处理等诸多领域的识别性能。尽管在上述领域取得了突破，其在医疗多模态大模型的分析与应用仍有待探索，尤其针对基准模型构建和多模态异构数据融合分析等。本次报告将分享我们在医疗多模态大模型研发与应用等方面的最新进展，以及在疾病诊断、疗效预测和预后等方面的应用和挑战。

实验室主页：https://smartlab.cse.ust.hk/

专家简介：陈浩，香港科技大学计算机科学与工程系，化学与生物工程系和生命科学部助理教授，医工交叉联合创新中心主任，研究兴趣包括大模型医疗，计算病理，多模态数据融合，医学图像分析，可解释深度学习，计算机辅助微无创诊疗等。在 Nature Biomedical Engineering、Nature Communications、Lancet Digital Health、Nature Machine Intelligence、MICCAI、IEEE-TMI、MIA、CVPR、ICCV 等顶级期刊和会议发表论文 200 余篇（谷歌学术引用 32600 余次，h-index 75），连续入选斯坦福大学全球排名前 2%科学家名单，科睿唯安全球高被引科学家等。曾获得 2023 年亚洲青年科学家、国家教育部优秀成果二等奖、北京市科技进步一等奖、2019 年人工智能医学影像顶级会议 MICCAI 青年科学家影响力奖等奖项，担任包括 IEEE TMI、TNNLS、J-BHI 和 CMIG 等期刊编委，担任 ICLR、CVPR、ACM MM、MICCAI 等多个国际会议的领域主席和程序委员，曾带领团队获得 15 余项国际医学图像分析的挑战赛冠军。



隋尧 北京大学

报告题目：神经成像质量与效率：面向临床的优化

报告摘要：传统神经成像方法受限于成像质量与成像效率的矛盾，严重影响临床新技术转化与应用。得益于近年来人工智能技术的快速发展，成像质量与效率的平衡与优化也随之迎来了新的契机。通过整合智能技术与成像方法创新，高效且临床切实可行的神经成像方法得以成功实现。本报告介绍以真实临床场景应用为导向的面向结构与功能的智能神经成像新技术新方法。

专家简介：隋尧，北京大学助理教授、博士生导师，健康医疗大数据国家研究院、医学技术研究院、人工智能研究院副研究员。博士毕业于清华大学电子工程系，图像处理专业。曾于哈佛大学医学院放射学系任教研系列讲师、哈佛大学附属波士顿儿童医院计算放射学实验室任研究科学家。现任国际医学图像计算与计算机辅助干预会议 MICCAI 领域主席、国际生物医学成像会议 ISBI 大会报告主席、国际人工智能会议 AAAI 资深组委会成员。研究工作聚焦于人工智能与医学成像的交叉领域，发表研究论文 50 余篇，其中一作 20 余篇。

张亚琴 中山大学附属第五医院

报告题目：人工智能助力乳腺疾病影像解读

报告摘要：人工智能辅助乳腺影像分析是基于医学影像诊断流程，结合深度学习与生成对抗网络等前沿技术，旨在实现乳腺疾病的早期精准筛查与个体化诊断。本报告聚焦人工智能在乳腺疾病影像分析中的应用，围绕《健康中国 2030》规划纲要及两癌筛查的总目标，重点介绍了团队在人工智能在合成乳腺高分辨率图像、自动分割、乳腺疾病良恶性分类及乳腺癌淋巴血管侵犯预测等领域的研究进展，强调了其可解释性和临床转化前景，助力乳腺影像智能诊断的发展。

专家简介：张亚琴，中山大学附属第五医院影像医学部主任兼放射科主任，博士生导师、博士后合作导师。在 BMJ 等权威期刊共发表学术论文 50 余篇；获广东省科技进步一等奖、二等奖各 1 项及华夏医学科技奖；主持及参与该领域国家级课题多项，参编多项指南及专著。创立人工智能创新工作室 6 年并担任珠海市劳模创新联盟领衔人，积极推动 AI 与影像组学在常见肿瘤等重大疾病早筛与疗效评估中的临床转化。担任中华医学会放射学分会、广东省医学会等多项学术委员会委员，并担任 RSNA、ER 通讯会员及 CJAR 等多个期刊编委。

张建鹏 阿里巴巴达摩院 浙江大学

报告题目：基于视觉语言预训练的开放场景医学影像理解

报告摘要：AI 在帮助放射科医生提高医学图像解释和诊断的效率和准确性方面显示出巨大的潜力。然而，一个通用的人工智能模型需要大规模的数据和全面的注释，这在医疗环境中通常是不切实际的。本次报告将介绍我们最近在开放场景医学影像理解方面的研究工作，重点关注视觉语言预训练方向，分析当前医学影像分析领域所面临的挑战，并介绍一些初步探索工作。

专家简介：张建鹏，阿里巴巴达摩院算法专家、浙江大学博士后研究员。主持浙江省博后择优资助项目，获 2024 年度中国图像图形学会优博，连续三年入选全球前 2%科学家榜单。研究方向为肿瘤早筛、多模态大模型，近五年在 IEEE-TPAMI/TMI、MedIA、ICLR、CVPR、ICCV、ECCV、MICCAI 等本领域顶级期刊/会议发表论文 30 余篇，谷歌学术被引用 5000 余次。担任 MICCAI 2024、2025 领域主席，以及 IEEETPAMI/TMI/CVPR/MICCAI 等十余个国际期刊和会议的评审人。

Workshop 3：开放视觉环境理解与生成

组织者：陈使明（阿联酋人工智能大学(MBZUAI)）、谢国森（南京理工大学）、杨小汕（中国科学院自动化研究所）

时间：6 月 7 日（周六） 8:30-12:00 地点：四楼大湾区厅 1

时间	主持人	内容
8:30-9:00	谢国森	讲者：鲍秉坤（南京邮电大学） 题目：记忆机制启发的视频行为理解
9:00-9:30	谢国森	讲者：彭玺（四川大学） 题目：噪声关联学习
9:30-9:40	杨小汕	企业宣讲：阿里妈妈（葛铁铮） 题目：计算机视觉 in 阿里妈妈
9:40-10:10	杨小汕	讲者：舒祥波（南京理工大学） 题目：开放场景下人体行为智能计算
10:10-10:30	中场休息	
10:30-11:00	陈使明	讲者：张正（哈尔滨工业大学（深圳）） 题目：高效能跨模态关联分析
11:00-11:10	陈使明	企业宣讲：集智进化（代季峰） 题目：集智进化宣讲
11:10-11:40	陈使明	讲者：谢伟迪（上海交通大学） 题目：生成式视觉-语言模型研究

组织者：陈使明 阿联酋人工智能大学(MBZUAI)

个人简介：陈使明，阿联酋人工智能大学 (MBZUAI) 研究科学家。曾任 CMU/ MBZUAI 的博士后研究员。他于 2022 年在华中科技大学获得博士学位，入选国家“优培计划”、CSIG 优秀博士论文奖提名、华为学术之星等。研究兴趣包括零样本学习、视觉-语言学习。在 TPAMI、NeurIPS、ICML、CVPR、ICCV 等人工智能权威会议和期刊上发表了 20 余篇论文。担任 TPAMI、IJCV、ICLR、ICML、CVPR 等权威期刊和会议的审稿人，任 PRCV’23 和 VALSE’23-25 领域主席。

组织者：谢国森 南京理工大学

个人简介：谢国森，南京理工大学计算机科学与工程学院教授，国家级青年人才，江苏特聘教授。2016 年于中国科学院自动化研究所获工学博士学位，先后在新加坡、阿联酋留学和工作。研究方向为计算机视觉、开放环境复杂图像视觉理解等，在领域内国际期刊/会议发表论文 80 余篇，涵盖 TPAMI、IJCV、TIP、NeurIPS、CVPR、ICCV、ECCV 等。2023-2024 年度全球前 2%顶尖科学家榜单；获国际会议 MMM 2016 最佳学生论文奖。担任 IEEE TIP、Pattern Recognition 等权威期刊编委和 ICLR 的领域主席。承担多项国家级和省部级科研项目，包括国自然面上/青年基金，江苏省人才项目等。

组织者：杨小汕 中国科学院自动化研究所

个人简介：杨小汕，中国科学院自动化研究所多模态人工智能系统全国重点实验室研究员、博士生导师，国家优青。近年来聚焦开放环境多媒体内容理解开展研究，在相关领域已发表 80 余篇论文，其中 TPAMI、TMM、TIP 等 IEEE/ACM Trans.期刊和 MM、CVPR、NeurIPS、ICML 等 CCF-A 类会议 56 篇，获中科院院长奖、中科院优博、腾讯卓创奖，负责国家优秀青年基金项目、面上项目、青年基金项目、科技委重点项目课题，相关算法为腾讯、咪咕、航天二院提供了重要的技术支持。

鲍秉坤 南京邮电大学

报告题目：记忆机制启发的视频行为理解

报告摘要：视频行为理解是计算机视觉领域的核心任务之一，旨在通过从连续的视频序列中识别、定位和解释目标行为，实现上下文关联和语义理解。随着深度学习技术和大规模预训练模型的快速发展，现有研究在行为理解任务性能方面取得了长足的进步。然而，在应对长时序视频、流式在线视频等真实世界任务数据时，现有方法仍然面临时-空建模方式低效、判别信息难以稳定保持等瓶颈难题。针对上述挑战，受人类记忆系统中情景记忆与语义记忆协同机制的启发，首先构建情景记忆启发的提示学习策略，通过快速产生历史行为摘要，实现长程时-空跨度下高效行为定位；然后设计语义记忆辅助的知识迁移框架，通过离线教师模型与在线学生模型间的可靠知识蒸馏，实现面向动态流式视频的精准行为检测；最后探讨视觉-语言预训练模型与上述两种记忆机制间的关联，指导建立情景记忆与语义记忆协同的视频行为理解方法，并验证其在异常行为侦测、装配行为在线预测等任务上的有效性。

讲者简介：鲍秉坤，南京邮电大学计算机学院、软件学院、网络空间安全学院院长，教授、博士生导师、国家杰青、中组部万人青拔、江苏省杰青、江苏省双创人才。从事多媒体计算、社交多媒体、计算机视觉等领域研究，发表高水平论文 100 余篇；主持新一代人工智能国家科技重大专项、国家自然科学基金重点项目等 10 余项国家级、省部级项目；荣获 2018 年度电子学会科学技术（自然科学类）一等奖，ACM TOMM 2016 年度最佳论文奖、IEEE MM 2017 年度最佳论文奖、Multimedia Modeling 2019 年度最佳论文 Runner Up 奖。CSIG 女工委副主任、学工委秘书长、多媒体专委会常委。担任 IEEE TMM/TCSVT、ACM TOMM 等期刊编委。

彭玺 四川大学

报告题目：噪声关联学习

报告摘要：针对多智能体多传感器的数据智能分析一直是 AI 领域的研究重点之一。过去诸多研究一般隐式假设这些数据已跨时空对齐，不存在错误匹配数据。然而，现实情况中不同传感器的信号传输速率存在差异，不同设备的数据采集过程存在时空异步，无论是机器还是人工都难以保证数据是完全正确对齐的。这样文不对题、答非所问的噪声关联 (Noisy Correspondence)数据一旦被当成正确对齐的训练数据，将难以获得理想结果。本次报告从模态、数据样本、样本属性等不同粒度探讨噪声关联学习的最新进展，特别是其在跨模态检索、行为重识别、图匹配、大模型预训练、长视频定位及检索、机器阅读理解、多视图聚类等不同任务场景中的特有表现形式和解决方案。此外，此次报告还希望和大家就噪声关联学习未来的发展方向进行交流。

讲者简介：彭玺，四川大学教授，博导，教育部“CJ 学者”特聘教授、“工程数值模拟基础算法与模型”全国重点实验室副主任。研究方向为机器学习理论及其在多学科交叉领域上的应用（AI4Science），在 Nature Communications、JMLR、TPAMI、IJCV 等国际权威刊物上发表学术论文百余篇。

舒祥波 南京理工大学

报告题目：开放场景下人体行为智能计算

报告摘要：开放场景中，由于分散数据利用率低、监督训练样本量少、行为语义复杂性高等因素，给人体行为智能计算带来了新的挑战。基于此，本报告将探讨多样化开放场景下的人体行为智能计算研究任务，重点介绍课题组近年在边云协同的模型预训练与微调、数据受限的人体行为鲁棒表征、细粒度/多粒度行为分析与推理等方面的技术同，并简要介绍相关技术的推广应用。

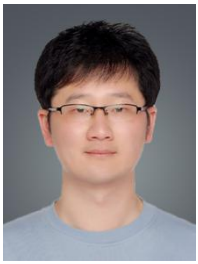
讲者简介：舒祥波，南京理工大学计算机科学与工程学院/人工智能学院教授、社会安全信息感知与系统工信部重点实验室副主任、国家优秀青年基金获得者、江苏省杰出青年基金获得者。近年主要研究兴趣为人体行为计算，在 TPAMI、CVPR、ICCV、ACM MM 等期刊/会议上发表论文 100 余篇，其中 ESI 高被引论文 8 篇；获中国电子学会自然科学一等奖、ACM MM 2015 最佳论文提名、MMM 2016 最佳学生论文奖、江苏省优博、中国人工智能学会优博、2024 年度江苏自然科学百篇优秀学术成果论文；入选全球前 2%顶尖科学家（2021-2024 年连续 4 年入选）；承担国家自然科学重点/面上/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、Pattern Recognition 等期刊编委。

张正 哈尔滨工业大学（深圳）

报告题目：高效能跨模态关联分析

报告摘要：大规模多源异构数据正在实时地产生、传输和处理，如何将多模态数据转化为智能，以实现高效能智能决策和自主分析是当前多模态人工智能研究的重点。本报告将汇报跨模态关联分析在多层次语义理解、跨模态精准对齐和可信赖模型推理等方面的最新研究成果，着重介绍本团队在跨模态关联分析在开放场景下数据、表征和模型的不确定性建模方法，以及高效大模型基础架构设计与微调技术，并对未来发展方向与趋势进行讨论与展望。

讲者简介：张正，哈尔滨工业大学（深圳）长聘教授，国家级青年人才，广东省珠江学者，深圳市优青。长期从事高效能多模态机器学习的研究，专注于高效与可信多模态大模型，出版中英文学术专著/编著 6 部，发表 IEEE/ACM 汇刊和 CCF A 类期刊/会议论文 100 余篇，谷歌引用一万余次。主持国家级和省部级自然科学基金、深圳市科技创新基金以及阿里巴巴创新研究计划、华为合作基金等科研项目 10 余项。受邀担任 IEEE TIFS、IEEE TAC、IEEE JBHI 等权威期刊编委，以组织委员会成员成功举办了多项权威学术会议，常年受邀担任 ICML、NeurIPS、ICLR、CVPR、ACM MM 等 A 类顶级学术会议的领域主席。



谢伟迪 上海交通大学

报告题目：生成式视觉-语言模型研究

报告摘要：近年来，生成式视觉-语言模型通过自由形式的文本和图像作为输入，能够以自然语言的方式与人类进行交互，展现出在统一处理复杂多模态任务方面的强大潜力。在本次报告中，我将分享生成式多模态大模型在以下几个方面的研究成果：1）多模态检索任务的统一处理：探索如何利用生成式框架统一解决多种多模态检索任务，实现无需额外训练即可处理未见复杂检索任务的强泛化能力。2）视频理解中的问答与定位：提出一种针对视频问答、多步推理和定位的统一框架，显著提升模型在长视频和实时流媒体处理中的性能。3）体育分析场景的应用：解析生成式大模型在体育视频理解中的应用案例，例如在足球视频分析中的表现。4）基于多智能体的视频问答：利用多智能体系统生成维链推理，通过知识蒸馏赋予生成式视觉-语言模型更强的时空逻辑推理能力，并刷新多项视频问答任务的性能记录。5）复杂医疗场景中的实践：探讨生成式模型在复杂医疗场景中的应用潜力与实际效果。

讲者简介：谢伟迪，上海交通大学长聘轨副教授，教育部 U40 获得者，国家级青年人才(海外)，上海市海外高层次人才计划获得者，上海市启明星计划获得者，科技部科技创新 2030—“新一代人工智能”重大项目青年项目负责人，国家基金委面上项目负责人。 博士毕业于牛津大学视觉几何组（Visual Geometry Group，VGG），首批 Google-DeepMind 全额奖学金获得者，China-Oxford Scholarship 获得者，牛津大学工程系杰出奖获得者。主要研究领域为计算机视觉，医学人工智能，共发表论文超 80 篇，包括 Nature Communications，NPJ Digital Medicine，CVPR，ICCV, NeurIPS, ICML, IJCV 等，Google Scholar 累计引用超 14500 余次，多次获得国际顶级会议研讨会的最佳论文奖和最佳海报奖，最佳期刊论文奖，MICCAI Young Scientist Publication Impact Award Finalist； Nature Medicine， Nature Communications 特邀审稿人，计算机视觉和人工智能领域的旗舰会议 CVPR，NeurIPS，ECCV 的领域主席。https://weidixie.github.io。

Workshop 4：低空智能感知与理解

组织者：丛润民（山东大学）、万人杰（香港浸会大学）、李锋（合肥工业大学）

时间：6 月 7 日（周六）8:30-12:00　　地点：四楼大湾区厅 2

时间	主持人	内容
8:30-9:00	丛润民	讲者：齐俊桐（上海大学） 题目：无人机集群感知控制技术与应用
9:00-9:30	丛润民	讲者：白慧慧（北京交通大学） 题目：无人机图像智能感知方法研究
9:30-10:00	万人杰	讲者：朱鹏飞（天津大学） 题目：低空智能感知关键技术及应用
10:00-10:20	中场休息	
10:20-10:50	万人杰	讲者：钟平（国防科技大学） 题目：无人机对地遥感目标识别与对抗
10:50-11:20	李锋	讲者：邹征夏（北京航空航天大学） 题目：三维遥感场景感知与新视角合成研究进展
11:20-11:50	李锋	讲者：赵健（中国电信集团） 题目：Anti-UAV：复杂场景低慢小目标智能感知

组织者：丛润民 山东大学

个人简介：山东大学教授、博士生导师，国家级青年人才、山东省泰山学者青年专家。主要研究方向包括模式识别与机器学习、计算机视觉、多媒体信息处理等。主持、参与了包括国家自然科学基金、国家重点研发计划、北京市科技新星计划在内的多项科研项目。在 TPAMI、IJCV、TIP、NeurIPS、CVPR、ICCV、ICML 等 CCF-A、IEEE/ACM Trans 论文 106 篇，ESI 热点论文 2 篇、ESI 高被引论文 18 篇，谷歌引用 12000 余次；授权国家发明专利 34 项。担任 IEEE Transactions on Neural Networks and Learning Systems 等 3 个 SCI 期刊编委，获中国图象图形学学会自然科学奖二等奖（序 1）、IEEE Chester W. Sall 奖、CVPR 运动引导视频目标分割挑战赛冠军、ECCV 指代视频目标分割挑战赛冠军、IEEE CVPR NTIRE 双目图像超分挑战赛亚军、IEEE ICME 最佳学生论文奖亚军、ACM SIGWEB 中国新星奖、中国图象图形学学会优秀博士学位论文奖、《信号处理》2020-2022 年度优秀论文奖等。

组织者：万人杰 香港浸会大学

个人简介：博士，香港浸会大学计算机科学系助理教授。2019 年于新加坡南洋理工大学 （NTU）交叉科学研究院获得博士学位。研究方向主要包括人工智能安全和计算成像。曾获得优秀博士论文奖（2019 年），瓦伦堡-南洋理工大学校长博士后奖学金（2020 年），VCIP 最佳论文奖（2020 年）。他目前为多个顶级期刊和会议审稿，并担任 ACM MM2024 及 2025 的领域主席。

组织者：李锋 合肥工业大学

个人简介：合肥工业大学副教授。主要研究方向为计算机视觉、图像视频增强及多媒体信息处理。主持国家自然科学基金青年基金，国家自然科学基金重点项目课题负责人，参与包括科技部重点研发计划等多项项目。获 IEEE CVPR NTIRE 双目图像超分挑战赛亚军。在包括 IEEE TPAMI、IJCV、CVPR、ICCV、ICLR 等 CCF-A 类会议、IEEE 汇刊上发表论文 30 余篇。

齐俊桐 上海大学

报告题目：无人机集群感知控制技术与应用

报告摘要：各类无人机系统近年来已经在生产生活各领域发挥着越来越重要的作用。随着工作任务的复杂性以及所处环境的动态不确定性，无人机向着智能化、自主化、集群化发展，无人机集群控制将成为该领域发展的重要使能技术。本报告将以无人机集群技术为重点，结合团队的研究进展，从集群感知与控制技术的特点、难点、技术突破、应用等方面进行分享。

讲者简介：齐俊桐，教授、博士生导师，上海大学人工智能研究院副院长，天津大学机器人研究所副所长。全国劳动模范，中国青年五四奖章获得者，国家万人计划领军人才，天津市第十七届、十八届人民代表大会代表。多年来致力于机器人与无人系统自主控制与集群控制技术研究，主持国家 973、863 重点项目、自然基金重点项目等 40 余项，获吴文俊人工智能科技进步一等奖、中国专利优秀奖、辽宁省科技进步一等奖、辽宁省自然学术成果一等奖等奖励。

白慧慧 北京交通大学

报告题目：无人机图像智能感知方法研究

报告摘要：传统图像感知研究通常局限于单一层次的分析框架，未能充分考虑无人机视角下特有的复杂多变的感知环境。无人机图像感知不仅是对常规图像感知范畴的扩展，更是对其进行了更为精细的子类划分。相较于传统图像感知任务，无人机图像感知面临诸多独特挑战，主要体现在：目标纹理特征的特殊性、外部环境的高度动态性、数据集的局限性以及任务层级的多样性等方面。具体而言，视点变化、光照条件差异、复杂背景干扰以及目标遮挡等因素，都给特殊目标的准确识别带来了显著困难。值得注意的是，无人机执行的顺序任务与并行任务往往呈现出明显的层次关联特征，这使得仅从单一层次入手的传统图像感知方法难以取得理想效果。本报告将系统性地介绍我们团队近年来在无人机图像智能感知领域所取得的研究进展和创新方法。

讲者简介：白慧慧，北京交通大学教授、博士生导师。主要研究方向是图像视频编码和增强等。已发表学术论文 70 余篇，包括 IEEE 汇刊 TPAMI 与 CCF A 类会议论文 CVPR 等 30 余篇。获国际专利授权 3 项、国家发明专利授权 15 项。曾主持国家自然科学基金重点/面上/青年项目、北京市自然科学基金-小米创新联合基金项目、江苏省自然科学基金项目等。获北京市自然科学一等奖、北京市科技进步二等奖、中国电子学会青年科学家奖、中国产学研合作创新成果奖二等奖等，入选中国人工智能学会教学激励计划一类成果、中国图象图形学学会教学激励计划二类成果。

朱鹏飞 天津大学

报告题目：低空智能感知关键技术及应用

报告摘要：智能无人系统依赖于多传感器对周围环境进行鲁棒的环境感知。团队构建了 VisDrone 大规模无人机视觉数据平台，包括可见光数据、双光数据以及多机协同数据等，覆盖目标检测、目标跟踪、群体分析和协同感知等任务。基于 VisDrone 数据平台，团队围绕数据算力受限条件下的低代价学习范式、多机多传感器不同步条件下的协同学习机理以及未知场景和类别条件下的进化学习机制开展研究，未来将主要聚焦无人机具身智能理论与方法，并在军事安防等场景开展应用。

讲者简介：朱鹏飞，天津大学智能与计算学部教授，国家优青和天津市杰青获得者。主要研究方向是智能无人机，构建了大规模无人机视觉开放数据平台 VisDrone，包含超过 2000 万图像/视频帧和 2000 万目标标注。已在 IEEE TPAMI 和 IJCV 等 CCF A 类和 IEEE 汇刊发表论文 80 余篇。获吴文俊人工智能科技进步一等奖等奖励。主持科技创新 2030-“新一代人工智能”重大项目和国家自然科学基金重点项目 10 余项。

邹征夏 北京航空航天大学

报告题目：三维遥感场景感知与新视角合成研究进展

报告摘要：基于多无人机和卫星的多视角图像对复杂场景进行三维渲染与分析，在实景三维构建、低空智能感知等方面有重要研究与应用价值。本次报告中，讲者将围绕三维遥感场景视角稀疏、场景尺度大的特点，介绍课题组近期在该方向的研究工作，包括稀疏视角和大范围 3D 遥感场景的 BEV 感知、Occupancy 感知、隐神经场新视角合成、BlockGaussian 新视角合成等。最后，将对该方向的潜在应用和未来方向进行讨论。

讲者简介：邹征夏，北京航空航天大学教授、博士生导师，国家级青年人才。主要研究方向为遥感图像处理、深度学习，以第一/通讯作者身份在 Proceedings of the IEEE、Nature Communications、TPAMI、TIP、CVPR、ICCV 等期刊和会议发表论文 30 余篇，论文引用 8000 余次，主持国家自然科学基金海外人才项目、面上项目。担任 IEEE 高级会员、TIP 期刊编委（Associate Editor）、Nature 旗下期刊 Communications Engineering 特刊编辑。研究成果被国家自然科学基金委、新华社、新科学人、麻省理工科技评论等媒体报道，收录于斯坦福大学著名公开课 CS231n 深度学习与计算机视觉，成果应用于高分一号、巴遥一号、委遥一号、高景一号 03 星等遥感卫星。

钟平 国防科技大学

报告题目：无人机对地遥感目标识别与对抗

报告摘要：无人机载遥感系统凭借其灵活性、高效性和低成本等优势，广泛应用于智能交通、场景监控、资源勘探等领域。然而，作为这些典型应用场景的核心关键技术，复杂环境下的目标识别面临诸多挑战。本次讲座将重点针对传感器弱对齐、环境高复杂、目标多变化、场景非通视等实际应用场景面临的难点问题，介绍无人机载对地遥感目标多模态特性分析，及其在多模态图像融合目标识别、跨模态目标重识

别、水下目标识别等领域的应用方法，并探讨针对典型无人机载智能目标检测识别模型的对抗攻击方法，实现对重要目标的有效防护。

讲者简介：钟平，国防科技大学自动目标识别全国重点实验室研究员，博士生导师。英国剑桥大学访问学者。全国优秀博士学位论文获得者，入选学校首批拔尖创新人才培育对象和教育部新世纪优秀人才支持计划，立二等功一次。某领域主题专家，中国图象图形学学会第六、七届理事，国家自然科学基金委、教育部学位中心、湖南省和江西省科技厅等单位评审专家，IEEE Senior Member。主要从事人工智能基础理论及其在卫星/无人机等多平台多模态图像分析和智能目标识别方面的应用研究。先后主持国家自然科学基金、装备重大基础研究项目课题、装备应用创新项目、国防 973 项目专题、“高分”重大专项课题、国防预研基金重点项目等国家和军队重要科研项目 20 余项。近年来以第一作者和通信作者发表相关学术论文 100 余篇，SCI 收录论文 60 余篇，出版学术专著 1 部。国际期刊 IEEE J-STARS、《电子学报》、《航空兵器》、《国防科技大学学报》等国内外期刊编委。

赵健，中国电信集团

报告题目：Anti-UAV：复杂场景低慢小目标智能感知

报告摘要：近年来，商用小型无人机飞速发展，其相比于载人机而言，具有体积小、成本低、机动性强等优势，可完成一些载人机无法完成的任务，已被广泛应用于航拍、勘探、救援、物流等诸多领域。然而，也有不法分子利用无人机对敏感区域进行侦查/监视或携带危险物品/武器对重要人物进行毁伤。因此，反无人机（Anti-UAV）技术的研究变得至关重要。我们首次在国际学界提出了“Anti-UAV”这一任务，并围绕该主题推出了一系列相关数据集，这些数据集为反无人机跟踪算法的训练和评估提供了重要平台。此外，我们还依托国际顶级会议（如 CVPR、ICCV）组织了四届 Workshop & Challenge，吸引了全球众多研究团队参与，极大地推动了该领域的研究进展。在技术层面，针对复杂场景下低慢小目标检测跟踪的难点挑战，我们也提出了相应的创新方法，有效提升了算法性能，技术成果成功应用于多个国家重要部门和重大活动。未来，我们将继续致力于提升反无人机系统的智能化水平，以应对日益复杂的无人机威胁。

讲者简介：赵健，中国电信人工智能研究院（TeleAI）多媒体认知学习实验室主任、资深研究科学家，北京图象图形学学会理事，国际知名期刊《Pattern Recognition》编委等，隶属李学龙教授（中国电信集团 CTO、首席科学家、TeleAI 院长）团队，2019 年博士毕业于新加坡国立大学，主要研究领域为临地安防、AI 治理。共发表 CCF-A 类国际期刊和会议论文 40 余篇，获得了国内外学术界的广泛关注和好评。入选了中国科协、北京市科协“青年人才托举工程”，并获 2023 年度中国人工智能学会吴文俊人工智能优秀青年奖、2022 年度中国人工智能学会吴文俊人工智能自然科学奖一等奖（第二完成人）、CCF-A 类国际会议 ACM MM 唯一最佳学生论文奖（一作，2018），8 次在国内外技术竞赛中夺冠。相关技术成果不但服务于多个国家重要部门、发挥了重要作用，同时也在中国电信、蚂蚁金服等 7 个科技领军企业得到应用，产生了显著效益。

Workshop 5: 大场景多对象的重建与生成

组织者：李坤（天津大学）、季梦奇（北京航空航天大学）、马健（天津大学）

时间：6 月 8 日（周日）8:30-12:00　地点：四楼大湾区厅 2

时间	主持人	内容
8:30-9:00	李坤	讲者：王程（厦门大学） 题目：城市空间中激光雷达感知定位与人身份识别
9:00-9:30	李坤	讲者：刘子纬（南洋理工大学） 题目：From Multimodal Generative Models to Dynamic World Modeling
9:30-10:00	季梦奇	讲者：马楠（北京工业大学） 题目：大场景下的移动智能体多对象目标检测与配准定位技术
10:00-10:30	季梦奇	讲者：李坤（天津大学） 题目：大场景下的群体三维重建与生成
10:30-11:00	马健	讲者：马月昕（上海科技大学） 题目：面向人-机-环境协同共生的感知与生成
11:00-11:30	马健	讲者：温建伟（北京拙河科技有限公司） 题目：亿像素光场重建与智能分析在产业界的应用探索
11:30-12:00	马健	Panel 嘉宾：王程（厦门大学）、刘子纬（南洋理工大学）、马楠（北京工业大学）、李坤（天津大学）、马月昕（上海科技大学）、温建伟（北京拙河科技有限公司）

组织者：李坤 天津大学

个人简介：李坤，天津大学英才教授，国家优青、天津市杰青、中国图象图形学学会石青云女科学家奖获得者。担任天津市人工智能学会副秘书长、SCI 一区期刊 CAAI TRIT 的编委、ACM MM 2021 大会领域主席等职务。以第一作者或通讯作者在知名国际期刊会议上发表论文 70 余篇，其中 IEEE Trans.长文/SCI 一区/CFF A 类期刊会议论文 38 篇（9 篇影响因子>11）。获多媒体领域知名国际会议 ICME 最佳论文奖（获奖率 0.8%），专利授权 31 项（转让 9 项）。主持国家重点研发计划等 17 项科研项目。

组织者：季梦奇 北京航空航天大学

个人简介：季梦奇，北航人工智能学院副教授。2019 年至 2021 年于清华大学成像与智能技术实验室从事博士后研究。本科与硕博分别毕业于北京科技大学和香港科技大学，期间曾访问德国慕尼黑工大、波恩大学、马普所等研究机构。围绕光场计算重建和弱小目标感知领域的基础问题开展多年研究，代表性成果包括 2 篇《Nature》子刊以及 IEEE T-PAMI、ICCV 等期刊会议。曾入选国家青托、北航小米学者、北航青拔人才等，曾主持国自然面上、博后基金特别资助、博后面上，以及启元、航天科技、中电科等项目。

组织者：马健 天津大学

个人简介：马健，天津大学智能与计算学部助理研究员，博士毕业于英国布里斯托大学（QS 世界大学排名第 55 名），获英国工程和物理科学研究委员会（EPSRC）资助。主持中国博士后面上资助项目，多篇学术论文发表在 IJCV、IEEE T-II、NeurIPS 等人工智能、计算机视觉等领域国际顶级期刊会议上，1 项技术应用在产业界，申请国家发明专利 5 项，获专利授权 1 项。

王程 厦门大学

报告题目：城市空间中激光雷达感知定位与人身份识别

报告摘要：城市空间智能的基础是准确、实时得到城市复杂场景中的人和车等动态要素的态势，激光雷达是获取城市空间态势信息的重要传感器。本报告将针对激光雷达感知定位和激光雷达人体动捕与生物识别两个前沿问题，介绍厦门大学 ASC 实验室的研究进展。在感知定位方面，隐式表达的类脑激光雷达视觉定位技术正从实验室验证向城市级规模应用推进，其核心突破体现在大范围场景建模效率与复杂环境鲁棒性的双重提升。在人体身份识别方面，激光雷达被证明是多人动作捕捉和身份重新识别的有效传感器，能够实现遮挡和人-物互动情况下的动作捕捉，进而能够实现在换装等复杂条件下的人体步态识别和身份识别。将二者的结合更加产生了对大范围城市场景中的人群态势感知的潜力。

讲者简介：王程，厦门大学南强重点岗位教授，国家级人才计划基金获得者，入选国家“万人计划”科技创新领军人才。是现任福建省智慧城市感知与计算重点实验室主任。研究兴趣包括计算机三维视觉，激光雷达，遥感智能处理，空间大数据分析，智慧城市。在 Nature Communication，ISPRS-JPRS，IEEE TGRS，CVPR，NeurIPS 等顶级期刊和会议上发表 300 余篇论文，被引用超过 14000 次。

刘子纬 南洋理工大学

报告题目：From Multimodal Generative Models to Dynamic World Modeling

报告摘要：Beyond the confines of flat screens, multimodal generative models are crucial to create immersive experiences in virtual reality, not only for human users but also for robotics. Virtual environments or real-world simulators, often comprised of complex 3D/4D assets, significantly benefit from the accelerated creation enabled by Gen AI. In this talk, we will introduce our latest research progress on multimodal generative models for objects, avatars, scenes, motions, and ultimately dynamic world models.

讲者简介：刘子纬，新加坡南洋理工大学助理教授，并获得南洋学者称号（Nanyang Assistant Professor）。他的研究兴趣包括计算机视觉、机器学习与计算机图形学。他在国际顶级会议及期刊（CVPR / ICCV / ECCV / NeurIPS / ICLR / TPAMI / TOG / Nature - Machine Intelligence）上发表文章 100 余篇，总引用量 2.8 万余次，获得专利 50 余项。他领导搭建了数个国际知名的基准数据库，例如 CelebA 和 DeepFashion 等。同时他也领导数个广泛使用的开源软件建设，例如 MMFashion 和 MMHuman3D 等。他获得过多个领域内奖项，包括微软小学者奖、香港政府博士奖、ICCV 青年学者奖、HKSTP 最佳论文奖、CVPR 最佳论文候选和 WAIC 云帆奖等。他是国际顶级会议 CVPR、ICCV、NeurIPS 和 ICLR 的领域主席（Area Chair）以及国际顶级期刊 IJCV 的编委（Associate Editor）。

马楠 北京工业大学

报告题目：大场景下的移动智能体多对象目标检测与配准定位技术

报告摘要：针对大场景下移动智能体（如无人车、无人机）的动态感知与高精度定位需求，本报告提出一种融合动态视角实时位姿估计与静态地图协同配准的创新技术框架。通过集成多传感器（相机、LiDAR、IMU）与智能体动力学模型，实现低成本、高频率的相机位姿实时解算，为动态视角感知提供时空一致性基础；同时，基于移动智能体预先采集的多模态数据构建高精度静态语义地图，并在实时任务中使用一种基于特征筛选和局部-全局优化的点云配准框架，利用点云几何特征进行高特征耦合筛选并利用局部适应性关键区域聚合和全局模态一致性融合进行点云位姿匹配，形成了一种新的从粗配准到精配准的跨模态点云配准流程，解决了大场景下点云配准精度受点云密度、精度、噪声差异影响的难题。

讲者简介：马楠，北京工业大学教授，博士生导师，青年北京学者，信息科学技术学院副院长，智能感知与自主控制教育部工程研究中心副主任，国家重点研发计划项目负责人，国家一流本科课程负责人，中国人工智能学会副秘书，北京市智能制造与机器人技术创新专项负责人，研究方向为交互认知、具身智能、无人驾驶与移动机器人。先后以第一完成人获得中国图象图形学学会科技进步一等奖、中国电子学会科学技术奖【技术发明类】二等奖，主持多项国家、省部级项目，承担北汽集团、东风悦享、云迹科技等 10 余项企业委托智能交互系统项目。带领团队多次在国内外人工智能、无人驾驶比赛中获得冠军，团队成果“无人驾驶云智能交互系统”获第二届中国“AI+”创新创业大赛总决赛特等奖；获第六届全国教育科学研究优秀成果奖二等奖和北京市教学成果一等奖等。

李坤 天津大学

报告题目：大场景下的群体三维重建与生成

报告摘要：大场景下的人群精准三维定位与姿态形状重建是实现公共安全精准分析与预警的关键。在动态大场景视频中，对象数量多、分布范围广、尺度差异大、遮挡频繁、且单视角存在深度歧义，因此，如何建立遮挡鲁棒、空时一致、全局定位的动态多对象重建方法是关键问题。本报告将重点介绍本研究组在动态大场景群体三维重建与生成方面的探索：从静态重建到动态重建、从特定相机到通用相机、从离线处理到在线推理、从群体重建到群体生成，逐步深入探索的方法创新与实践突破。

讲者简介：李坤，天津大学智能与计算学部英才教授、博士生导师。主要研究方向为三维视觉，尤其是以人为中心的智能重建与生成。以第一作者/通讯作者在国际知名期刊和会议上发表论文 70 余篇，部分研究成果实现了产业化应用。荣获中国图象图形学学会石青云女科学家奖、ICME'17 最佳论文奖等荣誉。主持了国家自然科学基金优秀青年科学基金、天津市杰出青年科学基金、国家重点研发计划青年科学家项目等 17 项科研项目。担任天津市人工智能学会副秘书长、ACM MM 2021 大会领域主席、Fundamental Research 等国际

期刊青年编委、VALSE 2022 大会本地主席等职务。

马月昕 上海科技大学

报告题目：面向人-机-环境协同共生的感知与生成

报告摘要：在通用人工智能迅速发展的当下，人-机-环境协同共生的研究步入关键发展时期。其核心理念为突破人、机、环境间的物理隔阂，构建起高效互动、深度交融的生态体系。在这一体系中，以人为中心的通用场景的感知、认知与规控技术，成为实现协同共生的关键支撑要素。本报告聚焦于人-机-环境共融领域的前沿研究进展，深入探究以人为中心的感知与生成的算法研究及其在具身智能中的应用，同时梳理讨论技术落地实践的成效与面临的挑战。

讲者简介：马月昕，上海科技大学研究员、助理教授、博导，博士毕业于香港大学。主要研究方向为三维视觉、具身智能、自动驾驶。共发表相关领域顶会或顶刊论文 80 余篇，其中一作与通讯论文 40 余篇，包括 Science Robotics、TPAMI、CVPR、ICCV、ECCV、SIGGRAPH 等，谷歌学术引用 5000 余次。参与指导的论文获 MICCAI 2024 唯一最佳论文奖，ACM MM 2024 最佳论文候选。曾获上海市海外高层次人才，上海市优秀教学成果（高等教育类）一等奖，曾获 SemanticKITTI、NuScenes、Argoverse 等多个国际自动驾驶挑战赛冠军和亚军。

温建伟 北京拙河科技有限公司

报告题目：亿像素光场重建与智能分析在产业界的应用探索

报告摘要：十亿像素（Giga）光数据记录、光场重建，以及基于 Giga 的视频大模型智能分析是计算摄像领域的热点和难点，在军事智能、安全生产、文娱直播等场景下均具有广泛的需求，本报告介绍了采用集成化的异构、分布式光学采集前端，计算光场全景并实时呈现，进一步地利用重建的全景光场进行全域智能分析的技术路线和工程实践，通过智慧城市运营、低空经济基础设施、交通能源等行业应用，展现 Giga 的技术发展趋势和广阔前景。

讲者简介：温建伟，正高级工程师，清华大学博士，中国指挥与控制学会低空产业工作委员会常务委员，中国人工智能学会元宇宙专委会委员，中国电子学会消费电子分会委员，中国合格评定国家认可委员会（CNAS）评审专家，北京市科技委奖励办评审专家。参与制定国家重大发展规划多项，发表学术论文 10 余篇，申请发明专利 50 余篇。现任北京拙河科技有限公司创始人、董事长，负责公司的战略规划、资本运作及业务开拓。获得 2019 年中国电子学会科技进步一等奖，2022 年北京市技术发明一等奖。

Workshop 6: 视觉通用模型

组织者：王立君（大连理工大学）、李冠彬（中山大学）、黄岩（中科院自动化所）

时间：6 月 8 日（周日）8:30-12:00　地点：四楼大湾区厅 1

时间	主持人	内容
8:30-9:00	王立君	讲者：王兴刚（华中科技大学） 题目：重建 vs 生成 - 解决潜在扩散模型中的优化困境
9:00-9:30	王立君	讲者：董超（中国科学院深圳先进技术研究院） 题目：通用底层视觉前沿探索
9:30-9:50	李冠彬	企业宣讲：OPPO 题目：面向 AI 手机的多模态生成与问答探索实践
9:50-10:20	李冠彬	讲者：代季峰（清华大学） 题目：多模态基础模型研究
10:20-10:40	中场休息	
10:40-11:10	黄岩	讲者：王栋（大连理工大学） 题目：通用视觉感知模型初探
11:10-11:40	黄岩	讲者：陈隆（香港科技大学） 题目：Narrowing the Gaps: Towards Real-World Multimodal Reasoning & Generation Models

组织者：王立君 大连理工大学

个人简介：王立君，大连理工大学未来技术/人工智能学院教授，博士生导师，国自然优秀青年基金获得者，主要研究方向聚焦于图像深度估计、视频理解、多模态大模型等。主持国自然联合重点、面上和青年项目，入选人社部“博士后创新人才支持计划”和大连市“科技人才创新支持计划”，在本领域顶级学术会议和期刊发表论文 40 余篇，谷歌学术总引用 1 万余次。相关研究成果获得辽宁省科技进步一等奖，中国图象图形学会自然科学二等奖，教育部自然科学二等奖，中国图象图形学学会优秀博士论文奖，以及辽宁省优秀博士论文奖。连续三年获得 VOT 国际视觉跟踪竞赛冠军。担任中国科协《科技导报》青年编委，VALSE 执行委员，CCF-CV 与 CSIG-MV 专委会执行委员等。

组织者：李冠彬，中山大学

个人简介：李冠彬，中山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。主要研究领域为人工智能领域的图像视频内容理解与生成。 迄今为止累计发表计算机学会 A 类/中科院一区论文 200 余篇，谷歌学术引用超过 15700 次，入选全球前 0.05%顶尖科学家榜单。曾获得中国图象图形学学会青年科学家奖、吴文俊人工智能优秀青年奖、ACM 中国新星提名奖、中国图象图形学学会科学技术一等奖、ICCV2019 最佳论文提名奖、CVPR2024 最佳论文候选、ICMR2021 最佳海报论文奖等荣誉。主持了包括国家自然科学基金优青、面上、青年、广东省杰青、CCF-腾讯犀牛鸟科研基金、CCF-快手科研基金、华为科研合作基金、美团北斗科研合作基金等 10 多项科研项目。担任广东省图象图形学会计算机视觉专委会主任、中国图象图形学学会青工委副秘书长、中国计算机学会青年科技论坛广州主席、广州计算机学会副秘书长。担任人工智能领域顶级会议 CVPR、ECCV、AAAI 等领域主席或高级程序委员，获得 8 项人工智能领域国际顶级会议竞赛冠军，研究成果应用于智能交通分析、智慧医疗诊断、数字人驱动的智慧教育等。

组织者：黄岩 中科院自动化所

个人简介：黄岩，国家优青，中科院自动化所副研究员。研究方向为多模态认知与机器人，在相关领域的国内外期刊和会议上发表论文共计 100 余篇，曾获国内外学术会议最佳论文奖 3 项、国内外主流竞赛冠军 4 项，担任 IEEE TIP 编委、CVPR 领域主席、10 余次国内外研讨会的组织主席。曾获得北京市自然科学一等奖、中国图象图形学学会青年科学家奖、中国科学院院长特别奖、NVIDIA 创新研究奖等。

王兴刚 华中科技大学

报告题目：重建 vs 生成 - 解决潜在扩散模型中的优化困境

报告摘要：我们的研究尝试解决潜在扩散模型中重建与生成的优化困境：通过视觉基础模型对齐的 VA-VAE（VF Loss）约束高维潜在空间，结合改进的 LightningDiT 架构，在 ImageNet 256×256 生成任务上实现 1.35 的 FID 新纪录，仅用 64 个训练周期（比原 DiT 快 21 倍），首次在不增加计算成本下同时优化了重建与生成性能。

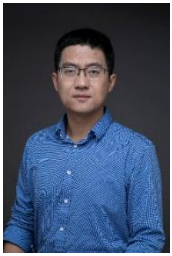
讲者简介：王兴刚，华中科技大学电信学院教授博导，国家级青年人才，现任 Image and Vision Computing 期刊共同主编。主要从事视觉表征学习、目标检测分割跟踪等领域研究，谷歌学术引用 3 万 8 千余次，1000+引用论文 6 篇，获湖北青年五四奖章、CSIG 青年科学家奖，CVMJ 最佳论文奖，MIR 期刊最被引用论文奖等。

董超 中国科学院深圳先进技术研究院

报告题目：通用底层视觉前沿探索

报告摘要：在通用人工智能 AGI 的背景下，研究通用底层视觉是大势所趋。但通用底层视觉仍然是一个全新的概念，它有哪些含义呢？理想的通用底层视觉算法应该可以实现任意图像空间之间的映射，这里包含了四个通用：场景通用——可以处理任意类型的输入图像；任务通用——可以生成任意类型的输出图像；数据通用——可以学习到任意图像空间的分布；模态通用——可以通过语言进行控制和交互。能实现这四种通用的算法就是理想的通用底层视觉算法。本次报告将会分享通用底层视觉方向的前沿探索成果，并为大家带来作者的新书《底层视觉之美——高清大片背后的人工智能》。

讲者简介：董超，博士生导师，中国科学院深圳先进技术研究院研究员，深圳理工大学教授，上海人工智能实验室双聘领军科学家。主要研究方向为底层计算机视觉，包括图像超分辨率、去噪和增强等，发表相关论文 100 余篇，谷歌引用量超过 4 万次。2014 年，在欧洲计算机视觉大会（ECCV）上发表论文 SRCNN，首次将深度学习引入图像超分辨领域。2017 年至今，多次带队参加国际图像超分辨率比赛，共获得 9 项冠军。2016 年-2018 年就职于商汤科技，带领商汤超分团队开发了世界首款基于深度学习的数码变焦软件。2021 年被斯坦福大学评选为世界前 2%顶尖科学家。2022 年被清华大学评为 AI 2000 人工智能全球最具影响力学者。2023 年获得上海市技术发明一等奖。



代季峰 清华大学

报告题目：多模态基础模型研究

报告摘要：在我们迅速发展的数字世界中，机器理解、解释和创造内容的能力是一个引人入胜的关键主题。今天，我们正见证一个非凡的时代，大型基础模型不仅仅是处理信息，它们正在学习理解和生成具有惊人精度和创造力的复杂语言和图像内容。多模态基础模型，正在重塑我们对人工智能能力的理解。这些模型无缝集成了多种形式的数据，如文本和视觉，它们不仅仅是工具，而是合作伙伴，增强人类的创造力，扩展机器能够实现的领域。在这次报告中，我们将探索这些模型的复杂工作原理，并报告我们研究团队在这个方向上的最新进展。我们将穿越语言和图像的领域，理解这些模型如何理解我们和我们的世界。

讲者简介：代季峰，清华大学电子工程系副教授，博士生导师。在 2009 年和 2014 年于清华大学自动化系分别获得工学学士和博士学位，博士导师周杰教授。2014 年至 2019 年在微软亚洲研究院视觉组工作，担任首席研究员、研究经理。2019 年至 2022 年在商汤科技研究院工作，担任执行研究总监。2022 年 7 月全职加入清华大学电子工程系。他的研究兴趣包括计算机视觉、深度学习等。他在相关领域发表国际期刊、会议文章 80 余篇，论文总引用 5 万余次。以可变形卷积为代表的多篇论文被选入深度学习权威框架 PyTorch 成为标准算子，在物体识别领域有较大影响力。他连续两年获得物体识别领域权威的 COCO 比赛冠军，之后历届冠军系统也使用了他提出的算法。他提出的算法获得自动驾驶感知领域权威的 Waymo 2022 竞赛冠军，获得 CVPR 2023 最佳论文奖。他是视觉领域顶刊 TPAMI 和 IJCV 的编委，和视觉领域顶会 NeurIPS , ICCV，CVPR，ECCV, ICLR 的领域主席，ICCV 2019 的宣传主席。

王栋 大连理工大学

报告题目：通用视觉感知模型初探

报告摘要：随着深度学习模型的快速发展和高质量标注数据的逐步丰富，视觉感知模型逐步呈现出从专用化向通用化的发展趋势。本报告主要介绍团队在通用视觉感知模型上的一些初步探索，如统一四个跟踪/分割相关任务的 Unicorn 模型、统一十个实例感知任务的 UniNext 模型、统一多样分割相关任务的 Spider 模型以及统一多模态视觉跟踪的 SUTrack 模型等。最后，介绍当前通用视觉感知模型存在的问题和未来潜在趋势。

讲者简介：王栋，大连理工大学信息与通信工程学院，教授、博导。迄今在本领域顶级会议(CVPR/ICCV)及期刊(TPAMI/IJCV)发表论文 50 余篇，Google Scholar 引用 1.5 万余次。获得国际视觉目标跟踪竞赛 VOT 冠军(10 次)、CCF 自然科学二等奖、教育部自然科学二等奖、CVPR2020 最佳论文提名等学术奖励。研究工作获得国家自然科学基金优秀青年科学基金、区域联合重点项目等资助。

陈隆 香港科技大学

报告题目： Narrowing the Gaps: Towards Real-World Multimodal Reasoning & Generation Models

报告摘要： Today's pretrained foundation models have demonstrated astonishing abilities in different applications. Hundreds of foundation models have been proposed during the past few years. Although significant progress has been achieved, there are still several challenges in designing stronger but more efficient foundation models. In this talk, I am going to share a series of recent works on multimodal reasoning and generation, which tries to narrow the gaps between today's foundation model research and real-world applications.

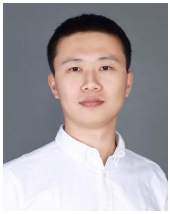
讲者简介： Dr. Long Chen is an assistant professor at the Computer Science and Engineering (CSE) department, Hong Kong University of Science and Technology (HKUST). He is leading the research group: LONG Group (https://long-group.cse.ust.hk/). Before joining HKUST, he was a postdoctoral research scientist at Columbia University. He obtained his Ph.D. degree from Zhejiang University, and he was also a visiting student at NTU & NUS. His primary research interests are Computer Vision, Machine Learning, and Multimedia. Specifically, he aims to build an efficient multimodal AI system that can realize "human-like" multimodal understanding and generation. By "human-like", we mean that the vision systems should be equipped with three types of abilities: 1) Explainable: The model should rely on (right) explicit evidences when making decisions, i.e., right for the right reasons. 2) Robust: The model should be robust to some situations with only low-quality training data (e.g., training samples are biased, noisy, or limited). 3) Universal: The model design is relatively universal, i.e., it is expected to be effective for various tasks.

Workshop 7：多模态认知计算

组织者：胡迪（中国人民大学）、刘武（中国科学技术大学）、赵健（中国电信集团）

时间：6 月 7 日（周六）13:30-17:00　　地点：四楼大湾区厅 1

时间	主持人	内容
13:30-14:00	胡迪	讲者：谢洪涛（中国科学技术大学） 题目：大模型时代下的场景文本检测、识别与编辑
14:00-14:30	胡迪	讲者：魏云超（北京交通大学） 题目：视觉智能推理
14:30-14:40	胡迪	企业宣讲：美团（张勇） 题目：本地生活场景中视觉生成的研究与应用
14:40-15:10	刘武	讲者：梁小丹（中山大学） 题目：“慢思考”在多模态基础模型与具身智能中的探索
15:10-15:40	刘武	讲者：沈为（上海交通大学） 题目：视觉语言的细粒度对齐
15:40~15:50	中场休息	
15:50-16:20	赵健	讲者：王正（武汉大学） 题目：视觉隐匿感知与对抗
16:20-16:50	赵健	讲者：徐文超（香港科技大学） 题目：多模态学习中的模态竞争及解决方案



组织者：胡迪 中国人民大学

个人简介：胡迪，中国人民大学高瓴人工智能学院副教授，博导。主要研究方向为机器多模态感知、交互与学习，以主要作者在TPAMI/ICML/CVPR/CoRL 等人工智能顶级期刊及会议发表论文 40 余篇。曾入选 CVPR Doctoral Consortium；荣获 2020 中国人工智能学会优博奖；荣获 2022 年度吴文俊人工智能优秀青年奖；入选第七届中国科协青托计划等。担任 AAAI、IJCAI Senior PC 等。

组织者：刘武 中国科学技术大学

个人简介：刘武，中国科学技术大学特任教授，博导，入选国家级青年人才，在重要国际会议和期刊上发表论文 100 余篇，曾获得 IEEE T-MM 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，以及天津市科技进步特等奖、ACM 中国新星奖、中科院优秀博士论文奖等，入选了《麻省理工科技评论》亚太区“35 岁以下科技创新 35 人”，北京市科技新星计划，斯坦福全球前 2%顶尖科学家榜单等，并担任了 IEEE T-MM Associate Editor，IEEE ICME 2022 和 ACM MM Asia 2021 技术委员会主席，IET Fellow 评审委员会委员等，主办/协办 20 多场国际顶级会议的多模态学习讲习班（Tutorial）和研讨会（Workshop）。

组织者：赵健 中国电信集团

个人简介：赵健，中国电信人工智能研究院(TeleAI)多媒体认知学习实验室(EVOL)主任、资深研究科学家，西北工业大学光电与智能研究院研究员、博士生导师，清华大学自动化系、复旦大学/浙江大学/中国科学技术大学计算机系博士生企业导师，博士毕业于新加坡国立大学，研究领域：多模态 AI Agent、AI+文旅、临地安防。

谢洪涛 中国科学技术大学

报告题目：大模型时代下的场景文本检测、识别与编辑

报告摘要：近年来，多模态大语言模型开始在文本图像处理的多个领域中崭露头角。本报告主要探索大模型时代下场景文本图像处理任务的研究趋势，分为大规模预训练和多任务统一模型两个方向。在大规模预训练上，讲述了基于自监督预训练、大模型蒸馏和扩散模型引导的训练技术。在多任务统一模型上，讲述了基于解耦表征学习的统一文本图像处理技术。通过介绍这一领域的最新技术进展并结合本实验室的相关研究成果，共同探讨场景文本检测、识别、编辑领域的技术难题，并对大模型时代下文本图像处理领域的技术发展趋势与未来研究方向进行讨论和展望。

讲者简介：谢洪涛，中国科学技术大学教授、博士生导师，国家杰青、优青，中科院青促会优秀会员。从事人工智能和多媒体内容安全方向的研究，以第一或通讯作者在国际一流期刊和会议上发表学术论文 100 余篇，担任 ACM TOMM 等四个国际著名期刊编委。主持科研项目 10 余项，含国家重点研发计划“网络空间安全治理”重点专项项目 1 项、国家自然科学基金联合重点项目 2 项。获 2023 年度国家技术发明奖二等奖，2019 年度国家自然科学奖二等奖，2022 年度教育部技术发明奖一等奖，2021 年度中国专利奖优秀奖，2018 年度中国电子学会自然科学奖一等奖，2022 年度中国图象图形学学会青年科学家奖。

魏云超 北京交通大学

报告题目：视觉智能推理

报告摘要：视觉智能推理旨在让机器像人一样从视觉数据中提取信息、理解场景，并基于此开展逻辑推理和决策。本次报告将着重介绍团队在探索视觉智能推理任务上的科研进展，包括 1）如何将大语言模型的复杂推理能力迁移到视觉感知任务当中，2）如何通过浏览大量视觉数据使得模型学习到一定的推理能力。

讲者简介：魏云超，北京交通大学二级教授，教育部长江学者。曾在 NUS、UIUC、UTS 从事研究工作，主要研究方向包括面向非完美数据的视觉感知、多模态数据分析与推理、生成式人工智能等，发表 TPAMI、CVPR 等顶级期刊/会议论文 100 多篇，Google 引用超 26000 次。入选 AI 100、MIT TR35 China、百度全球高潜力华人青年学者、《澳大利亚人》TOP 40 Rising Star，获世界互联网大会领先科技奖、教育部自然科学奖一等奖、ImageNet 目标检测冠军及多项 CVPR 竞赛冠军等奖励。主持国自然重大研究计划重点项目、国家重点研发计划青年科学家项目、北京市自然科学基金海淀联合基金重点项目等 10 余项。担任计算机学院科研副院长、“视觉智能交叉创新”教育部国际联合实验室副主任、“科幻音视频智能处理”北京市重点实验室副主任等职务。

梁小丹 中山大学

报告题目：“慢思考”在多模态基础模型与具身智能中的探索

报告摘要：近期，DeepSeek-V3/R1 引发了 AI 行业对扩展定律认知的范式转变，突显了业界对测试时推理能力发展的共识。本次报告中，我们将重点介绍在多模态推理领域的创新工作。该工作将“慢思考”能力融入大语言模型，以及通用具身智能系统的高层规划框架。与依赖直接快速推理的现有方法不同，我们的核心思路是逐步构建由原子动作组成的长思维链（CoT），引导多模态大模型执行复杂推理。同时，我们也将分享在通用机器人操作与导航的具身基础模型方面的最新研究成果。

讲者简介：梁小丹，中山大学教授，博导，国家万人计划青年拔尖人才，教育部 U40 人才。研究方向为具身智能和多模态大模型。在 Google Scholar 上被引用超过 30,000 次。现任《Image and Vision Computing》与《Neural Networks》期刊副主编，并担任 CVPR、ICML、ICCV、NeurIPS、ICLR、ECCV、ACM MM 等顶级会议领域主席，曾担任 CVPR 2023 监察委员会主席，组织 ICML 2023"AI for Math"研讨会。获得过 ACM 中国新星提名奖、阿里巴巴达摩院青橙奖、中国图象图形学学会-石青云女青年科学家奖、吴文俊人工智能优秀青年奖、中国科协青年人才托举工程、中国图象图形学学会科学技术一等奖、CCF 优秀博士学位论文奖、ACM 中国优秀博士学位论文奖等荣誉。

沈为 上海交通大学

报告题目：视觉语言的细粒度对齐

报告摘要：视觉语言的细粒度对齐是多模态人工智能研究中的一个重要问题，也是构建多模态大模型的关键技术之一。该问题的难点在于细粒度匹配的复杂性和数据标注的高成本。本次报告将介绍报告人团队近期从粒度匹配和数据管线两方面对该问题的探索：1）双曲空间下视觉语言的细粒度对齐及在开放词汇语义分割和多模态大模型高效训练等任务上的应用；2）自监督文本分割的大规模文本掩膜数据集构建及在此基础上的多模态文档理解大模型训练。

讲者简介：沈为，现任上海交通大学人工智能研究院教授，博士生导师，国家自然科学基金优青获得者。曾任约翰霍普金斯大学计算机系助理研究教授。研究方向为计算机视觉、深度学习与医学图像处理。在人工智能领域的顶级学术会议和期刊上发表论文 80 余篇，总学术引用 1 万多次，出版人工智能教材《动手学计算机视觉》一部。指导博士生论文获得医学图像处理顶级国际会议 MICCAI 2023 青年科学家奖。担任人工智能领域顶级国际会议 ICML 2025、NeurIPS 2023/2024/2025、ICCV 2025、CVPR 2022/2023 领域主席、SCI 一区期刊 Pattern Recognition 编委，上海计算机学会计算机视觉专委副主任。

王正 武汉大学

报告题目：视觉隐匿感知与对抗

报告摘要：机器通过相机和计算机视觉技术感知和理解物理世界，在现代科技中扮演着至关重要的角色。计算机视觉技术使得机器能够捕捉并分析来自环境中的图像和视频数据，从而获得对周围世界的理解。然而，除了处理和识别基本的可见视觉信息外，仍有大量的隐匿信息和意图隐藏在这些视觉数据背后。如何挖掘和识别这些隐藏的信息，成为当前研究中的重要课题。与此同时，随着隐私保护意识的提高，如何隐匿和防护敏感信息成为另一个不可忽视的挑战。本报告将详细介绍我们在“感知隐藏信息”和“隐匿敏感信息”方面的研究进展。

讲者简介：王正，武汉大学教授，计算机实验教学中心（网络安全国家级虚拟仿真实验教学中心）主任，入选国家级人才计划青年项目（海外），“武汉英才”产业领军人才。曾任新加坡国立大学高级访问学者、东京大学助理教授，荣获中国留日同学会最优秀青年学者奖、ACM 武汉新星奖。主要研究方向为多媒体内容分析、社会安全治理等，在 CCF-A 类期刊和会议发表论文 90 余篇。指导学生获得“挑战杯”黑科技专项赛道全国一等奖、中国国际大学生创新大赛全国银奖。现任 CCF-A 类期刊 IEEE Transactions on Image Processing 编委。

徐文超 香港科技大学

报告题目：多模态学习中的模态竞争及解决方案

报告摘要：多模态学习作为人工智能领域的重要研究方向，因其在语音、视觉、文本等多源数据融合中的广泛应用而备受关注。然而，多模态学习在实际应用中常面临“模态竞争”问题，即不同模态间由于数据质量、信息量或表示能力差异，导致某一模态在学习过程中占主导地位，进而抑制其他模态的信息表达，成为制约多模态系统发展的关键瓶颈。在此次报告中，基于本团队的近期研究，汇报人将详细介绍有关模态竞争的表现、成因以及各种复杂场景（域偏移，分布式）下的关键挑战。通过对现有方法和困难的分析，提出多种创新性的解决方案，旨在充分挖掘多模态知识的潜力，在复杂场景中实现效率与鲁棒性的双提升。

讲者简介：徐文超，中国香港科技大学综合系统与设计系助理教授，研究方向为边缘计算、多模态学习、无线网络等。在 IEEE TMC、CVPR、NeurIPS、ICLR、Proceedings of the IEEE 等期刊和会议上发表论文 100 余篇；担任 IEEE TCCN、Springer PPNA 等多个国际期刊编委，以及 ICML、CVPR 等国际会议区域主席。曾获得诺波特维娜综述奖、IEEE Globecom 等会议最佳论文奖。

Workshop 8：语言视觉协同生成

组织者：杨易（浙江大学）、武宇（武汉大学）、刘希慧（香港大学）

时间：6 月 8 日（周六） 8:30-12:00 地点：三楼报告厅

时间	主持人	内容
8:30-9:00	武宇	讲者：王亚星（南开大学） 题目：文生图模型中文本和图像表征的思考
9:00-9:30	武宇	讲者：饶安逸（香港科技大学） 题目：Bridging the Representation Gap between Humans and Computers for Video Production
9:30-9:40	杨易	企业宣讲：腾讯优图（曹浩宇，腾讯优图实验室高级研究员） 题目：迈向 GPT-4O 级实时音视频交互：全模态大模型的进展与趋势
9:40-10:10	杨易	讲者：王文海（上海人工智能实验室） 题目：书生·万象多模态大模型的技术演进与应用探索
10:10-10:30	中场休息	
10:30-11:00	刘希慧	讲者：杨思蓓（中山大学） 题目：面向语言-视觉智能的基础表征、跨模态生成与主动交互
11:10-11:40	刘希慧	讲者：李崇轩（中国人民大学） 题目：LLaDA：大语言模型新范式

组织者：杨易 浙江大学

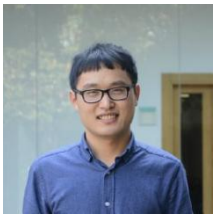
个人简介：杨易，浙江大学求是讲席教授（二级教授）、国家特聘专家。目前担任浙江大学计算机学院副院长、微软-教育部视觉感知重点实验室主任、人工智能省部共建协同创新中心副主任。主要研究方向为人工智能及其应用。所发论文 Google Scholar 引用 8 万余次，H-index 136，近 6 年连续入选 Clarivate Analytics 全球高被引学者。获教育部全国优秀博士论文（2010）、澳大利亚基金委青年研究职业奖（2013）、澳大利亚计算机学会颠覆创新金奖（2016）、谷歌学者研究奖（2016）、澳大利亚科研终身成就奖（2019）、亚马逊机器学习科研奖（2020）、AAAI 最具影响力论文（2021）、ACM MM 唯一最佳论文奖（2023）等多项 AI 领域国际奖项，以及 20 余次国际科研竞赛世界冠军。

组织者：武宇 武汉大学

个人简介：武宇，武汉大学计算机学院教授，国家海外优青获得者。2015 年在上海交通大学获得学士学位，2021 年在悉尼科技大学获得博士学位，2021-2022 年在普林斯顿大学从事博士后研究。主持国家科技创新 2030 重大项目课题，国自然面上等国家级项目，主要从事跨媒体机器学习、视觉-语言协同建模相关的研究，近 5 年，在 TPAMI、CVPR、NeurIPS 等 CCF A 类期刊会议上发表论文 50 余篇，谷歌引用 5000 余次。曾获 2020 年谷歌博士奖研金（Google PhD Fellowship）、2024 年 AAAI 学术新星奖（New Faculty Award）。长期担任 CVPR、NeurIPS、ICML 等人工智能顶会的领域主席，并受邀担任 CVPR 2023 大会的主要组织者、大会主席团成员。

组织者：刘希慧 香港大学

个人简介：Xihui Liu is an Assistant Professor at the Department of Electrical and Electronic Engineering (EEE) and Institute of Data Science (IDS), The University of Hong Kong. Before joining HKU, she was a postdoc Scholar at UC Berkeley, advised by Prof. Trevor Darrell. She obtained her Ph.D. degree from Multimedia Lab (MMLab), the Chinese University of Hong Kong, supervised by Prof. Xiaogang Wang and Prof. Hongsheng Li, and received her bachelor's degree from Tsinghua University. Her research interests cover computer vision, machine learning, and artificial intelligence, with special emphasis on visual synthesis, generative models, vision and language, and multimodal AI. She was awarded Adobe Research Fellowship 2020, MIT EECS Rising Stars 2021, and WAIC Rising Stars Award 2022. She serves as the area chairs for CVPR 2024, ACM MM 2024, ICLR 2025, and CVPR 2025.



王亚星 南开大学

报告题目：文生图模型中文本和图像表征的思考

报告摘要：SD 模型是一种依赖文本提示来生成图像的扩散模型，其核心优势在于能够精准描述目标图像的内容。不过，该模型在生成与文本语义高度一致的图像时存在一定的困难，并且推理过程相对缓慢。为应对这些挑战，本报告探讨优化文本嵌入的方法，通过移除不相关的信息来澄清复杂文本提示中主要对象之间的关系。此外，为了改善推理速度，本报告引入了特征共享机制，以减少处理时间并提高效率。

讲者简介：南开大学计算机学院副教授，博士生导师，入选海外高层次项目，南开“百名青年学科带头人培养计划”。西班牙巴塞罗那自治大学博士，曾在西班牙巴塞罗那自治大学从事博士后研究。研究方向为扩散模型、生成对抗网络、图像到图像翻译、迁移学习。在 IJCV，CVPR，NeurIPS 等期刊会议发表论文 30 余篇，谷歌学术引用 3000 余次。现担任 Computers, Materials & Continua 期刊编委，ECCV Workshop 组织者，在国际顶级期刊和会议 TPAMI、NeurIPS、CVPR、ICCV 等多次担任期刊和会议审稿人。多模态语言翻译国际竞赛（WMT16 Multimodal Machine Translation challenge）中 荣获第一名、2022 年粤港澳大湾区（黄埔）国际算法算例大赛（遥感目标检测赛道）亚军（2/116 队伍）。主持国家自然科学基金青年项目。

饶安逸 香港科技大学

报告题目：Bridging the Representation Gap between Humans and Computers for Video Production

报告摘要：Videos are a beautiful way to share our lives, ideas, stories, and emotions. Recent generative models (e.g. SORA) can generate photorealistic short videos. However, they remain far from being able to create complicated artwork (e.g. films) that requires more human creativity due to the huge gap between human mental representations on control signals and computer representations on pixels, timestamps. To address this challenge, we design controllable and reliable tools to bridge this gap such that creators interact with the tool with conceptual-friendly control signals and produce desired content in a more efficient way. Each module in the tool is reliable, steerable and explainable, which allows the creators to input their intentions, get their expected outputs, and know what happened within it. This allows creators to make iterative improvements in the video creation process rather than numerous trial-and-errors.

讲者简介：Anyi Rao is an Assistant Professor at HKUST. He was a Postdoctoral Scholar at Stanford. He studies human-centered AI for creativity and film, focusing on intelligent media editing and creation, aiming to build collaborative intelligence between AI and humans and unleash human creativity and productivity. His works include ControlNet, AnimateDiff, IC-Light, MovieNet, and Virtual Studio, with a Marr Prize (ICCV best paper award). He organized the Creative Visual Content Workshop at CVPR25, CVPR24, ICCV23, ECCV22, ICCV21, and the Generative Models Course at SIGGRAPH24. He curated the 2025 Hong Kong

AMC AI Film Festival and the 2023 Paris ShortFest AI Film Festival. He also serves as a co-chair of UIST24, UIST25, VINCI25, CVM25 and area chair (TPC) of SIGGRAPH Asia25.

王文海 上海人工智能实验室

报告题目：书生·万象多模态大模型的技术演进与应用探索

报告摘要：：随着大语言模型的兴起，多模态大模型也取得了显著进步，推动了复杂的视觉语言对话和交互，弥合了文本与视觉信息之间的鸿沟。本报告将探讨图文多模态大模型的基本原理和技术，探索如何利用开源套件构建强大的多模态大模型，扩展开源多模态模型的性能边界，以缩小开源模型与商业闭源模型在多模态理解方面的能力差距。

讲者简介：王文海，南京大学博士，香港中文大学博士后，上海人工智能实验室青年科学家，“书生”系列视觉和多模态基础模型核心开发者。研究成果获得了总共超 3 万次引用，单篇最高引用超 5000 次。研究成果分别入选 CVPR 2023 最佳论文，世界人工智能大会青年优秀论文奖，CVMJ 2022 最佳论文提名奖。入选斯坦福大学 2023-2024 年度全球前 2%顶尖科学家，CSIG 优博提名，世界人工智能大会云帆奖。担任 CSIG VI 编委，IJCAI 2021 的高级程序委员会委员，CVPR 2025 AC。

杨思蓓 中山大学

报告题目：面向语言-视觉智能的基础表征、跨模态生成与主动交互

报告摘要：本报告围绕语言-视觉智能的三个层次，从基础表征学习，到跨模态生成，再到动态环境中的主动交互闭环，逐层展开探讨：（1）表征学习方面，首先从第一性原理出发，突破 ViT 在多模态预训练中的表征瓶颈，其次引入适应不同学习条件的自适应机制，以增强真实场景下的泛化能力；（2）跨模态生成方面，针对图文生成，剖析大模型内在机制，协同提升识别、理解与推理能力，并有效缓解幻觉现象；在 4D 生成中，强化语义对齐与时空一致性，优化生成质量；（3）主动交互方面，构建从第一视角视频到仿真再到真实环境的迁移策略，结合主动探索与环境交互反馈，实现目标导向的行为规划与执行。

讲者简介：杨思蓓，计算机学院副教授，博士生导师，逸仙学者。分别于 2020 年和 2016 年在香港大学和浙江大学取得博士和学士学位。其主要研究领域为跨模态视觉感知、理解、生成与交互。迄今为止在 TPAMI、CVPR、ICCV 等期刊或会议发表 CCF A 类/中科院一区论文近 50 篇，引用 2000 余次。主持国自然青年、浦江人才、上海领军人才-海外等多项科研项目。担任 ICCV 等顶会领域主席（AC）。

李崇轩 中国人民大学

报告题目：LLaDA：大语言模型新范式

报告摘要：本次报告聚焦一个问题：自回归是否是通向当前乃至更高水平的生成式智能的唯一范式？本次报告首先从统一概率建模的视角总结当前基础生成模型的发展，并从这个视角出发指出大语言模型的性质（如可扩展性、指令追随、情景学习、对话、无损压缩）主要来自于生成式准则，而非自回归建模独有。基于这些洞察，介绍扩散语言模型最新进展，包括基础理论、扩展定律、规模训练、价值对齐和多模态理解等。LLaDA 系列模型通过非自回归的方式，展示了令人惊讶的可扩展性和多轮对话能力。这些结果不仅挑战了自回归的地位，更加深了我们对生成式人工智能的理解。

讲者简介：李崇轩，中国人民大学高瓴人工智能学院准聘副教授，主要研究领域为生成模型，领导研发扩散语言模型 LLaDA，部分成果部署于 DALL·E 2、Stable Diffusion、Vidu 等行业领先模型。获 ICLR 杰出论文奖、吴文俊优秀青年奖、北京市科技新星、吴文俊人工智能自然科学一等奖等，主持国家自然基金重大研究计划培育项目等。担任 IEEE TPAMI 编委（AE）和 ICLR、NeurIPS 等国际会议的领域主席（AC）。

Workshop 9: 生物特征识别遇上 AI+Science

组织者：朱翔昱（中科院自动化研究所）、雷震（中科院自动化研究所）、何晖光（中科院自动化研究所）

时间：6 月 7 日（周六）18:30-22:00　地点：四楼大湾区厅 2

时间	主持人	内容
18:30-19:00	朱翔昱	讲者：刘宁（中科院生物物理所） 题目：利用深度卷积神经网络理解熟悉面孔识别的神经机制
19:00-19:30	朱翔昱	讲者：常乐（中科院神经科学研究所） 题目：灵长类大脑对面孔信息的神经编码机制
19:30-19:40	雷震	讲者：李晓白（浙江大学） 题目：远程脉搏波测量及其在生物识别领域的应用
19:40~20:00	中场休息	
20:00-20:30	何晖光	讲者：李婧婷（中科院心理所） 题目：心理学实验驱动的微表情分析与情感理解
20:30-21:00	何晖光	讲者：朱翔昱（中科院自动化所） 题目：马尔视觉理论研究：2D-2.5D-3D 表征的涌现

组织者：朱翔昱 中科院自动化研究所

个人简介：朱翔昱，中国科学院自动化研究所项目研究员，从事生物特征识别、数字人、人工智能基础理论研究与应用。国际模式识别协会（IAPR）生物特征识别青年学者奖（YBIA）获得者（两年一次，每次从全球范围内评选 40 岁以下学者一名）。共发表论文 100 余篇，发表文章的 Google Scholar 总引用次数为 10000 余次。获得三次国际竞赛冠军以及四项最佳论文及提名奖。授权国家发明专利 16 项。获 2024 中国图象图形学学会自然科学二等奖（第一完成人），2021 中国电子学会科技进步二等奖、中国图象图形学学会优秀博士论文提名奖，任 CCF:A 类期刊 T-IFS、CCF:B 类期刊 PR 编委，中国图象图形学学会青托俱乐部副主席。

组织者：雷震 中科院自动化研究所

个人简介：雷震，中国科学院自动化研究所研究员，IEEE/IAPR Fellow，入选国家青年高层次人才计划，国家科技进步奖二等奖获得者，长期从事生物特征识别方面的研究，任国际生物特征识别顶刊（CCF-A 类）IEEE Transactions on Information Forensics and Security 副主编，任 Pattern Recognition, Neurocomputing，IET Computer Vision 等国际期刊副主编，发表学术论文 200 余篇，其中人工智能领域顶刊 IEEE TPAMI 9 篇，Google 学术引用超 2.5 万次，H 指数 75。，爱思唯尔中国高被引学者（2020，2021），入选 2022 全球前 2%顶尖科学家榜单，授权发明专利 22 项，撰写发布国家标准 2 项，国家公共安全行业标准 7 项。为表彰其在人脸生物特征方法、系统和数据方面的贡献，被评为 2019 国际模式识别协会（IAPR）青年学者奖（该奖项 2 年评选一次，每次从全球评选出一名 40 岁以下的学者）。获 2021 年国电子学会技术发明一等奖，2022 年中国图象图形学会自然科学奖二等奖。

组织者：何晖光 中科院自动化研究所

个人简介：中国科学院自动化研究所研究员，博士生导师，中国科学院大学岗位教授，上海科技大学特聘教授，国家高层次领军人才。先后主持包括多项国家自然科学基金（2 项重点）、863 项目、国家重点研究计划课题等多个重要项目。先后获得国家科技进步二等奖两项（排二、排三），北京市科技进步奖两项，教育部科技进步一等奖，中国科学院首届优秀博士论文奖等奖项。入选北京市科技新星，中国科学院“卢嘉锡青年人才奖”，中国科学院青年创新促进会优秀会员等。其研究领域为人工智能，脑-机接口、医学影像分析等，在 IEEE TPAMI, ICML 等发表文章 200 余篇。CCF/CSIG 杰出会员。建国七十周年纪念章获得者。

刘宁，中科院生物物理研究所

报告题目：利用深度卷积神经网络理解熟悉面孔识别的神经机制

报告摘要：面孔处理是认知神经科学和计算机视觉研究中的核心领域。我们的行为在很大程度上受到我们所遇到的人的影响，因此，识别熟悉的面孔对于社会互动至关重要。人类在识别熟悉面孔方面表现出卓越的能力，远超对不熟悉面孔的识别能力。我们探讨了深度卷积神经网络（DCNN）在处理熟悉面孔时与大脑神经表征的相似性。结果显示，经过面孔训练的 DCNN 模型，即 VGG-face 模型，在识别熟悉面孔时展现出显著优势，即使在复杂环境下亦如此。随着信息的积累，VGG-face 模型在识别熟悉面孔时的表现呈非线性加速，而对不熟悉面孔的识别能力则呈线性增长。此外，VGG-face 模型中存在熟悉面孔选择性单元，在处理熟悉面孔时表现出与猕猴脑中类似神经元的特征，并且是增强熟悉面孔识别能力的关键因素。此外，与用于物体分类训练的模型相比，专为面孔识别训练的 VGG 模型在熟悉面孔识别上表现出更明显的特征。我们的研究表明，对熟悉面孔的独特反应特征可能代表了一种经过精细调整的身份感知的优化结果。在视觉表征区域内，面孔的感知与记忆之间的重叠可能是自然结果。

讲者简介：中国科学院生物物理研究所认知科学与心理健康全国重点实验室研究员。主要以猕猴为实验动物，结合清醒猴脑功能磁共振成像技术与其他多种研究手段（如行为学、药理学、电生理、神经计算模型等），同时结合基于健康人群和脑疾病患者的观察与实验，从细胞、系统和行为等方面，多层次、多角度地研究社会认知及其相关脑疾病的神经机制，并针对相关脑疾病发展新的早期检测和有效治疗方案。已在 Cell, Nature Communications, PNAS 等期刊上发表相关学术论文 40 余篇。

常乐，中科院脑科学与智能技术卓越创新中心，神经科学研究所

报告题目：灵长类大脑对面孔信息的神经编码机制

报告摘要：面孔作为一类复杂物体在社会交互中扮演了重要作用。大脑如何加工和储存形形色色的面孔呢？基于非人灵长类动物模型的研究揭示了大脑对这类特殊视觉刺激的神经编码机制。在这个报告中，将介绍报告人对非人灵长类面孔信息加工的研究工作，涵盖面孔个体信息编码、信息有缺陷情景下的面孔编码（如遮挡面孔）等，并会讨论人工智能模型对面孔信息编码研究的贡献和影响。

讲者简介：中国科学院脑科学与智能技术卓越创新中心(神经科学研究所)高级研究员，博士生导师。2002-2006 年就读于南开大学数学系，获理学学士学位。2006-2013 年就读于中国科学院生物物理研究所，获理学博士学位。2013-2017 年在美国加州理工学院从事博士后研究工作。2018 年起任中国科学院脑科学与智能技术卓越创新中心研究员，研究方向为灵长类物体视觉的神经机制。他的研究通过结合功能核磁共振、电生理记录、人工智能等

技术手段，定量解析面孔和物体的神经编码机制。相关工作在 Cell，Nature Communications 等重要学术期刊发表。

李晓白，浙江大学

报告题目： 远程脉搏波测量及其在生物识别领域的应用

报告摘要： 心脏搏动可在面部造成微弱的皮肤颜色变化，这种微弱的变化肉眼不可见，但可通过算法远程重建出脉搏波信号（remote Plethysmography, rPPG），因其无接触的便捷性，可应用于情感识别、疾病检测、和生物特征等多领域。报告首先介绍 rPPG 测量的算法研究进展，其次介绍 rPPG 在生物特征领域的应用，包括 rPPG 用于人脸呈现攻击检测，和 rPPG 作为新型生物特征用于身份认证的探索性工作。

讲者简介：浙江大学百人计划研究员，博导，浙江省千人。本科毕业于北京大学，硕士毕业于中科院大学，博士毕业于芬兰奥卢大学。获得芬兰科学院博后奖金，2020 年至 2023 年 4 月在芬兰奥卢大学担任 tenure track 助理教授，获奥卢大学 2019 最具科学领导力的青年学者奖。2023 年加入浙江大学网络空间安全学院，兼任奥卢大学客座教授。研究领域包括机器视觉、机器学习、情感计算、生物识别等。具体方向有微表情识别、基于视频的远程生理信号测量、生物特征识别、人脸活体检测、对抗攻击和伪造检测、多模态情感识别和内容生成等等。发表期刊和会议文章 70 余篇，包括高水平期刊和会议文章如 IEEE TPAMI、TAC、SPM、PIEEE、IJCV、ICCV、CVPR 等十余篇，谷歌学术检索 H 指数 41，总引用 9100，入选 2022 至 2024 年全球 2%高被引学者。担任 IEEE-TCSVT、IEEE-TMM、CVIU 等期刊副编辑。关于微表情的研究被麻省理工科技评论报道，远程心率测量文章获 IEEE 芬兰区 2020 年最佳学生论文奖。

李婧婷，中国科学院心理研究所

报告题目：心理学实验驱动的微表情分析与情感理解

报告摘要：微表情作为揭示真实情绪与心理状态的核心非语言线索，在谎言识别、国家安全、司法审讯、临床诊疗及公共卫生等领域展现出重大应用价值。尽管深度学习技术已在多模态数据分析中取得突破，但微表情生成机制的理论模糊性与小样本数据稀缺性仍制约着其分析模型的性能提升。本报告首先基于面部肌电深入探究微表情表达过程中的面部肌肉运动模式，以期更全面地揭示微表情的本质；进而构建多个高压力交互场景中微表情与说谎行为、压力响应、认知负荷的关联规律；然后融合心理学实验揭示的面部情绪表达与识别规律，启发微表情智能分析的算法设计。最后，在认知科学层面，通过设计系列心理学对照实验，系统比较人类与大语言模型在面孔情感线索解析中的能力边界，为构建人机协同的智能情绪分析系统提供新的研究思路。

讲者简介：李婧婷，中国科学院心理研究所副研究员，博士生导师。获聘中国科学院心理研究所特聘骨干岗位，入选 2023 年中国科学院青年创新促进会成员。担任中国图象图形学学会女科技工作会委员、情感计算与理解专委会委员、机器视觉专委会委员，中国计算

机学会计算机视觉专委委员、人机交互专委委员。主持国家自然面上、青基等科研项目，于 IEEE TPAMI、TAC、TIP、ACMMM 等国内外期刊、会议发表微表情相关论文多篇，两篇论文进入 ESI 高被引论文清单，获 2023 年北京市科学技术奖自然科学奖二等奖。主要研究方向包括计算机视觉、情感计算，特别是智能人脸微表情分析。

朱翔昱，中科院自动化研究所

报告题目：马尔视觉理论研究：2D-2.5D-3D 表征的涌现

报告摘要：计算机视觉奠基人大卫·马尔（David Marr）指出，人类视觉系统通过 "二维草图（2D sketch）、二维半草图（2.5D sketch）、三维模型（3D model）" 的层级化处理流程，逐步构建 "部件-整体" 的物体表征。近年来，深度神经网络在视觉任务上的表现已接近人类水平，但其内部表征机制是否遵循马尔理论尚未得到充分探索。本报告将介绍围绕马尔理论展开的系列研究：我们以人脸感知为切入点，通过结合 神经网络可解释性分析 、 胶囊网络架构 及 无监督三维重建技术 ，系统探究了深度神经网络中间层表征的几何结构。实验表明，在人脸感知过程中，神经网络呈现出 "2D-2.5D-3D" 的表征构建流程，最终形成 "人脸-五官" 的层次化语义表征。这一发现支持了马尔视觉理论，也为理解视觉神经网络的内部工作机制提供了新的视角。

讲者简介：中国科学院自动化研究所项目研究员，从事生物特征识别、数字人、人工智能基础理论研究与应用。国际模式识别协会（IAPR）生物特征识别青年学者奖（YBIA）获得者（两年一次，每次从全球范围内评选 40 岁以下学者一名）。共发表论文 100 余篇，发表文章的 Google Scholar 总引用次数为 10000 余次。获得三次国际竞赛冠军以及四项最佳论文及提名奖。授权国家发明专利 16 项。获 2024 中国图象图形学学会自然科学二等奖（第一完成人），2021 中国电子学会科技进步二等奖、中国图象图形学学会优秀博士论文提名奖，任 CCF:A 类期刊 T-IFS、CCF:B 类期刊 PR 编委，中国图象图形学学会青托俱乐部副主席

Workshop 10：基于生成式 AI 驱动的具身智能

组织者：汪婧雅（上海科技大学）、贾奎（香港中文大学）、胡瑞珍（深圳大学）

时间：6 月 7 日（周六） 8:30-12:00 地点：四楼大湾区 3

时间	主持人	内容
8:30-9:00	胡瑞珍	讲者：盛律（北京航空航天大学） 题目：面向具身智能的单视图三维内容生成：从物体到场景的生成路径
9:00-9:30	胡瑞珍	讲者：黄思远（北京通用人工智能研究院） 题目：Empower Generalist Robot with Human-like Interactions: From Humans to Humanoids
9:30-10:00	汪婧雅	讲者：张直政(银河通用) 题目：合成数据开启端到端具身大模型训练新范式
10:00-10:10	汪婧雅	企业宣讲：银河通用（张直政） 题目：银河通用具身大模型机器人的跨行业应用
10:10-10:30	中场休息	
10:30-11:00	汪婧雅	讲者：石野（上海科技大学） 题目：扩散模型驱动的具身智能：理论与算法前沿突破
11:00-11:30	汪婧雅	讲者：穆尧（上海交通大学） 题目：从多模态认知到具身执行：大规模具身数据自动生成与具身大模型训练
11:30-12:00	汪婧雅	Panel 嘉宾：盛律（北京航空航天大学）、黄思远（北京通用人工智能研究院）、张直政(银河通用)、穆尧（上海交通大学）、石野（上海科技大学）

组织者：汪婧雅 上海科技大学

个人简介：汪婧雅博士现任上海科技大学信息科学与技术学院研究员、助理教授、博导。研究兴趣侧重于以人为中心的三维交互与具身智能。在计算机视觉顶级会议和期刊上发表论文 50 余篇，其中 CCF-A 类论文 40 余篇。担任 CVPR、NeurIPS、ICML、ICCV、ECCV、ACM MM 等会议的领域主席。攻博期间入选 CVPR Doctoral Consortium Award，第一作者论文入选 Computer Vision News Magazine 评比的 2018 Best of CVPR Paper。2023 年入选百度 AI 华人女性青年学者榜。获得 2024 年 ACM Design Automation Conference 最佳论文提名，2024 年 ACM Multimedia 最佳论文提名。

组织者：贾奎 香港中文大学

个人简介：贾奎教授现就职于香港中文大学（深圳）数据科学学院。他的主要研究领域是机器学习与计算机视觉，近期主要聚焦深度学习及其泛化、生成式三维建模与学习、三维感知大模型等方向。他的研究受到国家自然科学基金、广东省科技厅、华为、微软等机构和企业的资助，他的研究成果应用于奥比中光三维传感器产品及三星（美国）无人驾驶系统中。贾奎教授是跨维智能创始人，目前担任 Trans. on Machine Learning Research, IEEE Trans. on Image Processing 等期刊副主编。

组织者：胡瑞珍 深圳大学

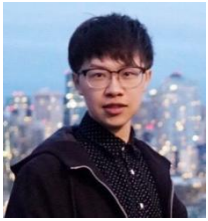
个人简介：胡瑞珍，深圳大学计算机与软件学院特聘教授，博士生导师，国家优秀青年科学基金、广东省杰出青年项目获得者。研究方向为计算机图形学，长期从事三维环境建模与交互方面的研究，发表 ACM SIGGRAPH/TOG 论文二十余篇；入选中科协青年人才托举工程；荣获亚洲图形学协会青年学者奖、全国几何设计与计算青年学者奖；担任期刊 IEEE TVCG、IEEE CG&A 和 Computers & Graphics 等国际期刊编委。

盛律 北京航空航天大学

报告题目：面向具身智能的单视图三维内容生成：从物体到场景的生成路径

报告摘要：构建高精度、物理合理且可编辑的三维场景，对在真实三维数据稀缺瓶颈下实现“虚实融合”训练，提升具身智能体对复杂环境的理解与适应性有重要价值。本次汇报将分享利用扩散模型从单视图构建高精度、可编辑三维视觉内容的系列工作，从三维物体的高精度生成到三维场景的组合式高效生成，仅用单张图片就能构建具有逼真外观、几何准确和物理合理的可编辑三维场景。基于这些工作，进一步介绍面向复杂具身感知任务的学习框架 RoboRefer，借助高精度的三维物体和可编辑三维场景构造海量数据，有效提升具身智能体对复杂动态具身感知任务的学习效率。

讲者简介：盛律，北京航空航天大学“卓越百人”副教授，入选小米青年学者和斯坦福 2024 年全球前 2%顶尖科学家排行榜单。主要研究方向为三维视觉和具身智能。在 IEEE TPAMI/IJCV 以及 CVPR/ICCV/NeurIPS/ICLR/ECCV 等重要国际期刊和会议发表论文超过 50 篇，Google Scholar 显示被引用数超 6600 次。组织 ICML 2024 Multimodal Foundation Models Meet Embodied AI 和 ICCV 2021 SenseHuman 等多个国际会议研讨会。现任 ACM Computing Surveys 副编辑，CVPR 2024-2025、ECCV 2024 和 ACM Multimedia 2024 领域主席，以及多个领域顶会顶刊审稿人和程序委员。任 CCF 和 CSIG 多个专委会执行委员，VALSE 执行领域主席。主持或参与多项国家自然科学基金、科技部重点研发计划和省部级重点研发计划项目。



黄思远，北京通用人工智能研究院，研究科学家

报告题目： Empower Generalist Robot with Human-like Interactions: From Humans to Humanoids

报告摘要：Creating general-purpose embodied robots is one of the ultimate goals of artificial intelligence research. However, most current models lack the ability to understand the 3D world and construct internal world models. Enabling agents to comprehend, reason about, and interact with the 3D world is a critical challenge to address and represents a major bottleneck on the path toward general artificial intelligence. In this talk, I will introduce our recent research efforts (StyleLoco, TRUMANS, LINGO, ManipTrans), which aim to tackle these bottlenecks by empowering general-purpose robots with human-like understanding and interaction capabilities in 3D environments, thereby unlocking a broader range of real-world tasks.

讲者简介：黄思远博士是北京通用人工智能研究院（BIGAI）的研究科学家，并担任通用视觉实验室主任，通院-宇树联合实验室主任。他在加州大学洛杉矶分校（UCLA）统计系获得博士学位，导师是朱松纯教授。他的研究旨在构建一个能够理解和与三维环境交互的类人通用智能体。为实现这一目标，他在以下方向做出了研究贡献：（1）开发可泛化的视觉表征以用于三维重建和语义落地，（2）建模并模仿人类与三维世界的复杂交互，（3）

构建擅长与三维世界和人类交互的具身智能体。他的研究发表于五十余篇会议及期刊论文，并曾获得 ICML Workshop 最佳论文，UCLA 优秀博士论文等奖项。他致力于开发能理解三维物理世界的具身智能体和视觉机器人。

张直政 银河通用联合创始人兼大模型负责人，智源学者

报告题目：合成数据开启端到端具身大模型训练新范式

报告摘要：具身动作数据的昂贵和不足是具身智能发展的主要瓶颈，而高质量的合成大数据为具身端到端大模型的泛化开启了一个训练新范式，即先通过大规模仿真数据预训练广泛学习通用技能，再通过少量真实样本后训练快速掌握专业知识并对齐场景要求。本报告以端到端抓取大模型 GraspVLA 和端到端导航大模型 NaVid 系列等工作为例，介绍如何通过合成大数据打破具身智能对于大规模真实数据的依赖，实现端到端视觉-语言-动作（VLA）大模型系统对于不同维度的全面泛化，并进一步探讨具身智能未来发展的重要方向。

讲者简介：张直政，银河通用联合创始人兼大模型负责人，智源学者，主导银河通用具身智能大模型研发，突破数据和泛化两大技术瓶颈，取得行业领先水平并获广泛关注和赞誉，因其对具身智能领域发展的卓越贡献于今年被评为“北京市劳动模范”。曾任微软亚洲研究院高级研究员，主导过多个基础模型和多模态大模型项目研发，有丰富的 AI 模型及系统的科研、产品化和管理经验。中国科学技术大学和哥伦比亚大学联合培养博士生，曾获中国电子教育学会优博、安徽省优博、中国科学技术大学优博、安徽省优秀毕业生等多个奖项。近三年在全球计算机视觉、人工智能顶级会议和期刊上发表论文 30 余篇。

石野，上海科技大学

报告标题：扩散模型驱动的具身智能：理论与算法前沿突破

报告摘要：本报告系统介绍我们团队近一年在扩散模型驱动具身智能领域的最新理论与算法成果。理论层面提出两大支柱：球面高斯约束扩散 DSG 首次建立损失引导误差下界理论，通过解析解实现零训练成本的流形约束加速；基于随机最优控制构建统一扩散桥框架 UniDB 系列，揭示传统方法为终端约束极限特例的普适规律。算法层面形成扩散强化学习双引擎：加权变分策略 QVPO 首创变分下界统一探索与利用的 off-policy 强化学习框架；可逆扩散策略 GenPO 通过精确反演机制建立首个扩散驱动的 on-policy 强化学习范式。验证层面实现跨物体泛化与跨地形泛化：AffordDP 通过可转移功能性建模，利用基础视觉模型与点云配准实现跨类别泛化，利用 DSG 引导扩散保持动作流形约束；DreamPolicy 框架通过地形感知扩散预测人形运动想象，在人形机器人复杂地形任务中实现零样本泛化。这些工作形成从生成建模、跨域推理到决策控制的完

整技术链条，为具身智能提供兼具理论深度与落地效能的解决方案。

讲者简介：石野博士，现任上海科技大学信息科学与技术学院助理教授、研究员、博导，YesAI 可信与通用智能实验室负责人。主要聚焦在可控、鲁棒、安全的人工智能理论算法及应用，近期系统研究了可控扩散模型的理论基础及其在具身智能上的应用。近 2 年来领导 YesAI 实验室以 80%+首投接收率发表顶会顶刊 30 余篇（NeurIPS, ICML, ICLR, CVPR, ICCV, TNNLS 等）。石野博士担任 NeurIPS 2025 领域主席，组织 ICCV 2025 人机交互与协作研讨会，曾入选上海市海外领军人才计划，上海市扬帆计划，主持国家自然科学基金，曾获得国家优秀留学生奖，IEEE ICCSCE 2016 最佳论文奖，ICLR 2025 生成式理论研讨会杰出论文奖。

穆尧 上海交通大学

报告题目：从多模态认知到具身执行：大规模具身数据自动生成与具身大模型训练

报告摘要：本报告系统性阐述基于多模态大模型的具身智能操作系统研究,构建从场景理解到物理执行的完整技术链路。基于视觉语言大模型,系统可从单次人类演示中分解复杂任务并生成原子技能代码,实现快速学习和场景泛化。为突破数据瓶颈,本研究构建虚实协同的数据生成系统:以物理仿真引擎建立动态场景知识库,结合大模型驱动的程序化内容生成,实现百万级交互轨迹数据自动构建,显著提升模型泛化能力。研究发现认知框架与数据系统相互增强:场景解构模型指导数据生成,物理仿真数据反向优化认知模块,形成自进化训练范式,为构建通用型物理世界智能体奠定基础。

讲者简介：穆尧，上海交通大学计算机学院人工智能研究院长聘教轨助理教授，在国际顶级期刊和会议发表论文 30 余篇，以第一作者或共同第一作者在计算机领域权威期刊会议上发表论文 12 篇，谷歌学术引用超 1400 余次。代表性成果获 2024 年 ECCV 协同具身智能研讨会最优论文奖、2024 年中国自动化学会自主机器人研讨会奖学金（全国 5 人）、2021 年 IEEE ICCAS2020 大会最优学生论文奖、IEEE IV2021 最优学生论文提名奖。入选 KAUST Rising Star 人才计划，曾获得香港政府博士奖学金、香港大学校长奖学金、连续 3 年获得国家奖学金。

Workshop 11：类脑成像：当事件相机遇到 AI

组织者：吴金建（西安电子科技大学）、余肇飞（北京大学）、杨文（西安电子科技大学）

时间：6 月 7 日（周六）18:30-22:00　地点：四楼大湾区厅 3

时间	主持人	内容
18:30-19:00	吴金建	讲者：戴玉超（西北工业大学） 题目：事件相机视觉：从运动感知与影像生成
19:00-19:30	吴金建	讲者：侯军辉（香港城市大学） 题目：Event Tracker：Shifting Object Tracking into Temporal-Continous Domain
19:30-20:00	余肇飞	讲者：郑乾（浙江大学） 题目：见微知著：基于事件信号的物理过程建模与环境感知方法
20:00~20:20		中场休息
20:20-20:50	杨文	讲者：周易（湖南大学） 题目：Enabling Faster and Safer Robots with Neuromorphic Event-based Vision
20:50-21:20	杨文	讲者：田永鸿（北京大学） 题目：神经形态智能感知计算理论与方法

组织者：吴金建 西安电子科技大学

个人简介：吴金建，西安电子科技大学教授，国家优青、陕西省青年人才、教育部霍英东青年基金获得者等。分别于 2008、2014 年获得西安电子科技大学学士、博士学位，2019 年破格晋升教授。主要研究高质量成像及图像智能处理，如事件相机成像系统设计、图像质量增强、图像质量评价、目标检测识别等。已发表学术论文 100 余篇，其中近五年发表 National Science Review, Adv. Sci., IEEE TPAMI、IEEE TIP、NeurIPS、CVPR、AAAI、ACM-MM 等中科院一区 SCI 或 A 类会议论文 50 余篇，两次获国际会议论文奖。主持国家部委项目、国家自然基金项目、科技部重点研发课题、教育部联合基金青年人才类等十余项（合计主持经费三千余万）。获国家自然科学二等奖、陕西省自然科学一等奖等荣誉。

组织者：余肇飞 北京大学

个人简介：余肇飞，北京大学人工智能研究院研究员、博士生导师，国家优秀青年基金获得者。分别于 2012 年、2017 年从重庆大学、清华大学获得工学学士、博士学位，2017-2020 年在北京大学从事博士后研究。主要研究方向为类脑计算、神经形态计算。在 Nature Biomedical Engineering、Science Advance、IEEE Transaction 汇刊和 NeurIPS、ICLR、ICML、CVPR、ICCV、ECCV、AAAI、IJCAI 等顶级会议上发表论文 60 余篇，主持科技部“脑科学与类脑研究”重大项目子课题、国家自然科学基金面上项目、北京市科技新星、博士后创新人才支持计划等项目，担任 ICML、ICLR、ACM MM 等会议领域主席（Area Chair），AAAI Senior PC Member，曾获北京市科学技术奖二。

组织者：杨文 西安电子科技大学

个人简介：杨文，西安电子科技大学助理研究员，分别于 2018、2024 年获得西安电子科技大学学士、博士学位。主要研究方向为事件相机感知成像、图像质量增强与评估，在 IEEE-TIP、IEEE TCSVT、AAAI、ACM-MM 等中科院一区 Top 期刊或 CCF-A 类国际会议上发表论文十余篇；主持国家自然科学基金青年项目、中国博士后科学基金面上项目、国家资助博士后研究人员计划、陕西省博士后科研项目资助等项目；获广东省人工智能产业协会科学技术奖技术发明一等奖、陕西省博士后创新创业大赛银奖、陕西省创新创业优秀博士后等奖项。

戴玉超 西北工业大学

报告题目：事件相机视觉:从运动感知与影像生成

报告摘要：事件相机(EventCamera)作为新型仿生视觉传感器，异步响应像素级亮度变化，突破了传统帧式相机在高速运动、高动态范围场景中的局限。事件相机在自动驾驶、机器人导航、军事国防、深空探测、高速工业检测等领域展现出巨大潜力。报告围绕课题组在基于事件相机的运动感知与影像生成方面的工作展开，涵盖二维与三维运动估计、长时点轨迹跟踪运动物体跟踪与分割、视频帧生成新视角生成等子任务，以打破现有基于帧的图像相机存在的感知瓶颈展现事件相机在复杂动态场景下的感知与生成潜力。

讲者简介：戴玉超，西北工业大学教授，国家级青年人才。主要研究工作集中在机器视觉、智能感知、图像处理、人工智能等领域，聚焦复杂动态场景的三维重建与感知、深度学习和几何模型融合的稠密匹配、新型仿生视觉传感器和计算成像等问题。主持国家自然科学基金、科技部科技创新 2030“新一代人工智能”重大研究计划子课题、JKW 领域基金重点项目等科研项目。近年来在 TPAMI、IJCV、ICCV、CVPR、NeurIPS 等国际顶级期刊和会议上发表论文 70 余篇，谷歌学术引用超过 7700 次，H 因子 43。先后获得 CVPR 2012 最佳论文奖（大陆高校 30 年来首次获得该奖项）、陕西省自然科学奖一等奖、中国图象图形学学会青年科学家奖、火箭军“智箭火眼”人工智能挑战赛全国第一名、IEEE CVPR 2020 最佳论文奖提名、ECCV 2020 鲁棒计算机视觉挑战赛双目深度估计赛道冠军和光流估计赛道亚军、CVPR 2017 非刚性结构与运动恢复挑战赛最佳算法奖、APSIPA 2017 年度峰会最佳深度学习/机器学习论文奖等奖项。担任 APSIPA 杰出讲者和 CVPR、ICCV、NeurIPS 等国际顶级会议领域主席，ACCV 2022 宣传主席，中国图象图形学报青年编委。

侯军辉 香港城市大学

报告题目： Event Tracker： Shifting Object Tracking into Temporal-Continous Domain

报告摘要： Event-based vision offers significant advantages for object tracking in high-speed, low-latency, and challenging illumination scenarios. This talk introduces advanced methodologies moving beyond traditional frame-based approaches toward continuous temporal tracking.Initially, it presents graph-based embedding methods that efficiently handle sparse and irregular event data, effectively capturing essential spatio-temporal relationships. Subsequently, the talk discusses leveraging inherent motion cues uniquely preserved within event streams, significantly enhancing predictive accuracy, especially in dynamic environments. Lastly, the talk explores high-rank cross-modal fusion strategies that bridge the gap between RGB imagery and event data for effective cross-modal tracking. Innovative techniques, such as plug-and-play transformer augmentations and cross-modal masking, harness modality-specific strengths to substantially boost tracking performance. Overall, this talk highlights the transformative shift toward continuous temporal tracking, providing key insights and directions for developing robust, real-time object tracking systems in complex real-world applications.

讲者简介：侯军辉，香港城市大学（城大）计算机科学系副教授，国家优青。2009 年获得中国广州华南理工大学信息工程（优才计划）学士学位，2012 年获得中国西安西北工业大学信号与信息处理工程硕士学位，2016 年获得新加坡南洋理工大学电气与电子工程学院博士学位。曾于 2015 年获得国家留学基金委颁发的中国政府优秀留学人才奖，并于 2018 年获得香港研究资助局颁发的早期事业奖（3/381）。近期主要研究视觉数据的三维/四维生成和基于扩散的建模。曾任/现任 IEEE TIP、IEEE TCSVT、SPIC 和 The Visual Computer 的编委，MSA-TC ，VSPC-TC，MMSP-TC 等国际学术组织的技术委员会成员，IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 的客座编辑， 以及 ACM MM'19/20/21/22、IEEE ICME'20、ICIP'22、VCIP'20 /21/22 和 WACV'21 的领域主席。已在 TPAMI、TIP、TNNLS、TGRS、TCSVT、CVPR、IJCV、ICCV 和 ECCV 等顶级期刊、会议发表论文几十余篇。

郑乾 浙江大学

报告题目：见微知著：基于事件信号的物理过程建模与环境感知方法

报告摘要：受生物视觉系统动态感知机制的启发，事件相机通过异步事件流突破了传统帧式相机的动态感知局限。其微秒级时间分辨率，使其能够精确捕捉环境中光照变化、相机/物体运动等物理过程，为环境感知提供了新的信息维度。本报告针对光源感知、高动态范围感知、本征分解、时空超分辨率等环境感知问题，系统探讨了不同任务下事件信号与物理过程的本质关联，通过建立基于事件信号的物理过程模型，提出针对性的深度学习解决方案，实现稳定的环境感知效果。

讲者简介：郑乾，浙江大学计算机学院、脑机智能全国重点实验室“百人计划”研究员，博士生导师，海外优青，中国视觉与学习青年学者研讨会（VALSE）执行领域主席委员会委员。研究方向主要为类脑智能算法。发表 CCF A 类论文 30 余篇，包括 TPAMI，CVPR，ICCV，NeurIPS，ICML 等，主持国家自然科学基金面上项目，担任国际期刊 Neurocomputing 副主编。

周易 湖南大学

报 告 题 目： Enabling Faster and Safer Robots with Neuromorphic Event-based Vision

报告摘要： The rapid growth of the mobile robotics market, driven by applications such as small drones and autonomous vehicles, demands higher levels of autonomy and the ability to operate in challenging high-speed scenarios. Traditional frame-based vision systems face challenges like motion blur, latency, and blind periods during exposures, while neuromorphic event-based sensors, such as event cameras, offer microsecond-resolution, asynchronous sensing for efficient, high-speed perception with minimal blur and low power, ideal for dynamic, resource-constrained robotics. However, leveraging the full potential of event-based vision for high-speed autonomy presents significant challenges. First, the differential working principle of event cameras complicates the recovery of motion and scene structure from spatio-temporal event data. Second, the asynchronous nature of event data creates difficulties in multi-sensor fusion due to non-equispaced temporal sampling. Third, existing algorithms for event-based state estimation and perception, which rely on parametric model fitting, are computationally inefficient and unsuitable for real-time applications, hindering the realization of event-enhanced high-speed autonomy. This talk explores how neuromorphic event-based vision can enable faster and safer robots by addressing these challenges. We discuss innovative sensing principles, advanced algorithms, and software tools designed to enhance tasks such as Simultaneous Localization and Mapping (SLAM) and dynamic-scene perception. By overcoming these barriers, event-based vision can unlock new possibilities for high-speed autonomy, paving the way for faster and safer mobile robots in the future.

讲 者 简 介： Yi Zhou is currently a Professor at Hunan University, where he directed the Neuromorphic Automation and Intelligence Lab (NAIL). He obtained his PhD from the Australian National University (ANU) in 2018. He was a visiting scholar at ETH Zurich (2017-2018) and was awarded the NCCR Fellowship Award by the Swiss National Science Foundation for his research on neuromorphic event-based 3D vision. From 2019 to 2021, he was a postdoc research fellow at the HKUST&DJI Innovation Joint Lab, where he proposed the world’s first open-source event-based stereo visual odometry (ESVO) system. His research interests consist of vSLAM, multi-view geometry theory, and dynamic vision sensor (DVS). He is an associate editor of the IEEE Robotics and Automation Letters.

田永鸿 北京大学

报告题目：神经形态智能感知计算理论与方法

报告摘要：人工智能的终极研究目标是研究用于模拟、延伸和扩展人类智能处理能力的理论、方法与技术，最终开发出能以与人类智能相似的方式做出反应的机器。然而，由于真实世界是高度复杂、变化无穷的动态开放场景，使得感知计算技术总会面临新的、从未见过的困难场景，如极端光照/天气条件、高速运动对象等。大脑是一个高效智能的超大规模生物脉冲网络，对开放动态环境具有很强的感知认知能力，因此迫切需要借鉴大脑机制发展神经形态感知计算理论与方法。本报告将围绕这一主题，重点分享采用模拟人类视网膜的新型神经形态传感器，结合模拟人类视觉皮层的脉冲神经网络，打造“神经形态视觉+脉冲神经网络”的感知计算新架构方面所面临的挑战问题与研究进展。

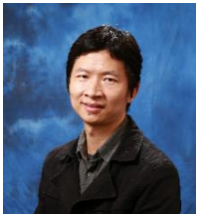
讲者简介：田永鸿，北京大学博雅特聘教授，博士生导师，IEEE Fellow，北京大学深圳研究生院副院长兼信息工程学院院长、科学智能学院执行院长，鹏城实验室智能计算部副主任兼云脑研究所所长，2018 年国家杰出青年基金获得者，2024 年首批国家杰出青年基金延续资助计划获得者。主要研究方向为分布式机器学习、类脑神经网络、神经形态视觉和 AI for Science。累计主持国家重点研发计划项目、国基金杰青/重点/重大仪器项目等国家、省部级与企业合作项目 40 余项，累计在 Nature/Science 子刊、IEEE Trans 等国际期刊和 ICML、NeurIPS 等国际会议发表学术论文 400 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、2024 年首届祖冲之奖重大成果奖、2025 年 IEEE Hans Karlsson 奖，是 2018 年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”总体专家组成员、广东省十四五重点专项“新一代人工智能”专家组成员。

Workshop 12: AI4Science，诺贝尔奖后的思考

组织者：欧阳万里（上海人工智能实验室）、周浩（清华大学智能产业研究院）、周东展（上海人工智能实验室）

时间：6 月 7 日（周六） 8:30-12:00 地点：四楼大湾区厅 5

时间	主持人	内容
8:30-9:00	欧阳万里	讲者：邵斌（中关村人工智能研究院） 题 目 ： Toward Protein Dynamics Simulations with Ab Initio Accuracy
9:00-9:30	欧阳万里	讲者：王笑楠（清华大学） 题目：AI+能源化工新材料:大小通专模型并进
9:30-9:40	周浩	企业宣讲：无问芯穹（吴保东） 题目：面向大模型科学研究的算力平权技术探索
9:40-10:10	周浩	讲者：肖文之（字节跳动） 题目：Protenix—生物分子复合物结构预测与设计
10:10-10:20	中场休息	
10:20-10:50	周浩	讲者：符天凡（南京大学） 题目：深度学习赋能的药物发现与开发
10:50-11:20	周东展	讲者：朱霖潮（浙江大学） 题目：AI 赋能科学智能仿真和设计
11:20-11:50	周东展	讲者：岳翔宇（香港中文大学） 题目：从 AI 到 AI4Science：我们的实践与探索



组织者：欧阳万里 上海人工智能实验室

个人简介：欧阳万里，上海人工智能实验室领军科学家、AI for Science 中心负责人，现任 VALSE AC。曾任悉尼大学电子信息工程学院研究主任。研究领域为模式识别、深度学习、计算机视觉、AI for Science。欧阳万里教授承担国家重大科研任务，主导开展人工智能驱动的交叉科学（AI4Science）研究工作，承担科技创新 2030——新一代人工智能重大项目（通用视觉方法体系与基础研发平台）通用视觉数据集与评测平台课题，成果显著，影响广泛。在 TPAMI、IJCV、CVPR、NeurIPS 等 CCF A 类期刊和会议论文发表 200 余篇，截止到 2024 年 11 月，谷歌学术引用达 50,000+, H-index 指数达 101。ICCV 最佳审稿人，担任人工智能领域顶级期刊 TPAMI 和 IJCV 副编，CVPR2023 资深领域主席，CVPR2021、ICCV2021 领域主席。

组织者：周浩 清华大学智能产业研究院

个人简介：周浩，清华大学智能产业研究院副研究员。研究方向是面向复杂符号系统的生成式人工智能，主要的应用包括超大规模语言模型，分子生成，蛋白质设计，新材料发现等。曾任字节跳动研究科学家和副总监，领导搭建了字节跳动的文本生成中台和 AI 辅助药物设计两个方向的研发团队。他长期担任 ICML、NeurIPS，ICLR，ACL 等人工智能顶级会议的领域主席，在人工智能顶级国际会议上发表论文 80 余篇。他作为负责人参与国家科技创新 2030 “新一代人工智能”重大项目课题《跨尺度化学大模型驱动的高效新材料智能设计》，作为中心主任牵头清华 AIR-字节跳动可拓展大模型智能联合研究中心。曾获 2019 年度中国人工智能学会优秀博士论文奖、自然语言处理领域顶级国际会议 ACL 2021 最佳论文奖 (1/3350)、2021 年度中国计算机学会 NLPCC 青年新锐学者奖、2024 年北京市科技新星等荣誉。

组织者：周东展 上海人工智能实验室

个人简介：周东展，上海人工智能实验室青年研究员。2023 年博士毕业于悉尼大学(QS 排名前 20)，当前主要研究方向为人工智能驱动的科学发现，包括科学大语言模型、多模态模型、科学具身智能等。在人工智能和物理顶级期刊会议发表论文十余篇，长期担任 TPAMI、IJCV、nature communication、CVPR 等顶级期刊会议的审稿人。

邵斌 中关村人工智能研究院

报告题目：Toward Protein Dynamics Simulations with Ab Initio Accuracy

报告简介：AlphaFold 获得 2024 年诺贝尔化学奖，其能准确预测蛋白质的静态结构。本报告将介绍作者团队在利用 AI 技术探索蛋白质高精度动力学模拟方面的工作 AI2BMD。AI2BMD 首次实现了对蛋白质生物大分子的准量子化学精度全原子高效大规模模拟。这项研究为探索蛋白质结构与功能关系、推进新药研发和疾病治疗策略提供了新的思路和工具。

讲者简介：邵斌博士现任中关村人工智能研究院院长。2010 年获复旦大学博士学位，曾任微软研究院科学智能中心资深首席研究员，主导微软分布式图引擎（Graph Engine）研发，并创立计算生物学研究组与科学计算研究组。其研究横跨人工智能、计算化学、生物分子动力学模拟、计算生物学及高性能科学计算等前沿领域，在《Nature》主刊及其子刊等国际顶级期刊与一流学术会议发表论文 50 余篇，曾获「ICDE 十年影响力论文奖」，蛋白质动力学模拟成果入选 2024 年度「中国生物信息学十大进展」。

王笑楠 清华大学

报告题目：AI+能源化工新材料:大小通专模型并进

报告摘要：AI for Science and Engineering 需解决能源、化工、材料等领域面临着一个共同挑战:如何将微观尺度物质特性与宏观尺度工程应用有效关联，实现高效精准的预测、优化和控制。对此，基于大小通专模型并进，AI 驱动的化工新材料与低碳过程设计可以结合大数据挖掘、深度主动学习、基础模型优化、数据增强等智能方法，有效指导新材料、新工艺、新系统的开发，探索高效的碳捕集、利用和催化转化体系。通过融合高通量计算和高通量实验等关键技术，结合机械自动化平台，高效开发关键材料和催化剂。本报告将探讨如何：1）构建数据驱动融合知识的新分子新材料建模与优化方法；2）发展关键新材料与器件的精准智能合成及表征策略；3）建立多尺度数字孪生与低碳智联系统，实现智能材料研制平台的转化应用。未来展望将从跨尺度、多模态、可通用、可解释 AI 的角度深入研究智能科学，建立大规模模块化领域人工智能基础模型，以数据为桥梁，实现理、实、算、数一体化闭环发展，迈向面向低碳绿色发展的通用智能。

讲者简介：王笑楠，清华大学化学工程系长聘副教授、博导，智能化工研究中心主任。新加坡国立大学荣誉副教授、博导，带领团队长期从事 AI+能源化工新材料的研究。在 Nat. Mach. Intell.、Nat. Synth.等期刊发表论文 170 余篇，被引 10900 余次，H-index 61。主持“新一代人工智能”国家科技重大专项“化学材料 AI 大模型赋能碳中和”（任首席科学家，项目负责人），入选科睿唯安 2024 年度“全球高被引科学家”、全球学者终身学术影响力榜、连续四年被 Elsevier 评为全球前 2% 顶尖科学家。担任 Applied Energy 等十本国际期刊副主编和编委，获美国化学会可持续化学与工程讲席奖、Cell Press 年度中国女科学家、青年北京学者、中国化工学会侯德榜化工科学技术奖“青年奖”等荣誉。

肖文之 字节跳动

报告题目：Protenix——生物分子复合物结构预测与设计

报告摘要：我们团队开源了复合物结构预测模型 Protenix，是全球领先的 Alphafold 3 复现开源工作。以此为基础，我们关注以下关键问题：1) 构建下一代结构预测模型，大幅提升各种体系上的预测精度，并进一步解决多构象、动态等问题。2) 将结构预测模型扩展为通用的基础模型，支持亲和力预测、分子设计等更多任务，解决广泛的生物和制药问题。我会介绍我们复合物结构建模和生成式设计的工作，分享其中的经验和关键挑战。

讲者简介：肖文之，字节跳动研究员，在 AI 落地应用方面有多年经验积累。目前负责生物复合物结构预测方向，团队成员来自于机器学习、计算化学、生物信息等领域，通过密切的团队合作解决生物场景的挑战性问题。

符天凡 南京大学

报告题目：深度学习赋能的药物发现与开发

报告摘要： 药物设计和开发是一个既漫长又昂贵的过程，涉及从分子发现到临床试验的多个复杂步骤。人工智能（AI）技术展示了巨大的潜力，可以显著加速这一过程并降低成本。在药物发现的初期阶段，

目标是识别具备理想药理特性的分子。本报告将深入探讨最新的药物设计方法，包括连续空间深度生成模型和离散空间药物设计路径搜索算法。这些先进的 AI 工具能够高效地探索化学空间，预测新化合物的活性和安全性，并优化候选药物的设计，以满足特定的治疗需求。进一步讲，在药物开发的后期阶段，重点转向了临床试验，这是评估药物对人体安全性和有效性的重要环节。为了提高临床试验的成功率和效率，本报告将介绍一系列最新的可信赖的方法，包括可解释性、不确定性感知的临床试验设计与预测技术。这些方法不仅能够模拟真实的临床试验过程，还能帮助科学家更好地理解潜在的风险和收益，从而做出更加明智的决策。

讲者简介：符天凡博士长期从事人工智能赋能的药物发现（AI for Drug）、人工智能赋能的科学发现（AI for Science）方面的研究。他本科硕士毕业于上海交通大学计算机科学与技术系，博士毕业于美国佐治亚理工学院计算机科学与工程系。曾任美国伦斯勒理工学院计算机科学系常任轨道助理教授。2024 年 12 月加入南京大学计算机科学与技术系，入选国家级青年人才项目。他在 Nature、Nature Chemical Biology、Nature Machine Intelligence、ICML、ICLR、NeurIPS、KDD、TKDE 等知名会议和期刊上发表学术论文 40 余篇。论文被国内外同行广泛引用，引用者来自斯坦福、麻省理工、哈佛、耶鲁、普林斯顿等国际著名机构，包括中、美、英、加、欧等国/地的 20 余位科学院/工程院院士和 50 余位 AAAI/ACM/IEEE Fellow。2017 年翻译了深度学习（“花书”），销量达 50 余万册。研究成果应用于多家生物医药企业。他还共同组织了前三届 AI for Science 研讨会。更多信息请参见个人主页：https://futianfan.github.io/

朱霖潮 浙江大学

报告题目：AI 赋能科学智能仿真和设计

报告摘要：AI 驱动的科学智能仿真与设计旨在融合数据和机理方法，显著提升科学计算设计效率和精度。在方程求解方面，本报告将介绍全息物理混合器，该方法有效统一了基于注意力和基于谱的方程求解方法，整合注意力机制的局部灵活性和谱方法的全局泛化能力。在逆向设计方面，本报告将介绍基于生成式 AI 的物质序列逆向设计方法，实现从功能到序列的高效反向映射。本报告将展示 AI 驱动技术如何在流体力学、材料科学、生物医学等领域革新科学仿真与设计流程，加速科学发现和工程创新。

讲者简介：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、斯坦福全球前 2%顶尖科学家、福布斯中国 30U30 等荣誉。曾在澳大利亚悉尼科技大学担任助理教授。主要研究方向为科学智能、智能仿真、人工智能通用基础模型等。曾获美国国家标准总局 TRECVID LOC 等 8 项国际竞赛冠军。担任 NeurIPS、ECCV、CVPR 等国际会议领域主席，并多次在国际会议上组织专题研讨会。

岳翔宇 香港中文大学

报告题目：从 AI 到 AI4Science：我们的实践与探索

报告摘要：近年来，人工智能（AI）技术正逐步从通用领域向科学计算（AI4Science）深度拓展，用 AI 赋能各个领域。本次报告中，我将结合我们 AI 的背景，介绍我们如何将 AI 技术与科学研究深度融合，推动跨学科的探索。我们以具体案例为基础，介绍在化学、智能制造、跨学科研究等领域的 AI4Science 应用。

讲者简介：岳翔宇现任香港中文大学信息工程系助理教授、博士生导师。他于 2014 年、2016 年和 2022 年分别在南京大学获得工学学士学位、斯坦福大学获得硕士学位、加州大学伯克利分校获得博士学位。自 2022 年起在香港中文大学任教，主要研究方向为人工智能、多模态学习、生成模型、AI4Science 等，已在国际权威期刊和会议发表论文 50 余篇，谷歌学术引用超 10000 次，H 指数 31。他曾荣获 Lotfi A. Zadeh 奖，并担任 CVPR 2025、NeurIPS 2025、ICML 2025 等计算机顶级会议的领域主席。

Workshop 13：大模型基础理论-从理论视角看待大模型技术发展

组织者：刘勇（中国人民大学）、李健（北京师范大学）

时间：6 月 7 日（周六） 8:30-12:00 地点：四楼大湾区厅 4

时间	主持人	内容
8:30-9:00	刘勇	讲者：车万翔（哈尔滨工业大学） 题目：迈向推理时代——长思维链机理及应用
9:00-9:30	刘勇	讲者：张奇（复旦大学） 题目：大语言模型能力来源与边界
9:30-9:40	刘勇	企业宣讲：九章云极（蒋宁） 题目：高性能高弹性普惠算力加速大模型科研
9:40-10:10	刘勇	讲者：龚铁梁（西安交通大学） 题目：领域泛化的信息论协同机制
10:10-10:30	中场休息	
10:30-11:00	李健	讲者：黄维然（上海交通大学/上海创智学院） 题目：矩阵信息论视角下的表征学习
11:00-11:30	李健	讲者：方聪（北京大学） 题目：随机梯度下降算法在高维回归问题中正则效应与泛化性能分析

组织者：刘勇 中国人民大学

个人简介：刘勇，中国人民大学，长聘副教授，博士生导师，国家级高层次青年人才。长期从事机器学习基础理论研究，共发表论文 100 余篇，其中以第一作者/通讯作者发表顶级期刊和会议论文近 50 篇，涵盖机器学习领域顶级期刊 JMLR、IEEE TPAMI、Artificial Intelligence 和顶级会议 ICML、NeurIPS 等。曾获中国人民大学“杰出学者”、中国科学院“青年创新促进会”成员、中国科学院信息工程研究所“引进优青”等称号。主持/参与国家自然科学面上/基金青年、科技部重点研发、北京市科技计划中央引导地方专项、北京市面上项目等项目。

组织者：李健 北京师范大学

个人简介：李健，北京师范大学副教授、博士生导师；曾入选中国科学院特别研究助理、中国科学院信息工程研究所优才计划 A 类/B 类、2024 年“微软亚洲研究院铸星计划”学者。研究方向为大语言模型、机器学习基础理论。在机器学习领域，已发表 30 余篇顶级期刊与会议论文，其中以第一作者在 TIT、JMLR、AI、 ICML 、NeurIPS 等机器学习领域顶级期刊与会议上发表 CCF-A 类论文 13 篇、CCF-B 类论文 3 篇。

车万翔 哈尔滨工业大学

报告题目：迈向推理时代——长思维链机理及应用

报告摘要：推理能力已经成为评估并提升模型智能水平的核心指标和重要途径。长思维链（Long Chain-of-Thought, Long CoT）作为推理大模型（RLLMs）处理复杂任务时的关键技术，能够显著增强模型的推理深度与广度。通过实现深度的逻辑推理、积极的探索行为和有效的反思机制，长思维链赋予模型解决更加复杂、多样化问题的能力。本报告将深入解析长思维链的基本原理及应用实践。具体内容包括系统地阐述长思维链的核心技术与机制，重点涵盖深度推理的逻辑形式与学习框架、反思能力的反馈与纠错机制、探索能力的扩展机制，以及内部与外部探索框架等关键内容。此外，报告还将详细分析长思维链中产生思维链边界与测试时扩展（Test Time Scaling）现象的根本原因。最后，报告将展望长思维链的未来发展趋势，特别关注多模态长思维链等前沿方向，以期为人工智能领域的研究者与实践者提供一定的参考与启示。

讲者简介：车万翔，哈尔滨工业大学计算学部长聘教授/博士生导师，人工智能研究院副院长，国家级青年人才，斯坦福大学访问学者。主要研究领域为自然语言处理、大语言模型。现任中国中文信息学会理事、计算语言学专业委员会副主任兼秘书长；国际计算语言学学会亚太分会（AACL）执委兼秘书长；国际顶级会议 ACL 2025 程序委员会共同主席。承担国家自然科学基金重点项目和专项项目、2030“新一代人工智能”重大项目课题等多项

科研项目。著有《自然语言处理：基于预训练模型的方法》一书。曾获 AAAI 2013 最佳论文提名奖。负责研发的语言技术平台（LTP）已授权给百度、腾讯、华为等公司付费使用。2024 年获中国人工智能学会吴文俊人工智能科技进步一等奖（排名第 1），2020 年获黑龙江省青年科技奖，2016 年获黑龙江省科技进步一等奖（排名第 2）。入选斯坦福大学和爱思唯尔发布的 2024 年度“全球前 2%顶尖科学家”榜单。

张奇 复旦大学

报告题目：大语言模型能力来源与边界

报告摘要：随着大语言模型能力的不断增强，剖析其能力来源与边界愈发关键。大语言模型为什么有“幻觉”？推理能力来源是强化学习吗？大模型真的具备推理能力吗？等等问题受到越来越多的关注。本次报告将从大模型能力边界的实践研究和大模型能力来源两个方面，结合国内外最新研究进展与报告人及团队工作进行介绍。

讲者简介：张奇，复旦大学计算机科学技术学院教授、博士生导师。兼任上海市智能信息处理重点实验室副主任，中国中文信息学会理事、CCF 大模型论坛常务委员、CIPS 信息检索专委会常务委员、CIPS 大模型专委会委员。主要研究方向是自然语言处理和信息检索，聚焦大语言模型、自然语言表示、信息抽取、鲁棒性和解释性分析等。在 ACL、EMNLP、COLING、全国信息检索大会等重要国际国内会议多次担任程序委员会主席、领域主席、讲习班主席等。近年来承担了国家重点研发计划课题、国家自然科学基金、上海市科委等多个项目，在国际重要学术刊物和会议发表论文 200 余篇，获得美国授权专利 4 项，著有《自然语言处理导论》和《大规模语言模型：理论与实践》，作为第二译者翻译专著《现代信息检索》。获得 WSDM 2014 最佳论文提名奖、COLING 2018 领域主席推荐奖、NLPCC 2019 杰出论文奖、COLING 2022 杰出论文奖。获得上海市“晨光计划”人才计划、复旦大学“卓越 2025”人才培育计划等支持，获得钱伟长中文信息处理科学技术一等奖、汉王青年创新一等奖、上海市科技进步二等奖、教育部科技进步二等奖、ACM 上海新星提名奖、IBM Faculty Award 等奖项。

龚铁梁 西安交通大学

报告题目：领域泛化的信息论协同机制

报告摘要：领域泛化旨在学习跨源域的不变表征，从而增强对分布外数据的泛化能力。虽然目前基于梯度或表征匹配算法在领域泛化方面取得成功，但这些方法通常缺乏泛化保证，或依赖于强假设，这对于理解分布匹配的底层机制仍存在巨大鸿沟。本报告将从一种概率视角来构建领域泛化，确保鲁棒性的同时避免过于保守的解。通过信息论理论分析，我们对梯度和表示匹配在促进泛化方面的作用提供了新的洞见。所得理论结果揭示了这两种匹配机制的互补关系，表明现有研究仅关注梯度或表示对齐不足以

解决领域泛化问题。基于理论结果，我们引入 IDM 策略来同时对齐域间梯度和表示，并结合用于复杂分布匹配的 PDM 方法，取得了 SOTA 的效果。

讲者简介：龚铁梁，西安交通大学计算机学院，陕西省大数据知识工程重点实验室副教授、博士生导师；渥太华大学博士后，密歇根大学访问学者。研究方向包括统计学习理论、信息论，机器学习等，并致力于设计具有理论保证的算法应用于实际问题。曾获获中国发明协会年度发明创新奖二等奖,华为 “揭榜挂帅”火花奖，西安交通大学医工交叉青年创新奖等。研究成果主要发表于 ICML, NeurIPS, ICLR, IEEE TIT, TSP 等国际会议及期刊上。目前担任国际期刊 IEEE TIT，JMLR, TNNLS,TMI 以及人工智能顶会 ICML,NeurIPS,ICLR 的审稿人，AAAI, IJCAI 高级程序委员，作为负责人主持国家自然科学基金、科技部 2030 新一代人工智能重大项目子课题，以及华为、中国电信等多项横向课题，并作为骨干成员参与国家自然科学基金重点及重大项目多项。

黄维然 上海交通大学/上海创智学院

报告题目：矩阵信息论视角下的表征学习

报告摘要：近年来，表征学习在监督学习、自监督学习和多模态对齐等任务中发挥着关键作用。尽管这些任务在训练目标和机制上存在差异，但它们都依赖于对输入数据的结构化表征与信息压缩。本报告从矩阵信息论的视角出发，统一分析不同学习范式中的表征过程，揭示其共性结构与信息流动机制，为理解不同学习范式中的表征机制提供了新视角。

讲者简介：黄维然，上海交通大学计算机学院副教授、博士生导师，上海创智学院全时导师，上海人工智能实验室科研顾问，入选 2024 年“微软亚洲研究院铸星计划”学者。先后在清华大学获得学士和博士学位。研究聚焦大模型的高效微调与持续学习研究，累计在 NeurIPS、ICLR、ICML 等国际顶级人工智能会议上发表论文 30 余篇，其中“噪声标签长尾问题”学术论文入选 ICCV 2023 最佳论文候选名单（17/8260）。

方聪 北京大学

报告题目：随机梯度下降算法在高维回归问题中正则效应与泛化性能分析

报告摘要：随机梯度下降算法是求解机器学习问题中的常见算法。在高维学习问题中，随机梯度下降算法的迭代次数往往低于模型参数量，算法对于模型的产生隐式正则效应是模型具有良好泛化的主要原因。本次讲座，我们将研究随机梯度下降算法在不同学习情境下求解线性与简单非线性模型的泛化性能，并进行定量比较。在线性模型中，我们将分别讨

论算法在不同学习尺度（即样本数与问题维度不同依赖关系）与协变量偏移条件下的学习效率，尝试理解算法对于学习问题的适应性与涌现发生的条件。在非线性模型，我们将阐明算法能够自适应问题结构，突破一阶算法在离线情形下面临的统计-计算鸿沟（statistical to computational gap）诅咒，并能够自动实现统计推断。

讲者简介：方聪，北京大学智能学院担任助理教授（博导）兼研究员。方聪于 2019 年在北京大学获得博士学位，先后在普林斯顿大学和宾夕法尼亚大学进行博士后研究。方聪的主要研究方向是机器学习基础理论与算法，已发表包括 PNAS、AoS、IEEE T.IT、JMLR、COLT、NeurIPS、PIEEE 等 30 余篇顶级期刊与会议论文，担任机器学习顶级会议 NeurIPS、ICML 领域主席（Area Chair），团队获得 2023 年度吴文俊人工智能自然科学奖一等奖。

Workshop 14：自动驾驶多模态世界模型

组织者：李镇（香港中文大学（深圳））、贾伟（合肥工业大学）、贾旭（大连理工大学）

时间：6 月 8 日（周日） 8:30-12:00 地点：四楼大湾区厅 4

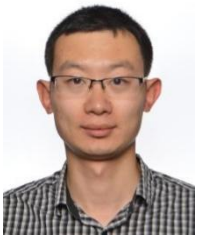
时间	主持人	内容
8:30-8:55	李镇	讲者：李弘扬（香港大学） 题目：迈向具备高泛化性的通用自动驾驶世界模型
8:55-9:20	李镇	讲者：马超（上海交通大学） 题目：自动驾驶多模态场景理解与生成
9:20-9:45	贾旭	讲者：王新宇（华为技术有限公司） 题目：面向未来智能辅助驾驶的 WEWA 架构
9:45-10:00	中场休息	
10:00-10:25	贾旭	讲者：王乃岩（小米） 题目：什么是适合物理世界 AI 的表征学习？
10:25-10:35	贾旭	企业宣讲：小米（王乃岩） 题目：走进小米汽车
10:35-11:00	贾伟	讲者：赵昊（清华大学） 题目：可控高效的时空世界模型在自动驾驶中的应用
11:00-11:25	贾伟	讲者：高晋（中国科学院大学） 题目：基于扩散模型的时空一致 4D 内容生成初探
11:25-11:50	贾伟	讲者：范略（中国科学院自动化研究所） 题目：生成和重建一体化的自动驾驶世界模型
11:50-12:15	李镇	Panel 嘉宾：金鑫（宁波东方理工大学）、李弘扬（香港大学）、马超（上海交通大学）、王新宇（华为）、王乃岩（小米）、赵昊（清华大学）、高晋（中国科学院大学）、范略（中国科学院自动化研究所）

组织者：李镇 香港中文大学（深圳）

简介：李镇博士，现任香港中文大学 (深圳)理工学院助理教授， 深圳市未来智联网络研究院助理院长，校长青年学者。李镇博士获得香港大学计算机科学博士学位 (2014-2018 年)，他还于 2018 年在芝加哥大学担任访问学者。李镇博士荣获 2023 年吴文俊人工智能优秀青年，2021 年中国科协第七届青年托举人才，2023 年 IROS 最佳论文 Finalist，6 次获得公开竞赛/数据集冠军等。李镇博士还获得了来自于国家、省市级以及工业界的科研项目（如华为青年科学家奖励捐赠、腾讯犀牛鸟项目等）。他领导了港中深的 Deep Bit Lab（https://mypage.cuhk.edu.cn/academics/lizhen/），其主要的研究方向是三维视觉，深度学习等基础理论算法研究，并致力于将人工智能算法推广应用于交叉学科，自动驾驶，具身智能，医学大数据分析等场景中，在该方向著名国际期刊和会议发表论文 80 余篇，包括顶级期刊 Cell Systems, Nature Communications, T-PAMI, IJCV, TMI, TVCG, TNNLS 等，以及顶级会议 CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML，IROS, ACM MM, AAAI, IJCAI, MICCAI 等。李镇博士担任 IEEE Transactions on Mobile Computing、IROS 副编、ICLR2024 AC 以及众多顶刊顶会的审稿人，李镇博士还是广东院士联合会脑科学与类脑智能专委委员，VALSE、MICS、CSIG-MV、3DV 专委会等学术组织的委员。

组织者：贾伟 合肥工业大学

简介：贾伟，合肥工业大学计算机与信息学院（人工智能学院）教授，博士生导师，智能科学与技术系系主任。VALSE 常务委员。中国图象图形学学会青年工作委员会副主任。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。已获得多项国家自然科学基金的资助。已在 TPAMI、IJCV、TIP、CVPR、ICCV 等国际顶级会议及权威期刊上发表 CCF A 类及中科院 1 区论文 60 多篇，论文被引近 7000 次，H 因子 39。担任《中国图象图形学学报》编委。2020-2024 连续 5 年入选全球前 2%顶尖科学家榜单。主要研究兴趣为计算机视觉、生物特征识别、自动驾驶等



组织者：贾旭 大连理工大学

简介：贾旭，大连理工大学副教授，专注于计算机视觉与人工智能领域的研究，在 TPAMI、TIP、CVPR、ICCV 等国际高水平期刊和会议上发表学术论文 60 余篇，谷歌学术累计引用达到 1 万余次，其中 4 篇引用超过 1000 次，H 影响因子 32，成果获得包括诺贝尔奖、多国院士等权威学者正面评价，已申请和授权国内外发明专利 20 余项。主持多项国家级项目或重点项目子课题，相关研究成果获得 CCF 自然科学二等奖（第一完成人）、华为火花奖、

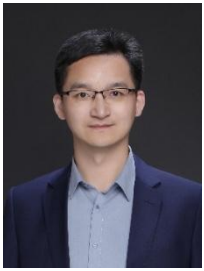
以及 CVPR 形状恢复挑战赛冠军等多项学术奖励。目前为 CCF、CSIG 和 CAAI 等多个专委会执委，以及 Valse 执行领域主席，多次担任 ICLR、ACM MM、IJCAI 等国际顶会领域主席或高级程序委员，并在 CVPR、ECCV 等国际顶会上组织多次研讨会。

李弘扬 香港大学

报告题目：迈向具备高泛化性的通用自动驾驶世界模型

报告摘要：为了打破高昂数据采集成本的限制，并提升模型的泛化能力，我们从互联网广泛获取驾驶视频数据，构建出的数据集涵盖全球范围内超过 2000 小时的驾驶视频，场景涵盖多种天气状况与复杂交通环境。GenAD、Vista 系列工作融合了最新的扩散模型，通过引入时序推理模块，能够有效地应对驾驶场景高度动态的挑战。实验表明，模型可以泛化至训练数据中没有的场景，性能超越现有通用或面向驾驶的视频预测模型。此外，模型还可以被用作动作条件预测模型或运动规划器，为真实世界中的规划任务服务。

讲者简介：李弘扬，香港大学数据科学研究院助理教授，OpenDriveLab 团队（opendrivelab.com）联合创始人。研究方向为端到端智能系统在机器人、自动驾驶的应用。他主导的端到端自动驾驶方案 UniAD 于 2022 年提出，获 IEEE CVPR 2023 最佳论文奖。UniAD 等系列工作产生了明显的社会经济效益，包括特斯拉于 2023 年推出的端到端 FSD。他构造的超大规模具身智能训练场 AgiBot World, 是业界首个百万真机、千万仿真数据集，系统研究具身 Scaling Law 方法论。他提出的俯视图感知方法 BEVFormer，获 2022 年百强影响力人工智能论文榜单，成为业界广泛使用的纯视觉检测基准。他多次担任 CVPR、NeurIPS、ICLR、ICCV、ICML、RSS 等国际会议领域主席（AC），其中获得 NeurIPS 2023 Notable AC。他是《自然·通讯》的审稿人、期刊《Automotive Innovations》客座编委。IEEE、CCF、CSIG 高级会员、IEEE 汽车委员会自动驾驶国际标准工作组组长。荣获 2024 年中国吴文俊人工智能青年科技奖、2023 年上海市东方英才计划领军项目。



马超 上海交通大学

报告题目：自动驾驶多模态场景理解与生成

报告摘要：多模态场景理解与生成是自动驾驶任务的核心与热点问题，本报告主要介绍多模态大模型理解与生成任务的最新进展，汇报如何提高多模态大模型对复杂场景里目标的指代推理能力，以及如何实现多种输入条件下统一的多模态大模型可控场景生成。

讲者简介：马超，上海交通大学人工智能研究院教授，博士生导师。上海市浦江人才、中国图象图形学学会优博。上海交通大学与加州大学默塞德分校联合培养博士。澳大利亚机器人视觉研究中心(阿德莱德大学)博士后研究员。主要研究计算机视觉问题。谷歌学术引用 1 万 3 千余次，连续入选爱思唯尔中国高被引学者（2020-2024）。任中国图象图形学学会优博俱乐部主席、青年工作委员会副秘书长。担任 CVPR、ICCV、ICLR 等会议领域主席，IEEE

Trans. on Multimedia (TMM)、Journal of Artificial Intelligence Research (JAIR)编委。主持国家自然科学基金青年项目(B 类)。获中国图象图形学学会青年科学家奖、MMM 2024 唯一最佳论文奖、华为技术合作领域 2021 年度优秀技术成果奖。

王新宇　华为技术有限公司

报告题目：面向未来智能辅助驾驶的 WEWA 架构

报告摘要：近年来，随着城区和高速智能辅助驾驶可用场景和使用体验快速提升，智能辅助驾驶已经得到消费者认可，并成为购车的重点考虑因素。面向未来 L3~L4 级别辅助驾驶产品，ADS 发布了 WEWA 架构。该架构包含云端 World Engine 和车端 World Action Model 两大部分。我们利用海量车辆驾驶数据（包含脱敏的多传感器原始数据和人类驾驶、接管、辅助驾驶行为数据）训练 World Foundation Model，在云端重建数据湖中每段驾驶数据的 4D 场景理解和驾驶行为决策思维链，同时通过高并发的仿真环境，支持车端模型端到端强化学习训练。车端 World Action Model 通过原生全模态训练，实现多传感器融合的 4D 表征、视听触觉全模态感知、复杂场景多步推理和多车多模行为规划，并通过自研 SOC 实现车端高效推理。新架构可助力 ADS 4.0 实现更类人、更好用的全场景辅助驾驶能力。

讲者简介：王新宇，华为 ADS 算法首席专家，ADS 算法研究与技术开发部部长。负责 ADS 业务和算法架构设计，下一代 ADS 算法预研。研究方向为智能驾驶端侧及云侧算法解决方案，具身智能。他主导了 ADS 1.0 城区高速智能辅助驾驶首版本业务架构设计交付，实现城区智能辅助驾驶首商用。主导 ADS 2.0 无图智驾完整方案设计交付，国内首家实现“全国都能开，有路就能开”。主导 ADS 3.0 端到端及交互模态架构交付，大幅提升城区智驾体验。曾获华为公司十大发明、国防科技进步一等奖等奖励。

王乃岩　小米

报告题目：什么是适合物理世界 AI 的表征学习？

报告摘要：在过去几年，多模态大语言模型（MLLM）很大程度革新了 AGI 认知。虽然 MLLM 在视觉理解和推理方面取得了显著的进步，但在自动驾驶与机器人领域仍有诸多难题有待突破。我们认为其核心原因在于现有的多模态特征提取器。现阶段常用的特征提取器多基于单帧海量互联网数据训练，强于语义理解。然而物理世界的 AI 需要的特征不仅仅止步于此，还需要对于几何、运动、时序的理解。现有的大规模视觉和 3D 预训练模型方法很难满足于这样的需求。在本场演讲中，我会介绍我们最近的两个相关工作，希望能够启迪后续的研究工作。

讲者简介：王乃岩——现为小米汽车自动驾驶杰出科学家。他曾认图森未来中国 CTO，在图森未来负责自动驾驶卡车技术研发，曾率领团队完成了中国第一个在公开高速上的全无人自动驾驶卡车演示。他于 2011 年本科毕业于浙江大学，2015 年博士毕业于香港科技大学。他还是 2014 Google PhD Fellow 计划入选者（中国仅四人入选）。多次在国际数据

挖掘(KDD Cup/Kaggle Challenge)和计算机视觉（ImageNet LSVRC）比赛中名列前茅，在计算机视觉与机器学习顶级会议与期刊上发表论文 70 余篇，发表论文引用次数已超过 20000 余次。

赵昊 清华大学

报告题目：可控高效的时空世界模型在自动驾驶中的应用

报告摘要：近年来，生成式世界模型作为数据驱动的仿真与训练手段，在自动驾驶系统中展现出巨大潜力。我们围绕“时空统一”“可控生成”“多模态支持”等关键需求，提出了两项创新性工作：UniScene 与 DiST-4D，分别从语义占据建模与度量深度建模两个核心几何表征出发，系统性提升了驾驶场景的多模态生成能力与时空一致性建模能力。UniScene 是首个以语义占据为中心的通用生成框架，采用分层建模范式，首先由 BEV 布局生成空间结构丰富的 3D 语义占据图，进而引导 RGB 视频与 LiDAR 点云的生成过程。借助高效的高斯渲染与稀疏建模策略，UniScene 显著提升了三模态数据的真实感与一致性。DiST-4D 则关注于具备预测能力的 4D 世界建模问题，首次提出时空解耦扩散建模框架，以度量深度作为核心桥梁，分别实现任意时间点的未来场景预测（DiST-T）与任意视角下的空间新视图合成（DiST-S）。为突破真实数据中轨迹分布有限的问题，DiST-4D 引入了自监督的循环一致性策略，增强了在未见轨迹上的泛化能力。整体而言，两项工作分别探索了以“语义占据”与“度量深度”为核心表征的可控生成范式，在生成质量、多模态一致性与下游任务适用性方面均取得了显著突破，构建了面向未来自动驾驶系统的高效、可扩展的时空世界建模范式。

讲者简介：赵昊，清华大学智能产业研究院（AIR）助理教授，于清华大学电子工程系获得学士和博士学位，在北京大学从事博士后研究，曾任英特尔中国研究院研究员。主要研究方向是与机器人相关的计算机视觉。在 CVPR / ICCV / ECCV 等顶级学术会议以及 T-PAMI / IJCV 等顶级学术期刊上发表了 50 余篇研究论文。主导研发了全球首个开源的模块化真实感自动驾驶仿真器 MARS，在 CICAI 2023 获得 Best Paper Runner-up 奖项。其研发的渲染阶段可调整精度速度的神经渲染方法 SlimmeRF 于 3DV 2024 获得 Best Paper 奖项。

高晋 中国科学院大学

报告题目：基于扩散模型的时空一致 4D 内容生成初探

报告摘要：视觉基座模型的快速发展使得面向动态物体或场景的 3D 重建和生成（也被称作 4D 生成）有了质的飞跃。这体现在，整个重建和生成过程不再依赖严格同步的多视角视频采集手段，抑或是特定场景下的人体或人脸模型，而是面向数据获取更加容易、物体类别更加广泛的动态开放场景来实现强大数据先验驱动的时空一致新视角预测。这对于未来面向通用人工智能、具身智能或自动驾驶合成大量数据、构建可交互世界模型至关重要。本次报告以 4D 内容生成辅助合成数据为切入点，

重点介绍所在团队在基于扩散模型的时空一致 4D 内容生成领域的两个初步探索工作，并简单介绍在自动驾驶领域的一些应用研究，来共同探讨如何促进相关领域的发展。

讲者简介：高晋，中国科学院自动化研究所多模态人工智能系统全国重点实验室研究员，硕士生导师。长期从事视觉目标自主感知与理解研究，在包括 IEEE TPAMI、IJCV、IEEE TIP、NeurIPS、ICML、CVPR、ICCV、ECCV 等重要国际期刊和国际会议发表学术论文 40 余篇。主持国家自然科学基金联合基金重点、青年科学基金（B 类）、北京市自然科学基金杰出青年科学基金等项目 10 余项。开发的时敏视觉目标自主感知和移动机器人视觉感知技术在国防和民用领域得到实际应用。

范略 中国科学院自动化研究所

报告题目：生成和重建一体化的自动驾驶世界模型

报告摘要：世界模型是自动驾驶视觉仿真的重要手段之一。当前该方向的主流方案通常基于视觉内容生成模型，以图像帧、条件描述以及动作控制等条件为输入，生成符合输入先验的视频/图像内容。由于缺乏显式的物理和几何约束，此类世界模型在时空一致性、精确视角控制能力等方面面临挑战。此外，显式 3D/4D 重建方法有着较好的时空一致性和绝对精准的视角控制，但其在未见自由视角上的泛化性较差。因此，我们团队结合两者优势，提出了 FreeVS、FreeSim 以及 FlexDrive 系列工作，实践了“生成重建一体化建模”的技术设想，突破了当前单一技术路线的局限性，一定程度上实现了视角高度自由、动作精准可控的自动驾驶视觉仿真。

讲者简介：范略，中国科学院自动化研究所模式识别实验室助理研究员，2024 年博士毕业于自动化所模式识别实验室，长期从事驾驶环境的感知与建模研究，在 IEEE TPAMI、CVPR、ICCV、NeurIPS、ICLR 等国际顶级会议刊物上发表论文 20 余篇。博士期间主导研发了用于针对高速自动驾驶环境的全稀疏 LiDAR 感知系列算法，支撑了超远距离（>500m）的实时感知，同时主导研发了首个超越人类平均标注水平的雷达点云物体标注平台，承担腾讯犀牛鸟等业界前沿课题，是科技创新 2030 重大项目-青年科学家项目的子课题负责人。

Panel 嘉宾：金鑫 宁波东方理工大学

简介：金鑫，新型研究型大学-宁波东方理工大学(暂名) 助理教授、博导，中科大博士、新国立访问学者、浙江省青年拔尖人才，曾获中科院院长特别奖、IEEE 电路与系统学会第二届视觉信号处理与通信新星奖 Rising Star 2024、ACM SIGAI China 国际计算机学会中国人工智能分会优博、安徽省优博。研究兴趣包括计算机视觉及多媒体技术，一作及通讯作者发表 CVPR、ICCV、ECCV、NeurIPS、TPAMI、TIP、TMM 等国际期刊与会议论文 40 余篇，论文总引约 5000 次，多项成果被微软、阿里、吉利汽车等企业集成采用。在 CVPR 2024 和 ECCV 2024 等 AI/CV 顶会，组织表征解耦学习与组合泛化 Tutorial，担任 IEEE VSPC 视觉信号处理与通信专委会、CSIG 多媒体专委会、CAAI 具身智能专委会首届委员、VALSE 执行 EAC、IEEE ICIP&VCIP 2024、IEEE ICME 2025 领域 AC。

Workshop 15: 人体动作理解与生成

组织者：徐婧林（北京科技大学）、曾润浩（深圳北理莫斯科大学）、涂志刚（武汉大学）

时间：6 月 7 日（周六）13:30-17:00　　地点：四楼大湾区厅 2

时间	主持人	内容
13:30-14:00	徐婧林	讲者：彭宇新（北京大学） 题目：基于多模态大模型的视觉内容理解与生成
14:00-14:30	徐婧林	讲者：林巍峣（上海交通大学） 题目：语义驱动的视频事件内容感知推理与编码
14:30-15:00	徐婧林	讲者：张姗姗（南京理工大学） 题目：知识驱动的人-物体交互理解与生成
15:00~15:10	中场休息	
15:10-15:40	曾润浩	讲者：秦杰（南京航空航天大学） 题目：“分久必合”：视频动作检测-分割-预测一体化方法研究
15:40-16:10	曾润浩	讲者：涂志刚（武汉大学） 题目：“以人为中心”视频行为识别与生成
16:10-17:00	徐婧林	Panel 嘉宾：彭宇新（北京大学）、林巍峣（上海交通大学）、张姗姗（南京理工大学）、秦杰（南京航空航天大学）、涂志刚（武汉大学）、曾润浩（深圳北理莫斯科大学）

组织者：徐婧林 北京科技大学

个人简介：徐婧林，北京科技大学智能科学与技术学院副教授，北京图象图形学学会理事、副秘书长，中国图象图形学学会青托俱乐部副主席。入选第九届中国科协青年人才托举工程、获 2024 年 CSIG 石青云女科学家奖、2022 年 CSIG 优秀博士学位论文奖、2023 年中国自动化学会自然科学一等奖（4/5）、2024 年 CSIG 自然科学二等奖（3/5）。主要研究方向为细粒度运动分析、视频理解等。已发表 ACM/IEEE Trans.和 CCF A 类论文 30 余篇。主持国家自然科学基金面上项目、青年基金、北京市自然科学基金面上、中国博士后科学基金面上等项目。担任《Chinese Journal of Electronics》青年编委、《电子与信息学报》编委、《计算机科学》青年编委等。

组织者：曾润浩 深圳北理莫斯科大学

个人简介：曾润浩，博士，深圳北理莫斯科大学人工智能研究院长聘副教授，博士生导师。广东省重大人才工程青年拔尖人才，深圳市科技创新人才，深圳市鹏城孔雀人才，广东潮博智库专家。研究领域涵盖机器学习、计算机视觉与多模态学习等，核心方向包括图结构化数据分析、视频动作识别、情绪识别和多模态大模型，在 IEEE TPAMI、IEEE TIP、CVPR 等国际顶级期刊会议发表论文 20 余篇，谷歌学术总引 1900 余次，单篇最高引 600 余次。在视频时序动作分析领域首创基于图结构的时空表示方法，在 THUMOS14 权威基准连续 14 个月排名全球第一，成果收录于科普教材《机器视觉》，已印 3000 册，获评第 31 届书博会少儿阅读节百种优秀图书并输出两种外语版权。近三年主持国家自然科学基金项目、广东省教育厅重点领域项目等纵向科研项目 7 项。获中国图象图形学学会优博提名奖（2023 全国 7 人），IEEE 杰出组织奖，成果入选计算机视觉国际顶级会议 CVPR2024 最佳论文候选（超 1 万篇投稿中的 24 篇之一）。受邀担任 NeurIPS、CVPR 等人工智能领域顶级会议和 TPAMI、TIP 等权威期刊的程序委员会委员和审稿人。担任国际会议 IEEE SmartIoT 2024 本地主席、CSIG 青科会 2023 论坛主席，广东图象图形学会计算机视觉专委会委员。

组织者：涂志刚 武汉大学

个人简介：涂志刚，武汉大学研究员，湖北省杰青，博士生导师。研究领域：人工智能、计算机视觉，聚焦“以人为中心”视频行为识别、重建与生成。发表高水平论文 80 余篇，第一/通讯作者中科院 1 区 Top SCI 期刊+CCF A 类顶会论文近 40 篇。获 2022 年湖北省自然科学二等奖（排名 1）等省部级科技奖励 3 项。主持国家重点研发课题、湖北省杰出青年基金、国家自然科学基金、教育部联合基金（青年人才类）、腾讯犀牛鸟基金

（技术创新奖）等科研项目。指导学生获 2024 中国国际大学生创新大赛-高教主赛道“全国金奖”、国家自然科学基金“青年学生项目”。担任中国仿真学会-视觉计算与仿真专委会副秘书长等职务。开发了视频人体行为智能识别系统，成功应用“第七届世界军人运动会开闭幕式”等多个领域，被央视新闻/体育频道采访报道。

彭宇新 北京大学

报告题目：基于多模态大模型的视觉内容理解与生成

报告摘要：多模态大模型在视觉内容理解与生成的协同进化上展现出巨大潜力，也面临关键挑战。在视觉内容理解上，真实世界的细粒度和多模态特性对大模型提出挑战；在视觉内容生成上，如何生成内容真实、逻辑合理且语义一致的视觉内容是需要研究的关键问题。围绕上述难题，本团队在细粒度多模态大模型、AIGC 等方面进行了相关研究，推动多模态大模型赋能视觉内容的理解与生成。

讲者简介：彭宇新，北京大学二级教授、博雅特聘教授、CAAI/CIE/CSIG Fellow、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、中国工程院“人工智能 2.0”规划专家委员会专家、中国人工智能产业创新联盟专家委员会主任、中国图象图形学学会副秘书长、奖励委员会副主任、北京图象图形学学会副理事长。主要研究方向为多媒体分析、计算机视觉、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、国家自然基金重点、北京自然基金联合基金重点、发改委专项等 40 多个项目。发表 TPAMI、IJCV、CVPR、NeurIPS、ICML 等 ACM/IEEE Trans.和 CCF A 类论文 130 多篇，获最佳论文奖 2 次。10 次参加由美国国家标准技术局 NIST 举办的国际评测 TRECVID 视频搜索比赛，均获第一名。成果应用于国家网信办、公安部、国家广播电视总局等重要单位以及华为、腾讯、快手、蔚来、美团、中国电信、中国铁塔等头部企业。IEEE TCSVT 高级领域编委、IEEE TMM 等期刊编委，培养博士生获中国计算机学会、中国电子学会等优博。

林巍峣 上海交通大学

报告题目： 语义驱动的视频事件内容感知推理与编码

报告摘要： 随着多媒体应用与服务的迅速发展，视频中的事件、行为、属性等语义信息在大规模多媒体系统中的应用日益重要，因此，对语义信息的精准提取及高效压缩等需求，变得日益显著。在本次报告中，将介绍本课题组在大规模事件语义信息提取与压缩方面的一些工作。首先，介绍目标行为和事件语义提取方面的工作，通过对当前的行为事件识别与定位架构进行的重新建模，并提出从全局到局部的渐进行为事件提取架构。其次，介绍复杂事件步骤对齐及因果推理的工作，通过对复杂事件的步

骤挖掘、对齐，并对事件中的因果关系进行推理分析，实现对视频中复杂事件的全面理解。第三，本报告还将介绍课题组在事件语义信息压缩编码方面的工作，设计了面向关键点序列及因果关系图等关键视频语义内容的压缩编码架构，实现了平均 60%以上的码率节省。最后，将介绍一些在实际场景的应用演示。

讲者简介：林巍峣，上海交通大学特聘教授，主要研究方向为人工智能，视频分析，视觉表征编码等。在相关领域发表权威期刊和会议论文 100 余篇，获发明专利 25 项。获 DanceTrack、GigaDetection 等多项国际权威评测第一；多项技术被国际和国内权威视频编码标准采纳，并牵头制定视觉特征编码团体标准。获得中国高交会、工博会优秀产品奖，中国产学研合作创新奖。入选 IEEE 多媒体计算中期职业成就奖、IEEE ICME 多媒体学术新星奖。任多个权威期刊编委、会议领域主席及权威标准化工作组组长。获国家杰出青年科学基金、国家自然科学基金联合重点基金等项目资助。

张姗姗　南京理工大学

报告题目：知识驱动的人-物体交互理解与生成

报告摘要：基于视觉的人-物体交互检测是计算机视觉领域的一项重要任务，旨在从图像或视频中识别人体动作、交互物体以及它们之间的交互关系，相关技术在视频监控、智能机器人等领域具有广泛的应用价值。通过分析我们发现，没有直接接触的人-物体对之间由于人体动作不明晰，导致其交互关系难以辨识，是该领域的一项主要挑战。本报告将介绍我们课题组利用先验知识引导模型理解复杂交互关系的相关工作，并进一步探索如何利用多视角、多姿态人体图像生成技术辅助人-物体交互理解。

讲者简介：张姗姗，南京理工大学计算机学院（人工智能学院）教授、博士生导师，国家优青、江苏省杰青获得者，研究领域为模式识别与计算机视觉。博士毕业于德国波恩大学，并曾在德国马普计算机研究所担任博士后研究员。2018 年入选中国科协“青年人才托举工程”、微软“铸星学者”计划；2021 年获得中国图象图形学学会石青云女科学家奖；2022-2024 连续三年入选爱思唯尔中国高被引学者；2024 年获得 CAAI-华为昇思 MindSpore 学术基金优秀项目奖励。目前担任模式识别权威期刊 Pattern Recognition 编委、CVPR、ICCV 领域主席、中国人工智能学会模式识别专委会副秘书长、江苏省“社会安全图像与视频理解”重点实验室副主任、VALSE 常务领域主席。

秦杰 南京航空航天大学

报告题目：“分久必合”：视频动作检测-分割-预测一体化方法研究

报告摘要：视频动作检测、分割与预测是人体动作理解与视频内容分析的关键问题，在智能安防、智慧体育、自动驾驶、具身智能等领域有着重要的应用价值。本报告首先聚焦时序动作检测、分割与定位等任务的难点问题，分别介绍团队在

上述方面的研究成果，主要包括基于上下文感知的动作检测网络 ACGNet，基于预测对比编码的动作分割网络 PACE，以及提名无关的动作定位优化框架 RefineTAD；在此基础上，进一步发掘上述任务间的共性与互补性，介绍近期在视频动作检测-分割-预测一体化方法研究方面的一些初步尝试。

讲者简介：秦杰，南京航空航天大学人工智能学院教授、博士生导师、院长助理，脑机智能技术教育部重点实验室副主任，国家级青年人才，中国科协海智特聘专家，中国计算机学会高级会员。本科/博士毕业于北京航空航天大学，博士后师从“马尔奖”得主 Luc Van Gool。目前主要从事人工智能、计算机视觉、具身智能和多媒体等领域的基础理论与关键技术研究。已在国际权威期刊和会议上发表论文 100 余篇，其中 CCF A 类国际顶级期刊和会议论文近 50 篇（包括 4 篇 IEEE-TPAMI、2 篇 IJCV、4 篇 IEEE-TIP、14 篇 CVPR 等），Google Scholar 引用 5100 余次，H 指数 37。获中国图象图形学学会自然科学奖二等奖（排名 1）、CCF A 类会议 ACM MM 2023 唯一荣誉提名奖（1/3072，南航本科生一作）、CCF B 类会议 ICME 2024 最佳论文提名（南航硕士生一作）、CCF T1 类期刊《计算机研究与发展》优秀论文奖等。担任 CCF A 类期刊 IJCV 客座编委、CCF B 类期刊 Neural Networks 副主编、CCF A 类会议 NeurIPS/AAAI/IJCAI/ACM MM 领域主席、计算机视觉顶级会议 ECCV 研讨会主席、CCF B/C 类会议 ECAI/IJCB/PRCV 领域主席等。主持国家海外高层次人才引进计划青年项目、国家自然科学基金面上项目、江苏省自然科学基金青年项目等国家级/省部级课题。

涂志刚 武汉大学

报告题目：“以人为中心”视频行为识别与生成

报告摘要：报告摘要：视频理解作为计算机视觉领域的核心任务成为推动人工智能发展的重要力量，其中“以人为中心”的视频行为识别与生成作为视频理解的关键技术具有广泛应用而一直都备受关注。首先，报告将以人体骨骼姿态为引线，介绍武汉大学行为理解与视觉感知研究组（HUVPRLab）在“以人为中心”视频行为识别与生成方面的系列工作，主要包括人体姿态估计、人体行为识别、人体行为生成与迁移，形成了以人体“骨骼姿态为载体+运动捕捉为动力”的视频行为理解研究范式。最后，报告将总结和展望“以人为中心”视频行为识别与生成的发展趋势。

讲者简介：涂志刚，武汉大学研究员，湖北省杰青，博士生导师。研究领域：人工智能、计算机视觉，聚焦“以人为中心”视频行为识别、重建与生成。发表高水平论文 80 余篇，第一/通讯作者中科院 1 区 Top SCI 期刊+CCF A 类顶会论文近 40 篇。获 2022 年湖北省自然科学二等奖（排名 1）等省部级科技奖励 3 项。主持国家重点研发课题、湖北省杰出青年基金、国家自然科学基金、教育部联合基金（青年人才类）、腾讯犀牛鸟基金（技术创新奖）等科研项目。指导学生获 2024 中国国际大学生创新大赛-高教主赛道“全国金奖”、国家自然科学基金“青年学生项目”。担任中国仿真学会-视觉计算与仿真专委会副秘书长等职务。开发了视频人体行为智能识别系统，成功应用“第七届世界军人运动会开闭幕式”等多个领域，被央视新闻/体育频道采访报道。

Workshop 16: 深度连续学习研讨会

组织者：刘夏雷（南开大学）洪晓鹏（哈尔滨工业大学）张幸幸（清华大学）

时间：6 月 7 日（周六）13:30-17:00　　地点：四楼大湾区厅 4

时间	主持人	内容
13:30~14:00	洪晓鹏	讲者：邓成（西安电子科技大学） 题目：基于脑启发的深度增量学习
14:00~14:30	洪晓鹏	讲者：周宇（南开大学） 题目：增量目标检测技术研究
14:30~15:00	张幸幸	讲者：叶翰嘉（南京大学） 题目：基于模型兼容的增量学习方法探究
15:00~15:30	张幸幸	讲者：余璐（天津理工大学） 题目： 持续学习中的少遗忘、强鲁棒和可解释研究
15:30~15:40	中场休息	
15:40~16:10	刘夏雷	讲者：朱飞（中国科学院香港创新研究院人工智能与机器人创新中心） 题目： 多模态持续学习
16:10~16:40	刘夏雷	讲者：张幸幸（清华大学） 题目：预训练下的持续学习理论、方法与应用
16:40~17:00	刘夏雷	Panel 嘉宾：邓成（西安电子科技大学）、周宇（南开大学）、叶翰嘉（南京大学）、余璐（天津理工大学）、朱飞（中国科学院香港创新研究院人工智能与机器人创新中心）、许可乐（国防科技大学）

组织者：刘夏雷 南开大学

个人简介：刘夏雷，南开大学计算机学院副教授，研究方向为开放环境视觉连续学习。入选南开大学“百名青年学科带头人培养计划”，入选第九届中国科协青年托举计划，博士生导师。博士毕业于西班牙巴塞罗那自治大学，博士后工作于英国爱丁堡大学。长期从事连续学习、无监督学习和小样本学习等面向开放环境的机器学习和计算机视觉问题。至今共发表学术论文 40 余篇，谷歌学术引用 4500 余次。包含国际顶级期刊和会议 TPAMI、NeurIPS、CVPR、ICCV 等，一篇文章入选 CVPR 2022 Best Paper Finalists。担任 VALSE 2022-2025 组委会成员，组织 CVPR 2023 年连续学习 Workshop，获第二届粤港澳大湾区 (黄埔)国际算法算例大赛“序列任务的连续学习”冠军。

组织者：洪晓鹏 哈尔滨工业大学

个人简介：洪晓鹏，哈尔滨工业大学教授，IEEE 资深会员，VALSE 资深 AC。斯坦福大学全球前 2%顶尖科学家年度榜单入选者。已在顶级国际和国内刊物和国际会议上发表文章 80 余篇，相关工作见诸美国《麻省理工技术评论》等技术媒体专文报道。2 次获得领域内国际权威期刊和会议的优秀论文奖，5 次带队获得国际评测冠军。作为负责人主持了国家重点研发计划课题、国家自然科学基金、芬兰信息学会博士后基金等 10 余个项目。多次受邀担任 ACM MM 、 AAAI/IJCAI 的领域主席或资深程序委员以及 CVIU 和 IVC 等国际主流期刊的编委。中国图像与图形学学会情感计算与理解专业委员会副秘书长，黑龙江省计算机学会学术工作委员会副主任、奖励工作委员会秘书长。目前的研究领域包括：多模态目标感知、深度连续学习、智能决策与任务分配等。

组织者：张幸幸 清华大学

个人简介：张幸幸，清华大学助理研究员，中国电子学会优博，北京市优秀毕业生。主要研究方向为数据-时空高效的机器学习及其在具身智能中的应用，在 IROS 2024 足式机器人挑战赛中获得亚军。相关成果发表在 Nature Machine Intelligence (封面文章)、TPAMI、NeurIPS (Spotlight)、ICLR、ECCV 等国际顶级期刊与会议，累计发表论文近 50 篇，授权专利 10 余项，主持多项国家级和省部级科研项目。曾系统梳理持续学习领域的发展脉络并提出未来方向，其综述在 TPAMI 发布后受到人工智能社区广泛关注，相关内容在 Twitter 平台浏览量超过 5 万。

邓成，西安电子科技大学

报告题目：基于脑启发的深度增量学习

报告摘要：人类与生俱来具有终身不断获取、整合和迁移知识的能力，这种学习能力被称之为增量学习（Incremental Learning）。在机器学习领域，特别是深度学习模型提出以来，增量学习致力于解决模型训练的一个普遍缺陷——灾难性遗忘（Catastrophic Forgetting），即在新任务上训练时，在旧任务上的表现通常会显著下降。本报告以人脑学习与记忆机制为引入，分析探讨现有增量学习主流范式，并提出增量学习未来发展方向。

讲者简介：邓成，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，中国图像图形学会高级会员、中国计算机学会高级会员。主要研究方向为增量学习理论与方法、多模态数据协同计算理论与方法，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。现担任国际知名期刊 Pattern Recognition、Neurocomputing 副编辑；担任多个国际学术会议的高级程序委员/程序委员，如 ICML、NeurIPS、ICCV、CVPR、KDD、AAAI、IJCAI 等。研究成果获 2019 年、2023 年陕西省自然科学一等奖、2016 年国家自然科学二等奖。

周宇 南开大学

报告题目：增量目标检测：关键挑战与技术进展

报告摘要：增量目标检测任务要求目标检测模型在持续学习新类别的同时，有效避免对旧类别知识的遗忘，从而兼具快速适应新知识的“可塑性”与保持原有知识的“稳定性”。然而，在实际增量场景中，模型往往面临标注数据稀缺、数据存储空间受限、数据隐私保护需求等诸多挑战。针对上述问题，本报告将围绕以下几个方面展开深入探讨：(1) 首先明确增量目标检测的任务定义，剖析其中存在的核心技术难点与瓶颈；(2) 分析知识蒸馏方法在增量检测任务中的应用，阐述如何有效地从旧模型中挖掘知识，并迁移指导新任务的学习；(3) 介绍数据重放策略的最新进展，探讨目标级数据重放如何在资源受限条件下高效缓解灾难性遗忘问题；(4) 讨论基于 Transformer 的增量检测方法，分析如何利用其架构有效地增强新旧任务的交互能力，并借助参数高效微调优化其域适应能力，进一步提升模型整体性能。最后，本报告将结合增量目标检测技术当前的发展现状与趋势，提出对未来研究方向的深入思考和展望，以期为增量目标检测领域的发展提供有价值的启发。

讲者简介：周宇，南开大学计算机学院教授、博导；CSIG 文档图像分析与识别专委会常务委员、副秘书长。研究方向为 OCR、终身学习及多模态智能等。近年来在 CCF-A 类/SCI 一区会议期刊发表论文近 40 篇，获 CCF-A 类会议 ACM MM 2021 最佳论文提名奖（5/1942 篇）。团队核心技术获得 CSIG 2022 票据识别与分析挑战赛冠军、2020 年“中国人工智能·多媒体信息识别技术竞赛”手写&印刷文本 OCR 两项高校组冠军、ICDAR ReST 2023 印章主

体文字检测第三名等近 10 项学术竞赛奖项。研发的场景文本提取系统、特定目标检测系统、钓鱼网站检测系统等应用于多个国家部委及企业，发挥关键作用。主持国家重点研发计划课题&子课题、国家自然科学基金面上&青年基金项目、国家部委重大工程课题、中国博士后科学基金、企业委托等项目/课题多项

叶翰嘉 南京大学

报告题目：基于模型复用与兼容的持续学习

报告摘要：持续学习关注当新知识不断增加时如何利用已有模型和新增数据进行模型高效更新，并防止模型“灾难性遗忘”。本报告从持续学习不同阶段模型间“兼容性”的角度出发，提出通过模型复用在类别、领域分布发生变化时轻量的更新方法。探究当面临新数据时，如何有效实现当前模型和历史模型在模型权重、特征表示层面的兼容，减缓模型遗忘并平衡模型对已有分布和新分布的判别能力。模型间的“兼容性”也适用于大模型能力的演进，本报告分析视觉信息引入后多模态大模型注意力的变化，设计模态专家模块补偿注意力分布，在多模态大模型持续更新时保持其纯文本问答能力；也将多模态大模型的能力轻量化扩展至多语言方面，应对多语言图文问答。实验结果展现出上述基于复用与兼容思路设计的大小模型在不同持续学习场景中的能力，验证了这一思路的有效性。

讲者简介：叶翰嘉，现任南京大学人工智能学院副教授，在南京大学机器学习与数据挖掘研究所 (LAMDA)从事学术研究工作，研究方向包括表示学习、预训练模型复用等领域。叶翰嘉在人工智能领域发表《IEEE Trans. PAMI》等学术论文 60 余篇，受邀担任国际重要会议 ICML/ NeurIPS/ ICLR/ CVPR/ IJCAI 领域主席、中国计算机学会高级会员、中国计算机学会大模型论坛执行委员；主持国家重点研发计划专项项目、国家自然科学基金面上项目，获中国人工智能学会吴文俊人工智能青年科技奖、中国计算机学会优秀博士学位论文奖。

余璐 天津理工大学

报告题目：持续学习中的少遗忘、强鲁棒和可解释研究

报告摘要：持续学习旨在使模型能够持续获取新知识，同时避免遗忘已有知识，提升模型的泛化能力和长期适应能力。本报告围绕持续学习中的三个关键挑战展开：一是少遗忘，探讨如何通过参数正则化、记忆重放等机制缓解灾难性遗忘；二是强鲁棒，重点介绍模型在对抗攻击下保持稳定表现的方法；三是可解释性，聚焦于如何理解模型的知识更新过程和决策依据，从而提升模型的可信度与可控性。

讲者简介：余璐，天津理工大学“明理学者”副教授，主持国家自然科学基金面上项目、青年项目。主要研究领域为连续学习、鲁棒性研究等，共发表论文 20 余篇，包括 CVPR、

NeurIPS、ICLR 等。担任中国计算机学会多媒体技术专业委员会执行委员。获 China MM 2023 最佳海报奖和 CVPR 2022“持续学习”论坛最佳论文奖。

朱飞 中国科学院香港创新研究院人工智能与机器人创新中心

报告题目：多模态持续学习

报告摘要：多模态持续学习作为持续学习的重要分支，不仅拓展了单一模态下的知识积累能力，还能促进跨模态信息的融合与传递，从而应对复杂场景中任务多样性与模态协同的问题。本报告围绕视觉-语言模型，探讨多模态持续学习思路，并介绍我们在该问题上的初步研究成果。最后，我们对持续学习研究进行总结和展望。

讲者简介：朱飞，中国科学院香港创新研究院人工智能与机器人创新中心助理研究员。2018 年和 2023 年分别在清华大学、中国科学院自动化研究所获学士和博士学位。曾获中国科学院院长特别奖、中国科学院和北京市优秀博士学位论文奖。研究兴趣包括模式识别、机器学习、持续学习与不确定性估计等，在 IEEE-TPAMI、IJCV、Neural Networks、Pattern Recognition、自动化学报等期刊与 CVPR、ECCV、NeurIPS、ICLR 等会议上发表论文多篇。

张幸幸 清华大学

报告题目：从预训练到持续演进：理论、技术与应用

报告摘要：多预训练模型在显著缓解持续学习中灾难性遗忘问题的同时，也大幅提升了模型的泛化能力；而持续学习方法则进一步增强了预训练模型在多样化下游任务中的适应与表现能力。特别是在具身智能领域，持续学习正逐步展现出巨大的研究价值与应用潜力，成为构建通用智能体的重要路径之一。本报告将围绕预训练与持续学习的融合必要性与关键挑战展开，重点介绍脑启发的持续学习机制、参数高效的持续微调方法（涵盖单模态与多模态场景），以及面向具身任务的持续强化学习等前沿研究方向。通过这些方法，有望推动智能体在动态环境中不断习得复杂行为、实现跨任务组合与泛化，最终迈向可持续演进、适应任意任务与本体的通用学习算法。

讲者简介：张幸幸，清华大学助理研究员，中国电子学会优博，北京市优秀毕业生。主要研究方向为数据-时空高效的机器学习及其在具身智能中的应用，在 IROS 2024 足式机器人挑战赛中获得亚军。相关成果发表在 Nature Machine Intelligence (封面文章)、TPAMI、NeurIPS (Spotlight)、ICLR、ECCV 等国际顶级期刊与会议，累计发表论文近 50 篇，授权专利 10 余项，主持多项国家级和省部级科研项目。曾系统梳理持续学习领域的发展脉络并提出未来方向，其综述在 TPAMI 发布后受到人工智能社区广泛关注，相关内容在 Twitter 平台浏览量超过 5 万。

许可乐 国防科技大学

Panel 嘉宾简介：许可乐，国防科技大学计算机学院副研究员，入选省部级人才计划，Kaggle Grandmaster。长期从事多模态机器学习研究。主持国家、军队级项目十余项。在相关智能领域的公认的会议和期刊发表论文 100 余篇（包括 NeurIPS、ICML、AAAI、ACM MM、ASE、TASLP、TGRS、TMI 等 CCF A/B 类论文 60 余篇）。获军队科技进步奖一等奖一项。

Workshop 17: 人工智能大模型安全

组织者：彭春蕾（西安电子科技大学）、刘晗（大连理工大学）、刘艾杉（北京航空航天大学）

时间：6 月 7 日（周六）18:30-22:00　地点：四楼大湾区厅 4

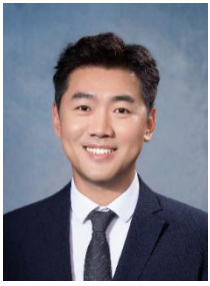
时间	主持人	内容
18:30-19:00	彭春蕾	讲者：韦星星（北京航空航天大学） 题目：多模态大模型可信度评估及增强
19:00-19:30	彭春蕾	讲者：蔺琛皓（西安交通大学） 题目：人工智能大模型安全的完整与可用
19:30-19:40	彭春蕾	企业宣讲：百度（王珂尧） 题目：面向可解释性的多模态人脸活体检测方法研究
19:40-20:10	刘晗	讲者：郭青（南开大学） 题目：面向多模态模型的攻击与防御方法研究
20:10-20:40	刘晗	讲者：吴海威（电子科技大学） 题目：信息泡沫时代如何练就火眼金睛
20:40~20:50	中场休息	
20:50-21:20	刘艾杉	讲者：周文柏（中国科学技术大学） 题目：大模型生成内容安全：从文本到多模态生成内容的安全防护体系
21:20-21:50	刘艾杉	讲者：董胤蓬（清华大学） 题目：面向多模态大模型的安全性与可信性

组织者：彭春蕾 西安电子科技大学

个人简介：彭春蕾，西安电子科技大学教授，博士生导师，西电杭州研究院挂职副院长，网络与信息安全学院导学第四党支部书记。近年来从事视觉身份可信识别方面的研究，具体包括伪造检测、智能生成和鲁棒识别等。发表论文 60 余篇，主持国家自然科学基金面上项目、陕西省重点研发计划、CCF-腾讯犀牛鸟基金等，以排名第二参与 2 项国家自然科学基金重点项目。入选中国科协青年人才托举工程、陕西省青年科技新星，获吴文俊人工智能优秀青年奖、湖北省技术发明一等奖 (3/6)、中国图象图形学学会自然科学二等奖 (3/5)、陕西省优秀博士学位论文、中国图象图形学学会优秀博士学位论文奖。指导研究生入选中国电子学会硕士学位论文激励计划。担任国际期刊 Visual Computer Associate Editor，《网络空间安全科学学报》青年编委等。

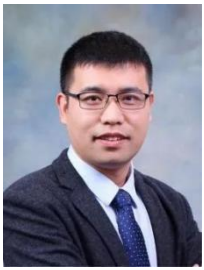
组织者：刘晗 大连理工大学

个人简介：刘晗，大连理工大学副教授，博士生导师，AI+教育研究所副所长。长期从事垂直领域大模型/知识图谱/智能体领域相关研究。在 IEEE TKDE，NeurIPS，KDD，ACL，CVPR，SIGIR，WWW，AAAI，IJCAI 等国际高水平期刊或会议发表论文 70 余篇（一作/通讯 40 余篇）。主持国家自然科学基金，某科委创新特区，教育部春晖计划等各类科研项目十余项。曾获得 CCF 科技成果奖自然科学二等奖，教育部-华为智能基座栋梁之师，ACM 中国学术新星（大连），小米青年学者，字节跳动安全 AI 挑战赛全国冠军等。已联合发布支持软硬件全栈国产化的智能化工、医疗、船务三项垂直领域大模型。



组织者：刘艾杉 北京航空航天大学

个人简介：刘艾杉，北京航空航天大学计算机学院副教授，从事安全可信人工智能方面的研究。累计在国际顶级会议/期刊发表论文 70 余篇 (一作/通信 40 余篇)，申请专利 20 余项，编制国标 10 项、专著/教材 3 本，谷歌学术引用 3000 余次。获省部级科技进步一等奖 1 项、省部级科技进步二等奖 1 项，安全领域四大顶会 (CCF-A 类)ACM CCS 2024 杰出研究成果奖，有关建议被中共中央办公厅采纳。常年担任论坛主席在包括 CVPR、AAAI、IJCAI 等 CCF-A 类国际会议上组织 10 余次“智能安全”系列主题国际研讨会和竞赛，常年担任 NeurIPS、ICML 等 CCF-A 类会议领域主席 (Area Chair)，担任 Pattern Recognition 等 SCI 期刊客座编辑，推动智能安全社区健康发展。



韦星星 北京航空航天大学

报告题目：多模态大模型可信度评估及增强

报告摘要：本报告旨在介绍多模态大模型的可信度评估及增强方法，主要从可信度评估和增强两个方面展开。首先，针对多模态大模型的可信度评估，从事实性、安全性、鲁棒性、公平性和隐私性五个维度设计了详细的评测实例和工具，全面评估并排序了现有主流的多模态大模型。基于此，提出了名为 MultiTrust 的可信度评估框架，该框架能够系统地量化和比较不同模型在各个维度上的表现。其次，基于评估结果，进一步探讨了多模态大模型可信度缺失的内在原因，发现当前多模态大模型普遍依赖的视觉-语言预训练模块存在鲁棒性和安全性方面的欠缺。为了提升多模态大模型的可信度，介绍一种基于特征一致性的视觉语言预训练模型增强方法，通过加强不同模态间表征学习的一致性，优化多模态模型的可信度。

讲者简介：韦星星，北京航空航天大学人工智能学院教授，博士生导师，国家级青年人才。主要研究方向为计算机视觉、安全可信人工智能，在 TPAMI, IJCV 和 CVPR, ICCV 等人工智能领域顶级期刊和会议发表学术论文 80 余篇 (其中 TPAMI、IJCV 论文 10 篇，ESI 高被引论文 5 篇)，授权发明专利 20 余项。多次受邀担任人工智能领域顶级国际会议的程序委员会委员和 TPAMI、IJCV 等期刊的审稿人。作为负责人主持科技部“新一代人工智能”重大项目课题、国家自然科学基金面上项目、军兵种预研项目、华为/腾讯校企合作项目。荣获国防技术发明一等奖 (3/6)，在国际顶会举办的重要竞赛中获得冠亚军 5 项，优秀奖 5 项，被 CCTV 和环球网报道。

蔺琛皓 西安交通大学

报告题目：人工智能大模型安全的完整与可用

报告摘要：随着深度学习、大模型、生成式人工智能等技术快速发展和成熟应用，其可信、安全、可控也广受关注。本报告围绕人工智能大模型的安全可信问题，从信息安全 CIA 三要素中的安全完整性及安全可用性出发，介绍人工智能大模型及其相关应用面临的安全风险和前沿技术。

讲者简介：蔺琛皓，西安交通大学网络空间安全学院教授、博士生导师，教育部青年长江学者，教育部工程研究中心副主任。毕业于西安交通大学、美国哥伦比亚大学、香港理工大学。长期从事人工智能安全、大模型安全、智能身份安全等相关研究，在该方向发表论文 60 余篇，包括 IEEE TPAMI、TDSC、TIFS、TIP、USENIX Security、S&P、ICML、NeurIPS、AAAI、CVPR 等，获得了 IJCAI-DCM 等国际学术会议最佳论文奖 2 次；先后获得了达摩院青橙奖最具潜力奖、吴文俊人工智能优秀青年奖、中国自动化学会自然科学一等奖、人社部高层次留学人才回国资助等；入选了陕西省高层次人才 (青年)、小米青年学者、思源学者等；主持了重点研发计划项目课题，科技创新 2030-“新一代人工智能”重大项目课题，国家自然科学基金重点类项目课题、面上、青年项目等；担任中国自动化学会人工智能与

安全专委会副主任委员，以及 ACM SIGSAC China、中国人工智能学会人工智能与安全等多个专委会委员。

郭青 南开大学

报告题目：面向多模态模型的攻击与防御方法研究

报告摘要：随着多模态模型如视觉-语言模型在自动驾驶、智能助手和内容生成等领域的广泛应用，其安全性挑战日益突出。本报告系统地探讨了针对多模态模型的攻防研究进展，从数字环境到物理世界的攻防测试。在攻击研究方面，我们揭示了增强视觉-语言攻击迁移能力的关键机制，设计了由 LLM 智能体驱动的实体世界攻击方法，并深入剖析了生成式多模态模型的安全漏洞，包括"生成-编辑"协作式越狱攻击和基于个性化特征的后门攻击。在防御研究方面，我们提出了语义感知的隐式表示框架，并基于此研发了鲁棒的视觉重采样技术，并创新性地引入语言引导的连续表示机制增强模型抵御能力。通过这些研究，我们不仅揭示了多模态模型面临的复杂安全挑战，也为构建更安全可靠的下一代多模态系统提供了基础和实用解决方案。

讲者简介：郭青，南开大学，教授，博士生导师，国家级青年人才（海外），入选斯坦福全球 Top 2%科学家。2019 年加入新加坡南洋理工大学任博士后研究员，并于 2020 年获聘为瓦伦堡-南洋理工大学校长博士后（全球 500 选 5 人），2022 年加入新加坡科技研究(A*STAR)前沿人工智能研究中心（CFAR)任高级研究员，2023 年兼职新加坡国立大学助理教授。曾获 ICME 最佳论文奖、ACM 优秀博士论文奖等多项荣誉，在新加坡期间主持多项重大科研项目，累计科研经费约 400 万新币。主要研究方向可靠视觉感知及人工智能安全。在 ICML、NeurIPS、ICLR、CVPR、ICCV、TPAMI、IJCV 等 A 类会议及期刊上发表论文 60 余篇。现担任 ICML、ICLR、NeurIPS、ICCV、IJCAI 领域主席，AAAI Senior PC，VALSE 2023 执行 AC。

吴海威 电子科技大学

报告题目：信息泡沫时代如何练就火眼金睛

报告摘要：在信息泡沫时代，虚假信息泛滥成灾，如同汹涌暗流冲击着我们的认知防线。AI 生成内容的兴起，更使虚假信息鉴别难度倍增。本次报告聚焦“信息泡沫时代下的真伪内容鉴别”，以 AI 治 AI，借助 AI 强大的运算与分析能力，打造智能“防御盾牌”。将深入探讨相关技术方法，助力大家在这场“信息保卫战”中练就火眼金睛。

讲者简介：吴海威，电子科技大学，计算机科学与工程学院（网络空间安全研究院)教授。从事人工智能与网络安全交叉领域研究，聚焦多媒体数字取证与人工智能安全。研究成果"面向社交网络的多媒体安全和取证关键技术"获 2022 年澳门自然科学奖；作为中国通信标准化协会技术参考；应用于阿里巴巴侵权纠纷等业务场景；于 2024 年"外滩大会·全球 Deepfake 攻防挑战赛"获冠军（1500 支队伍)。

周文柏 中国科学技术大学

报告题目：大模型生成内容安全：从文本到多模态生成内容的安全防护体系

报告摘要：大语言模型、文生图和文生视频等大模型的快速发展，这些模型与技术在创造价值的同时也带来了全新的内容安全挑战。本报告聚焦人工智能大模型安全领域的前沿研究，系统性地介绍了从文本到多模态生成内容的威胁评估与防御体系。在语言模型方面，聚焦于有害文本生成问题，提出了创新的评估和缓解框架，致力于构建更加安全可靠的语言模型生态；对于图像生成，我们深入研究了对抗攻击下的安全防护机制，设计了兼顾生成质量与防御能力的解决方案；在视频生成领域，我们探索了基于内容理解的安全控制方法，实现了对生成过程的智能调控与风险管理。这些研究在不同模态中展现了技术创新与实用价值的结合，形成从理论到实践、从文本到多模态的大模型安全防护体系。团队工作为 AI 技术的负责任发展和安全部署提供了切实可行的解决方案也为人工智能的健康发展探索了新的道路。

讲者简介：周文柏，中国科学技术大学网络空间安全学院副教授，博士生导师，中国科大“墨子杰出青年”，微软“铸星计划”青年学者。中国图象图形学会数字媒体取证与安全专委会副秘书长。主要研究兴趣包括信息隐藏和人工智能安全，发表安全领域四大顶会 ACM CCS、IEEE TPAMI、CVPR 等高水平论文五十余篇，获得 ACM CCS 杰出论文奖等多个国际顶会的最佳论文奖项。参与中央 JW 政工部牵头主办的“众智 2022”创新应用大赛获得第一名，参与由 Facebook 发起的全球最高水准与最大规模人脸深度伪造检测挑战赛 DFDC，获全球第二，赢得 30 万美元奖金，入选 2014 年以来中国人工智能安全领域的 8 项创新成果；理论研究成果入选《斯坦福人工智能报告 2022》。主导研发全球最具影响力深度伪造工具 DeepFaceLab，与 OpenAI 研发的 GPT2 共同入选 2020 年十大 Github 开源项目，主导“合成现实技术”，被央视官方解读为“新质生产力”代表性技术，并发布首个“合成现实数字人钱学森”，获得中央电视台、新华网、人民日报等主流媒体专访报道。主持某重点项目、国家自然科学基金等多项课题，项目总经费 2000 余万元。获得国家授权发明专利 5 项，担任 TPAMI、AAAI、CVPR、ICCV、IJCV 等多个顶级期刊会议的组委会成员与审稿人。

董胤蓬 清华大学

报告题目：面向多模态大模型的安全性与可信性

报告摘要：多模态大模型近年来取得了快速发展，深刻改变了人们理解和生成图像、文本等数据的方式，并催生了如 GPT-4o、Gemini、Sora 等代表性成果。然而，尽管多模态大模型取得了巨大成功，其在安全性和可信性方面仍然面临着严峻的挑战。例如，这些模型很容易被诱导生成有害内容，易受对抗性攻击的干扰，且存在显著的隐私风险。本报告将介绍多模态大模型所面临的安全风险以及基于红队对抗的大模型风险高效挖掘方法；进而讨论如何降低大模型的安全风险，提升其安全性；最后介绍面向多模态大模型的可信评测基准。

讲者简介：董胤蓬，清华大学人工智能学院助理教授，本科和博士毕业于清华大学计算机系，主要研究方向为机器学习、人工智能基础理论与安全。发表国际顶级学术会议和期刊论文六十余篇，谷歌学术引用 11000 余次，担任国际学术会议 ICML、NeurIPS、ICLR 领域主席。曾获得 CCF 优秀博士学位论文激励计划、清华大学优秀博士后等。

Workshop 18: 面向移动终端的 AI 图像增强

组织者：程明明（南开大学）、高广谓（南京邮电大学）、郭鑫（华为技术有限公司公司）

时间：6 月 7 日（周六）13:30-17:00　　地点：四楼大湾区厅 5

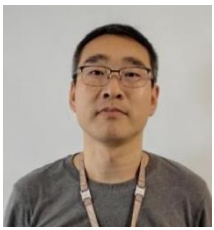
时间	主持人	内容
13:30-13:40	程明明	讲者：郭鑫(华为技术有限公司) 题目：移动终端影像算法中的业务难题
13:40-14:10	程明明	讲者：任文琦（中山大学） 题目：扩散模型驱动图像超分辨技术探索
14:10-14:40	程明明	讲者：张宇伦（上海交通大学） 题目：图像重建轻量化
14:40-14:50	高广谓	企业宣讲：上海合合信息（郭丰俊） 题目：AI 浪潮下的图像内容安全
14:50-15:20	高广谓	讲者：左旺孟（哈尔滨工业大学） 题目：人脸和文本图像复原方法研究
15:20-15:50	高广谓	讲者：王诗淇（香港城市大学） 题目：提升低光图像质量：从客观质量到主观感知的转变
15:50~16:00	中场休息	
16:00-16:30	郭鑫	讲者：郭春乐（南开大学） 题目：面向移动终端的影像 AI 娱乐化应用
16:30-17:00	郭鑫	讲者：周尚辰（南洋理工大学） 题目：视觉内容补全与提取

组织者：程明明 南开大学

个人简介：程明明，南开大学二级教授，新一代人工智能发展战略研究院副院长，媒体计算团队学术带头人。主持承担了国家杰出青年科学基金、优秀青年科学基金项目、科技部重大项目课题等。他的主要研究方向是人工智能、计算机视觉和计算机图形学，在 SCI 一区/CCF A 类刊物上发表学术论文 100 余篇（含 IEEE TPAMI 论文 40 余篇），h-index 为 100，论文谷歌引用 6 万余次，单篇最高引用 5 千余次，多次入选全球高被引科学家和中国高被引学者。技术成果被应用于华为、国家减灾中心等多个单位的旗舰产品。获得教育部自然科学一等奖 2 项、其他省部级科技奖 2 项。培养的 4 名博士生获得省部级优秀博士论文奖。现担任天津市视觉计算与智能感知重点实验室主任、中国图象图形学学会副秘书长、天津市人工智能学会副理事长和顶级期刊 IEEE TPAMI, IEEE TIP 和《中国科学：信息科学》编委。

组织者：高广谓 南京邮电大学

个人简介：高广谓，南京邮电大学先进技术研究院教授，南京理工大学国家重点学科模式识别与智能系统专业博士，研究方向为基于人工智能和深度学习的高效能视觉信息表示、重建以及理解。主持国家级、省部级项目 10 余项，包括国家自然科学基金项目 2 项（青年和面上）、江苏省自然科学基金项目 2 项（青年和优青），以骨干身份参与国家自然科学基金重点项目、科技创新 2030-“新一代人工智能”重大项目各 1 项。在国际权威期刊 IEEE TIP/TMM/TCSVT/TIFS/TITS、PR 以及权威会议 CVPR、AAAI、IJCAI 等上发表学术论文 70 余篇（ESI 高被引论文 4 篇），Google 学术引用 3000 余次。获江苏省科学技术奖一等奖 1 项（7/11），指导研究生获得江苏省优秀专业学位硕士学位论文奖 1 项。担任 IEEE/CCF/CSIG/CAAI 高级会员，中国计算机学会计算机视觉专委会执行委员，中国人工智能学会模式识别专委会委员，中国自动化学会模式识别与机器智能专委会委员，中国图象图形学学会机器视觉专委会委员，江苏省人工智能学会模式识别专委会常务委员，中国图象图形学学会青工委委员，VALSE 执行 AC。



组织者：郭鑫 华为技术有限公司

个人简介：郭鑫，2014 年博士毕业于北京邮电大学，毕业后加入华为公司，一直从事 camera 算法的研发工作。本次 workshop 中将对“华为手机 camera 业务难题”进行介绍，引发讲者们围绕难题进行讨论。

任文琦 中山大学

报告题目：扩散模型驱动图像超分辨技术探索

报告摘要：图像超分辨率作为 AI 图像增强的重要研究方向，在移动终端中的图像细节恢复与视觉质量提升中具有重要应用价值，尤其在移动拍摄、实时视频处理和云游戏等场景中展现出广阔的前景。近年来，扩散模型因其强大的生成能力在图像超分任务中取得显著进展，但在面向移动设备的部署中仍面临推理成本高、先验建模不足和视觉信息遗失等挑战。本报告将围绕扩散模型驱动的图像超分技术展开，介绍团队在扩散模型恢复先验增强、DiT 架构下的双提示图像恢复模型和退化引导的单步扩散模型等方面的研究成果，重点探讨如何在保证重建质量的同时提升模型的效率与轻量化水平。最后，将结合腾讯云游戏中的实际应用，分享图像超分模型在资源受限设备上落地的工程经验与技术优化思路，为移动终端图像增强技术的发展提供新的实践路径与研究方向。

讲者简介：任文琦，中山大学教授，主持国家自然科学优秀青年基金，天津大学与美国加州大学 Merced 分校联合培养博士，从事计算机视觉与多媒体内容安全领域的研究。在 CCF-A 类期刊和会议长文发表学术论文 80 余篇（8 篇 ESI 高被引论文，2 篇热点论文），Google 学术引用 18000 余次，近三年连续入选爱思唯尔中国高被引学者，全球前 2%顶尖科学家榜单。多次担任 CVPR、ICCV、ICLR、NeurIPs 等 AI 和 CV 领域学术会议的领域主席。主持国家自然科学优秀青年基金、联合重点项目、面上，深圳市优秀科技创新人才培养项目，华为、腾讯等企业合作项目 20 余项。获中国计算机学会优博论文奖、吴文俊人工智能优秀青年奖、中国电子学会自然科学一等奖（2/5）、中国图象图形学学会自然科学一等奖（3/5）。

张宇伦 上海交通大学

报告题目：图像重建轻量化

报告摘要：图像重建轻量化旨在在保障重建质量的前提下，显著降低模型参数量与计算成本，提升其在资源受限设备上的部署效率。该方向应用广泛，尤其适用于移动终端、边缘设备及智能监控等场景，有助于推动图像增强、医学影像处理及无人系统感知等任务的实时智能化发展。本报告介绍图像重建轻量化的模型设计、模型压缩与加速、及其端侧部署的科研进展。

讲者简介：张宇伦，上海交通大学，任长聘教轨副教授，入选国家海外高层次青年人才。主要研究方向是计算机视觉和机器学习，具体包括图像/视频复原与合成，模型压缩，计算成像，多模态计算，大语言模型等。在计算机视觉，机器学习，多媒体，人工智能等领域的顶级国际期刊和会议上发表学术论文 100 余篇。论文 Google 学术引用 25000 余次，一作论文单篇最高引用 5800 余次。获得 2015 年 IEEE VCIP 最佳学生论文奖，2019 年 IEEE ICCV RLQ Workshop 最佳论文奖，全球 AI 华人新星百强（2021 年），连续多年入选斯坦福“全球前 2%顶尖科学家”榜单（2021-2024 年），入选 2024 年爱思唯尔“中国高被引学者”。

近年来担任顶级会议 CVPR, ICCV, ECCV, ICLR, NeurIPS, ICML, ACM MM, IJCAI 领域主席。

左旺孟 哈尔滨工业大学

报告题目：人脸和文本图像复原方法研究

报告摘要：人脸和文本是图像内容的重要组成部分，往往更容易成为人们关注的重点，其增强和复原因而有其独特的必要性和研究应用价值。相对于自然图像，人脸和文本呈现出更为显著和稳定的结构化特点（如：脸部组件、笔画）和更易于获得参考图像，因而更容易取得较自然图像更好的复原效果。报告拟回顾和总结团队近年来在人脸和文本图像复原方面的研究进展，主要包括：(1) 基于参考图像的人像复原方法，(2) 人脸/文字的先验建模与复原方法，(3) 人脸和文本图像复原的拓展，包括人脸图像复原驱动的自然图像复原与图像畸变矫正、艺术字生成等。

讲者简介：左旺孟，哈尔滨工业大学计算学部教授。主要从事底层视觉、视觉生成、视觉理解和多模态学习等方面的研究。在 CVPR/ICCV/ECCV/NeurIPS/ICLR 等顶级会议和 IEEE T-PAMI、IJCV 及 IEEE Trans.等期刊上发表论文 200 余篇。曾任 CVPR、ICCV、ECCV 等会议领域主席，现任 IEEE T-PAMI、T-IP、中国科学-信息科学和自动化学报等期刊编委。

王诗淇 香港城市大学

报告题目：提升低光图像质量：从客观质量到主观感知的转变

报告摘要：由于光线不足，场景信号可能被系统噪声掩盖。低光图像增强方法旨在将低光条件下捕获的图像恢复到正常状态，从而提高可见度、对比度并抑制噪声，以改善视觉质量，并为高级计算机视觉任务提供良好的基础。传统的图像增强方法主要集中于优化像素级差异，虽然能够在像素级上产生正确的增强结果，但常常缺乏与人眼质量相关的感知监督。这种方法往往忽视了人眼视觉系统的特性，导致增强后的图像在主观质量上与人类感知存在显著差异。随着深度学习的迅速发展，低光图像增强面临的挑战有了更优雅的解决方案。关键在于如何将人眼对图像质量的主观感知融入增强过程，以确保提升后的图像不仅在客观质量上有所改善，还能更好地符合人类的视觉体验。

讲者简介：王诗淇，香港城市大学计算机系副教授。从事多媒体信号质量评价、压缩、增强处理及分析方面的研究。近年来，专注于多媒体质量评价及视觉增强，在 IEEE 汇刊如 TPAMI, TMM, TCSVT, TIP 等旗舰期刊上发表论文 150 余篇，相关研究成果在谷歌学术上被引用超过 17000 次，H 指数达到 63。担任 IEEE TCSVT, TIP, TMM, TCYB 等国际期刊编委。2024 年香港城市大学工学院研究卓越奖，2023 年第 48 届日内瓦国际发明展金奖和银奖，2021 年 IEEE 多媒体新星奖，2021 年香港城市大学校长奖，并且多次获得最佳论文奖。

郭春乐 南开大学

报告题目：面向移动终端的影像 AI 娱乐化应用

报告摘要：移动终端娱乐化需求日益增长，注重可玩性，交互性的娱乐化功能越来越重要。消费者影像娱乐化的需求主要包括定制化、可交互、多样性这三个维度。一方面影像算法需要更加智能，能够适应用户习惯，产生符合用户预期的结果，另一方面，也需要满足用户能够介入影像产生的过程的需求，使其获得创作的乐趣。更进一步的，影像 AI 娱乐化算法需要能够拓宽维度，创造新的娱乐需求。本报告将围绕“面向移动终端的影像 AI 娱乐化应用”这一主题，介绍本课题组在定制化人像修复、可交互的图像修饰与编辑、以及基于 RAW 数据的实时 HDR 视图合成等任务中的研究进展与实践探索。

讲者简介：郭春乐，南开大学计算机学院副教授，博导，入选南开大学“百名青年学科带头人培养计划”，第十届中国科协青年托举计划，2024 斯坦福全球前 2%顶尖科学家榜以及爱思唯尔中国高被引学者。博士毕业于天津大学，攻读博士期间曾前往伦敦玛丽女王大学和香港城市大学交流访问。主持包括国家自然科学基金、华为、三星等资助的多项科研项目，相关多项专利技术完成成果转化。他的主要研究内容包括计算成像、图像复原与复原、图像生成与编辑等。作为第一（通讯）作者在 TPAMI、TIP、CVPR 等国际学术期刊及会议上发表论文 30 余篇，谷歌学术引用 9000 余次，其中一作论文单篇最高引用 2000 余次。参与组织 2022-2025 年 VALSE 大会，曾任 BMVC2022 领域主席，参与组织 CVPR 2024 年 MIPI Workshop。现担任 SCI 二区期刊 IEEE Journal of Oceanic Engineering 编委。

周尚辰 南洋理工大学

报告题目：视觉内容补全与提取

报告摘要：在自然图像与视频中，视觉内容通常由多个语义图层构成，如背景、前景及光影等。为了支持用户在移动终端上实现更高自由度与更强交互性的内容编辑，亟需解决从观测图像中解耦并重建目标图层的技术挑战。实现高质量的背景补全与精细的前景提取，已成为推动手机端智能影像编辑能力演进的关键问题。本报告围绕“视觉内容补全与提取”主题，介绍我们在夜景光晕去除、带光影物体去除、视频补全与视频人像抠图等任务中的研究进展与实践探索。

讲者简介：周尚辰，新加坡南洋理工大学 MMLab@NTU 研究助理教授，在 CVPR、ICCV、NeurIPS、TPAMI 等会议与期刊上发表论文 40 余篇，谷歌引用超 5000 次；担任 NeurIPS 领域主席，曾于 ECCV 和 CVPR 大会组织“移动智能摄影与成像（MIPI）”系列研讨会。主要研究方向为底层视觉与多模态生成。

Workshop 19：空间智能的感知与决策

组织者：苏航（清华大学）、马月昕（上海科技大学）、赵恒爽（香港大学）

时间：6 月 7 日（周六）18:30-22:00　地点：四楼大湾区厅 5

时间	主持人	内容
18:30-19:00	苏航	讲者：蒋树强（中国科学院大学） 题目：基于具身场景记忆的视觉导航
19:00-19:30	苏航	讲者：杨睿刚（上海交通大学） 题目：Simulating the World from the World
19:30-20:00	马月昕	讲者：Jianlan Luo （UC Berkeley） 题目：Intelligent Physical Agents: High-Performance Learning for Generalist Robots
20:00-20:30	马月昕	讲者：王越（浙江大学） 题目：从视觉反馈控制到 VLA
20:30~20:40	中场休息	
20:40-21:10	赵恒爽	讲者：刘希慧（香港大学） 题目：3D Object Parsing: From Surface Segmentation to Generative Amodal Segmentation
21:10-21:40	赵恒爽	讲者：苏航（清华大学） 题目：跨越虚实鸿沟：基座模型驱动的具身智能泛化之路

组织者：苏航 清华大学

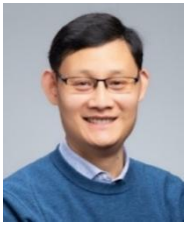
个人简介：苏航，清华大学计算机系副研究员，入选国家“万人计划”青年拔尖人才，主要研究鲁棒机器学习和具身决策等相关领域，发表 CCF 推荐 A 类会议和期刊论文 100 余篇，谷歌学术论文引用 15000 余次，受邀担任人工智能领域顶级期刊 IEEE TPAMI 和 Artificial Intelligence 的编委，IEEE 生成式大模型安全工作组主席，获得吴文俊人工智能自然科学一等奖，ICME 铂金最佳论文、MICCAI 青年学者奖和 AVSS 最佳论文等多个学术奖项，曾率队在 NeurIPS2017 对抗攻防等多个国际学术比赛中获得冠军。现任中国图像图形学会青工委执委、曾任 VALSE 执行 AC 委员会主席，NeurIPS21 的领域主席（Area Chair）、AAAI22 Workshop Co-Chair 等。

组织者：马月昕 上海科技大学

个人简介：马月昕，上海科技大学研究员、助理教授、博导。主要研究方向为三维视觉、具身智能、自动驾驶，在国际顶级期刊和会议上发表论文 60 余篇，其中一作和通讯论文 30 余篇，学术引用 4000 余次。参与指导的论文获 MICCAI 2024 唯一最佳论文奖。

组织者：赵恒爽，香港大学

个人简介：赵恒爽，香港大学计算机科学系助理教授，国家优青。主要研究方向为多模态场景理解与生成，在国际顶级会议和期刊上发表论文 80 余篇，学术引用 34000 余次，曾担任 CVPR、ECCV、NeurIPS 和 ICLR 的领域主席，以及 Pattern Recognition 的副编辑和 IEEE TCSVT 的客座编辑。



蒋树强 中国科学院大学

报告题目：基于具身场景记忆的视觉导航

报告摘要：具身智能是真实物理世界中人工智能的重要表现形态，具身导航是指智能体根据任务目标，感知与理解周围环境并执行移动动作完成任务，这是具身智能系统与真实世界交互的关键技术之一。生理学与认知科学研究表明，记忆在导航任务中起着至关重要的作用，智能体不仅通过当前观察来感知环境，还能通过过往经验和记忆对未观测的环境进行推测，如何将场景记忆与导航任务相结合，为智能体提供预想与决策能力，是一项值得研究的重要问题。本报告将首先介绍具身智能、具身记忆与具身导航的研究背景，并汇报基于具身场景记忆的视觉导航研究进展，包括基于场景知识图的物体导航、结构化具身知识增强的视觉语言导航、基于网格记忆地图的视觉语言导航等具身导航技术，最后介绍具身导航从虚拟到真实环境的适配并给出演示。

讲者简介：蒋树强，中国科学院大学特聘教授，博士生导师，先后担任期刊《IEEE TMM》、《ACM ToMM》、《IEEE Multimedia》、《计算机研究与发展》、《JCST》、《CAD 学报》编委，中国人工智能学会具身智能专委会主任、中国计算机学会多媒体专委会副主任、中国自动化学会网络计算专委会副主任、ACM SIGMM 中国分会副主席，研究方向是多媒体内容分析和具身智能技术。主持承担科技创新 2030-“新一代人工智能”重大项目、国家自然科学基金青年基金 A 类（杰青）、B 类（优青）、重点等项目 20 余项，发表论文 200 余篇，获授权专利 20 余项，多项技术应用到实际系统中，先后获省部级或学会奖励 5 项。

杨睿刚 上海交通大学

报告题目：Simulating the World from the World

报告摘要：AI 面临的一个巨大障碍是数据问题，尤其是在涉及物体和环境操作的具身 AI 领域，收集与文本规模相当的交互数据非常困难。因此，许多研究人员试图通过仿真数据来缓解这一问题。然而，在许多具身 AI 任务中，仿真与真实场景之间的领域差距仍然难以克服。

与纯仿真不同，我们探索了 Real2Sim2Real 的路径，即通过捕获真实世界的数据，在仿真中将其虚拟化以生成多种变体，然后将学习到的策略应用于现实世界。这种方法既保留了现实世界的真实性，又能够利用仿真生成多样性。我们特别关注柔软/可变形物体（例如布料）的捕获与仿真。我将介绍新的传感器设计以及精确且完全可微分的仿真技术，并展示一些初步成果。

讲者简介：杨睿刚 上海交通大学教授， IEEE fellow, 2003 年于美国北卡罗莱纳大学教堂山分校获博士学位，主修计算机科学。曾任美国肯塔基大学计算机系终身教授, 百度研究院机器人和自动驾驶实验室主任, 赢彻科技 CTO, 杨睿刚博士在包括 IJCV、IEEE T-PAMI、SIGGRAPH、CVPR、ICCV 在内的计算机视觉和图形学领域顶级期刊和会议上发表论文 150 余篇，Google Scholar 引用超过两万次，H 指数 74。

Jianlan Luo　UC Berkeley

报告题目：Intelligent Physical Agents: High-Performance Learning for Generalist Robots

报告摘要：Robot learning has advanced significantly in recent years, positioning it as an effective tool for achieving scalable, flexible robotic autonomy. However, the large-scale real-world adoption of such learning-based robotic systems remains challenging, for which they must fulfill stringent real-world performance criteria to be viable. In this talk, I will describe algorithms and principles for building high-performance robotic learning systems. I'll start by examining a range of high-performance "robot specialist" systems.

These systems are tailored to address key deployment factors such as reliability, robustness, and cycle time, which has ultimately paved the way for their industrial adoption. I will then proceed to describe mechanisms to build "robot generalist" foundation models by bootstrapping the aforementioned robot specialists. To conclude, I'll further discuss the connections between these two types of systems and methods for enabling these systems to execute complex, long-horizon tasks suitable for open-world deployment.

讲者简介：Jianlan Luo is a postdoctoral scholar in the Department of Electrical Engineering and Computer Sciences at UC Berkeley, working with Sergey Levine. Before moving back full-time to academia in 2022, he spent two years as a researcher in the robotics division of Google X. He received his MS/Ph.D. from UC Berkeley in 2020. His research interests lie in the intersection between machine learning, robotics, and controls, with a focus on developing high-performance learning-based robotic systems. His work has won ICRA best paper awards, led to several industrial adoptions, and has been featured in numerous media posts.

王越 浙江大学

报告题目：从视觉反馈控制到 VLA

报告摘要：视觉反馈控制是机器人实际应用中必不可缺的组成部分。然而，当前的视觉反馈控制部署流程复杂，成本较高，导致业务切换缓慢。本报告将从提升视觉反馈控制的通用性出发，尝试克服这些不足，并以此为起点介绍垂直领域的机器人视觉反馈控制的框架，探索视觉-语言-动作（VLA）的必要性。

讲者简介：王越，教授，浙江大学控制科学与工程学院。近五年来以通讯作者发表 Nature Communication、IJRR、TPAMI、TRO、IJCV 等期刊论文 30 多篇，发表 RSS、ICRA、IROS 等领域顶会 30 多篇，论文入选 ESI 高被引，获机器人领域旗舰会议 ICRA 最佳视觉论文，IROS 最佳设计论文提名等论文奖 5 次，入选年度全球前 2%顶尖科学家榜单。担任期刊 IEEE Robotics and Automation Letters 编委、ICRA/IROS 编委、Frontiers in Robotics and AI、Electronics 邀请编委、IROS2025 本地主席、ARTS2025 本地主席等。获中国发明协会创业创新奖二等奖（排名 1）等。主持科技部重点研发计划课题、国家自然科学基金、浙江省自然科学基金重大项目等。

刘希慧 香港大学

报告题目：3D Object Parsing: From Surface Segmentation to Generative Amodal Segmentation

报告摘要：While recent advances in 3D generation have demonstrated impressive progress in generating holistic shapes, modeling from the perspective of parts remains underexplored. Part-level understanding is fundamental to enabling controllable and compositional 3D content creation. In this talk, I will present our recent efforts toward part-centric 3D generation, focusing on two works, SAMPart3D and HoloPart. We investigate how to scale part segmentation to large and diverse 3D datasets without relying on predefined labels, and how to infer complete, amodally segmented parts that remain consistent with global object geometry. These efforts lay the groundwork for enabling more compositional, editable, and generalizable 3D generation systems.

讲者简介：Xihui Liu is an Assistant Professor at the Department of Electrical and Electronic Engineering and Institute of Data Science, The University of Hong Kong. Before joining HKU, she was a postdoc Scholar at UC Berkeley. She obtained her Ph.D. degree from Multimedia Lab (MMLab), the Chinese University of Hong Kong and received her bachelor's degree from Tsinghua University. Her research interests cover computer vision, machine learning, and artificial intelligence, with special emphasis on visual synthesis, generative models, and multimodal AI. She was awarded Adobe Research Fellowship 2020, MIT EECS Rising Stars 2021, and WAIC Rising Stars Award 2022. She serves as area chairs for CVPR 2024, ACM MM 2024, ICLR 2025, CVPR 2025, and NeurIPS 2025.

苏航 清华大学

报告题目：跨越虚实鸿沟：基座模型驱动的具身智能泛化之路

报告摘要：泛化能力不足是具身智能走向现实场景的核心瓶颈。大规模基座模型（Foundation Models）的兴起为突破这一瓶颈提供了关键支撑。本报告围绕具身智能基座模型的构建与泛化应用，从仿真数据、真实机器人数据和视频数据三个维度，系统阐述如何通过跨模态数据融合与动作表征解耦，推动具身智能实现真正的泛化能力。具体而言，我们首先探讨如何基于仿真数据，以 ManiBox 框架通过边界框引导的策略蒸馏技术，有效降低 Sim2Real 泛化差距，并发现了空间泛化所需数据量与空间规模呈现出显著的幂律关系；其次，我们介绍了基于真实机器人数据构建的 RDT 模型，通过提出可物理解释的统一动作空间与大规模跨本体预训练，RDT 显著提高了双臂操作任务的泛化能力。最后，以大规模视频数据为基础，通过扩散模型预训练和掩码动作预测模型构建统一的视频动作基座模型，仅需极少示教数据即可实现跨平台泛化。综上，本报告以基座模型为核心，系统梳理数据融合、表征解耦与泛化迁移的新范式，为具身智能的工业化应用与理论研究提供重要参考。

讲者简介：苏航，清华大学计算机系副研究员，入选国家“万人计划”青年拔尖人才，主要研究鲁棒机器学习和具身决策等相关领域，发表 CCF 推荐 A 类会议和期刊论文 100 余

篇，谷歌学术论文引用 15000 余次，受邀担任人工智能领域顶级期刊 IEEE TPAMI 和 Artificial Intelligence 的编委，IEEE 生成式大模型安全工作组主席，获得吴文俊人工智能自然科学一等奖，ICME 铂金最佳论文、MICCAI 青年学者奖和 AVSS 最佳论文等多个学术奖项，曾率队在 NeurIPS2017 对抗攻防等多个国际学术比赛中获得冠军。现任中国图像图形学会青工委执委、曾任 VALSE 执行 AC 委员会主席，NeurIPS21 的领域主席（Area Chair）、AAAI22 Workshop Co-Chair 等。

Workshop 20: 优秀学生论坛

组织者：韩晓光（香港中文大学（深圳））、李永露（上海交通大学）、于茜（北京航空航天大学）、胡迪（人民大学）

时间：6 月 8 日（周日） 8:30-12:10 地点：四楼大湾区厅 5

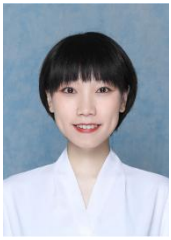
时间	环节	主持人	内容
8:30-8:40	前沿报告	李永露	讲者：陈科研（北京航空航天大学） 题目：遥感图像高效视觉基础模型研究
8:40-8:50			讲者：刘世隆（清华大学） 题目：LLaVA-Plus：学习使用工具创建多模态智能体
8:50-9:00			讲者：卫雅珂（中国人民大学） 题目：平衡多模态学习
9:00-9:10			讲者：黄文柯（武汉大学） 题目：多模态大模型的领域微调
9:10-9:20			讲者：刘欣鹏（上海交通大学） 题目：无中生有的人体运动理解
9:20-9:30			讲者：王健伊（新加坡南洋理工大学） 题目：视频超分大模型
9:30-9:40			讲者：倪旻恒（香港理工大学） 题目：基于多模态认知的视觉推理与生成
9:40-9:50			讲者：窦志扬（香港大学） 题目：From Static 3D Geometry to Dynamic 4D Content: Analysis, Recovery, and Generation
9:50-10:00	中场休息		
10:00-11:00	主题辩论	韩晓光 于茜	辩题：“工程化”论文被顶会录取对视觉领域发展是利大于弊还是弊大于利？
11:00-12:00	导师面对面 & 自由讨论		

组织者：韩晓光 香港中文大学（深圳）

个人简介：韩晓光博士，现任香港中文大学（深圳）理工学院助理教授。他于 2017 年获得香港大学计算机科学专业博士学位。其研究方向包括计算机图形学和三维视觉等，在该方向包括著名国际期刊和会议已发表论文 100 余篇，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。曾担任 CVPR、ICCV、ECCV、NeurIPS 等领域主席，Siggraph Asia 程序委员，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，曾两次入选 CVPR 最佳论文列表，曾入选世界人工智能大会青年优秀论文提名奖。他也积极组织和参与各种学术活动，目前担任 GAMES 秘书长，负责 GAMES 平台的运营，他也曾策划和组织论文背后的故事（PaSS）系列在线科研分享活动。

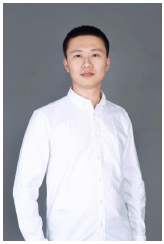
组织者：李永露 上海交通大学

个人简介：李永露博士，上海交大助理教授、上海创智学院全时导师，博导，研究具身智能、物理推理、行为理解，在 TPAMI、CVPR、NeurIPS 等发表研究成果 50 余篇，引用 100+论文 8 篇，ESI 高被引论文 3 篇，一篇独立通讯论文获 ICRA 2025 Best Paper Award on HRI；开源项目获 Github star 1.3 万+；代表工作 HAKE（引用 1.3k+，Github Star 2.18k+，官网全球访问 15.8 万+次）、AlphaPose（引用 630+，Github Star 8.4k+）。任 NeurIPS’24/25 Area Chair，上海交大 ACM 班《计算机视觉》课程教师， VALSE EACC，中国人工智能学会-具身智能专委会副秘书长。主持、参与多项国家级项目，如青基、科技部重点研发计划等。获上海市海外高层次人才、中国人工智能学会吴文俊人工智能科学技术奖-优秀博士学位论文、WAIC 云帆奖-璀璨明星、明日之星、AI100 青年先锋、NeurIPS’20/21 杰出审稿人、百度奖学金、华人 AI 新星百人等。



组织者：于茜 北京航空航天大学

个人简介：于茜，“卓越百人”副教授，博士生导师，入选第九届中国科协青年人才托举工程。2018 年博士毕业于英国伦敦大学玛丽女王学院，之后在美国加州大学伯克利分校从事博士后研究。研究方向是计算机视觉和深度学习，主要聚焦于 AIGC、草图理解与应用、矢量图生成等。目前发表学术论文 40 余篇，Google Scholar 引用 2500 余次。曾荣获 2015 年英国机器视觉大会(BMVC)的最佳论文奖，相关成果受到 BBC 在内的十余家海内外媒体的报道。自 2020 年入职以来，曾/现主持国自然青年基金项目、CCF-百度松果基金项目、北航-华为关键软件项目等，作为项目骨干参与科技部重大项目两项；担任 ACM Multimedia 2024 的社交媒体主席和 CVPR 2024/2025 的领域主席。



组织者：胡迪 人民大学

个人简介：胡迪，中国人民大学高瓴人工智能学院副教授，博导。主要研究方向为 机器多模态感知、交互与学习，以主要作者在 TPAMI/ICML/CVPR/CoRL 等人工智能顶级期刊及会议发表论文 40 余篇。曾入选 CVPR Doctoral Consortium；荣获 2020 中国人工智能学会优博奖；荣获 2022 年度吴文俊人工智能优秀青年奖；入选第七届中国科协青托计划等。担任 AAAI、IJCAI Senior PC 等。

陈科研，北京航空航天大学

报告题目：遥感图像高效视觉基础模型研究

报告摘要：遥感技术的进步显著提升了影像分辨率，但现有模型在跨任务适应性与高分辨率数据处理方面仍存在局限性。本研究针对遥感图像中关键目标占比小且空间分布稀疏的特点，提出动态视觉感知基础模型 DynamicVis。通过构建基于选择性状态空间模型的动态区域感知网络，本方法有效平衡局部细节特征与全局语义关联，同时结合元嵌入表征的多实例学习机制，增强模型在跨任务场景中的知识迁移能力。DynamicVis 在保持较低计算资源消耗的前提下，能够实现了高分辨率遥感图像的高效处理，并在九类下游任务中展现出优秀的泛化性能。

讲者简介：陈科研，北京航空航天大学博士研究生，主要研究方向为遥感图像智能解译。在 Proceedings of the IEEE、IEEE TPAMI、IEEE TGRS、CVPR 等国际权威期刊/会议发表论文 30 余篇，其中一作论文 10 篇，ESI 热点论文 3 篇，ESI 高被引论文 6 篇，3 篇论文入选权威期刊最受欢迎论文榜单前三，谷歌学术引用 5000 余次。主持首批博士生国家自然科学基金，参与国家重点研发计划、联合重点等多项重要课题项目，相关成果在我国遥感卫星地面系统中得到应用。曾获得宝钢奖学金特等奖、国家奖学金，北航和北京市优秀本科毕业论文，北航优秀硕士学位论文等奖励和荣誉。个人网页：chenkeyan.top。

刘世隆 清华大学

报告题目： LLaVA-Plus：学习使用工具创建多模态智能体

报告摘要： LLaVA-Plus 是一款功能强大的多模态智能助手，能够同时处理文字与图像相关的任务。相比于之前的 LLaVA，LLaVA-Plus 拥有更强的能力，不仅可以理解和生成文字及图像内容，还能够根据用户需求灵活调用合适的工具完成任务，包括图像分析、生成、外部知识检索以及多种功能的组合应用。其显著特点在于，图像信息在整个交互过程中始终被深度关联并主动参与，这大幅提升了工具使用的效率和效果，同时也拓展了更多复杂场景下的应用可能性。

讲者简介：刘世隆是清华大学的博士生，由朱军教授指导，也长期在 IDEA 研究院接受张

磊教授指导。他的研究领域包括计算机视觉、多模态学习以及多模态智能体。刘世隆曾在 IDEA 研究院、NVIDIA、微软雷德蒙德研究院和生数科技等机构实习。他是 Grounding DINO 的第一作者，这是 Hugging Face 上下载量最高的零样本目标检测器；同时，他也是 DINO 的共同第一作者，这是第一个在 COCO 数据集上达到 SOTA 的 DETR 类目标检测器。他的研究成果广受认可，迄今为止已获得超过 8,000 次引用和 30,000 个 GitHub 星标。他荣获了多项奖项，包括 2024 WAIC 云帆奖明日之星和 2023 CCF-CV 学术新锐奖等。

卫雅珂，中国人民大学

报告题目：平衡多模态学习

报告摘要：在一般的多模态学习的范式中，不同模态数据通常采用统一的多模态学习目标进行学习和优化。但不同模态数据天然的异质性给多模态联合学习带来了困难。例如，在判别性任务中，对于标签为“drawing picture”的视听样本，相较音频数据，具有更丰富的判别性信息的视觉数据更容易被学习，被模型所偏好，最终使得使得模型对各个模态数据的利用程度存在不平衡，只有个别模态被充分学习，阻碍了多模态学习的潜力。因此，如何平衡并促进对异质多模态数据的充分挖掘与学习在近年来引起了广泛关注。在报告中，讲者将对这一问题相关研究进行总结与梳理，并展望该问题在更广泛背景下的潜在发展。

讲者简介：卫雅珂，中国人民大学高瓴人工智能学院博士生，导师为胡迪老师。在博士期间的主要研究聚焦于多模态学习机制，以第一/学生第一/共一作者身份在中国计算机学会（CCF）推荐的期刊和会议上发表学术论文 8 篇，其中第一作者 4 篇。包括两篇国际顶级期刊 TPAMI 论文和多篇国际顶级会议如 ICML、CVPR、ECCV 论文等。博士在读期间曾获 2024 年度百度奖学金（全球遴选，仅 10 人入选）、博士生国家奖学金等。在研究工作中，于 CVPR2022 发表的 Oral 论文中于国际上首次提出“平衡多模态学习（Balanced Multimodal Learning）”这一研究方向，并以此为主线展开科研探索，于实验观察、算法及理论在内的多个层面上对该问题进行了系统性研究，发表了系列论文。

黄文柯 武汉大学

报告题目：多模态大模型的领域微调

报告摘要：随着人工智能技术的快速发展，多模态大模型在融合和解析不同模态信息方面展现出强大潜力，显著推动了视觉问答、图文检索等多模态理解任务的进步。微调多模态大模型以适配具体应用场景已成为主流解决方案，但在微调过程中，模型常常面临通用任务知识遗忘与多任务冲突等新挑战。针对上述问题，本汇报探讨了通用与专业任务间的平衡策略，进一步提出了面向多任务微调的优化思路，提升多模态大模型的专

业性与泛化性能。

讲者简介：黄文柯，武汉大学计算机学院 2023 级博士生，师从叶茫教授和杜博教授。研究方向为联邦学习、多模态大模型。目前已于 CCF-A 类会议和期刊上发文多篇论文，包含 IEEE TPAMI 2 篇， CVPR Oral 2 篇。申请人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)和首届中国科协青年人才托举工程-博士生专项，获得雷军卓越奖学金。

刘欣鹏 上海交通大学

报 告 题 目： Something out of Nothing for Human Motion Understanding (无中生有的人体运动理解)

报告摘要：人体运动是计算机视觉的重要研究对象之一，在近年取得了大量进展。本次报告将从对运动的人-世界的二元结构出发，并进一步将人体运动拆解为行为意图/语义-人体动力学-人体运动学的层级化系统，简要介绍报告人对语义-动力-运动-世界这一四元结构的探索。在深度学习的背景下，充分、高质量的数据成为探索这一结构的必要。然而，相较于图文数据，尽管纯人体运动数据在量级上日渐增长，但各元素仍难保持完备。就此，报告人以将以“无中生有”：从人体运动数据中挖掘出看似不存在的有用信息为线索，介绍报告人近期在人体运动理解领域的一系列工作。

讲者简介：刘欣鹏是上海交通大学-上海创智学院学院联合培养的三年级博士生，导师是卢策吾老师。他的主要研究兴趣是人体运动理解，行为理解和具身智能。他目前已经在 CVPR, ICLR, AAAI, NeurIPS, ECCV，T-PAMI 等顶级会议/期刊发表了十余篇论文，其中一作/共一 8 篇，Google Scholar 引用破 1000。他曾获杨元庆奖学金，吴文俊荣誉博士等荣誉。

王健伊 新加坡南洋理工大学

报告题目：视频超分大模型

报告摘要： 视频超分旨在将低质量输入转化为高质量视频。其核心意义在于突破硬件限制，提升视觉质量并降低传输成本，在自然视频和 AIGC 视频的处理中都是不可或缺的一环。然而，面对复杂多样的视频内容和质量损失，如何提升视频分辨率的同时生成符合人类感知的高频细节仍然是现有技术面临的重大挑战。本报告围绕“视频超分大模型”主题，介绍我们借助大数据+大模型在视频超分领域的研究进展与实践探索。

讲者简介：王健伊，新加坡南洋理工大学 MMLab@NTU 四年级博士，导师吕健勤教授，在 CVPR、ECCV、AAAI、IJCV 等会议与期刊上发表论文 15 篇，谷歌引用 1800 余次，代表作包括 CLIP-IQA, StableSR 等；AISG PhD Fellowship。主要研究方向为底层视觉与视频生成。

倪旻恒（香港理工大学）

报告题目：基于多模态认知的视觉推理与生成

报告摘要：随着大模型技术的快速发展，多模态场景逐渐成为人工智能领域的前沿方向。然而，当前生成与推理模型仍存在显著割裂，生成模型难以有效利用推理能力，而推理模型则未能充分结合生成知识，限制了其在复杂场景中的任务适应性与动态交互能力。针对这一挑战，本报告聚焦多模态认知框架下的视觉推理与生成方法构建，介绍我们基于多模态认知在视觉理解、编辑合成、以及机器人控制等领域中的相关工作，探讨复杂多模态任务中应用的可能性。

讲者简介：倪旻恒现为香港理工大学（PolyU）三年级博士研究生，师从左旺孟教授与张雷教授。他已在顶级国际会议与期刊上发表了 14 篇论文。其研究兴趣主要集中在多模态生成与理解中的协同机制，涵盖大型语言模型（LLMs）、多模态视觉语言模型（VLMs）以及扩散模型（Diffusion Models）等方向。近期，他亦积极探索人工智能伦理与认知相关课题。

窦志扬 香港大学

报告题目：From Static 3D Geometry to Dynamic 4D Content: Analysis, Recovery, and Generation

报告摘要： 4D 内容的分析、重建和生成，其中 4D 包括三个空间维度(x,y,z)和一个时间维度(t)，如形状和运动。该研究不仅涉及静态物体，还覆盖了随时间变化的动态过程，旨在全面理解空间与时间的变化。这些技术在虚拟现实、具身人工智能和机器人等领域具有广泛应用。该报告将主要汇报以下进展：从图形学管线角度对三维静态物体的表示与生成进行分析、建模，并进一步扩展至面向四维动态内容（如动画）的获取与生成方法。在此基础上，提出融合几何、拓扑与物理先验的高效四维内容恢复与分析框架，旨在显著提升四维内容生成的效率、可用性与质量。

讲者简介：窦志扬是香港大学计算机科学系的研究生硕士，师从王文平教授和 Taku Komura 教授。他的研究兴趣包括计算机图形学、几何处理、角色动画和物理仿真动画。他的研究成果发表在 SIGGRAPH、SIGGRAPH ASIA、EUROGRAPHICS、ACM TOG、TVCG、SGP、CVPR、ICCV、ECCV、ICLR 等国际顶级会议和期刊，曾获 SIGGRAPH 最佳论文奖、CGF 年度高被引论文奖、港大基金会 2023/24 学年优秀博士生奖、Meshy AI Fellowship Finalist 等。他现阶段的研究重点是将几何、拓扑和物理的先验融入到 4D 数据的获取、分析与生成过程之中。

Poster 交流论文一览表

A01	Gang Wu, Junjun Jiang, Yijun Wang, Kui Jiang, Xianming Liu	哈尔滨工业大学	AAAI 2025
	Debiased All-in-one Image Restoration with Task Uncertainty Regularization		
A02	Yuanzhi Wang, Yong Li, Mengyi Liu, Xiaoya Zhang, Xin Liu, Zhen Cui, Antoni B. Chan	南京理工大学	AAAI 2025
	Re-Attentional Controllable Video Diffusion Editing		
A03	Tianyu Huang, Haoze Zhang, Yihan Zeng, Zhilu Zhang, Hui Li, Wangmeng Zuo, Rynson W. H. Lau	哈尔滨工业大学	AAAI 2025
	DreamPhysics: Learning Physics-Based 3D Dynamics with Video Diffusion Priors		
A04	Junyi Li, Zhilu Zhang, Wangmeng Zuo	哈尔滨工业大学	AAAI 2025
	Rethinking Transformer-Based Blind-Spot Network for Self-Supervised Image Denoising		
A05	Zhengpeng Duan, Jiawei Zhang, Siyu Liu, Zheng Lin, Chunle Guo, Dongqing Zou, Sijie Ren, Chongyi Li	南开大学	AAAI 2025
	A Diffusion-Based Framework for Occluded Object Movement		
A06	Junkai Fan, Kun Wang, Zhiqiang Yan, Xiang Chen, Shangbing Gao, Jun Li, Jian Yang	南京理工大学	AAAI 2025
	Depth-Centric Dehazing and Depth-Estimation from Real-World Hazy Driving Video		
A07	Feng Han, Kai Chen, Chao Gong, Zhipeng Wei, Jingjing Chen, Yugang Jiang	复旦大学	AAAI 2025
	DuMo: Dual Encoder Modulation Network for Precise Concept Erasure		
A08	Yabo Zhang, Yuxiang Wei, Xianhui Lin, Zheng Hui, Peiran Ren, Xuansong Xie, Wangmeng Zuo	哈尔滨工业大学	AAAI 2025
	VideoElevator: Elevating Video Generation Quality with Versatile Text-to-Image Diffusion Models		
A09	Tongshun Zhang, Pingping Liu, Ming Zhao, Haotian Lv	吉林大学	ACMMM 2024
	DMFourLLIE: Dual-Stage and Multi-Branch Fourier Network for Low-Light Image Enhancement		
A10	Haijie Yang, Zhenyu Zhang, Hao Tang, Jianjun Qian, Jian Yang	南京理工大学	ACMMM 2024
	ConsistentAvatar：Learning to Diffuse Fully Consistent Talking Head Avatar with Temporal Guidance		
A11	Hongtao Wu, Yijun Yang, Huihui Xu, Weiming Wang, Jinni Zhou, Lei Zhu	香港科技大学（广州）	ACMMM 2024
	RainMamba: Enhanced Locality Learning with State Space Models for Video Deraining		
A12	Weiqi Li, Shijie Zhao, Bin Chen, Xinhua Cheng, Junlin Li, Li Zhang, Jian Zhang	北京大学	ACMMM 2024

	ResVR: Joint Rescaling and Viewport Rendering of Omnidirectional Images		
A13	Zongsheng Yue, Kang Liao, Chen Change Loy	西安交通大学	CVPR 2025
	Arbitrary-steps Image Super-resolution via Diffusion Inversion		
A14	Juncheng Wang, Chao Xu, Cheng Yu, Lei Shang, Zhe Hu, Shujun Wang, Liefeng Bao	香港理工大学	CVPR 2025
	Synchronized Video to Audio Generation via Mel Quantization and Continuum Decomposition		
A15	Duocheng Chen, Shihao Zhou, Jinshan Pan, Jinglei Shi, Lishen Qu, Jufeng Yang	南开大学	CVPR 2025
	A Polarization-aided Transformer for Image Deblurring via Motion Vector Decomposition		
A16	Zhu Liu, Zijun Wang, Jinyuan Liu, Fanqi Meng, Long Ma, Risheng Liu	大连理工大学	CVPR 2025
	DEAL: Data-Efficient Adversarial Learning for High-Quality Infrared Imaging		
A17	Chengyou Jia, Changliang Xia, Zhuohang Dang, Weijia Wu, Hangwei Qian, Minnan Luo	西安交通大学	CVPR 2025
	ChatGen: Automatic Text-to-Image Generation From FreeStyle Chatting		
A18	Junyang Chen, Jinshan Pan, Jiangxin Dong	南京理工大学	CVPR 2025
	FaithDiff: Unleashing Diffusion Priors for Faithful Image Super-resolution		
A19	Chen Zhao, Zhizhiu Chen, Yunzhe Xu, Enxuan Gu, Jian Li, Zili Yi, Qian Wang, Jian Yang, Ying Tai	南京大学	CVPR 2025
	From Zero to Detail: Deconstructing Ultra-High-Definition Image Restoration from Progressive Spectral Perspective		
A20	Lve Fan, Hao Zhang, Qitai Wang, Hongsheng Li, Zhaoxiang Zhang	中科院自动化所	CVPR 2025
	FreeSim: Toward Free-viewpoint Camera Simulation in Driving Scenes		
A21	Shenghai Yuan, Jinfa Huang, Xianyi He, Yunyang Ge, Yujun Shi, Liuhan Chen, Jiebo Luo, Li Yuan	北京大学	CVPR 2025
	Identity-Preserving Text-to-Video Generation by Frequency Decomposition		
A22	Tianhao Qi, Jianlong Yuan, Wanquan Feng, Shancheng Fang, Jiawei Liu, Siyu Zhou, Qian He, Hongtao Xie, Yongdong Zhang	中国科学技术大学 字节跳动	CVPR 2025
	Maskξ^2ξDiT: Dual Mask-based Diffusion Transformer for Multi-Scene Long Video Generation		
A23	Haowen Bai, Jiangshe Zhang, Zixiang Zhao, Yichen Wu, Lilun Deng, Yukun Cui, Tao Feng, Shuang Xu	西安交通大学	CVPR 2025
	Task-driven Image Fusion with Learnable Fusion Loss		
A24	Jiayi Fu, Siyu Liu, Zikun Liu, Chunle Guo, Hyunhee Park, Ruiqi Wu, Guoqing Wang, Chongyi Li	南开大学	CVPR 2025
	Iterative Predictor-Critic Code Decoding for Real-World Image Dehazing		
A25	Haojie Yan, Zhan Lu, Zehao Chen, De Ma, Huajin Tang, Qian Zheng, Gang Pan	浙江大学脑机智	AAAI 2025

		能全国重点实验室 浙江大学计算机学院	
	EvSTVSR: Event Guided Space-Time Video Super-Resolution		
A26	Hao Zhao, Mingjia Li, Qiming Hu, Xiaojie Guo	天津大学	CVPR 2025
	Reversible Decoupling Network for Single Image Reflection Removal		
A27	Jianlong Jin, Chenglong Zhao, Ruixin Zhang, Sheng Shang, Jianqing Xu, Jingyun Zhang, Shaoming Wang, Yang Zhao, Shouhong Ding, Wei Jia, Yunsheng Wu	合肥工业大学	CVPR 2025
	Diff-Palm: Realistic Palmprint Generation with Polynomial Creases and Intra-Class Variation Controllable Diffusion Models		
A28	Zeyu Xiao, Dachun Kai, Yueyi Zhang, Zhengjun Zha, Xiaoyan Sun, Zhiwei Xiong	中国科学技术大学	ECCV 2024
	Event-Adapted Video Super-Resolution		
A29	Jinliang Xiao, Tingzhu Huang, Liangjian Deng, Guang Lin, Zihan Cao, Chao Li, Qibin Zhao	电子科技大学	CVPR 2025
	Hyperspectral Pansharpening via Diffusion Models with Iteratively Zero-Shot Guidance		
A30	Peishan Cong, Ziyi Wang, Yuexin Ma, Xiangyu Yue	上海科技大学，香港中文大学	CVPR 2025
	SemGeoMo: Dynamic Contextual Human Motion Generation with Semantic and Geometric Guidance		
A31	Shuoyan Wei, Feng Li, Shengeng Tang, Yao Zhao, Huihui Bai	北京交通大学	CVPR 2025
	EvEnhancer: Empowering Effectiveness, Efficiency and Generalizability for Continuous Space-Time Video Super-Resolution with Events		
A32	Tianyi Zhu, Dongwei Ren, Qilong Wang, Xiaohe Wu, Wangmeng Zuo	哈尔滨工业大学	CVPR 2025
	Generative Inbetweening through Frame-wise Conditions-Driven Video Generation		
A33	Yuheng Xu, Shijie Yang, Xin Liu, Jie Liu, Jie Tang, Gangshan Wu	南京大学	CVPR 2025
	AutoLUT: LUT-Based Image Super-Resolution with Automatic Sampling and Adaptive Residual Learning		
A34	Hongbin Lin, Zilu Guo, Yifan Zhang, Shuaicheng Niu, Yafei Li, Ruimao Zhang, Shuguang Cui, Zhen Li	香港中文大学（深圳）	CVPR 2025
	DriveGEN: Generalized and Robust 3D Detection in Driving via Controllable Text-to-Image Diffusion Generation		
A35	Kendong Liu, Zhiyu Zhu, Hui Liu, Junhui Hou	香港城市大学	CVPR 2025
	Acc3D: Accelerating Single Image to 3D Diffusion Models via Edge Consistency Guided Score Distillation		
A36	Haoyue Liu, Jinghan Xu, Yi Chang, Hanyu Zhou, Haozhi Zhao, Lin Wang,	华中科技大学	CVPR 2025

	Luxin Yan		
	TimeTracker: Event-based Continuous Point Tracking for Video Frame Interpolation with Non-linear Motion		
A37	Hesong Li, Ziqi Wu, Ruiwen Shao, Tao Zhang, Ying Fu	北京理工大学	CVPR 2025
	Noise Calibration and Spatial-Frequency Interactive Network for STEM Image Enhancement		
A38	Ziteng Cui, Xuangeng Chu, Tatsuya Harada	东京大学	CVPR 2025
	Luminance-GS: Adapting 3D Gaussian Splatting to Challenging Lighting Conditions with View-Adaptive Curve Adjustment		
A39	Yunlong Lin, Zixu Lin, Haoyu Chen, Panwang Pan, Chenxin Li, Sixiang Chen, Yeying Jin, Wenbo Li, Xinghao Ding	厦门大学	CVPR 2025
	JarvisIR: Elevating Autonomous Driving Perception with Intelligent Image Restoration		
A40	Yujun Liu, Ruisheng Wang, Shangfeng Huang, Guorong Cai	深圳大学	CVPR 2025
	EdgeDiff: Edge-aware Diffusion Network for Building Reconstruction from Point Clouds		
A41	Yuanbo Wang, Zhaoxuan Zhang, Jiajin Qiu, Dilong Sun, Zhengyu Meng, Xiaopeng Wei, Xin Yang	大连理工大学	CVPR 2025
	Touch2Shape: Touch-Conditioned 3D Diffusion for Shape Exploration and Reconstruction		
A42	Mingde Yao, Menglu Wang, Jingwen Tan, Lingen Li, Tianfan Xue, Jinwei Gu	香港中文大学	CVPR 2025
	PolarFree: Polarization-based Reflection-free Imaging		
A43	Shunlin Lu, Jingbo Wang, Zeyu Lu, Linghao Chen, Wenxun Dai, Junting Dong, Zhiyang Dou, Bo Dai, Ruimao Zhang	中山大学 香港中文大学(深圳) 上海人工智能实验室	CVPR 2025
	ScaMo: Exploring the Scaling Law in Autoregressive Motion Generation Model		
A44	Feiyang Shen, Hongping Gan	西北工业大学	CVPR 2025
	HUNet: Homotopy Unfolding Network for Image Compressive Sensing		
A45	Jiaxiu Jiang, Yabo Zhang, Kailai Feng, Xiaohe Wu, Wenbo Li, Renjing Pei, Fan Li, Wangmeng Zuo	哈尔滨工业大学	CVPR 2025
	mc^2: multi-concept guidance for customized multi-concept generation		
A46	Xingyuan Li, Zirui Wang, Yang Zou, Zhixin Chen, Jun Ma, Zhiying Jiang, Long Ma, Jinyuan Liu	大连理工大学	CVPR 2025
	DifIISR: Diffusion Model with Gradient Guidance for Infrared Image Super-Resolution		
A47	Jinyuan Liu, Bowei Zhang, Qingyun Mei, Xingyuan Li, Yang Zou, Zhiying Jiang, Long Ma, Risheng Liu, Xin Fan	大连理工大学	CVPR 2025
	DCEvo: Discriminative Cross-Dimensional Evolutionary Learning for Infrared and Visible Image Fusion		
A48	Chuanbo Tang, Zhuoyuan Li, Yifan Bian,	中国科学技术大	CVPR 2025

	Li Li, Dong Liu	学	
	Neural Video Compression with Context Modulation		
A49	Siyuan Li, Luyuan Zhang, Zedong Wang, Juanxi Tian, Cheng Tan, Zicheng Liu, Chang Yu, Qingsong Xie, Haonan Lu, Haoqian Wang, Zhen Lei	OPPO	CVPR 2025
	MergeVQ: A Unified Framework for Visual Generation and Representation with Disentangled Token Merging and Quantization		
A50	Zilong Chen, Yikai Wang, Wenqiang Sun, Feng Wang, Yiwen Chen, Huaping Liu	清华大学	CVPR 2025
	MeshGen: Generating PBR Textured Mesh with Render-Enhanced Auto-Encoder and Generative Data Augmentation		
E01	Xihua Wang, Ruihua Song, Chongxuan Li, Xin Cheng, Boyuan Li, Yihan Wu, Yuyue Wang, Hongteng Xu, Yunfeng Wang	中国人民大学高瓴人工智能学院	CVPR 2025
	Animate and Sound an Image		
E02	Lang Nie, Chunyu Lin, Kang Liao, Shuaicheng Liu, Yun Zhang, Yao Zhao	北京交通大学	ECCV 2024
	Eliminating Warping Shakes for Unsupervised Online Video Stitching		
E03	Runyi Li, Xuhan Sheng, Weiqi Li, Jian Zhang	北京大学	ECCV 2024
	OmniSSR: Zero-shot Omnidirectional Image Super-Resolution using Stable Diffusion Model		
E04	Zongliang Wu, Ruiying Lu, Ying Fu, Xin Yuan	浙江大学/西湖大学	ECCV 2024
	Latent Diffusion Prior Enhanced Deep Unfolding for Snapshot Spectral Compressive Imaging		
E05	Yixiang Qiu, Hao Fang, Hongyao Yu, Bin Chen, Meikang Qiu, Shu-Tao Xia	清华大学	ECCV 2024
	A Closer Look at GAN Priors: Exploiting Intermediate Features for Enhanced Model Inversion Attacks		
E06	Renlong Wu, Zhilu Zhang, Shuohao Zhang, Longfei Gou, Haobin Chen, Lei Zhang, Hao Chen, Wangmeng Zuo	哈尔滨工业大学	ECCV 2024
	Self-Supervised Video Desmoking for Laparascopic Surgery		
E07	Lan Yao, Chaofeng Chen, Xiaoming Li, Zifei Yan, Wangmeng Zuo	哈尔滨工业大学	ECCV 2024
	Combining Generative and Geometry Priors for Wide Angle Portraits Correction		
E08	Hai Jiang, Ao Luo, Xiaohong Liu, Songchen Han, Shuaicheng Liu	四川大学	ECCV 2024
	LightenDiffusion: Unsupervised Low-Light Image Enhancement with Latent-Retinex Diffusion Models		
E09	Hang Yao, Ming Liu, Zhicun Yin, Zifei Yan, Xiaopeng Hong, Wangmeng Zuo	哈尔滨工业大学	ECCV 2024
	GLAD: Towards Better Reconstruction with Global and Local Adaptive Diffusion Models for Unsupervised Anomaly Detection		
E10	Shunkun Liang, Banglei Guan, Zhenbao Yu, Pengju Sun, Yang Shang	国防科技大学	ECCV 2024
	Camera Calibration Using a Collimator System		

E11	Yang Liu, He Guan, Chuanchen Luo, Lue Fan, Junran Peng, Zhaoxiang Zhang	中国科学院自动化研究所	ECCV 2024
	CityGaussian: Real-time High-quality Large-Scale Scene Rendering with Gaussians		
E12	Jian Ma, Chen Chen, Qingsong Xie, Haonan Lu	OPPO	ECCV 2024
	PEA-Diffusion: Parameter-Efficient Adapter with Knowledge Distillation in non-English Text-to-Image Generation		
E13	Tao Liu, Kai Wang, Senmao Li, Joost van de Weijer, Fahad Shahbaz Khan, Shiqi Yang, Yaxing Wang, Jian Yang, Ming-Ming Cheng	南开大学	ICLR 2025
	One-Prompt-One-Story: Free-Lunch Consistent Text-to-Image Generation Using a Single Prompt		
E14	Zhilu Zhang, Shuohao Zhang, Renlong Wu, Zifei Yan, Wangmeng Zuo	哈尔滨工业大学	ICLR 2025
	Exposure Bracketing Is All You Need For A High-Quality Image		
E15	Zheng Zhong, Xiao Dong, Haoxiang Li, Shiyue Zhang, Wenqing Zhang, Xujie Zhang, Hanqing Zhao, Dongmei Jiang, Xiaodan Liang	中山大学	ICLR 2025
	CatVTON: Concatenation Is All You Need for Virtual Try-On with Diffusion Models		
E16	Shengyuan Zhang, Ling Yang, Zejian Li, An Zhao, Chenye Meng, Changyuan Yang, Guang Yang, Zhiyuan Yang, Lingyun Sun	浙江大学	ICLR 2025
	Distribution Backtracking Builds A Faster Convergence Trajectory for Diffusion Distillation		
E17	Yizhuo Lu, Changde Du, Chong Wang, Xuanliu Zhu, Liuyun Jiang, Xujin Li, Huiguang He	中国科学院自动化研究所	ICLR 2025
	Animate Your Thoughts: Reconstruction of Dynamic Natural Vision from Human Brain Activity		
E18	Hongyin Zhang, Pengxiang Ding, Shangke Lv, Ying Peng, Donglin Wang	浙江大学	ICLR 2025
	GEVRM: GOAL-EXPRESSIVE VIDEO GENERATION MODEL FOR ROBUST VISUAL MANIPULATION		
E19	Chengyu Fang, Chunming He, Fengyang Xiao, Yulun Zhang, Longxiang Tang, Yuelin Zhang, Kai Li, Xiu Li	清华大学	NeurIPS 2024
	Real-world Image Dehazing with Coherence-based Pseudo Labeling and Cooperative Unfolding Network		
E20	Chunming He, Chengyu Fang, Yulun Zhang, Longxiang Tang, Jinfa Huang, Kai Li, Zhenhua Guo, Xiu Li, Sina Farsiu	杜克大学	ICLR 2025
	Reti-Diff: Illumination Degradation Image Restoration with Retinex-based Latent Diffusion Model		
E21	Long Peng, Wenbo Li, Renjing Pei, Jingjing Ren, Jiaqi Xu, Yang Wang, Yang	中国科学技术大学	ICLR 2025

	Cao, Zheng-Jun Zha		
	Towards Realistic Data Generation for Real-World Super-Resolution		
E22	Xiang Liu, Bin Chen, Zimo Liu, Yaowei Wang, Shutao Xia	清华大学深圳国际研究生院	ICLR 2025
	An Exploration with Entropy Constrained 3D Gaussians for 2D Video Compression		
E23	Shilin Lu, Zihan Zhou, Jiayou Lu, Yuanzhi Zhu, Adams Wai-Kin Kong	新加坡南洋理工大学	ICLR 2025
	Robust Watermarking Using Generative Priors Against Image Editing: From Benchmarking to Advances		
E24	Hanqiao Ye, Yuzhou Liu, Yangdong Liu, Shuhan Shen	中国科学院大学人工智能学院 中国科学院自动化研究所	ICLR 2025
	NeuralPlane: Structured 3D Reconstruction in Planar Primitives with Neural Fields		
E25	Xiaoyu Zhou, Xingjian Ran, Yajiao Xiong, Jinlin He, Yongtao Wang, Deqing Sun, Ming-Hsuan Yang	北京大学	ICML 2024
	GALA3D: Towards Text-to-3D Complex Scene Generation via Layout-guided Generative Gaussian Splatting		
E26	Xiaole Tang, Xin Hu, Xiang Gu, Jian Sun	西安交通大学	ICML 2024
	Residual-Conditioned Optimal Transport: Towards Structure-Preserving Unpaired and Paired Image Restoration		
E27	Zhu Xu, Qingchao Chen, Yuxin Peng, Yang Liu	北京大学	ICML 2024
	Semantic-Aware Human Object Interaction Image Generation		
E28	Haoyu Deng, Zijing Xu, Yule Duan, Xiao Wu, Wenjie Shu, Liangjian Deng	电子科技大学	ICML 2024
	Exploring the low-pass filtering behavior in image super-resolution		
E29	Chunyang Cheng, Tianyang Xu, Xiaojun Wu, Hui Li, Xi Li, Josef Kittler	江南大学	IJCV 2025
	FusionBooster: A Unified Image Fusion Boosting Paradigm		
E30	Zengxi Zhang, Zhiying Jiang, Long Ma, Jinyuan Liu, Xin Fan, Risheng Liu	大连理工大学	IJCV 2025
	HUPE: Heuristic Underwater Perceptual Enhancement with Semantic Collaborative Learning		
E31	Taihang Hu, Linxuan Li, Joost van de Weijer, Hongcheng Gao, Fahad Shahbaz Khan, Jian Yang, Ming-Ming Cheng, Kai Wang, Yaxing Wang	南开大学	NeurIPS 2024
	Token Merging for Training-Free Semantic Binding in Text-to-Image Synthesis		
E32	Haiyu Zhao, Lei Tian, Xinyan Xiao, Peng Hu, Yuanbiao Gou, Xi Peng	四川大学	NeurIPS 2024
	AverNet: All-in-one Video Restoration for Time-varying Unknown Degradations		
E33	Min Zhao, Guande He, Yixiao Chen, Hongzhou Zhu, Chongxuan, Jun Zhu	清华大学	ICML 2025
	RIFLEx: A Free Lunch for Length Extrapolation in Video Diffusion Transformers		
E34	Jiahao Wang, Caixia Yan, Haonan Lin,	西安交通大学	NeurIPS

	Weizhan Zhang, Mengmeng Wang, Tieliang Gong, Guang Dai, Hao Sun		2024
	OneActor: Consistent Subject Generation via Cluster-Conditioned Guidance		
E35	Qiming Hu, Hainuo Wang, Xiaojie Guo	天津大学	NeurIPS 2024
	Single Image Reflection Separation via Interactive Dual-Stream Transformers		
E36	Huayang Huang, Yu Wu, Qian Wang	武汉大学	NeurIPS 2024
	ROBIN: Robust and Invisible Watermarks for Diffusion Models with Adversarial Optimization		
E37	Haoye Dong, Aviral Chharia, Wenbo Gou, Francisco Vicente Carrasco, Fernando De la Torre	卡内基梅隆大学	NeurIPS 2024
	Hamba: Single-view 3D Hand Reconstruction with Graph-guided Bi-Scanning Mamba		
E38	Qiankun Gao, Jiarui Meng, Chengxiang Wen, Jie Chen, Jian Zhang	北京大学	NeurIPS 2024
	HiCoM: Hierarchical Coherent Motion for Streamable Dynamic Scene with 3D Gaussian Splatting		
E39	Xinyang Li, Zhangyu Lai, Linning Xu, Yansong Qu, Liujuan Cao, Shengchuan Zhang, Bo Dai, Rongrong Ji	厦门大学	NeurIPS 2024
	Director3D: Real-world Camera Trajectory and 3D Scene Generation from Text		
E40	Tianchen Zhao, Tongcheng Fang, Haofeng Huang, Enshu Liu, Rui Wan, Xianying Chen, Shiyao Li, Zinan Lin, Guohao Dai, Shengen Yan, Huazhong Yang, Xuefei Ning, Yu Wang	清华大学	ICLR 2025
	ViDiT-Q: Efficient and Accurate Quantization of Diffusion Transformers for Image and Video Generation		
E41	Senmao Li, Taihang Hu, Joost van de Weijer, Fahad Shahbaz Khan, Tao Liu, Linxuan Li, Shiqi Yang, Yaxing Wang, Ming-Ming Cheng, Jian Yang	南开大学	NeurIPS 2024
	Faster Diffusion: Rethinking the Role of the Encoder for Diffusion Model Inference		
E42	Jiayi Chen*, Yubin Ke*, Lin Peng, He Wang	北京大学	RSS 2025
	Dexonomy: Synthesizing All Dexterous Grasp Types in a Grasp Taxonomy		
E43	Zhiying Jiang, Zengxi Zhang, Jinyuan Liu, Xin Fan, Risheng Liu	大连海事大学	TIP 2024
	Multispectral Image Stitching via Global-Aware Quadrature Pyramid Regression		
E44	Yao Jiang, Xin Li, Keren Fu, Qijun Zhao	四川大学	TIP 2025
	Transformer-based Light Field Salient Object Detection and Its Application to Autofocus		
E45	Xiang Chen, Jinshan Pan, Jiangxin Dong, Jinhui Tang	南京理工大学	TPAMI 2025
	Towards Unified Deep Image Deraining: A Survey and A New Benchmark		
E46	Xingming Long, Jie Zhang, Shiguang Shan	中国科学院计算技术研究所	TPAMI 2024

	Generalized Face Liveness Detection via De-fake Face Generator		
E47	Jiahong Fu, Qi Xie, Deyu Meng, Zongben Xu	西安交通大学	TPAMI 2024
	Rotation Equivariant Proximal Operator for Deep Unfolding Methods in Image Restoration		
E48	Ziyuan Luo, Anderson Rocha, Boxin Shi, Qing Guo, Haoliang Li, Renjie Wan	香港浸会大学	TPAMI 2025
	The NeRF Signature: Codebook-Aided Watermarking for Neural Radiance Fields		
E49	Tong Li, Hansen Feng, Lizhi Wang, Lin Zhu, Zhiwei Xiong, Hua Huang	北京理工大学	TPAMI 2024
	Stimulating Diffusion Model for Image Denoising via Adaptive Embedding and Ensembling		
E50	Zeke Xia, Ming Hu, Dengke Yan, Ruixuan Liu, Anran Li, Xiaofei Li, Mingsong Chen	华东师范大学	AAAI 2025
	MultiSFL: Towards Accurate Split Federated Learning via Multi-Model Aggregation and Knowledge Replay		
B01	Xiaochuan Liu, Xin Cheng, Yuchong Sun, Xiaoxue Wu, Ruihua Song, Hao Sun, Denghao Zhang	中国人民大学高瓴人工智能学院	AAAI 2025
	EyEar: Learning Audio Synchronized Human Gaze Trajectory Based on Physics-Informed Dynamics		
B02	Yihan Wu, Yichen Lu, Yifan Peng, Xihua Wang, Ruihua Song, Shinji Watanabe	中国人民大学	AAAI 2025
	Enhancing Audiovisual Speech Recognition Through Bifocal Preference Optimization		
B03	Yihao Huang, Le Liang, Tianlin Li, Xiaojun Jia, Run Wang, Weikai Miao, Geguang Pu, Yang Liu	新加坡南洋理工大学	AAAI 2025
	Perception-guided Jailbreak against Text-to-Image Models		
B04	Zimeng Wu, Jiaxin Chen, Yunhong Wang	北京航空航天大学 计算机学院	AAAI 2025
	Unified Knowledge Maintenance Pruning and Progressive Recovery with Weight Recalling for Large Vision-Language Models		
B05	Haotian Peng, Jiawei Liu, Jinsong Du, Jie Gao, Wei Wang	中国科学院沈阳自动化研究所	AAAI 2025
	BearLLM: A Prior Knowledge-Enhanced Bearing Health Management Framework with Unified Vibration Signal Representation		
B06	Chao Pang, Xingxing Weng, Jiang Wu, Jiayu Li, Yi Liu, Jiaxing Sun, Weijia Li, Shuai Wang, Litong Feng, Gui-Song Xia, Conghui He	武汉大学	AAAI 2025
	VHM: Versatile and Honest Vision Language Model for Remote Sensing Image Analysis		
B07	Jiandong Jin, Xiao Wang, Qian Zhu, Haiyang Wang, Chenglong Li	安徽大学	AAAI 2025
	Pedestrian Attribute Recognition: A New Benchmark Dataset and A Large Language Model Augmented Framework		
B08	Baichuan Zhou, Haote Yang, Dairong Chen, Junyan Ye, Tianyi Bai, Jinhua Yu,	上海人工智能实验室	AAAI 2025

	Songyang Zhang, Dahua Lin, Conghui He, Weijia Li		
	UrBench: A Comprehensive Benchmark for Evaluating Large Multimodal Models in Multi-View Urban Scenarios		
B09	Kaifan Zhang, Lihuo He, Xin Jiang, Wen Lu, Di Wang, Xinbo Gao	西安电子科技大学	AAAI 2025
	CognitionCapturer: Decoding Visual Stimuli from Human EEG Signals with Multimodal Information		
B10	Fan Liu, Wenwen Cai, Jian Huo, Chuanyi Zhang, Delong Chen, Jun Zhou	河海大学	AAAI 2025
	Making Large Vision Language Models to be Good Few-shot Learners		
B11	Jingyu Liu, Minquan Wang, Ye Ma, Bo Wang, Aozhu Chen, Quan Chen, Peng Jiang, Xirong Li	中国人民大学	AAAI 2025
	D&M: Enriching E-commerce Videos with Sound Effects by Key Moment Detection and SFX Matching		
B12	Jingjing Xie, Yuxin Zhang, Jun Peng, Zhaohong Huang, Liujuan Cao	厦门大学媒体分析与计算实验室	AAAI 2025
	TextRefiner: Internal Visual Feature as Efficient Refiner for Vision-Language Models Prompt Tuning		
B13	Hangzhou He, Lei Zhu, Xinliang Zhang, Shuang Zeng, Qian Chen, Yanye Lu	北京大学	AAAI 2025
	V2C-CBM: Building Concept Bottlenecks with Vision-to-Concept Tokenizer		
B14	Yang Liu, Mengyuan Liu, Shudong Huang, Jiancheng Lv	四川大学	AAAI 2025
	Asymmetric Visual Semantic Embedding Framework for Efficient Vision-Language Alignment		
B15	Yue Ma, Hongyu Liu, Hongfa Wang, Heng Pan, Yingqing He, Junkun Yuan, Ailing Zeng, Chengfei Cai, Xiangyang Shen, Wei Liu, Qifeng Chen	香港科技大学 腾讯混元	ACM Siggraph 2024
	Follow Your Emoji: Fine-Controllable and Expressive Freestyle Portrait Animation		
B16	Ruilin Yao, Shengwu Xiong, Yichen Zhao, Yi Rong	武汉理工大学	ACMMM 2024
	Visual Grounding with Multi-modal Conditional Adaptation		
B17	Yansong Qu, Shaohui Dai, Xinyang Li, Jianghang Lin, Liujuan Cao, Shengchuan Zhang, Rongrong Ji	厦门大学	ACMMM 2024
	GOI: Find 3D Gaussians of Interest with an Optimizable Open-vocabulary Semantic-space Hyperplane		
B18	Yipo Huang, Xiangfei Sheng, Zhichao Yang, Quan Yuan, Zhichao Duan, Pengfei Chen, Leida Li, Weisi Lin, Guangming Shi	西安电子科技大学	ACMMM 2024
	AesExpert: Towards multi-modality foundation model for image aesthetics perception		
B19	Zhixuan Liang, Yao Mu, Yixiao Wang, Tianxing Chen, Wenqi Shao, Wei Zhan, Masayoshi Tomizuka, Ping Luo, Mingyu	香港大学	CVPR 2025

	Ding		
	DexHandDiff: Interaction-aware Diffusion Planning for Adaptive Dexterous Manipulation		
B20	Fengxiang Wang, Hongzhen Wang, Zonghao Guo, Di Wang, Yulin Wang, Mingshuo Chen, Qiang Ma, Long Lan, Wenjing Yang, Jing Zhang, Zhiyuan Liu, Maosong Sun	国防科技大学	CVPR 2025
	XLRS-Bench: Could Your Multimodal LLMs Understand Extremely Large Ultra-High-Resolution Remote Sensing Imagery?		
B21	Zelin Peng, Zhengqin Xu, Zhilin Zeng, Zhangsong Wen, Yu Huang, Menglin Yang, Feilong Tang, Wei Shen	上海交通大学	CVPR 2025
	Understanding Fine-tuning CLIP for Open-vocabulary Semantic Segmentation in Hyperbolic Space		
B22	Hairui Ren, Fan Tang, He Zhao, Zixuan Wang, Dandan Guo, Yi Chang	吉林大学	CVPR 2025
	Beyond Words: Augmenting Discriminative Richness via Diffusions in Unsupervised Prompt Learning		
B23	Wenlong Yu, Qilong Wang, Chuang Liu, Dong Li, Qinghua Hu	天津大学	CVPR 2025
	CoE: Chain-of-Explanation via Automatic Visual Concept Circuit Description and Polysemanticity Quantification		
B24	Gen Luo, Xue Yang, Wenhan Dou, Zhaokai Wang, Jiawen Liu, Jifeng Dai, Yu Qiao, Xizhou Zhu	上海人工智能实验室	CVPR 2025
	Mono-internvl: Pushing the boundaries of monolithic multimodal large language models with endogenous visual pre-training		
B25	Xin Wang, Kai Chen, Jiaming Zhang, Jingjing Chen, Xingjun Ma	复旦大学	CVPR 2025
	TAPT: Test-Time Adversarial Prompt Tuning for Robust Inference in Vision-Language Models		
B26	Yikun Liu, Pingan Chen, Jiayin Cai, Xiaolong Jiang, Yao Hu, Jiangchao Yao, Yanfeng Wang, Weidi Xie	上海交通大学	CVPR 2025
	LamRA: Large Multimodal Model as Your Advanced Retrieval Assistant		
B27	Kang Liu, Zhuoqi Ma, Xiaolu Kang, Yunan Li, Kun Xie, Zhicheng Jiao, Qiguang Miao	西安电子科技大学	CVPR 2025
	Enhanced Contrastive Learning with Multi-view Longitudinal Data for Chest X-ray Report Generation		
B28	Yanhao Wu, Haoyang Zhang, Tianwei Lin, Lichao Huang, Shujie Luo, Rui Wu, Congpei Qiu, Wei Ke, Tong Zhang	西安交通大学	CVPR 2025
	Generating Multimodal Driving Scenes via Next-Scene Prediction		
B29	Lianyu Wang, Meng Wang, Huazhu Fu, Daoqiang Zhang	南京航空航天大学	CVPR 2025
	Vision-Language Model IP Protection via Prompt-based Learning		
B30	Yuejiao Su, yi wang, qiongyang hu, chuang yang, lap-pui chau	香港理工大学	CVPR 2025

	ANNEXE: Unified Analyzing, Answering, and Pixel Grounding for Egocentric Interaction		
B31	Hongyu Li, Jinyu Chen, Ziyu Wei, Shaofei Huang, Tianrui Hui, Jialin Gao, Xiaoming Wei, Si Liu	北京航空航天大学	CVPR 2025
	LLaVA-ST: A Multimodal Large Language Model for Fine-Grained Spatial-Temporal Understanding		
B32	Shuo Li, Fang Liu, Zehua Hao, Xinyi Wang, Lingling Li, Xu Liu, Puhua Chen, Wenping Ma	西安电子科技大学	CVPR 2025
	Logits DeConfusion with CLIP for Few-Shot Learning		
B33	Jian Liang, Wenke Huang, Guancheng Wan, Qu Yang, Mang Ye	武汉大学	CVPR 2025
	LoRASculpt: Sculpting LoRA for Harmonizing General and Specialized Knowledge in Multimodal Large Language Models		
B34	Yuanmin Tang, Jue Zhang, Xiaoting Qin, Jing Yu, Gang Xiong, Gaopeng Gou, Qingwei Lin, Saravan Rajmohan, Dongmei Zhang, Qi Wu	中国科学院大学网络空间安全学院 中国科学院信息工程研究所	CVPR 2025
	Reason-before-Retrieve: One-Stage Reflective Chain-of-Thoughts for Training-Free Zero-Shot Composed Image Retrieval		
B35	Luyao Tang, Yuxuan Yuan, Chaoqi Chen, Zeyu Zhang, Yue Huang, Kun Zhang	厦门大学	CVPR 2025
	OCRT: Boosting Foundation Models in the Open World with Object-Concept-Relation Triad		
B36	Shaoyu liu, Jianing Li, Guanghui Zhao, Yunjian Zhang, Xin Meng, Fei Richard Yu, Xiangyang Ji, and Ming Li	光明实验室	CVPR 2025
	EventGPT: Event Stream Understanding With Multimodal Large Language Models		
B37	Jinpeng Wang, Tianci Luo, Yaohua Zha, Yan Feng, Ruisheng Luo, Bin Chen, Tao Dai, Long Chen, Yaowei Wang, Shutao Xia	清华大学深圳国际研究生院	CVPR 2025
	Embracing Collaboration Over Competition: Condensing Multiple Prompts for Visual In-Context Learning		
B38	Enshen Zhou, Qi Su, Cheng Chi, Zhizheng Zhang, Zhongyuan Wang, Tiejun Huang, Lv Sheng, He Wang	北京航空航天大学	CVPR 2025
	Code-as-Monitor: Constraint-aware Visual Programming for Reactive and Proactive Robotic Failure Detection		
B39	Shijia Zhao, Qiming Xia, Xucheng Guo, Pufan Zou, Maoji Zheng, Hai Wu, Chenglu Wen, Cheng Wang	厦门大学	CVPR 2025
	SP3D: Boosting Sparsely-Supervised 3D Object Detection via Accurate Cross-Modal Semantic Prompts		
B40	Geng Li, Jinglin Xu, Yunzhen Zhao, Yuxin Peng	北京大学	CVPR 2025
	DyFo: A Training-Free Dynamic Focus Visual Search for Enhancing LMMs in Fine-Grained Visual Understanding		

B41	Kai Chen, Yunhao Gou, Runhui Huang, Zhili Liu, Daxin Tan, Jing Xu, Chunwei Wang, Yi Zhu, Yihan Zeng, Kuo Yang, Dingdong Wang, Kun Xiang, Haoyuan Li, Haoli Bai, Jianhua Han, Xiaohui Li, Weike Jin, Nian Xie, Yu Zhang, James T. Kwok, Hengshuang Zhao, Xiaodan Liang, Dit-Yan Yeung, Xiao Chen, Zhenguo Li, Wei Zhang, Qun Liu, Jun Yao, Lanqing Hong, Lu Hou, Hang Xu	香港科技大学	CVPR 2025
	EMOVA: Empowering Language Models to See, Hear and Speak with Vivid Emotions		
B42	Fan Yang, Ru Zhen, Jianing Wang, Yanhao Zhang, Haoxiang Chen, Haonan Lu, Sicheng Zhao, Guiguang Ding	OPPO AI Center	CVPR 2025
	HEIE: MLLM-Based Hierarchical Explainable AIGC Image Implausibility Evaluator		
B43	Zhaochen Liu, Limeng Qiao, Xiangxiang Chu, Lin Ma, Tingting Jiang	北京大学	CVPR 2025
	Towards Efficient Foundation Model for Zero-shot Amodal Segmentation		
B44	Liting Lin, Heng Fan, Zhipeng Zhang, Yaowei Wang, Yong Xu, Haibin Ling	鹏城实验室	ECCV 2024
	Tracking Meets LoRA: Faster Training, Larger Model, Stronger Performance		
B45	Chao Gong, Kai Chen, Zhipeng Wei, Jingjing Chen, Yugang Jiang	复旦大学	ECCV 2024
	Reliable and Efficient Concept Erasure of Text-to-Image Diffusion Models		
B46	Zeyu Liu, Weicong Liang, Zhanhao Liang, Chong Luo, Ji Li, Gao Huang, Yuhui Yuan	清华大学	ECCV 2024
	Glyph-ByT5: A Customized Text Encoder for Accurate Visual Text Rendering		
B47	Ting Pan, Lulu Tang, Xinlong Wang, Shiguang Shan	北京智源人工智能研究院	ECCV 2024
	Tokenize Anything via Prompting		
B48	Linlan Huang, Xusheng Cao, Haori Lu, Xialei Liu	南开大学	ECCV 2024
	Class-Incremental Learning with CLIP: Adaptive Representation Adjustment and Parameter Fusion		
B49	Yunfei Liu, Lei Zhu, Lijian Lin, Ye Zhu, Ailing Zhang, Yu Li	粤港澳大湾区数字经济研究院（IDEA 研究院）	ICLR 2025
	TEASER: Token-EnAanced Spacial Modeling for Expression Reconstruction		
B50	Zhi Gao, Bofei Zhang, Pengxiang Li, Xiaojian Ma, Yue Fan, Tao Yuan, Yuwei Wu, Yunde Jia, Songchun Zhu, Qing Li	北京大学	ICLR 2025
	Multi-modal Agent Tuning: Building a VLM-Driven Agent for Efficient Tool Usage		
F01	Hong Li, Nanxi Li, Yuanjie Chen, Jianbin Zhu, Qinlu Guo, Cewu Lu, Yonglu Li	上海交通大学	ICLR 2025
	The Labyrinth of Links: Navigating the Associative Maze of Multi-modal LLMs		
F02	Shufan Shen, Zhaobo Qi, Junshu Sun,	中国科学院计算	ICLR 2025

	Qingming Huang, Qi Tian, Shuhui Wang	技术研究所	
	ENHANCING PRE-TRAINED REPRESENTATION CLASSIFIABILITY CAN BOOST ITS INTERPRETABILITY		
F03	Yifeng Xu, Zhenliang He, Shiguang Shan, Xilin Chen	中国科学院计算技术研究所	ICLR 2025
	CtrLoRA: An Extensible and Efficient Framework for Controllable Image Generation		
F04	Jie Zhang, Zhongqi Wang, Mengqi Lei, Zheng Yuan, Bei Yan, Shiguang Shan, Xilin Chen	中科院计算所	ICLR 2025
	DYSCA: A DYNAMIC AND SCALABLE BENCHMARK FOR EVALUATING PERCEPTION ABILITY OF LVLMS		
F05	Guo Chen, Yicheng Liu, Yifei Huang, Yuping He, Baoqi Pei, Jilan Xu, Yali Wang, Tong Lu, Limin Wang	南京大学	ICLR 2025
	Cg-bench: Clue-grounded question answering benchmark for long video understanding		
F06	Guangtao Lv, Chenghao Xu, Jiexi Yan, Muli Yang, Cheng Deng	西安电子科技大学	ICLR 2025
	Towards Unified Human Motion-Language Understanding via Sparse Interpretable Characterization		
F07	Shangzhe Di, Zhelun Yu, Guanghao Zhang, Haoyuan Li, Tao Zhong, Hao Cheng, Polin Li, Wanggui He, Fangxun Shu, Hao Jiang	上海交通大学	ICLR 2025
	Streaming Video Question-Answering with In-context Video KV-Cache Retrieval		
F08	Yuming Chen, Jiangyan Feng, Haodong Zhang, Lijun Gong, Feng Zhu, Rui Zhao, Qibin Hou, Ming-Ming Cheng, Yibing Song	南开大学	ICLR 2025
	Re-Aligning Language to Visual Objects with an Agentic Workflow		
F09	Zhanwei Zhang, Shizhao Sun, Wenxiao Wang, Deng Cai, Jiang Bian	浙江大学	ICLR 2025
	FlexCAD: Unified and Versatile Controllable CAD Generation with Fine-tuned Large Language Models		
F10	Kai Gan, Bo Ye, Minling Zhang, Tong Wei	东南大学	ICLR 2025
	Semi-Supervised CLIP Adaptation by Enforcing Semantic and Trapezoidal Consistency		
F11	Haochen Wang, Anlin Zheng, Yucheng Zhao, Tiancai Wang, Xiangyu Zhang, Zhaoxiang Zhang	中国科学院自动化研究所	ICLR 2025
	Reconstructive Visual Instruction Tuning		
F12	Zhipei Xu, Xuanyu Zhang, Runyi Li, Zecheng Tang, Qing Huang, Jian Zhang	北京大学	ICLR 2025
	FakeShield: Explainable Image Forgery Detection and Localization via Multi-modal Large Language Models		
F13	Chongjie Si, Zhiyi Shi, Shifan Zhang, Xiaokang Yang, Hanspeter Pfister, Wei Shen	上海交通大学	ICLR 2025

	Unleashing the Power of Task-Specific Directions in Parameter Efficient Fine-tuning		
F14	Minheng Ni, Yutao Fan, Lei Zhang, Wangmeng Zuo	哈尔滨工业大学	ICLR 2025
	Visual-O1: Understanding Ambiguous Instructions via Multi-modal Multi-turn Chain-of-thoughts Reasoning		
F15	Jiabo Ye, Haiyang Xu, Haowei Liu, Anwen Hu, Ming Yan, Qi Qian, Ji Zhang, Fei Huang, Jingren Zhou	阿里巴巴通义实验室	ICLR 2025
	mPLUG-Owl3: Towards Long Image-Sequence Understanding in Multi-Modal Large Language Models		
F16	Guangkai Xu, Yongtao Ge, Mingyu Liu, Chengxiang Fan, Zhiyue Zhao, Hao Chen, Chunhua Shen	浙江大学	ICLR 2025
	What Matters When Repurposing Diffusion Models for General Dense Perception Tasks?		
F17	Zhengyao Lv, Chenyang Si, Junhao Song, Zhenyu Yang, Yu Qiao, Ziwei Liu, Kwan-Yee K. Wong	香港大学	ICLR 2025
	FasterCache: Training-Free Video Diffusion Model Acceleration with High Quality		
F18	Chang Zou, Xuyang Liu, Ting Liu, Siteng Huang, Linfeng Zhang	四川大学	ICLR 2025
	Accelerating Diffusion Transformers with Token-wise Feature Caching		
F19	Hulingxiao He, Geng Li, Zijun Geng, Jinglin Xu, Yuxin Peng	北京大学	ICLR 2025
	Analyzing and Boosting the Power of Fine-Grained Visual Recognition for Multi-modal Large Language Models		
F20	Xiaojun Jia, Tianyu Pang, Chao Du, Yihao Huang, Jindong Gu, Yang Liu, Xiaochun Cao, Min Lin	Nanyang Technological University	ICLR 2025
	Improved techniques for optimization-based jailbreaking on large language models		
F21	Yanqi Dai, Huanran Hu, Lei Wang, Shengjie Jin, Xu Chen, Zhiwu Lu	中国人民大学高瓴人工智能学院	ICLR 2025
	MMRole: A Comprehensive Framework for Developing and Evaluating Multimodal Role-Playing Agents		
F22	Kairui Yang, Zihao Guo, Gengjie Lin, Haotian Dong, Zhao Huang, Yipeng Wu, Die Zuo, Jibin Peng, Ziyuan Zhong, Xin Wang, Qing Guo, Xiaosong Jia, Junchi Yan, Di Lin	天津大学	ICLR 2025
	Trajectory-LLM: A Language-based Data Generator for Trajectory Prediction in Autonomous Driving		
F23	Yuqi Yang, Pengtao Jiang, Qibin Hou, Hao Zhang, Jinwei Chen, Bo Li	南开大学	ICLR 2025
	Multi-Task Dense Predictions via Unleashing the Power of Diffusion		
F24	Jiang-Xin Shi, Tong Wei, Zhi Zhou, Jie-Jing Shao, Xin-Yan Han, Yu-Feng Li	南京大学	ICML 2024
	Long-Tail Learning with Foundation Model: Heavy Fine-Tuning Hurts		

F25	Zeyu Wang, Libo Zhao, Jizheng Zhang, Rui Song, Haiyu Song, Jiana Meng, Shidong Wang	大连民族大学	IJCV 2025
	Multi-Text Guidance Is Important: Multi-Modality Image Fusion via Large Generative Vision-Language Model		
F26	Jianing Qiu, Jian Wu, Hao Wei, Peilun Shi, Minqing Zhang, Yunyun Sun, Lin Li, Hanruo Liu, Hongyi Liu, Simeng Hou, Yuyang Zhao, Xuehui Shi, Junfang Xian, Xiaoxia Qu, Sirui Zhu, Lijie Pan, Xiaoniao Chen, Xiaojia Zhang, Shuai Jiang, Kebing Wang, Chenlong Yang, Mingqiang Chen, Sujie Fan, Jianhua Hu, Aiguo Lv, Hui Miao, Li Guo, Shujun Zhang, Cheng Pei, Xiaojuan Fan, Jianqin Lei, Ting Wei, Junguo Duan, Chun Liu, Xiaobo Xia, Siqi Xiong, Junhong Li, Kyle Lam, Benny Lo, Yih Chung Tham, Tien Yin Wong, Ningli Wang, Wu Yuan	香港中文大学	NEJM AI 2024
	Development and Validation of a Multimodal Multitask Vision Foundation Model for Generalist Ophthalmic Artificial Intelligence		
F27	Pengxiang Li, Zhi Gao, Bofei Zhang, Tao Yuan, Yuwei Wu, Mehrtash Harandi, Yunde Jia, Songchun Zhu, Qing Li	北京理工大学	NeurIPS 2024
	FIRE A Dataset for Feedback Integration and Refinement Evaluation of Multimodal Models		
F28	Fangcheng Liu, Yehui Tang, Zhenhua Liu, Yunsheng Ni, Douyu Tang, Kai Han, Yunhe Wang	华为	NeurIPS 2024
	Kangaroo: Lossless Self-Speculative Decoding for Accelerating LLMs via Double Early Exiting		
F29	Chenyu Zheng, Wei Huang, Rongzhen Wang, Guoqiang Wu, Jun Zhu, Chongxuan Li	中国人民大学	NeurIPS 2024
	On Mesa-Optimization in Autoregressively Trained Transformers: Emergence and Capability		
F30	Jiaqi Tang, Hao Lu, Ruizheng Wu, Xiaogang Xu, Ke Ma, Cheng Fang, Bin Guo, Jiangbo Lv, Qifeng Chen, Yingcong Chen	香港科技大学	NeurIPS 2024
	Hawk: Learning to Understand Open-World Video Anomalies		
F31	Junyang Wang, Haiyang Xu, Haitao Jia, Xi Zhang, Ming Yan, Weizhou Shen, Ji Zhang, Fei Huang, Jitao Sang	阿里巴巴	NeurIPS 2024
	Mobile-Agent-v2: Mobile Device Operation Assistant with Effective Navigation via Multi-Agent Collaboration		
F32	Cheng Chen, Junchen Zhu, Xu Luo, Hengtao Shen, Lianli Gao, Jingkuan Song	电子科技大学	NeurIPS 2024
	CoIN: A Benchmark of Continual Instruction tuNing for Multimodel Large Language Model		
F33	Tao Zhang, Xiangtai Li, Hao Fei, Haobo	武汉大学	NeurIPS

	Yuan, Shengqiong Wu, Shunping Ji, Chen Change Loy, Shuicheng Yan		2024
	OMG-LLaVA: Bridging Image-level, Object-level, Pixel-level Reasoning and Understanding		
F34	Dailing Zhang, Shiyu Hu, Xiaokun Feng, Xuchen Li, Meiqi Wu, Jing Zhang, Kaiqi Huang	中国科学院自动化研究所	NeurIPS 2024
	Beyond accuracy: Tracking more like human via visual search		
F35	Chuanhao Li, Zhen Li, Chenchen Jing, Shuo Liu, Wenqi Shao, Yuwei Wu, Ping Luo, Yu Qiao, Kaipeng Zhang	上海人工智能实验室	NeurIPS 2024
	SearchLVLMs: A Plug-and-Play Framework for Augmenting Large Vision-Language Models by Searching Up-to-Date Internet Knowledge		
F36	Chenxi Zhao, Jinglei Shi, Liqiang Nie, Jufeng Yang	南开大学	NeurIPS 2024
	To err like human: Affective bias-inspired measures for visual emotion recognition evaluation		
F37	Lianyu Hu, Tongkai Shi, Wei Feng, Fanhua Shang, Liang Wan	天津大学	NeurIPS 2024
	Deep Correlated Prompting for Visual Recognition with Missing Modalities		
F38	Jiazhao Zhang, Kunyu Wang, Shaoan Wang, Minghan Li, Haoran Liu, Songlin Wei, Zhongyuan Wang, Zhizheng Zhang, He Wang	北京大学	RSS 2025
	Uni-NaVid: A Video-based Vision-Language-Action Model for Unifying Embodied Navigation Tasks		
F39	Yizhou Wang, Yixuan Wu, Weizhen He, Xun Guo, Feng Zhu, Lei Bai, Rui Zhao, Jian Wu, Tong He, Wanli Ouyang, Shixiang Tang	香港中文大学	TPAMI 2025
	Hulk: A Universal Knowledge Translator for Human-Centric Tasks		
F40	Da-Wei Zhou, Yuanhan Zhang, Yan Wang, Jingyi Ning, Han-Jia Ye, De-Chuan Zhan, Ziwei Liu	南京大学	TPAMI 2025
	Learning without Forgetting for Vision-Language Models		
F41	Ke Liang, Lingyuan Meng, Meng Liu, Yue Liu, Wenxuan Tu, Siwei Wang, Sihang Zhou, Xinwang Liu, Fuchun Sun, Kunlun He	国防科技大学	TPAMI 2024
	A Survey of Knowledge Graph Reasoning on Graph Types: Static, Dynamic, and Multi-Modal		
F42	Di Wang, Meiqi Hu, Yao Jin, Yuchun Miao, Jiaqi Yang, Yichu Xu, Xiaolei Qin, Jiaqi Ma, Lingyu Sun, Chenxing Li, Chuan Fu, Hongruixuan Chen, Chengxi Han, Naoto Yokoya, Jing Zhang, Minqiang Xu, Lin Liu, Lefei Zhang, Chen Wu, Bo Du, Dacheng Tao, Liangpei Zhang	武汉大学	TPAMI 2025
	HyperSIGMA: Hyperspectral Intelligence Comprehension Foundation Model		
F43	Huan Ma, Jingdong Chen, Guangyu	天津大学	其他 2025

	Wang, Changqing Zhang		
	Estimating LLM Uncertainty with Logits		
F44	Huanjin Yao, Jiaxing Huang, Wenhao Wu, Jingyi Zhang, Yibo Wang, Shunyu Liu, Yingjie Wang, Yuxin Song, Haocheng Feng, Li Shen, Dacheng Tao	清华大学	其他 2024
	Mulberry: Empowering mllm with o1-like reasoning and reflection via collective monte carlo tree search		
F45	Chenyangguang Zhang, Alexandros Delitzas, Fangjinhua Wang, Ruida Zhang, Xiangyang Ji, Marc Pollefeys, Francis Engelmann	清华大学	CVPR 2025
	Open-Vocabulary Functional 3D Scene Graphs for Real-World Indoor Spaces		
F46	Minghang Zheng, Xinhao Cai, Qingchao Chen, Yuxin Peng, Yang Liu	北京大学	ECCV 2024
	Training-free Video Temporal Grounding using Large-scale Pre-trained Models		
F47	Ge Wu, Xin Zhang, Zheng Li, Zhaowei Chen, Jiajun Liang, Jian Yang, Xiang Li	南开大学，旷视科技	ECCV 2024
	Cascade Prompt Learning for Vision-Language Model Adaptation		
F48	Yake Wei, Yu Miao, Dongzhan Zhou, Di Hu	中国人民大学	其他
	MokA: Multimodal Low-Rank Adaptation for MLLMs		
C01	Jiazhong Cen, Jiemin Fang, Chen Yang, Lingxi Xie, Xiaopeng Zhang, Wei Shen, Qi Tian	上海交通大学	AAAI 2025
	Segment Any 3D Gaussians		
C02	Jiawei Li, Hongwei Yu, Jiansheng Chen, Xinlong Ding, Jinlong Wang, Jinyuan Liu, Bochao Zou, Huimin Ma	北京科技大学	AAAI 2025
	A2RNet: Adversarial Attack Resilient Network for Robust Infrared and Visible Image Fusion		
C03	Chuanguang Yang, Xinqiang Yu, Han Yang, Zhulin An, Chengqing Yu, Lipo Huang, Yongjun Xu	中国科学院计算技术研究所	AAAI 2025
	Multi-Teacher Knowledge Distillation with Reinforcement Learning for Visual Recognition		
C04	Chen Duan, Qianyi Jiang, Pei Fu, Jiamin Chen, Shengxi Li, Zining Wang, Shan Guo, Junfeng Luo	美团	AAAI 2025
	InstructOCR: Instruction Boosting Scene Text Spotting		
C05	Chunyu Zhao, Wentao Mu, Xian Zhou, Wenbo Liu, Fei Yan, Tao Deng	西南交通大学	AAAI 2025
	SalM²: An Extremely Lightweight Saliency Mamba Model for Real-Time Cognitive Awareness of Driver Attention		
C06	Shiyi Zheng, Peizhi Zhao, Zhilong Zheng, Peihang He, Haonan Cheng, Yi Cai, Qingbao Huang	广西大学	AAAI 2025
	Look Around Before Locating: Considering Content and Structure Information for Visual Grounding		
C07	Zeqin Yu, Jiangqun Ni, Jian Zhang, Haoyi	中山大学	AAAI 2025

	Deng, Yuzhen Lin		
	Reinforced Multi-teacher Knowledge Distillation for Efficient General Image Forgery Detection and Localization		
C08	Chenxin Li, Xinyu Liu, Wuyang Li, Cheng Wang, Hengyu Liu, Yifan Liu, Zhen Chen, Yixuan Yuan	香港中文大学	AAAI 2025
	U-KAN Makes Strong Backbone for Medical Image Segmentation and Generation		
C09	Xin Chen, Ben Kang, Wanting Geng, Jiawen Zhu, Yi Liu, Dong Wang, Huchuan Lu	香港城市大学	AAAI 2025
	SUTrack: Towards Simple and Unified Single Object Tracking		
C10	Nan Yang, Yang Wang, Zhanwen Liu, Meng Li, Yisheng An, Xiangmo Zhao	长安大学	AAAI 2025
	SMamba: Sparse Mamba for Event-based Object Detection		
C11	Yiming Yang, Yueru Luo, Bingkun He, Erlong Li, Zhipeng Cao, Chao Zheng, Shuqi Mei, Zhen Li	香港中文大学深圳	AAAI 2025
	Topo2Seq: Enhanced Topology Reasoning via Topology Sequence Learning		
C12	Yuxuan Zhao, Weijian Ruan, He Li, Mang Ye	武汉大学	AAAI 2025
	NightReID: A Large-Scale Nighttime Person Re-Identification Benchmark		
C13	Kunpeng Wang, Keke Chen, Chenglong Li, Zhengzheng Tu, Bin Luo	安徽大学	AAAI 2025
	Alignment-Free RGB-T Salient Object Detection: A Large-Scale Dataset and Progressive Correlation Network		
C14	Sijia Chen, En Yu, Wenbing Tao	华中科技大学	AAAI 2025
	Cross-View Referring Multi-Object Tracking		
C15	Yulin He, Wei Chen, Siqi Wang, Tianci Xun, Yusong Tan	国防科技大学	AAAI 2025
	Achieving Speed-Accuracy Balance in Vision-based 3D Occupancy Prediction via Geometric-Semantic Disentanglement		
C16	Qijian Tian, Xin Tan, Yuan Xie, Lizhuang Ma	上海交通大学	AAAI 2025
	DrivingForward: Feed-forward 3D Gaussian Splatting for Driving Scene Reconstruction from Flexible Surround-view Input		
C17	Duanyang Yuan, Sihang Zhou, Xiaoshu Chen, Dong Wang, Ke Liang, Xinwang Liu, Jian Huang	国防科技大学	AAAI 2025
	Knowledge Graph Completion with Relation-Aware Anchor Enhancement		
C18	Jiancheng Pan, Yanxing Liu, Yuqian Fu, Muyuan Ma, Jiahao Li, Danda Pani Paudel, Luc Van Gool, Xiaomeng Huang	清华大学	AAAI 2025
	Locate Anything on Earth: Advancing Open-Vocabulary Object Detection for Remote Sensing Community		
C19	Chengyang Ye, Geyunzhi Zhu, Pingping Zhang	大连理工大学	AAAI 2025
	Towards Open-Vocabulary Remote Sensing Image Semantic Segmentation		
C20	Shide Du, Zihan Fang, Yanchao Tan, Changwei Wang, Shiping Wang,	福州大学	AAAI 2025

	Wenzhong Guo		
	OpenViewer: Openness-Aware Multi-View Learning		
C21	Haoke Xiao, Lv Tang, Pengtao Jiang, Hao Zhang, Jinwei Chen, Bo Li	vivo Mobile Communication Co., Ltd, Shanghai, China	AAAI 2025
	Boosting Vision State Space Model with Fractal Scanning		
C22	Yuanhao Guo, Yanqiang Huo, Ning Cheng, Zongjun Pan, Xiaoming Yi, Jiankun Cao, Haoyu Sun, Jianqing Wu	交通运输部公路科学研究院 中公高科养护科技股份有限公司 山东大学	Comput.-Aided Civ. Infrastruct. Eng 2025
	Deep Line Segment Detection for Concrete Pavement Distress Assessment		
C23	Songsong Duan, Xi Yang, Nannan Wang	西安电子科技大学	CVPR 2025
	Multi-label Prototype Visual Spatial Search for Weakly Supervised Semantic Segmentation		
C24	Xin Zhang, Xue Yang, Yuxuan Li, Jian Yang, Ming-Ming Cheng, Xiang Li	南开大学	CVPR 2025
	RSAR: Restricted State Angle Resolver and Rotated SAR Benchmark		
C25	Chenyu Zhang, Kunlun Xu, Bing Su, Xu Zou, Yuxin Peng, Kunlun Xu	北京大学	CVPR 2025
	STOP: Integrated Spatial-Temporal Dynamic Prompting for Video Understanding		
C26	Yingping Liang, Yutao Hu, Wenqi Shao, Ying Fu	北京理工大学	CVPR 2025
	Distilling Monocular Foundation Model for Fine-grained Depth Completion		
C28	Yisi Luo, Xile Zhao, Kai Ye, Deyu Meng	西安交通大学	CVPR 2025
	STINR: Deciphering Spatial Transcriptomics via Implicit Neural Representation		
C29	Yunze Liu, Li Yi	清华大学交叉信息研究院	CVPR 2025
	MAP: Unleashing Hybrid Mamba-Transformer Vision Backbone's Potential with Masked Autoregressive Pretraining		
C30	Jiayuan Rao, Haoning Wu, Hao Jiang, Ya Zhang, Yanfeng Wang, Weidi Xie	上海交通大学	CVPR 2025
	Towards Universal Soccer Video Understanding		
C31	Guangqian Guo, Yong Guo, Xuehui Yu, Wenbo Li, Yaoxing Wang, Shan Gao	西北工业大学	CVPR 2025
	Segment Any-Quality Images with Generative Latent Space Enhancement		
C32	Kaiyu Li, Ruixun Liu, Xiangyong Cao, Xueru Bai, Feng Zhou, Deyu Meng, Zhi Wang	西安交通大学	CVPR 2025
	SegEarth-OV: Towards Training-Free Open-Vocabulary Segmentation for Remote Sensing Images		
C33	Yongliang Wu, Xinting Hu, Yuyang Sun, Yizhou Zhou, Wenbo Zhu, Fengyun Rao, Bernt Schiele, Xu Yang	东南大学	CVPR 2025
	Number it: Temporal Grounding Videos like Flipping Manga		
C34	Ruihai Wu, Ziyu Zhu, Yuran Wang, Yue Chen, Jiarui Wang, Hao Dong	北京大学	CVPR 2025

	GarmentPile: Point-Level Visual Affordance Guided Retrieval and Adaptation for Cluttered Garments Manipulation		
C35	Mengqiu Xu, Kaixin Chen, Heng Guo, Yixiang Huang, Ming Wu, Zhenwei Shi, Chuang Zhang, Jun Guo	北京邮电大学	CVPR 2025
	MFogHub: Bridging Multi-Regional and Multi-Satellite Data for Global Marine Fog Detection and Forecasting		
C36	Haifeng Zhang, Qinghui He, Xiuli Bi, Weisheng Li, Bo Liu, Bin Xiao	重庆邮电大学	CVPR2025
	Towards Universal AI-Generated Image Detection by Variational Information Bottleneck Network		
C37	Yuzhou Liu, Lingjie Zhu, Hanqiao Ye, Shangfeng Huang, Xiang Gao, Xianwei Zheng, Shuhan Shen	中国科学院自动化研究所	CVPR 2025
	BWFormer: Building Wireframe Reconstruction from airborne LiDAR point clouds with Transformer		
C38	Pan Yin, Kaiyu Li, Xiangyong Cao, Jing Yao, Lei Liu, Xueru Bai, Feng Zhou, Deyu Meng	西安交通大学	CVPR 2025
	Towards Satellite Image Road Graph Extraction: A Global-Scale Dataset and A Novel Method		
C39	Xiang Xu, Lingdong Kong, Hui Shuai, Liang Pan, Ziwei Liu, Qingshan Liu	南京航空航天大学	CVPR 2025
	LiMoE: Mixture of LiDAR Representation Learners from Automotive Scenes		
C40	Zhe Shan, Yang Liu, Lei Zhou, Cheng Yan, Heng Wang, Xia Xie	海南大学	CVPR 2025
	ROS-SAM: High-Quality Interactive Segmentation for Remote Sensing Moving Object		
C41	Chuanyu Sun, Jiqing Zhang, Yang Wang, Huilin Ge, Qianchen Xia, Baocai Yin, Xin Yang	大连理工大学	CVPR 2025
	Exploring Historical Information for RGBE Visual Tracking with Mamba		
C42	Changbin Zhang, Yujie Zhong, Kai Han	香港大学	CVPR 2025
	Mr. DETR: Instructive Multi-Route Training for Detection Transformers		
C43	Sitong Gong, Yunzhi Zhuge, Lu Zhang, Zongxin Yang, Pingping Zhang, Huchuan Lu	大连理工大学	CVPR 2025
	The Devil is in Temporal Token: High Quality Video Reasoning Segmentation		
C44	Yuhao Wang, Yongfeng Lv, Pingping Zhang, Huchuan Lu	大连理工大学	CVPR 2025
	IDEA: Inverted Text with Cooperative Deformable Aggregation for Multi-modal Object Re-Identification		
C45	Bowen Yin, Jiaolong Cao, Ming-Ming Cheng, Qibin Hou	南开大学计算机学院	CVPR 2025
	DFormerv2: Geometry Self-Attention for RGBD Semantic Segmentation		
C46	Hao Tan, Zichang Tan, Jun Li, Ajian Liu, Jun Wan, Zhen Lei	中国科学院自动化研究所	CVPR 2025
	Recover and Match: Open-Vocabulary Multi-Label Recognition through Knowledge-Constrained Optimal Transport		

C47	Yuxin Yao, Zhi Deng, Junhui Hou	香港城市大学	CVPR 2025
	RigGS: Rigging of 3D Gaussians for Modeling Articulated Objects in Videos		
C48	Jiahao He, Keren Fu, Xiaohong Liu, Qijun Zhao	四川大学计算机学院	CVPR 2025
	Samba: A Unified Mamba-based Framework for General Salient Object Detection		
C49	Dachong Li, Li Li, Zhuangzhuang Chen, Jianqiang Li	深圳大学	CVPR 2025
	ShiftwiseConv: Small Convolutional Kernel with Large Kernel Effect		
C50	Yunxiang Fu, Meng Lou, Yizhou Yu	香港大学	CVPR 2025
	SegMAN: Omni-scale Context Modeling with State Space Models and Local Attention for Semantic Segmentation		
G01	Hao Ren, Yiming Zeng, Zetong Bi, Zhaoliang Wan, Junlong Huang, Hui Cheng	中山大学	CVPR 2025
	Prior Does Matter: Visual Navigation via Denoising Diffusion Bridge Models		
G02	Xun Huang, Jinlong Wang, Qiming Xia, Siheng Chen, Bisheng Yang, Cheng Wang, Chenglu Wen	厦门大学	CVPR 2025
	V2X-R: Cooperative LiDAR-4D Radar Fusion for 3D Object Detection with Denoising Diffusion		
G03	Jingtao Li, Yingyi Liu, Xinyu Wang, Yunning Peng, Chen Sun, Shaoyu Wang, Zhendong Sun, Tian Ke, Xiao Jiang, Tangwei Lu, Anran Zhao, Yanfei Zhong	武汉大学	CVPR 2025
	HyperFree: A Channel-adaptive and Tuning-free Foundation Model for Hyperspectral Remote Sensing Imagery		
G04	Runmin Zhang, Jun Ma, Siyuan Cao, Lun Luo, Beinan Yu, Shujie Chen, Junwei Li, Huiliang Shen	浙江大学	ECCV 2024
	SCPNet: Unsupervised Cross-modal Homography Estimation via Intra-modal Self-supervised Learning		
G05	Zhongyu Xia, Zhiwei Lin, Xinhao Wang, Yongtao Wang, Yun Xing, Chengxiang Qi, Nan Dong, Ming-Hsuan Yang	北京大学	ECCV 2024
	HENet: Hybrid Encoding for End-to-end Multi-task 3D Perception from Multi-view Cameras		
G06	Peiqi Jiao, Yuecong Min, Xilin Chen	中国科学院计算技术研究所	ECCV 2024
	Visual Alignment Pre-training for Sign Language Translation		
G07	Kaishen Yuan, Zitong Yu, Xin Liu, Weicheng Xie, Huanjing Yue, Jingyu Yang	天津大学 大湾区大学	ECCV 2024
	AUFormer: Vision Transformers are Parameter-Efficient Facial Action Unit Detectors		
G08	Ziye Wang, Yiran Qin, Lin Zeng, Ruimao Zhang	中山大学深圳校区	ICLR 2025
	High-Dynamic Radar Sequence Prediction for Weather Nowcasting Using Spatiotemporal Coherent Gaussian Representation		
G09	Letian Huang, Jiayang Bai, Jie Guo,	南京大学	ECCV 2024

	Yuanqi Li, Yanwen Guo		
	On the Error Analysis of 3D Gaussian Splatting and an Optimal Projection Strategy		
G10	Yufan Liu, Wanqian Zhang, Dayan Wu, Zheng Lin, Jingzi Gu, Weiping Wang	中国科学院信息工程研究所， 中国科学院大学	ECCV 2024
	Prediction Exposes Your Face: Black-box Model Inversion via Prediction Alignment		
G11	Xinhao Luo, Man Yao, Yuhong Chou, Bo Xu, Guoqi Li	中国科学院大学自动化研究所	ECCV 2024
	Integer-Valued Training and Spike-Driven Inference Spiking Neural Network for High-performance and Energy-efficient Object Detection		
G12	Junliang Chen, Huaiyuan Xu, Yi Wang, Lap-Pui Chau	香港理工大学	ICLR 2025
	OccProphet: Pushing Efficiency Frontier of Camera-Only 4D Occupancy Forecasting with Observer-Forecaster-Refiner Framework		
G13	Jinyang Li, En Yu, Sijia Chen, Wenbing Tao	华中科技大学	ICLR 2025
	OVTR: End-to-End Open-Vocabulary Multiple Object Tracking with Transformer		
G14	Chaodong Xiao, Minghan Li, Zhengqiang Zhang, Deyu Meng, Lei Zhang	香港理工大学（The Hong Kong Polytechnic University）	ICLR 2025
	Spatial-Mamba: Effective Visual State Space Models via Structure-Aware State Fusion		
G15	Wenhui Tan, Boyuan Li, Chuhao Jin, Wenbing Huang, Xiting Wang, Ruihua Song	中国人民大学高瓴人工智能学院	ICLR 2025
	Think Then React: Towards Unconstrained Action-to-Reaction Motion Generation		
G16	Yunheng Li, Zhongyu Li, Quansheng Zeng, Qibin Hou, Ming-Ming Cheng	南开大学	ICML 2024
	Cascade-CLIP: Cascaded Vision-Language Embeddings Alignment for Zero-Shot Semantic Segmentation		
G17	Xinhang Wan, Jiyuan Liu, Xinwang Liu, Yi Wen, Hao Yu, Siwei Wang, Shengju Yu, Tianjiao Wan, Jun Wang, En Zhu	国防科技大学	ICML 2024
	Decouple then Classify: A Dynamic Multi-view Labeling Strategy with Shared and Specific Information		
G18	Shi Yin, Shijie Huang, Shangfei Wang, Jinshui Hu, Tao Guo, Bing Yin, Baocai Yin, Cong Liu	合肥综合性国家科学中心人工智能研究院	IJCAI 2024
	1DFormer: A Transformer Architecture Learning 1D Landmark Representations for Facial Landmark Tracking		
G19	Shiyu Hu, Xin Zhao, Kaiqi Huang	(NTU), Singapore	IJCV 2024
	SOTVerse: A User-defined Task Space of Single Object Tracking		
G20	Xizhou Zhu, Xue Yang, Zhaokai Wang, Hao Li, Wenhan Dou, Junqi Ge, Lewei Lu, Yu Qiao, Jifeng Dai	上海人工智能实验室， 上海交通	NeurIPS 2024

		大学	
	Parameter-Inverted Image Pyramid Networks		
G21	Yuxuan Li, Xiang Li, Weijie Li, Qibin Hou, Li Liu, Ming-Ming Cheng, Jian Yang	南开大学	NeurIPS 2024
	SARDet-100K: Towards Open-Source Benchmark and ToolKit for Large-Scale SAR Object Detection		
G22	Yaohua Zha, Naiqi Li, Yanzi Wang, Tao Dai, Hang Guo, Bin Chen, Zhi Wang, Zhihao Ouyang, Shutao Xia	清华大学深圳国际研究生院	NeurIPS 2024
	LCM: Locally Constrained Compact Point Cloud Model for Masked Point Modeling		
G23	Yanmin Wu, Jiarui Meng, Haijie Li, Chenming Wu, Yahao Shi, Xinhua Cheng, Chen Zhao, Haocheng Feng, Errui Ding, Jingdong Wang, Jian Zhang	北京大学	NeurIPS 2024
	OpenGaussian: Towards Point-Level 3D Gaussian-based Open Vocabulary Understanding		
G24	Tieyuan Chen, Huabin Liu, Tianyao He, Yihang Chen, Chaofan Gan, Xiao Ma, Cheng Zhong, Yang Zhang, Yingxue Wang, Hui Lin, Weiyao Lin	上海交通大学	NeurIPS 2024
	MECD: Unlocking Multi-Event Causal Discovery in Video Reasoning		
G25	Linhui Xiao, Xiaoshan Yang, Fang Peng, Yaowei Wang, Changsheng Xu	中国科学院自动化研究所	NeurIPS 2024
	OneRef: Unified One-tower Expression Grounding and Segmentation with Mask Referring Modeling		
G26	Zhu Yu, Runmin Zhang, Jiacheng Ying, Junchen Yu, Xiaohai Hu, Lun Luo, Siyuan Cao, Huiliang Shen	浙江大学	NeurIPS 2024
	Context and Geometry Aware Voxel Transformer for Semantic Scene Completion		
G27	Hao Deng, Kunlei Jing, Shengmei Chen, Cheng Liu, Jiawei Ru, Bo Jiang, Lin Wang	西北大学	NeurIPS 2024
	LinNet: Linear Network for Efficient Point Cloud Representation Learning		
G28	Dingkang Liang, Xin Zhou, Wei Xu, Xingkui Zhu, Zhikang Zou, Xiaoqing Ye, Xiao Tan, Xiang Bai	华中科技大学	NeurIPS 2024
	PointMamba: A Simple State Space Model for Point Cloud Analysis		
G29	Yilin Wei, Jianjian Jiang, Chengyi Xing, Xiantuo Tan, Xiaoming Wu, Hao Li, Mark Cutkosky, Weishi Zheng	中山大学	NeurIPS 2024
	Grasp as You Say: Language-guided Dexterous Grasp Generation		
G30	Jiangming Shi, Xiangbo Yin, Yachao Zhang, Zhizhong Zhang, Yuan Xie, Yanyun Qu	厦门大学	NeurIPS 2024
	Learning commonality, divergence and variety for unsupervised visible-infrared person re-identification		
G31	Yao Wu, Mingwei Xing, Yachao Zhang, Xiaotong Luo, Yuan Xie, Yanyun Qu	厦门大学	NeurIPS 2024

	UniDSeg: Unified Cross-Domain 3D Semantic Segmentation via Visual Foundation Models Priors		
G32	Kun Wang, Zhiqiang Yan, Junkai Fan, Wanlu Zhu, Jun Li, Jian Yang	南京理工大学	NeurIPS 2024
	DCDepth: Progressive Monocular Depth Estimation in Discrete Cosine Domain		
G33	Yanping Fu, Wenbin Liao, Xinyuan Liu, Hang Xu, Yike Ma, Yucheng Zhang, Feng Dai	中国科学院计算技术研究所	NeurIPS 2024
	TopoLogic: An Interpretable Pipeline for Lane Topology Reasoning on Driving Scenes		
G34	Chunhui Zhang, Li Liu, Guanjie Huang, Hao Wen, Xi Zhou, Yanfeng Wang	上海交通大学	NeurIPS 2024
	WebUOT-1M: Advancing Deep Underwater Object Tracking with A Million-Scale Benchmark		
G35	Jialong Zuo, Ying Nie, Hanyu Zhou, Huaxin Zhang, Haoyu Wang, Tianyu Guo, Nong Sang, Changxin Gao	华中科技大学	NeurIPS 2024
	Cross-Video Identity Correlating for Person Re-identification Pre-training		
G36	Qishuai Wen, Chunguang Li	北京邮电大学	NeurIPS 2024
	Rethinking Decoders for Transformer-based Semantic Segmentation: A Compression Perspective		
G37	Hanyu Zhou, Yi Chang, Zhiwei Shi, Wending Yan, Gang Chen, Yonghong Tian, Luxin Yan	National University of Singapore	TPAMI 2024
	Adverse Weather Optical Flow: Cumulative Homogeneous-Heterogeneous Adaptation		
G38	Jinglin Xu, Yongming Rao, Jie Zhou, Jiwen Lu	北京科技大学	TPAMI 2025
	Transferable Unintentional Action Localization with Language-guided Intention Translation		
G39	Xin Liu, Rong Qin, Junchi Yan, Jufeng Yang	南开大学	TPAMI 2024
	NCMNet: Neighbor Consistency Mining Network for Two-View Correspondence Pruning		
G40	Chao Xiao, Wei An, Yifan Zhang, Zhuo Su, Miao Li, Weidong Sheng, Matti Pietikäinen, Li Liu	国防科技大学电子科学学院	TPAMI 2024
	Highly Efficient and Unsupervised Framework for Moving Object Detection in Satellite Videos		
G41	Tongkun Guan, Wei Shen, Xiaokang Yang	上海交通大学	TPAMI 2025
	CCDPlus: Towards Accurate Character to Character Distillation for Text Recognition		
G42	Huafeng Li, Zengyi Yang, Yafei Zhang, Wei Jia, Zhengtao Yu, Yu Liu	昆明理工大学	TPAMI 2025
	MulFS-CAP: Multimodal Fusion-supervised Cross-modality Alignment Perception for Unregistered Infrared-visible Image Fusion		
G43	Yan Lu, Xinzhu Ma, Lei Yang, Tianzhu Zhang, Yating Liu, Qi Chu, Tong He, Yonghui Li, Wanli Ouyang	香港中文大学	TPAMI 2025

	GUPNet++: Geometry Uncertainty Propagation Network for Monocular 3D Object Detection		
G44	Yong Li, Yufei Sun, Zhen Cui, Pengcheng Shen, Shiguang Shan	东南大学， 中国科学院计算技术研究所	TPAMI 2025
	Instance-Consistent Fair Face Recognition		
G45	Yimin Fu, Zhunga Liu, Jialin Lv	香港浸会大学	TPAMI 2025
	Reason and Discovery: A New Paradigm for Open Set Recognition		
G46	Siyu Ren, Junhui Hou, Xiaodong Chen, Hongkai Xiong, Wenping Wang	香港城市大学	TPAMI 2025
	DDM: A Metric for Comparing 3D Shapes Using Directional Distance Fields		
G47	Yingda Yin*, Jiangran Lv*, Yang Wang, Haoran Liu, He Wang, Baoquan Chen	北京大学	TPAMI 2025
	Towards Robust Probabilistic Modeling on SO(3) via Rotation Laplace Distribution		
G48	Xuying Zhang, Bowen Yin, Zheng Lin, Qibin Hou, Dengping Fan, Ming-Ming Cheng	南开大学	TPAMI 2025
	Referring camouflaged object detection		
G49	Rongchang Li, Zhenhua Feng, Tianyang Xu, Linze Li, Xiaojun Wu, Muhammad Awais, Sara Atito, Josef Kittler	江南大学	ECCV 2024
	C2C: Component-to-composition learning for zero-shot compositional action recognition		
G50	Hongjian Liu, Qingsong Xie, Tianxiang Ye, Zhijie Deng, Chen Chen, Shixiang Tang, Xueyang Fu, Haonan Lu, Zhengjun Zha	OPPO	AAAI 2025
	SCott: Accelerating Diffusion Models with Stochastic Consistency Distillation		
D01	Wenyao Ni, Jiangrong Shen, Qi Xu, Gang Pan, Huajin Tang	西安交通大学	AAAI 2025
	ALADE-SNN: Adaptive Logit Alignment in Dynamically Expandable Spiking Neural Networks for Class Incremental Learning		
D02	Wentao Yu, Shuo Chen, Yongxin Tong, Tianlong Gu, Chen Gong	南京理工大学	AAAI 2025
	Modeling Inter-Intra Heterogeneity for Graph Federated Learning		
D03	Qiwei Li, Jiahuan Zhou	北京大学王选计算机研究所	AAAI 2025
	CAPrompt: Cyclic Prompt Aggregation for Pre-Trained Model Based Class Incremental Learning		
D04	Qi Deng, Shuaicheng Niu, Ronghao Zhang, Yaofu Chen, Runhao Zeng, Jian Chen, Xiping Hu	华南理工大学	AAAI 2025
	Learning to Generate Gradients for Test-Time Adaptation via Test-Time Training Layers		
D05	Yuhong Chen, Ailin Song, Huifeng Yin, Shuai Zhong, Fuhai Chen, Qi Xu, Shiping Wang, Mingkun Xu	福州大学	AAAI 2025
	Multi-View Incremental Learning with Structured Hebbian Plasticity for		

	Enhanced Fusion Efficiency		
D06	Yudong Zhang, Xu Wang, Xuan Yu, Zhaoyang Sun, Kai Wang, Yang Wang	中国科学技术大学	AAAI 2025
	Drawing Informative Gradients from Sources: A One-stage Transfer Learning Framework for Cross-city Spatiotemporal Forecasting		
D07	Zhuohang Dang, Minnan Luo, Chengyou Jia, Guang Dai, Xiaojun Chang, Jingdong Wang	西安交通大学	AAAI 2024
	Noisy correspondence learning with self-reinforcing errors mitigation		
D08	Yanru Sun, Zongxia Xie, Dongyue Chen, Emadeldeen Eldele, Qinghua Hu	天津大学	AAAI 2025
	Hierarchical Classification Auxiliary Network for Time Series Forecasting		
D09	Zhuang Qi, Lei Meng, Zhaochuan Li, Han Hu, Xiangxu Meng	山东大学	AAAI 2025
	Cross Silo Feature Space Alignment for Federated Learning on Clients with Imbalanced Data		
D10	Qinglun Zhang, Zhen Liu, Haoqiang Fan, Guanghui Liu, Bing Zeng, Shuaicheng Liu	电子科技大学	AAAI 2025
	FlowPolicy: Enabling Fast and Robust 3D Flow-based Policy via Consistency Flow Matching for Robot Manipulation		
D11	Yuanbin Qian, Shuhan Ye, Chong Wang, Xiaojie Cai, Jiangbo Qian, Jiafei Wu	宁波大学	AAAI 2025
	UCF-Crime-DVS: A Novel Event-Based Dataset for Video Anomaly Detection with Spiking Neural Networks		
D12	Ziheng Chen, Yue Song, Xiaojun Wu, Gaowen Liu, Nicu Sebe	University of Trento	ICLR 2025
	Understanding Matrix Function Normalizations in Covariance Pooling through the Lens of Riemannian Geometry		
D13	Luoyi Sun, Xuenan Xu, Mengyue Wu, Weidi Xie	浙江大学 上海人工智能实验室	ACMMM 2024
	Auto-ACD: A Large-scale Dataset for Audio-Language Representation Learning		
D14	Tongtong Feng, Xin Wang, Feilin Han, Leping Zhang, Wenwu Zhu	清华大学	ACMMM 2024
	U2UData: A Large-scale Cooperative Perception Dataset for Swarm UAVs Autonomous Flight		
D15	Yisu Liu, Jinyang An, Wanqian Zhang, Dayan Wu, Jingzi Gu, Zheng Lin, Weiping Wang	中国科学院信息工程研究所 中国科学院大学	ACMMM 2024
	Disrupting Diffusion: Token-Level Attention Erasure Attack against Diffusion-based Customization		
D16	Hongcheng Li, Yucan Zhou, Xiaoyan Gu, Bo Li, Weiping Wang	中国科学院信息工程研究所	ACMMM 2024
	Diversified Semantic Distribution Matching for Dataset Distillation		
D17	Shining Wang, Yunlong Wang, Ruiqi Wu, Bingliang Jiao, Wenxuan Wang, Peng Wang	西北工业大学	CVPR 2025
	SeCap: Self-Calibrating and Adaptive Prompts for Cross-view Person Re-Identification in Aerial-Ground Networks		

D18	Xiaomin Li, Yixuan Liu, Takashi Isobe, Xu Jia, Qinpeng Cui, Dong Zhou, Dong Li, You He, Huchuan Lu, Zhongdao Wang, Emad Barsoum	大连理工大学	CVPR 2025
	ReNeg: Learning Negative Embedding with Reward Guidance		
D19	Huitong Chen, Yu Wang, Yan Fan, Guosong Jiang, Qinghua Hu	天津大学	CVPR 2025
	Reducing Class-wise Confusion for Incremental Learning with Disentangled Manifolds		
D20	Jiyuan Liu, Xinwang Liu, Chuankun Li, Xinhang Wan, Hao Tan, Yi Zhang, Weixuan Liang, Qian Qu, Yu Feng, Renxiang Guan, Ke Liang	国防科技大学	CVPR 2025
	Large-scale Multi-view Tensor Clustering with Implicit Linear Kernels		
D21	Junlong Cheng, Bin Fu, Jin Ye, Guoan Wang, Tianbin Li, Haoyu Wang, Ruoyu Li, He Yao, Junren Chen, Jingwen Li, Yanzhou Su, Min Zhu, Junjun He	四川大学 上海人工智能实验室	CVPR 2025
	Interactive Medical Image Segmentation: A Benchmark Dataset and Baseline		
D22	Wei Shang, Dongwei Ren, Wanying Zhang, Yuming Fang, Wangmeng Zuo, Kede Ma	哈尔滨工业大学	ECCV 2024
	Arbitrary-Scale Video Super-Resolution with Structural and Textural Priors		
D23	Kunyu Wang, Xueyang Fu, Xin Lu, Chengjie Ge, Chengzhi Cao, Wei Di, Zhengjun Zha	中国科学技术大学	CVPR 2025
	Efficient Test-time Adaptive Object Detection via Sensitivity-Guided Pruning		
D24	Zhenyu Cui, Jiahuan Zhou, Yuxin Peng	北京大学	CVPR 2025
	DKC: Differentiated Knowledge Consolidation for Cloth-Hybrid Lifelong Person Re-identification		
D25	Boyun Li, Haiyu Zhao, Wenxin Wang, Peng Hu, Yuanbiao Gou, Xi Peng	四川大学	CVPR 2025
	MaIR: A Locality- and Continuity-Preserving Mamba for Image Restoration		
D26	Zijian Gao, Wangwang Jia, Xingxing Zhang, Doulan Zhou, Kele Xu, Dawei Feng, Yong Dou, Xinjun Mao, Huaimin Wang	国防科技大学	CVPR 2025
	Knowledge Memorization and Rumination for Pre-trained Model-based Class-Incremental Learning		
D27	Run He, Kai Tong, Di Fang, Han Sun, Ziqian Zeng, Haoran Li, Tianyi Chen, Huiping Zhuang	华南理工大学	CVPR 2025
	AFL: A Single-Round Analytic Approach for Federated Learning with Pre-trained Models		
D28	Guanyao Wu, Haoyu Liu, Hongming Fu, Yichuan Peng, Jinyuan Liu, Xin Fan, Risheng Liu	大连理工大学	CVPR 2025
	Every SAM Drop Counts: Embracing Semantic Priors for Multi-Modality Image Fusion and Beyond		
D29	Zhaoyi Tian, Feifeng Wang, Shiwei Wang, Zihao Zhou, Yao Zhu, Liquan Shen	上海大学	CVPR 2025

	High Dynamic Range Video Compression: A Large-Scale Benchmark Dataset and A Learned Bit-depth Scalable Compression Algorithm		
D30	Ziming Huang, Xurui Li, Haotian Liu, Feng Xue, Yuzhe Wang, Yu Zhou	华中科技大学	CVPR 2025
	AnomalyNCD: Towards Novel Anomaly Class Discovery in Industrial Scenarios		
D31	Yuan Wang, Ouxiang Li, Tingting Mu, Yanbin Hao, Kuien Liu, Xiang Wang, Xiangnan He	中国科学技术大学	CVPR 2025
	Precise, Fast, and Low-cost Concept Erasure in Value Space: Orthogonal Complement Matters		
D32	Qing Zhou, Junyu Gao, Qi Wang	西北工业大学	CVPR 2025
	Scale Efficient Training for Large Datasets		
D33	Rong Qin, Xingyu Liu, Jinglei Shi, Jing Lin, Jufeng Yang	南开大学	CVPR 2025
	Boosting the Dual-Stream Architecture in Ultra-High Resolution Segmentation with Resolution-Biased Uncertainty Estimation		
D34	Yanbiao Ma, Wei Dai, Wenke Huang, Jiayi Chen	西安电子科技大学	CVPR 2025
	Geometric Knowledge-Guided Localized Global Distribution Alignment for Federated Learning		
D35	Ming Hu, Jianfu Yin, Zhuangzhuang Ma, Jianheng Ma, Feiyu Zhu, Bingbing Wu, Ya Wen, Meng Wu, Cong Hu, Bingliang Hu, Quan Wang	中国科学院西安光学精密机械研究所	CVPR 2025
	β-FFT: Nonlinear Interpolation and Differentiated Training Strategies for Semi-Supervised Medical Image Segmentation		
D36	Zhilin Zhu, Xiaopeng Hong, Zhiheng Ma, Weijun Zhuang, Yaohui Ma, Yong Dai, Yaowei Wang	哈尔滨工业大学	ECCV 2024
	Reshaping the Online Data Buffering and Organizing Mechanism for Continual Test-Time Adaptation		
D37	Xiangtao Zhang, Sheng Li, Ao Li, Yipeng Liu, Fan Zhang, Ce Zhu, Le Zhang	电子科技大学	CVPR 2025
	Subspace Constraint and Contribution Estimation for Heterogeneous Federated Learning		
D38	Yulu Bai, Jiahong Fu, Qi Xie, Deyu Meng	西安交通大学	CVPR 2025
	A Regularization-Guided Equivariant Approach for Image Restoration		
D39	Wenxin Su, Song Tang, Xiaofeng Liu, Xiaojing Yi, Mao Ye, Chunxiao Zu, Jiahao Li, Xiatian Zhu	上海理工大学	CVPR 2025
	Domain Adaptive Diabetic Retinopathy Grading with Model Absence and Flowing Data		
D40	Chen Tang, Xinzhu Ma, Encheng Su, Xiufeng Song, Xiaohong Liu, Wei-Hong Li, Lei Bai, Wanli Ouyang, Xiangyu Yue	香港中文大学	CVPR 2025
	UniSTD: Towards Unified Spatio-Temporal Prediction across Diverse Disciplines		
D41	Jie Liu, Tiexin Qin, Hui Liu, Yilei Shi, Lichao Mou, Xiao Xiang Zhu, Shiqi Wang, and Haoliang Li	香港城市大学	CVPR 2025
	Q-PART: Quasi-Periodic Adaptive Regression with Test-time Training for		

	Pediatric Left Ventricular Ejection Fraction Regression		
D42	Yifeng Yang, Lin Zhu, Zewen Sun, Hengyu Liu, Qinying Gu, Nanyang Ye	上海交通大学	CVPR 2025
	OODD: Test-time Out-of-Distribution Detection with Dynamic Dictionary		
D43	Leilei Ma, Shuo Xu, Mingkun Xie, Lei Wang, Dengdi Sun, Haifeng Zhao	安徽大学	CVPR 2025
	Correlative and Discriminative Label Grouping for Multi-Label Visual Prompt Tuning		
D44	Junjie Zhou, Jiao Tang, Yingli Zuo, Peng Wan, Daoqiang Zhang, Wei Shao	南京航空航天大学	CVPR 2025
	Robust Multimodal Survival Prediction with the Latent Differentiation Conditional Variational AutoEncoder		
D45	Rui Wang, Shaocheng Jin, Ziheng Chen, Xiaoqing Luo, Xiaojun Wu	江南大学	CVPR 2025
	Learning to Normalize on the SPD Manifold under Bures-Wasserstein Geometry		
D46	Kai Wang, Mingjia Shi, Yukun Zhou, Zekai Li, Zhihang Yuan, Yuzhang Shang, Xiaojiang Peng, Hanwang Zhang, Yang You	新加坡国立大学	CVPR 2025
	A Closer Look at Time Steps is Worthy of Triple Speed-Up for Diffusion Model Training		
D47	Yushun Tang, Shuoshuo Chen, Zhihe Lu, Xinchao Wang, Zhihai He	南方科技大学	ECCV 2024
	r Domain Shift Correction in Online Test-time Adaptation		
D48	Cong Wu, Xiaojun Wu, Linze Li, Tianyang Xu, Zhenhua Feng, Josef Kittler	江南大学	ECCV 2024
	Efficient Few-Shot Action Recognition via Multi-level Post-reasoning		
D49	Zhongqi Wang, Jie Zhang, Shiguang Shan, Xilin Chen	中科院计算所	ECCV 2024
	T2IShield: Defending Against Backdoors on Text-to-Image Diffusion Models		
D50	Sensen Gao, Xiaojun Jia, Xuhong Ren, Ivor Tsang, Qing Guo	MBZUAI	ECCV 2024
	Boosting Transferability in Vision-Language Attacks via Diversification along the Intersection Region of Adversarial Trajectory		
H01	Zhengyuan Xie, Haiquan Lu, Jiawen Xiao, Enguang Wang, Le Zhang, Xialei Liu	南开大学	ECCV 2024
	Early Preparation Pays Off: New Classifier Pre-tuning for Class Incremental Semantic Segmentation		
H02	Xiang An, Kaicheng Yang, Xiangzi Dai, Ziyong Feng, Jiankang Deng	格灵深瞳 华为诺亚	ECCV 2024
	Multi-label Cluster Discrimination for Visual Representation Learning		
H03	Shiyuan Meng, Wenchao Meng, Qihang Zhou, Shizhong Li, Weiye Hou, Shibo He	浙江大学	ECCV 2024
	MoEAD: A Parameter-efficient Model for Multi-class Anomaly Detection		
H04	Zhaoxin Wang, Handing Wang, Cong Tian, Yaochu Jin	西安电子科技大学	ECCV 2024
	Preventing catastrophic overfitting in fast adversarial training: A bi-level optimization perspective		

H05	Yunfeng Diao, Baiqi Wu, Ruixuan Zhang, Ajian Liu, Xiaoshuai Hao, Xingxing Wei, Meng Wang, He Wang	合肥工业大学	ICLR 2025
	TASAR: Transfer-based Attack on Skeletal Action Recognition		
H06	Hongyu Qu, Jianan Wei, Xiangbo Shu, Wenguan Wang	南京理工大学	ICLR 2025
	Learning Clustering-based Prototypes for Compositional Zero-shot Learning		
H07	Qingyang Zhang, Yatao Bian, Xinke Kong, Peilin Zhao, Changqing Zhang	天津大学	ICLR 2025
	COME: Test-time Adaption by Conservatively Minimizing Entropy		
H08	Song Tang, Wenxin Su, Yan Gan, Mao Ye, Jianwei Dr., Xiatian Zhu	上海理工大学	ICLR 2025
	Proxy Denoising for Source-Free Domain Adaptation		
H09	Qi Zhang, Yifei Wang, Jingyi Cui, Xiang Pan, Qi Lei, Stefanie Jegelka, Yisen Wang	北京大学	ICLR 2025
	Beyond Interpretability: The Gains of Feature Monosemanticity on Model Robustness		
H10	Zihan Ye, Shreyank N Gowda, Shiming Chen, Xiaowei Huang, Haotian Xu, Fahad Shahbaz Khan, Yaochu Jin, Kaizhu Huang, Xiaobo Jin	西交利物浦大学	ICLR 2025
	Zerodiff: Solidified Visual-semantic Correlation in Zero-Shot Learning		
H11	Pengxin Guo, Shuang Zeng, Yanran Wang, Huijie Fan, Feifei Wang, Liangqiong Qu	香港大学	ICLR 2025
	Selective Aggregation for Low-Rank Adaptation in Federated Learning		
H12	Zonghui Guo, Yingjie Liu, Jie Zhang, Haiyong Zheng, Shiguang Shan	中国海洋大学	CVPR 2025
	Face Forgery Video Detection via Temporal Forgery Cue Unraveling		
H13	Chunlei Li, Yilei Shi, Jingliang Hu, Xiaoxiang Zhu, Lichao Mou	脉得智能科技（无锡）有限公司	ICLR 2025
	Scale-Aware Contrastive Reverse Distillation for Unsupervised Medical Anomaly Detection		
H14	Haobin Li, Peng Hu, Qianjun Zhang, Xi Peng, Xiting Liu, Mouxing Yang	四川大学	ICLR 2025
	Test-time Adaptation for Cross-modal Retrieval with Query Shift		
H15	Yi Zhang, Siwei Wang, Jiyuan Liu, Shengju Yu, Zhibin Dong, Suyuan Liu, Xinwang Liu, En Zhu	国防科技大学	ICLR 2025
	DLEFT-MKC: Dynamic Late Fusion Multiple Kernel Clustering with Robust Tensor Learning via Min-Max Optimization		
H16	Yingqi Liu, Qinglun Li, Jie Tan, Yifan Shi, Li Shen, Xiaochun Cao	中山大学	ICLR 2025
	Understanding the Stability-based Generalization of Personalized Federated Learning		
H17	Yunfan Li, Peng Hu, Dezhong Peng, Jiancheng Lv, Jianping Fan, Xi Peng	四川大学	ICML 2025
	Image Clustering with External Guidance		
H18	Yichuan Mo, Hui Huang, Mingjie Li, Ang Li, Yisen Wang	北京大学	ICML 2024

	TERD: A Unified Framework for Safeguarding Diffusion Models Against Backdoors		
H19	Xinjie Yao, Yu Wang, Pengfei Zhu, Wanyu Lin, Jialu Li, Weihao Li, Qinghua Hu	天津大学	ICML 2024
	Socialized learning: making each other better through multi-agent collaboration		
H20	Yu Lu, Yuanzhi Liang, Linchao Zhu, Yi Yang	浙江大学	NeurIPS 2024
	FreeLong: Training-Free Long Video Generation with SpectralBlend Temporal Attention		
H21	Yaohua Liu, Jiaxin Gao, Xuan Liu, Xianghao Jiao, Xin Fan, Risheng Liu	大连理工大学	IJCAI 2024
	Advancing generalized transfer attack with initialization derived bilevel optimization and dynamic sequence truncation		
H22	Zhijie Rao, Jingcai Guo, Xiaocheng Lu, Jingming Liang, Jie Zhang, Haozhao Wang, Kang Wei, Xiaofeng Cao	香港理工大学	IJCAI 2024
	Dual Expert Distillation Network for Generalized Zero-Shot Learning		
H23	Yiyang Wang, Yuchen Han, Yuhan Guo	大连海事大学	IJCAI 2024
	Self-adaptive Extreme Penalized Loss for Imbalanced Time Series Prediction		
H24	Jianqing Zhang, Yang Liu, Yang Hua, Hao Wang, Tao Song, Zhengui Xue, Ruhui Ma, Jian Cao	上海交通大学	JMLR 2025
	PFLlib: A Beginner-Friendly and Comprehensive Personalized Federated Learning Library and Benchmark		
H25	Peiliang Gong, Mohamed Ragab, Min Wu, Zhenghua Chen, Yongyi Su, Xiaoli Li, Daoqiang Zhang	南京航空航天大学	KDD 2025
	Augmented Contrastive Clustering with Uncertainty-Aware Prototyping for Time Series Test Time Adaptation		
H26	Ming Hu, Zhihao Yue, Xiaofei Xie, Cheng Chen, Yihao Huang, Xian Wei, Xiang Lian, Yang Liu, Mingsong Chen	Singapore Management University	KDD 2024
	Is aggregation the only choice? federated learning via layer-wise model recombination		
H27	Huijun Xing, Rui Sun, Jinke Ren, Jun Wei, Chunmei Feng, Xuan Ding, Zilu Guo, Yu Wang, Yudong Hu, Wei Wei, Xiaohua Ban, Chuanlong Xie, Yu Tan, Xian Liu, Shuguang Cui, Xiaohui Duan, Zhen Li	香港中文大学（深圳）	Nature Communications 2025
	Achieving Flexible Fairness Metrics in Federated Medical Imaging		
H28	Yiming Zhong, Qi Jiang, Jingyi Yu, Yuexin Ma	上海科技大学	CVPR 2025
	DexGrasp Anything: Towards Universal Robotic Dexterous Grasping with Physics Awareness		
H29	Junshu Sun, Chenxue Yang, Xiangyang Ji, Qingming Huang, Shuhui Wang	中国科学院计算技术研究所	NeurIPS 2024
	Towards Dynamic Message Passing on Graphs		
H30	Yan Fan, Yu Wang, Pengfei Zhu, Dongyue	天津大学	NeurIPS

	Chen, Qinghua Hu		2024
	Persistence Homology Distillation for Semi-supervised Continual Learning		
H31	Yanxin Yang, Chentao Jia, Dengke Yan, Ming Hu, Tianlin Li, Xiaofei Xie, Xian Wei, Mingsong Chen	华东师范大学	NeurIPS 2024
	SampDetox: Black-box Backdoor Defense via Perturbation-based Sample Detoxification		
H32	Wenjun Miao, Guansong Pang, Jin Zheng, Xiao Bai	北京航空航天大学	NeurIPS 2024
	Long-Tailed Out-of-Distribution Detection via Normalized Outlier Distribution Adaptation		
H33	Xingkui Zhu, Yiran Guan, Dingkang Liang, Yuchao Chen, Yuliang Liu, Xiang Bai	华中科技大学	NeurIPS 2024
	MoE Jetpack: From Dense Checkpoints to Adaptive Mixture of Experts for Vision Tasks		
H34	Xinyu Fang, Kangrui Mao, Haodong Duan, Xiangyu Zhao, Yining Li, Dahua Lin, Kai Chen	浙江大学 & 上海人工智能实验室	NeurIPS 2024
	MMBench-Video: A Long-Form Multi-Shot Benchmark for Holistic Video Understanding		
H35	Yunpeng Gong, Zhun Zhong, Yansong Qu, Zhiming Luo, Rongrong Ji, Min Jiang	厦门大学	NeurIPS 2024
	Cross-modality perturbation synergy attack for person re-identification		
H36	Qiufeng Wang, Xu Yang, Fu Feng, Jing Wang, Xin Geng	东南大学	NeurIPS 2024
	Cluster-Learngene: Inheriting Adaptive Clusters for Self-Attention		
H37	Chenyu Huang, Peng Ye, Tao Chen, Tong He, Xiangyu Yue, Wanli Ouyang	复旦大学	NeurIPS 2024
	EMR-Merging: Tuning-Free High-Performance Model Merging		
H38	Xinting Liao, Weiming Liu, Pengyang Zhou, Fengyuan Yu, Jiahe Xu, Jun Wang, Wenjie Wang, Chaochao Chen, Xiaolin Zheng	浙江大学	NeurIPS 2024
	FOOGD: Federated Collaboration for Both Out-of-distribution Generalization and Detection		
H39	Honglin Liu, Peng Hu, Changqing Zhang, Yunfan Li, Xi Peng	四川大学	NeurIPS 2024
	Interactive Deep Clustering via Value Mining		
H40	Yanyan Huang, Weiqin Zhao, Yihang Chen, Yu Fu, Lequan Yu	香港大学	NeurIPS 2024
	Free Lunch in Pathology Foundation Model: Task-specific Model Adaptation with Concept-Guided Feature Enhancement		
H41	Houlun Chen, Xin Wang, Hong Chen, Zeyang Zhang, Wei Feng, Bin Huang, Jia Jia, Wenwu Zhu	清华大学	NeurIPS 2024
	VERIFIED: A Video Corpus Moment Retrieval Benchmark for Fine-Grained Video Understanding		
H42	Bin-Bin Gao	腾讯优图	NeurIPS 2024

	MetaUAS: Universal Anomaly Segmentation with One-Prompt Meta-Learning		
H43	Xincheng Yao, Zixin Chen, Chao Gao, Guangtao Zhai, Chongyang Zhang	上海交通大学	NeurIPS 2024
	ResAD: A Simple Framework for Class Generalizable Anomaly Detection		
H44	Xiaoqian Liu, Peng-Fei Zhang, Xin Luo, Zi Huang, Xinshun Xu	山东大学	TIP 2024
	Noisy-Aware Unsupervised Domain Adaptation for Scene Text Recognition		
H45	Yanran Zhu, Xiao He, Chang Tang, Xinwang Liu, Yuanyuan Liu, Kunlun He	中国地质大学（武汉）	TKDE 2024
	Multi-View Adaptive Fusion Network for Spatially Resolved Transcriptomics Data Clustering		
H46	Yingxi Li, Xiaowei Bai, Liang Xie, Xiaodong Wang, Feng Lu, Feitian Zhang, Ye Yan, Erwei Yin	北京大学 军事科学院国防科技创新研究院 天津（滨海）人工智能创新中心	TMC 2024
	Real-time Gaze Tracking via Head-eye Cues on Head Mounted Devices		
H47	Dingwei Fan, Junyong Zhao, Chunlin Li, Xinlong Wang, Ronghan Zhang, Qi Zhu, Mingliang Wang, Haipeng Si, Daoqiang Zhang, Liang Sun	南京航空航天大学	TMI 2024
	MA-SAM: A Multi-atlas Guided SAM Using Pseudo Mask Prompts without Manual Annotation for Spine Image Segmentation		
H48	Shulan Ruan, Huijie Liu, Zhao Chen, Bin Feng, Kun Zhang, Caleb Chen Cao, Enhong Chen, Lei Chen	清华大学	TOIS 2025
	CPWS: Confident Programmatic Weak Supervision for High-Quality Data Labeling		
H49	Peirong Zhang, Yuliang Liu, Songxuan Lai, Hongliang Li, Lianwen Jin	华南理工大学	TPAMI 2025
	Privacy-Preserving Biometric Verification with Handwritten Random Digit String		
H50	Man Yao, Xuerui Qiu, Tianxiang Hu, Jiakui Hu, Yuhong Chou, Keyu Tian, Luzimo Leng, Bo Xu, Guoqi Li	中国科学院自动化研究所	TPAMI 2025
	Scaling Spike-driven Transformer with Efficient Spike Firing Approximation Training		

合作单位简介

铂金合作单位

AutoDL

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 2w 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。上线 3 年多来，根据开票数据统计，AutoDL 累计为超过 70w 开发者，142 所 985、211、双一流高校，540 多所其他高校，以及 3000 余家企业提供了弹性、省钱、好用的普惠 AI 云算力服务。

铂金合作单位

华为

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

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

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

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

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

金牌合作单位

腾讯优图实验室

优图实验室是腾讯在人工智能领域的重要布局，于 2012 年成立，先后荣获上海市科技进步奖一等奖、福建省科技进步一等奖等殊荣。腾讯优图实验室专注于人工智能领域的前沿技术研究和产业实践落地，涵盖了视觉感知理解、多模态理解、大语言模型、AIGC、计算加速等技术研究领域；腾讯优图实验室在多个领域开展技术研究和应用落地，同时为腾讯内外部超 500+产品提供技术支持。此外，优图实验室汇聚来自国内外的顶尖人才，并与世界顶级院校和机构合作，拥有超过 1800 项全球专利，1000 余篇论文被国际顶会收录。

金牌合作单位

上海合合信息科技股份有限公司

合合信息成立于 2006 年，是一家人工智能及大数据科技企业，基于自主研发的领先的智能文字识别及商业大数据核心技术，为全球 C 端用户和多元行业 B 端客户提供数字化、智能化的产品及服务。

在 C 端产品方面，截至 2023 年 12 月底，公司扫描全能王、名片全能王、 启信宝 3 款 APP 在 App Store 与 Google Play 应用市场的全球用户累计首次下载量合计超过 9.4 亿；截至 2023 年 12 月的月活合计约 1.5 亿；

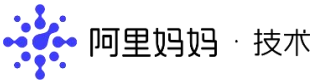
在 B 端业务方面，公司智能文字识别与商业大数据服务已覆盖了银行、证券、保险、政府、物流、制造、地产、零售等近 30 个行业的众多头部客户。《财富》杂志 2023 年发布的世界 500 强公司名单中，公司客户已覆盖超过 130 家。

公司业务具有三大独特性：

（1）是行业内少有的在人工智能、大数据两个领域均具有行业领先的核心自主研发技术的科技企业；

（2）是行业内少有的在 C 端产品与 B 端服务同时拥有完善布局矩阵的企业；

（3）是行业内少有的在国内、国际市场同时布局且均取得了规模化用户和产值的企业。



金牌合作单位

阿里妈妈

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

阿里妈妈作为全球优秀的品牌数智经营阵地，是淘宝和天猫商家优选的消费者投资平台。通过阿里妈妈 AI 技术与智能投放产品，能够帮助商家获取全域数智经营解决方案，让消费者经营与货品运营的每一个环节可视化，让每一份经营都算数。旗下淘宝联盟是国内电子商务效果营销联盟的佼佼者，以淘宝客的私域经营为中心，连接站外私域场景和商家电商场景。

金牌合作单位

美团

美团是一家科技零售公司。美团以“零售+科技”的战略践行“帮大家吃得更好，生活更好”的公司使命。

自 2010 年 3 月成立以来，美团持续推动服务零售和商品零售在需求侧和供给侧的数字化升级，和广大合作伙伴一起努力为消费者提供品质服务。

美团始终以客户为中心，不断加大在新技术上的研发投入。美团会和大家一起努力，更好承担社会责任，更多创造社会价值。

美团科研合作：

美团科研合作致力于搭建美团技术团队与学术界合作的桥梁，让学术前沿落地应用，让真实场景支撑研究。

至今，我们已与数十所知名高校及科研机构的学者，围绕人工智能、机器人、自动驾驶、运筹优化、大数据、信息基础设施等领域开展了百余项课题合作；并在相关领域国际会议、期刊发表数百篇论文，在国际顶级赛事多次获得冠亚军，部分合作成果已在美团业务场景中落地。

深圳市美团机器人研究院：

深圳市美团机器人研究院于 2022 年 7 月在深圳市民政局支持下正式注册成立。研究院依托美团生活服务丰富场景与数据积累，结合深圳及大湾区的科研优势，开展面向机器人共性关键技术的研发，引领机器人学科前沿和技术创新方向，加快科研成果的落地转化，在大湾区打造机器人技术“政产学研用”全方位结合的开放协同创新平台。

依托研究院，美团展开了一系列的科研合作活动，如您希望联系我们，请发邮件至：meituan.oi@meituan.com。

金牌合作单位

百度

百度是拥有强大互联网基础的领先 AI 公司，以“用科技让复杂的世界更简单”为使命，坚持技术创新，致力于“成为最懂用户，并能帮助人们成长的全球顶级高科技公司”。百度公司 2000 年 1 月 1 日创立于中关村，创始人李彦宏拥有“超链分析”技术专利，使中国成为美国、俄罗斯、和韩国之外，全球仅有的 4 个拥有搜索引擎核心技术的国家之一。百度每天响应来自 100 余个国家和地区的数十亿次搜索请求，是网民获取中文信息和服务的最主要入口，服务 10 亿互联网用户。

多年来，百度在深度学习、自然语言处理、语音、视觉等人工智能技术方面持续积累，大量技术成果应用于搜索引擎等产品。近 10 年百度在深度学习、大模型、自动驾驶、AI 芯片等领域持续投入，成为全球为数不多在“芯片-框架-模型-应用”四层全栈布局的人工智能公司。从高端芯片昆仑芯，到飞桨深度学习框架，再到文心预训练大模型，到搜索、智能云、自动驾驶、小度等应用，各个层面都有领先业界的自研技术。层层领先，可以实现端到端优化，大幅提升效率。

在芯片层，百度自主研发的昆仑芯 1 代 AI 芯片在 2018 年推出，已在百度搜索引擎、小度等业务中部署数万片，并服务数十家外部客户。昆仑芯 2 代 AI 芯片也于 2021 年 8 月正式发布并量产，性能比昆仑芯 1 代提升 2-3 倍，目前已经在金融、工业、交通、教育等客户的业务中被广泛部署和使用。

在框架层，飞桨是中国首个自主研发、功能丰富、开源开放的产业级深度学习平台。飞桨位列中国深度学习市场应用规模第一。截至 2025 年 4 月，飞桨文心开发者数量已达 2185 万，服务了 67 万家企业，创建了 110 万个模型。

在模型层，文心大模型是百度自主研发的产业级知识增强大模型体系，具备知识增强和产业级两大特色。目前，文心大模型已大规模应用于搜索、信息流、智能音箱等互联网产品，并通过飞桨深度学习开源开放平台、百度智能云赋能工业、能源、金融、通信、媒体、教育等各行各业。

2025 年 3 月 16 日，文心大模型 4.5 和文心大模型 X1 正式推出。文心大模型 4.5 是百度自主研发的新一代原生多模态基础大模型，通过多个模态联合建模实现协同优化，多模态理解能力优秀；具备更精进的语言能力，理解、生成、逻辑、记忆能力全面提升，去幻觉、逻辑推理、代码能力显著提升。文心大模型 X1 具备更强的理解、规划、反思、进化能力，并支持多模态，是首个自主运用工具的深度思考模型。2025 年 4 月 25 日，百度正式发布文心大模型 4.5 Turbo 和文心大模型 X1 Turbo，具备多模态、强推理、低成本三大特性。

2023 年 3 月 16 日，新一代知识增强大语言模型文心一言正式开启邀请测试，在全球大厂中拔得头筹。2023 年 8 月 31 日，文心一言率先向全社会全面开放。截至 2024 年 11 月，文心一言累计用户规模已达 4.3 亿。

无问芯穹

金牌合作单位

无问芯穹

无问芯穹（Infinigence AI）作为国际领先的 AI 基础设施企业，致力于成为大模型时代首选的算力运营商。依托“多元异构、软硬协同”的核心技术优势，打造了连接“M 种模型”和“N 种芯片”的“M×N”AI 基础设施新范式，实现多种大模型算法在多元芯片上的高效协同部署。无问芯穹 Infini-AI 异构云平台基于多元芯片算力底座，向大模型开发者提供极致性价比的高性能算力和原生工具链，为大模型从开发到部署的全生命流程降本增效。

无问芯穹以“释放无穹算力，让 AGI 触手可及”为使命，通过不断的技术创新实现普惠 AI，让算力成本实现万倍下降，如同水电煤一般为千行百业注入新质生产力。

oppo

金牌合作单位

OPPO

OPPO 是提供移动智能科技产品与服务的全球品牌。我们打造拥有易简设计与智慧流畅体验的产品和服务，让每一个人都能轻松灵动地创作表达，享受创造的乐趣。

GALBOT
银 河 通 用

金牌合作单位

银河通用

北京银河通用机器人有限公司成立于 2023 年 5 月，是一家专注于通用具身多模态大模型机器人研发的创新企业，致力于为全球用户提供智能机器人产品，在商业、家庭等环境中为广泛应用，服务千行百业、千家万户。

银河通用现设有北京、深圳和苏州三地研发中心，深度融合产、学、研等各界力量，与北京大学、北京智源人工智能研究院分别成立了具身智能联合实验室、研究中心，不断推动解决产业界技术瓶颈，并持续为具身智能领域的创新发展提供方向建议。公司凝聚海内外一众顶尖人才，现有算法、软件、硬件研发团队 100+人，正迅速地发展壮大。创始团队成员在具身大模型领域拥有百篇国际前沿学术论文以及十余年的资深人工智能领域从业经历，兼具全球领先的具身智能研发实力和千万级智能硬件产品量产经验，具备强大的商业化“实战”能力。

阿里云

金牌合作单位

阿里云计算有限公司

全球领先的云计算及人工智能科技公司

阿里云创立于 2009 年，是全球领先的云计算及人工智能科技公司。基于自研的飞天云计算操作系统，阿里云向全球客户提供基于基础设施即服务（IaaS）、平台即服务（PaaS）和模型即服务（MaaS）三层架构的全方位云服务。目前，阿里云是亚太第一、中国最大的公共云服务提供商。

阿里巴巴自研大模型通义千问是全球领先的大模型之一，目前已开源多个尺寸的系列模型，以支持更多企业客户实现 AI 创新。

金牌合作单位

小米

小米集团成立于 2010 年 4 月，2018 年 7 月 9 日在香港交易所主板挂牌上市(1810.HK)，是一家以智能手机、智能硬件和 IoT 平台为核心的消费电子及智能制造公司。

胸怀“和用户交朋友，做用户心中最酷的公司”的愿景，小米致力于持续创新，不断追求极致的产品服务体验和公司运营效率，努力践行“始终坚持做感动人心、价格厚道的好产品，让全球每个人都能享受科技带来的美好生活”的公司使命。

小米是全球领先的智能手机品牌之一，智能手机出货量稳居全球前三。截至 2024 年 12 月，全球月活跃用户 7.02 亿。同时，小米已经建立起全球领先的消费级 AIoT（人工智能和物联网）平台，截至 2024 年 12 月 31 日，小米 AIoT 平台已连接的 IoT 设备（不包括智能手机、笔记本电脑及平板）数达到 9.05 亿。集团业务已进入全球逾 100 个国家和地区。2024 年 8 月，小米集团连续六年进入《财富》“世界 500 强排行榜”(Fortune Global 500)。

小米集团目前为恒生指数、恒生中国企业指数、恒生科技指数及恒生神州 50 指数成份股。

Miromind

金牌合作单位

MiroMind

集智进化(Miromind)是有长期稳定资本支持的新型 AGI 研发组织,致力于成为人工智能创新领域的全球领军者,专注于基础模型以及下一代智能关键技术的前沿探索，实现人工智能和人类智能的结合。

九章云极 DataCanvas

金牌合作单位

北京九章云极科技股份有限公司

九章云极（DataCanvas）是领先的人工智能基础设施及智算云提供商，自主研发构建了完整的自研 AIDC 技术栈集、智算云操作系统、智算产业链，面向 AI 训练和推理提供高性能计算、云服务和人工智能软件。

旗下囊括九章智算云 AIaya NeW Cloud、九章智算操作系统 AIaya NeW OS、算力包、九章开源社区等业界知名品牌，云平台汇聚百万 AI 开发者，超 4000P 智能算力储备，客户数量上千家，凭借自主创新开创性定义算力单位，实现算力普惠。

趋动云 VirtAI Cloud 连接算力·连接人

银牌合作单位

趋动科技（上海）有限公司

趋动科技作为软件定义 AI 算力技术的领导厂商，专注于为全球用户提供国际领先的数据中心级 AI 算力池化和虚拟化软件及解决方案，已完成中关村高新、国高新、“专精特新”等企业认证。趋动科技的 OrionX AI 算力池化软件能够帮助用户提高资源利用率和降低 TCO，提高算法工程师的工作效率。趋动科技的双子座 GEMINI AI 训练平台，为客户提供强大的 AI 算力管理服务以及高效的算法开发和训练支持，能够化繁为简，帮助企业建好 AI 平台、管好 GPU、用好 AI 服务。依托全球领先的 AI 算力池化技术，趋动科技重磅推出趋动云 VirtAI Cloud，为万千企业和 AI 开发者带来又便宜、又好用的 AI 算力池化云服务。

极市 EXTREME MART

银牌合作单位

极视角科技

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

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

ZhenFund 真格基金

银牌合作单位

真格基金

真格基金创立于 2011 年，是国内最早的天使投资机构之一。自创立伊始，真格基金一直积极在人工智能、芯片与半导体、机器人与硬件、医疗健康、企业服务、新能源、跨境出海、消费生活等领域寻找最优秀的创业团队和引领时代的投资机会。

真格基金从早期陪伴了小红书、Manus、月之暗面、Nuro、Momenta、出门问问、晶泰科技、地平线、云天励飞、禾赛科技、亿航智能、格灵深瞳、逸仙电商等团队一路成长。

自 2014 年清科「中国股权投资年度排名」设立早期投资机构排名以来，真格基金已连续 9 年获得「中国早期投资机构 30 强」TOP3。真格基金创始人徐小平从 2016 年起连续 5 年入选福布斯「全球最佳创投人榜单（The Midas List）」，在 2019 年榜单中排名第 11 位。真格基金创始合伙人兼 CEO 方爱之从 2019 年起连续 6 年上榜「全球最佳创投人」榜单，并在 2022 年排名第 12 位，在女性投资人中位列第 1。同年，方爱之登顶福布斯「全球最佳天使创投人榜单（The Midas Seed List）」。

真格基金总部位于北京，并陆续布局上海和深圳。我们相信创新，相信冒险，相信好的创业者稀少而珍贵。十几年来，我们一直坚定地与创业者站在一起，推动他们缔造引领科技创新并改变世界的伟大公司。

金山办公 KINGSOFT OFFICE

银牌合作单位

金山办公

金山办公（688111.SH）是国内领先的办公软件产品和服务提供商，于 2019 年在上海证券交易所上市，是中国“硬科技”的代表性企业。

秉持“技术立业”和“用户第一”理念，金山办公为来自全球 220 多个国家和地区的用户提供办公服务，旗下主要产品包括 WPS Office、WPS 365 等。

思腾合力 SITONHOLY 加速AI梦想

银牌合作单位

思腾合力

思腾合力（天津）科技有限公司，自 2009 年成立以来，聚焦高性能计算及算力底座解决方案。我们紧跟 AI 发展趋势，专注于为 AI 应用提供高效算力支持，致力于成为卓越的算力解决方案引领者，构筑 AI 世界的无限可能。思腾建立了完善的研发、生产、制造基地，通过自主创新的软硬件结合，打造了完整且自主可控的产品生态系统，涵盖深度学习、高性能计算、虚拟化、分布式存储、AI 集群管理及云计算等多个领域。

我们坚持客户至上，核心产品线包括自研 AI 管理软件、算力服务器（GPU、信创）及云计算服务（裸金属、公有云）。同时，我们还拓展了算力运营、维保、垂直大模型一体机、定制化及液冷产品等五项业务，以满足客户多样化需求。

展望未来，思腾合力将继续秉持创新精神，紧跟 AI 发展步伐，提升技术实力和服务水平，为构建更加智能、美好的未来贡献 AI 力量。

银牌合作单位

美图

美图公司成立于 2008 年，是一家以美为内核，以人工智能为驱动的科技公司。秉承着“让艺术与科技美好交汇”的使命，美图公司致力于打造优秀的影像与设计产品，让图像、视频、设计的制作变得更简单，并通过美业解决方案助力产业数字化升级。美图公司于 2016 年 12 月在香港联合交易所主板挂牌上市，股票代码：1357.HK。

自 2010 年美图影像研究院（MT Lab）成立起，美图公司持续进行 AI 领域的探索与布局。作为美图公司技术中枢，美图影像研究院（MT Lab）在 CVPR、ECCV、ICCV、AAAI、ACM MM、TPAMI、NIPS 等国际顶级会议及期刊上发表学术论文超 50 篇，并在多项国际人工智能竞赛上斩获冠军。

近年来，美图公司持续推动 AI 应用落地。旗下生活场景产品以美图秀秀、美颜相机、Wink 为代表，连续斩获全球多个国家和地区的应用榜单冠军。

同时，通过美图设计室、开拍、美图云修、WHEE、MOKI 等产品逐步深入生产力场景，并于 2024 年 3 月完成对中国领先的视觉创意平台站酷收购。

美图奇想大模型（MiracleVision）于 2024 年 1 月通过《生成式人工智能服务管理暂行办法》备案后，持续升级图像生成与视频生成能力，全面应用于美图旗下影像与设计产品，并助力电商、广告、游戏、影视、动漫五大行业的工作流提效。在自研视觉大模型基础上，美图公司积极部署调用优质模型能力，形成优势互补效应，从而持续提升产品能力和盈利能力。

银牌合作单位

爱诗科技

•人才：爱诗团队成员来自清华、北大、中科院等顶级学府，曾任职于字节、微软亚洲研究院、快手、腾讯等头部机构的核心技术团队，拥有世界一流的计算机视觉算法攻坚能力和解决系统工程问题的经验。

• 产品：爱诗是国内最早发布公开可用的视频生成产品的中国企业，旗下产品 PixVerse 截至目前已经在世界范围内积累了六千万用户、单月月活超 1600 万，国内版产品拍我 AI 也于 2025 年 5 月 App 及 Web 双端上线。

• 发展规划：未来爱诗也将持续探索 AI 视频领域的具体应用场景，不仅只在生产端为内容生产者提供好用的工具，更将积极探索在内容消费端的可能性，打造 AI 原生的视频生产和消费平台。

银牌合作单位

GpuGeek

首都在线（证券代码：300846）成立于 2005 年，总部位于北京，在美国、香港、新加坡、上海、广州、海南等地拥有 18 家分支机构，业务范围遍及 50 多个国家，在国内以及海外三大核心地域美洲、欧洲、亚太设有 24 个区域、52 个可用区、 94 个数据中心、上千个边缘算力节点，只需 5 分钟即可完成全球业务的多点部署，众多互联网百强企业都在使用首都在线的产品和服务。

首都在线致力于成为一家覆盖全球的云计算及智算服务的综合服务商，面向全球客户提供优质的云计算、大数据、人工智能等技术产品与服务，打造贴近客户业务场景的行业解决方案，以云服务赋能数字经济，成为与客户共同应对变化的伙伴。2023 年，首都在线全面向智算转型，正式发布“一体两翼”发展战略，凭借全球互联、异构池化、统一调度等核心能力，依托智能高效的云端算力服务，以“云+智+网”一体化的发展路径赋能千行百业。

银牌合作单位

将门

将门是一家以技术创新为驱动的新型创投机构，旗下设有将门投资基金、将门大企业服务以及将门 TechBeat 技术社区三大服务。

将门投资基金聚焦智能科技领域早期投资，截至目前，基金已累计投资了近百家技术驱动型创业公司，在人工智能赛道形成显著生态优势，投资项目包括智谱 AI、文远知行、格灵深瞳、暗物智能、杉数科技、禾赛科技、迪英加科技等标杆企业。

将门大企业服务致力于为全球行业龙头构建创新战略落地体系，依托对前沿科技趋势的深度洞察与企业战略需求的精准匹配，提供「技术发掘 — 场景验证 — 规模化落地」全链条服务。目前已携手博世、宝马、保时捷、联合利华、巴黎欧莱雅、本田、资生堂、新世界集团等世界 500 强及中国领军企业，推动创新技术与真实业务场景的深度融合。

将门 TechBeat 技术社区专注于打造全球智能科技人才的深度交互平台，涵盖了计算机视觉、自然语言处理、机器学习、机器人等前沿领域，目前已汇聚超 40,000 名横跨学术界与产业界的学者、技术专家。通过高频技术分享、跨界场景研讨与产学研资源链接，构建从基础研究到产业落地的全链条成长支持。

将守四方，门聚豪杰。将门期待着，和全球优秀的科技创新者与创业者一起，打开通往将来之门！

银牌合作单位

易方达基金管理有限公司

易方达基金是一家综合性资产管理公司。公司成立于 2001 年，依托资本市场，通过专业化运作，为境内外各类投资者提供综合性资产管理解决方案，实现长期可持续的投资回报。公司坚守客户利益至上原则，坚持做值得各方长期托付的、负责任的公司。

经过 20 多年的发展，易方达基金在坚持稳健、规范的基础上，不断锐意进取、开拓创新，目前拥有包括公募、社保、基本养老保险、年金、特定客户资产管理、QDII、投资顾问等在内的多类业务资格，在主动权益、指数投资、债券、多资产、另类资产等投资领域全面布局。截至 2025 年一季度末，易方达基金及下属机构资产管理规模超 3.5 万亿元，客户包括个人投资者及社保基金、企业年金和职业年金、银行、保险公司、境外央行及养老金等机构投资者。

公司以“发现价值、创造未来”为使命，坚持规范、稳健、开放的经营理念，坚持“深度研究驱动、时间沉淀价值”的投资理念，不断提高服务实体经济与国家战略、服务资本市场改革发展、服务居民财富管理需求的能力努力打造一流投资机构，以自身高质量发展服务经济社会高质量发展。

银牌合作单位

思谋科技

思谋科技（SmartMore），智能制造的持续创新者，基于先进视觉技术，以深度学习和机器视觉作为技术引擎，应用领先的光学、机械、工业自动化、大模型等技术，助力加快发展新质生产力，扎实推进高质量发展。 经过二十多年的技术积累与沉淀，结合独特的商业思考，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，以大模型和工业 AI-AOI 为核心，推动探索工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，服务了卡尔蔡司、空客、博世、佳能、大陆集团、舍弗勒等来自全球超过 200 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。

思谋由计算机视觉国际顶尖专家创立，自成立以来一直坚持国际化发展路线，成为了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

银牌合作单位

摩尔线程

摩尔线程成立于 2020 年 10 月，以全功能 GPU 为核心，致力于向全球提供加速计算的基础设施和一站式解决方案，为各行各业的数智化转型提供强大的 AI 计算支持。

我们的目标是成为具备国际竞争力的 GPU 领军企业，为融合人工智能和数字孪生的数智世界打造先进的加速计算平台。我们的愿景是为美好世界加速。

银牌合作单位

深圳元戎启行科技有限公司

作为国际领先的人工智能企业，元戎启行致力于打造“物理世界的通用人工智能”，以创新技术打造 AI 司机，实现 RoadAGI，引领人工智能行业变革。

元戎启行由 CEO 周光博士带领团队于 2019 年创立，总部位于深圳，在北京设有研发中心，在硅谷、德国、新加坡、上海、成都、重庆、保定等地均有业务落地。元戎启行已完成 6 轮融资，累计融资金额超 5 亿美元。2024 年 11 月，元戎启行完成由国内头部主机厂独家投资的 C1 轮融资。

元戎启行汇聚了全球顶尖学府和科研机构的专业人才。目前，元戎启行集团总人数超 1000 人，研发人员占比达 84%。凭借前瞻性技术布局，元戎启行获评国家高新技术企业，荣登 2023 年科创中国“新锐企业榜”，获 2023 年度深圳市科技进步奖一等奖，入选胡润百富榜 2023 年度人工智能独角兽企业等。

元戎启行始终坚持自主创新，成功推出不依赖高精度地图、应用端到端模型的智能驾驶平台 DeepRoute IO，以及新一代 VLA 模型(视觉-语言-动作模型)。凭借富有竞争力的产品和服务，元戎启行已与多家车企达成量产合作，共同推进十余款组合辅助驾驶汽车落地。预计到 2025 年，将有超 20 万辆搭载元戎启行组合辅助驾驶方案的车辆进入消费者市场。元戎启行相信智能驾驶技术将为通用人工智能的实现带来全新契机，开启人类智能发展的新篇章。

银牌合作单位

深圳云天励飞技术股份有限公司

深圳云天励飞技术股份有限公司（688343.SH）成立于 2014 年，致力于为人工智能的规模化应用提供坚实的国产算力底座，是 AI 推理芯片的领军企业。基于自主研发的神经网络处理器芯片平台和多模态大模型平台，云天励飞打造了面向消费者、企业和行业的各类 AI 推理计算产品和解决方案，推动 AI 深入千行百业，加速推理 AI 时代的发展，让智能无处不在。

银牌合作单位

数据堂（北京）科技股份有限公司

数据堂（股票代码：831428）成立于 2010 年，是全球知名的人工智能数据服务企业，致力于为人工智能及大数据领域公司提供训练数据集、数据采集与标注定制服务、标注平台部署等一体化数据解决方案。

公司通过构建“场景化数据工厂”，在全球部署专业采集基地，积累了涵盖千余种细分场景的专业数据集，覆盖文本、语音、图像、视频、3D 点云等多种数据资源。依托全球化采集团队与自研 AI 标注工具，实现数据采集与智能标注一体化服务，涵盖 45 套成熟的标注工具，可满足自动驾驶、医疗影像、金融智能语音等多场景需求。为提升企业数据生产效率，其自主研发的智能标注平台支持私有化部署，显著降低企业数据处理成本。同时公司建立数据脱敏、加密及权限管控体系，保障数据合法合规流通。

凭借高质量数据服务体系，数据堂已帮助全球上千家企业提升 AI 模型性能。未来，数据堂将继续专注于人工智能数据服务，推动人工智能技术、应用和产业的创新，赋能全球人工智能产业高效、安全、可持续发展。

银牌合作单位

英博数科

北京英博数科科技有限公司(简称“英博数科”)是鸿博股份有限公司(股票代码:002229)于 2022 年 6 月设立的全资子公司，是业务包括：智算中心建设运维、GPU 容器服务、先进算力实验室和产业孵化器等，助力企业加速 AI 技术研发和业务发展。

旗下产品英博云专注提供高效益、多样化的 GPU 智算服务，支持弹性 K8s 集群、IB 高速网络与全闪存储，提供训练、推理、模型优化全流程技术支持及私有可用区、MLOps 定制服务，满足企业多元智算需求。

英博数科以四大优势赋能未来科技：

- 稳定算力供应链保障
- 大集群建设运维专长
- 多维度性能评测优化
- 优秀团队的智慧基因

为 AIGC、高校科研、企业服务等众多领域提供全方位智算服务，推动“AI+产业”深度融合。

UNITREE 宇树

银牌合作单位

杭州宇树科技股份有限公司

杭州宇树科技是一家世界知名的机器人公司，专注于消费级、行业级高性能通用足式/人形机器人及灵巧机械臂的自主研发、生产和销售，曾受邀参加 2021 牛年央视春晚、2022 年冬奥会开幕式、2023 Super Bowl 赛前表演、2023 年杭州亚运会和亚残运会以及 2025 蛇年央视春晚等，并多次受到央视新闻联播等权威媒体报道，在机器人核心零部件、运动控制、机器人感知决策、机器人 AI 等综合领域持续深耕，引领行业。

宇树是全球首家公开零售高性能四足机器人并最早实现行业落地的公司，研发和生产均位于杭州高新区（滨江）。其四足机器人销量占全球出货量的 60-70%，大尺寸通用人形机器人出货量全球领先，业务范围覆盖全球 50%以上的国家和地区。

自 2016 年起，宇树致力于消费级、行业级高性能通用足式机器人的自主研发，坚持软、硬件自主创新，在足式机器人领域达到全球技术领先。目前累计提交国内外专利申请 200 余项，其中授权专利 180 余项。

部分组织单位简介

主办单位：中国图象图形学学会

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

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

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

承办单位：中山大学

中山大学由孙中山先生于 1924 年亲手创办，是现代综合性大学 ，中央直管高校，国家“双一流”建设高校，入选“985 工程”和“211 工程”。中山大学组建了人文学部、社会科学学部、经济与管理学部、理学部、工学部、信息学部、医学部等 7 个学部，统筹相同和相近学科的学术标准与学术评价，加强跨学院（系）的合作协同，形成高质量发展合力。中山大学现有国家级科研创新平台 42 个（包括国家超级计算广州中心、天琴中心、南方海洋科学与工程广东省实验室（珠海）、“中山大学”号海洋综合科考实习船和“中山大学极地”号破冰科考船）、省部级平台 268 个，学校着力推进理、工、医科重大科研平台建设。

VALSE——学术华尔兹

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

截至目前，VALSE已成功举办14届，分别为VALSE2011（杭州），VALSE2012（西安），VALSE2013（南京），VALSE2014（青岛），VALSE2015（成都），VALSE2016（武汉），VALSE2017（厦门），VALSE2018（大连），VALSE2019（合肥），VALSE2020（线上），VALSE2021（杭州），VALSE2022（天津），VALSE 2023（无锡），VALSE2024（重庆）。VALSE2025将于2025年6月6-8日在珠海举行，由中国图象图形学学会主办，中山大学承办，中国图象图形学学会青年工作委员会、AutoDL协办。

除大会演讲之外，VALSE年度大会还不断推陈出新，逐渐增加了 Poster/Spotlight、Tutorial、年度进展评述(APR)、Workshops、重要学术进展和工业界技术分享等环节，参会人数也逐渐增长到了5000人以上。在此过程中，逐渐形成了由山世光、潘纲、刘青山、颜水成、李学龙、周志华、徐东、马毅、周昆、高新波、何晓飞、余凯、杨健、黄华、白翔等15名青年学者组成的指导委员会。2018年，VALSE确立了指导委员会委员45岁退休的原则，故高新波、马毅、杨健、周志华、汤进、吴小俊六位老师进入顾问委员会，同时吸纳了华刚、汪萌、虞晶怡和张敏灵四位老师进入指导委员会。此外，VALSE大会也吸引了越来越多的企业参加，已成为计算机视觉与机器学习领域“产-学-研”合作交流的重要平台。

为配合VALSE系列年度研讨会，VALSE主要发起人之一山世光于2014年6月18日创建了VALSE专业学术交流QQ群，即VALSE-A群。此后，逐渐开通了VALSE-B-R群。从而形成了一个近两万两千人的视觉与学习青年学者在线社区。在白翔、程明明、孟德宇等青年学者的支持下，自2014年9月开始VALSE每周或隔周定期举办VALSE Webinar线上学术报告会。此外，在已有品牌活动的基础上，还陆续推出了《VALSE短教程》、《VALSE论文速览》等，以经济、便捷的在线形式，将众多青年学者的最新工作和学术思想呈现给世界各地的华人青年学者和研究生。自2020年4月以来，活动迁移至B站直播。VALSE Webinar迄今已组织385期的在线学术报告。特别是众多知名青年学者的亲临报告（如：颜水成、王晓刚、屠卓文、凌海滨、沈春华、张磊、朱军、李纯明、印卧涛、熊红凯、刘利刚、齐国君、刘烨斌、毕彦超、Philip Torr、华刚、刘小明等），更大大激励了VALSE Webinar的发展，目前VALSE B站有5.4万粉丝，B站所存放的视频是VALSE 2020年以来的Webinar录制视频，视频的累计播放量161.1万，单个视频的最高播放量在5.6万次。逐渐形成了一个独具特色、经济高效、便捷实用的在线学术交流舞台。VALSE历史视频都会更新在B站空间，欢迎在B站搜索VALSE_Webinar关注我们！也可以通过链接直接观看：https://space.bilibili.com/562085182/。

VALSE Online/Webinar 是青年学者自组织、自管理的舞台。其兴起不仅得益于VALSE指导委员会成员的大力支持，更离不开逐渐形成的VALSE Online组织团队。除发起人山世

光之外，越来越多的青年学者参与了进来，特别是白翔（华中科技大学）、程明明（南开大学）、孟德宇（西安交通大学）、彭玺（四川大学)、贾伟（合肥工业大学）、郑海永（中国海洋大学）、纪荣嵘（厦门大学）、姬艳丽（中山大学）、张利军（南京大学）、章国锋（浙江大学）、左旺孟（哈尔滨工业大学）、张兆翔（自动化所）、何晖光（自动化所）、禹之鼎（CMU)、苏航（清华大学）、欧阳万里（悉尼大学）、郭裕兰（中山大学）、郭宗辉（中国海洋大学）等，都为VALSE Online的发展付出了大量心血。为了更好地组织VALSE年度及在线活动，VALSE成立了常务AC委员会(LACC)，资深AC委员会(SACC)，执行AC委员会(EACC)（名单参见后面的委员会名单)，220名（不含已退役AC）青年学者积极参与到了相关活动的组织中。

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

借此机会，我们要诚挚感谢本届VALSE大会的合作单位，包括：华为、AutoDL、百度、OPPO、腾讯优图、美团、合合信息、阿里妈妈、无问芯穹、银河通用、阿里云、小米、TCCI、九章云极、极视角、真格基金、美图、金山办公、思腾合力、思谋科技、爱诗科技、趋动科技、首都在线（GpuGeek）、将门、易方达、摩尔线程、数据堂、英博数科、元戎启行、云天励飞、杭州宇树科技。感谢这些公司的负责人和联系人为赞助VALSE而做出的努力，谢谢您们！

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

VALSE 在线活动参与方法介绍

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

直播地址：

https://live.bilibili.com/22300737；

历史视频观看地址：

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

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

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

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

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

VALSE 各委员会

顾问委员会

高新波	西安电子科技大学	马　毅	UC Berkeley
周志华	南京大学	杨　健	南京理工大学
汤　进	安徽大学	吴小俊	江南大学

指导委员会

白　翔	华中科技大学	何晓飞	浙江大学
华　刚	Wormpex AI Research	黄　华	北京师范大学
李学龙	中国电信集团	刘青山	南京邮电大学
潘　纲	浙江大学	山世光	中国科学院计算技术研究所
汪　萌	合肥工业大学	徐　东	香港大学
颜水成	新加坡国立大学	虞晶怡	上海科技大学
余　凯	地平线机器人	张敏灵	东南大学
周　昆	浙江大学		

常务 AC 委员会（LACC）

主席： 白　翔　华中科技大学

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

资深 AC 委员会（SACC）

主席： 姬艳丽　中山大学

副主席： 严骏驰	上海交通大学	谢凌曦	华为技术有限公司
张姗姗	南京理工大学		
资深 AC： 陈　涛	复旦大学	樊　彬	北京科技大学
丛润民	山东大学	高常鑫	华中科技大学
洪晓鹏	哈尔滨工业大学	高陈强	中山大学
胡　鹏	四川大学	江　波	安徽大学

	连宙辉	北京大学	林巍峣	上海交通大学
	刘　偲	北京航空航天大学	李冠彬	中山大学
	明　悦	北京邮电大学	马　超	上海交通大学
	潘金山	南京理工大学	任传贤	中山大学
	沈　为	上海交通大学	王利民	南京大学
	王文冠	浙江大学	王云鹤	华为诺亚方舟实验室
	魏云超	北京交通大学	徐　畅	悉尼大学
	许永超	武汉大学	夏　勇	西北工业大学
	谢凌曦	华为技术有限公司	严骏驰	上海交通大学
	杨　猛	中山大学	张　磊	重庆大学
	张　林	同济大学	赵　健	军事科学院
	郑伟诗	中山大学	郑　乾	浙江大学

执行 AC 委员会（EACC）

主席：	郭裕兰	中山大学		
副主席：	胡　鹏	四川大学	郭宗辉	中国海洋大学
	李冠彬	中山大学	郑　乾	浙江大学
执行 AC：	白　磊	上海人工智能实验室	白亚龙	京东
	曹　兵	天津大学	曹相湧	西安交通大学
	陈冠英	中山大学	陈　浩	香港科技大学
	陈使明	美国卡耐基梅隆大学、阿联酋人工智能大学	陈威华	阿里巴巴达摩院
	程光亮	英国利物浦大学	崔兆鹏	浙江大学
	代季峰	清华大学	邓　欣	北京航空航天大学
	丁恒辉	复旦大学	丁明宇	加州大学伯克利分校
	丁长兴	华南理工大学	董宣毅	Google
	董胤蓬	清华大学	范鹤鹤	浙江大学
	冯　婕	西安电子科技大学	冯尊磊	浙江大学
	傅雪阳	中国科学技术大学	高广谓	南京邮电大学
	高　林	中国科学院计算技术研究所	宫　辰	南京理工大学
	宫明明	墨尔本大学	顾　实	电子科技大学
	顾舒航	电子科技大学	郭春乐	南开大学
	郭　青	新加坡科技研究局	郭晓杰	天津大学
	郭宗辉	中国海洋大学	韩　波	香港浸会大学
	韩　凯	华为诺亚方舟实验室	韩晓光	香港中文大学（深圳）
	贺　通	上海人工智能实验室	胡　迪	中国人民大学
	胡建芳	中山大学	胡庆拥	军事科学院
	黄　雷	北京航空航天大学	黄维然	上海交通大学
	黄文炳	中国人民大学	霍　静	南京大学
	贾　旭	大连理工大学	江　波	安徽大学
	焦剑波	伯明翰大学	金　鑫	宁波东方理工大学（暂名）
	康国梁	北京航空航天大学	况　琨	浙江大学

雷柏英 深圳大学
李崇轩 中国人民大学
李 坤 天津大学
李 爽 北京理工大学
李永露 上海交通大学
林 迪 天津大学
刘 晗 大连理工大学
刘晋源 大连理工大学
刘希慧 香港大学
刘 洋 北京大学
刘 勇 中国人民大学
马月昕 上海科技大学
牛玉磊 哥伦比亚大学
齐梦实 北京邮电大学
屈靓琼 香港大学
任冬伟 哈尔滨工业大学
盛 律 北京航空航天大学
眭亚楠 清华大学
孙奕帆 百度
唐彦嵩 清华大学
田亚鹏 德克萨斯大学达拉斯分校
万人杰 香港浸会大学
王福田 安徽大学
王国庆 电子科技大学
王立君 大连理工大学
王 伟 北京交通大学
王 鑫 清华大学
王一帆 大连理工大学
韦星星 北京航空航天大学
吴庆波 电子科技大学
谢雨彤 澳大利亚阿德莱德大学
徐天阳 江南大学
闫庆森 西北工业大学
杨二昆 西安电子科技大学
杨文瀚 鹏城实验室
杨 旭 西安电子科技大学
叶翰嘉 南京大学
弋 力 清华大学
于乐全 香港大学
余肇飞 北京大学
张弘扬 滑铁卢大学

李 策 兰州理工大学
李皓亮 香港城市大学
李雷达 西安电子科技大学
李 响 西安交通大学
李 镇 香港中文大学（深圳）
刘 峰 墨尔本大学
刘 昊 宁夏大学
刘 伟 上海交通大学
刘夏雷 南开大学
刘瑶瑶 约翰霍普金斯大学
陆 昊 华中科技大学
倪张凯 同济大学
彭春蕾 西安电子科技大学
秦 杰 南京航空航天大学
任 博 南开大学
任文琦 中山大学
舒祥波 南京理工大学
隋 尧 北京大学
唐 厂 中国地质大学（武汉）
田春伟 西北工业大学
涂志刚 武汉大学
汪婧雅 上海科技大学
王高昂 浙江大学
王 鹤 北京大学
王思为 军事科学院
王 啸 北京航空航天大学
王亚星 南开大学
王奕森 北京大学
吴金建 西安电子科技大学
武 宇 武汉大学
徐婧林 北京科技大学
徐 易 大连理工大学
晏轶超 上海交通大学
杨 帅 北京大学
杨 曦 西安电子科技大学
杨 旭 东南大学
叶 茫 武汉大学
易 冉 上海交通大学
于 茜 北京航空航天大学
元玉慧 微软亚洲研究院
张 健 北京大学

	张　力	复旦大学	张　璐	大连理工大学
	张平平	大连理工大学	张　青	中山大学
	张瑞茂	中山大学	张宇伦	上海交通大学
	赵恒爽	香港大学	赵思成	清华大学
	赵文达	大连理工大学	周天飞	苏黎世联邦理工学院
	周　毅	东南大学	朱霖潮	浙江大学
	朱翔昱	中国科学院自动化研究所		

秘书处：

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

会场及酒店

会场位置

会议地点：珠海国际会议中心-C 座
交通地址：珠海十字门中央商务区湾仔片区（香洲区银湾路 1663 号）

交通信息

交通情况概览

珠海九洲机场

打车路线【全程估计 22 分钟】

距离会场约 14.6 公里，出租车约 25 元

珠海金湾机场

打车路线【全程估计 29 分钟】

距离会场约 30.5 公里，出租车约 50 元，高速费约 14 元。

十字门站

打车路线【全程约 6 分钟】

距离会场约 373 米

珠海站

打车线路【全程估计 14 分钟】

距离会场约 8.4 公里，出租车约 12 元。

前山站

打车线路【全程估计 15 分钟】

距离会场约 8.8 公里，出租车约 12 元。

深圳市

高铁线路【全程估计 2 小时 4 分钟】

深圳北站-珠海站（车程约 1 小时 50 分钟，高铁二等座费用约 144.5 元）；
珠海站-珠海国际会展中心（车程约 14 分钟，出租车约 12 元）。

驾车线路【全程估计 1 小时 40 分钟】

深圳市-珠海国际会展中心（全程约 120 公里，高速费约 114 元）

广州市

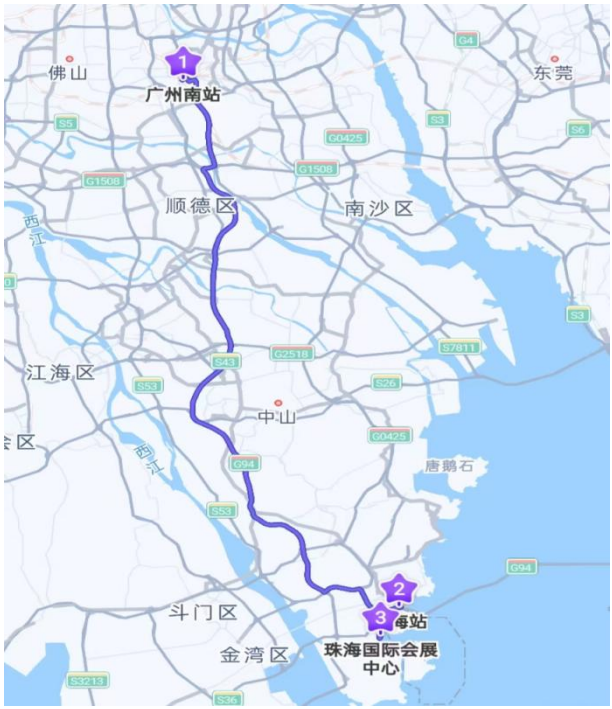
动车线路【全程估计 1 小时 14 分钟】

广州南站-珠海站（车程约 1 小时，动车二等座费用约 70 元）；

珠海站-珠海国际会展中心（车程约 14 分钟，出租车约 12 元）。

驾车线路【全程估计 1 小时 50 分钟】

广州市-珠海国际会展中心（全程约 135 公里，高速费约 69 元）。



酒店信息

大会协议酒店信息

为方便各位老师和同学参会，组委会前期调研了会场周边的酒店，珠海国际会议中心周围酒店列表如下。同时也非常建议参会者自行通过携程等公共平台自行搜索会场附近酒店。

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

酒店名称	星级	房型	协议价	早餐供给	房间数量	联系人及电话	预定方式	距离
珠海瑞吉酒店	五星	豪华大床房	900	含单早	70间	李重余 13726252063	自行联系销售，报会议名称订房	400米，步行5分钟
		十字门江景标间	1000	含双早	10间			
华发行政公寓	四星	高级城景双床房	450	含双早	80间	酒店咨询电话：0756-6819999（总机预订部）	扫描小程序码： 1.选择对应日期 2.点击“更多选项”并输入促销代码：20242202	690米，步行9分钟
		商务城景一房一厅	400	含单早	80间			
		复式海景一房一厅	450	含单早	130间			
		豪华城景两房一厅	700	含双早	15间			
珠海希尔顿酒店	五星	单间	600	含单早	60间	朱誉 13167026802	自行联系销售，报会议名称订房，预定时需付款，且预定不可取消	1.6公里，步行21分钟
		标间	600	含双早	60间			
全季珠海保仟口岸国际会展中心店	四星	单间	530	含双早	106间	伍旺旭 13672788442	自行联系销售，报会议名称订房	730米，步行10分钟
		标间	580	含双早	28间			
上午国际酒店	二星	单间	350	不含早（早餐30元/份）	20	酒店前台：16676668806	自行联系酒店，报会议名称订房	660米，步行9分钟
		标间	350	不含早（早餐30元/份）	20			
银华酒店	三星	单间	400	含单早	24	雷儿 18198791399	自行联系销售，报会议名称订房	1.4公里，步行17分钟
		标间	400	含双早	56			
凯瑞海会展酒店公寓	二星	单间	350	不含早（早餐30元/份）	40	酒店前台：13928090217	自行联系酒店，报会议名称订房	670米，步行9分钟
		标间	350	不含早（早餐30元/份）	20			
朵尔曼国际酒店公寓	二星	单间	388	不含早餐	40	魏经理 13417859599	自行联系销售，报会议名称订房	890米，步行12分钟
		标间	428	不含早餐	20			
朵尔曼悦享公寓酒店	二星	单间	388	不含早餐	70	魏经理 13417859599	自行联系销售，报会议名称订房	960米，步行12分钟
		标间	428	不含早餐	40			
盛捷酒店（现改名丽呈酒店）	四星	标间/单间	400	均含双早	单标间各70间	喻巍 13728619753	自行联系销售，报会议名称订房	2.8公里，车程5分钟
翰江都城酒店	四星	时尚房标间/单间（周日-周四）	24 8	均含双早	时尚房标间：71间 时尚房单间：6间	何国伏 13750087216	自行联系销售，报会议名称订房	18公里，车程19分钟
		时尚房标间/单间（周五-周六）	268	均含双早				
		风雅房标间/单间	318	均含双早	风雅房标间：73间 风雅房单间：31间			
		都会房标间/单间	338	均含双早	都会房标间：18间 都会房单间：12间			
易八高尔夫酒店	三星	高级标间/高级单间	268	均含双早	高级标间：27间 高级单间：12间	肖强 13750042023	自行联系销售，报会议名称订房	6.6公里，车程9分钟
		豪华标间/豪华单间	288	均含双早	豪华标间：36间 豪华单间：27间			
拱北智选假日酒店	四星	单间	380	含单早	64间	张秀 13697796607	自行联系销售，报会议名称订房	8.4公里，车程13分钟
		标间	380	含双早	36间			
		城市景观单间	440	含单早	32间			
		城市景观标间	440	含双早	46间			
珠海南航明珠大酒店	四星	精选单间	370	含单早	100间	Candy 18923017077	通过扫描二维码，关注公众号，点击链接后输入会议预定代码9132	3.4公里，车程5分钟
		商务单间	420	含单早	30间			
		标间	450	含双早	40间			
珠海凤凰谷假日酒店	四星	A座豪华标间/单间（平日）	350	均含双早	豪华标间：16间 豪华单间：19间	张甜 13726281975	自行联系销售，报会议名称订房	10公里，车程16分钟
		A座豪华标间/单间（周末）	380	均含双早				
		A座行政单间（平日/周末）	410/460	含双早	5间			
		A座豪华行政单间（平日/周末）	470/500	含双早	5间			
		B座高级标间/单间（平日）	270	均含双早	单标间各24间			
		B座高级标间/单间（周末）	300	均含双早				
		D座标间/单间（平日）	240	含双早	标间：50间 单间：38间			
		D座标间/单间（周末）	270	含双早				
雅辰悦居酒店	四星	单间	440	含单早	20间	吴经理 13172258029	通过扫描二维码自行预定房间，会议期间可以安排大巴车送到	7.1公里，车程12分钟
		标准间	500	含双早	80间			
棋帆酒店	四星	普通单间	360	含单早	15间	马丽芳 13798965997	自行联系销售，报会议名称订房	3.6公里，车程7分钟
		海景单间	400	含单早	25间			
		普通标间	400	含双早	15间			
		海景标间	450	含双早	15间			
希悦博雅连锁公寓	二星	单间	150	无早餐	50间	彭经理 13169658585	自行联系销售，报会议名称订房	3公里，车程5分钟
		标间	180	无早餐	50间			
万隆国际公寓	二星	单间	150	无早餐	50间	薛经理 15875669921	自行联系销售，报会议名称订房	4.4公里，车程7分钟
		标间	200	无早餐	50间			
珠海北京酒店	三星	单间	218	无早餐	35间	酒店前台：0756-8666288 滕经理 18023011260	自行联系销售，报会议名称订房	12公里，车程19分钟
		标间	218	无早餐	26间			


VALSE