主办 CCF计算机视觉专业委员会

COMPUTER VISION NEWSLETTER

CCF 计算机视觉专委会简报

04 2024

总第 42 期

CCF 计算机视觉专委会

COMPUTER VISION NEWSLETTER

计算机视觉专委会简报

2024 年第 04 期

总第 42 期

主　办　CCF 计算机视觉专业委员会

编委会

CONTENTS
简报目录

CCF 计算机视觉专委会
CCFCV.CCF.ORG.CN
CCFCVN@GMail.com

CCF-CV 走进高校系列报告会

第 139 期 河南科技学院

2024 年 9 月 28 日，由中国计算机学会主办，中国计算机学会计算机视觉专委会（CCF-CV）、河南科技学院软件学院联合承办的第 139 期 CCF-CV 走进高校系列报告会——“计算机视觉前沿技术及应用” 学术论坛在河南科技学院行政楼 206 报告厅成功举行。本次活动邀请了南京理工大学李泽超教授、西安电子科技大学吴金建教授、北京交通大学王伟教授 3 名国家级人才计划入选者作特邀报告。本期活动的执行主席是河南科技学院软件学院院长古乐声教授和河南科技学院软件学院副院长马玉琨副教授。报告由古乐声教授主持。

活动由吴金建教授、王伟教授、李泽超教授分别做主题报告。从开放环境下视觉内容异常检测、新型事件相机系统设计及异步信号智能处理、数字人脸生成与可控编辑等方面进行了分享。三位专家与师生积极互动，展开了热烈的讨论，针对师生们提出的一系列问题，专家们给予了详尽的解答，分享了大量宝贵的学术观点与见解，现场氛围活跃。此次报告极大地鼓舞了师生们的科研动力，为师生搭建了一个难得的学习与交流平台，促进了知识与思想的碰撞与融合。

第 140 期 苏州大学

2024 年 10 月 12 日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、苏州大学承办的第 140 期 CCF-CV 走进高校系列报告会——“计算机视觉前沿技术及应用”学术论坛在苏州大学天赐庄校区敬贤堂成功举行。本次活动邀请了中科院计算所陈熙霖研究员、中山大学赖剑煌教授、南京邮电大学刘青山教授、华中科技大学白翔教授、北京工业大学毋立芳教授、中科院计算所王瑞平研究员等众多计算机视觉领域的顶尖专家进行专题报告和交流研讨。本次活动的执行主席是苏州大学计算机学院副院长黄河教授与青年教师樊佳庆博士，活动开幕式由苏州大学计算机科学与技术学院李凡长教授主持。

活动首先由苏州大学党委书记张晓宏教授致欢迎词。随后，黄河教授和陈熙霖研究员进行致辞。在学术报告环节，赖剑煌教授、白翔教授、毋立芳教授分别做主题报告。在专题报告之后，陈熙霖、刘青山、白翔、毋立芳、王瑞平五位专家围绕 AI 技术革新下计算机视觉发展前景与挑战开展了圆桌论坛。参加本次活动的老师和同学认真聆听了报告，并与报告嘉宾进行了交流互动。

CCF-CV 走进高校系列报告会

第 141 期 河海大学

2024 年 11 月 22 日下午，CCF-CV 走进河海大学@水利部水利大数据重点实验室学术报告会于河海大学牛首山科技园会议室顺利召开。本次论坛由中国计算机学会计算机视觉专业委员会（CCF-CV）主办，河海大学计算机与软件学院、水利部水利大数据重点实验室承办。河海大学原副校长、水利部水利大数据重点实验室首席教授朱跃龙，中国科学院大学吕科教授、哈尔滨工业大学左旺孟教授、中国科学院自动化研究所杨小汕研究员、CCF 计算机视觉专委会副秘书长、南京理工大学潘金山教授、上海大学王日英教授等国内知名专家学者出席活动，会议由计算机与软件学院副院长刘凡教授主持。

活动首先由朱跃龙教授致辞，他详细介绍了水利部水利大数据重点实验室取得诸多关键技术突破和显著成果。随后杨小汕研究员、吕科教授、左旺孟教授分别围绕多模态大模型、机器视觉图像处理和智能学习做主题报告。计算机与软件学院的教师和研究生代表也参加了本次会议，同特邀嘉宾进行了深入交流与探讨。

第 142 期 嘉应学院

2024 年 11 月 23 日，由中国计算机学会计算机视觉专委会（CCF-CV）主办、嘉应学院承办的“CCF-CV 走进高校系列报告会”在嘉应学院百年纪念大楼 121 报告厅成功举行。本次活动以“AI 大模型时代的计算机视觉前沿技术与应用”为主题，邀请了中山大学郑伟诗教授、西北工业大学王琦教授、重庆大学张磊教授、南开大学杨巨峰教授、深圳大学贾森教授等 5 位计算机视觉领域的专家学者做特邀报告。会议由嘉应学院数学学院院长黄可坤教授和计算机学院院长叶忠文共同主持。

活动首先由嘉应学院廖志成副校长发表致辞。随后，杨巨峰教授代表专委会进行了致辞。在学术报告环节，郑伟诗教授、王琦教授、张磊教授、杨巨峰教授、贾森教授分别做主题报告。报告会之余，特邀专家们还与嘉应学院青年教师开展学科建设及科研项目申报的指导工作。为青年教师的项目申报提供了宝贵的意见，更让他们对未来的科研方向和申报策略有了更深刻的认识。

责任编委 毋立芳、黄岩

CCF-CV 走进企业系列交流会

第 29 期 合合信息

随着 AI 逐渐渗透到各行各业，其安全风险也在与日俱增，AI 技术“野蛮生长”引发公众担忧。12 月 11 日，由中国计算机学会计算机视觉专委会主办，合合信息承办，中国运筹学会数学与智能分会协办的《打造大模型时代的可信 AI》论坛（简称“论坛”）顺利举行。特邀来自上海交通大学、电子技术标准化研究院、中国科学技术大学、中科院、合合信息等机构与企业的专家们，分享 AI 安全领域最新研究成果。

此次活动旨在联合产学研机构共同探寻高效的 AI 安全治理道路。会议开始，合合信息智能技术平台事业部副总经理丁凯博士发表致辞，对各位嘉宾、各位参加本次活动的朋友们表达了热烈欢迎。随后，中国计算机学会计算机视觉专委会副秘书长潘金山教授进行了致辞，针对 AI 时代“有图未必有真相”的现状，鼓励通过学术交流活动推动 AI 安全和计算机视觉领域技术发展。

致辞结束后，与会嘉宾们就 AI 安全治理问题展开了热烈探讨。上海交通大学教授金耀辉、合合信息图像算法研发总监郭丰俊、中国电子标准院网安中心测评实验室副主任何延哲、中国科学技术大学教授谢洪涛、中国科学院自动化研究所研究员赫然博士分别进行主题报告。

生成式人工智能发展日新月异，技术革新与安全治理缺一不可，面对 AI 的潜在风险，加强行业内部自律，从源头做好安全措施是守护 AI 健康成长的第一道防线。本次活动是产学研联合探索 AI 安全治理的一次有效尝试。未来，合合信息会持续深耕 AI 视觉安全领域，积极推动行业合作与交流。

责任编委 潘金山

第 20 期 以人为中心的视觉感知：由表及里

CCF-CV 视界无限系列研讨会

2024 年 10 月 12 日，由中国计算机学会（CCF）主办、CCF 计算机视觉专业委员会和南京师范大学计算机与电子信息学院/人工智能学院承办的 CCF-CV “视界无限”系列研讨会——“以人为中心的视觉感知：由表及里”在南京师范大学芳菲楼报告厅成功举办。

此次会议特邀南京师范大学学科建设处处长张晓锋教授及 CCF-CV 秘书长、中国科学院计算技术研究所王瑞平研究员发表致辞。

张晓锋教授在致辞中介绍了南京师范大学概况，并表达了对此次活动能够有力推动计电学院相关领域研究的深切期望，期待其为计算机学科及其相关学科带来新的灵感与启发。

王瑞平研究员介绍了 CCF-CV 系列学术活动的背景与现状，并对与会的领导、嘉宾、老师及同学们表达了诚挚的感谢，预祝研讨会取得圆满成功。

中国科学院计算技术研究所山世光研究员、中国科学院自动化研究所张兆翔研究员、浙江大学李玺教授、上海交通大学林巍峣教授、南京大学单彩峰教授、南开大学杨巨峰教授做主题报告。在主题研讨环节中，南开大学刘夏雷副教授、清华大学唐彦嵩研究员、香港中文大学唐诗翔博士、南京大学张振宇副教授，分别就各自的研究领域做了观点分享。此外，多位报告专家与参会的青年学者还围绕相关议题展开了热烈而深入的自由讨论，共同促进了学术的碰撞与交融。会议由南京师范大学钱建军教授、周俊生教授、顾彦慧教授、杨琬琪副教授，以及南京理工大学张姗姗教授共同主持。

山世光研究员的报告以“视线估计与跟踪研究”为题，深入浅出地介绍了两项前沿的视线估计技术：一是利用三维视线估计技术，实现对面部图像的眼神精准跟踪；二是结合二维与三维视线追踪技术，对场景图像中的眼神进行高效捕捉。接着，他分享了其团队在眼神行为分析领域的研究成果，特别是针对自闭症谱系障碍儿

童的眼神注意分析，以及社交交流中视线行为的深度剖析。最后，山世光研究员还就数据驱动方法在数据收集方面的实践以及当前研究面临训练和测试数据不足等挑战进行了探讨。

张兆翔研究员的报告以“以人为中心的生成式模型初探：由里及表”为题，巧妙地从人工智能和以人为本的角度深入剖析了“里”与“表”的概念，并以此为基础，探索构建兼具人性与风格化的智能体。随后逐一详述了角色扮演语言风格模型、角色扮演意识风格模型及角色扮演动作生成模型一系列创新技术路径，针对当前模型的发展现状与存在的关键问题进行了探讨。最后，对由里及表的全方位生成的未来进行展望并强调了由表及里的感知与由里及表的生成之间的双向促进作用。

李玺教授的报告以“多模态视觉结构学习”为题，首先从目标视觉感知特性、视觉特征表达、深度学习器构建机制、高层语义理解等多维度视角进行了深入剖析，并引入了大规模多模态特征学习所涉及的核心研究问题挑战与技术方法。随后，他系统回顾了多模态特征表达和学习领域的发展历程，分享了近几年来其团队利用特征学习进行视觉语义分析与理解方面的一系列代表性的研究工作及其应用，包括球面几何感知的 Transformer 在全景语义分割中的应用，以及通过语言自适应推理，通过迭代扫描进行引用表达理解等。最后李玺教授还深入探讨了多模态特征学习当前面临的开放性问题与难题。

林巍峣教授的报告以“语义驱动的大规模视频内容感知与编码”为题，首先介绍了目标行为与事件语义提取方面的工作，通过重构当前行为事件识别与定位的框架，提出了从全局至局部的渐进式行为事件提取架构。紧接着，他深入阐述了复杂事件步骤对齐及因果推理的研究进展，通过对复杂事件的步骤挖掘、对齐实现对目标对象和事件类型的感知，并对事件中的因果逻辑进行推理分析实现对视频中复杂事件的全面理解。此外，林教授还详细介绍了其课题组在语义信息压缩编码领域的研究成果，设计了一套针对关键点序列及因果关系图等核心视频语义内容的压缩编码架构，实现了对象、交互、事件多层次语义的高效联合编码。最后，他分享了这一研究方向在实际应用场景中的广阔前景与案例。

单彩峰教授的报告以“Camera-based Physiological Measurement”为题，首先介绍了基于相机的非接触式生命体征监测技术的研究背景，并着重阐述了非接触式光学成像相较于传统接触式生物传感器的显著优势。随后，他详细列举了该技术的实际应用场景，包括新生儿远程心率监护、血压测量、灌注检测、体内器官组织的血液流动情况监测等，这些应用已在多个复杂的医疗场景中验证了非接触式多维生命体征监测的可行性。紧接着，单教授分享了其团队最新的研究进展，提出了 Camera-SCG 技术，该技术利用激光干涉现象构建三维离焦散斑探测系统，实现了对心脏三维运动的精准非接触光学重建。最后，他探讨了非接触式生命体征监测研究中面临的诸多挑战，为与会者提供了宝贵的启示。

杨巨峰教授的报告以“情智兼备数字人与机器人关键技术初探”为题，首先剖析了该科学问题的起源与背景，分析了国内外机构在该领域的研究现状，通过对比两者的发展历程，阐述了情智兼备的数字人与机器人在未来科技中的重要地位及其所面临的关键难题与挑战。紧接着，分享了五个可研究的方向，包括情感机理、情感大模型、多模态情智融合解译、个性化情感表征与动态计算、情感表达与连续交互。最后，杨巨峰教授对情智兼备技术的一体两面及其蕴含的研究机遇进行了探讨，为与会者提供了宝贵的洞见与启发。

在主题研讨环节，刘夏雷副教授在题为“基于图文预训练模型的连续学习方法研究”的报告中表示连续学习是新一代人工智能系统的关键技能之一，阐述连续学习的概念并指出其面临最大的挑战--灾难性遗忘，随后以图文预训练模型为基础，探索知识引导的连续学习方法，分别从判别式模型与生成式模型两个角度，剖析并展示解决连续学习难题的新思路。

唐彦嵩研究员在题为“关于人体动作理解与生成的一些思考”的报告中介绍了不同维度下的人体动作理解任务，通过长时程视频动作问答，关注细粒度人体动作的描述在推理阶段进行更加高效设计。随后介绍了不同控制条件下的人体动作生成方法并展示了音乐驱动动作生成的研究成果，最后对未来人体动作理解与生成领域面临的挑战进行了思考与讨论。

唐诗翔博士在题为“通用行人检索大模型研究”的报告中指出大规模多模态行人重识别的核心挑战在于构建一个能够支持多样化多模态指令并具备强大场景泛化能力的通用行人检索模型。为此，唐博士介绍了一系列研究成果，包括多模态通用行人重识别数据集、行人重识别模型与行人检索大模型。

张振宇副教授在题为“表里先验引导的三维数字人重建与生成”的报告中介绍三维数字人的背景与意义。随后从“表”先验与显式约束、“里”先验的规律探索，以及两者的统一与结合三个方面，系统介绍了这两类先

验在三维数字人重建与生成问题中的重要应用。深入分析了这两类先验的适用性，指出了在解决三维数字人重建与生成难题中的独特优势，同时也讨论了它们可能带来的局限性与挑战。

本次研讨会环节邀请了李玺教授、单彩峰教授、林巍峣教授、杨明教授、刘夏雷副教授、唐彦嵩研究员、唐诗翔博士、张振宇副教授进行自由讨论。对话由张珊珊教授主持。

围绕“应该怎样随着不断涌现的热点去选择一些新的任务和问题进行研究”这个问题，各位老师结合自己的研究领域进行了深入探讨。与会的老师和同学积极提问，同各位老师进行了深入交流与探讨。

研讨会在大家激烈的讨论中落下帷幕。每一位专家的分享都提供了宝贵的思路和启示。希望大家不断探索未知领域，共同应对技术发展的挑战，为实现更加智能、更加美好的世界贡献智慧和力量。

第 21 期 多模态结构模式表征与识别

CCF-CV 视界无限系列研讨会

2024 年 11 月 23 日，由中国计算机学会（CCF）计算机视觉专业委员会（CCF-CV）主办，安徽省多模态认知计算安徽省重点实验室和安徽省安全人工智能重点实验室承办的 CCF-CV “视界无限”系列研讨会第 21 期——“多模态结构模式表征与识别”在安徽大学磬苑校区理工 A 楼报告厅成功举办。研讨会由安徽大学计算

机学院江波教授、安徽大学人工智能学院李成龙教授、郑爱华教授共同担任执行主席。

会议邀请了安徽大学科学技术处处长郑春厚教授和安徽大学计算机学院院长仲红教授致辞。郑春厚教授在致辞中简要地介绍了安徽大学的历史发展与现状，并表示安徽大学近年来正以奋斗的姿态抓住人工智能新浪潮，以此为契机将安徽大学的工科发展推向新的发展阶段，最后郑春厚教授希望这次活动能够进一步加强安徽大学与计算机领域优秀的同仁之间的联系，为计算机领域的发展带来新的启发与思考。

仲红教授介绍了安徽大学计算机学院的情况，并对一直以来 CCF 视觉专委会和各兄弟院校对安徽大学计算机学科发展提供的帮助表达了衷心的感谢，并预祝此次研讨会圆满成功同时祝愿CCF-CV视界无限系列研讨会能够越办越好！

四川大学吕建成教授、浙江大学李玺教授、中国科学院自动化研究所王亮研究员、大连理工大学卢湖川教授、南京理工大学张姗姗教授做主题报告。会议由安徽大学计算机学院江波教授、安徽大学人工智能学院郑爱华教授共同主持。会议吸引了来自中国科学技术大学、西北工业大学、安徽理工大学、安徽建筑大学、合肥大学、中科院合肥物质科研研究院、江淮前沿科技协同创新中心等高校以及企业的师生报名参加。

吕建成教授在其题为“面向边缘智能的联邦学习”的报告中，深入探讨了神经网络的发展趋势、联邦学习与边缘智能的结合，以及基于神经网络的联邦学习优化等关键议题。

首先，吕教授从神经网络的学习问题入手，详细讨论了学习算法的演变、学习效率的提升以及学习过程中遇到的挑战。

接着，吕教授分析了特征提取问题，强调了在不同应用场景下如何有效地从数据中提取有用信息的重要性，并讨论了特征提取技术的最新进展。

吕教授详细介绍了其团队针对这些问题的研究工作，包括算法优化、通信策略改进、模型鲁棒性增强等方面，并分享了团队取得的研究成果。

李玺教授在其题为“多模态特征表达与学习”的报告中，为听众呈现了一场关于多模态数据处理和特征学习的深刻讲座。报告不仅回顾了多模态特征表达和学习

领域的发展历程，还深入探讨了数据驱动的大规模视觉特征学习任务，涉及视觉感知特性、视觉特征表达、深度学习器构建机制、高层语义理解等多个维度。

首先，李教授从视觉计算理论与应用的角度出发，详细阐述了多模态特征表达的重要性。他特别提到了近年来在深度学习领域取得的一些突破性进展，并探讨了这些技术在多模态特征学习中的应用潜力。李教授还介绍了团队近年来在多模态特征表达与学习领域的一系列代表性研究工作，包括在图像识别、视频分析、自然语言处理等方面的应用。

最后，李教授从图像生成的角度探讨了多模态认知交互任务中多模态特征学习的特点。他分析了机器如何通过学习多模态特征来理解和交互世界，以及这一过程中所面临的挑战和机遇。

王亮研究员在其题为“多模态认知计算”的报告中，为听众提供了一个全面而深入的视角，探讨了多模态数据在认知计算中的应用和挑战。

王研究员首先明确了多模态认知计算的定义，即利用视觉、听觉等多种模态的数据，研究模态表征、融合、对齐、生成等关键技术，以实现从多模态感知到多模态理解的技术跨越。接着，王研究员分析了现有多模态模型与方法在认知功能方面的不足，特别是模型架构同质化的问题。

针对这些问题，王研究员介绍了他的团队在深度认知网络框架指导下开展的一系列研究工作。这些研究涵盖了注意机制、记忆机制、推理机制、决策机制等四个层面，旨在提升模型在多模态认知能力。他详细阐述了团队在多模态共享记忆、多模态概念知识、多模态知识压缩、多模态幻觉检测等方面的研究成果，这些研究不仅推动了多模态认知计算技术的发展，也为相关领域的应用提供了新的思路和工具。

卢湖川教授分享了题为“视觉内容感知生成”的报告。报告详细介绍了团队基于生成模型所开展的相关研究以及取得的最新进展与应用，展现了生成技术在视觉内容创造中的潜力和前景。

首先，卢教授介绍了团队在生成模型框架优化方面的创新工作。通过深入研究和技术创新，团队显著提升了大规模模型的训练效率，这不仅降低了生成技术的应用门槛，也使得生成技术能够为更多用户提供帮助。卢教授详细阐述了优化策略，包括算法改进以及分布式训练等，这些策略共同作用，使得生成模型的训练和部署变得更加高效和经济。接着，卢教授讲述了团队如何将图像、音频、时频等多模态数据融入到生成模型中。这种多模态数据的融合，不仅丰富了生成模型的生成内容，更符合用户的需求，同时也构建了生成模型对多模态数据理解的新范式。

在报告的后半部分，卢教授针对现有生成模型的局限性进行了深入分析。他指出，许多现有的生成模型仅关注个体主题的建模，而忽视了主题之间的物理关系或互动等问题。为了解决这一问题，卢教授介绍了团队在定制化内容生成技术方面的研究成果。

张姗姗教授在她的题为“资源受限条件下的物体检测”的报告中，深入探讨了在实际应用场景中物体检测所面临的困难与挑战。

张教授首先指出，基于深度学习的物体检测模型若要实现理想的检测精度，往往需要依赖大量带标签的数据和高性能计算设备。然而，在实际应用中，这些条件难以满足，例如大规模带标签的数据难以获取、高能耗模型难以部署于移动终端等小型应用设备等。这些问题限制了物体检测技术在更广泛场景下的应用。

针对这些挑战，张教授详细介绍了团队在资源受限条件下物体检测问题上的研究工作。这些工作主要包括个性化的半监督学习方法、无监督自适应方法以及模型轻量化方法等。

最后，张教授对领域内未来需要解决的问题进行了深刻的总结与展望。她指出，尽管目前已有一些进展，但在数据极端受限的场景下，如何进一步提升模型的性能仍然是一个挑战。此外，开放场景下未知语义的识别也是一个重要的研究方向。同时，如何将物体检测技术与下游任务更好地联动学习，也是提升整体系统性能的关键。

研讨会的最后是自由讨论与提问环节，与会专家与在场老师和同学们针对“对于质量欠佳的输入，大模型存在的概念理解偏差问题有何比较好的解决方法？”，“在对生成模型进行训练时，如何平衡模型对细节与全局生成质量控制？”，“未来针对复杂场景有趣的基于大模型的 tracking topic，有哪些值得研究的方向？”“基于图的推理模型在未来具有怎样的情景？”等问题进行了深入的交流与讨论。

最后，研讨会在主持人郑爱华教授的宣布下落下帷幕，大家在热烈的讨论氛围中陆续离场。此次研讨会围绕多模态结构模式表征与识别主题，汇聚了该领域的专家学者，大家齐聚一堂共同探讨了该领域的最新进展和未来发展趋势，通过深入的交流和思想碰撞，为推动该领域的创新和发展提供了新的思考和动力。参与者们纷纷表示，这次研讨会不仅增进了彼此之间的了解和合作，也为解决实际问题提供了新的思路和方法，期待未来能有更多这样的交流机会。

责任编委 杨巨峰

2024 年度 CCF-CV 专委工作会议顺利召开

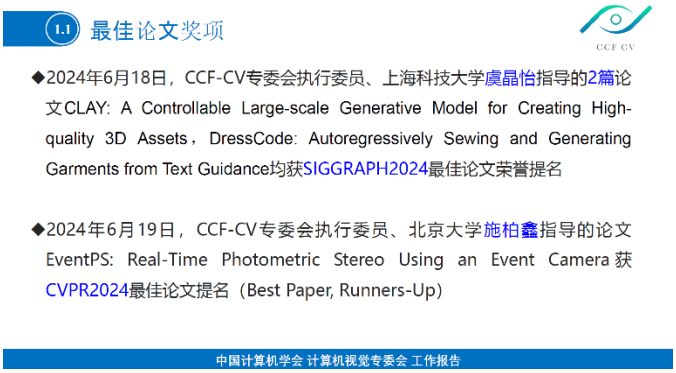
2024 年 10 月 18 日，中国计算机学会计算机视觉专委会（CCF-CV）2024 年度工作会议在新疆国际会展中心成功召开，专委会秘书长、中国科学院计算技术研究所王瑞平研究员主持会议。

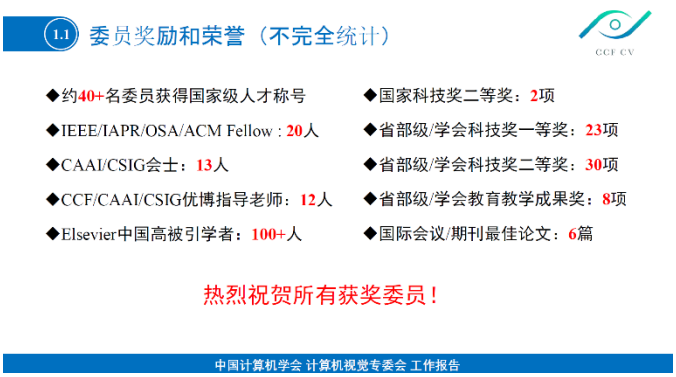
首先，中国科学院计算技术研究所所长、专委会主任陈熙霖研究员致辞。陈老师感谢了全体委员在过去一年中为专委会发展做出的杰出贡献。在全体同仁的共同努力下，专委会组织开展了走进高校、走进企业、“视界无限”、讲习班、RACV、PRCV 等多场精彩纷呈的学术活动，惠及计算机视觉领域广大师生和研究人员。陈老师对以上活动的承办单位和组织者表示感谢，预祝全体委员在未来取得更大成绩。

随后，CCF 专委工委委员、华南理工大学副校长许勇教授代表学会发言。许老师肯定了专委会换届以来开展的工作和取得的成绩，尤其是 RACV 的系列特色品牌活动在学会和兄弟专委会产生了良好示范效应，在计算机视觉领域广大从业者中赢得了口碑。许老师鼓励全体委员在专委会领导的带领下更上一层楼，祝愿工作会议圆满成功。

接下来，专委会秘书长王瑞平研究员向与会嘉宾和执行委员做了 2024 年度专委会工作报告。报告简要介绍了目前专委会的组织结构，概述了常委会议及秘书处工作会议的内容，持续加强组织建设。

他还提到，多名委员参与完成的项目获得了国家级和省部级科技奖项，参与完成的论文获得了重要国际会议的最佳论文奖等，展现了专委会的学术实力。接着，他全面回顾了专委会过去一年的学术交流活动，包括走进高校、走进企业、视界无限、RACV 等，以及通报了专委会委员积极参与的学会活动，包括组织 CNCC 的专题论坛、协助 CCF 发布科学发展报告。然后，展示了专委会简报、网站和公众号等宣传渠道及取得的效果。王秘书长还提出了未来的工作计划，秉承“一切为委员服务”的宗旨，将进一步明确各项学术活动主题、突出专委会特色，通过多种方式进一步提升委员参与专委治理的活跃度。





随后，进入 CCF-CV 颁奖环节。颁奖仪式由专委会副主任、南京邮电大学刘青山教授主持。2024 年度 CCF-CV 终身学术贡献学者授予北京交通大学阮秋琦教授。阮老师曾任国务院学位委员会学科评议组成员、北京交通大学学位委员会副主席、中组部联系的党内高级专家。主要研究方向包括数字图像处理、计算机视觉、虚拟现实技术及多媒体信息处理等。阮老师致辞感谢专委会同仁给予的认可，并对专委会在推动领域发展方面发挥的重要作用表示了肯定，鼓励年轻学者百尺竿头更进一步，持续创新，再建新功。

2024 年度 CCF-CV 杰出成就学者授予香港中文大学王晓刚教授。王老师是国内较早开展基于深度学习视

觉算法研究的学者，带领团队建立人像识别领域的第一个深度学习算法，首次超越肉眼的识别精度，对推动计算机视觉的产业应用也做出了杰出贡献。王老师在致辞中分享了科研经历，讲述了几个代表性工作的缘起和故事，对来自专委会的肯定表示了感谢。

2024 年度 CCF-CV 持久影响力工作授予发表于 ECCV 2014 的论文《Learning a Deep Convolutional Network for Image Super-Resolution》，作者是 Chao Dong、Chen Change Loy、 Kaiming He、Xiaoou Tang。该工作开启了用深度学习实现图像超分辨率的先河，在理论上连接了稀疏编码和卷积网络，仅用三层网络就超越了传统算法，其核心模块被后续的底层视觉广泛采用，论文被国际同行引用超过 1.6 万余次。

2024 年度 CCF-CV 学术新锐授予大连理工大学陈鑫和上海交通大学贾萧松两位同学。

2024 年度 CCF-CV 服务贡献学者授予储珺、崔海楠、库尔班·吾布力、李实英、李晓旭、潘金山等六位老师。

根据会议日程，工作会议进行了执行委员的新增选举工作，由专委会秘书长王瑞平研究员主持。本年度共有 80 名计算机视觉领域的学者和企业研究人员申请加入专委会。在委员增选环节，申请人逐一上台介绍了个人情况和亮点工作。经常委会无记名投票，共有 61 人被遴选为新委员，大家纷纷表示将积极参与专委会组织的各项活动，为专委会建设和发展贡献力量。

委员建言献策环节由专委会副主任、中国科学院自动化研究所王亮研究员主持。原专委会主任、北京大学查红彬教授提议在今后专委会举办的活动中进一步突

出“特色”，能够通过小而精、主题集中的学术交流活动使得青年学者和学生们更加受益。专委会常务委员、航天宏图首席科学家王涛博士建议利用网站、公众号等平台对委员的研究工作进行展示与宣传，也便于在走进高校、走进企业等活动中对讲者的选拔。北京交通大学阮秋琦教授建议大家共同努力创办一种刊物，使之成为专委的一面旗帜。北京大学彭宇新教授建议后续专委会可以多多组织一些论坛形式的交流活动。委员们热情洋溢地发言把会议推向了高潮。

最后，专委会秘书长王瑞平研究员对会议进行了总结。王老师再次感谢参加会议的特邀嘉宾和执行委员。会后，全体委员进行合影留念，CCF-CV 2024年度工作会议圆满落幕。

责任编委 朱安娜

CCF-CV 常务委员会 2024 年度第二次工作会议顺利召开

2024 年 10 月 13 日中国计算机学会计算机视觉专委会 (CCF-CV) 常务委员会 2024 年度第二次工作会议在 CCF 业务总部&学术交流中心举行，本次工作会议由专委会主任陈熙霖研究员主持会议，常务委员会委员参会，秘书处成员列席。

首先，专委会主任陈熙霖研究员围绕党的二十届三中全会，带领大家学习了二十届三中全会公报的内容。

随后，王瑞平秘书长围绕走进高校、走进企业、视界无限、专委简报、RACV 2024 研讨会、计算机视觉前沿讲习班等专委会特色品牌活动介绍了专委会的工作进展。针对前三季度的活动进展，介绍了专委会后续工作开展的一些规划等。

接下来，常委会委员们围绕专委会新委员增选、专委会奖励、RACV 研讨会征集、PRCV 发展规划、专委会委员参与 CCF 事务、以及专委会各项活动事宜等议题展开了热烈讨论，形成了具体可行的指导性建议。

CCF-CV 常务委员会 2024 年度第二次工作会议顺利召开

最后，会议在热烈的交流讨论氛围中结束。

责任编委 潘金山

专题综述

神经网络剪枝中的对比原则与金标准剪枝的可靠性

西湖大学　王欢

本文试图探讨神经网络剪枝（neural network pruning）领域中的对比原则[1]与金标准剪枝（oracle pruning）的可靠性[2]，及其和目前剪枝领域的多种费解现象之间的关系。限于篇幅，本文更多起到总览的作用，指出关键概念、结论、及其相关的论文，而不赘述细节（细节请参考[1] [2]）。读者可以通过本文了解神经网络剪枝领域的一些重要概念和当前发展的主要问题。水平有限，如有纰漏，欢迎指正。

一、神经网络剪枝基本概念及历史

剪枝（pruning）指将神经网络中不必要的部分移除掉，实际中可以等价为置零（zeroing）操作，所形成的是一个稀疏网络（sparse network），因此在文章中常见的神经网络“剪枝”和“稀疏化”是类似的含义。按照剪枝的粒度（granularity）可以分为结构化剪枝（structured pruning）和非结构化剪枝（unstructured pruning）。前者指整块地移除参数，带来的是缩小版的密集网络（smaller dense network）：宽度（或称通道数、滤波器数）减少或层数减少，有利于实际硬件加速；后者指单个地移除参数，带来的是同样大小但稀疏的网络（same-sized sparse network），有利于参数压缩但不利于硬件加速。剪枝一般分为三阶段：预训练（pretrain）、剪枝（prune）、重训练（retrain）[9]。这三阶段中，第二阶段被广泛认为是最关键的，核心科学问题是寻找定义参数重要与不重要的准则，被称为剪枝准则（pruning criterion）。

剪枝历史久远，几乎和神经网络本身的发展同步。早期的剪枝文章可追溯到至少1988年[23][24][25][26]，LeCun等人早期也曾研究过剪枝（如其经典工作 OBD[9]）。剪枝的早期研究（指1995年前后第二次神经网络衰落之前）和现代研究（指2012年前后深度学习兴起之后）的出发点所有不同：早期关注利用剪枝来达到更好的泛化性能（侧重点在性能）[23][24][25]，而现代研究关注利用剪枝来降低模型训练和推理的代价（侧重点在效率）。

现代剪枝方法的常规操作和观点，如需要重训练（也就是三阶段剪枝流程的第三步）、迭代式剪枝性能更好，在早期就被充分认识到（如OBD[9]）；现代剪枝算法中提出的多种剪枝准则（pruning criterion）也在早期被提出过[4]。从早期到现代，剪枝领域的变化主要是对象从小模型变成如今的深度模型所引发的种种变化，从剪枝算法本身上来说，并没有大的变化。

二、神经网络剪枝现代研究的若干阶段与困局

在深度学习刚兴起时，人们就注意到其网络参数中的冗余性，进而提出剪枝算法来将冗余参数置零。代表性工作由Han等人在2015-2016年奠定[7][8]，其提出的深度压缩[8]算法更是摘得ICLR16最佳论文奖，此时关注的重点在于非结构化剪枝。随后人们进一步追求加速，关注点从全连接层转移到卷积层，因而有大量结构化剪枝文章[17][18][19][20]，各种剪枝准则被提出[21]，剪枝成为小众热门课题。

2019年开始出现新的范式：初始阶段剪枝（pruning at initialization (PaI), or, pruning before training）[13]，即对一个未训练的随机网络进行剪枝。与之对应的传统剪枝范式可以被称为训练后剪枝（pruning after training(PaT)）。新范式的兴起源于两篇文章，SNIP[11]和LTH[3]，后者摘得ICLR19最佳论文

奖，因而使得PaI这一范式迅速兴起。与此同时，多篇论文发现对于传统剪枝范式而言，很多复杂剪枝算法的性能并没有比简单基线方法（幅值剪枝）更好，这是令人费解的一个现象。到大模型时代，模型参数剪枝被认为非常具有挑战性，因为往往需要重训练，而大模型重训练代价巨大，剪枝算法遂转向对令牌（tokens）而非网络参数的剪枝[16]。本文所探讨的依然是模型参数剪枝，而非令牌剪枝，并试图去回答为什么很多复杂剪枝算法的性能并没有比简单基线方法（幅值剪枝）更好。

神经网络剪枝现代研究的困局。目前学界对于剪枝的兴趣相比之前呈下降趋势，主要原因在于剪枝尚未能带来显著的实际收益：

- PaI的困境是，（1）剪枝中的关键在于区分开重要、不重要参数，对一个随机网络来说，其中的参数都是随机初始化的，其重要、不重要的区分是否真的有意义还有待验证。（2）PaI的实际价值在于省掉剪枝三阶段中的第一阶段、从而达到节省训练代价的目的，但到目前为止，所有PaI方法的性能都比传统幅值剪枝方法要差，也就是节省训练成本的代价是性能折损[5]，得不偿失。
- PaT的困境是，（1）近来越来越多的研究者注意到不同剪枝准则得到的网络的性能差异不大，声称“当前最佳”的方法并没有比基线方法（幅值剪枝）要明显更好；（2）PaT中的非结构化剪枝带来的实际加速有限（这一点部分原因是软硬件系统对于稀疏矩阵运算的支持并不充分）；而结构化剪枝由于剪枝粒度较大，性能折损严重，一般都需要重训练，在大模型时代并不适合。

本文试图去解释PaT中的第一点困局：为何不同剪枝准则得到的网络性能差别不大，以及如何合理评估过去几年中剪枝领域的实际发展。

三、神经网络剪枝方法的公平对比原则

在探索不同剪枝方法的重训练阶段学习率设置之前，首先要明确不同剪枝方法的公平对比原则。公平对比是评估任何科技领域发展的基础，但是由于各种原因，在剪枝领域，这种公平对比一直未能很好实现。

Method	#Epochs	LR schedule
SSL [84]$_{\text{NeurIPS'16}}$ (CIFAR10)	–	0.01
L_1-norm [43]$_{\text{ICLR'17}}$	20	0.001, fixed
DCP [90]$_{\text{NeurIPS'18}}$	60	0.01, step (36/48/54)
GAL-0.5/1 [48]$_{\text{CVPR'19}}$	30	0.01, step decay (10/20)
Taylor-FO [56]$_{\text{CVPR'19}}$	~25	0.01, step decay (10/20)
Factorized [44]$_{\text{CVPR'19}}$	90	0.01, step decay (30/60)
CCP-AC [62]$_{\text{ICML'19}}$	100	0.001, step decay (30/60/90)
HRank [47] $_{\text{CVPR'20}}$	30×#layers	0.01, step decay (10/20)
GReg-1/2 [81] $_{\text{ICLR'21}}$	90	0.01, step decay (30/60/75)
ResRep [14] $_{\text{ICCV'21}}$	180	0.01, cosine annealing
L_1-norm [43] $_{\text{ICLR'17}}$ **(our reimpl.)**	90	0.01, step decay (30/60/75)

表 1　不同剪枝方法的重训练阶段采用的超参数

如表 1 所示，不同剪枝方法的重训练阶段采用的超参数有很大差异，尤其是 LR schedule（本文稍后解释为何重训练 LR schedule 值得重点关注）。除此之外，还有更多无关变量影响公平对比，具体可参见文章[4]。

这种混乱的对比情况使得我们难以判断剪枝领域的实际发展程度[4]，为了解决这一问题，我们首先需要定义公平对比原则（Fairness Principle）。在该问题上，我们需要认识到，由于剪枝方法本身的差异，很难只用唯一的对比原则让所有算法进行对比来实现所谓的“公平性”。因此，公平对比原则有层次之分：越相似的算法对比起来我们可控制的程度越高，从而可以更细致地知道一个算法相比于另一个算法好在哪里；而对于差异比较大的剪枝算法，我们无法对比那么详细，因此需要放弃一些比较细节的对比规则而保持整体上的公平性。

最基础的公平对比原则是：总迭代次数一样。总迭代次数一样意味着算法的时间成本（大致）相同（注：此前有工作[13]认为总训练成本的定义是总浮点乘加计算量的个数（FLOPs），由于被剪枝之后网络变小、FLOPs降低，因此为了维持总浮点乘加计算量的个数，重训练阶段的迭代次数就应该按比例延长。我们认为该定义的模糊性太高，在实际中的指导意义不大。譬如实际中我们更关心总训练时长，而FLOPs降低和加速之间的关系并不显著，且严重依赖于具体的软硬件平台，引入FLOPs作为训练成本的度量并没有比用迭代次数更实用，却更难以定义、加大了不同剪枝算法之间的对比难度，反而背离了试图矫正剪枝基准测试的初衷）。对于任何剪枝算法，无论剪枝算法本身如何设计，这一点都可以做到，因此实际意义和可操作性都很高。

第二条公平对比原则：控制无关变量原则。后提出的算法应合理控制变量，和前算法在无关因素上应尽量保持一致。这一原则的贯彻有赖于实践者的经验。下面我们以对比基于重要性（importance-based）的剪枝算法和基于正则化（regularization-based）的剪枝算法为例，来说明这一原则。这两类剪枝算法涵盖了现有多数剪枝方案，因此讨论的必要性和参考价值都很高。基于重要性的剪枝算法的代表是幅值剪枝[7][20]，基于正则化的剪枝算法选择的是递增正则化（GReg-1）[15]。

这里对比的难点在于，幅值剪枝算法是单次剪枝（one-shot pruning），不涉及任何迭代训练，而递增正则化需要花费一定迭代次数来训练网络，这多出来的正则化迭代（regularization iterations）次数需要摊到合适的区间中。有如下两种方案：

- 方案1：把正则化迭代放到预训练阶段，算法模型示意图如图1所示。这样也就假定剪枝算法有权选择从不同时间点的预训练模型开始剪枝（那么，实施剪枝算法的基础模型是不一样的），但这样的好处是可以维持重训练阶段的整个LR schedule完全一致，符合控制无关变量原则。

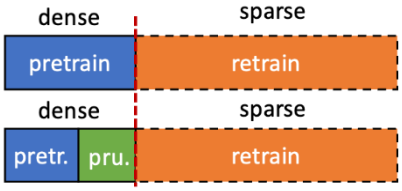


图 1　正则化迭代放到预训练阶段示意图

- 方案2：把正则化迭代放到重训练阶段，示意图如图2所示。这样也就假定剪枝所基于的基础模型已经给定，不能获取到之前的权重，剪枝算法可以设计自己的剪枝过程，但所引入的迭代成本必须摊销在重训练阶段内。这种情况下，基础模型被控制得完全一样，但是无法保证重训练LR schedule一样。

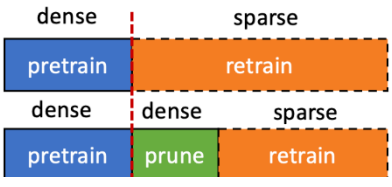


图 2　正则化迭代放到重训练阶段示意图

以上两种对比规则我们认为都是公平的，只是侧重点不同，最终剪枝算法的使用者可以根据他们具体的情况来选择合适的对比设置。

同时，我们也可以看出，在对比基于重要性的剪枝算法与基于正则化的剪枝算法时，控制基础模型和控制重训练过程之间存在冲突，无法保证二者同时一样，根据具体情况有所侧重是必须的：譬如对于模型预训练者而言，他们存有不同阶段的权重，那么方案1、方案2对他们都适用；而对于普通剪枝算法开发者来说，基础模型往往是给定的，因此方案2更适用于他们的情况，方案1则不适用。

剪枝算法开发者也完全有可能提出如下的策略，即如图3所示同时改变了基础模型和重训练过程。那么应该解释为什么所提出的剪枝算法没有控制无关变量，如无必要，不推荐有这样的设计。

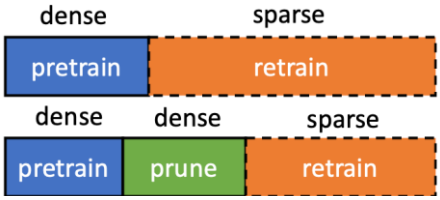


图 3　改变基础模型和重训练过程示意图

剪枝方法对比的现状。遗憾的是，已发表的剪枝算法中，其实连最基础的维持总训练成本的原则也没有达成（如表1所示，有的算法采用20 epochs重训练，而有的则高达180 epochs），这其中有大量历史原因（譬如当事人在开发算法时没有充足的计算资源）。如今来看，我们也很难一一去矫正这些结果（尽管有人在进行类似的尝试，如[4]）。

第三条公平对比原则：最优性能原则。在剪枝三阶段中，应采用已知最优或标准（实际上很难知道什么是最优设置，领域不断发展，最优设置往往在改变，所以这里说的“最优”设置更准确的说法是“不明显差”的设置。但即便如此，如[1]指出，部分文章的超参设置（如重训练学习率）属于“明显差”的类型，加剧了剪枝方法对比混乱）的设置（包括超参数设置）以期达到最优性能。譬如预训练阶段，对于 pytorch 模型来说采用标准的 torchvision 预训练模型是标准操作，使用自训练

的、明显低于 torchvision 预训练模型性能的是不合适的。这一点在早期剪枝文章中存在问题（由于彼时预训练模型开源并不完善，尤其是使用 Caffe 框架的剪枝文章），近些年的剪枝文章中已经比较少出现基础模型不一致的问题。

四、为何重训练阶段学习率的影响如此之大？

随着pytorch等框架流行、开源模型日趋完善，因此在剪枝算法中控制基础模型是容易的。那么，剩下影响对比公平性的主要来源是重训练（retrain）阶段，而本文要强调的一点核心观察就是：既往剪枝方法对于重训练阶段的重视和讨论严重不足。重训练阶段一直被认为不是剪枝算法的核心组成部分而遭到轻视，很多文章中只是一笔带过，超参数等并未详述。下面我们就以学习率为例说明重训练阶段超参数的影响之大及其原因，这其中涉及到和（稀疏）神经网络可训练性（trainability）的关系。

表2是在ImageNet100数据集上剪枝ResNet34模型，基础模型训练了60 epochs，使用幅值剪枝[20]，随后重训练了60 epochs。基础模型一样、所被剪掉的参数位置也完全一样，区别仅在重训练阶段使用不同的剪枝率设置。从表中可见，使用大的学习率可以显著“提升”性能。

这一看似神奇的现象的原因是什么呢？尤其是，大学习率从根本上就比小学习率好吗？

如图4所示，我们通过分析重训练阶段的学习曲线发现，学习率小的模型并非根本上性能差，而是因为收敛速度变慢，导致模型并未被训练充分。如果给与模型充分的训练时间，即使使用小学习率，也能达到和大学习率类似的性能。因此，大学习率并非从根本上“提升”了模型性能，只是加速了收敛，使更高性能被提前看到。

层剪枝率	50%	70%	90%	95%
学习率 0.001	82.90	81.24	77.29	70.53
学习率 0.01	**83.67**	**82.96**	**80.78**	**77.81**
Top1 准确率差异	**+0.77**	**+1.72**	**+3.49**	**+7.28**

表 2　重训练阶段学习率对性能的影响

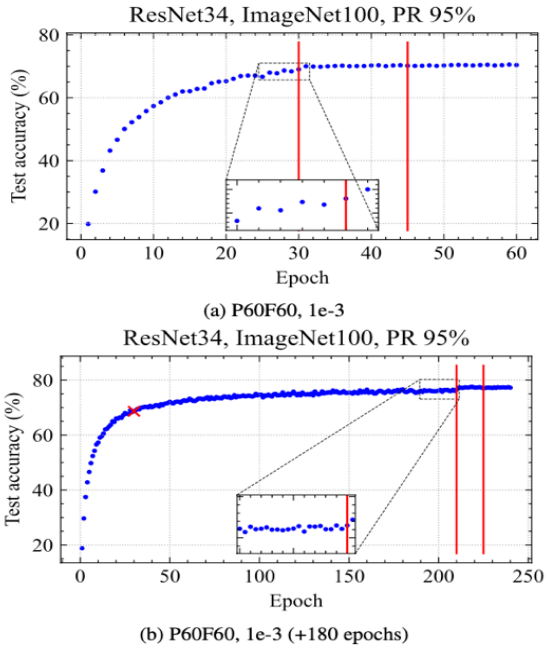


图 4　重训练阶段的学习曲线对比图

上述现象的实际问题是：为什么重训练阶段网络收敛速度变慢？这就是网络可训练性（trainability）的问题：剪枝后的网络由于参数丢失，尽管继承了参数，但表达能力是折损的，且往往由于参数被剪枝，剩余参数的分布不是正常的（近似于正态分布的）分布，进而影响梯度下降过程中的梯度传播，导致收敛变慢。而这一点，在以往的剪枝工作中并未被充分认识到（一些工作认为剪枝后的模型继承了现有参数、训练就类似于“微调”（finetuning），因此不能使用大学习率，应该使用小的学习率，而实际情况与此恰恰相反，由于可训练性变差，收敛速度变慢，采用大学习率更有利于加速收敛，得到更优性能。），导致不恰当地选择了明显不好的超参数，使得幅值剪枝算法的性能被严重低估。

五、为什么不同剪枝准则的差异不大？

在明晰可训练性在神经网络剪枝后的重训练阶段的影响之后，我们可以对幅值剪枝采用更合理的超参数（学习率设置），进而得到了更好的性能。令人意外的是，其性能相当卓越，以致于能和其他更复杂的、基于不同剪枝准则的方法相媲美，如表3所示。

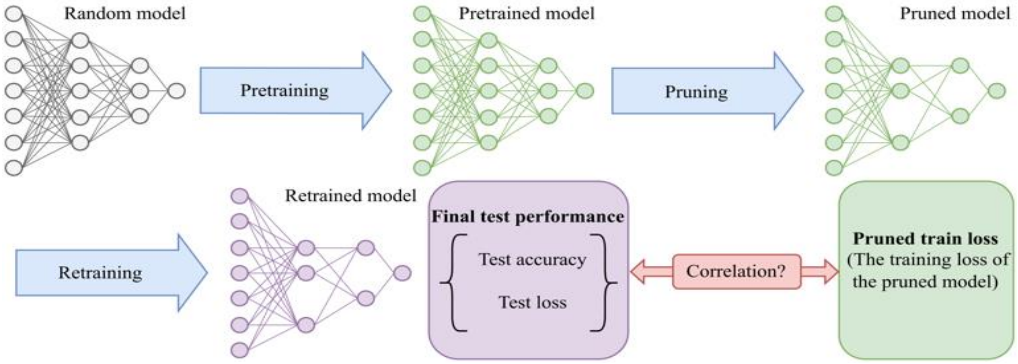


图 5　金标准剪枝可靠性的分析框架示意图

这一观察和多篇论文的结论形成呼应（如[6]），由此我们可以比较肯定地下结论：幅值剪枝，尽管想法简单，但仍是当前最好的剪枝方法（之一）。

那么，接下来要问的问题是：从1990s[9]开始，就不断有研究者提出幅值剪枝过于简单、幅值大小不能准确反映参数重要性（譬如没有考虑参数所在的损失函数曲率等）等论点，进而提出了大量理论上更好的剪枝准则，为何它们实际中却收效颇微？

我们感觉这个问题的根源可能在于剪枝中的第三阶段：重训练阶段。剪枝准则的推导（几乎）都是基于刚剪枝完的、没有经过重训练的模型性能，而实际中，我们最终对比的是经过了重训练的模型性能，这二者之间的关系大吗？这个问题即是对金标准剪枝（oracle pruning）可靠性的质疑。

Method	Pruned acc. (%)	Speedup
SFP [29] IJCAI'18	74.61	1.72×
DCP [90] NeurIPS'18	74.95	2.25×
GAL-0.5 [48] CVPR'19	71.95	1.76×
Taylor-FO [56] CVPR'19	74.50	1.82×
CCP-AC [62] ICML'19	75.32	2.18×
ProvableFP [46] ICLR'20	75.21	1.43×
HRank [47] CVPR'20	74.98	1.78×
GReg-1 [81] ICLR'21	75.16	2.31×
GReg-2 [81] ICLR'21	75.36	2.31×
CC [45] CVPR'21	**75.59**	2.12×
L_1-norm [43] ICLR'17 **(our reimpl.)**	75.24	**2.31**×
GAL-1 [48] CVPR'19	69.88	2.59×
Factorized [44] CVPR'19	74.55	2.33×
LFPC [28] CVPR'20	74.46	2.55×
GReg-1 [81] ICLR'21	74.85	2.56×
GReg-2 [81] ICLR'21	**74.93**	2.56×
CC [45] CVPR'21	74.54	**2.68**×
L_1-norm [43] ICLR'17 **(our reimpl.)**	74.77	2.56×

表 3　重训练超参数完善后的幅值剪枝与其他方法对比

金标准剪枝（oracle pruning）[2]是指将待剪枝的参数置零，记录下造成的损失函数上升程度，损失函数上升越小的参数被认为越不重要，应该优先被剪掉。金标准剪枝的实施需要穷举式对比所有的剪枝情况，这一过程复杂度太大（组合爆炸），因此过往方法都是采用对损失函数进行近似，从而得到对一个或一组参数的重要性分数公式。几乎现存所有的剪枝准则都基于金标准剪枝的思想。

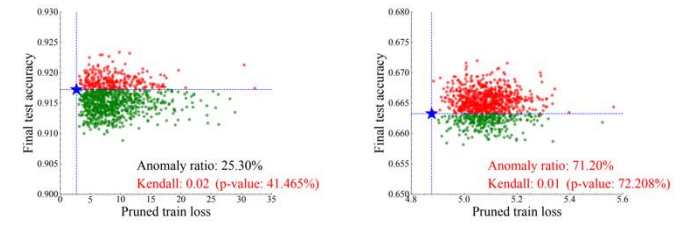
为了对金标准剪枝进行检验，我们对一个模型进行随机剪枝，记录下其重训练前后的性能，得到大量数据点之后，分析相关性。分析框架如图5所示。

我们在不同规模的模型（LeNet5-mini，ResNet56/VGG19，ResNet18，ViT-B/16，TinyLLaVA-3.1B），数据集（MNIST，CIFAR，ImageNet-1K，多模态大模型数据集）进行了实验，训练了37,000多个模型，得到一个令人意外的观察：

对于稍微实用点的模型（从CIFAR数据集上训练的模型起）来说，模型刚剪枝完的准确率和重训练后的准确率之间的相关性相当微弱（如图6所示）。

这一经验性观察意味着金标准剪枝的想法并不成立：用刚剪枝完模型的准确率挑选出来的“好模型”，无法保证其在重训练完之后也是“好模型”。进而，基于金标准剪枝思想得到的各种剪枝准则的有效性也因此值得怀疑。

以上结论或可解释为什么不同剪枝准则（pruning criterion）尽管在理论上比幅值剪枝更优但实际上性能



(1) ResNet56 / CIAFR10 (Pr. 0.5, 1K samples) (2) VGG19 / CIFAR100 (Pr. 0.5, 1K samples)

图6 剪枝完重训练前、后的模型准确率散点图

却差不多：我们“期待”更复杂的剪枝准则能够带来更优性能，这一“期待”本身就是不成立的。

六、结论与讨论

本文开篇介绍了神经网络剪枝的基本概念、历史、以及当前困局；接着解释了不同剪枝方法的公平对比原则，以便正确评估剪枝领域的发展；随后讨论剪枝领域对于重训练阶段的轻视所带来的问题，指出可训练性在重训练阶段的影响；最后解释在公平对比下为什么很多更复杂的剪枝准则性能却没有比简单的幅值剪枝明显更优，并指出一个惊人观察：目前复杂剪枝准则的基础——金标准剪枝——对于现代模型的剪枝来说本身就是不成立的。

金标准剪枝不成立的观察对于无需重训练的剪枝方法没有影响，而对于需要重训练的剪枝方法的影响颇大，可能动摇其剪枝准则的理论根基。在未来，对于需要重训练的剪枝方法，继续基于金标准剪枝提出新的剪枝准则的意义似乎不大。剪枝准则可以回归到简单的幅值剪枝（且已经经验性地观察到幅值剪枝的效果可以媲美复杂的剪枝准则），而更多地去考量重训练阶段的可训练性问题，以带来新的性能突破或简化整个剪枝算法的流程、增强其扩展性和易用性。

责任编委 魏秀参

参考文献

[1] Wang, H., Qin, C., Bai, Y., and Fu, Y. Why is the state of neural network pruning so confusing? on the fairness, comparison setup, and trainability in network pruning. arXiv preprint arXiv:2301.05219, 2023.

[2] Sicheng Feng, Keda Tao, and Huan Wang. Is Oracle Pruning the True Oracle? Under Review, 2024.

[3] Jonathan Frankle and Michael Carbin. The lottery ticket hypothesis: Finding sparse, trainable neural networks. In ICLR, 2019.

[4] Davis Blalock, Jose Javier Gonzalez, Jonathan Frankle, and John V Guttag. What is the state of neural network pruning? In MLSys, 2020.

[5] Jonathan Frankle, Gintare Karolina Dziugaite, Daniel M Roy, and Michael Carbin. Pruning neural networks at initialization: Why are we missing the mark? In ICLR, 2021.

[6] Trevor Gale, Erich Elsen, and Sara Hooker. The state of sparsity in deep neural networks. arXiv preprint arXiv:1902.09574, 2019.

[7] Song Han, Jeff Pool, John Tran, and William J Dally. Learn- ing both weights and connections for efficient neural network. In NeurIPS, 2015.

[8] Song Han, Huizi Mao, and William J Dally. Deep com- pression: Compressing deep neural networks with pruning, trained quantization and huffman coding. In ICLR, 2016.

[9] Y. LeCun, J. S. Denker, and S. A. Solla. Optimal brain dam- age. In NeurIPS, 1990.

[10] B. Hassibi and D. G. Stork. Second order derivatives for network pruning: Optimal brain surgeon. In NeurIPS, 1993.

[11] Namhoon Lee, Thalaiyasingam Ajanthan, and Philip Torr. Snip: Single-shot network pruning based on connection sensitivity. In ICLR, 2019.

[12] Namhoon Lee, Thalaiyasingam Ajanthan, and Philip Torr. Snip: Single-shot network pruning based on connection sensitivity. In ICLR, 2019.

[13] Zhuang Liu, Mingjie Sun, Tinghui Zhou, Gao Huang, and Trevor Darrell. Rethinking the value of network pruning. In ICLR, 2019.

[14] Wang, H., Qin, C., Zhang, Y., and Fu, Y. Recent advances on neural network pruning at initialization. In IJCAI, 2022.

[15] Wang, H., Qin, C., Zhang, Y., and Fu, Y. Neural pruning via growing regularization. In ICLR, 2021.

[16] Keda Tao, Can Qin, Haoxuan You, Yang Sui, Huan Wang. DyCoke: Dynamic Compression of Tokens for Fast Video Large Language Models. arXiv preprint arXiv:2411.15024, 2024.

[17] Wen, W., Wu, C., Wang, Y., Chen, Y., and Li, H. Learning structured sparsity in deep neural networks. In NeurIPS, 2016.

[18] He, Y., Zhang, X., and Sun, J. Channel pruning for accelerating very deep neural networks. In ICCV, 2017.

[19] Molchanov, P., Tyree, S., and Karras, T. Pruning convolu- tional neural networks for resource efficient inference. In ICLR, 2017.

[20] Li, H., Kadav, A., Durdanovic, I., Samet, H., and Graf, H. P. Pruning filters for efficient convnets. In ICLR, 2017.

[21] Molchanov, P., Mallya, A., Tyree, S., Frosio, I., and Kautz, J. Importance estimation for neural network pruning. In CVPR, 2019.

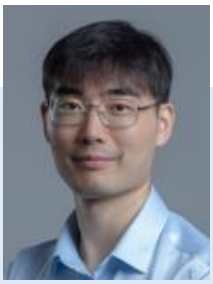
[22] Janowsky, S. A. Pruning versus clipping in neural networks. Physical Review A, 39(12):6600, 1989.

[23] Mozer, M. C. and Smolensky, P. Skeletonization: A technique for trimming the fat from a network via relevance assessment. In NeurIPS, 1988.

[24] Baum, E. and Haussler, D. What size net gives valid generalization? In NeurIPS, 1988.

[25] Chauvin, Y. A back-propagation algorithm with optimal use of hidden units. In NeurIPS, 1988.

[26] Hanson, S. and Pratt, L. Comparing biases for minimal network construction with back-propagation. In NeurIPS, 1988.



王 欢

王欢，浙江大学学士、硕士，美国东北大学博士。2024 年 6 月加入西湖大学工学院任助理教授，创立高效智能计算实验室（Efficient Neural Computing and Design Lab, ENCODE Lab），担任独立 PI、博导。研究领域为高效人工智能、神经渲染、计算机视觉；专注于高效人工智能相关的理论、算法、应用研究，致力于让前沿 AI 算法落地。曾在 Google / Snap / MERL / Alibaba 等业界研究机构实习。发表顶会顶刊论文 30 余篇。在西湖大学教授《计算机和程序设计基础》本科生通识课程。

Email: wanghuan@westlake.edu.cn

热点追踪

融合特征轨迹的相机全局平移估计方法

陶沛霖　崔海楠　荣梦琪　申抒含
中国科学院自动化研究所

本文是中国科学院自动化研究所三维视觉研究组团队的研究成果，发表在 CVPR 2024 的工作 HETA[1]。在三维视觉领域中，准确估计摄像机位姿并从图像集生成场景点云是基础性任务，在自动驾驶[20]、增强现实[18]、和神经辐射场[19]等领域，具有广泛的应用。一般而言，运动恢复结构算法作为实现这些目标的一种常见且有效的方法脱颖而出。论文研究的问题是如何通过融合特征轨迹约束，解决全局式运动回复结构算法中在短基线、相机近似共线运动场景下，平移平均精度低鲁棒性差的问题。先前的方法要么仅依赖相机间的相对平移，在相机近似共线移动时无法准确估计相机间的相对尺度；要么仅依赖特征轨迹约束，对特征轨迹中的错误匹配敏感。针对上述问题，团队提出一种新颖的混合约束显式平移平均框架(HETA)，通过混合使用相对平移和特征轨迹约束，同时显式优化场景三维点。通过在顺序 KITTI 测距基准和互联网无序数据集 1DSfM 上测试，发现我们的方法表现超过了许多最先进的全局运动恢复结构算法。

一、引言

从运动恢复结构方法的基本步骤是首先通过特征点检测和匹配构建一个场景图[12,13]，图中节点表示摄像机，边连接具有充足特征匹配的摄像机。随后，估计摄像机位姿并三角化得到场景三维结构。

通常摄像机位姿估计的主要方法是采用增量式模式，例如 COLMAP[10]。该方法首先通过精心选择图像对来创建初始模型。然后，使用 PnP 算法[17]注册包含足够数量 2D-3D 对应关系的图像。最后，通过迭代的方法，包括三角测量、捆绑调整 (BA) [11]和 PnP 步骤，同时估计场景结构和摄像机位姿。尽管增量式方法在精度和对抗异常值的稳健性方面表现出色，但它们对图像注册顺序的变化较为敏感，可能导致误差积累和漂移。此外，重复的非线性捆绑调整显著影响效率，使其不适用于大规模场景。为解决增量式方法中的这些问题，全局式方法[2-9]被提出，通过从摄像机的相对位姿估计其全局旋转和平移以实现对所有摄像机进行注册，随后，对场景结构进行三角测量和优化，其全部流程如图 1 所示。由于仅需一次的捆绑调整优化，从而显著提高了效率，并实现了所有相机间均匀的误差分布。具体而言，摄像机的全局姿态$\boldsymbol{R}_i, \boldsymbol{t}_i$和相对姿态$\boldsymbol{R}_{ij}, \boldsymbol{t}_{ij}$满足以下方程：

$$\boldsymbol{R}_j\boldsymbol{R}_i^T = \boldsymbol{R}_{ij}, \quad \frac{\boldsymbol{t}_i - \boldsymbol{t}_j}{\|\boldsymbol{t}_i - \boldsymbol{t}_j\|_2} = \boldsymbol{R}_j^T\boldsymbol{t}_{ij} = \boldsymbol{v}_{ij} \qquad (1)$$

其中符号$\boldsymbol{v}_{ij}$表示全局坐标系中的相对平移。对于全局旋转估计，现有基于李代数的方法[15]已经得到充分研究。然而，由于相对平移估计对低视差特征匹配敏感[9]，并且存在尺度不确定性问题，这使得全局平移估计比全局旋转估计更为困难。仅依赖相对平移的方法局限于注册在平行刚性图中的摄像机，并在摄像机经历共线运动时遇到退化问题。即使摄像机运动轨迹几乎共线，相对平移中的轻微扰动也可能导致估计摄像机位置发生显著变化，使得仅使用相对平移无法实现准确的估计。

为解决这些挑战，一些方法将特征轨迹的约束纳入目标函数中。根据在优化过程中是否估计了特征轨迹对应的三维点，这些方法可以被分为隐式方法和显式方法。大多数隐式方法估计摄像机基线尺度，或者利用来自隐式三维点的相机到点的约束，但这些方法都对相对平移的异常值敏感。例如，LiGT[8]旨在仅使用特征轨迹构建约束。然而，由于特征轨迹通常表现出比相对平移更高

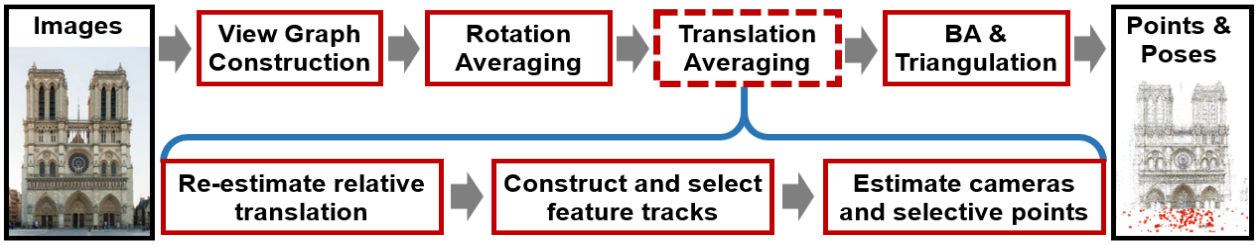


图 1　使用 HETA 的全局式 SfM 方法流程框架。其中 HETA 算法的流程为：给定相机的全局姿态，首先重新估计相机相对平移，然后构造并挑选一些特征轨迹，最后使用一种鲁棒的目标函数同时估计相机和部分三维点的位置。

的外点率，因此其鲁棒性仍然不足。1DSfM[6]作为一种经典的显式方法，同时利用摄像机到摄像机和摄像机到点的约束，以估计三维点和摄像机位置，但在处理特征轨迹异常值时也会产生嘈杂的解。1DSfM[6]的失败主要归因于使用不恰当的目标函数和不准确的观测值作为输入。在本研究中，我们重新审视了这些问题，并引入了一种新颖的混合约束显式平移平均框架(HETA)。

我们的贡献涵盖了三个关键方面：(1) 将使用特征轨迹的全局平移估计方法分为显式和隐式方法，并重新审视它们的优势和劣势。(2) 对两种形式的线性目标函数进行了比较分析，并引入了一种新颖的混合约束显式方法，通过鲁棒的L_1范数优化，随后进行无偏的L_2范数优化，在两个步骤中同时估计摄像机和点。(3) 为了提高相对平移的准确性，我们使用极线几何中的共面约束对其进行了重新估计。为了增强这种重新估计鲁棒性，我们分析了视差角影响，并滤除了不稳定特征匹配。

二、融合特征轨迹的平移平均相关工作

我们重新审视了显式和隐式方法，并对它们的优势和劣势进行了彻底分析。假设从三维点$\boldsymbol{P}_k$发出的光线在n个相机平面上生成n个投影特征点。对于一个特征点，我们将其在局部归一化摄像机坐标系中的坐标表示为$\boldsymbol{X}_{ki}^T=(x_{ki},y_{ki},1)^T$，其中k是特征轨迹或三维点的索引，i是相机的索引。在全局相机坐标系中，三维点和相机之间的关系满足：

$$\frac{\boldsymbol{P}_k-\boldsymbol{t}_i}{\|\boldsymbol{P}_k-\boldsymbol{t}_i\|_2}=\boldsymbol{R}_i^T\frac{\boldsymbol{X}_{ki}}{\|\boldsymbol{X}_{ki}\|_2}=\boldsymbol{f}_{ki} \tag{2}$$

其中t_i表示相机位置，f_{ki}表示从相机t_i到三维点P_k的归一化特征射线。从公式（1）和公式（2）可知，相机到相机约束和相机到点约束的数学表达式是等价的。显式方法的核心思想是使用与估计相机平移相同的目标函数

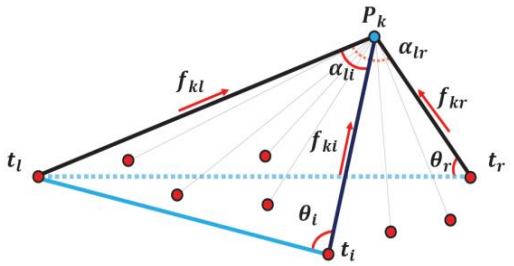


图 2 摄像机与摄像机和摄像机与三维点间的约束示例。其中红色的点表示摄像机，蓝色点表示一个一个特征轨迹对应的三维点，红色的箭头代表特征射线的方向，基相机$\boldsymbol{t_l},\boldsymbol{t_r}$在特征轨迹中拥有最大的视差角。

来估计三维点。与低视差场景下误差较大的相对平移相比，特征射线作为从图像中导出的原始信息，自然表现出更高的精度。因此，在显式方法中使用特征射线在理论上可以提供比仅依赖相对平移的方法更高的精度。隐式方法主要通过两种方式约束相机。一类方法利用特征点的深度一致性来计算相机基线尺度。另一类方法用已有的观测量如特征射线和相对平移线性表示三维点，并基于这些三维点及其相应的特征射线来约束相机。对于第一类方法，我们以 2016 年 Cui 等人[2]的方法为例。如图 2 所示，通过三维点到其可见相机的连接构造了两个相邻三角形$\{P_k-t_l-t_r\}$和$\{P_k-t_l-t_i\}$。根据正弦定理，两个相机基线尺度的比例为：

$$\frac{\|\boldsymbol{t_l}-\boldsymbol{t_r}\|_2}{\|\boldsymbol{t_l}-\boldsymbol{t_i}\|_2}=\frac{\sin\theta_i\cdot\sin\alpha_{lr}}{\sin\theta_r\cdot\sin\alpha_{li}} \tag{3}$$

然而，在低视差场景中，这种方法对异常值非常敏感。一方面，在低视差场景中，相对平移估计不准确，导致诸如θ_i和θ_r等角度不正确。另一方面，分母中的低视差角，例如α_{li}，导致数值不稳定性。这意味着视差角的轻微变化会导致比率计算中的显著变化。

对于第二类隐式方法，为看到共同场景点的相机导出了线性约束。然而，在 2015 年 Cui 等人[3]的方法和 PGILP[9]中，对三维点的表示仍然依赖于相对平移，其误差累积到所表示的三维点中。为了解决这些问题，2021 年 Cai 等人提出了一种 LiGT[8]约束，通过在每个特征轨迹中选择具有最大视差角的两个基相机中的特征射线，线性表示其对应的三维点。如图 2 所示，对于一个特征轨迹，其中两个基相机t_l, t_r的，特征点在相机 l 中的深度计算为：

$$\begin{aligned}\|\boldsymbol{P_k}-\boldsymbol{t_l}\|_2 &= \frac{\|\boldsymbol{t_l}-\boldsymbol{t_r}\|_2 \cdot sin\,\theta_r}{sin\,\alpha_{lr}} \\ &= \frac{\|\boldsymbol{f_{kr}}\times(\boldsymbol{t_l}-\boldsymbol{t_r})\|_2}{\|\boldsymbol{f_{kl}}\times\boldsymbol{f_{kr}}\|_2} \\ &= \frac{\left((\boldsymbol{f_{kl}}\times\boldsymbol{f_{kr}})\times\boldsymbol{f_{kr}}\right)\cdot(\boldsymbol{t_l}-\boldsymbol{t_r})}{\|\boldsymbol{f_{kl}}\times\boldsymbol{f_{kr}}\|_2^2\cdot}\end{aligned} \quad (4)$$

由公式 (4) 可知，三维点可由两个基相机及其特征射线表示：

$$\boldsymbol{P_k}=\boldsymbol{t_l}+\frac{\boldsymbol{f_{kl}}\left((\boldsymbol{f_{kl}}\times\boldsymbol{f_{kr}})\times\boldsymbol{f_{kr}}\right)^T}{\|\boldsymbol{f_{kl}}\times\boldsymbol{f_{kr}}\|_2^2}(\boldsymbol{t_l}-\boldsymbol{t_r}) \quad (5)$$

结合公式（2），可以通过隐式表达三维点和其对应特征轨迹里其他相机（如相机 i）建立相机和点之间的约束。

我们从两个方面比较显式与隐式方法。在鲁棒性方面，显式方法通常优于隐式方法，因为在显式方法中，每个特征轨迹的三维点是通过所有特征射线进行估计的，而在隐式方法中，三维点仅由两个基相机的特征射线表示。在效率方面，尽管隐式方法避免了显式地优化三维点，但新引入摄像机到点约束增加了相机之间的连接性，从而破坏了传统全局平移平均方法的优化矩阵的稀疏特性。通过实验发现显式方法与隐式方法效率相当。

三、混合约束显式平移平均框架(HETA)

尽管引入特征轨迹约束带来了很多好处，但当特征轨迹中存在很多异常值时，求解的相机平移也会包含较大的噪声。与特征射线相比，相对平移可以在摄像机之间提供更直接和严格的约束。为了增强鲁棒性和效率，与 PGILP[9]和 LiGT[8]等方法使用所有的特征轨迹不同，我们不但利用相对平移来提供相机之间直接约束，而且选择更可靠特征轨迹提供相机与点之间约束。

具体思路为，首先构建一个场景及轨迹图$\mathrm{G}=\{V\cup P, E_v\cup E_p\}$，其中 V 中的每个节点表示一个相机，P 中的每个节点表示一个三维点，E_v中的每个边连接 V 中的相机对，E_p中的每个边表示相机到三维点的特征射线。设$\mathbb{C}$为相机到相机约束，$\mathbb{P}$为相机到点约束。目标函数可以表达为:

$$\min_{V;P}\sum_{E_v}\rho\left(\left|\|\mathbb{C}\|\right|_p\right)+\sum_{E_p}\rho\left(\left|\|\mathbb{P}\|\right|_p\right) \quad (6)$$

其中 p 表示优化范数，ρ(·)表示鲁棒估计函数。考虑到鲁棒性，有三个主要任务：(1) 增强低视差场景中相对平移的精度；(2) 选择可靠的特征轨迹；(3) 为两种类型的约束定义鲁棒的目标函数。我们将在接下来几小节中解决这些任务，并在最后给出我们方法的完整框架。

3.1 相对平移的重估计

在对极几何中，共面约束可以表示为$\boldsymbol{X_{kj}}\cdot(\boldsymbol{t_{ij}}\times\boldsymbol{R_{ij}}\boldsymbol{X_{ki}})=\boldsymbol{0}$。在给定全局相机姿态时，可以改写为：

$$\begin{aligned}&(\boldsymbol{R_i^T}\boldsymbol{X_{ki}}\times\boldsymbol{R_j^T}\boldsymbol{X_{kj}})\cdot\boldsymbol{R_j^T}\boldsymbol{t_{ij}}=\boldsymbol{0} \\ &\Leftrightarrow(\boldsymbol{f_{ki}}\times\boldsymbol{f_{kj}})\cdot\boldsymbol{v_{ij}}=\boldsymbol{0}.\end{aligned} \quad (7)$$

由公式 (7) 可知，每个相对平移可以使用极线平面法向量重新估计，该法向量由$\boldsymbol{f_{ki}}\times\boldsymbol{f_{kj}}$计算得到。由于相机内参数和全局旋转的不准确性，从特征射线估计的法向量不可避免地具有一些角度误差。LUD[5]中提出一种方法通过最小化相对平移和法向量之间余弦角来使用所有归一化法向量重新估计相对平移。然而，由于法向量的准确性也受到视差角的影响，因此在估计过程中对每个法向量采用相同的权重是不合理的。与 LUD[5]中归一化法向量$\boldsymbol{f_{ki}}\times\boldsymbol{f_{kj}}$的方法相比，通过理论推导和实验发现，在估计过程中使用非归一化的法向量估计相对平移，可以为每个特征匹配保留合理的权重$\|\boldsymbol{f_{ki}}\times\boldsymbol{f_{kj}}\|_2=\sin\alpha$，改进的目标函数如下所示：

$$\min_{v_{ij}}\sum_k\rho\left(\|(\boldsymbol{f_{ki}}\times\boldsymbol{f_{kj}})\cdot\boldsymbol{v_{ij}}\|_2\right)\quad \text{s.t.}\quad \|\boldsymbol{v_{ij}}\|_2=1 \quad (8)$$

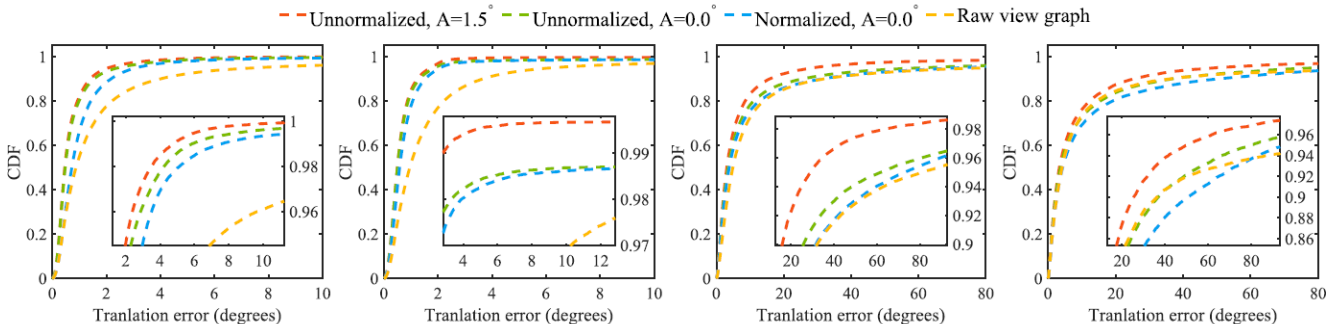


图 3　在部分 KITTI[16]和 1DSfM[6]数据集上，比较原始场景图、是否归一化法向量，以及过滤小视差角相对平移后相对平移角度误差的累积分布函数结果，从左至右，数据集分别为 KITTI-06,KITTI-09,1DSfM-PIC,1DSfM-ROF。

此外，当低视差角的法向量误差变得非常大时，基于共面性一致性来估计相对平移或验证特征匹配变得无效。因此，在估计相对平移之前，我们预先定义一个阈值 A 过滤视差角低于该阈值的特征匹配。

3.2 特征轨迹的挑选策略

通过重新估计相对平移，我们筛选出违反共面性或可见性约束的特征匹配，并使用并查集算法构建特征轨迹。由于特征轨迹可能包含高比例的特征射线异常值，因此只选择部分特征轨迹以提高效率和鲁棒性。根据公式 (8)，具有较大视差角的特征匹配的共面性一致性更可靠。因此，根据它们最大视差角，对所有特征轨迹进行降序排序。然后，依次检查每个特征轨迹，确定它是否能与覆盖次数不足的图像之间建立连接。此过程一直持续，直到所选特征轨迹子集至少覆盖所有相机 N 次。

3.3 优化目标函数的定义

相机到相机和相机到点的约束都可以表示为以下公式：$\boldsymbol{s_i - s_j = \|s_i - s_j\|_2 \cdot s_{ij}}$，其中$\boldsymbol{s_i, s_j}$表示相机或点，$\boldsymbol{s_{ij}}$表示从$\boldsymbol{s_j}$到$\boldsymbol{s_i}$的已知归一化向量，例如特征射线或相对平移。我们比较了两种线性目标函数，包括叉积形式$\|\boldsymbol{s_{ij}}\times(\boldsymbol{s_i - s_j})\|_2$和尺度形式$\|\boldsymbol{s_i - s_j} - \lambda_{ij}\boldsymbol{s_{ij}}\|_2$，其中$\lambda_{ij}$是一个尺度变量，×表示为叉乘。为了消除尺度和方向的不确定性，分别对叉积形式和尺度形式的目标函数使用不等式约束$\boldsymbol{s_{ij}\cdot(s_i - s_j)} \geq 1$和$\lambda_{ij} \geq 1$。

假设$\boldsymbol{s_{ij}^G}$是$\boldsymbol{s_{ij}}$的真值。在大多数情况下，当$\boldsymbol{s_{ij}\cdot s_{ij}^G} \geq \boldsymbol{0}$时，两个不等式约束都定义了正确的可行区域。在最优解的情况下，两个目标函数中残差的大小是相同的。与此同时，尺度形式中的λ_{ij}等于$\boldsymbol{s_{ij}\cdot(s_i - s_j)}$，表示$\boldsymbol{s_i - s_j}$在$\boldsymbol{s_{ij}}$上的投影的大小。因此，$\lambda_{ij}$是一个冗余变量，因为它完全由当前的$\boldsymbol{s_i, s_j}$和已知的$\boldsymbol{s_{ij}}$确定。此外，当尺度变量的变化范围很大，例如在基线长度或特征射线深度差异较大的情况下，它们通常难以收敛到最优解。这显著影响了实际优化的整体精度和效率。

当$\boldsymbol{s_{ij}}$存在显著误差导致$\boldsymbol{s_{ij}\cdot s_{ij}^G < 0}$时，尺度形式中的$\lambda_{ij}$等于最低界限 1，以最小化惩罚。叉积形式的不等式约束提供了一个错误的可行区域，导致一个偏置的解决方案。然而，通过我们的相对平移重新估计，整体相对平移的精度得到提升。此外，1DSfM[6]过滤方法通过多个随机的一维投影很好地过滤方向误差较大的相对平移，可以过滤掉大多数存在显著误差的相对平移。因此，使用叉积形式的目标函数以获得更好的收敛性。

3.4 优化框架

为了避免来自特征射线的冗余和不正确的约束，仅使用相机到相机的不等式约束来消除尺度和方向的模糊性。为了提高鲁棒性，目标函数在L_1范数下使用 ADMM 方法[14]进行优化，如下所示：

$$\min_{\substack{\boldsymbol{t_i}, i\in V;\\ \boldsymbol{P_k}, k\in P}} \sum_{\boldsymbol{ij}\in \boldsymbol{E_v}} \|\boldsymbol{v_{ij}}\times(\boldsymbol{t_i - t_j})\|_1 + \sum_{\boldsymbol{ki}\in \boldsymbol{E_p}} \|\boldsymbol{f_{ki}}\times(\boldsymbol{P_k - t_i})\|_1$$
$$\boldsymbol{s.t.} \quad \sum_{\boldsymbol{i}\in V} \boldsymbol{t_i} = \boldsymbol{0}, \quad \boldsymbol{v_{ij}\cdot(t_i - t_j)} \geq \boldsymbol{1}, \quad \forall \boldsymbol{ij} \in \boldsymbol{E_v} \qquad (9)$$

其中这两个约束被用来消除内在的位置和尺度模糊。然而，正如 BATA[7]中所提到的，对于相机基线和特征点深度尺度不同的目标函数，这个优化结果是有偏的。因此，可以使用一个无偏的基于角度的目标函数来进一步优化当前的结果。在每次迭代中，使得：$\hat{\boldsymbol{v}}_{\boldsymbol{ij}} = \frac{\boldsymbol{t_i - t_j}}{\|\boldsymbol{t_i - t_j}\|_2}$，

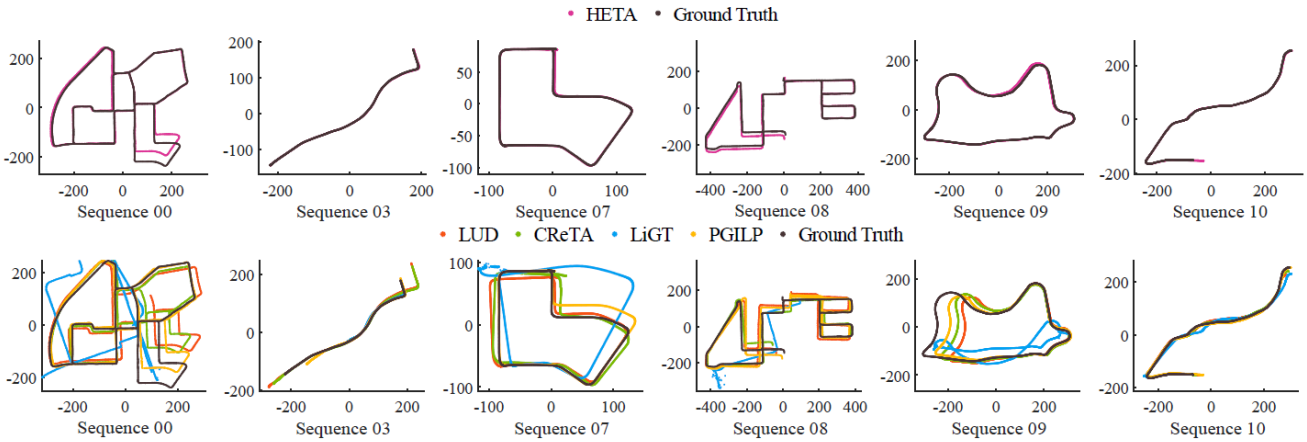


图 4　KITTI 数据集[16]上估计的相机运动轨迹。对比方法包括 LUD[5]、CReTA[4]、PGILP[9]和 LiGT[10]。

$\hat{\boldsymbol{f}}_{\boldsymbol{ki}} = \frac{\boldsymbol{P_k - t_i}}{||\boldsymbol{P_k - t_i}||_2}$，然后，使用 IRLS 方法来优化目标函数。

$$\min_{\substack{\boldsymbol{t_i}, i\in V;\\ \boldsymbol{P_k}, k\in P}} \sum_{ij\in E_v} \boldsymbol{\rho}\left(\boldsymbol{H}(\hat{\boldsymbol{v}}_{ij})\right) + \sum_{ki\in E_p} \boldsymbol{\rho}\left(\boldsymbol{H}(\hat{\boldsymbol{f}}_{ki})\right), \quad \boldsymbol{s.t.} \sum_{i\in V} \boldsymbol{t_i} = \boldsymbol{0},$$
$$\text{where} \quad \boldsymbol{H}(\hat{\boldsymbol{s}}_{ij}) = \begin{cases} \|\boldsymbol{s}_{ij}\times\hat{\boldsymbol{s}}_{ij}\|_2, & \boldsymbol{s}_{ij}\cdot\hat{\boldsymbol{s}}_{ij} \geq \boldsymbol{0}; \\ \boldsymbol{1}, & \boldsymbol{s}_{ij}\cdot\hat{\boldsymbol{s}}_{ij} < \boldsymbol{0}. \end{cases} \tag{10}$$

四、实验结果

4.1 相对平移的重估计

我们进行实验证明了使用未归一化法线向量和过滤低视差角特征匹配对重新估计的相对平移的影响。在公式（9）中，误差损失宽度β设置为$\sin 1° \cdot \sin 5°$，表明期望当视差角α等于1°时，错误角度γ应小于5°。如图 3 所示，相对平移误差的累积分布函数（CDF）表明使用未归一化法线向量显著提高了相对平移的准确性。通过滤视差角低于 1.5 度的特征匹配，进一步提高了准确性。如图 3 所示，1DSfM-ROF 数据中，由于相机全局旋转精度不足，使用归一化法线向量估计的相对平移的准确性不佳。然而，在这种情况下，我们的方法仍然产生了更好的结果。

4.2 有序数据集上的评估结果

KITTI 数据集[16]是使用安装在驾驶汽车上的两台摄像机收集的。因此，大多数特征匹配的视差角受到限制，相机运动轨迹倾向于近似共线，这对全局平移估计系统提出了重大挑战。估计的相机运动轨迹如图 4 所示。HETA 实现了最高的准确性。尽管使用了非常精确的相对平移，LUD[5]和 CReTA-BATA[4]都难以产生准确的结果。LiGT[8]方法基于矩阵分解估计相机平移，提高了效率但缺少鲁棒性。相比之下，PGILP[9]通过在L_1范数下优化每个相机到点的约束产生更好的结果。通过这些比较，我们可以得出结论，我们的方法 HETA 在准确性和鲁棒性方面均超过了所有比较方法。

Data		LUD			CReTA-BATA			LiGT			PGILP			1DSfM			HETA						
		BA			BA			BA			BA			BA			L_1		L_2		BA		
Name	N_t	$\tilde{e}$	$\bar{e}$	N_c	$\tilde{e}$	$\bar{e}$	N_c	$\tilde{e}$	$\bar{e}$	N_c	$\tilde{e}$	$\bar{e}$	N_c	$\tilde{e}$	$\bar{e}$	N_c	$\tilde{e}$	$\bar{e}$	$\tilde{e}$	$\bar{e}$	$\tilde{e}$	$\bar{e}$	N_c
ALM	497	0.1	0.5	483	0.1	0.4	487	0.3	1.8	422	0.1	0.5	486	0.3	4.6	389	0.5	1.2	0.5	1.2	0.1	0.4	**488**
ELS	217	0.2	0.4	212	0.2	0.4	215	0.2	0.4	204	0.2	0.4	**216**	0.2	0.4	196	2.4	3.9	2.1	3.8	0.2	0.4	**216**
GDM	590	0.1	3.4	560	0.1	3.6	561	5.1	1e3	504	0.2	4.1	556	0.4	66.5	475	2.8	10.4	2.2	10.1	0.2	2.7	**564**
MDR	178	0.2	6.3	168	0.2	5.6	170	8.6	16.4	137	0.2	9.7	170	0.8	9.8	122	1.4	9.7	1.4	9.6	0.2	7.0	**174**
MND	403	0.1	0.1	399	0.1	0.1	399	0.1	0.1	383	0.1	0.1	398	0.1	0.3	363	0.5	1.0	0.5	1.0	0.1	0.1	**400**
ND	479	0.1	0.6	457	0.1	0.3	468	6.2	7.3	397	0.1	0.3	462	0.1	47.5	374	0.3	1.4	0.3	0.9	0.1	0.3	**476**
NYC	296	0.1	0.2	290	0.1	0.3	**294**	0.1	1.5	222	0.1	0.2	285	0.1	5.7	261	0.7	1.8	0.5	1.5	0.1	0.1	290
PDP	295	0.1	0.4	287	0.1	0.1	286	7.4	2e2	108	0.1	0.3	290	0.1	1.9	249	1.1	2.9	1.1	2.9	0.1	0.2	**291**
PIC	1838	0.1	0.5	1797	0.1	0.5	**1811**	12.3	81.8	649	0.1	0.5	1774	0.5	3.1	1621	0.9	1.9	0.7	1.7	0.1	0.4	1807
ROF	918	0.1	0.2	875	0.1	0.2	899	0.6	3.1	732	0.1	0.5	892	0.8	23.5	725	1.9	3.9	1.2	3.3	0.1	0.1	**907**
TFG	3989	0.9	2.5	3864	0.7	1.8	3913	37.5	45.8	789	1.2	4.3	3860	12.1	18.9	3348	3.3	6.4	2.6	5.8	0.7	2.4	**3951**
TOL	396	0.5	3.6	**391**	0.2	4.8	387	70.3	76.0	152	0.3	4.3	380	3.2	7e2	276	2.5	4.9	2.1	4.5	0.4	1.5	387
USQ	637	0.3	2.8	582	0.3	4.4	603	6.7	1e2	336	0.5	4.5	602	0.4	1e2	505	4.2	7.5	3.6	7.2	0.2	2.1	**619**
VNC	713	0.2	5.7	672	0.2	9.7	**702**	20.5	28.2	474	0.2	9.1	664	0.2	4.5	556	1.8	4.2	1.7	4.0	0.1	0.8	686
YKM	337	0.1	0.1	327	0.1	0.1	329	0.1	0.3	318	0.1	0.2	323	2.9	23.9	262	1.2	2.4	1.1	2.1	0.1	0.2	**333**

表 1　1DSfM 数据集上相机位置误差。N_t表示场景图中摄像机的数目，N_c为捆绑调整(BA)后图像的注册数目，其最优结果被加粗表示。$\tilde{e}$和$\bar{e}$分别表示摄像机位置误差的中值和均值，单位为米。

4.3 无序数据集上的评估结果

1DSfM 数据集[6]是通过许多不同类型的摄像机收集的。由于提供的相机内参数的有限精度和大量错误的特征匹配，估计的相对姿态存在较大误差。因此，由Chatterjee方法估计[14]的全局旋转的精度低于KITTI数据集[16]，使得全局平移估计更具挑战性。BA 后的标定结果如表 1 所示。从这个比较中，LUD[5]、CReTA-BATA[4]和 HETA 估计的相机位置的准确性相当。然而，对于大多数数据，HETA 注册的图像数量最多，表明与 LUD[5]和 CReTA-BATA[4]相比，它具有更高的鲁棒性。对于依赖于所有特征轨迹作为输入的隐式方法 LiGT[8]和 PGILP[9]，它们的性能不及 HETA。这种差异是由于在优化过程中包含了大量来自特征轨迹外点的错误约束。

五、总结

我们重新审视了利用特征轨迹进行全局平移估计的问题，并提出了一种新颖混合约束显式平移平均框架。我们的方法在顺序和无序数据集上表现优越，超过了许多现有的最先进方法。然而，特征匹配异常值的普遍存在仍然限制了全局运动恢复结构算法广泛应用。在未来，我们期望利用神经网络，从数据中学习更多先验信息，以提高特征轨迹的匹配精度，提升算法的鲁棒性。

责任编委 王金甲

参考文献

[1] Tao P, Cui H, Rong M, et al. Revisiting Global Translation Estimation with Feature Tracks. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2024: 20686-20696.

[2] Hainan Cui, Shuhan Shen, and Zhanyi Hu. Robust global translation averaging with feature tracks. In IEEE International Conference on Pattern Recognition, pages 3727–3732, 2016.

[3] Zhaopeng Cui, Nianjuan Jiang, Chengzhou Tang, and Ping Tan. Linear global translation estimation with feature tracks. In British Machine Vision Conference, pages 46.1–46.13, 2015.

[4] Lalit Manam and Venu Madhav Govindu. Correspondence reweighted translation averaging. In European Conference on Computer Vision, pages 56–72. Springer, 2022.

[5] Onur Ozyesil and Amit Singer. Robust camera location estimation by convex programming. In IEEE Conference on Computer Vision and Pattern Recognition, pages 2674–2683, 2015.

[6] Kyle Wilson and Noah Snavely. Robust global translations with 1dsfm. In European Conference on Computer Vision, pages 61–75. Springer, 2014.

[7] Bingbing Zhuang, Loong-Fah Cheong, and Gim Hee Lee. Baseline desensitizing in translation averaging. In IEEE Conference on Computer Vision and Pattern Recognition, pages 4539–4547, 2018.

[8] Qi Cai, Lilian Zhang, Yuanxin Wu, Wenxian Yu, and Dewen Hu. A pose-only solution to visual reconstruction and navigation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1):73–86, 2021.

[9] Liyang Liu, Teng Zhang, Brenton Leighton, Liang Zhao, Shoudong Huang, and Gamini Dissanayake. Robust global structure from motion pipeline with parallax on manifold bundle adjustment and initialization. IEEE Robotics and Automation Letters, 4(2):2164–2171, 2019.

[10] Johannes Lutz Schonberger and Jan-Michael Frahm. Structure-from-motion revisited. In IEEE Conference on Computer Vision and Pattern Recognition, pages: 4104-4113,2016.

[11] Bill Triggs, Philip F McLauchlan, Richard I Hartley, and Andrew W Fitzgibbon. Bundle adjustment—a modern synthesis. In Vision Algorithms: Theory and Practice: International Workshop on Vision Algorithms, pages 298–372. Springer, 2000.

[12] Federica Arrigoni, Andrea Fusiello, Elisa Ricci, and Tomas Pajdla. Viewing graph solvability via cycle consistency. In IEEE International Conference on Computer Vision, pages 5540–5549, 2021.

[13] Federica Arrigoni, Tomas Pajdla, and Andrea Fusiello. Viewing graph solvability in practice. In IEEE/CVF International Conference on Computer Vision, pages 8147–8155, 2023.

[14] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends® in Machine learning, 3 (1):1–122, 2011.

[15] Avishek Chatterjee and Venu Madhav Govindu. Efficient and robust large-scale rotation averaging. In IEEE International Conference on Computer Vision, pages 521–528, 2013.

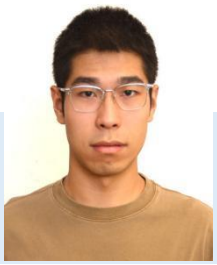
[16] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. The International Journal of Robotics Research, 32(11):1231–1237, 2013.

[17] Xiao-Shan Gao, Xiao-Rong Hou, Jianliang Tang, and Hang-Fei Cheng. Complete solution classification for the perspective-three-point problem. IEEE Transactions on Pattern Analysis and Machine Intelligence, 25(8):930–943, 2003.

[18] Rodrigo Chac´on Quesada and Yiannis Demiris. Design and evaluation of an augmented reality head-mounted display user interface for controlling legged manipulators. In IEEE International Conference on Robotics and Automation, pages 11950–11956, 2023.

[19] Linning Xu, Yuanbo Xiangli, Sida Peng, Xingang Pan, Nanxuan Zhao, Christian Theobalt, Bo Dai, and Dahua Lin. Grid-guided neural radiance fields for large urban scenes. In IEEE Conference on Computer Vision and Pattern Recognition, pages 8296–8306, 2023.

[20] Eric Brachmann, Martin Humenberger, Carsten Rother, and Torsten Sattler. On the limits of pseudo ground truth in visual camera re-localisation. In IEEE International Conference on Computer Vision, pages 6218–6228, 2021.



陶沛霖

中国科学院自动化研究所硕士生，研究方向：全局式运动恢复结构，三维重建。
Email: taopeilin2023@ia.ac.cn

崔海楠

中国科学院自动化研究所副研究员，研究方向：大规模场景三维建模，从运动恢复结构和视觉定位。
Email: hncui@nlpr.ia.ac.cn

荣梦琪

中国科学院自动化研究所助理研究员，研究方向为三维场景理解，包括细粒度三维语义分割、多模态三维分割基础模型等。

Email: mengqi.rong@ia.ac.cn

申抒含

中国科学院自动化研究所研究员，研究领域为三维计算机视觉理论与应用，包括大规模场景三维重建、智能机器人三维环境感知、场景三维语义理解等。

Email: shshen@nlpr.ia.ac.cn

顶会观察

ACM MM 2024

厦门大学 吴垚 曲延云

国际计算机图形学与多媒体领域会议（ACM International Conference on Multimedia，ACM MM）是计算机视觉和模式识别领域最重要的会议之一。这一会议由 ACM 主办，自 1993 年首次举办以来，已成为该领域内学术界和工业界交流的重要平台，也是中国计算机学会（CCF）推荐的 A 类国际学术会议。ACM MM 专注于多媒体技术领域的最新研究成果、技术创新和行业趋势。会议涵盖了多媒体内容的创建、处理、传输和交互等多个方面，旨在促进学术界和工业界在多媒体技术应用和产品开发方面的交流与合作。

一、会议概况

ACM MM 每年举办一次，此次为第 32 届多媒体国际学术会议。于 2024 年 10 月 28 日~ 11 月 1 日在澳大利亚墨尔本（Melbourne）市召开。本次会议的投稿数量、审稿人数和参会人数都是前所未有的。首先，ACM MM 2024 公布了今年的论文录用结果：2024 年共有 4385 篇投稿进入审稿阶段，相比去年增加了 42%，最终 1149 篇论文被录用，录取率为 26.20%。而在录用的投稿中，仅有 174 篇被进一步评选为 Oral，其接受率为 3.97%。会议议程紧凑而丰富，涵盖了多媒体技术领域的多个关键议题。会议举办了多场主题演讲、研讨会、技术教程等活动，深入探讨了机器学习与人工智能、多语言处理、云计算与虚拟化等前沿话题。

二、特邀报告

笔者聆听并记录了 3 个特邀报告。来自香港科技大学的 Pascale Fung 教授带来了主题为《From Assistants to Agents in the LLM Era》的演讲。自 20 世纪 90 年代以来，AI 助手已经融入我们的生活。基于大语言模型（LLMs）构建的 AI 代理改变了我们进行研究、开发和部署的方式。这些代理是具备不同程度推理和规划能力的 AI 系统，能够预见用户需求，自主确定满足这些需求的步骤，访问大语言模型之外的知识和工具，以完成用户的任务。它们标志着大语言模型时代的新一代 AI 助手。在本次演讲中，Pascale Fung 教授概述 AI 助手向 AI 代理演进的过程，当前仍面临的挑战以及未来开发涵盖不同模态和界面的代理家族所带来的机遇。随着未来不同模型架构的出现，LLMs 将继续作为构建代理的有效工具，取代手动设计。

随后，来自悉尼大学的 Judy Kay 教授带来了主题为《Empowering People to Harness and Control their Multimodal Data》的演讲。随着技术的发展，日益丰富的数字设备生态系统渗透我们的生活中，这些设备能够捕获大量的长期个人数据。在本次演讲中，Judy Kay 教授分享一系列案例研究，他指出利用他的研究可以“创建系统和界面，使人们能够掌握和控制这些数据及其使用”，在此基础上介绍了未来计划。首先，第一组案例探讨了如何利用来自可穿戴设备（如智能手表）的数据，为个人信息化界面提供支持，帮助我们长期了解自身情况，包括对大型数据集（超过 14 万人）的分析，以及用于锻炼的虚拟现实游戏。其次，第二组案例来自正式教育环境，针对来自个人数据界面的数据进行学习，探讨了开放学习者模型（Open Learner Models，简称 OLMs）是如何利用学习数据的。最后，Judy Kay 教授分享一些关键见解，这些见解对于研究议程具有重要意义，包括：终身学习的 OLMs；快速思考与慢速、审慎思考所需的不同界面的本质；传达不确定性；帮助人们从多模态数据中真正了解自己；以及这些如何与人

工智能时代教育面临的紧迫挑战，如假新闻和真相衰减等问题相联系。

最后，来自罗切斯特大学的 Jiebo Luo 教授带来了主题为《Multimodal LLMs as Social Media Analysis Engines》的演讲。近期的 AI 研究揭示了多模态大型多模态模型（Multimodal Large Multimodal Models，简称 MLMMs）在各种通用视觉和语言任务中的卓越能力。对于 MLMMs 在更专业领域的表现，学术界的兴趣日益浓厚。社交媒体内容本质上是多模态的，融合了文本、图像、视频，有时还包括音频。要有效理解此类内容，模型必须解释这些不同通信模式之间的复杂互动及其对所传达信息的影响。理解社交多媒体内容仍然是当代机器学习框架面临的一个挑战性问题。为了评估 MLMMs 在社交多媒体分析方面的能力，Jiebo Luo 教授选择了五个代表性任务进行概述，包括情感分析、仇恨言论检测、假新闻识别、人口统计推断以及政治意识形态检测。并介绍了研究使用现有基准数据集对每个任务进行初步定量分析，随后是对结果的仔细审查以及选择能够展示 GPT-4V 在理解多模态社交媒体内容方面潜力的定性样本。GPT-4V 在这类任务中表现出显著的效果，展示了诸如图像-文本对的联合理解、情境和文化意识以及广泛常识知识等方面的优势。除了已知的幻觉问题外，Jiebo Luo 教授进一步介绍了几种尝试，以改善某些任务的性能。并强调了 MLMMs 在未来通过分析多模态信息来加深我们对社交媒体内容及其用户理解方面的光明前景。

三、ACM MM 颁奖

在本届 ACM MM 被评选为 Oral 论文中，有 26 篇论文被提名为 ACM MM 2024 最佳论文。其中，由杭州电子科技大学、中国科学院计算所、杭州电子科技大学丽水研究院、澳大利亚阿德莱德大学和麦考瑞大学合作的论文《From Speaker to Dubber: Movie Dubbing with Prosody and Duration Consistency Learning》[1]荣获 ACM Multimedia 2024 最佳论文奖。考虑到电影配音数据集规模有限（由于版权问题）和背景噪声的干扰，直接从电影配音数据集中学习限制了学习模型的发音质量。为了解决这个问题，作者提出了一

图 1　最佳论文奖汇报现场

种两阶段的配音方法，让模型先学习发音知识，然后再进行电影配音练习。

最佳学生论文奖分别由蒙纳士大学、科廷大学、印度理工学院、科廷理工大学合作的论文《AV-Deepfake1M: A Large-Scale LLM-Driven Audio-Visual Deepfake Dataset》[2]和由南加利福尼亚大学，谷歌合作的论文《An In-depth Study of Bandwidth Allocation across Media Sources in Video Conferencing》[3]荣获。前者针对嵌入真实视频中的音视频操控小片段定位问题，模拟了此类内容生成的过程，并提出了 AV-Deepfake1M 数据集。该数据集包含针对超过 2000 个个体的内容驱动型（i）视频操控、（ii）音频操控以及（iii）音视频联合操控，总计生成超过 100 万段视频。本文对所提出的生成数据流进行了详尽描述，并附有对生成数据质量的严格分析，为构建下一代深度伪造定位方法发挥至关重要的作用。后者通过分析 Zoom、Webex 和 Google Meet 等平台的带宽分配策略，重点关注其对用户体验质量（Quality of Experience, QoE）的影响。作者提出了一种通用的 QoE 预测模型。该研究是一项开创性的工作，旨在评估不同场景和网络条件下的多媒体传输效果，超越了以往仅关注单一媒体类型的研究所能达到的范围。

图 2　各类最佳论文颁奖现场

最佳 demo 奖由延世大学的论文《DanceMimic: Awaken Your Dancing Instinct through a Real-time Dance Imitation Capture System》[4]荣获。除此之外，还有 7 篇论文获得荣誉奖。来自中国的研究者获得了其中 6 个。

四、总结

在 ACM MM 2024 录用的全部论文中，第一作者来自中国大陆的论文占到整体录取论文近 70%左右。这充分说明，我国当前计算机图形学与多媒体领域的研究已经走到了世界舞台的前列，让国际计算机学界听到了更多来自中国的声音。笔者衷心希望肩负使命的研究者在多媒体处理、机器学习、计算机视觉等多个方面做出更多出彩的成绩，对中国乃至世界的科技发展产生更加深远的影响。

责任编委 **崔海楠**

参考文献

[1] Zhang, Z., Li, L., Cong, G., Yin, H., Gao, Y., Yan, C., van den Hengel, A., & Qi, Y. (2024). From Speaker to Dubber. Movie Dubbing with Prosody and Duration Consistency Learning. In MM '24 (pp. 7523-7532). ACM. https://doi.org/10.1145/3664647.3680777

[2] Cai, Z., Ghosh, S., Adatia, A.P., Hayat, M., Dhall, A., Gedeon, T., & Stefanov, K. (2024). AV-Deepfake1M: A Large-Scale LLM-Driven Audio-Visual Deepfake Dataset. In MM '24 (pp. 7414-7423). ACM. https://doi.org/10.1145/3664647.3680795

[3] Zejun Zhang, Xiao Zhu, Anlan Zhang, and Feng Qian. 2024. An In-depth Study of Bandwidth Allocation across Media Sources in Video Conferencing. In MM'24. (pp.7696–7704).ACM. https://doi.org/10.1145/3664647.3681007

[4] Seongjean Kim, Jungwoo Huh, Yeseung Park, Jungsu Kim, and Sanghoon Lee. 2024. DanceMimic: Awaken Your Dancing Instinct through a Real-time Dance Imitation Capture System. In MM'24 .(pp.11267–11269).ACM. https://doi.org/10.1145/3664647.3684991

吴 垚

厦门大学信息学院，博士研究生。主要研究方向计算机视觉，涵盖 3D 场景理解，多模态学习等方向。

Email: wuyao@stu.xmu.edu.cn

曲延云

厦门大学信息学院计算机科学与技术系教授、博导，教工党支部书记（全国样板支部），中国自动化学会混合智能专委会副主任兼秘书长，CCF 高级会员， IEEE Senior Member。获得福建省自然科学二等奖，入选“2023 年度科学影响力排行榜”全球前 2%顶尖科学家。长期从事计算机视觉、模式识别研究，发表论文 150 余篇，其中中国计算机学会推荐的 CCF-A 类期刊会议论文五十余篇，研究成果发表在计算机视觉顶刊 TPAMI 和 IJCV，图像处理、机器学习领域 JCR 主流期刊 TIP、TNNLS、TYCB、PR、TKDE 等，领域顶级会议 CVPR、ICCV、NeurIPS、AAAI、IJCAI、ECCV、ACM MM。

Email: quyanyun@gmail.com

南京邮电大学周全教授访谈

2024 年 12 月 23 日，《CCF-CV 专委简报》在线采访了南京邮电大学博士生导师周全教授。下面是采访实录。

问题 1：周老师，您好！首先，请您分享一下您的个人学习和研究经历。

感谢您的采访，与很多从国外回国的优秀“海龟”不同，我属于地地道道的“土鳖”！我本科毕业于中国地质大学（武汉），学习电子信息工程专业。硕士和博士都毕业于华中科技大学，师从刘文予教授。由于我是湖北鄂州人，和当时还在 UCLA 的朱松纯教授是老乡，所以博士期间在朱老师创建的莲花山研究院有过短暂的学术访问。首先，要感谢刘老师和朱老师的指导，使我逐步了解了计算机视觉和模式识别领域的国际学术前沿，也使我热爱上了这门学科；同时博士阶段的学习经历也使得我掌握了独立从事科研的能力，为后面的理论研究工作打下了坚实的基础。2013 年博士毕业后，我入职了南京邮电大学，在通信与信息工程学院从事教学科研工作，致力于将我的学术和行业经验传授给学生，并持续推动计算机视觉领域的研究进展。可能是为了弥补在求学过程中没有国外经历的短板吧，在工作期间，我先后访问了瑞典于默奥大学、日本九州工业大学，以及美国天普大学。这些国外的访问经历不仅丰富了我的科研阅历，也为我积累了广泛的人脉和志同道合的合作伙伴。

问题 2：您的学习和研究经历比较丰富，能否分享一下在此过程中令您难忘的一些片段呢？

要说最令我难忘的事，还是 2020 年底即将结束美国访学的时候，由于新冠疫情原因我回国的事情一波三折，差点没能回得来。其实我 2020 年年中就买了国航年底回国的票，但当时国内入境管理非常严格，我回国的航班正好熔断了。为了能够顺利回来，我紧急联系了留学基金委，他们帮我及时改签了达美的航班。但是后来由于美国人要过感恩节，他们那趟航班最终也停飞了。我那个时候公寓也退了，酒店又不敢去住。好在一位朋友介绍到一家本地华人家里临时过渡了一周，才改签到美联航回来。当时一张经济舱的票都 4 到 5 万人民币，是平时的 10 多倍，最后留学基金委也帮我们报销了。我后来得知国家为了让像我们这样的国外访学老师和博士生回国，花了很多钱。这即表明了国家对我们这些科研人员的爱护和重视，也让我们真实体会到国家的强大！这件事情我经常讲给我身边的朋友和我的学生听，同时也激励我自身，要把在国外所学报效和回馈给国家，才不负国家对我的培养！

问题 3：您有过美国、日本、瑞典三个国家高校的访学经历，能对这三个国家在研究环境上的差异做一些评价么？

我在美国访问的时间比较长，有一年时间。我在日本和瑞典访问的时间比较短，分别只有 2 个月和 4 个月。如果说三个国家在科研环境上的差异我觉得可能和东西方国家的文化差异有关。日本的学术等级制度还是很独特的，学生见到老师，或者资历浅的老师见到资深老师都是毕恭毕敬的，见面会鞠躬打招呼，显得特别恭敬。瑞典和美国属于西方国家，相对比较开放和随意一

些。老师和学生之间属于亦师亦友的关系，彼此之间也可以开玩笑。另外，我还发现一个有意思的现象。日本高校学院内不同团队老师之间的交流很少，好像他们都是各做各的，而美国和瑞典比较注重学科交叉，比如我在瑞典访问期间，所在的学院就是工程和自动化学院，里面就包含机械、控制、视觉、信号处理等很多相关方向，美国天普大学也是这样。但不管怎样，他们都很重视技术的落地和实际应用。

问题 4：您的研究工作深耕于计算机视觉中的目标检测、语义分割和图像理解等任务。请问，在这些多维度的探索中，您认为自己的研究工作具备哪些独到的特色与优势？

是的，我从博士阶段就一直从事图像/视频语义理解方面的研究工作。2013 年博士毕业之后，正是卷积神经网络开始兴起的时候，当时我就果断从传统的图像标注方法转到利用卷积神经网络解决目标检测和语义分割任务。实际上，很多理论上的问题也是从实际应用中来的，比如我们在一些实际的横向课题中，发现客户需要在终端设备中部署一些神经网络模型。但是这些终端设备受到算力，存储空间和能耗等多方面的限制，导致传统的目标检测和语义分割网络很难部署。因此，我们又转向设计轻量化的网络模型以满足实际的需要。我们团队也是国内较早开展轻量化工作的团队之一。这期间，我们团队相继提出了 LEDNet、DPNet，以及 BANet 等一系列代表性工作。前一段时间，我们团队 2019 年提出的 LEDNet 还首次获得了 IEEE ICIP 的最有影响力论文奖，该工作率先提出了轻量级非对称编码器-解码器网络架构，在语义分割精度、实时性、模型大小以及计算复杂度等各个维度都取得了较好的平衡。此外，我们团队还致力于解决图像/视频理解中非常具有挑战性的工作，也就是俗称的“硬骨头、卡脖子”问题，如弱监督语义分割，开放集语义分割，以及噪声标签下的语义分割问题等等。目前视觉-语言大模型的轻量化也是未来研究的必然趋势。您可以看出我们的工作既注重理论前沿，又重视解决实际的应用问题。

问题 5：您入选了江苏省科技副总，这应该来源于您在成果转化方面的成就，能介绍一下您在成果转化方面的一些成果么？未来您在这方面又是如何规划的？

前面已经提到我们和很多企业都有横向课题的联系，但说来也惭愧，我是 2022 年才入选的江苏省科技副总。如果说在这些成果中有哪些是让我印象最深刻的，应该是我们在 2018 年和华为合作的“Research on new algorithm for robust deep neural networks”项目，这也是我们课题组的第一个横向合作的成果转化项目。当时华为希望将一些深度卷积模型写到芯片中的 crossbar 结构单元中，但是发现在模型写入和读取的过程中存在读入和写入的误差问题。于是，他们找到我们课题组，看能不能帮他们解决这个问题。我们从知识蒸馏的角度，通过约束 crossbar 读写参数和模型真实参数的一致性来解决读写误差的问题，取得了较好的效果。说起来，这个项目和华为海思还有一定关系，可以说是华为备胎计划的一部分。当时并没有想到第二年美国就对华为实施了全面制裁，我们也为能够为国内科技进步贡献一部分力量而感到自豪。

目前，我们也在积极转化图像理解领域的成果到相关领域。我们已经和江苏省人民医院、中国电信、四川省农科院开展合作。未来还会在智慧农业，远程医疗和自动驾驶领域持续发力。

问题 6：您不仅在学术领域深耕细作，还荣任了国际知名期刊如 Pattern Recognition 和 Computer and Electrical Engineering 的编辑，以及 IEEE Transactions on Multimedia 等多个期刊的客座编辑，同时担任了 70 多个国际 SCI 的期刊审稿人，以及一系列国际知名学术会议和论坛的技术委员会主席和领域主席。请问从这些丰富多样的职务与活动中，您积累了哪些宝贵的经验？又有哪些深刻的感悟，能够激励和启迪学术界同仁？

好的，很高兴和大家分享。首先，要确立好服务学术社会的意识，想到才能够做到，下一步才是如何做到。其实我觉得年轻老师或者博士需要把握好科学研究和

学术兼职之间的度。对于青年学者而言，需要把主要精力放在研究工作一线上面，修好内功，因为这个阶段正是个人一生中学术思想的活跃期。同时，以学术活动和社会兼职为辅，通过开拓学术视野、开展学术交流，正向推动自身的研究工作，并逐渐积累相关学术兼职的经验和人脉，修好外功。我自己承担 PR 和 CAEE 的编委，组织一些 IEEE Trans 的特刊，包括现在也在申请 IEEE TNNLS 的编委，也是这么一路走过来的。当然这也只是我的一家之言，仅供各位老师借鉴参考。

问题 7：近年来，您已在国际学术顶级期刊（IEEE TIP/TITS/TMM/TMI/TNNLS，PR 等）和国际重要学术会议（IEEE ICASSP，IEEE ICIP，ICPR）上发表 SCI/EI 论文 90 余篇，其中 ESI 高被引论文和热点论文各 1 篇。于 2024 年获得 IEEE ICIP 最具影响力论文奖，2024 年、2022 年和 2007 年分别获 IEEE/SPIE ISAIR 最佳展示论文奖、会议突出贡献奖和最佳学生论文奖，以及 2018 年获 EAI ROSENET 国际会议最佳论文奖，指导研究生获得江苏省人工智能学会优秀硕士论文提名奖。请问您如何做到在论文方面如此高产出的，能分享一下您的经验么？

惭愧，和很多优秀的老师相比，我也不是那么“高产”。但我始终坚信在读博期间我的导师刘文予教授教导我们的话：“任何事情都不是一蹴而就的，都是慢慢积累，水到渠成的。”虽然工作后我们团队开展了很多新的研究方向，但几乎都是以图像/视频语义理解为中心展开的。这既有利于把一个问题做深做透，又有利于成果持续产出、积累和沉淀。另外一点经验就是要多交流，无论是参加学术研讨会还是出国访学，往往不同思想的激烈碰撞会产生很好的 idea。再一个就是要解决实际问题，我们组很多好的工作也是从实际问题中来的。前面提到的获得 IEEE ICIP 最具影响力论文奖的工作就是很好的例子。

问题 8：您曾入选南京“青蓝工程”青年骨干教师，主讲《信号与系统》、《数字图像处理》和《深度学习与计算机视觉》课程，请问能分享一下您的教学理念和教学方法么？

其实教学和科研并不矛盾，两者相辅相成、相互促进。我长期讲授《信号与系统》课程，喜欢把一些图像处理、计算机视觉中的一些例子引入到这门课程中，这样学生学起来也不会那么枯燥。在《深度学习与计算机视觉》这门课程中，我喜欢先讲一些实验效果，给同学们一些直观上的感受和认识，再讲背后的原理，这样可以吸引学生的兴趣。学生一旦感兴趣，就愿意加入我的团队，作为硕士或博士研究生，从事相关方向的研究工作，最终做到科研促进教学。

问题 9：在繁忙的工作之余，您有哪些爱好，以给自己放松和充电呢？同时，您又是如何平衡工作与个人家庭生活，确保两者和谐共生的？

我平时喜欢游泳，闲暇下来也会爬爬山。平时工作虽然很忙，但是我会尽可能挤出时间来陪伴家人，周末或者节假日一家人出去郊游，做一些亲子活动，享受美好生活之后可以更好地投身到事业中去。

问题 10：如果吐露研究工作者的心声，您最想说的是什么？

就以我读博士期间实验室墙上的一个座右铭和大家分享一下吧。与各位同仁共勉：科学的道路上没有平坦的大路可走，只有那些在崎岖小路上攀登不畏劳苦的人，才有希望到达光辉的顶点！

责任编委 余烨 赵振兵

周全

南京邮电大学通信与信息工程学院教授，博士生导师。美国天普大学和日本九州工业大学兼职教授。研究领域包括图像/视频理解，神经网络轻量化及其在实际中的应用。目前担任中国计算机学会杰出会员、IEEE 高级会员、IAPR 高级会员、图像图形学学会高级会员、CCF-CV 执行委员、CAA-PRMI 专委会委员、CAAI-PR 专委会委员、CSIG-MV 专委会委员、CSIG-VBD 专委会委员、江苏省人工智能学会常务委员。先后承担国家自然科学基金、江苏省自然科学基金、华为科技项目等课题 30 余项，参与国家自然科学基金重点项目和重大研发计划。先后入选江苏省科技副总，江苏省“青蓝工程”青年骨干教师等。近年来，在国际学术顶级期刊（IEEE TIP/TITS/TMM/TMI/TNNLS、PR 等）和国际重要学术会议（IEEE ICASSP/ICIP、ICPR）发表 SCI/EI 论文 90 余篇，其中 ESI 高被引论文和热点论文各 1 篇。于 2024 年获得 IEEE ICIP 最具影响力论文奖，2024 年、2022 年和 2007 年分别获 IEEE/SPIE ISAIR 最佳展示论文奖、会议突出贡献奖和最佳学生论文奖，以及 2018 年获 EAI ROSENET 国际会议最佳论文奖，指导研究生获得江苏省人工智能学会优秀硕士论文提名奖。授权国家发明专利 10 余项。目前担任国际 SCI 期刊 Pattern Recognition、Computers & Electrical Engineering 副编辑，以及 IEEE Transactions on Fuzzy Systems、IEEE Transactions on Multimedia、Multimedia Tools And Applications、Visual Intelligence 等 SCI 期刊客座编辑。

委员好消息

✪ 2024 年 4 月 24 日，北京市人民政府做出了关于 2023 年度北京市科学技术奖励的决定，CCF-CV 专委会 7 位执行委员的 5 项成果获奖：北京航空航天大学**徐迈**等完成的“人类视觉系统启发的视频感知质量优化理论与方法”获自然科学一等奖，中国科学院心理研究所**王甦菁**、**李婧婷**等完成的“微表情表达机制与稀疏深度运动感知研究”获自然科学二等奖，北京大学**连宙辉**等完成的“中国文字的字体智能计算方法与自动生成关键技术”和北京邮电大学**明悦**等完成的“面向宽带网络的跨媒体感知计算技术与应用”获技术发明二等奖，中国科学院空天信息创新研究院**孙显**、北京科技大学**殷绪成**等完成的“多体制卫星遥感数据智能解译关键技术与重大应用”获科技进步一等奖。

✪ 2024 年 8 月，亚洲青年科学家基金项目（Asian Young Scientist Fellowship）发布了在生命科学、物质科学、数学与计算机科学三大领域的 12 位 2024 年度研究员名单，CCF-CV 专委会执行委员、清华大学**黄高**入选。

✪ 2024 年 9 月 28 日，2024 年度“CCF 科技成果奖”评选结果发布，CCF-CV 专委会 4 位执行委员的成果获奖，国防科技大学**郭裕兰**等完成的“大规模三维几何数据的学习、理解与生成理论与方法”、重庆邮电大学**肖斌**等完成的“不确定性知识的多粒度表示与发现理论与方法”获自然科学一等奖，大连理工大学**贾旭**、**卢湖川**等完成的“动态可引导的视觉内容生成与增强”获自然科学二等奖。

✪ 2024 年 10 月 9 日，MICCAI2024 最佳论文奖揭晓，CCF-CV 专委会执行委员、上海科技大学**马月昕**指导的论文 RoCoSDF: Row-Column Scanned Neural Signed Distance Fields for Freehand 3D Ultrasound Imaging Shape Reconstruction 获最佳论文奖。

✪ 2024 年 10 月 30 日，IEEE ICIP2024 最具影响力论文奖揭晓，CCF-CV 专委会执行委员、南京邮电大学**周全**指导的论文 LEDNet: A Lightweight Encoder-Decoder Network for Real-Time Semantic Segmentation 获最具影响力论文奖。

✪ 2024 年 11 月 17 日，在 2024 世界青年科学家峰会上，第十八届中国青年科技奖正式揭晓。CCF-CV 专委会执行委员、哈尔滨工业大学（深圳）**聂礼强**入选。

✪ 2024 年 12 月 5 日，2024 年度 IAPR Fellows 公布，CCF-CV 专委会执行委员、厦门大学**纪荣嵘**、浙江大学**李玺**、南京邮电大学**刘青山**、浙江大学**杨易**、中国科学院自动化研究所**张兆翔**、中山大学**郑伟诗**当选。

责任编委 刘海波

基于深度学习的医学图像分割方法及其开源代码

兰州理工大学 李策 张建伟

医学图像分割算法作为一种重要的医学图像处理技术，其旨在将医学图像中正常组织器官及病变区域检测并区分出来，并从分割的区域分析病变在图像上的表现特征，使临床诊断的准确性和可靠性得到有效提高，为患者的病情判断、疾病的临床诊疗和预后管理等提供可靠的依据。在临床诊断中，医学图像分割主要依赖放射科医生的经验和主观判断，这不仅耗时，而且影响分割结果的准确性和一致性。因此，自动准确的脑肿瘤分割对于提升临床诊断的效率和准确性至关重要。

近年来，相较于传统方法，基于深度学习的医学图像分割方法具有更高的准确性和鲁棒性，现已成为研究人员进行医学图像分析的热点。鉴于此，本文总结并介绍了最近几年流行的医学图像分割方法及其开源代码。

1、U-Mamba 方法

介绍： 卷积神经网络(Convolutional Neural Network, CNN)和 Transformer 是生物医学图像分割领域的主流方法，但由于卷积的固有局部性问题和 Transformer 的计算复杂性问题，这两类方法在处理长距离依赖关系方面存在一定的局限性。为了解决这一问题，本文将 CNN 和 Mamba 模型结合，提出了一种用于 2D 和 3D 医学图像的通用分割方法，称为 U-Mamba。U-Mamba 结合了卷积和状态空间序列模型的优势，可以捕获局部特征的同时进行全局特征提取，并且相较于基于 Transformer 的方法有更小的计算复杂度。此外，U-Mamba 具有自配置机制，能够在无需人工干预的情况下自动适应各种数据集，在 CT 和 MRI 图像中 3D 腹部器官分割、内窥镜图像中的器械分割以及显微镜图像中的细胞分割四种任务上均优于现有的基于 CNN 和 Transformer 的分割方法。下面将对 U-Mamba 医学图像分割方法进行简单介绍。

U-Mamba 方法的整体框架如图 1 所示，是一个基于编码器-解码器的“U”形结构。其中编码器使用图 1(a)所示的 Mamba 模块构建，以捕获局部特征和远程依赖关系。解码器由残差模块和转置卷积组成，进行特征提取和特征分辨率恢复，同时使用跳跃连接，将编码器和解码器的多尺度特征连接起来，然后进行上采样，最终生成分割结果。

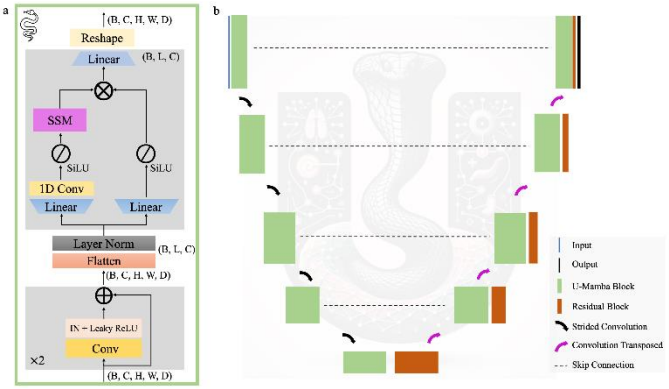


图 1 U-Mamba 方法框架图

图 2 展示了本文所提方法与其他医学图像分割方法的部分定性结果。可以看出，相较于现存的医学图像分割方法，U-Mamba 的分割区域更加完整，分割边界更加准确，展示出较为优越的分割性能。

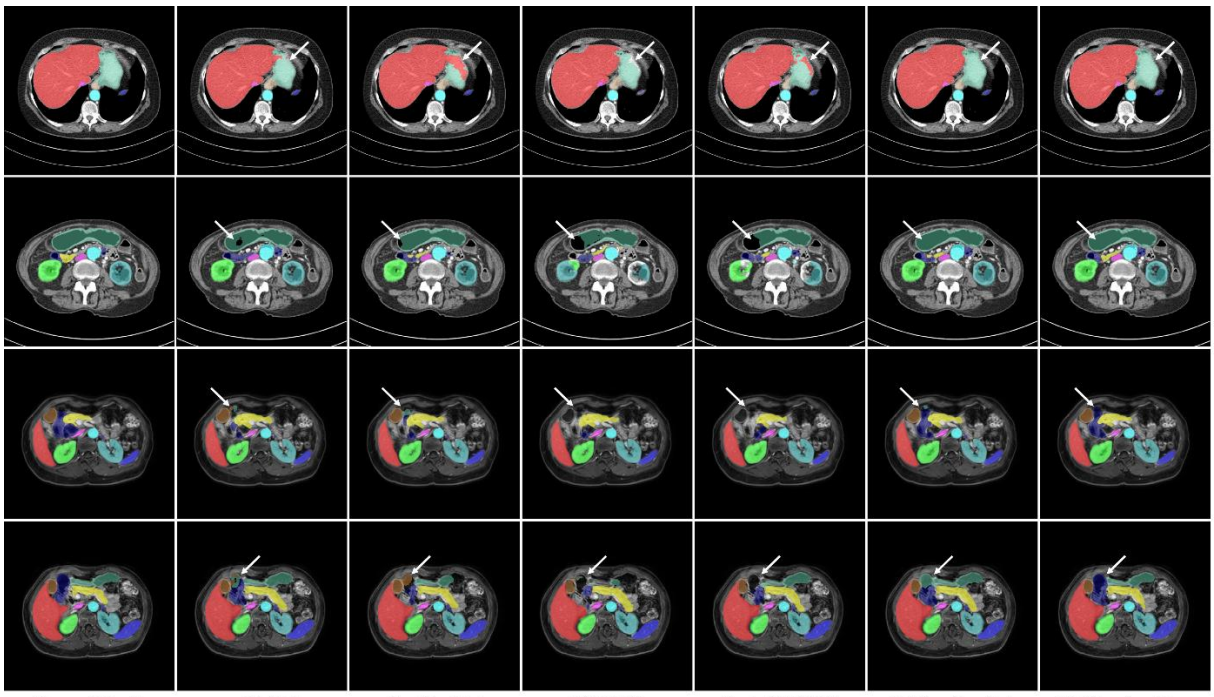


图 2 U-Mamba 方法部分定性结果展示

论文地址：
https://arxiv.org/abs/2401.04722
代码地址：
https://wanglab.ai/u-mamba.html

2、Swin-UMamba 方法

介绍：基于 Mamba 的医学图像分割模型相较于其他方法，具有：准确率更高、显存占用更少和计算开销更低的优势。然而，这些工作主要关注于视觉模型结构设计，没有进一步探讨预训练 Mamba 模型在医学图像分割中的作用。考虑到许多医学图像数据集数据量较少，使用预训练模型可以有效缓解模型过拟合问题并提升模型泛化能力。本文提出了 Swin-Umamba，希望借助 ImageNet 数据集进行预训练，进一步提升基于 Mamba 的模型在医学图像分割任务中的性能。

图 3 所示为 Swin-Umamba 方法的整体结构，类似于 UNet 的结构，包括编码器、解码器以及两者之间的跳跃连接。Swin-UMamba 的编码器部分基本遵循 VMamba-Tiny 的网络结构设计以便加载预训练模型参数，主要由 Patch Merging 模块和 VSS 模块构成。对于解码器部分，Swin-UMamba 在上采样模块中加入了一个额外的残差卷积模块用来处理跳跃连接的特征，并在每个尺度上都加入了一个额外的分割头用于深度监督，上采样模块如图 3 右上角所示。

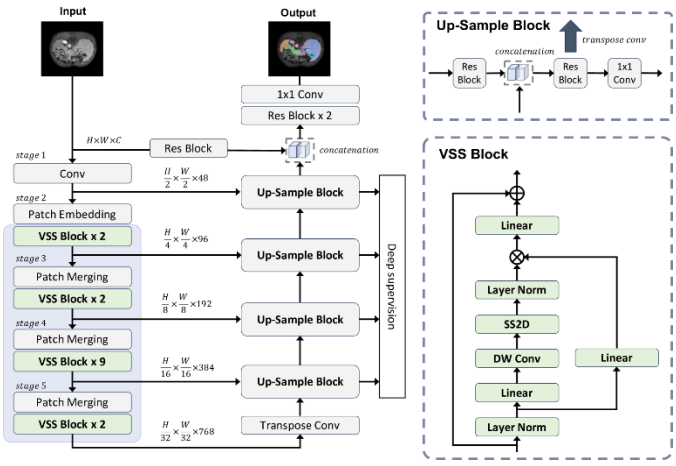


图 3 Swin-UMamba 方法框架图

图 4 展示了本文所提方法与其他医学图像分割方法分别在 AbdomenMRI(腹部 MRI 图像多器官分割)、

Endoscopy(内窥镜手术器械分割)和 Microscopy(细胞分割)三个数据集上的部分主观结果对比结果。可以看出，相较于现存的医学图像分割方法，Swin-UMamba 展示出较为优越的分割性能。

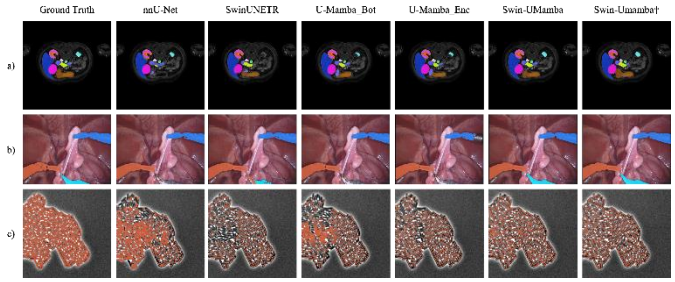


图 4 Swin-UMamba 方法部分定性结果展示

论文地址：
https://arxiv.org/abs/2402.03302
代码地址：
https://github.com/JiarunLiu/Swin-UMamba

3、VM-UNet 方法

介绍：基于 CNN 和 Transformer 的模型各有局限性。CNN 因局部感受野限制而难以捕获长距离信息，导致分割结果不佳；而 Transformer 虽然在全局建模上表现出色，但自注意力机制的计算复杂度较高。近期，以 Mamba 为代表的状态空间模型(State-Space Models, SSMs)不仅在全局建模上表现出色，而且具有线性复杂度，在各种视觉任务中表现出色。

本文基于 SSMs，提出了一种用于医学图像分割模型，命名为 VM-UNet(Vision Mamba UNet, VM-Unet)。引入了视觉状态空间(VSS)作为捕获上下文信息的基础块，构建了一个编码器-解码器结构，并使用预训练好的 VMamba-S 初始化其权重。图 5 所示为 VM-UNet 方法的整体结构。

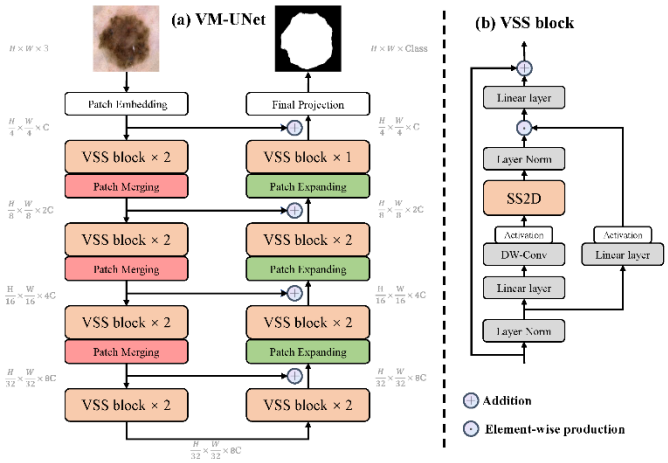


图 5 VM-UNet 方法框架图

在 ISIC17、ISIC18 和 Synapse 数据集上进行了全面的实验，结果表明 VM-UNet 在医学图像分割任务中具有竞争力。

论文地址：
https://arxiv.org/pdf/2402.02491
代码地址：
https://github.com/JCruan519/VM-UNet

责任编委 王田 樊鑫

李 策

教授，博士生导师，兰州理工大学电气工程与信息工程学院从事教育与科研工作，担任网络与信息中心主任。研究方向为计算机视觉、医学影像分析，智能机器人等。

张建伟

兰州理工大学电气工程与信息工程学院 硕士研究生，研究方向为医学影像分析、计算机视觉等。

医疗多模态数据集

东北大学 贾同 贾娜娜

医学多模态数据具有多样化的数据来源和形式，如影像、基因组、临床数据和生理信号等。这些数据在结构、维度和复杂性上存在差异，处理时需要考虑数据缺失、噪声和时间依赖性等问题。

不同模态的数据往往能够互为补充，结合分析能提供更全面的疾病诊断和预测。接下来本文将介绍几种医学多模态数据集。

1、BraTS 数据集

BraTS（Brain Tumor Segmentation Challenge）数据集是一个广泛使用的医学影像数据集，专门用于脑肿瘤的自动分割和分类任务。它是由多个医学研究机构和大学合作开发，旨在推动脑肿瘤分割和医学图像分析领域的研究。

BraTS 数据集的核心目的是为脑肿瘤（尤其是胶质母细胞瘤）提供高质量的标注数据，供研究人员开发和评估自动分割算法，尤其是在使用磁共振成像（MRI）进行肿瘤检测和治疗计划时。

BraTS 数据集中通常包含每个患者四种不同模态的 MRI 扫描数据，这些模态提供了对肿瘤不同方面的补充信息。

T1 加权图像（T1-weighted images）：用于显示解剖结构，提供详细的脑组织结构信息。T1 加权对比增强图像（T1-weighted contrast-enhanced images）：增强后的 T1 图像，可以显示肿瘤区域的血管化结构，帮助识别肿瘤。T2 加权图像（T2-weighted images）：对比度高，可以帮助识别肿瘤和水肿区域，特别是在脑白质和灰质之间。FLAIR 图像（Fluid-attenuated inversion recovery）：用于消除脑脊液的信号，以提高肿瘤和脑白质的对比度，特别适合检测肿瘤周围的水肿区域。图 1 所示为四种图像模态示意图。

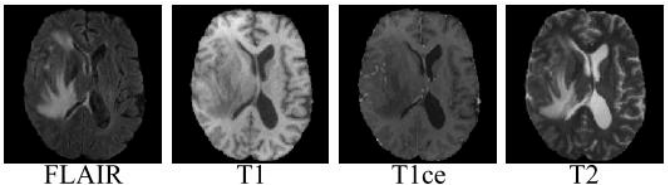


图 1 脑 MRI 成像中的四种模态

数据集的每个图像都有对应的标注数据，由医学专家手工标注。标注的目标是识别和分割肿瘤的不同区域，通常分为以下几种：肿瘤核心区域（Tumor core）：肿瘤的主要部分，通常包括坏死组织；增强区域（Enhancing tumor）：肿瘤中活跃生长的部分，通常由对比剂增强显现；全肿瘤区域（Whole tumor）：包括所有肿瘤相关的区域，包括肿瘤核心、增强区域和水肿区。图 2 所示为肿瘤标注区域。

图 2 脑肿瘤标注

BraTS 数据集主要用于以下几个应用场景：肿瘤分割，BraTS 数据集为自动肿瘤分割算法的研究提供了标准数据集，肿瘤分割不仅可以帮助临床医生快速定位肿瘤，还能用于肿瘤生长跟踪、治疗评估等任务；多模态学习，由于数据集提供了四种不同的 MRI 模态，研究者可以使用这些多模态信息进行深度学习模型的训练，改善肿瘤分割精度；疾病进展预测，通过分析肿瘤的体积、形态和增长速度等，可以预测肿瘤的进展情况，评估不同治疗方法的效果；临床决策支持，自动分割肿瘤区域可以为医生提供有效的决策支持，帮助其制定个性化治疗计划；模型验证与评估，BraTS 数据集为新算法提供了一个标准化的测试平台，研究人员可以基于此评估来比较不同分割算法的性能。

数据集下载地址：
https://www.synapse.org/Synapse:syn27046444/wiki/616571

BraTS 数据集是脑肿瘤自动分割领域最重要的资源之一，为研究人员提供了大量的高质量标注数据，有助于推动医学影像分析算法的发展，尤其是在脑肿瘤的自动诊断和治疗中具有广泛应用。通过 BraTS 数据集，研究人员能够设计出更精确、鲁棒的算法，从而改善临床决策和病人预后。

2、OASIS 数据集

OASIS（Open Access Series of Imaging Studies）数据集是一个广泛使用的神经影像学数据集，专门用于研究与阿尔茨海默病（Alzheimer's disease, AD）以及其他认知障碍相关的神经退行性疾病。该数据集由华盛顿大学的神经成像与计算机科学实验室开发，并公开发布，旨在为脑部疾病的研究者提供高质量的影像数据，支持脑结构、功能及其变化与认知能力之间的关系研究。

OASIS 数据集包括了多种类型的数据，具体包括以下内容。

1. MRI 图像数据

结构性 MRI（Structural MRI）：OASIS 数据集主要提供高分辨率的结构性 MRI 图像，这些图像可以帮助研究人员分析大脑的形态学特征，如大脑皮层的厚度、大脑区域体积、脑灰质和白质的结构变化等。

模态：数据集包括 T1 加权的 MRI 图像，这是脑部成像中最常用的一种方式，能够提供高分辨率的大脑解剖结构信息。如图所示为脑 MRI 图像。

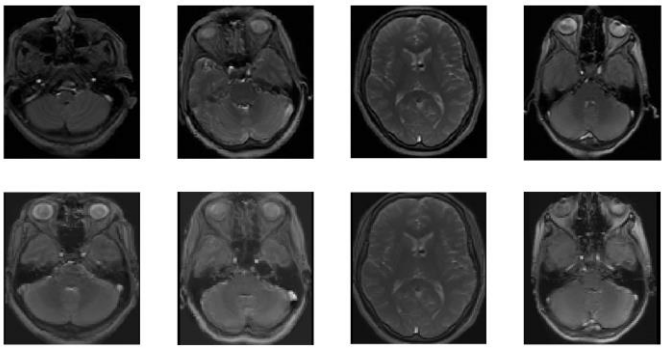


图 3 结构性 MRI 图像

2. 认知测试数据

OASIS 数据集包含多项认知测试的结果，例如：

Mini-Mental State Examination (MMSE)：评估受试者认知功能的简易测试。其他认知能力评估：包括语言、记忆、执行功能、注意力等方面的评估。通过认知测试数据，研究人员能够将大脑影像学特征与受试者的认知功能进行关联，帮助理解脑结构变化与认知能力之间的关系。

3. 年龄、性别等人口统计数据：数据集提供了受试者的年龄、性别、教育背景等人口统计信息，这些信息对于分析认知变化与不同因素的关系非常重要。

4. 阿尔茨海默病和轻度认知障碍的标签：数据集中提供了关于每个受试者认知状态的标签，包括正常、轻度认知障碍（MCI）以及阿尔茨海默病（AD）。这些标签使得研究者可以根据认知障碍的不同阶段来分析和比较大脑结构变化。

OASIS 数据集广泛应用于以下领域的研究：阿尔茨海默病与轻度认知障碍研究，通过分析 MRI 图像和认

知测试结果，研究人员能够更好地理解阿尔茨海默病的早期标志，从而推动早期诊断的研究；脑结构与认知功能的关联，研究人员通过对比不同认知状态下的受试者的脑结构变化（例如大脑皮层厚度、脑体积等），探讨脑部结构变化如何与认知能力下降相关；机器学习与算法开发，通过结合 MRI 影像数据和认知测试数据，研究人员可以开发用于预测阿尔茨海默病早期标志的模型，甚至尝试通过机器学习预测疾病的进展；临床决策支持系统，利用 OASIS 数据集开发的算法可以为医生提供辅助诊断和治疗决策支持，帮助医生根据患者的大脑影像和认知能力更好地评估病情。

数据集下载地址：https://sites.wustl.edu/

OASIS 数据集是一个强大且广泛应用于阿尔茨海默病及认知障碍研究的重要资源。它为研究者提供了高质量的脑影像学数据和相关的认知测试结果，有助于推动脑部疾病的早期诊断、疾病进展分析、脑结构功能关系研究等方面的研究。通过这个数据集，研究人员能够探索大脑如何随年龄变化及认知障碍的发展而发生结构性改变，并开发出更先进的预测和诊断方法。

3、CPTAC-UCEC 数据集

CPTAC-UCEC （Cancer Proteome Atlas – Uterine Corpus Endometrial Carcinoma） 数据集是由美国国家癌症研究所（NCI）资助的癌症蛋白质组图谱（CPTAC）计划的一部分，专门针对子宫内膜癌（Uterine Corpus Endometrial Carcinoma, UCEC）的多组数据集。该数据集包含了子宫内膜癌（UCEC）患者的临床数据、基因组数据、转录组数据、蛋白质组数据等，旨在通过多维度的数据帮助研究人员深入理解子宫内膜癌的分子机制，并推动癌症早期诊断、治疗和个性化医学的研究。

子宫内膜癌（UCEC） 是一种起源于子宫内膜的恶性肿瘤，是女性常见的生殖器癌症之一。CPTAC-UCEC 数据集主要致力于揭示 UCEC 的分子特征、潜在的生物标志物以及其分子机制。该数据集整合了以下几类数据。

1. 临床数据

患者信息：包括年龄、性别、种族、肿瘤分期、治疗信息、随访数据等。病理数据：包括组织学类型、分级、肿瘤大小、转移情况等。生存数据：例如总生存期（OS）、无病生存期（PFS）等。

2. 蛋白质组数据

全蛋白质组数据：CPTAC-UCEC 数据集提供了子宫内膜癌患者组织中的蛋白质表达谱，包括正常组织与肿瘤组织之间的差异。这些数据有助于研究蛋白质的表达水平、翻译后修饰（如磷酸化、泛素化等），并且能揭示肿瘤的潜在生物标志物和治疗靶点。定量蛋白质组数据：这些数据可以用来识别与癌症发生、发展相关的蛋白质，进而研究它们在癌症治疗中的潜力。

3. 基因组数据

突变数据：包括肿瘤样本中的基因突变、插入/缺失（INDELs）、拷贝数变异（CNVs）等。基因表达数据：提供肿瘤组织和正常组织之间的基因表达差异。这些数据有助于揭示与肿瘤发生相关的基因，并可作为潜在的生物标志物。甲基化数据：CPTAC-UCEC 还包括了 DNA 甲基化的相关数据，能够揭示基因调控的表观遗传学机制。

4. 转录组数据

RNA-seq 数据：包括肿瘤组织和正常组织的转录组数据，帮助分析与肿瘤相关的基因表达变化。这些数据有助于揭示基因表达模式、转录水平变化与癌症进展之间的关系。

5. 图像数据

成对的放射性图像和病理图像用于患者的生存期预测，如图 4 所示为图像模态。

CPTAC-UCEC 数据集广泛应用于以下领域的研究：生物标志物发现、癌症分子机制研究、靶向治疗、个性化医学。

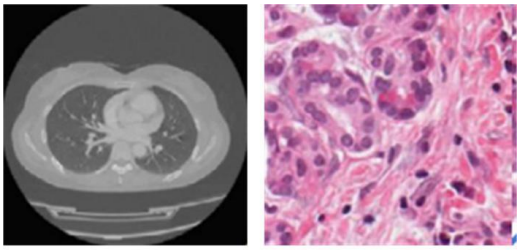


图 4 CPTAC-UCEC 多模态图像

数据集下载地址：https://proteomics.cancer. gov/programs/cptac

CPTAC-UCEC 数据集是一个丰富且全面的多组学数据集，为研究人员提供了大量关于子宫内膜癌的基因组、转录组、蛋白质组和临床数据。该数据集的目标是帮助科学家深入了解 UCEC 的分子机制、开发新的诊断标志物和治疗靶点，并推动癌症个性化治疗的发展。通过多层次、多角度的分析，CPTAC-UCEC 数据集将在癌症研究中发挥重要作用。

责任编委 李策 樊鑫

贾 同

东北大学信息科学与工程学院教授、博士生导师，智能感知与机器人研究所所长。研究方向计算机视觉、模式识别与机器学习等。电子邮箱：jiatong@ise.neu.cn

贾娜娜

博士研究生，东北大学信息科学与工程学院，研究方向为医学影像处理。
电子邮箱：2010284@stu.neu.edu.cn

好 文 推 荐

大连理工大学与 OPPO 联合提出的”LOVD: Large-and-Open Vocabulary Object Detection”最新成果发表在 ACM MM 2024。

论文：Shiyu Tang, Zhaofan Luo, Yifan Wang, Lijun Wang, Huchuan Lu, Weibo Su, Libo Liu. LOVD: Large-and-Open Vocabulary Object Detection, ACM MM 2024, 9321-9329, 2024

现有开放词表目标检测器通常需要在推理过程中预先定义一个准确且紧凑的词表。然而，在现实场景中，词表通常是不确定的，并且词表所含词汇类别数目是呈指数增长的。因而，为模拟更真实场景，本文提出了一种大规模开放词表目标检测算法。这种算法通过使用包含数千未见类别的大规模词表来测试检测器。大量的未见类别不可避免地导致干扰项类别的增加，严重阻碍了识别过程，并导致检测结果不令人满意。为了解决这一挑战，本文提出了一个大型开放词表检测器（Large-and-Open Vocabulary Detector, LOVD），其主要包含两个核心组件，即图像-区域过滤（Image-to-Region Filtering，IRF）模块和交叉视图验证（Cross-View Verification, CV2）机制。IRF 和 CV2 在全局和局部视图中进行图像-区域过滤以缓解大规模词表类别干扰。与先前的工作相比，本文提出的 LOVD 方法更具可扩展性，面对大规模词表输入时有更加稳健的表现，并可与主要的开放词表检测方法无缝集成，以提高其开放词表检测性能。

具体而言，如图 1 所示，其主要由一个类别无关的定位模块、一个图像-区域过滤（IRF）模块和一个两分支投票模块组成。给定一个输入图像，本文首先使用预训练视觉语言模型的视觉编码器提取特征图。之后，通过区域提议网络定位所有潜在对象，并通过 RoI Align 从图像特征图中提取其特征。然后，IRF 模块通过交叉视图验证（CV2）机制对每个提议对象进行识别。交叉视图验证机制由并行的、基于匹配和基于查询的两种开放词表分类器组成，两种分类器将输出两个不同但可互补验证的结果作为 IRF 的输出。最后，通过一个投票过程，将 IRF 输出的两分支识别类别结合，计算得出最终结果。本文在 COCO，LVIS，Pascal VOC 等数据集上的广泛实验展现出 LOVD 相较于现有方法的独特优势。

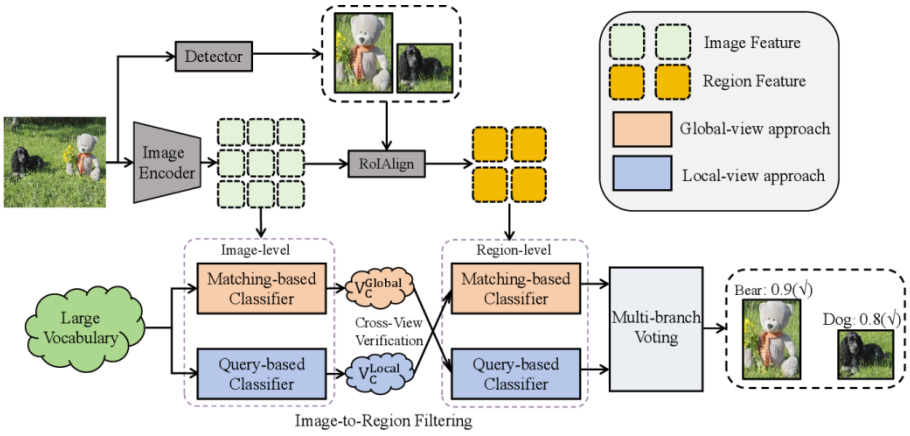


图 1 LOVD 检测流程图

责任编委 贾同 王田

好 文 推 荐

中国科学院大学“Fully Sparse Fusion for 3D Object Detection”的最新成果发表在 IEEE TPAMI 2024。

论文：Yingyan Li, Lue Fan, Yang Liu, Zehao Huang, Yuntao Chen, Naiyan Wang. Fully Sparse Fusion for 3D Object Detection, IEEE TPAMI, 46(11): 7217-7231 (2024)

众所周知，3D 目标检测是汽车自动驾驶重要的组成部分。当前，LiDARs 和像机是 3D 感知主要使用的两大传感器。感知过程中，将两类传感器的信息数据结合不仅可以发挥激光雷达精确深度估计的优势，还可以充分利用相机丰富的语义信息。现存主流的多模态方法往往依赖于密集检测器，该类检测器通过构建稠密的鸟瞰（Bird’s-Eye-View， BEV）特征图检测目标。BEV 特征图的大小会随着感知范围的扩大而成倍增长，这无疑极大增加了检测成本。

为了解决上述问题，本文设计了一个不引入任何密集 BEV 特征的多模态检测器(Fully Sparse Fusion, FSF)，方法流程如图 1 所示。首先，在具有丰富语义信息的 2D 图片上，本文使用一个 2D 实例分割模型生成相应图片的实例掩模。而后，根据像机的内外参数将每个 2D 实例掩模转换为 3D 视锥，包含在这个视锥内的点构成了与掩模相对应的 3D 实例。这些来自图像模态的 3D 实例与基于雷达的实例分割生成的 3D 实例是互补的。最后，基于所提双模态实例的预测模块将上述两类实例进行融合并做出最终预测。

FSF 在流行的 3D 检测数据集 nuScenes、Waymo 和 Argoverse2 中取得了最先进的性能。值得一提的是，在远程检测 Argoverse2 数据集中，FSF 比之前最先进的多模态检测器快 2.7 倍。

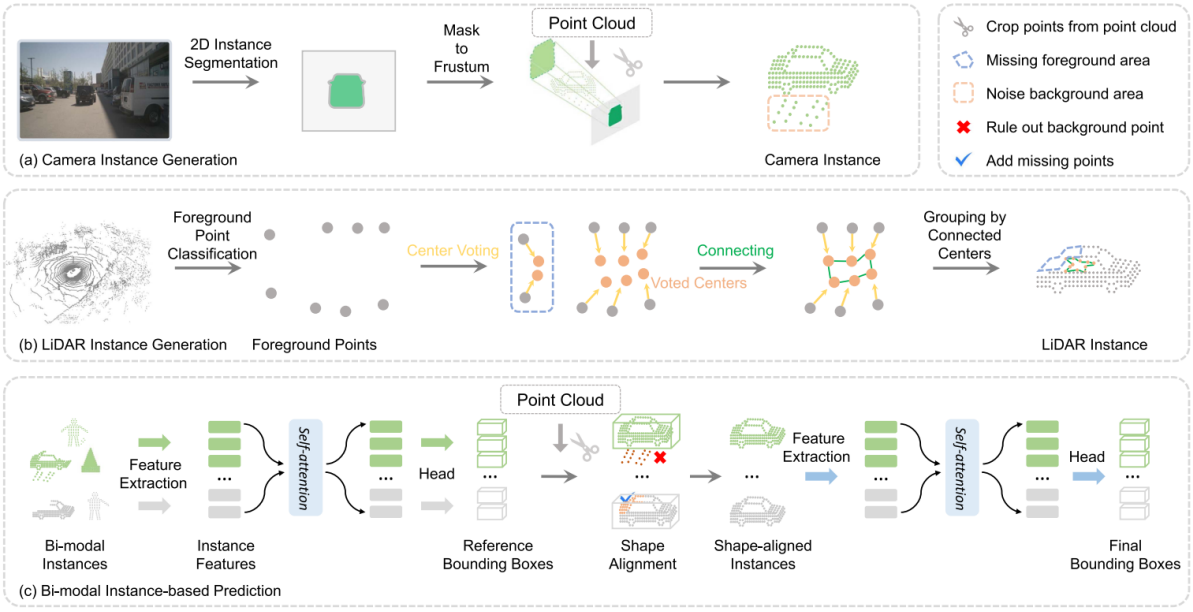


图 1 FSF 模型结构流程图

责任编委 樊鑫 贾同

好 文 推 荐

中国科学技术大学、阿里巴巴集团、上海交通大学、蚂蚁集团的“CCM: Real-Time Controllable Visual Content Creation Using Text-to-Image Consistency Models“最新成果发表在 ICML 2024。

论文：Jie Xiao, Kai Zhu, Han Zhang, Zhiheng Liu, Yujun Shen, Zhantao Yang, Ruili Feng, Yu Liu, Xueyang Fu, Zheng-Jun Zha. CCM: Real-Time Controllable Visual Content Creation Using Text-to-Image Consistency Models, ICML 2024.

一致性模型（Consistency Models, CMs）是一种近年来崭露头角的生成模型，能够在较少采样步骤条件下生成高质量图像。CMs 可以通过一致性蒸馏技术（Consistency Distillation）从预训练的扩散模型（Diffusion Models, DMs）中获得，也可以通过一致性训练技术（Consistency Training）从数据中独立训练。然而，目前尚未探索如何为预训练的 CMs 添加额外控制条件，限制了其在可控生成任务中的应用范围。

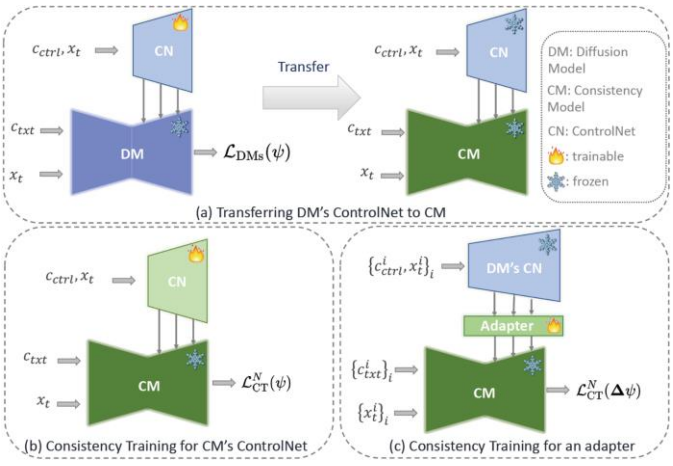


图 1 三种基于 CMs 的 ControlNet 训练策略

受到 ControlNet 在文本到图像生成任务中成功实践的启发，本文研究了将 ControlNet 与 CMs 相结合的方法，以实现高效且可控的图像生成。具体而言，如图 1 所示，本文设计并探讨了三种训练方案：(a) 直接迁移基于 DMs 的 ControlNet 到 CMs；(b) 针对 CMs 训练专属的 ControlNet，通过一致性约束进一步增强生成效果；(c) 借助适配器（Adapter），在 CMs 中更好地利用基于 DMs 的 ControlNet，并通过一致性训练缓解 DMs 与 CMs 之间的性能差距。

本文通过理论分析和实验验证得出了以下结论：方案(a)能够实现从 DMs 到 CMs 的可控语义信息迁移，但在生成图像细节方面表现不足；方案 (b) 通过一致性训练技术实现了专属训练，在图像语义控制和细节生成方面均展现出显著优势；方案 (c) 能提升基于 DMs 的 ControlNet 迁移到 CMs 的可控生成效果。

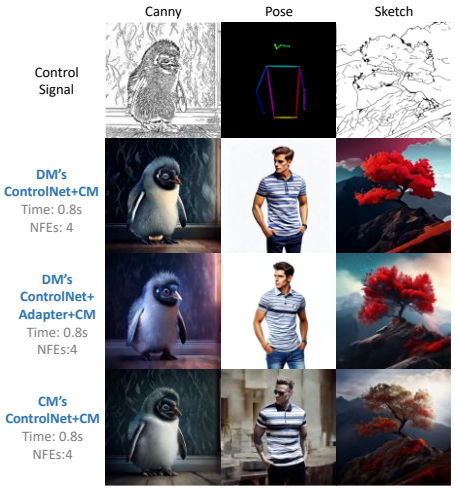


图 2 基于所提出训练策略的图像生成效果

针对多种条件控制生成任务（如草图转图像、姿势图转图像等），本文提出的 ControlNet for CMs 方法不仅在高层语义控制上表现出色，同时能够高效实现细节生成，为可控图像生成领域提供了新方案。图 2 展示了基于本文所提出训练策略的图像生成效果。

责任编委 李策 王田

征文通知

1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 Image and Vision Computing，Pattern Recognition 和 IEEE Journal of Biomedical and Health Informatics。

3 会议简介

中国模式识别与计算机视觉学术会议 PRCV (Chinese Conference on Pattern Recognition and Computer Vision)，由中国计算机学会（CCF）、中国自动化学会（CAA）、中国图象图形学学会（CSIG）和中国人工智能学会（CAAI）联合主办，定位国内顶级的模式识别和计算机视觉领域学术盛会。

第八届 PRCV 将于 2025 年 10 月 16 日至 10 月 19 日在上海举办，由上海交通大学承办。本届会议将秉持团结模式识别与计算机视觉领域科技工作者的宗旨，进一步推动开放合作，广泛吸引学术界和工业界的人才，提升会议的国际化水平，力求打造一个高品质的学术交流平台。大会的举办将为学术界与工业界提供更多产学研合作机会，推动模式识别与计算机视觉领域的协同创新和可持续发展。

责任编辑：刘帅奇

表 1 计算机视觉领域相关国内外会议

会议名称	会议时间	会议地点	截稿日期	会议网站
ICLR 2025	2025.07.13-19	Vancouver, Canada	2025.01.30	https://icml.cc/Conferences/2025
ICCV 2025	2025.10.19-25	Hawaii, United States	2025.03.07	https://iccv2025.thecvf.com/
IJCAI 2025	2025.08.16-22	Guangzhou, China	2025.01.23	https://2025.ijcai.org/
ACL 2025	2025.07.27-8.01	Vienna, Austria	2025.02.15	https://2025.aclweb.org/

表 2 计算机视觉领域相关国内外期刊专刊

期刊名称	专刊题目	投稿网址	截稿日期
IVC	Advancing Transparency and Privacy: Explainable AI and Synthetic Data in Biometrics and Computer Vision	https://www.sciencedirect.com/special-issue/314030/advancing-transparency-and-privacy-explainable-ai-and-synthetic-data-in-biometrics-and-computer-vision	2025.02.07
PR	From bench to the wild: Recent Advances in Computer Vision methods (WILD-VISION)	https://www.sciencedirect.com/special-issue/315871/from-bench-to-the-wild-recent-advances-in-computer-vision-methods-wild-vision	2025.03.31
JBHI	Identification and Correction of False Medical Data in Large Language Models: Enhancing Reliability and Accuracy of Healthcare AI Systems	https://www.embs.org/jbhi/wp-content/uploads/sites/18/2024/11/JBHI_AIGC.-Identification.pdf	2025.02.01

COMPUTER VISION NEWSLETTER

04 2024

总第 42 期

计算机视觉专委会简报

CCF 计算机视觉专委会