

主办 CCF 计算机视觉专业委员会

COMPUTER
VISION
NEWSLETTER

CCCF 计算机视觉 专委会简报

03 2020

总第 25 期



CCCF 计算机视觉
专委会

COMPUTER VISION NEWSLETTER



计算机视觉专委会 简报

2020年第03期

总第25期



CCF 计算机视觉
专委会

主 办 编委会

CCF 计算机视觉专业委员会

/专委动态/

荣誉主编	王 亮	中国科学院自动化研究所
主 编	马占宇	北京邮电大学
执行主编	李实英	上海科技大学
主 编	毋立芳	北京工业大学
编 委	黄 岩	中国科学院自动化研究所
	任传贤	中山大学
	杨巨峰	南开大学

/科技前沿/

主 编	王金甲	燕山大学
编 委	储 珺	南昌航空大学
	崔海楠	中国科学院自动化研究所
	任桐炜	南京大学

/委员风采/

主 编	余 焯	合肥工业大学
编 委	刘海波	哈尔滨工程大学
	赵振兵	华北电力大学

/学术资源/

主 编	李 策	兰州理工大学
编 委	樊 鑫	大连理工大学
	贾 同	东北大学
	沈沛意	西安电子科技大学

/海外学者/

主 编	金 鑫	北京电子科技学院
编 委	刘帅奇	河北大学
	张汗灵	湖南大学

CONTENTS

简报目录

| 专委动态

- 04 走进高校活动系列
- 05 CCF-CV 视界无限系列研讨会
- 07 CCF-CV 走进企业系列交流会
- 08 RACV 2020 计算机视觉前沿进展研讨会圆满召开
- 10 CCF-CV 专委会多名委员当选 IAPR Fellow
- 11 PRCV 2022 征集承办单位
- 13 CCF-CV “持久影响力论文奖”
- 14 CCF-CV “中科视拓 Seeta 学术新锐奖”

| 科技前沿

- 15 基于元学习的泛化人脸识别
- 18 多图的离线协同与在线增量匹配
- 20 上下文感知的密集视频描述生成
- 22 CVPR 2020

| 委员风采

- 25 哈尔滨工业大学左旺孟教授访谈
- 28 东南大学杨万扣副教授访谈
- 31 委员好消息

| 学术资源

- 32 艺术字风格化开源代码
- 35 可解释性数据集
- 38 好文推荐

| 海外学者

- 42 加州大学圣迭戈分校苏昊教授团队
- 48 征文通知

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 88 期 河北工业大学



2020年6月21日上午,由中国计算机学会计算机视觉专委会(CCF-CV)主办,河北工业大学省部共建电工装备可靠性与智能化国家重点实验室承办的 CCF-CV 走进高校系列报告会——河北工业大学站,通过线上直播的方式举行。本期报告会的主题是“智能医学影像和计算机视觉前沿技术与应用”,由厦门大学纪荣嵘教授、中国科学院自动化研究所赫然研究员、北京航空航天大学徐迈教授、清华大学高跃教授、上海大学施俊教授 5 位专家担任本次报告会的特邀讲者,由河北工业大学刘晓明教授、徐桂芝教授、许铮铨教授担任本次报告会的执行主席,河北工业大学许铮铨教授主持了本次报告会。在本次报告会上,专家们就智能医学影像和计算机视觉领域的前沿技术进行了深入的交流和思想的碰撞。在随后的圆桌讨论中,五位专家还就“后疫情时代智能医疗的开展和落地”及“智能医学工程专业的建设”这两个议题分享了各自的真知灼见。

最后,刘教授再次对来自全国各地的专家、学者表示热烈的欢迎。刘教授表示,希望通过此次线上报告进一步促进电气工程学院和国重实验室与各兄弟院校的学术交流,推动科研合作发展。

第 89 期 上海交通大学



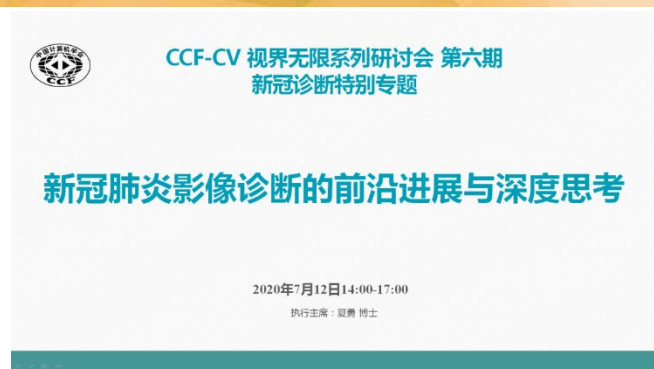
2020年7月18日下午,由中国计算机学会计算机视觉专委会(CCF-CV)主办,上海交通大学人工智能研究院承办的 CCF-CV 走进高校系列报告会——上海交通大学站,通过线上直播的方式举行。本期报告会邀请了北京大学信息科学技术学院智能科学系查红彬教授、清华大学副教务长和研究生院院长周杰教授、上海科技大学信息科学与技术学院执行院长虞晶怡教授、大连理工大学创新学院院长卢湖川教授 4 位特邀讲者,由上海交通大学人工智能研究院杨小康教授、严骏驰副教授、马超助理教授担任本次报告会的执行主席,杨小康教授、严骏驰副教授主持了本次报告会。上海交通大学人工智能研究院常务副院长杨小康教授首先进行开场致辞。在本次报告会上,专家们就计算机视觉领域的前沿技术进行了深入的交流和思想的碰撞。在随后的圆桌讨论中,五位专家围绕“计算机视觉的六大关系”分享了各自的真知灼见。

最后,杨小康教授再次对来自全国各地的专家、学者表示欢迎,希望通过线上报告会,进一步促进人工智能研究院与各兄弟院校的学术交流,推动科研合作发展。

责任编辑 毋立芳

第 6 期 新冠肺炎影像诊断的前沿进展与深度思考

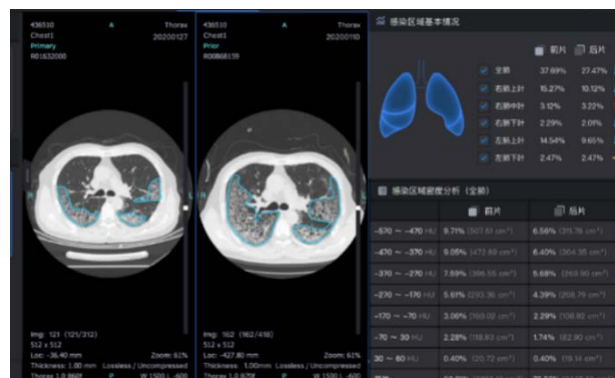
CCF-CV 视界无限系列研讨会



2020年7月12日，由中国计算机学会计算机视觉专委会主办的第6期 CCF-CV “视界无限”系列活动特别专题——“新冠肺炎影像诊断的前沿进展与深度思考”研讨会在线上成功举行。本期研讨会由计算机视觉专委会副秘书长、南开大学杨巨峰教授组织，西北工业大学夏勇教授担任执行主席。研讨会邀请了计算机视觉专委会主任、北京大学查红彬教授致辞，南开大学程明明教授、华中科技大学王兴刚副教授、上海交通大学王乾副教授、天津大学张长青副教授、阿联酋起源人工智能研究院付华柱研究员做主题报告并参与圆桌讨论。计算机视觉专委会 B 站公众号对会议进行了全程直播，直播人气峰值达到 3648。

首先，查红彬教授代表计算机视觉专委会致辞。查老师指出了本次活动的两个重要意义：其一，自新冠疫情爆发以来，计算机视觉领域众多专家学者积极投身抗疫一线，在新冠肺炎影像诊断方面开展了卓有成效的研究工作，为疫情防控贡献了 AI 的力量，借此机会全面总结相关领域的研究成果和最新进展恰逢其时。其二，计算机视觉专委会主办的“视界无限”系列活动定位于就计算机视觉领域的某一具体问题进行深入细致的研

讨，本次活动在总结经验之余也对计算机视觉社区如何在未来助力应对公共卫生事件进行深入的思考和交流，其成果有望产生长远和深刻的影响。希望线上研讨这种新形式吸引更多同行参与到活动中来，帮助大家更好地理解知识、激发创意；祝愿研讨会圆满成功。



程明明教授的报告题目是“基于医学影像的疾病筛查”。他结合抗疫实践，介绍了医学影像在疾病筛查中的作用和意义，以及相关识别检测技术，包括基于 CT 影像的新冠肺炎检测与分割技术，以及基于 X 光图像的肺结核检测技术。程老师在报告最后对医学影像人工智能筛查，特别是新冠肺炎 CT 影像的 AI 筛查，做了一些总结和反思，重新整理了下一步探索的思路。

王兴刚副教授以“胸部 CT 影像中的弱监督 COVID-19 分类与病灶定位”为题，深入浅出地介绍了他们团队在疫情爆发早期数据资源和计算资源同时受限的困难情况下，开发的一套诊断算法。该算法使用普通的 GPU 设备，仅用 1.93 秒即可处理单个患者的 CT 数据。在训练过程中无需注释病变，采用弱监督深度学习模型可以准确地预测 COVID-19 的感染可能性，并发现胸部 CT 的病变区域。这种易于训练的高性能深度学习算法提供

了一种识别 COVID-19 患者的快速方法, 在新冠疫情的防控中具有良好的应用价值。

王乾副教授的报告题目是“Dual-sampling Attention Network for Diagnosis of COVID-19 from Community Acquired Pneumonia”。王老师指出新冠肺炎的早期诊断对治疗病人和防控疫情扩散均有重要意义, 他们提出了一种基于胸部 CT 影像的双采样注意力网络解决此问题。具体地, 他们在 3D 卷积神经网络基础上设计了一种新型的在线注意力模块以快速定位和识别肺部感染区域。在目前已知的最大规模数据集上取得了最优结果, 并对数据的不平衡性有较好的适应能力。

张长青副教授的报告题目是“基于多视图学习的新新冠肺炎诊断”, 为相关研究提供了一种新的思路。针对早期的新冠患者的 CT 影像与普通肺炎具有很大的相似性的问题, 研究可以较好区分两者的方法。张老师就其团队如何利用 CT 影像的多种特征准确、快速、稳定地对新冠肺炎患者进行鉴别诊断进行了详细介绍。

付华柱研究员的报告题目是“Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images”, 重点介绍了近期提出的肺部感染区域分割神经网络 Inf-Net。针对在有限时间内快速开发该系统面临的数据饥渴难题, 进一步提出了半监督的分割框架, 在少量标注数据和大量未标注数据上学习能力和分割性能都得到了验证。

在 Panel 环节, 与会嘉宾就“核酸检测假阴性比较多, 影像诊断是不是也有这样的问题? 影像诊断的优势是什么?”、“目前新冠肺炎的 AI 影像诊断是否能够



应用于临床? 不足主要在哪里?”、“新冠肺炎的 AI 影像诊断研究中, 如何更好地利用数据?”、“最近有医生表示, ‘疫情在中国能控制下来, 核心还是靠传统人工办法和智慧来实现的, 人工智能没有深度参与’。在后新冠时代, 面对可能出现的类似公卫危机, 人工智能和计算机视觉能做什么?”等问题展开热烈讨论, 网络参会者也在直播平台就自己关心的问题与嘉宾交换了观点。



最后, 第 6 期“视界无限”研讨会在下午 5 点 15 分圆满结束。

责任编辑 杨巨峰

第 21 期 联影医疗

CCF-CV 走进企业系列交流会

2020 年 8 月 8 日，由中国计算机学会计算机视觉专委会主办的第二十一期 CCF-CV 走进企业系列交流会——“CCF-CV 走进联影医疗”在线上成功举行。本期研讨会由计算机视觉专委会副秘书长、中科院自动化所黄岩副研究员、南开大学杨巨峰教授、联影医疗金晓燕女士等组织。



研讨会邀请了计算机视觉专委会副主任、上海科技大学虞晶怡教授和联影集团联席 CEO、联影研究院院长陈群博士致辞，计算机视觉专委会主任、北京大学查红彬教授、西北工业大学夏勇教授、南京航空航天大学张道强教授、上海科技大学何旭明副教授作主题报告并参与圆桌讨论。本次研讨会由联影医疗资深科学家曹拓宇博士主持，计算机视觉专委会 B 站公众号对会议进行了全程直播，直播人气峰值达到 5000+。

北京大学查红彬教授带来关于《视觉 SLAM:在线学习的途径》的报告。本报告介绍了如何最大限度地利用传感数据的时空一致性与三维地图的几何不变性，在现有多视点几何计算的基础上强化 SLAM 算法的系统性与泛化性，为解决传统 SLAM 方法重建误差积累严重、处理成本高昂等问题提供了新思路。

西北工业大学夏勇教授带来关于《基于深度学习的

医学影像分割技术》的报告。本报告以皮肤镜和病理组织显微图像分割为例，系统介绍了夏老师研究组进行医学图像分割方面的经验和体会，同时也探讨了在医学图像小数据上进行深度学习研究所面临的机遇与挑战。

南京航空航天大学张道强教授带来关于《Deep Learning for MR Image Reconstruction, Segmentation and Classification》的报告。本报告系统介绍了张老师研究组基于深度学习在 MR 图像压缩感知重建、大脑 MR 图像分割，以及脑疾病诊断分类方面的最新进展。

上海科技大学何旭明副教授带来关于《Weak Supervision Strategies for Semantic Segmentation and Alignment》的报告。为了缓解像素级场景理解需要大规模密集标注数据的问题，本报告介绍了何老师研究组从自监督学习和小样本学习的角度进行弱监督语义分割的最新成果。

联影医疗资深科学家曹拓宇博士带来关于《联影集团和联影研究院介绍》的报告。该报告系统回顾了联影集团的组织架构、业务范围、产品市场、发展战略、以及产学研协同创新方面的情况，让大家对公司的各个方面有了比较全面的认识。

随后的圆桌讨论环节由曹拓宇博士主持，圆桌讨论主要围绕：弱监督学习中如何能有效地训练网络和避免对少量标记样本的过拟合；如何实现高效的点云分割和降采样算法；计算机视觉的瓶颈与未来落地方向等联影较为感兴趣的话题，展开了热烈讨论。

责任编辑 黄岩

RACV2020 计算机视觉前沿进展研讨会圆满召开



2020年8月29日，中国计算机学会计算机视觉专委会（CCF-CV）年度学术研讨会 RACV（Recent Advances on Computer Vision）在福建厦门圆满召开。从去年开始，RACV会议在组织形式与研讨内容上进行了全新改版，定位为国内计算机视觉领域的小规模精品研讨会，通过定向邀请方式汇集领域专家，深度研讨计算机视觉领域中的若干核心问题并形成进展报告。研讨会试图通过务实、开放与平等的对话与讨论，深入发掘相关研究领域潜在的问题，为广大的科研人员提供观察问题的新视角与新观点。

本次会议开幕式由专委会副主任、上海科技大学虞晶怡教授主持，专委会主任、北京大学查红彬教授和厦门大学王萑子教授进行开幕式致辞。根据常委委员前期的讨论票选，本次会议设置了3项研讨主题。每项主题首先由特邀嘉宾们进行主题发言，之后所有与会人员自由讨论。

29日早上首先进行了主题一“视觉与语言：现状与趋势”的研讨。该主题由专委会副主任、中科院自动化

所王亮研究员、清华大学鲁继文副教授、西北工业大学王鹏教授3位委员负责组织，邀请了中国人民大学金琴教授、京东技术副总裁梅涛博士、北京大学彭宇新教授、电子科技大学申恒涛教授共4位嘉宾进行主题发言。自2012年以来，深度学习的兴起为视觉与语言领域的发展带来了蓬勃生机，放眼未来，视觉与语言的核心挑战和关键应用有哪些？几位嘉宾围绕预训练模型、认知机制建模、模型评测、工业界应用等前沿方向的未来发展趋势进行了精彩的观点分享。



29日下午首先进行了主题二“计算机视觉训练集如何构建”的研讨。该主题由爱奇艺资深科学家王涛博

士、华中科技大学白翔教授、华中科技大学杨欣教授 3 位委员负责组织，邀请了爱奇艺科学家路香菊博士、百度资深研发工程师孙逸鹏博士、南京理工大学魏秀参教授、武汉大学夏桂松教授共 4 位嘉宾进行主题发言。近年来，以 ImageNet 等为代表的视觉数据集极大地推动了视觉算法的发展。如今各种新型数据集层出不穷，那么计算机视觉数据集如何进行高效构建、有哪些经验值得借鉴？嘉宾们围绕数据集高效设计、数据集均衡性、未知新样本泛化性等议题展开了深入探讨。



最后，研讨会闭幕式由专委会副主任、南京信息工程大学刘青山教授主持。本次研讨会从任务、数据、算法三方面探讨了本领域最前沿研究问题，主题发言视角广阔且多元化，自由讨论热情激烈、你来我往，参会嘉宾们纷纷表示本次会议内容丰富，收获颇丰。按照计划，组委会后续将整理相关主题的发言与讨论文稿，形成观点性文档进行发布，把讨论从线下延伸到线上，欢迎更多专家学者积极参与。本次研讨会由厦门大学王菡子教授与纪荣嵘教授主要负责组织，专委会秘书处成员协助会务组织的相关事宜。



然后，29 日下午还进行了主题三“网络结构自动搜索”的研讨。该主题由厦门大学纪荣嵘教授、北京邮电大学马占宇教授、厦门大学孙晓帅副教授 3 位委员负责组织，邀请了暗物智能科技研究副总监苏江博士、鹏程实验室陈杰博士、旷视研究院张祥雨博士共 3 位嘉宾进行主题发言。网络结构自动搜索是近期最受关注的研究方向之一，如何自动获得网络结构对于深度学习实际应用具有重大意义，那么相关方向的研究进展与未来趋势有哪些？嘉宾们围绕最优结构评估、网络自适应改进、生物大脑之间联系等议题展开了深入探讨。



责任编辑 任传贤

CCF-CV 专委会多名委员当选 IAPR Fellow

国际模式识别协会是由世界各国模式识别方面的专家和团体组成的国际性学术组织，成立于 1978 年。国际模式识别协会会员 (IAPR Fellow) 于 1994 年设立，授予模式识别领域的杰出科研工作者及活动组织者。作为模式识别领域最权威的国际性学术组织之一，该协会会员每两年评选一次，当选率少于会员总数的 0.25%。2020 年，全球共 23 人当选 IAPR Fellow，其中包括 CCF 计算机视觉专委会白翔、赫然、鲁继文共 3 位委员，以表彰他们在模式识别与计算机视觉领域做出的贡献！

白翔 华中科技大学电信学院教授、副院长，国家防伪工程技术研究中心副主任，国家自然科学基金优秀青年科学基金获得者。主要研究方向为计算机视觉与模式识别、文档分析等。已在计算机视觉与模式识别领域一流国际期刊和会议如 PAMI、CVPR 等发表论文 60 余篇。由于在场景文字检测与识别取得的突出成就，2019 年在国际模式识别协会主办的国际文档分析与识别会议上获得 IAPR/ICDAR Young Investigator Award (青年学者奖)。现任期刊 PR, PRL, Frontier of Computer Science, 中国科学信息科学、自动化学报编委。曾担任 CVPR20、AAAI20、CVPR19、AAAI19、ICPR18 等人工智能主流国际会议的领域主席/资深技术程序委员或竞赛主席，并获得 AAAI-2019 Outstanding SPC Award。他是 IEEE 高级会员，中国计算机学会计算机视觉专委会常务委员和青年工作组组长，于 2014-2018 连续 5 年入选 Elsevier 中国高被引学者榜。

赫然 中国科学院自动化研究所模式识别国家重点实验室研究员，中国科学院脑科学与智能技术卓越创新

中心年轻骨干。2009 年于中国科学院自动化研究所获模式识别与智能系统工程博士。2010 年至今历任模式识别国家重点实验室任助理研究员、副研究员、研究员。从事模式识别应用基础理论研究，并应用到生物特征识别和人工智能安全。近期聚焦在概率深度学习以及图像生成中的瓶颈问题，展开图像模式分析基础理论研究。出版信息理论学习专著 1 部；在 IEEE TPAMI、TIP、TIFS、TNNLS、TKDE、IJCV 等权威国际期刊以及 NIPS、ICCV、CVPR、ECCV、IJCAI、AAAI 等权威国际会议发表论文 150 篇，研究工作获国家自然科学基金优秀青年科学基金、北京自然科学基金杰出青年科学基金和中科院青年促进会优秀会员资助。受邀参加 70 周年国庆阅兵观礼 (2019)、中央电视台科教频道《透视新科技》科普栏目 (2020)。

鲁继文 清华大学自动化系副教授，博士生导师，国家自然科学基金优秀青年科学基金获得者，主要研究兴趣包括计算机视觉、模式识别和智能机器人。第一/通讯作者发表 IEEE 汇刊论文 60 余篇 (其中 PAMI 论文 15 篇) 和 CVPR/ICCV/ECCV 会议论文 60 余篇，谷歌学术引用 11000 余次，H-index 56，主持国家自然科学基金联合重点基金 1 项和国家重点研发计划课题 1 项。担任国际期刊 Pattern Recognition Letters 主编，IEEE T-IP/T-CSVT/T-BIOM 和 PR 编委，国际会议 IEEE FG 2023、IEEE VCIP 2022、IEEE AVSS 2021 和 IEEE ICME 2020 程序委员会主席，CVPR 2020/2021 领域主席，中国计算机学会计算机视觉专委会常务委员，中国人工智能学会模式识别专委会常务委员。

责任编辑 黄岩

PRCV 2022 征集承办单位



中国模式识别和计算机视觉大会 (PRCV, Chinese Conference on Pattern Recognition and Computer Vision) 是全国模式识别学术会议 (CCPR)、中国计算机视觉大会 (CCCV) 等相关会议合并后联合举办的全国性学术会议。

全国模式识别学术会议 (CCPR) 由中国自动化学会、人工智能学会和模式识别国家重点实验室联合主办，2007 年、2008 年在北京举办 (中科院自动化所承办)，2009 年在南京举办 (南京理工大学承办)，2010 年在重庆举办 (重庆大学承办)，2012 年在北京举办 (中科院自动化所承办)，2014 年在长沙举办 (湖南大学承办)，2016 年在成都举办 (电子科技大学承办)。中国计算机视觉大会 (CCCV) 由中国计算机学会主办、计算机视觉专委会承办。首届 CCCV 于 2015 年在西安举办 (西安电子科技大学承办)；第二届 CCCV 于 2017 年 10 月 12-14 日在天津举办 (中国民航大学承办，天津大学和南开大学协办)。

中国人工智能学会模式识别专委会、中国计算机学会计算机视觉专委会、中国自动化学会模式识别与机器智能专委会、中国图象图形学学会视觉大数据专委会四

个专委会的委员名单重叠度高，每年各专委会组织的相关学术年会或研讨会 (如 CCPR、CCCV 等) 内容相近。因此，为了推进本领域的学科发展，打造一个高质量的学术交流与合作平台，经讨论一致同意合并过去各个专委会的类似会议，联合举办中国模式识别与计算机视觉大会。

2018 年起，中国模式识别和计算机视觉大会 (PRCV) 由中国人工智能学会、中国计算机学会、中国自动化学会和中国图象图形学学会联合主办，每年举办一次。2018 年由中山大学 (广州) 等承办 (参会人数 1600+)，2019 年由西北工业大学 (西安) 等承办 (参会人数 1800+)，2020 年由南京理工大学 (南京) 等承办，2021 年由北京科技大学 (北京) 等承办，现征集 2022 年承办单位。

会议地点

中国模式识别和计算机视觉大会 (PRCV) 将在全国各地城市轮流举办。2022 年的举办地应该是在广州、西安、南京、北京以外的地区。

会议时间

会议召开时间应该在气候比较适宜的季节，一般在秋季，以 9-11 月为宜。2022 年 PRCV 应该避开领域内主要的相关国际学术会议的时间。

关于组委会

承办单位应该有一个实验室或部门在国内模式识别和计算机视觉界有一定学术基础和组织能力。大会主席和程序委员会主席应由本领域的知名学者担任 (根据 PRCV 指导委员会建议，程序委员会主席每个专委会各派一人担任)。请不要把非本学术领域的领导列为大会主

席或程序委员会主席，担任实际组织工作的组织委员会主席除外。

承办单位在申请时不能确定大会主席和程序委员会主席的，可先空缺，在确定承办会议后，由协调指导委员会推荐指定大会主席和程序委员会主席。

会议程序

PRCV 论文集将由 Springer 正式出版。

会议将公开征集论文，录用论文分口头报告(Oral)和张贴报告(Poster)。要求每篇录用论文至少有一名作者注册参加会议并报告论文，否则该论文将从论文集删除。

主会会议日程一般为 3 天。会议上除了特邀报告和论文报告，还可以组织专题论坛、讲习班等形式的交流和讨论。

财务预算

会议积极争取企业赞助，同时适当收取注册费，用于支付论文集出版和会议举办过程中的实际费用。对人工智能学会、计算机学会、自动化学会、图象图形学学会会员，以及学生，注册费应至少优惠 20%。

会议结束后，组委会应向第一主办学会和协调指导委员会提交详细的财务收支报告，会议结余资金应向联合主办学会上交。

承办申请书

申请书应包含以下内容：

1. 会议时间、地点（高校、酒店或会议中心）
2. 承办单位研究基础、会议设施、接待设施介绍
3. 组委会主要人员及其研究经历、组织经验
4. 程序安排基本想法介绍
5. 财务预算，包括注册费标准、预计收支情况等

请在 2020 年 10 月 1 日前将申请书发给中国模式

识别和计算机视觉大会协调指导委员会联系人王亮。申请人将被邀请在 2020 年 PRCV 上介绍会议申请报告。如有超过一个申请单位，将在会上由参会代表投票决定承办单位。

附：中国模式识别和计算机视觉大会 (PRCV) 协调指导委员会

主席

谭铁牛（中科院自动化所）

委员（按姓氏拼音排序）

陈熙霖（中科院计算所）

刘成林（中科院自动化所）

权 龙（香港科技大学）

芮 勇（联想集团）

查红彬（北京大学）

郑南宁（西安交通大学）

周 杰（清华大学）

秘书长

王 亮（中科院自动化所）

联系人：北京中科院自动化所模式识别国家重点实验

王 亮（中科院自动化所）

邮编：100190

电话：010-82544643

邮箱：wangliang@nlpr.ia.ac.cn

责任编辑 黄岩

CCF-CV "持久影响力论文奖"

为了鼓励原创性、基础性并能够产生持久影响力的创新性研究，中国计算机学会计算机视觉专委会（CCF-CV）自 2020 年度开始设立了“持久影响力论文奖”。根据《CCF-CV 奖励条例》和《CCF-CV 持久影响力论文奖评选条例》，经过形式审查、提名与奖励工作组初筛和评审委员会专家终评，评选出首届（2020 年度）CCF-CV “持久影响力论文奖”获奖论文。

获奖论文 《Guided Image Filtering》

作者：Kaiming He, Jian Sun, Xiaoou Tang

出处：ECCV 2010 (Oral)

推荐理由：

Image filtering plays an important role in image processing and graphics. This ECCV paper presents a general-purpose image filter named guided filter, which generates the filtering output by considering the content of a guidance image. It can perform as a fast edge-preserving smoothing operator with better behavior near the edges. These make the guided filter both effective and efficient in a great variety of computer vision and computer graphics applications such as noise reduction, detail smoothing, HDR compression, image matting, haze removal. Concededly, the technologies of this paper are included in many official image/graphics processing toolboxes, e.g.,

Matlab2014, OpenCV3.0, Pytorch. Undoubtedly, this paper has a significant impact on the fields related to computer vision and computer graphics. Many academic researchers and commercial companies have benefited greatly from the methods presented and the results reported in this ECCV paper.

In addition, this ECCV paper has received more than 5000 citations in 10 years since its publication. More remarkably, the citation number has been going up year over year, showing its lasting influence. It has been extensively cited as a seminal work in low-level vision tasks. Many industrial product such as Adobe Photoshop, Google Photo, have adopted the technology developed based on idea of guided filtering. Overall, this paper stands out as the winner due to its long-lasting and wide-reaching impact.

评审委员会名单：

卢湖川（组长），华刚，罗杰波，山世光，沈春华，陶大程，颜水成，查红彬

中国计算机学会计算机视觉专委会

提名与奖励工作组

责任编辑 任传贤

CCF-CV "中科视拓 Seeta 学术新锐奖"

根据《CCF-CV 奖励条例》和《CCF-CV 学术新锐奖评选条例》，经过形式审查、提名与奖励工作组初评和评审委员会专家终评，评选出三名 2020 年度 CCF-CV "中科视拓 Seeta 学术新锐奖" 获奖者。

2020 年 CCF-CV "中科视拓 Seeta 学术新锐奖" 名单

获奖者姓名：李铎

性别：男

单位：香港科技大学

博士生类型：硕博

入学时间：2019 年 9 月

导师：陈启峰

推荐人 1：章国峰

推荐人 2：王瑞平

获奖者姓名：高尚华

性别：男

单位：南开大学

博士生类型：硕士

入学时间：2018 年 9 月

导师：程明明

推荐人 1：颜水成

推荐人 2：王亮

获奖者姓名：彭思达

性别：男

单位：浙江大学

博士生类型：硕博

入学时间：2018 年 9 月

导师：周晓巍

推荐人 1：章国峰

推荐人 2：郭裕兰

颁奖将于 2020 年 10 月 16 日晚上 CCF-CV 专委会的工作会议上举行。入围终评者将在 10 月 17 日下午 PRCV2020 大会的“博士生论坛”上介绍其研究工作。

“学术新锐奖”评审委员会名单

卢湖川（组长）、白翔、刘偲、鲁继文、王井东

中国计算机学会计算机视觉专委会

提名与奖励工作组

责任编辑 任传贤

热点追踪

基于元学习的泛化人脸识别

中科院自动化研究所 郭建珠 朱翔昱 雷震

人脸识别模型通常需要部署在未知的场景中，识别未知的人群。由于训练数据和目标场景的分布偏差，模型在未知场景的识别性能往往表现很差，比如，在网图训练的模型部署在监控场景中，识别的准确率会大打折扣，在这对模型的泛化性能带来了极大的挑战。为解决该技术难题，本文提出了一种基于元学习的泛化人脸识别框架。该框架在训练采样过程中，通过采样元训练域和元测试域来模拟场景的分布偏差，并通过一种提出的元优化的损失，使得模型同时在元训练和元测试域上都能有提升，进而提升模型在真实未知场景下的泛化性。相关成果被 CVPR2020 录取为口头报告。

近些年得益于深度学习的发展，人脸识别性能在一些通用测试集上得到了极大提升^[1,2,3,4,5]，然而这些测试集与训练集有着类似的数据分布。当识别模型部署在实际场景中时，由于目标场景与训练集数据分布不一致，模型的性能会显著下降。并且目标场景通常是未知的，数据亦不可获取，无法使用目标场景的数据对模型进行微调。针对这个挑战，我们提出了泛化人脸识别问题：在目标场景未知的条件下，如何设计一个有效的训练策略或方法，让模型能在未知场景下取得较好的泛化性，

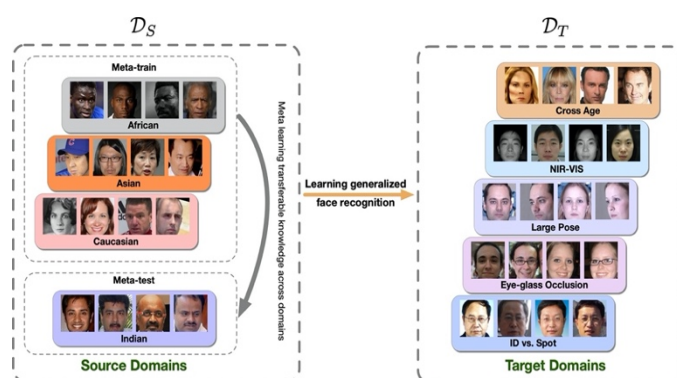


图 1 泛化人脸识别

如图 1 所示。

针对泛化人脸识别问题，我们提出了一种基于元学习的人脸识别框架 MFR (Meta Face Recognition)。MFR 主要包括三部分：(1) 跨域采样；(2) 多域分布优化；(3) 元优化。整体的框架如图 2 所示。

首先，跨域采样是为了模拟训练场景和测试场景的分布偏差，每次迭代时，根据训练集的域标签，将训练集分为元训练域和元测试域，并在两个域中分别采样一定人数。其次，在多域分布优化中，我们使用了三种损

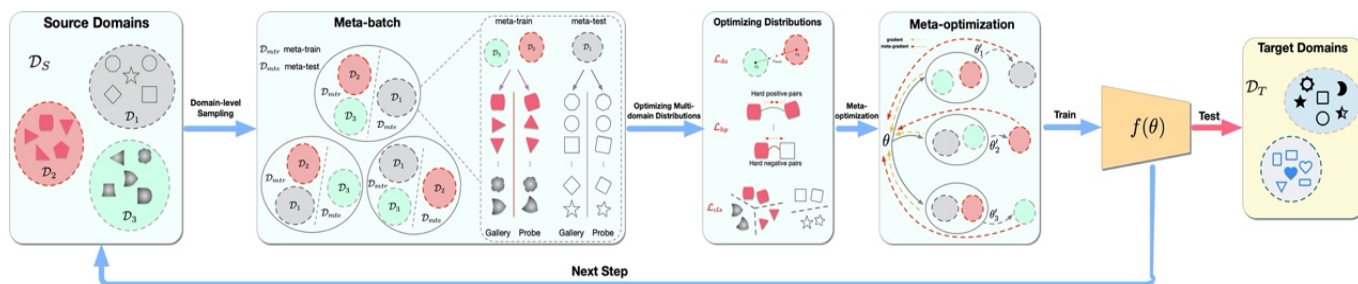


图 2 MFR 整体框架

失函数，包括难样本损失、软分类损失和域对齐损失，来学习具有判别性和域不变性的人脸表征。最后，元优化通过三个步骤对模型进行优化。(1) 元训练：对元训练域进行优化，并得到梯度更新后的模型参数；(2) 元测试：在元测试域上，对元训练更新后的模型参数进行二次更新；(3) 对元优化和元测试的损失进行加权，对原始模型的参数进行梯度反传更新。元优化的示意图如图 3 所示。

为了验证模型的泛化性，我们提出了两个不同难度的测试协议，如表 1 所示。一个是跨种族（印度人、非洲人、亚洲人，高加索人）测试协议 GFR-R；另一个是跨场景测试协议 GFR-V，其更接近实际场景，也更具挑战性。在协议中，目标域的数据在训练中是未知的，用于模拟未知的应用场景。通过表 2 可以看出，通过元学习，本方法在跨种族和跨场景的测试中均取得了最好的性能。

Protocol	Source Domains	Target Domain(s)	
GFR-R	I	Caucasian	Indian
		Asian	
		African	
		Indian	
	II	Caucasian	African
		Asian	
		Indian	
		Indian	
	III	Caucasian	Asian
		African	
		Indian	
		Indian	
	IV	Asian	Caucasian
		African	
		Indian	
		Indian	
GFR-V	Caucasian	CACD-VS	
	Asian	CASIA NIR-VIS 2.0	
	African	MultiPIE	
	Indian	MeGlass	
	Indian	Public-IvS	

表 1 GFR-R 跨人种测试协议和 GFR-V 跨域测试协议

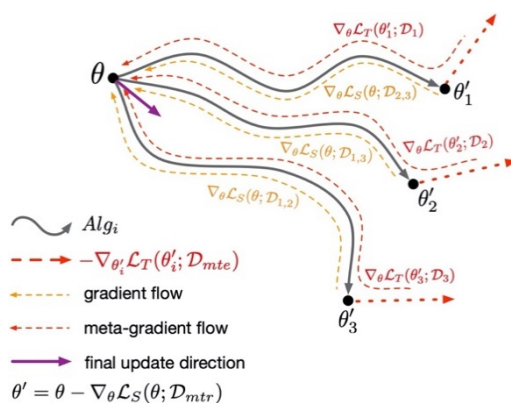


图 3 元优化

Protocol	Method	VR (%)			Rank-1 (%)
		FAR=1%	FAR=0.1%	FAR=0.01%	
GFR-R I (Indian)	Base	94	82.2	64.65	80.3
	Base-Agg	94.1	80.9	65.3	81
	Base-FT rnd.	62.5	39	21.05	39.3
	Base-FT imp. [34]	87	69.9	51.2	69.6
	MLDG [12]	94.2	83	66.3	80.5
	MFR (Ours)	95.4	86.1	71.4	83.1
GFR-R II (African)	Base	91.6	74.5	55.4	73.1
	Base-Agg	90.5	74.8	56.3	74
	Base-FT rnd.	26.2	10.9	3.5	21
	Base-FT imp. [34]	78.7	56.6	36.45	57.9
	MLDG [12]	91.9	74.8	55.7	73.8
	MFR (Ours)	92.3	79.4	60.8	75.2
GFR-R III (Asian)	Base	91.89	77.98	60.86	75.98
	Base-Agg	91.49	78.08	59.41	76.28
	Base-FT rnd.	40.44	17.32	7.67	27.53
	Base-FT imp. [34]	80.58	57.56	39.79	61.86
	MLDG [12]	92.29	78.28	60.3	76.68
	MFR (Ours)	93.49	80.7	62.56	78.68
GFR-R IV (Caucasian)	Base	96.6	89.6	78.6	86.6
	Base-Agg	97	88.1	79.1	86.8
	Base-FT rnd.	61.1	36.2	18.9	36.7
	Base-FT imp. [34]	91.5	78.2	63.4	76.8
	MLDG [12]	96.8	89.6	79.15	86.3
	MFR (Ours)	98.2	92.9	81.1	88.9
GFR-V (MeGlass)	VR (%)				
		FAR=0.01%	FAR=0.001%	FAR=0.0001%	Rank-1 (%)
	Base	85.92	71.96	53.5	97.6
	Base-Agg	86.77	73.5	54.96	97.69
	MLDG [12]	85.54	69.23	49.32	97.81
	Face Syn. [31]	90.14	80.32	66.92	96.73
MFR (Ours)	90.79	80.86	66.15	98.57	

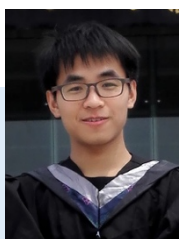
表 2 GFR-R 跨人种测试协议和 GFR-V 跨域测试协议结果

责任编辑 崔海楠

参考文献

- [1] Yaniv Taigman, Ming Yang, Marc'Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1701–1708, 2014.
- [2] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1891–1898, 2014.

- [3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823, 2015.
- [4] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphreface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 212–220, 2017.
- [5] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In CVPR, 2019.



郭建珠

中科院自动化研究所博士生。主要研究方向为人脸识别和三维人脸。
Email: jianzhu.guo@nlpr.ia.ac.cn



朱翔昱

中科院自动化研究所副研究员。主要研究方向为人脸识别、人脸对齐和三维人脸。
Email: xiangyu.zhu@nlpr.ia.ac.cn



雷震

中科院自动化研究所研究员。主要研究方向为人脸检测、人脸对齐、人脸识别、计算机视觉、模式识别、人工智能、图像处理等。
Email: zlei@nlpr.ia.ac.cn

热点追踪

多图的离线协同与在线增量匹配

上海交通大学 严骏驰

图 匹配问题一直是计算机视觉领域中一个十分基础而重要的问题，其宗旨在于寻找两个或多个存在复杂关系的图集合间元素的对应关系。图匹配最经典的应用之一在于将不同图片中的特征点进行匹配，这是包括图像匹配、图像识别与理解、三维多视角重建、几何物体形状匹配、运动结构理解等很多计算机视觉问题的关键一环。另一方面，图匹配一般是指在给定相似性模型下，寻找两个或多个图的节点之间的对应关系。其二图匹配的形式一般被规约成劳勒的二阶指派问题（Lawler's Quadratic Assignment Programming），这已经被证明为 NP 难问题，而基于二图匹配提出的多图匹配形式则更加复杂，难度也更大。图匹配领域的突破也在为研究

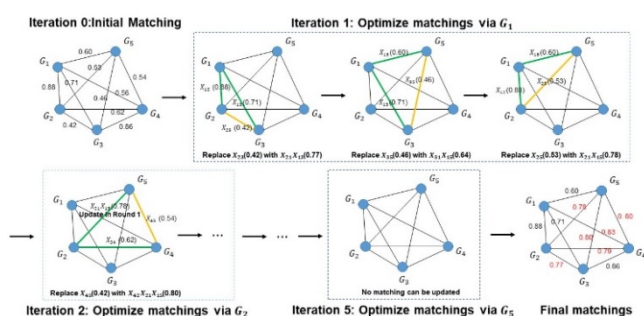


图 2 MGM-Floyd 算法流程

不错的效果。该方法遵循匹配链优化的思想来解决多图匹配问题，通过搜索匹配之间的组合来优化原有的二图匹配。通过超图的视角看待问题，超图的顶点表示需要匹配的图，超图的边表示两个图之间的匹配，搜索匹配链的组合方式相当于在超图上搜索最优的路径。研究者证明了由相似度得分和循环一致性组成的匹配质量函数具有与距离函数相似的性质，并通过严密的推理论证最短路算法与多图匹配存在结合的理论基础。基于这一点，融合了经典 Floyd 算法的离线多图求解器 MGM-Floyd 与融合了单源最短路 SPFA 算法的在线多图求解器 MGM-SPFA、FastSPFA 被提出。相较于现有

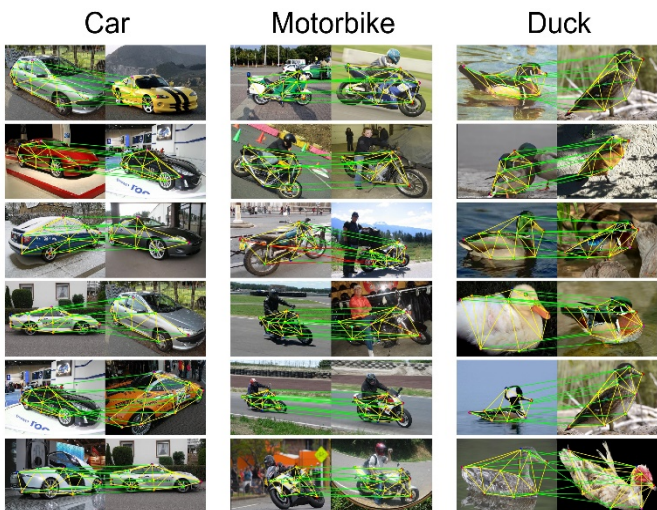


图 1 Willow Object 数据集上 Floyd 方法匹配效果

NP 难的组合优化问题作出贡献。

近期，基于超图上的最短路径优化算法引入多图匹配领域，同时在离线协同匹配与在线增量匹配领域取得

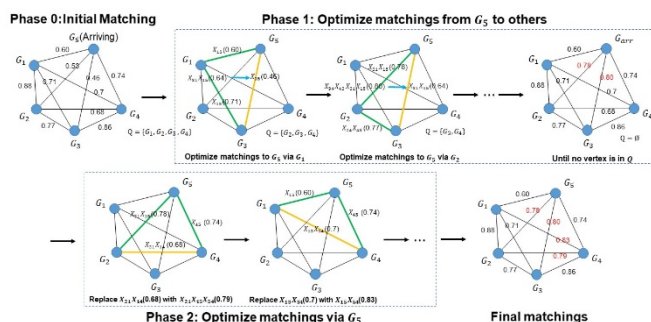


图 3 MGM-SPFA 算法流程

category	Hungarian	Floyd-pc	Floyd-uc	Floyd-c	Spectral [20]	MatchLift [21]	MatchALS [22]	MISM [5]
Car	0.503	0.844	0.840	0.850	0.601	0.665	0.629	0.750
Duck	0.442	0.803	0.800	0.793	0.485	0.554	0.525	0.732
Motorbike	0.317	0.821	0.817	0.843	0.255	0.296	0.310	0.653
Face	0.854	1.000	1.000	1.000	0.927	0.931	0.934	0.937
Winebottle	0.543	0.934	0.930	0.931	0.630	0.700	0.669	0.814
Time (Car)	1.263	9.377	10.960	14.981	1.469	17.695	2.622	3.966

表 1 不同离线多图匹配求解器在 Willow-Object 数据集上的匹配精度以及匹配速度的表现

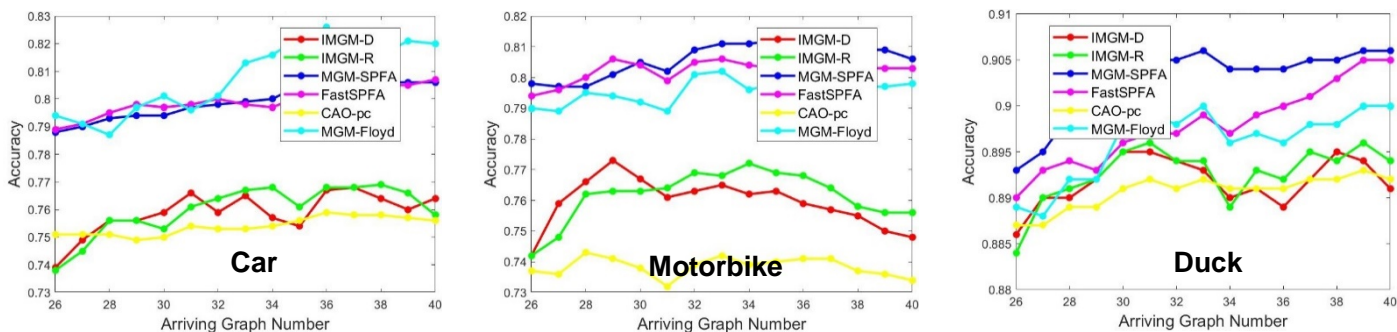


图 4 不同在线多图匹配求解器在 Willow-Object 数据集上的匹配精度的表现

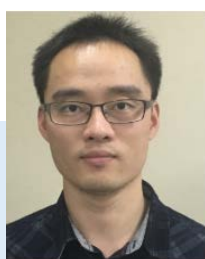
的求解器，这些算法在时间、精度等多个方面都有较大的提升。

这些算法都在仿真数据集和真实数据集上都进行了实验。仿真测试数据集包括形变测试，外点测试，遮罩测试等多种环境下的效果测试；真实数据集则包括 Willow-ObjectClass 和 Sub-Rome16k 两个真实数据集。从综合的角度来看，本文提出的方法在所有数据集的离线和在线实验中都拥有着最佳的性能。在离线模式下，MGM-Floyd 的综合表现上比较突出，相较于 SOTA，在对应模式下或是精度或是速度都有着明显的优势，在与其他连续空间离线多图匹配求解器比较的时候精度优势则更大，精度有着 10%-20% 的提升。在线模式中，MGM-SPFA 在在线多图增量匹配问题中精度最高，而

FastSPFA 精度上与 MGM-SPFA 比较接近，速度上比其他的在线多图匹配求解器快了将近 1-2 倍。与此同时，经过试验检验，以上所有算法都有着相当高的稳定性与鲁棒性，可以适应复杂的现实场景图片匹配应用。

以上工作已发表于国际期刊 IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)。具体题目为：Unifying Offline and Online Multi-graph Matching via Finding Shortest Paths on Supergraph。三位作者是蒋泽天 (本科生)，王天哲 (本科生)，严骏驰 (副教授)，均来自上海交通大学。

责任编辑 储璐



严骏驰

上海交通大学计算机系与人工智能研究院长聘教授。主要研究方向为机器学习与计算机视觉。

Email: yanjunchi@cs.sjtu.edu.cn

热点追踪

上下文感知的密集视频描述生成

中国人民大学 陈师哲 金琴

视频描述生成是全面理解视频语义内容的一个重要任务。传统视频描述生成任务主要是为视频生成单句话的文本描述。然而，长时间视频中往往包含多个事件，难以用单句话概括。因此，越来越多的研究开始关注于密集视频描述生成(Dense Video Captioning)任务^[1]。这一任务旨在检测长时间视频中的所有事件，并分别为每一事件生成文本描述。图 1 展示了密集视频描述的一个示例。目前的方法主要将这一任务分解成两部分进行解决，分别是事件检测和事件描述生成，其中事件检测模型负责在长时视频中检测出有意义的事件，事件描述模型为检测的每一个事件生成一句文本描述。

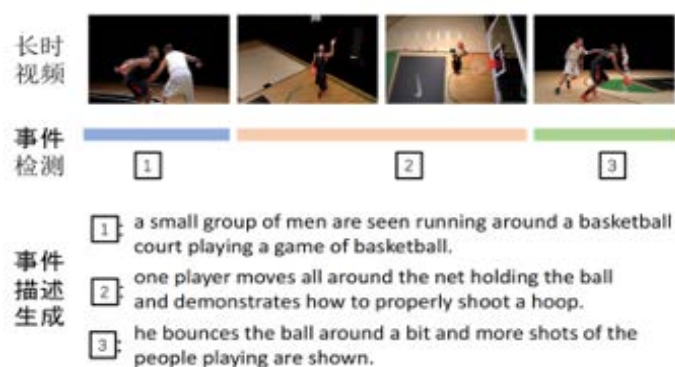


图 1 密集视频描述示例

为了推动密集视频描述这一研究任务的发展，CVPR ActivityNet Workshop 从 2017 年开始连续 4 年举办了 Dense-Captioning Events in Videos 比赛。该挑战赛由斯坦福大学组织，基于 ActivityNet 密集视频描述数据集，共有 20000 个视频，平均每个视频包含 3.7 句事件描述。该竞赛吸引了卡内基梅隆大学、北京大学、清华大学、复旦大学、浙江大学等国内外科研

院校，以及百度、京东、腾讯、三星等公司和研究院参加。我们团队在该比赛上从 2018 年到 2020 年连续 3 年获得了冠军。下面简单介绍我们提出的方案——上下文感知的密集视频描述生成模型。

上下文关系对于检测和理解长视频中的不同事件至关重要。首先，视频中的不同事件不是独立的，上下文为理解每个事件提供更全面的信息，使得事件语义理解更为准确；其次，上下文关系有利于区分不同事件，不仅可以使得检测的事件片段更为多样化，而且使得生成的事件文本描述更具有区分性。

对于事件检测，我们提出了基于时序上下文的事件检测模型^[2]，图 2 展示了所提出模型的结构框架。该模型采用 GRU 网络直接进行事件时间戳的预测，输入为前一步所生成的事件片段的位置和语义特征，从而可以根据上文信息大大减少生成事件片段的冗余性。为了更有效利用事件的双向时序信息，我们融合从前往后以及从后往前的事件时间戳预测结果。实验结果表明我们提出的模型可以生成更加准确的事件片段候选。

对于事件描述生成，我们系统地探讨和比较了不同上下文特征和模型的影响^[3]。我们提出五种上下文类型，分别是上下文增强的帧级特征、局部上下文、全局上下文、事件上下文和句子上上下文，以及两种上下文编码模型，分别是事件独立模型和事件相关模型。图 3 展示了上下文感知的事件独立模型示意图。实验结果表明，利用上下文增强的帧级特征和局部上下文的事件独立模型在速度和性能上达到较好平衡，在事件描述准确性和多样性两个方面均有提高。

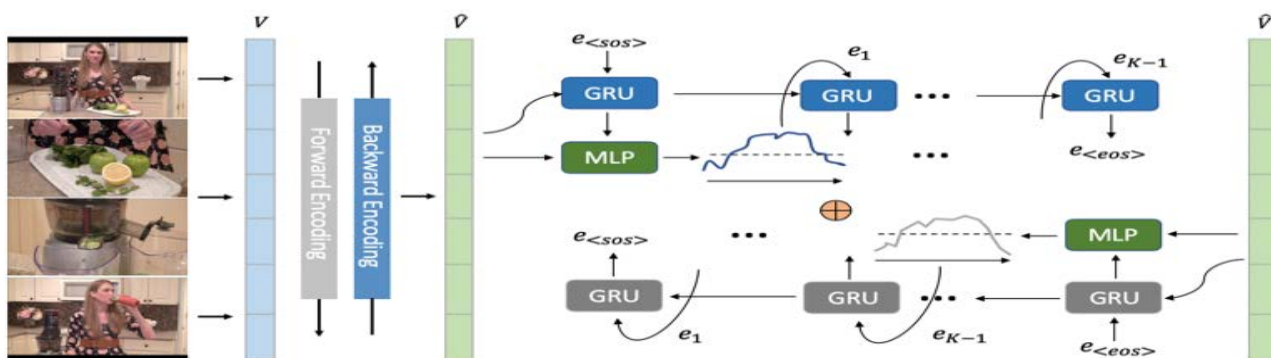


图2 基于时序上下文的事件检测模型

密集视频描述为全面和细粒度的视频语义理解提供了新的途径，但目前的研究还较为初步，还存在着许多问题值得探索。例如，使得生成的不同事件描述更具有连续性，形成逻辑连贯的段落描述，而不是独立的事件描述。现有评测标准也具有局限性，仅从检测的精度和描述准确率考虑，缺乏对事件检测的召回率和描述多样性等不同方面的综合考量。

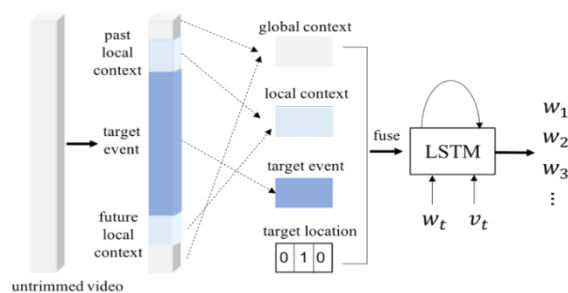
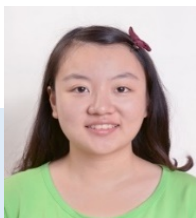


图3 上下文感知的事件描述生成模型

责任编辑 王金甲 任桐炜

参考文献

- [1] Krishna, Ranjay, Kenji Hata, Frederic Ren, Li Fei-Fei, and Juan Carlos Niebles. "Dense-captioning events in videos." In Proceedings of the IEEE international conference on computer vision, pp. 706-715. 2017.
- [2] Song, Yuqing, Shizhe Chen, Yida Zhao, and Qin Jin. "Team RUC_AIM3 Technical Report at Activitynet 2020 Task 2: Exploring Sequential Events Detection for Dense Video Captioning." arXiv preprint arXiv:2006.07896 (2020).
- [3] Chen, Shizhe, Yuqing Song, Yida Zhao, Qin Jin, Zhaoyang Zeng, Bei Liu, Jianlong Fu, and Alexander Hauptmann. "Activitynet 2019 task 3: Exploring contexts for dense captioning events in videos." arXiv preprint arXiv:1907.05092 (2019).



陈师哲

中国人民大学信息学院博士，主要研究方向为视觉与自然语言、视频语义理解和多模态计算。
Email: cszhel@ruc.edu.cn



金琴

中国人民大学信息学院教授，主要研究领域包括多媒体智能计算、人机交互等。
Email: qjin@ruc.edu.cn

顶会观察

CVPR 2020

中山大学 李冠彬

国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 是计算机视觉及模式识别的顶级学术会议, 与 ICCV、ECCV 并称为计算机视觉领域三大顶会。CVPR 有着非常高的学术影响力, 在中国计算机学会推荐国际学术会议被评为人工智能领域的 A 类会议; 在 Google Scholar 发布的 2020 年学术指标中, 以高达 299 的 h5 指数雄踞人工智能领域的榜首、位列所有领域出版物的第五位。因此, CVPR 每年都吸引来自世界各地的参会者, 其中不乏学术界和工业界的泰斗大牛、广罗人才的企业机构, 以及奋战于科研一线的研究生、博士生、教授、研究员等。

CVPR 2020 原定于 2020 年 6 月 14 日至 19 日于美国华盛顿州西雅图举行, 由于 COVID-19 新冠病毒在世界各地陆续爆发, 本届 CVPR 转为完全线上虚拟会议方式进行, 给予了各个参会人员全新的参会体验。

一、国际计算机视觉与模式识别会议的亮点

在大会举办日期的前几个月, 新冠肺炎疫情在世界各地各个国家陆续爆发, 因此 CVPR 2020 首次采用了完全数字虚拟会议的方式, 让世界各地的参会者通过互联网参加虚拟会议。会议收录的论文作者需要准备各种形式的材料, 包括摘要、图片、论文讲解视频等, 作为虚拟会议的交流媒介。会议主办方通过邮件进行协调, 并建立了虚拟会议的网址, 以供参会人员进行展示及技术交流。每篇论文的简介及讲解视频会根据论文主题进行分类, 在网页上进行展示。且每篇论文安排了两次线上交流时间, 允许作者与感兴趣的参会者通过文本聊天或线

上会议的方式进行答疑与交流, 这为不同时区的参会者提供了非常大的便利。本次虚拟会议在保证安全的前提下, 最大限度地保障了参会者的自由度、便利程度, 让大家获得了很好的参会体验及交流经历。

二、论文录用情况

CVPR 2020 总共收到了 6656 篇有效投稿, 其中 1470 篇论文被接收, 接收率约为 22%, 从 CVPR 2018 开始论文接受率连续两年下降, 而今年创下近十年来 CVPR 会议的最低接收纪录。其中仅有 335 篇论文入选 oral presentation, oral 率约为 5.7%。从大会公布的数据来看, 今年大会接收到的注册及有效投稿数量都有显著的提高, 可见被接受论文的质量之高。

本次 CVPR 2020 收录论文中, 中国作者占论文作者总人数的 39.2%, 占比为所有国家中最高。从大学/机构来看, 来自清华大学的作者人数最多, 排名第一。其他贡献突出的中国高校还有上海交通大学、北京大学、浙江大学等。除了学术界的科研力量, 来自中国工业界的各大互联网企业也获得了不俗的成绩, 其中包括华为、阿里巴巴、商汤、百度、腾讯、旷视等。其中, 商汤及联合实验室共 62 篇论文入选, 华为研究团队有 34 篇论文入选, 阿里巴巴、旷视分别有 16 篇论文入选。这些被录用的论文在许多重要工业应用领域有着重大的突破, 包括网络结构搜索、三维点云理解、网络模型训练加速与量化、视频理解领域等。此外, 谷歌在本次会议表现亮眼, 共有近 70 篇论文入选, 涵盖了诸多热门领域, 如图像识别、自动驾驶、语义分割、目标追踪等。

CVPR 2020 会议涵盖的方向包括目标检测、目标

跟踪、图像分割、人脸识别、姿态估计、三维点云、视频分析、模型加速、GAN 和 OCR 等，论文关键词包括迁移学习 (Transfer Learning)、表示学习 (Representation Learning)、机器人 (Robotics)、三维数据分析、无监督学习 (Unsupervised Learning) 和自监督学习 (Self-supervised learning) 等。

三、主题演讲

今年 CVPR 2020 邀请到了两位重量级嘉宾进行炉边谈话，他们分别是微软 CEO Satya Nadella 与亚马逊云服务 SVP Charlie Bell。Satya 在与微软公司前执行副总裁沈向洋的对话中，分享了对计算机视觉、人工智能研究与应用前景的思考与展望。Satya 首先回顾了自己从印度生活开始，到开始在美国学习计算机科学并最终成为微软 CEO 的经历，指出“予力全球每一人、每一组织，成就不凡”这一微软使命是鞭策他每天前行努力工作的根基。Satya 也列举了计算机视觉当今的一些精彩应用，以及未来发展的热点方向。计算机视觉技术为工业界许多场景提供了有力的支持，包括计算机视觉在 3D 场景下的应用，在云和边缘设备上的创新应用，在疾病筛查与诊断场景下的应用。关于计算机视觉技术的未来发展热点，Satya 指出了三个突破方向，包括 4D Understanding (4D 理解)，Background Matting (背景替换)，Holoportation (全息瞬移)。4D 理解能够让计算机理解跟踪物品、人、互动行为及群组活动；背景替换能够在不适用绿幕情况下将演讲者的图像投影到虚拟舞台上，制造更真实的会议体验；全息瞬移则能超越时空及语言局限，使得演讲者能够用多种语言和形象进行流利的演讲。此外，Satya 还对人工智能接下来的发展方向、应用领域、社区间的国际合作、行业机遇、AI 设计的道德问题等分享了自己的观点。

值得一提的是，本届 CVPR 特意缅怀了 84 岁华人计算机视觉泰斗 Thomas S. Huang (黄煦涛)。黄煦涛教授是首位担任 CVPR 程序主席的华人，也是华人计算机视觉史上的关键人物，他为华人 CV 领域留下丰厚的学术遗产。经 PAMITC 执行委员会批准，CVPR 设立计算机视觉领域的 Thomas S. Huang 纪念奖，从 2021 年开始每年在 CVPR 上颁发。

四、会议热点论文与教程

本次会议涌现了许多热点论文，它们具有非常高的学术价值与应用价值。其中，CVPR 2020 的最佳论文^[1]、最佳学生论文^[2]一作均为华人，且均聚焦于 3D 领域。最佳论文^[1]亮点在于仅给定一张单目图像，可在解构拍摄视角的同时，将其深度 (3D)、光照等分解出来，采用无监督的方式估计 3D 对象，还原物体原始的 3D 面貌。传统的 3D 重建方法需要有监督的信息 (如多视图、深度图、轮廓、关键点)；当训练数据足够多，模型训练成功以后，才可基于双目图像估计出深度图，该类方法通常需要耗费高昂的成本。而本文基于自动编码器解构给定图像的视角、深度、texture 等维度，再经过组合渲染，重构 3D 模型。该方法在多个不同类别的物体重建中证明了其潜在的应用价值，不仅可根据人脸图像进行重建，还可对绘画作品、动物等进行重建。

最佳学生论文^[2]提出一种解决多边形网格生成问题的新思路。该工作受计算机图形学经典空间数据结构 Binary Space Partitioning (BSP) 的启发，对空间进行递归细分以获得凸集，从而使网络在无监督条件下可以通过凸分解学习 3D 形状高质量表示。BSP-Net 能自动生成最少的多边形，合成外形尽量完美、真实的三维物体。相比其它神经网络模型，BSP-Net 所用的多边形数量大幅减少，镶嵌效果更好。且通过 BSP-Net 训练的凸面可以轻松提取以形成多边形网格，无需进行昂贵的等值曲面处理。

另外，来自国内学术界和工业界的多篇论文也引起了广泛的讨论。香港中文大学推出面向细粒度动作分析的高质量数据集 FineGym^[3]，通过对现有动作识别方法进行分析与对比，为未来细粒度动作相关研究任务提供了方向。腾讯提出神经网络滤波器嫁接技术^[4]，通过引入外部信息激活无效的滤波器，从而大幅度提高模型性能。香港科技大学与腾讯优图实验室提出基于注意力机制及多关系检测器的小样本物体检测^[5]，基于注意力机制构建的 RPN、多关系检测器及三元组对比训练方法，使得网络能够不经重新训练就能用于新类别检测，并提出了 1000 类的小样本物体检测数据集。华为诺亚方舟实验室提出 AdderNet^[6]，在深度学习领域开创性地提

出了加法神经网络，将卷积计算中的乘法操作替换成了轻量级的加法计算，在国内外引起了热烈的讨论。商汤 EIG 算法中台团队针对物体检测中的尺度问题，设计了更具有等变性的特征金字塔以及物体检测器结构^[7]，使得单阶段检测器的性能在 COCO 数据集上大幅度提升。香港大学和中山大学合作提出了 Ref-Reasoning 数据集^[8]，包含真实图像和具有不同推理布局的语义丰富的表达式，是一个可用于结构化的指称表达式推理的大规模真实数据集。该论文同时提出了一种场景图引导的模块网络，提升了基于语言描述的视觉目标定位这一问题求解过程的可解释性。

此外，CVPR 2020 也举办了很多高质量的研讨会 (Workshops) 与教程 (Tutorials)。涉及的主题包括可解释机器学习、视频建模教程、自动化深度学习 (覆盖超参搜索及网络结构搜索)、图网络结构表示学习、视觉-语言研究进展、自动驾驶等。这些研讨会与教程为大家提供了很好的学习交流的平台，起到了全面的指导作用。

参考文献

- [1] Wu, Shangzhe, Christian Rupprecht, and Andrea Vedaldi. "Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild." IEEE/CVF CVPR 2020.
- [2] Chen, Zhiqin, Andrea Tagliasacchi, and Hao Zhang. "BSP-Net: Generating Compact Meshes via Binary Space Partitioning." IEEE/CVF CVPR 2020.
- [3] Shao, Dian, et al. "FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding." IEEE/CVF CVPR 2020.
- [4] Meng, Fanxu, et al. "Filter Grafting for Deep Neural Networks." IEEE/CVF CVPR 2020.
- [5] Fan, Qi, et al. "Few-shot Object Detection with Attention-RPN and Multi-relation Detector." IEEE/CVF CVPR 2020.
- [6] Chen, Hanting, et al. "AdderNet: Do We Really need Multiplications in Deep Learning?" IEEE/CVF CVPR 2020.
- [7] Wang, Xinjiang, et al. "Scale-Equalizing Pyramid Convolution for Object Detection." IEEE/CVF CVPR 2020.
- [8] Yang, Sibe, et al. "Graph-Structured Referring Expression Reasoning in the Wild." IEEE/CVF CVPR 2020.



李冠彬

中山大学数据科学与计算机学院副教授。主要研究方向为计算机视觉和机器学习。
Email: liguanbin@mail.sysu.edu.cn

五、总结与展望

回顾近三年的 CVPR 最佳论文，我们可以发现，CVPR 越来越青睐致力于解决真实场景下存在的视觉问题的方法或工作，包括对真实场景的建模、实现各类视觉子任务 (分类、检测、语义分割) 的统一融合，从而模拟人类视觉系统对真实物理世界的认知。从其关键词来看，实现该目的的概念方法 (meta learning、life-long learning、robust learning) 也是当前及若干年内可以想见的热门领域。

从 Staya 的炉边谈话以及最近举办的 AAAI/ICLR/ICASSP 等会议可以看出，线上虚拟会议一方面能够确保参会者的交流与体验，另一方面能够较大程度上节约参会者的开销与时间。此外，这种方式无疑在新冠肺炎疫情肆虐的这段时间最大程度保障了学术活动的正常进行。在计算机视觉技术日益发展的今日，或许这种方式可以成为未来科研交流的主要形式。

责任编辑 王金甲

哈尔滨工业大学左旺孟教授访谈

2020年6月22日,《CCF-CV专委简报》在线采访了哈尔滨工业大学计算机学院博士生导师左旺孟教授。下面是采访实录。

左老师,您好!首先,请您介绍一下您的个人学习和研究经历。

我的经历非常简单,从1995年大学入学开始,学习和工作都一直在哈尔滨工业大学。唯一可以说的就是本科和硕士学的是材料学,从博士入学起就一直在计算机学院学习和工作。

您近年来做了一些生成式对抗网络和迁移学习方面的工作,能否介绍一下您在这几个方面的研究情况和相关研究成果?

很大程度上我们对视觉应用研究得更多一些。对于生成式对抗网络(GAN),我们其实是在研究图像去噪和图像超分辨过程中,发现使用标准的MSE损失往往会使得重建图像有一定的over-smoothing现象。因而从2016年开始考虑能否与生成式对抗网络相结合,利用对抗损失改善重建图像质量,同时借鉴底层视觉之前的损失函数来避免GAN训练的不稳定性。对于迁移学习,我们其实也是考虑到深度学习的成功很大程度上依赖于大量的标注数据,觉得有必要关注和开展些迁移学习和自监督学习方面的工作。

实际上,我们在生成式对抗网络方面主要是依托人脸属性编辑、引导人脸增强和图像修复等具体任务在做,

我们在迁移学习和自监督学习方面的工作也主要是依托领域自适应、图像压缩和图像去噪等具体任务在做。

您在底层视觉和图像生成等领域的应用方面也颇有建树,请您谈一下这些应用方面我们国家现在的研究进展如何?相比国际,我们的研究地位如何?

无论是从学术界还是产业界的角度来看,底层视觉和图像生成都属于我们国家与国外差距相对较小的领域。尤其是底层视觉方面,港中文的SRCNN和我们的DnCNN对近年来图像超分辨和图像去噪的发展也都起到了一定的推动作用。图像生成方面国内的论文数量和质量也并不比国外差,但还是缺少一些有影响力的模型和框架如Pix2pix、CycleGAN和StyleGAN等,希望大家今后能在这些方面继续努力。

您在顶级期刊和顶级会议上发表了论文100余篇,在论文发表方面取得了突出成绩,能跟大家分享一下您是如何做到持续产出高水平论文的呢?

不敢说有什么经验。我会倾向于给学生较大的自由度,在自己还能给些指导的前提下尽可能让他们做自己感兴趣的方向,但会鼓励他们一定要去关注应用或方法层面上有价值的问题和给出有新意的解决方案。另外,在做的过程自己也会尽可能参与进去,一起讨论、调整和完善方案,起到互相促进的效果。不过我自己也不确定这种方法是否有普遍性或者是不是一种正确的方式。

您的论文具有很高的引用次数，能跟大家分享一下您在学术影响力方面做了哪些努力呢？

论文的引用次数其实有很大的偶然性。很多时候同一个想法可能会同时出现在多篇论文中，但往往只有其中的一两篇会得到大部分的关注。我自己更倾向于尽可能保证论文的贡献和完成质量，这种做法一方面有助于提升论文被大家关注的可能性，另一方面万一真得到较多的关注时也不至于太尴尬。如果非要分享些经验的话，可能就是尽可能做有用的研究。

您担任过多个顶级学术会议的领域主席，能介绍一下您参与组织这些会议的经验吗？另外，在与国内外同行交流时，怎样才能获得更好的效果呢？

我目前只担任过 ICCV 和 CVPR 的领域主席，体会比较深的就是 AC Meeting 时领域主席会认真讨论每篇论文的原创性和启发性，会更愿意推荐有新意的论文。这在一定程度上也促进和保证了计算机视觉领域的繁荣和多样性，许多学者也愿意做一些有趣的和基础性的研究尝试。在跟国内外同行交流时，会鼓励学生能够提前读相关的论文，通过面对面的交流澄清自己对论文内容的疑惑，就具体开展工作时的经验体会和对领域的认识等也会进行一些沟通。

可否请您谈一下在后深度学习时代，计算机视觉技术将如何发展？面临哪些挑战？哪些研究方向会特别有价值呢？

我自己也很认同深度学习的浪潮肯定会逐渐过去，但深度学习时代发展起来的一些方法和思维模式，如通过设计恰当的网络和损失函数来训练一个具备某种功能的模型，在后深度学习时代应该会被保留下来被大家继续使用。至于后深度学习时代的研究，我可能会建议大家能更关注于解决计算机视觉自身的问题（如新型视觉传感器、内容生成、视觉认知、视觉与语言/机器人的结合），同时也能以比较开放的心态拥抱和利用学习领域的最新进展。

请问您是怎么管理研究生的？您对他们的要求是什么？

刚开始时会希望他们能围绕一个我比较熟悉的任务开展研究，便于有问题及时发现和调整，也锻炼下他们的代码、实验和写作能力。随后，会逐渐加强研究生的自主性，希望他们能够根据自己的兴趣定义一些有趣的问题，做出些有新意和有价值的工作。对研究生的论文发表并没有确定性的规定，主要是满足学校和学院毕业要求，但会希望他们能锻炼出独立和合作做研究的能力，以及保证学位论文的高质量完成。

如果吐露研究工作者的心声，您最想说的是什么呢？

希望大家都能做出更有价值和影响力的成果，一起努力推动中国乃至世界计算机视觉领域的发展。

责任编辑 赵振兵 余烨



左旺孟

哈尔滨工业大学计算机学院教授、博士生导师。主要关注生成式对抗网络、迁移学习模型及其在底层视觉、图像生成、视觉跟踪、图像分类等领域的应用。在 CVPR/ICCV/ECCV 等顶级会议和 T-PAMI/IJCV、IEEE Trans. 等顶级期刊上发表论文 100 余篇，谷歌学术引用 15,000 余次，谷歌学术 H-Index 为 62。发表 ESI 高被引论文 8 篇。提出的 DnCNN 模型被正式收入 MATLAB 2017b 及后续版本的 Image Processing 和 Deep Learning Toolbox，并被 CV-News Magazine 作为 Main Story 专题报道，目前最高单篇引用 1900 余次。担任人工智能学会模式识别专委会常委、中国图象图形学会机器视觉专委会常委、中国图象图形学会青工委执委，国家自然科学基金重点项目负责人。担任 ICCV 2019、CVPR 2020/2021 等 CCF-A 类领域主席及 IJCAI、AAAI 等 CCF-A 类会议高级程序委员，担任 ECCV AIM 2020 竞赛单元主席之一。

东南大学杨万扣副研究员访谈

2020年8月上旬,《CCF-CV专委简报》在线采访了东南大学自动化学院博士生导师杨万扣副研究员。下面是采访实录。

杨老师,您好!首先,请您介绍一下您的个人学习和研究经历。

我在南京理工大学计算机系及后来的计算机学院分别获得学士、硕士和博士学位。硕士毕业后,我在北京铁道部下属一家国企工作一年,作为主力参与了公司承担的铁道部“TFDS”项目,即火车故障的自动检测工作。在公司工作有两点体会:(1)自己的知识储备和能力还有欠缺;(2)公司上班要求打卡,不自由。于是我选择回南京理工大学继续读博,打算毕业后进高校工作。博士期间有幸到香港理工大学做研究助理,开阔了视野,并认识了一群好朋友。博士毕业到东南大学自动化学院做博士后,中途去了美国北卡大学威明顿分校做博士后。博士后出站留在东南大学自动化学院工作到现在。2015年申请到CSC资助前往悉尼科技大学学习一年。博士期间主要做子空间学习,后面逐步转向动作识别、目标检测、跟踪和图像分割,并陆续开发了一些应用系统。这么多年的工作和发展,离不开工作单位、母校、母院及师长和朋友们的关心和大力支持。

您在目标检测与跟踪、行人理解、人脸识别和环境感知等方面有很多的研究成果,能否跟大家分享一下您的这些成果?

总的说来,我们的工作还比较浅显。做论文之余,我组织学生搭建了一些软硬件结合的实际系统,如基于机器视觉的显示屏外观检测系统、布匹瑕疵检测系统和红外人脸测温系统等,其中一些系统有推广。关于目标检测,我们主要做了一些轻量级的快速目标检测方法,如我们设计了一个模型大小为300K的人脸检测算法,在工业界得到了应用;我们对YOLO算法进行了修改和模型压缩,成功应用到实际系统中。关于目标跟踪,我们基于深度框架下卷积滤波器的优化和利用时空记忆网络做了一些单目标跟踪算法,在多个标准数据集上取得了当前比较好的性能,和其他优势方法相比,我们的方法需要训练的数据量非常少。关于图像语义分割,我们利用NAS构建一些轻量快速的网络模型。关于人体姿态估计,我们利用人体先验结构知识辅助搜索针对各个部件的特定神经网络结构。基于这些研究工作,我们陆续完成了一些应用项目,并逐步搭建了无人车和无人机平台。

今年7月份的一则新闻“2020年春节期间,东南大学老师杨万扣及他的团队通过昼夜奋战突击开发了“非接触式智能精准测温筛查系统”,成功部署在江苏省委统战部、江苏省公安厅办公大厅、东南大学、南京29中、凤凰集团、徐庄软件园、汤泉街道、永辉超市等场所,直接服务于抗“疫”一线,得到了中共江苏省委统战部的表扬。”家喻户晓,请问杨老师能分享一下该系统的开发过程,及开发过程的一些花絮么?

春节期间，新型冠状病毒肺炎疫情爆发，全国同舟共济、众志成城，抗击疫情。我们正好在热红外图像和白光图像方面做过一些工程项目，有一定积累，感觉可以做测温筛查系统为抗击疫情做点事。于是组织研究生，通过昼夜奋战突击开发了“非接触式智能精准测温筛查系统”。自动化专业的学生对软件、硬件和算法都有涉猎，开发软硬件结合的系统相对有优势。我们系统从白光图像中定位人脸，然后映射到红外图像中人脸区域，利用红外图像测量温度，通过大数据分析自动锁定温度异常者并自动报警。由于我们的系统没有采用黑体标定，一开始大家还挺担心。设备使用中发现黑体只是有助于测量绝对温度，我们系统没有黑体是可行的。我们算是市场上第一批做测温设备的队伍，和海康、大华等企业进行同台竞技，经受住了市场的考验。在系统开发过程中，一家帮我们制作底板的工厂，被邻居举报非法开工，后来辗转好几家加工厂，才制作好底板。

您参与完成的项目“鲁棒人脸视觉特征的提取、建模与识别的理论和方法研究”获中国教育部自然科学二等奖，能否跟大家分享一下这个项目相关的研究工作、研究成果及研究花絮？

本项目由东南大学郑文明教授负责完成，我是第三完成人。我和郑文明教授有多年的合作关系。本项目围绕人脸图像中的表情、身份以及亲属关系识别等科学问题，深入开展鲁棒人脸视觉特征的提取、建模与识别的理论和方法研究，重点解决视觉场景中光照与人脸姿态变化下的鲁棒人脸视觉特征提取与建模问题。在项目中我的贡献是提出人脸视觉特征抽取与识别的多流形鉴别分析理论。

能否介绍一下您在美国北卡罗纳大学威尔明顿分校和澳大利亚悉尼科技大学的访学经历和期间的研究工作呢？

博士期间有幸去了香港理工大学做研究助理，发现很多同事都在申请出国，于是我投简历申请出国。由于2008和2009年国际金融危机等原因，当时我没有收

到海外博士后职位 offer。在东南大学博士后期间，才陆续收到几个 offer。考虑有海外经历好找工作，同时东南大学博士后合作导师也支持我出国，于是就去了美国北卡大学威明顿分校 Face Aging Group 小组。当时在美国主要是帮着开发一些系统，如人脸年龄估计、性别分类和人脸年龄模拟等。后来顺利留在了东南大学，就从美国回国了。非常感谢美国 Karl Ricanek 教授对我的资助。

2015年我申请了CSC资助前往悉尼科技大学学习一年。在悉尼科技大学学习期间主要做了关于稀疏表示的工作，同时认识了很多青年才俊。

作为一名优秀的青年学者，请问您对未来研究趋势是如何把握，以及如何规划未来的研究工作的？

作为一名普通老师，我的科研水平比很多同行落后不少。目前主要是一边做些理论，一边搭建实际系统。关于未来的研究工作，主要通过以下两个方式获取：(1) 参加学术会议，和学术大师交流，了解学术前沿；(2) 跟工业界和研究所多交流，了解工业界需求，调整研究方向和资源分配。

您在教学上也获得了很多成就，如：“自主系统的智能自适应控制方法”获得2018年江苏省教育厅教育教学与科研成果奖（研究类）一等奖（高校自然科学研究类）等，能否分享一下您的这些教学成果，以及您如何在教学和科研两方面进行兼顾的？

作为一名普通老师，我承担本科生的《C++程序设计》和研究生的《模式识别原理》两门课，每年教学量大概150学时。借鉴MIT的《程序设计》课程教学方法，东南大学自动化学院在《C++程序设计》教学过程中，设计平时个人项目、团队项目和上机考试三种考核方式，对学生培养实施过程化考核，效果显著。我们会经常安排一些图像处理等专业方面的应用性小项目给学生完成。以我们自动化学院程序设计教学组为主力，拿到了东南大学教学成果一等奖和江苏省教学成果一

等奖。上《模式识别原理》课程时，以公式推导为主，考核方式是个人项目和团队项目。以我所在的导航与控制研究所部分研究成果，拿到了江苏省教育厅教学成果一等奖（研究类），该奖项中我的贡献较小。通过给本科生上课，能吸引学生主动联系读研。

能否介绍一下您现在的研究团队，以及您团队的管理情况和研究方向规划？

作为一名普通老师，我这谈不上团队，自己和几个学生一起捣鼓吧。大家定期讨论交流。不定期邀请校外学者来交流指导。在平时，我们培养学生的团队责任感、使命感和荣誉感。希望学生毕业后都能有个好出路和好未来。会和学生一起讨论我们往后的方向，初定面向自主无人机系统研究环境感知和路径规划相关技术。

“家庭”与“工作”同样重要，请问您在这两个方面分别投入了多少时间？占比关系如何？您是如何处理这两方面的复杂关系的？

作为一名普通老师，家庭压力和工作压力都很大。家里主要是老人帮忙照顾小孩，大部分时间是在工作上吧。主要是家里人比较理解和支持！

如果吐露研究者心声，您最想说的的是什么？

作为一名普通老师，希望生活压力小一点，做点自由的研究，能落地更好。往后，请各位老师和朋友多多关心和支持！

责任编辑 余焯 赵振兵



杨万扣

东南大学副研究员，博士，博士生导师，模式识别与智能系统专业，研究方向为模式识别、计算机视觉和自主无人系统。从攻读硕士学位期间开始进行图像处理、模式识别领域的研究工作，已有 10 多年的研究积累。最近正在从事目标检测、行人检测和动作识别、图像分割等方面的理论研究及面向自主无人系统的应用。在国际期刊如《IEEE TPAMI》、《IEEE TNNLS》、《IEEE TIP》、《Pattern Recognition》及国际会议如 CVPR、AAAI、IJCAI、ICPR、PRCV 上发表论文多篇，其中 20 余篇被 SCI 检索。Google Scholar 上检索的次数为 1000 余次，ESI 高被引论文两篇。博士论文《人脸识别中的部分特征抽取技术研究》获 2010 年度江苏省优秀博士论文。先后主持多项国家自然科学基金青年基金和面上项目。曾获得教育部自然科学二等奖和江苏省教学成果一等奖各一项。担任国际 SCI 源期刊《Neural Processing Letters》编委。曾担任 IScIDE2012 组委会主席、IScIDE2013 本地组织主席和 PRCV2020 宣传主席等。现为中国计算机学会计算机视觉专委会委员、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秘书长和江苏省人工智能学会智能驾驶技术专委会副主任。

委员好消息

2020年6月18日,教育部公示了关于2020年享受政府特殊津贴推荐人选,CCF-CV专委会主任、北京大学查红彬教授和CCF-CV专委会常务委员、中山大学赖剑煌教授入选。

2020年6月14-19日CVPR 2020在线举办,会议期间举办了系列竞赛。CCF-CV专委会委员、厦门大学曲延云教授团队在NTIRE 2020 Challenge on Image Demoiring: Methods and Results赛道中夺得冠军。团队结合数据集的特点,采用全局上下文模块来学习图像的自相似性依赖;针对网络计算复杂度大,采用DWT和IDWT对特征图进行变换;针对训练集和测试集存在的差异,通过对齐源域和目标域特征的二阶统计量来进行网络微调。CCF-CV专委会委员、中国人民大学金琴副教授团队在ActivityNet Challenge-Task 2 Dense-Captioning Events in Videos赛道中夺得冠军。团队充分考虑了视频中可描述事件的时序关系,提出了一种新的简单且有效的视频时序动作检测方法来提取视频中的事件,随后充分利用事件前后的上下文信息对所检测出来的事件序列生成密集描述。

2020年6月19日CVPR 2020落下帷幕,CCF-CV专委会委员、中科院自动化所雷震研究员指导的论文Bridging the Gap Between Anchor-based and Anchor-free Detection via Adaptive Training Sample Selection和CCF-CV专委会常务委员、大连理工大学卢湖川教授指导的论文High-Performance Long-Term Tracking with Meta-Updater获得论文奖提名。

2020年7月14日,ICML 2020杰出论文出炉,CCF-CV专委会委员、北京理工大学付莹教授和黄华教授指导的论文Tuning-free Plug-and-Play Proximal Algorithm for Inverse Imaging Problems获杰出论

文奖。该论文由北京理工大学和剑桥大学合作完成,提出了一个无需调参的即插即用近似算法,可自动确定内部参数。该方法的核心是建立一个用于自动搜索参数的策略网络,可以通过混合无模型和基于模型的深度强化学习进行有效学习。ICML 2020共评出杰出论文奖和杰出论文荣誉提名奖各2篇,时间检验奖1篇。

2020年7月15日获悉,CCF-CV专委会委员、华中科技大学白翔教授、中国科学院自动化研究所赫然研究员、清华大学鲁继文副教授当选IAPR Fellow。

2020年7月17日,浙江省召开科学技术奖励大会颁布2019年度浙江省科学技术奖获奖,CCF-CV专委会委员、浙江工业大学陈胜勇教授的“视觉系统中的目标物运动状态分析”项目荣获自然科学一等奖。该项目在解决陌生复杂环境中对象分割与场景描述问题、多目标跟踪与运动状态分析问题等方面均有突破。

2020年7月22日,北京市科学技术委员会公布了2020年北京市科技新星计划拟入选人员名单,CCF-CV专委会委员、中科院自动化所副研究员黄岩入选。

2020年8月11日,国家自然科学基金委员会公布了2020年度国家杰出青年科学基金建议资助项目申请人名单,CCF-CV专委会委员、中科院信息工程研究所操晓春研究员的智能视觉内容安全项目、厦门大学纪荣嵘教授的紧致化计算机视觉分析项目和北京理工大学杨健教授的手术导航理论、关键技术及应用项目获得资助。

2020年9月10日,CCF-CV专委会委员、北京科技大学殷绪成教授主持完成的网络图像视频大数据的智能识别关键技术及应用项目荣获2019年度北京市科技技术进步奖一等奖。

责任编辑 刘海波

艺术字风格化开源代码

北京大学 杨 帅 刘家瑛

文字字效是文字上诸如颜色、描边、阴影、反射和纹理等的艺术效果，具有字效的艺术字看上去更生动有吸引力。艺术字风格化旨在根据参考样例字效或者风格图像为文字渲染字效生成艺术字，可被广泛地应用于广告、杂志、海报等平面设计中。然而艺术字的人工制作过程繁琐，需要一定的技术。因此全自动艺术字风格化技术具有巨大的研究价值和商用价值。

艺术字风格化是图像风格化领域中一个新兴的研究方向，相较于一般的图像风格化问题，文字字效的风格元素具有良好的定义（例如描边、阴影效果），风格化目标明确，因此适合研究、分析与评价。自 2017 年提出的字效迁移模型 T-Effect，已经发展出若干有效的艺术字风格化模型，并建立了相应的数据集。本文将重点介绍几个开源的艺术字风格化模型以及数据集。

1、Awesome Typography: Statistics-Based Text Effects Transfer

工作：该论文提出了字效迁移任务：如图 1，根据文字 S 及其艺术字 S' 学习字效并迁移至文字 T ，得到艺术字 T' 。对应地提出了一个非参数化的字效迁移模型 T-Effect。该论文根据对文字字效图像的分析，归纳总结了字效的空间分布先验，用以约束提出的基于局部图像块的风格化目标优化方程。目标方程从字效的局部纹理细节、全局纹理分布和纹理自然性三方面进行设计，能更好地保持局部纹理细节和全局纹理分布与参考风格的一致性。



图 1 字效迁移示意图及 T-Effect 的艺术字生成

论文：<https://arxiv.org/abs/1611.09026>

代码：<https://github.com/williamyang1991/Text-Effects-Transfer/>

2、Multi-Content GAN for Few-Shot Font Style Transfer

工作：该论文提出了一个参数化的艺术字风格化模型 MC-GAN，根据已有的几个带有风格的字母，生成未知的带有此种风格的其他字母。如图 2 所示，模型基于生成对抗网络，结合了字体迁移子网络与字效迁移子网络，能同时迁移字体和字效。

论文：<https://arxiv.org/abs/1712.00516>

代码：<https://github.com/azadis/MC-GAN>

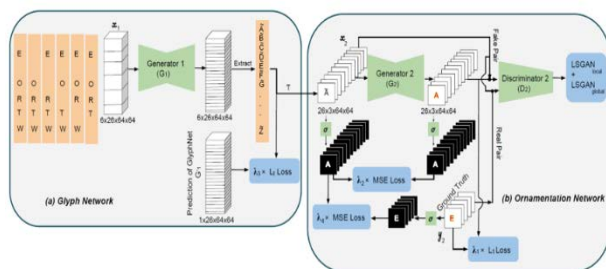


图 2 MC-GAN 的网络结构及艺术字生成结果

3、Typography with Decor: Intelligent Text Style Transfer

工作: 该论文重点关注了艺术字中常出现的装饰品, 这些装饰品相较于规则的字效更加复杂和多样。如图 3 所示, 论文首先训练网络检测参考风格中的装饰品, 从而将问题转换为了标准的字效迁移和装饰品/艺术字组合问题。在组合过程中, 文章同时考虑了装饰品的空间分布与重要程度, 从而能根据文字将装饰品自适应地摆放到合理位置。

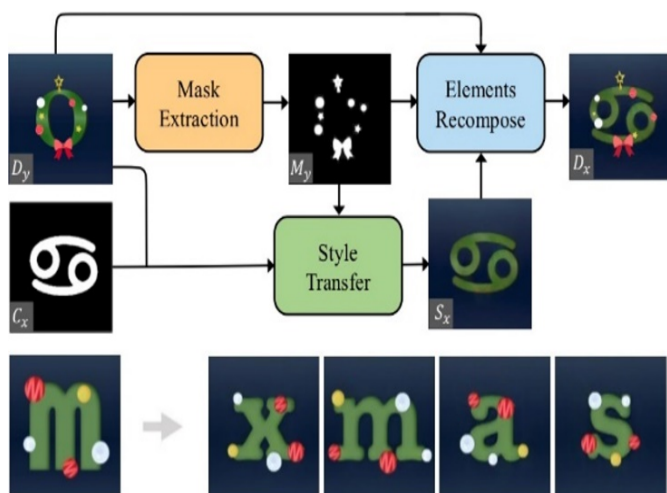


图 3 艺术字装饰物处理框架及艺术字生成结果

论文: <https://github.com/daoshee/Typography2019/blob/master/3159.pdf>

代码: <https://github.com/daoshee/Typography-with-Decor>

4、Controllable Artistic Text Style Transfer via Shape-Matching GAN

工作: 该论文研究另一类艺术字风格化任务: 不使用现有的艺术字作为参考, 而是直接利用任意风格图像渲染艺术字。该任务的关键在于调整文字的轮廓以模拟风格图像中的目标主体的形态, 如图 4 所示, 要使火焰在文字上烧起来, 文字边缘势必要进行形状的调整。该论文提出了双向结构匹配的框架来学习文字的形变。在反向结构迁移阶段, 提取风格图像的结构, 利用高斯尺度空间表示, 反向将文字的形状风格迁移到结构图上, 获得简化结构图。在正向风格迁移阶段, 正向学习该上述过程的逆过程, 即将简化结构映射到原始结构再进一步映射回风格图像, 从而学会结构迁移和纹理迁移。

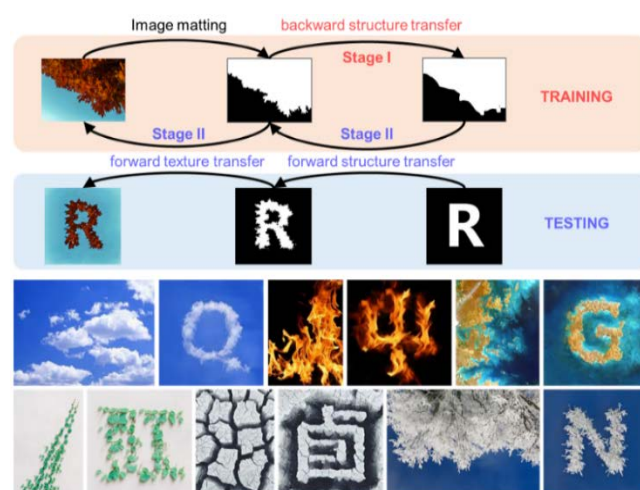


图 4 双向结构匹配框架及艺术字生成结果

论文: <https://arxiv.org/abs/1905.01354>

代码: <https://github.com/TAMU-VITA/ShapeMatchingGAN>

5、TE141K: Artistic Text Benchmark for Text Effects Transfer

工作: 该论文提出基于字形字效编解码的字效迁移生成对抗网络 TET-GAN 对文字进行风格化, 如图 5 所示, 通过显式的字形和字效特征提取, 单个网络模型能进行多达 152 种字效风格的迁移。同时为了促进字效迁移相

关研究的发展，亟待建立一个字效迁移的基准用以评价不同图像风格化模型在字效迁移任务上的性能。该论文提出了新的大规模字效数据集 TE141K。图 6 展示了部分字效风格，表 1 给出了数据集的统计信息，包含 141081 张成对的文字/字效图像，涵盖汉字、英文字母、数字、日语假名等不同的字形。在 TE141K 的基础上，该论文为 14 种不同的图像风格化模型在多风格字效迁移、单样本监督字效迁移、单样本无监督字效迁移三个任务上建立了字效迁移的基准，为后续研究评价和分析字效迁移模型的性能提供有力的支撑。除了字效迁移任务，该数据集还可用于多领域迁移、图像转换等多种模型的研究。

论文: <https://arxiv.org/abs/1905.03646>

代码: <https://daooshee.github.io/TE141K/>



图 6 TE141K 数据集的艺术字示意图

表 1. 典型开源的关系检测数据集

数据集	字效	字形	图像
TE141K-E	67	988 (英文字母)	59280
TE141K-C	65	837 (中英文+数字)	54405
TE141K-S	20	1024 (小语种+符号)	20480
合计	152	2849	141081

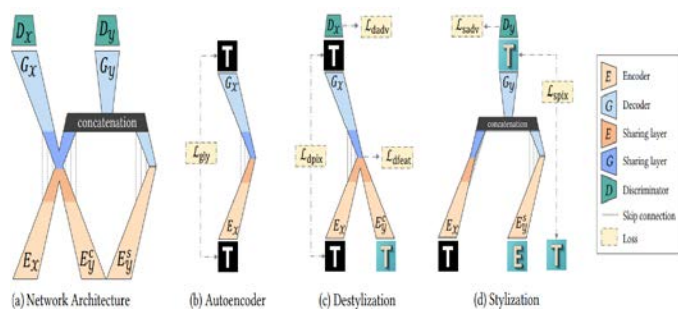
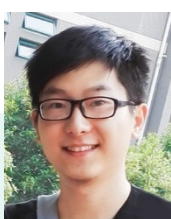


图 5 TET-GAN 的网络结构及训练任务

责任编辑 沈沛意



杨 帅

北京大学王选计算机研究所博士研究生，研究方向为图像风格化和图像修复。
个人主页: <https://williamyang1991.github.io/>



刘家瑛

北京大学王选计算机研究所副教授。主要研究方向为多媒体信号处理，压缩和计算机视觉。
研究组主页: www.wict.pku.edu.cn/struct/

可解释性数据集

大连理工大学 付陈平 樊鑫

神经网络可解释性问题近年深受国内外研究者关注，其主要目的在于对深度神经网络进行解释方法的开发。当前在金融投资、军事、医疗、自动驾驶等高风险决策领域，缺乏透明度成为深度神经网络实践应用的障碍，研究其可解释性能够帮助人们从网络模型中抽取其学习到的知识，从而学习到新见解。对深度神经网络模型进行研究时，相较于黑箱模型，人们对可解释的模型更容易进行弱点分析。可解释性降低了模型偏差检测的难度，更好地帮助人们验证和改善模型。深度神经网络可解释性的研究促进深度网络理论和应用的发展，同时有助于人机互信、人机协同，推动神经网络在多领域的实际应用。

近年来有研究指出，对于相同任务，具有相同的分类性能的不同深度神经网络在使用哪些特征作为决策基础方面存在很大差异，这表明确定最合适的网络需要可解释性。深度神经网络解释方法的定性评价需要由人来进行主观评价最终的解释结果，其主要方法为可视化解释方法，从输入变量的角度对预测结果进行解释，因此使用含有丰富细节特征的图像数据集对研究深度神经网络可解释性是十分重要的。分析网络内部激励层对图像中的特征产生的响应是研究网络可解释性的主要方法，从内部学习到的知识表征、输入输出等多个角度来评估模型的鲁棒性和泛化能力。本文介绍了深度神经网络可解释性应用于不同领域常用的数据集，包括 ADE20K、CUB200-2011、ApolloScape、VQA 和 VQA v2.0 数据集。

1、VQA 和 VQA v2.0

介绍： VQA (Visual Question Answering)数据集和 VQA v2.0 是两个视觉问题回答数据集，包含关于图像的开放式问题。这些问题需要对视觉、语言和常识的理解来回答。这两个数据集均由 Devi Parikh 和她的团队进行开发。

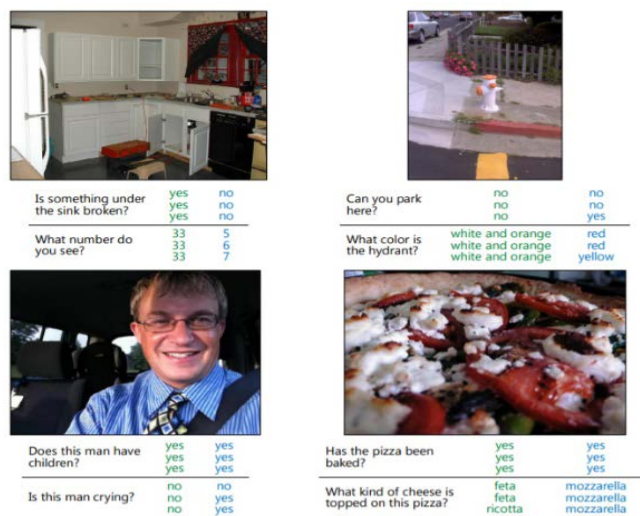


图 1 VQA 数据集

VQA 数据集包含 265,016 幅图像(具体和抽象场景)，每幅图像至少 3 个问题(平均 5.4 个问题)，每一个问题对应 10 个答案，其任务包含开放式任务和选择题任务。VQA v2.0 数据集包含 204,721 张具体图像，1105904 个问题，每一个问题对应 10 个答案，其数据集中的每个问题都有两个互补的图像，它们看起来很相似，但对问题的答案不同，其构造上比 VQA 数据集更

加平衡。在神经网络可解释性研究中，其用于多模态解释方法。VQA 是一个被广泛研究的多模态任务，它需要视觉和文本理解以及常识知识。VQA v2.0 对两个语义相似但答案不同的图像提出相同的问题。由于这两个图像在语义上相似，所以 VQA 模型必须采用细粒度的推理来正确地回答问题，这个过程适于研究网络解释。再比较互补对的解释，我们可以更容易地确定我们的解释是否集中在做出决定的重要因素上。图 1 和图 2 分别展示了 VQA 数据集和 VQA v2.0 数据集的部分例子。

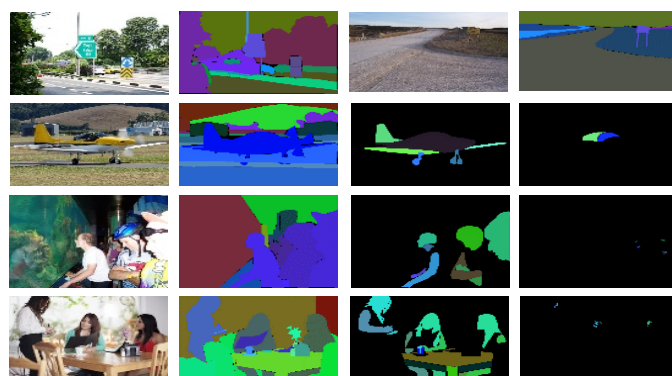


图 3 ADE20K 数据集

数据集地址: <http://groups.csail.mit.edu/vision/datasets/ADE20K/>



图 2 VQA v2.0 数据集

数据集地址: <https://visualqa.org/>

2、ADE20K 数据集

介绍: ADE20K 数据集是用于场景解析的数据集，该数据集由 MITCSAIL 研究组于 2017 年发布和维护，其包含 150 种物体类型，可用于场景感知、解析、分割、多物体识别和语义理解。

DE20K 数据集的训练集中有 20210 幅图像，验证集中有 2000 幅图像，测试集中有 3000 幅图像。所有的图像中的对象都被进行标注，许多对象也用它们的部件进行注释。对于图像中的对象或对象部件都有对应的掩模图像。ADE20K 中的任何图像都至少包含 5 个对象，单张图像包含的最大对象实例数达到 273 个。在神经网络可解释性研究中，通常以场景识别作为任务训练模型，从模型中提取激励层映射到输入图像中，结合数据集中对象掩模和对象部件掩模进行网络可解释研究。图 3 展示了 ADE20K 数据集集中的部分对象。

3、CUB200-2011 数据集



图 4 CUB200-2011 数据集

介绍: CUB200-2011 数据集是一个鸟类识别细粒度数据集，其常用于细粒度分类识别研究，其中图片由加州理工学院提供。

CUB200-2011 数据集包含 200 个鸟类子类，共有 5994 张训练图片和 5794 张测试图片，每张图像均提供了图像类标记信息，图像中鸟的边界框、关键部位信息和属性信息。每张图提供了鸟的 15 个关键部位(头、身体、冠、颈背、眼睛、嘴、喉咙、前额、背、胸、肚子、翼、腿、尾巴)的中心坐标。在神经网络可解释性研究中，由于数据集中对象部位类别已经确定，可用于训练可解释的神经网络。使用数据集中所提供的对象部位，限定神经网络中每个滤波器代表一个特定的对象部位，通过训练将滤波器推向对象部位的表示，使

网络高层编码更多语义上有意义的知识。图 4 展示了 CUB200-2011 数据集中的部分对象。

数据集地址 <http://www.vision.caltech.edu/visipedia/CUB-200-2011.html>

4、ApolloScape 数据集

介绍: ApolloScape 是一个大规模自动驾驶数据集, 包含了轨迹数据集, 三维感知激光雷达目标检测和跟踪数据集, 其中图像采集工作由百度完成, 并在 2018 年 3 月应需开放。

ApolloScape 数据集包括约 100K 图像帧, 80k 激光雷达点云和 1000km 城市交通轨迹。其中的场景解析数据集包含 146997 图像帧与相应的像素级注释, 姿态信息和深度地图的静态背景。数据集内包括高分辨率图像序列和逐像素标注的 RGB 视频, 带有语义分割的调查级密集三维点。在神经网络可解释性研究中, 该数据集用于探讨自动驾驶系统中深度神经感知和控制网络的视觉解释。具体做法为对数据集提供的丰富语义标签, 例如行人、车辆、标示牌、信号灯等, 在深度神经网络

中探究潜在影响网络输出的图像区域, 这将产生更简洁的可视化解释, 并更准确地暴露网络的行为。图 5 展示了 ApolloScape 数据集中部分场景解析图像。

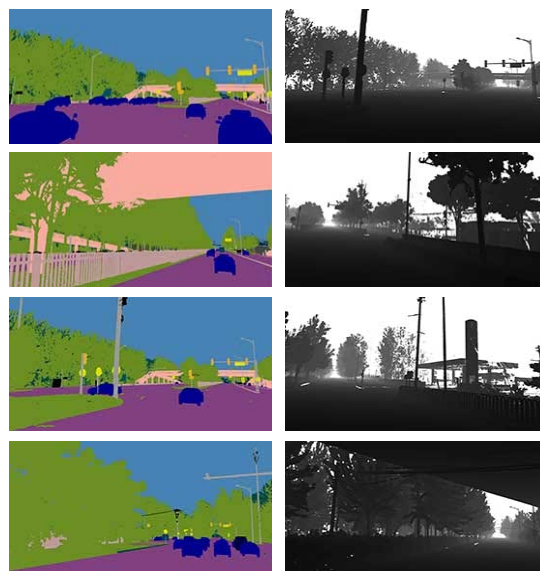


图 5 ApolloScape 数据集

数据集地址 <http://apolloscape.auto/>

责任编辑 李策



付陈平

博士研究生, 大连理工大学国际信息与软件学院, 研究方向为计算机视觉。



樊鑫

博士生导师, 大连理工大学国际信息与软件学院从事教学与科研工作, 担任中日国际信息与软件学院院长。研究方向为计算机视觉与图像处理、医学影像分析。

个人主页: http://faculty.dlut.edu.cn/Xin_Fan/zh_CN/index.htm

好文推荐

西北大学和西安电子科技大学团队的“基于深度学习技术的灵长类动物个体身份识别”最新成果发表在 Cell 子刊 iScience-2020。

论文: Songtao Guo, Pengfei Xu, Qiguang Miao, Guofan Shao, Colin A Chapman, Xiaojiang Chen, Gang He, Dingyi Fang, He Zhang, Yewen Sun, Zhihui Shi, Baoguo Li. Automatic Identification of Individual Primates with Deep Learning Techniques, iScience, vol. 23, no. 8, pp. 1-32, Aug. 2020

个体识别是动物行为学和动物生态学基础。在动物行为规律研究的过程中,最重要的一个环节就是对研究动物进行个体识别。而近半个世纪以来,动物身份识别主要借助动物本身特征(如斑纹、颜色、伤疤等)或人为标记特征(如烙印、染色、无线电项圈及遗传标记等)进行个体识别,导致数据采集的可靠性和效率较低,实践和精准度受到人力、时间和空间条件的限制,极大地制约了动物行为研究。

论文针对动物个体身份的自动识别问题,设计并完成了研究利用 AI 技术进行动物面部检测、追踪和识别的系统——Tri-AI(图 1)。该系统针对不同灵长类动物面部特征相似性较高的问题,提出端到端的具有关注机制的深度神经网络模型(图 2)。该模型具有三个特征提取通道,分别关注动物面部全局特征、面部皮肤区域特征以及局部细节特征,最后联合三通道的特征信息进行准确的身份识别。Tri-AI 系统颠覆了传统依靠动物个体特征或人为标记特征进行动物身份识别的方法,实现了对野生个体的准确身份识别和连续跟踪采样的功能(图 3)。Tri-AI 系统进一步扩展应用于多个类群的不同物种的个体身份识别,且同样适用于动物夜视图像的个体识别。目前该系统已经在灵长类的 41 个代表性物种和 4 种食肉动物群体(如图 3)进行了适用性验证, Tri-AI 系统对个体数大于 19 的 17 个灵长类物种的平均识别精度达 94.1%,且每秒识别 31 张图片,极大的提高了动物个体数据分析效率,未来可作为大数据分析的核心组件,为动物学研究提供新的技术方案,也为实现野生动物保护和智能管理提供可靠的技术支撑。

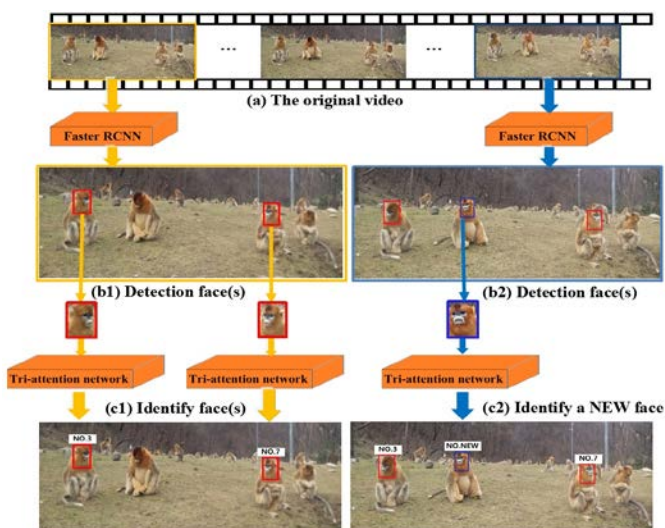


图 1 Tri-AI 动物个体识别系统框架图

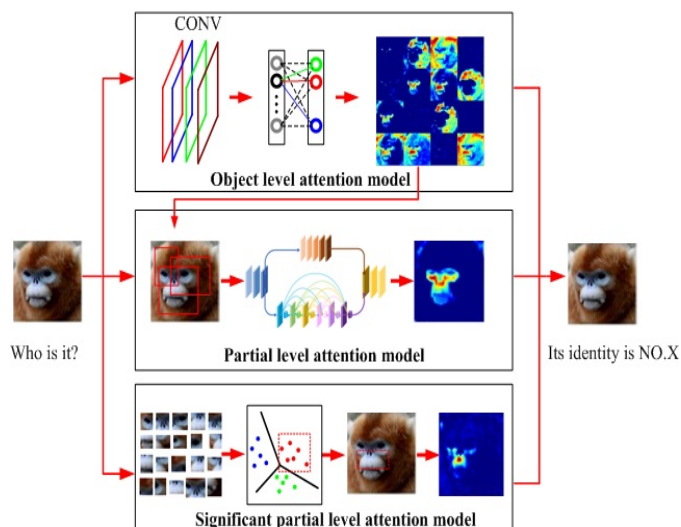


图 2 基于关注机制的深度神经网络模型

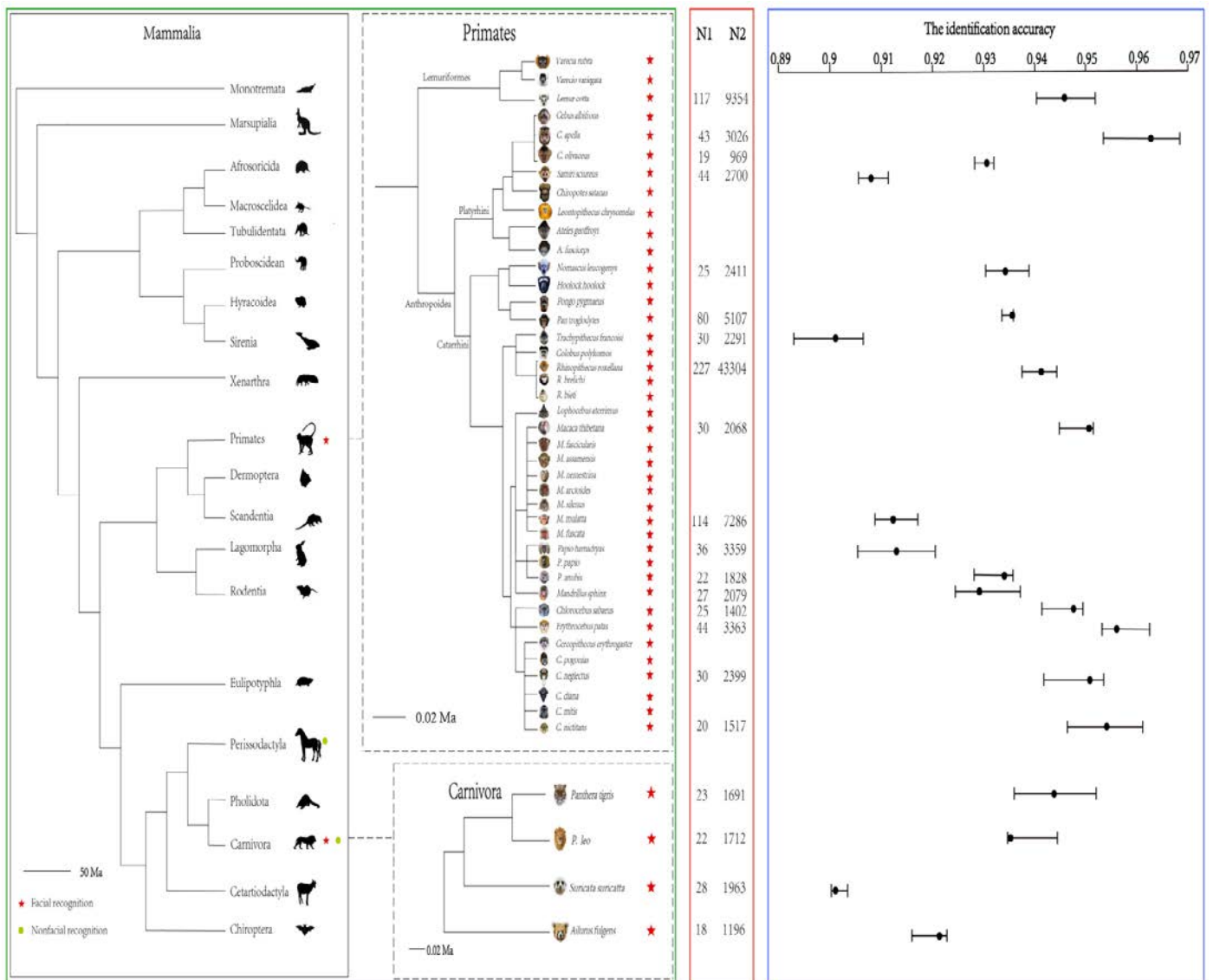


图 3 Tri-AI 对多物种进行个体识别

责任编辑 沈沛意 贾同

好文推荐

中科院和中国科学院大学团队的“基于对抗交叉光谱人脸补全的 NIR-VIS 人脸识别”最新成果发表在 IEEE TPAMI-2020。

论文: Ran He, Jie Cao, Lingxiao Song, Zhenan Sun, Tieniu Tan. Adversarial Cross-Spectral Face Completion for NIR-VIS Face Recognition, IEEE TPAMI, vol. 42, no. 5, pp. 1025-1037, May. 2020

近红外-可见(NIR-VIS)异质人脸识别是指将近红外与可见人脸图像进行匹配的过程。目前的异质人脸识别方法都是通过从近红外图像中合成普通的人脸图像,然后将传统的人脸识别方法扩展到近红外光谱中。然而,由于自遮挡和传感间隙的存在,近红外人脸图像相较于普通的人脸图像往往是不完整的。因此,本文提出了一种基于对抗交叉光谱的人脸补全算法,并将其应用到了 NIR-VIS 人脸识别任务当中。

如图 1 所示, NIR-VIS 的面部补全问题可以看作是在对抗学习框架中学习人脸图像从近红外域到可见域的映射。其中,生成器包括位姿校正网络 G_p , 纹理渲染网络 G_t 和一个变形融合网络 G_f 。输入一个近红外的人脸图像, G_p 用来估计一个稠密对应的紫外线(UV)场; G_t 的目的是生成一个姿态不变的面部纹理图; G_f 融合这两个输出, 得到最后的结果。(b)中的多尺度判别器通过区分可见域中真实图像和转换后的图像来保证补全图像的视觉质量。(c)中的细粒度判别器通过引导生成器以一种对抗的方式最小化输入的内差来监督网络的解耦过程。

大量实验表明, 该网络不仅能够生成高分辨率的人脸图像, 而且有利于提高异构人脸识别的准确率。

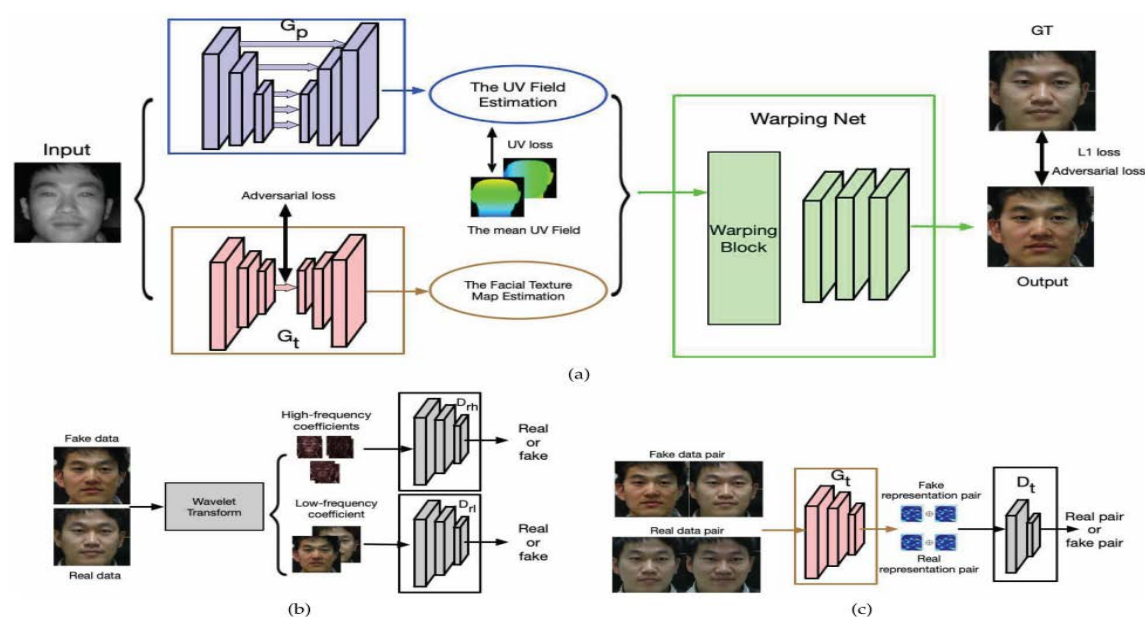


图 1 NIR-VIS 人脸补全网络示例。(a)为生成器的前向传播过程, (b)为多尺度判别器, (c)为细粒度判别器。

责任编辑 贾同 樊鑫

好文推荐

中山大学团队的“基于深度非对称度量嵌入的无监督行人重识别”最新成果发表在 IEEE TPAMI-2020。

论文: Hongxing Yu, Ancong Wu, Weishi Zheng. Unsupervised Person Re-identification by Deep Asymmetric Metric Embedding, IEEE TPAMI, vol. 42, no. 4, pp. 956-973, Apr. 2020.

在针对行人重新识别(Re-ID)算法研究中,现有的基于有监督的算法往往需要大量的标注信息,然而现实世界中仅存在大量的无标签图像信息,这在一定程度上限制了实际应用的可扩展性。一般的基于无监督行人重识别算法在解决这种扩展性问题的同时,存在着动态的相机视角变化引起的具体视角偏差,这种具体视角的偏差极大的影响了在无监督场景中跨视角行人判别信息的获取。因此,本文提出了一种基于跨视角聚类的无监督非对称距离度量学习方式来解决上述问题。非对称距离度量为每个相机视角进行具体的特征变化来应对具体的特征扰动。并在此基础上提出了一种新的无监督损失函数嵌入非对称度量到神经网络中,构建了一种新颖的无监督网络架构(DECAMEL)。DECAMEL网络通过学习一个更紧凑的跨视角的聚类结构,可以缓解具体视角

偏差,并且促进潜在跨视角判别信息的挖掘来进行无监督的行人重识别。

DECAMEL 算法流程图如图 1 所示,沿着棕色箭头,具体的流程步骤如下:

- 1、首先通过深度网络来提取行人图像特征;
- 2、由于初始化的特征空间存在着具体的视角偏差,因此通过一种基于聚类的非对称度量学习(CAMEL)来学习一个初始化的非对称矩阵;
- 3、接下来通过优化所提的无监督损失,DECAMEL以一种端到端的方式联合学习特征表达与非对称矩阵。
- 4、在测试阶段,通过采用非对称度量变化下的特征来计算成对采样之间的距离,进行行人重识别。

图 1 中 U_v 表示在第 v 个摄像头视角下,对行人图像特征进行的具体特征变换, $d(x_1, x_2)$ 表示成对采样特征 x_1, x_2 之间的距离度量。本文中的 f_{loss} 由三部分组成:聚类误差,跨视角的一致性正则误差以及一种约束条件误差。

本文算法模型在 7 种基准数据集上进行了大量实验。实验结果表明,本文算法均优于其他无监督的行人重识别算法模型,进一步表明了本文所提算法的有效性。

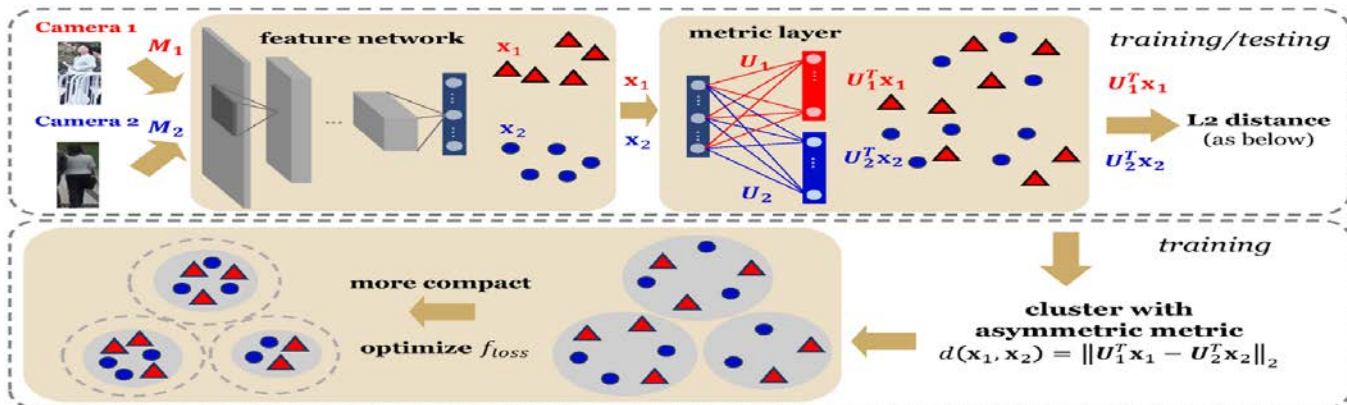


图 1 DECAMEL 算法流程图

责任编辑 樊鑫 贾同

加州大学圣迭戈分校苏昊教授团队

UC San Diego
SU Lab

美国加州大学圣迭戈分校苏昊教授团队研究聚焦于人工智能相关的基础问题，包括机器学习、计算机视觉、计算机图形学、机器人等。

团队发展历史

苏昊教授团队的研究方法并重现代人工智能研究的两大支柱：(1) 大规模数据集的建设；以及(2) 有理论支撑的核心机器学习算法设计。其研究以视觉数据的感知为支点，逐渐拓展到与物理过程相关的决策、推理与知识发现等认知问题。该团队的工作多次推动了人工智能新思潮的进步。

苏昊教授曾参与 ImageNet^[1]及领导 ShapeNet^[2]等大型数据集的建设工作。所提出的 RenderForCNN^[3]首次使用计算机合成图像方法生成数据用于神经网络训练，搭建了计算机图形学与深度学习两大领域之间的桥梁，这一思路目前已被学界与业界用于其他深度学习任务中。所提出的 PointNet^[4,5]及其系列工作首次提出使用神经网络直接对点云数据进行处理，是当今三维深度学习领域的开创奠基论文之一。

在 ImageNet^[1]数据集出现前，最大的图像数据集包含不足 100 种类的 10000 张图片，而 ImageNet^[1]共包含近 20000 种类的 20000000 张图片。ImageNet^[1]数据集被用于一系列图像分类、目标检测挑战赛，其结果证明了大型数据集对于提升深度学习能

力的重要性。ImageNet^[1]数据集构建的最大难点在于如何在保证标注质量的同时尽可能降低标注成本。为了克服这一困难，苏昊教授在博士研究生期间（普林斯顿大学、斯坦福大学），在李飞飞教授的支持下，探索并验证了通过在 Amazon Mechanical Turk 平台进行众包的方法收集图像检测标注的有效性。这也是第一个通过众包方式进行标注搭建的数据集。

为了推动深度学习在计算机图形学及三维计算机视觉的发展，苏昊教授在斯坦福大学读博期间，在 Leonidas Guibas 教授的支持下，发起并领导了 ShapeNet^[2]数据集的建设工作，并在加州大学圣迭戈分校任教期间，深入拓展了此项工作。ShapeNet^[2]数据集包含 4000 种类约 300000 个不同的几何体，在其中一个包含 50000 个不同物体的子集上标注有语义信息（物品种类及部件），几何信息（对称性），物理信息（质量，尺寸）等。ShapeNet^[2]目前已被学界与业界广泛采用，用于包括三维重建，三维形状分析，三维打印等多种任务中。

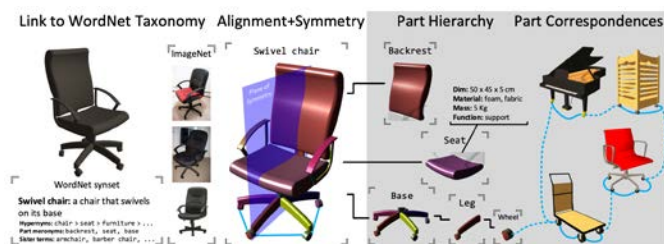


图 1: ShapeNet^[2]标注工具图示

对于视点估计、三维重建等深度学习任务，从真实

物体和传感器获得训练数据十分耗时。在 RenderForCNN^[3]工作前，学界和业界需要雇佣大量人力进行数据的采集和标注，同时数据的质量难以保证。为解决这一问题，苏昊教授与其合作者提出使用计算机图形学自动生成大规模带有准确标注的训练数据集，并提出一种新型的 CNN 网络结构用于视点估计任务。该方法基于此前所搭建的 ShapeNet^[2]数据集。此文及其提出的域随机化 (Domain Randomization) 技术，展示了通过虚拟环境训练的智能体在真实世界被有效使用的可行性，后来被广泛应用于生成机器人问题的训练数据（包括图像分割、强化学习等应用场景）。



图 2: RenderForCNN^[3]合成图像图示

在 PointNet^[4,5]系列工作前，三维深度学习方法一般将三维点云转化为三维体素网格，进而与二维图像类似地使用卷积神经网络进行处理。但由于三维体素网格的显存占用正比于分辨率的立方，使得此类方法难以用于实际应用场景。在 PointNet^[4,5]系列工作中，苏昊教授与其合作者提出使用对称函数的新型网络结构，可以直接对三维点云进行处理，提高网络的计算效率与灵活性，并在包括物体分类，部件分割，场景语义分割等任务中取得超出此前基于三维体素网格方法的表现。此外，还对 PointNet^[4,5]网络结构进行了理论分析，以进一步理解此类方法的能力与鲁棒性原因。

团队目前研究方向

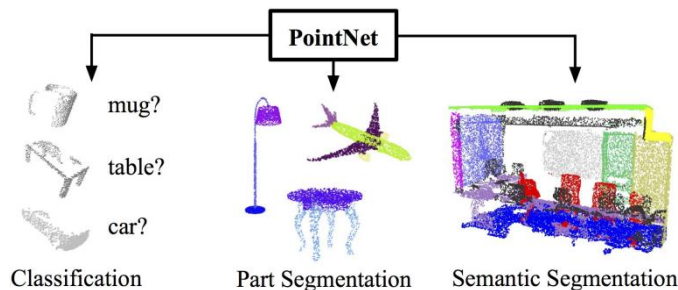


图 3: PointNet^[4,5]应用任务

苏昊教授团队的目前研究主要面向智能体在物理环境中主动及连续学习的理论与技术。具体研究内容包括计算机视觉与图形学、机器学习理论、机器人学等。

研究方向 1: 计算机视觉与图形学

苏昊教授团队始终致力于建设大规模数据集以推进深度学习与计算机视觉的进一步发展。此前，数据驱动的深度学习结果表明大规模数据集对于提升深度学习的能力有极大帮助，但是简单的监督学习方法可能使得网络泛化能力较弱，即对于训练中未见过的物体种类无法进行合理的预测。由三维物体的整体认知转变为由不同部件形状组合而成的组合体认知对于理解三维物体及其交互具有十分重要的意义。为了通过深度学习方法实现部件级别的组合体认知，苏昊教授团队在 ShapeNet^[2]的基础上构造了 PartNet^[6]。PartNet^[6]包含具有层级结构的精细实例级别的三维物体部件分割，并且保证了不同类别物体间部件分类的一致性。

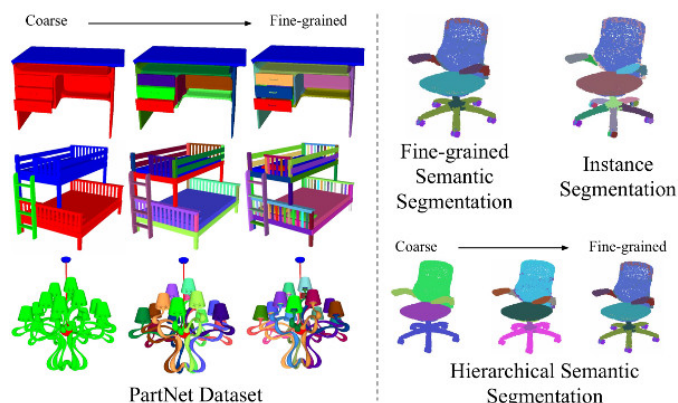


图 4: PartNet^[6]数据集图示

除了从大规模数据集中学习部件级别的组合体认知外，苏昊教授团队还研究如何通过部件之间的相对运动实现自监督的部件分割^[7]，从而使得机器人有可能通过对世界的观察，自主发现物体的组成结构。此任务的难点在于其物体种类以及部件数量是未知的，同时输入数据可能存在残缺或噪音。为了解决这一问题，苏昊教授团队提出匹配模块、移动模块与分割模块的迭代分割网络结构，如图 5 所示。其中首先使用 PointNet++ 网络预测部件运动前后的匹配关系，然后使用 PointNet++ 网络预测运动前部件各点的位移，进而根据各点的位移信息对物体进行部件分割。

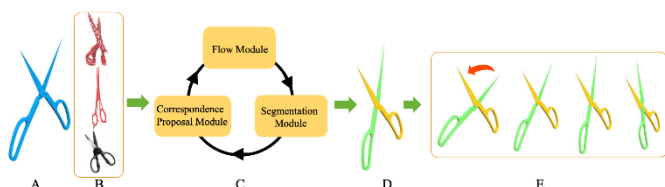


图 5：物体部件分割自监督学习方法

除部件分割外，苏昊教授团队对于三维重建、场景语义分割、新视角图像合成等问题也有深入研究。现有的基于深度学习的多视角三维重建方法均利用三维成本体实现多视角图像特征的融合以及深度预测。其不足在于三维成本体显存占用高，无法进行高分辨率、高精度的深度预测，同时现有方法没有考虑不可见视角图像特征对深度预测的负面影响，影响深度预测精度。针对以上问题，苏昊教授团队提出基于点云深度学习的可见性感知的多视角三维重建网络框架^[8,9]。其首先利用可见性感知的多视角图像特征聚合模块对每个视角中各像素的可见性进行独立预测，进而仅使用可见视角进行特征聚合，避免了不可见视角对深度预测的影响；且利用点云作为重建物体表面的表征形式，并使用假设点和图神经网络将网络计算集中于物体表面邻域，减少了其他对最终深度预测没有帮助的计算，提高了计算效率及深度预测的分辨率和精度。

研究方向 2：机器学习理论

在机器学习理论研究方面，苏昊教授团队研究利用

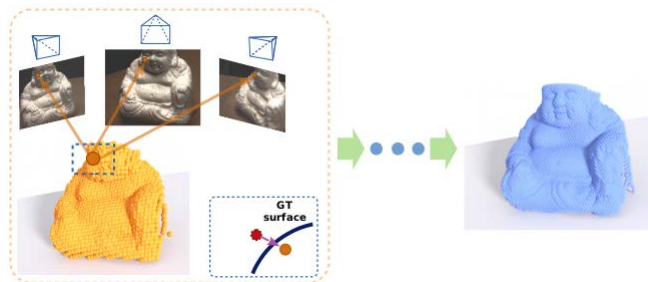


图 6：基于点云深度学习的可见性感知的多视角三维重建网络框架^[8,9]

模仿学习、迁移学习、结构化模型等方法提高强化学习的采样效率以及探索能力，并在机器人抓取、视频游戏等多种挑战性场景中验证其方法的有效性，以及研究对于二维卷积神经网络和三维点云神经网络的对抗攻击防守理论、局部最小值表征方法等。

一个充分理解环境的智能体应能够将所学习的知识用于任意给定目标，因此其应当学习预测所有状态与目标的累计奖励映射。但是现有的强化学习方法很难对长远视野目标的价值函数进行准确估计，导致最终策略失败。针对这一问题，苏昊教授团队提出对环境进行显式的层级建模^[10]，其高层使用动态地标建立访问过的状态空间的抽象映射，低层使用价值网络实现准确的局部决策。实验证明，使用所提出的方法可以使智能体在训练初期阶段达到长期目标，并在多个具有挑战性的任务上取得比现有强化学习方法更好的结果。

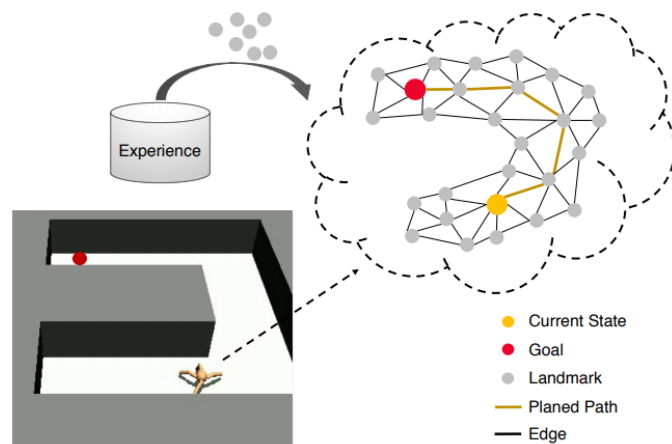


图 7：用路标进行状态空间映射实现全局目标到达^[10]

现有的模仿学习方法当模仿者与专家具有不同的动

力学模型时通常会导致学习失败，其主要原因在于现有方法专注于模仿动作而非状态。为了解决这一问题，苏昊教授团队提出新的基于状态对齐的模仿学习方法^[11]，以训练模仿者尽可能跟随专家演示中的状态序列。状态对齐来自于局部和全局模型，并使用策略更新目标将其与强化学习模型进行结合。实验表明，所提出的方法对于模仿者和专家具有不同动力模型的情形可以更好地实现模仿学习。

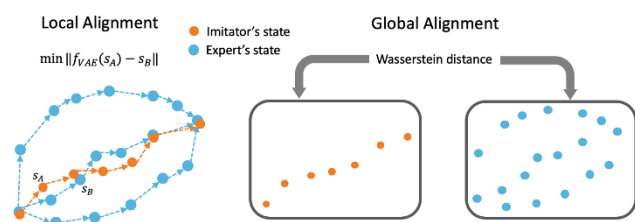


图 8：基于状态对齐的模仿学习方法^[11]

深度学习理论的最近发展引发了对于深度神经网络的不同局部极小值的泛化性的研究。现有工作一般专注于发现良好的局部极小值的性质或研究正则化方法使得网络可以收敛于更好的局部极小值，但尚无可以同时解决这两个问题的方法。苏昊教授团队提出一种基于 Fisher 信息的度量方法^[12]，既可以衡量局部极小值的泛化性能，也可以作为一种有效的正则化方法，从而以一个统一的框架解决这两个问题。对于所提出方法，进行了泛化性能范围的理论分析，并在 CIFAR-10 与 CIFAR-100 上对各种网络结构进行实验验证。

研究方向 3：机器人学

长期以来，建造家庭助理机器人一直是视觉和机器人研究者的追求。为了实现此任务，具有物理逼真的模拟、足够的关节物体以及可移植到真实机器人的模拟环境是必不可少的。现有的机器人仿真环境一般仅满足以上的部分需求，均存在一定的简化，使其无法满足所有的需求。因此，苏昊教授团队领导搭建了 SAPIEN (<https://sapien.ucsd.edu/publication>)^[13]，一个包含各种生活常见物体的物理仿真环境。同时为了增进此仿真环境的真实性，SAPIEN^[13]中所有物体均通过人工

标注实现可活动关节部件的划分，使得 SAPIEN^[13]可用于更复杂且真实的任务学习，例如，打开冰箱门，取出可乐并拧开瓶盖。SAPIEN^[13]中整合了 PhysX 物理引擎与 ROS 控制接口，使得此环境更便于由仿真环境到真实机器人应用的迁移。苏昊教授团队认为 SAPIEN^[13]仿真环境对于推进机器人通过与环境交互实现主动连续学习具有重要的意义，并希望计算机视觉以及机器人社区可以以此为基础共同开发。

除了仿真环境外，团队也研究真实机器人应用场景中基础重要任务，如物品抓取与操作。在 S4G^[14]中，团队研究利用点云深度学习，由包含噪音的单视角三维测量点云中单阶段生成可行的平行夹爪抓取位姿，相比于现有的双阶段采样检测方法，大幅提高了检测速度与抓取成功率。S4G^[14]使用计算机合成的抓取数据集进行训练，并在未进行任何微调的情况下可以直接应用于真实抓取场景，证明了点云深度学习具有较好的泛化性能。

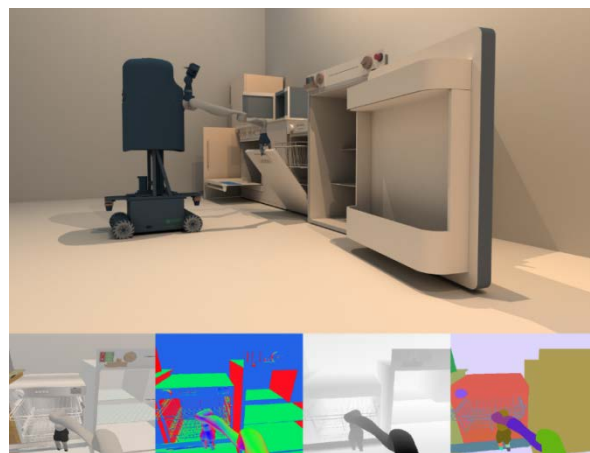
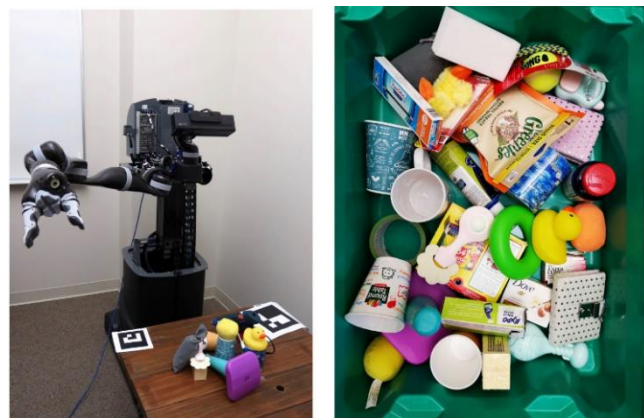


图 9：SAPIEN^[13]机器人物理仿真环境

未来研究工作

面向智能体在物理环境中主动以及连续学习，现有深度学习方法虽然取得了很大的进展，但仍有许多基础性问题，比如因果推理等问题依旧没有解决。苏昊教授团队希望可以在现有工作的基础上继续推进，提高机器人的推理能力以及与真实物理环境进行交互、执行复杂任务的能力。

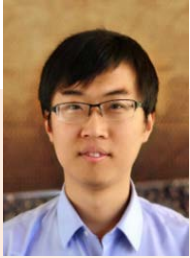
图 10: SAPIEN^[13]支持的机器人操作任务图 11: S4G^[14]机器人抓取实验

责任编辑 张汗灵

参考文献

- [1] Russakovsky, Olga, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang et al. "Imagenet large scale visual recognition challenge." *International Journal of Computer Vision* 115, no. 3 (2015): 211-252.
- [2] Chang, Angel X., Thomas Funkhouser, Leonidas Guibas, Pat Hanrahan, Qixing Huang, Zimo Li, Silvio Savarese et al. "Shapenet: An information-rich 3d model repository." *arXiv preprint arXiv:1512.03012* (2015).
- [3] Su, Hao, Charles R. Qi, Yangyan Li, and Leonidas J. Guibas. "Render for CNN: Viewpoint estimation in images using CNNs trained with rendered 3d model views." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2686-2694. 2015.
- [4] Qi, Charles R., Hao Su, Kaichun Mo, and Leonidas J. Guibas. "Pointnet: Deep learning on point sets for 3d classification and segmentation." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 652-660. 2017.
- [5] Qi, Charles Ruizhongtai, Li Yi, Hao Su, and Leonidas J. Guibas. "Pointnet++: Deep hierarchical feature learning on point sets in a metric space." In *Advances in Neural Information Processing Systems*, pp. 5099-5108. 2017.
- [6] Mo, Kaichun, Shilin Zhu, Angel X. Chang, Li Yi, Subarna Tripathi, Leonidas J. Guibas, and Hao Su. "Partnet: A large-scale benchmark for fine-grained and hierarchical part-level 3d object understanding." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 909-918. 2019.
- [7] Yi, Li, Haibin Huang, Difan Liu, Evangelos Kalogerakis, Hao Su, and Leonidas Guibas. "Deep part induction from articulated object pairs." *ACM Transactions on Graphics (TOG)* 37, no. 6 (2019): 209.
- [8] Chen, Rui, Songfang Han, Jing Xu, and Hao Su. "Point-based multi-view stereo network." In *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1538-1547. 2019.
- [9] Chen, Rui, Songfang Han, and Jing Xu. "Visibility-aware point-based multi-view stereo network." *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2020).
- [10] Huang, Zhiao, Fangchen Liu, and Hao Su. "Mapping state space using landmarks for universal goal reaching." In *Advances in Neural Information Processing Systems*, pp. 1940-1950. 2019.
- [11] Liu, Fangchen, Zhan Ling, Tongzhou Mu, and Hao Su. "State alignment-based imitation learning." *arXiv preprint arXiv:1911.10947* (2019).

- [12] Jia, Zhiwei, and Hao Su. "Information-theoretic Local minima characterization and regularization." arXiv preprint arXiv:1911.08192 (2019).
- [13] Xiang, Fanbo, Yuzhe Qin, Kaichun Mo, Yikuan Xia, Hao Zhu, Fangchen Liu, Minghua Liu et al. "SAPIEN: A Simulated Part-based Interactive ENvironment." arXiv preprint arXiv:2003.08515 (2020).
- [14] Qin, Yuzhe, Rui Chen, Hao Zhu, Meng Song, Jing Xu, and Hao Su. "S4g: Amodal single-view single-shot se (3) grasp detection in cluttered scenes." arXiv preprint arXiv:1910.14218 (2019).



苏昊(Hao Su)

苏昊，曾于斯坦福大学获得计算机科学博士学位，北京航空航天大学获得数学博士学位。现任美国加州大学圣迭戈分校计算机系助理教授，并与机器人研究所和视觉计算中心保持合作关系。苏昊教授曾担任 ICCV、CVPR、AAAI、Pacific Graphics、3DV 等国际会议的领域主席 (Area Chair) 或高级程序委员会主席 (Senior Program Chair)。苏昊教授团队研究聚焦于人工智能相关的基础问题，包括机器学习、计算机视觉、计算机图形学、机器人等。苏昊教授的 ShapeNet、PointNet 系列工作等被视为三维深度学习领域的开创性工作。

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。

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, Pattern Recognition Letters 和 International Journal of Approximate Reasoning。

3 会议简介

国际计算机视觉与模式识别会议 (CVPR, IEEE Conference on Computer Vision and Pattern Recognition) 是 IEEE 一年一度的学术性会议，会议的主要内容是计算机视觉与模式识别技术。CVPR 是世界顶级的计算机视觉会议 (三大顶会之一)。

2021 年 CVPR 在美国纳什维尔举行，本届会议将汇聚全世界从事计算机视觉理论与应用研究的广大科研工作者及工业界同仁，共同分享计算机视觉领域的最新理论和技术成果，为大家提供精彩的学术盛宴。

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
WACV 2021	2021.01.05-06	Waikoloa, USA	2020.09.30	http://wacv2021.thecvf.com/home
ICLR 2021	2021.05.04-05	Vienna, Austria	2020.10.02	http://iclr.cc/
AAMAS 2021	2021.05.03-07	London, UK	2020.10.10	https://aamas2021.soton.ac.uk/
AISTATS 2021	2021.03.11-15	California, USA	2020.10.15	https://www.aistats.org/aistats2021/
CVPR 2021	2021.06.19-25	Nashville, USA	2020.11.17	http://cvpr2021.thecvf.com/

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
IJAR	Probability and Statistics: Foundations and History. In honor of Glenn Shafer	https://www.journals.elsevier.com/international-journal-of-approximate-reasoning/call-for-papers/foundations-and-history-in-honor-of-glenn-shafer	2020.10.21
IVC	Advances in Domain Adaptation for Computer Vision	https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/advances-in-domain-adaptation-for-computer-vision	2020.11.2
IJAR	Time Series Clustering and Classification	https://www.journals.elsevier.com/international-journal-of-approximate-reasoning/call-for-papers/special-issue-on-time-series-clustering-and-classification	2020.12.1
PR	AI for Combating COVID-2019	https://www.journals.elsevier.com/pattern-recognition/call-for-papers/special-submission-stream-on-ai-for-combating-covid-2019	2020.12.31
PRL	Deep Learning for Precise and Efficient Object Detection	https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/deep-learning-for-precise	2020.12.31

COMPUTER VISION NEWSLETTER

03 2020
总第 25 期



计算机视觉专委会简报



CCF 计算机视觉
专委会