

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

01 2020

总第 23 期



CCF 计算机视觉
专委会

COMPUTER VISION NEWSLETTER



计算机视觉专委会 简报

2020 年第 01 期
总第 23 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 走进高校系列活动
- 06 视界无限系列研讨会
- 08 CCF-CV 常委委员会工作会议成功召开
- 09 CCF-CV 专委工作会议暨换届选举顺利召开
- 10 CCF-CV 走进高校系列报告会活动细则
- 12 CCF-CV 企业交流活动方案更新
- 13 CCF-CV 成立专委顾问委员会
- 14 CCF-CV 秘书处工作会议圆满召开

| 科技前沿

- 15 专题综述：食物识别研究进展
- 19 热点追踪：多模态医学图像合成与重建
- 20 热点追踪：基于 AI 的新冠肺炎肺部 CT 影像量化分析
- 21 顶会观察：ACM Multimedia 2019

| 委员风采

- 24 北京航空航天大学李甲教授访谈
- 28 华南理工大学许勇教授访谈
- 30 委员好消息

| 学术资源

- 32 场景图生成开源代码
- 34 迁移学习数据集
- 37 好文推荐

| 海外学者

- 39 美国纽约州立大学石溪分校凌海滨教授团队
- 43 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 82 期 河南科技学院



2019年11月29日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、河南科技学院信息工程学院与人工智能学院共同承办的第82期CCF-CV走进高校系列报告会活动——“计算机视觉前沿技术及应用”在河南科技学院行政楼206报告厅成功举行。本期报告会邀请了北京工业大学毋立芳教授、武汉大学张洪艳教授、华中科技大学刘文予教授、北京航空航天大学李文玲副教授和中科院计算所韩琥副研究员五位专家学者做了特邀报告。本次报告由河南科技学院信息工程学院院长李国厚教授担任执行主席，来自信息工程学院、人工智能学院、河南师范大学的150多名师生聆听了报告。

报告会现场气氛热烈，在场的师生们就专家的报告纷纷提出自己感兴趣的问题和学术研究遇到的困惑，现场专家也给出了详细且针对性的解答，极大地鼓舞了我校师生们的科研热情，报告会在热烈的气氛中圆满结束。

第 83 期 悉尼大学



2019年12月4日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，悉尼大学电子信息工程系联合承办的第八十三期CCF-CV走进高校系列报告会在悉尼大学ABS Case Study Lecture Theatre 2140成功举行。本期报告会邀请了电子科技大学孟凡满副教授、北京大学邹月娴教授、悉尼大学徐畅博士、华南理工大学许勇教授、悉尼科技大学朱霖潮博士、中科院计算所纪雯研究员、美国密苏里大学李竹副教授七位专家学者做特邀报告。悉尼大学欧阳万里博士担任本次报告会的执行主席。

本期CCF-CV走进高校系列报告会持续了三个多小时，吸引了来自悉尼大学等单位的师生前来聆听，七位计算机视觉领域的专家学者通过幽默风趣的讲解，给大家系统介绍了自己的研究领域和最新研究成果。报告内容精彩、反响热烈，极大地激发了师生们的科研热情。最后，参加报告会的现场听众对七位特邀讲者表示由衷的欢迎与感谢，报告会在雷鸣般的掌声中圆满结束。

第 84 期 广东石油化工学院



2019年12月4日下午，由中国计算机学会计算机视觉专委会 (CCF-CV) 主办，广东石油化工学院自动化学院与计算机学院联合承办的第 84 期 CCF-CV 走进高校系列报告会——“人工智能前沿技术及应用”在广东石油化工学院图书馆报告厅二厅举行。本期报告会特邀中山大学数据科学与计算机学院赖剑煌教授、哈尔滨工业大学计算机学院左旺孟教授、江西财经大学方玉明教授、西北工业大学计算机学院王鹏教授四位专家学者做报告。广东石油化工学院自动化学院院长康文雄教授和计算机学院院长荆晓远教授担任本次报告会的执行主席。

本期报告会持续了三个多小时，来自广东石油化工学院、华南理工大学，武汉大学和贵州民族大学的上百名师生聆听了四位专家在该领域的最新动态介绍以及他们各自的最新研究成果与心得。会议气氛热烈，参会师生积极与专家交流互动。

最后，荆晓远教授进行简要总结，感谢四位专家带来的精彩报告，祝贺本次报告会取得圆满成功！

第 85 期 北京工业大学



2020年1月7日上午，由中国计算机学会计算机视觉专委会 (CCF-CV) 主办，北京工业大学信息学部联合承办的第 85 期 CCF-CV 走进高校系列报告会——“计算机视觉前沿技术及应用”在北京工业大学人文楼 514 会议室举办。本期报告会特邀清华大学刘永进教授、哈尔滨工业大学左旺孟教授、中国科学院计算技术研究所王瑞平研究员、北京航空航天大学李甲研究员、南开大学杨巨峰教授 5 位专家学者做报告。北京工业大学贾克斌教授、孙艳丰教授、杨震教授担任本次报告会的执行主席。

此次报告会持续了四个多小时，五位专家展示了计算机视觉的前沿技术，为与会师生研究指明了方向。报告会现场气氛热烈，报告会在热烈的气氛中圆满结束。

最后，中国计算机学会计算机视觉专委会马占宇秘书长代表专委会讲话，对专家的报告致以感谢，也对承办单位的组织表示感谢！其次介绍了 CCF-CV 专委会的活动，最后感谢大家对 CCF-CV 专委会长期以来的支持。

责任编辑 毋立芳

第 5 期 上海科技大学

CCF-CV 视界无限系列研讨会



2020年1月4日，由中国计算机学会计算机视觉专委会主办、上海科技大学信息科学与技术学院承办的第5期 CCF-CV “视界无限”系列活动——“虚拟现实/增强现实 (VR/AR) 的前沿进展与未来趋势”研讨会在上海科技大学成功举行。本期研讨会由计算机视觉专委会副主任、上海科技大学虞晶怡教授组织和主持。研讨会邀请了北京大学查红彬教授、美国加州理工学院 Mathieu Desbrun 教授、清华大学刘焯斌副教授、浙江大学章国锋教授、上海交通大学严骏驰副教授、微软亚洲研究院王井东研究员做引导发言，大连理工大学卢湖川教授、南开大学程明明教授、爱奇艺王涛研究员、上海科技大学何旭明副教授、高盛华副教授、Laurent Kneip 博士参加了深度研讨环节。来自全国高校和企业的 60 余名参会者汇聚申城，就 VR/AR 的理论与应用进行了热烈的交流和深入细致的探讨。

会议议程依次包括引导发言、深度研讨、系统展示三部分。首先，虞晶怡教授致辞。虞老师回顾了上海科技大学 2016 年承办计算机视觉专委会 RACV 会议的情景，欢迎与会嘉宾和师生再次来到上科大，共同探索计

算机视觉技术在 VR/AR 领域的应用。希望上科大新的设施和环境帮助大家更好地理解知识、激发创意；希望通过深入研讨推动 VR/AR 同行之间的交流、促进合作、提升影响力，并祝愿研讨会圆满成功。



查红彬教授的引导发言题目是“SLAM 研究的新进展：从多视点几何计算到在线学习”。他指出，传统的 SLAM 技术充分利用多视点几何与 SfM (Structure from Motion) 等领域的高效算法，并通过与深度传感器、惯性传感器等下一代传感设备的数据融合，在传感器轨迹计算精度与三维场景重建质量方面取得了显著的进步。但在实际应用中，仍存在重建误差积累严重、计算成本高昂等问题，影响了机器系统的在线响应速度以及对复杂环境的自适应能力。针对这些问题，应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能力，进一步改善其基本性能。

Mathieu Desbrun 教授以“Towards Virtual Physical Reality”为题，深入浅出地介绍了虚拟物理现

实的任务定义、主要难点和研究现状。作为计算机图形学领域国际知名的专家，他风趣地分享了研究该问题的缘起，并从学术与工业双重视角出发，给出了该领域当前的开放问题以及下一阶段值得关注的重点。

刘焯斌副教授的报告题目是“人体动态三维重建技术与应用”。人体对象是视觉场景信息中最重要的部分，具有复杂的纹理外观特性、非刚性形变和高动态特性。人体对象的动态三维重建可应用于全息三维影像通信、3D 人体试衣、增强现实、安防视频人体行为分析、各类智能机器三维视觉分析、影视娱乐游戏等。报告围绕人体三维重建的精准、实时、便捷、大范围、语义化等技术目标，介绍了近年在多视点相机、单深度相机、单 RGB 图像/视频的人体三维重建等方面的研究工作，并对未来潜在应用进行了讨论。

章国锋教授的报告题目是“面向增强现实应用的视觉 SLAM 与高精度地图构建”。他指出，在过去的十多年里视觉 SLAM 和视觉惯性 SLAM 技术已经取得了很大的成功，并成功应用于 AR 产品。但在实际应用中，开发一个鲁棒高效的 SLAM 系统仍然很有挑战性。例如，在实际增强现实应用中，很容易遇到各种各样很有挑战性的情况，如快速运动、强旋转、严重的运动模糊、动态干扰等等，如何在这种复杂非配合环境下做到稳定运行，从而保证良好的增强现实体验？如何高效鲁棒地对室内外大规模场景进行扫描和构建高精度三维地图？如何在一个低功耗的移动设备上也能做到实时精准的跟踪定位？章老师的报告为这些关键问题提供了解决方案，并展示了一系列增强现实应用。

严骏驰副教授的报告题目是“Robust Correspondence Recovery via Neural Graph Matching”，重点介绍了近年来提出的多个可用于增强现实的超图匹配和多图匹配方法，并对图匹配与图割、聚类、链接预测等经典问题的结合方式进行了展示。最后，严老师展望了图匹配与机器学习问题的未来交叉研究方向。

王井东研究员的报告题目是“Face alignment, human pose estimation and parsing with high-resolution networks”。人体图像分析是 VR/AR 领域

的重要研究对象之一。他介绍了在人体图像分析中几个关键问题的最新工作，包括基于高分辨率网络 HRNet 的人脸关键点检测和人体姿态估计、以及基于 HRNet 和物体上下文 (object context) 的人体部件分割等。

在 Panel 环节，与会嘉宾与现场师生就“VR/AR 的挑战性问题和发展趋势是什么？”、“VR/AR 今后的发展，是往云端走还是终端走？云边端分别计算什么？”、“VR/AR 的某些场景中传统特征提取方法会被取代吗？”、“RGB 和 RGB-D，哪个更加有用？”等问题展开热烈讨论，参会者也就自己关心的问题与嘉宾交换了观点。



在系统展示环节，与会者参观了上海科技大学多学科人工现实工作室 (MARS) 的虚拟现实/增强现实应用成果，并与系统开发者进行了面对面的交流互动。



最后，第 5 期“视界无限”研讨会在热烈的掌声中圆满结束。

责任编辑 杨巨峰

CCF-CV 常务委员会工作会议成功召开



2020年1月4日，中国计算机学会计算机视觉专委会（CCF-CV）常务委员会工作会议于上海成功召开。

本次常委会工作会议是专委会换届后的首次会议，因此对于专委会的持续蓬勃发展具有承前启后的重大意义。会议系统地梳理了专委会各项工作的进展、存在的不足，明确了专委会后续发展的重点，特别为专委会系列品牌学术活动的继续创新突破指明了方向。

会议首先由专委会主任查红彬教授做主题发言，他充分肯定了专委会过去一年取得的各项进展，对接下来的工作提出了多项切实可行的发展建议。

陈熙霖研究员代表第二届专委会回顾总结了过去四年的工作，他特别感谢谭铁牛院士倡议发起和带领筹建计算机视觉专委会并积极领导组织各项活动，感谢所有委员、常委、秘书处及每一位参与活动的人。他指出上一届专委会为国内计算机视觉社区和年轻学者及学生做了很多贡献，尤其在培养年轻人方面起了很多正面

作用。祝愿专委会在新一届主任带领下把各项工作做的更好，取得更大的成绩。

之后，常委会委员们针对专委会发展相关议题展开了高效深入的讨论，并逐一形成了具体可行的指导性建议。相关议题均由专委会委员们积极建言献策而来，秘书处进行了分类汇总，涉及专委会日常工作、学术活动、组织建设、领域发展等4个方面近30个议题。其中，值得广大委员们期待的专委会发展举措包括：调整专委会简报形式和板块内容、增加对企业的推荐和介绍、开设专委会网站英文版、增设顾问委员并优化增选委员规则等。

会议最后，查红彬主任向第二届常务委员会委员颁发了感谢证书，并作总结发言。会议结束后，全体与会委员合影留念。



责任编辑 杨巨峰

CCF-CV 专委工作会议暨换届选举顺利召开



中国计算机学会计算机视觉专委会（CCF-CV）2019年度工作会议于2019年11月9日在西安曲江国际会议中心成功举办，来自全国高校、科研院所、企业的现任委员300余人参加了工作会议。会议由专委会秘书长、中科院自动化所王亮研究员主持。

会议首先由专委会副主任、中科院计算所陈熙霖研究员致辞。陈熙霖研究员代表专委会主任谭铁牛院士感谢委员们从祖国各地到场参会，他回顾了本届专委会四年来的发展历程和主要成绩，积极肯定了专委会在过去一年中在学术交流及社会服务等各方面取得的丰硕成果，鼓励委员们在当前人工智能热潮中聚焦学术前沿、做出高质量研究工作，希望专委会大家庭一起精诚合作，各项活动进一步提升品牌质量，打造精品学术交流活动，为推进计算机视觉学术研究与产业发展继续发挥积极引领作用。

之后，中国计算机学会理事、专委工委委员、西北工业大学於志文教授代表学会致辞。於志文教授对专委



工作会议的召开表示祝贺，对专委开展的各项活动给予了积极评价，对专委换届选举作的重要事项提出了指导意见。他最后预祝2019年计算机视觉专委会全体会议及换届选举圆满成功。

随后，专委会秘书长王亮研究员向与会委员做年度工作报告。报告介绍了专委会组织结构及过去5年获得的计算机学会评优奖励，通报了4月份专委会常委会议和5月份秘书处工作会议提出和实施的若干新举措，全面回顾了专委会过去一年的各项学术交流活动，指出了工作中存在的问题和拟采取的改进方案，最后介绍了下一年度专委会工作计划。

颁奖仪式由专委会副主任、专委提名与奖励工作组组长、北京大学查红彬教授主持，查红彬教授详细介绍了本年度设立的奖项评选范围与评选规则，包括：终身学术贡献奖、杰出成就奖、服务贡献奖、学术新锐奖。4个奖项的获奖人既有早年投入我国计算机视觉基础性研究的老一代科学家，又有新一代崭露头角的学术新星。

责任编辑 任传贤

CCF-CV 走进高校系列报告会活动细则

自 2015 年 11 月开展以来, CCF-CV 走进高校系列报告会活动得到了高校、讲者、听众的大力支持。截至 2019 年 11 月底, 走进高校系列报告会活动已成功举办了 85 期, 遍及祖国大陆。目前, 活动已邀请讲者做分享专题报告 320 余场, 活动现场人均听众 200 人次, 微信公众号平均阅读 1000 余次, 在征得讲者同意的前提下, 于爱奇艺网站上传报告会的现场视频, 并于专委会网站分享讲者的 PPT, 在国内外计算机视觉领域引起了强烈反响。在这里, 向对活动开展提供支持和帮助的承办方以及分享报告的讲者们表示由衷的感谢!

在新一届专委会领导班子的领导下, 我们拟针对活动存在的“单向输出”的问题进行改进, 加强讲者与承办单位的深度交流, 发挥领域资深讲者对承办单位的指导作用。并于春节假期期间发放了调查问卷, 得到大家的踊跃响应和积极反馈。

我们对 CCF-CV 走进高校系列报告会活动进行了全面总结, 并在反馈意见的基础上, 对活动内容、活动申请、讲者邀请等进行了完善。以下是修订的活动细则, 欢迎各位踊跃申请, 也感谢大家长期以来的关注和支持!

CCF-CV 走进高校系列报告会活动细则

(2020 年 2 月修订)

1. 宗旨

为了更好地推动计算机视觉学科领域的学术与技术交流, 促进学者间的了解与合作, 推动计算机视觉的发展, 提升我国计算机视觉国际领域的影响力, 中国计算机学会计算机视觉专委会拟在全国范围的高校

和科研院所开展 CCF-CV 走进高校系列报告会活动。

2. 活动内容

活动内容包括如下:

1) 邀请报告 (必选): 邀请 3 名左右讲者就领域前沿问题作报告;

2) 研究交流 (可选):

a. 研究点评: 承办单位 1-2 名在读博士生或青年教师汇报自己的研究工作, 专家点评, 并提出研究建议;

b. 答疑解惑: 讲者和承办单位师生以座谈会的方式进行交流, 专家就承办单位师生提出的人才培养、科学研究等方面的问题进行解答并给出建议;

c. 现场参观: 专家参观承办单位的工作, 提出建议;

d. 其他活动: 承办单位根据实际需求提出, 和秘书处协商确定。

3. 活动形式

a. 活动由 CCF-CV 主办, CCF-CV 秘书处负责协调和指导工作;

b. 每次活动由一家单位承办, 主要负责安排论坛主题、联系讲者、推广宣传、场地预定、通知听众、撰写新闻稿等;

c. 每次活动邀请三名左右的讲者, 活动内容包括邀请报告、研究交流两个环节;

d. 每月原则上组织不超过 2 次活动;

e. 原则上同一城市两次活动间隔在三个月以上。

4. 活动申请

- a. 活动采取自愿申请的方式，有意承办活动的单位须向秘书处提交书面申请；
- b. 承办方需提前至少三个月向秘书处提出申请，秘书处将在接到申请两周内商定并给予答复。

5. 讲者邀请

- a. 活动讲者采取邀请为主、推荐辅助的方式，由秘书处与承办方共同完成，讲者确定后，由秘书处统一给讲者发正式邀请邮件；
- b. 为了更好地宣传计算机视觉及相关领域的原创性工作，鼓励国内外学者将自己的工作以“CCF-CV 走进高校活动讲者信息表”的形式发给秘书处活动联系人，我们将根据活动主题进行匹配推荐；
- c. 最终讲者名单由执行主席确定。

6. 活动准备与宣传

- a. 活动开始日期前至少 1 个月启动准备工作；
- b. 活动时间地点由秘书处和承办方协商确定；
- c. 秘书处联系讲者填好资料：CCF-CV 走进高校活动讲者信息表和个人照片等资料发给承办方，承办方参考宣传材料模板制作本次活动宣传材料；
- d. 承办方须提前两周将制作完成的宣传材料和一张承办单位的横向照片发至专委会秘书处，以便在专委会网站和微信公众号上统一宣传。

7. 活动现场

- a. 活动现场悬挂横幅“CCF-CV 走进高校系列报告会（第 x 期）”，电子横幅也可；
- b. 活动开始前，承办方负责提醒讲者填写“CCF-CV 走进高校系列报告会讲者报告视频公开知情书”，其中甲方由活动执行主席代表专委签字；
- c. 活动开始前，参会人员签到，签到表由承办单位自行设计，注意签到表中必须包括信息：姓名、单位、身份、是否 CCFCV 专委委员，是否 CCF 会员；
- d. 活动开始前及休息间隙放映 CCF-CV 宣传页

PPT 和专委宣传片。

- e. 活动期间，承办方须按照讲者的意愿安排录制现场活动的音视频。

8. 活动总结

- a. 活动结束后两日内承办方须提交以下材料给秘书处：1) 新闻稿；2) 讲者签字版知情书扫描件；
- b. 活动结束后两周内承办方须提交以下材料给秘书处：1) 讲者 PPT；2) 录像资料（与知情书一致）；3) 签到表电子版和学生、教师、CCF 会员、CCF-CV 专委委员统计数据。
- c. 秘书处负责把讲者同意分享的 PPT 放到专委会网站，并将讲者同意分享的视频放到爱奇艺视频空间。

9. 注意事项

- a. 活动的时间和地点由秘书处和承办方协商；
- b. 承办方需承担活动场地、宣传等相关费用，支付讲者课时费以及承担讲者活动期间的本地交通和餐饮费用。对于有条件的承办方，可在上述费用基础上，额外承担讲者城市间往来交通、住宿等费用。原则上讲者自己承担交通和住宿费用。具体费用承担情况，承办方、讲者和秘书处在活动前协商一致即可；
- c. 承办方有义务为讲者出具官方的邀请信等证明材料，以便讲者报销差旅费用；
- d. 活动开始至少一个月前，秘书处与承办方确定讲者、场地等事宜。
- e. 承办方负责录制现场活动的音视频，并在活动结束后将论坛的影音资料和活动报道等发送给秘书处，经审核后发布。

10. 其他

- a. 该活动以学术交流为主要目的，严禁以活动名义进行非法集会，传播非学术内容；
- b. 若有赞助等商业活动，须与秘书处提前商定。

责任编辑 毋立芳

CCF-CV 企业交流活动方案更新

从 2015 年底至今，CCF-CV 企业交流活动得到了企业、讲者、听众的大力支持。截至 2019 年底，CF-CV 企业交流活动已成功举办了 20 期，相继走进了多家国内外知名企业。在这里，向对活动顺利开展提供支持和帮助的承办方以及分享精彩报告的讲者们表示由衷的感谢！在新一届专委领导班子的领导下，我们拟对现有活动形式进行升级，以进一步加强企业与专委的沟通交流。以下是活动细则，欢迎各位踊跃申请，也感谢大家长期以来的关注和支持！

1. 宗旨

为了更好地推动 CCF-CV 专委与计算机视觉领域高科技企业之间的交流与合作，最大程度发挥双方在人才、资源等方面的优势，实现共享互赢，专委拟开展形式多样的 CCF-CV 企业交流互动活动。

2. 活动内容

活动分为以下三个类型，每次择一种开展：

- 1) 专委走进企业：高科技企业举办交流会，邀请研发主题相关的专委委员介绍他们的研究成果，并与企业技术人员探讨产品研发过程中遇到的实际问题及解决思路；
- 2) 企业走进专委：专委组织的走进高校等活动中，邀请企业研究人员以学术报告的形式介绍他们的研发成果和产品、解决的问题；
- 3) 专委 meets 企业：在 PRCV 或 RACV 等专委主办/承办的大会期间，设立互动专区或专题竞赛，专委委员及企业研究人员现场讲解各自的研究成果，探讨研究升

级思路及合作机会。

3. 活动组织

- 1) 疫情期间暂缓实地参观活动，以企业 Webinar 为主；
- 2) 活动由 CCF-CV 和企业联合主办，CCF-CV 秘书处负责指导和协调工作；
- 3) 活动尝试开拓不同地区的企业，对于每个地区，寻找合适的常委牵头组织，专委会进行协助；
- 4) 活动包括特定主题的报告、座谈、展示、体验、竞赛等不同形式；
- 5) 规模大小不限，人数不限，由举办活动的企业或高校自定活动场地、人数以及活动形式；
- 6) 活动原则上每季度不超过 2 次；
- 7) 活动举办方至少三周前提供材料，专委公众号、网站和简报配合发布活动信息及企业展示资料，帮助推广宣传；
- 8) 活动举办方录制活动现场音视频，活动后三天以内整理活动材料发送秘书处，秘书处审核后在专委公众号、网站发布；
- 9) 通过专委简报宣传企业信息，企业宣传内容由企业自主提供，采用简报模板编辑。

4. 注意事项

- 1) CCF-CV 企业交流活动以学术/技术交流为目的，严禁以活动名义进行非法集会，传播非学术内容；

2) CCF-CV 企业交流活动欢迎学生参加,可以增加有益于学生加深对企业了解的内容和环节,比如参观、体验,但在形式上应与企业招聘会区别开来;

3) 举办 CCF-CV 企业交流活动的企业,在专委奖项或专委承办的 PRCV 大会奖项的冠名时可以享受优先权;

4) 活动相关费用由承办方负担,包括活动场地、宣传等相关费用;如果邀请外部讲者,则需要承担讲者课时费、活动期间的本地交通和餐饮费,原则上讲者自己承担城市间往来交通、本地住宿费用;有条件的承办方,出于

自愿可额外承担讲者的这部分费用;

5) 活动举办方有义务配合讲者需要,出具官方的邀请信等证明材料,以便讲者报销差旅费用。

5. 活动申请

活动采取自愿申请的方式,有意承办活动的企业需提前至少两个月向秘书处提出申请,秘书处将在接到申请两周内给予答复并配合跟进。

责编黄岩

CCF-CV 成立专委顾问委员会

为了更好地发挥资深委员的作用,CCF-CV 专委会成立顾问委员会,旨在进一步推动专委会服务学术、联系业界等工作开展。顾问委员会委员可出席专委会常委会议、秘书处工作会议等,并对专委会的工作给予指导建议。

经过提名并报常委会审议,日前通过了顾问委员会名单如下:

主任:

谭铁牛 院士 中科院自动化所

委员:

陈熙霖 研究员 中科院计算所

贾云得 教授 北京理工大学

王蕴红 教授 北京航空航天大学

张艳宁 教授 西北工业大学

感谢顾问委员会委员对专委会发展作出的巨大贡献,同时也对他们的入选表示热烈祝贺!

责编黄岩

CCF-CV 秘书处工作会议圆满召开

中国计算机学会计算机视觉专委会
第三届常委会
第一次秘书处工作会议
2020年3月25日 在线会议



2020年3月25日，中国计算机学会计算机视觉第三届常委会（CCF-CV）第一次秘书处工作会议（2020年度第一次会议）在腾讯会议平台在线召开。专委秘书处全体成员参会，会议由秘书长马占宇教授主持，专委副主任王亮研究员全程参加并指导。

本次会议主要讨论如何落实专委会常务委员会第五次会议的各项决议，并制定秘书处下一阶段的工作计划。过去的2019年，专委各项学术活动和组织建设稳步推进，走进高校、走进企业、视界无限、专委简报与网站的活动形式与内容不断推陈出新，继续展示CCF-CV专委的组织活力与特色引领性。新一年度，各项活动如何更有实效地开展，如何更加广泛地激发委员们参与专委建设的积极性，如何办好专委年度学术大会PRCV2020与学术论坛RACV2020，大家围绕这些议题展开了热烈讨论，并形成了具体可行的执行方案。

PRCV2020的会议筹备工作已经启动，秘书处将积极配合组委会需要，利用专委会网站、微信公众号、简报等不同方式跟进宣传。专委会品牌活动RACV2020论坛继续定位为国内计算机视觉领域的小规模精品研讨会，拟定于暑期期间举行，由厦门大学承办。专委简报成立新一届编委会并全新改版。专委网站成立新一届网站运行团队并增加英文版、历任领导等版块，增加委员个人主页链接功能。专委拟建立委员贡献评价制度，对有突出贡献的志愿服务人员进行表彰和奖励。走进高校活动将延续过去4年多85期报告会的火爆，进一步创新活动形式，在活动申请、讲者征集、活动形式等方面将更为开放，加强邀请专家对承办单位的指导作用，提升活动对计算机视觉前沿研究的宣传力度。走进企业活动持续推进专委与业界领头企业的产学研交流合作，在学术成果落地与企业技术需求之间搭建更为便捷的桥梁。2019年启动的“视界无限”全新品牌活动已成功举办5期专题研讨会，后续将进一步围绕视觉研究前沿与热点问题提升活动质量与影响力，形成独具特色的品牌效应。

责任编辑 毋立芳

专题综述

食物识别研究进展

复旦大学 陈静静 香港城市大学 朱斌 杨宗桦

一、引言

随着我国经济的快速发展，人民的生活水平不断提高，饮食健康问题越来越受关注。据数据显示，全球每年有 1100 万人死于不健康饮食。因此，为了提高饮食健康状况，记录饮食，分析饮食营养状况是一项十分重要的研究任务。鉴于饮食健康管理的重要性，食物识别与营养分析的相关研究受到越来越多的关注。

作为健康管理的重要组成部分，食物识别的目的是帮助用户自动构建饮食日记，避免手动记录食物名字带来的麻烦。相比于通用图像识别，细粒度食物图像识别有其特有挑战。首先，由于切菜，烹饪方法的差异导致同类食物图片视觉上的差异很大。其次，同类食物的组成原料会随着地域，季节的改变而变化，使得很多食物没有固定的组成原料。图 1 展示了三种食物由于所用食材的差异造成的食物图片的类内视觉差异大的例子。类内视觉差异大使得细粒度食物图片识别成为一个比较难的问题。

早期的食物识别研究主要基于手工特征，如 SIFT, HoG 等，在这些特征的基础上再训练 SVM 分类器进行分类。由于食物图片类内差异大，手工特征难以取得满意的分类效果。近几年内，由于深度学习模型在图像识别上取得了巨大成功，近期的食物识别研究主要基于深度卷积网络。除了常见的用在通用物体分类上的深度卷积模型（如 VGG, ResNet）外，针对食物图像分类的深度卷积模型，例如，Wide slice Residual Networks^[1] 也被提出。目前，在包含上百个类的中等大小的数据集上，深度卷积神经网络能够取得高达 80% 的识别准确率。



图 1 所用食材差异造成的食物图片的类内视觉差异。第一排，第二排以及第三排图片分别属于“黄瓜炒鸡蛋”，“香辣藕丁”以及“手撕杏鲍菇”。

尽管深度卷积网络的出现使得食物图像识别的性能有了大幅度的提升，然而现有的基于卷积神经网络的食物识别模型离实际应用仍有一段距离。首先，现有食物识别模型都是在中小数据上训练得到，仅能识别百种食物图片，难以处理实际场景中上千万种类别的食物识别。其次，现有的食物识别模型依赖于大量手工标注的食物图片进行训练。由于手工标注成本高，这些识别模型的难以通过直接扩充训练数据集的方式提高模型的泛化能力。

为了解决实际应用中的大规模食物识别问题，现有的工作主要采用两种途径。一是对食物图片进行原料识别。食物由多种原料组成，与食物类别数量相比，原料种类的数量要少很多。因此，相比于食物类别识别，识

别食物原材料对于大规模的食物识别更加有效。另外一种途径是通过跨媒体菜谱检索进行食物识别。给定一个食物图片，通过跨媒体菜谱检索找到对应菜谱。由于菜谱包含食物名字，食材，甚至营养成分，因此，通过菜谱检索能够有效的进行食物识别以及营养分析。同时，由于网络上有大量的包含食物图片的菜谱，直接利用这些菜谱数据训练跨媒体检索模型能够节省人工标注，进而有效应对大规模食物图像识别问题。

二、研究现状

本章将从三个方面对现有食物识别工作进行介绍，数据集，食物识别以及跨媒体菜谱检索。

2.1. 数据集

食物数据集是食物识别的基础和先决条件。研究者们为此建立了大量的基准数据集，包含中国菜[2,3]、日本菜[4,5]和西式菜[6,7]等。中国自古以来就是对美食有极度追求的国家，也是食物种类、跨度最大的国家之一。[1]从众多的中文菜谱网站中爬取了大量的菜谱和食物图片数据，建立了首个大规模中国菜数据集：Vireo Food-172。该数据集总共包含 172 种不同类型的食物类型和 353 中不同的食材，同时图片数量达到 110,241 张。西式菜中属 Recipe1M+^[5]最大，包含百万菜谱和超过一千万的图片，但是需要注意的是，该数据集的标签并不是人工标注得到，而是通过对菜谱标题进行聚类自动生成。因此，该数据集提供的食物种类和食材的标注包含很多噪声。

2.2. 食物识别

现有工作对食物识别的研究主要包含两个部分：食物分类和食材识别。给定一张食物图片，食物分类模型需要能够判断是那种食物类型，是一个多类问题；而食材识别则更复杂些，需要识别图片中的食物是由哪些食材制作出来的，是一个多标签分类问题。[2]提出了不同的面向食物分类与食材识别的多任务学习模型。[2]中的结果显示，同时学习食物分类和食材识别能够比单任务学习食材识别好很多。值得一提的是，中国菜食物分类的 Top-5 准确率能够达到近 96%，食材识别也取得了不俗的表现。

众所周知，食材的切法和烹饪方法对食物最终食物的视觉效果产生巨大的影响，只考虑食材来做识别显然有其缺陷性。因此，[8]对食材的切法和烹饪方法进行了进一步的标注，通过引入食材切法以及烹饪方法为辅助任务进行多任务学习，使得食材识别的效果有了进一步的提升。[8]的模型结构如图 2 所示。此外，除了多任务学习，[8]为了处理由于相机到食物距离以及不同的切菜方法造成的食材大小差异问题，引入多尺度局部原料识别，实验证明该方法能够有效提高原料识别准确率。

2.3. 跨媒体菜谱检索

除了原料识别，应对大规模食物识别的另外一种有效方法为跨媒体菜谱检索。该方法不需要人工进行数据标注，因此近几年来吸引了不少研究者。相比与其他的跨媒体检索问题，菜谱检索有其独特的挑战性。首先菜谱是结构性文本，通常包含标题、食材和准备步骤三块信息。如何对菜谱的进行有效建模，从而学习到菜谱表征是一个难题。其次，菜谱步骤是教学性语言，食物图片是该教学语言产生的结果。然而，并不是每一个步骤都会对食物造成视觉上可见效果。因此，如何对菜谱步骤的视觉相除了原料识别，应对大规模食物识别的另外一种有效方法为跨媒体菜谱检索。该方法不需要人工进行数据标注，因此近几年来吸引了不少研究者。相比与其他的跨媒体检索问题，菜谱检索有其独特的挑战性。首先菜谱是结构性文本，通常包含标题、食材和准备步骤三块信息。如何对菜谱的进行有效建模，从而学习到菜谱表征是一个难题。其次，菜谱步骤是教学性语言，食物图片是该教学语言产生的结果。然而，并不是每一个步骤都会对食物造成视觉上可见效果。因此，如何对菜谱步骤的视觉相关性进行建模以建立起菜谱到实物图片的映射关系也是一个难题。

针对跨媒体菜谱检索，现有的方法大都是针对菜谱文本和食物图像学习一个联合嵌入空间，在该联合嵌入空间中计算图片与菜谱的相似度，以此进行菜谱检索。在学习联合嵌入空间的过程中，[9][10]引入注意力机制 (Attention Model) 使得学习到的联合嵌入空间更好，

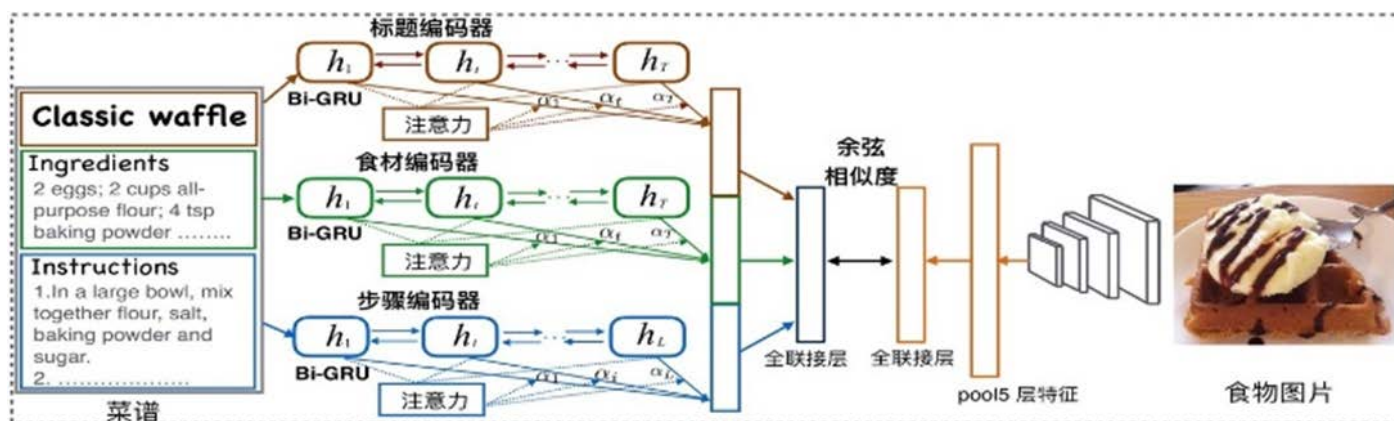


图 2 基于层次注意力机制的跨媒体菜谱检索模型

从而有效提高菜谱检索的准确率。为了对菜谱到食物图片的视觉相关性进行建模, [10]引入了层次注意力机制。如图 2 所示, 在菜谱表征学习中, 层次注意力机制会对每一个单词以及每一个句子学习到相应的权重。权重的大小表示该单词或句子的重要性。引入层次注意力能够帮助过滤掉一些视觉无关的句子或者单词, 使得学习到的联合嵌入空间的过程能够更好的进行菜谱检索。R2GAN^[11]探索了一种基于生成对抗网络(GAN)的检索模型, 通过专门设计的一个生成器两个识别器结构, 让检索模型能够学到更有效的特征表达。并且, 由菜谱特征重建出的食物图像能够协助用户更快速的找到相应的菜谱, 对检索中的排序结果也起到了解释的作用。

三、未来发展方向

尽管通过原料识别, 跨媒体菜谱检索能够一定程度上解决大规模食物识别问题。然而, 这些方法仍有一定的局限性。首先, 目前的识别模型难以应对跨地区识别以及检索。例如, 对于原料识别, 在中国菜上训练的模型难以应用到日本菜或者西方菜的原料识别上。同时, 在西式菜上训练的跨媒体检索模型也难以搬迁到中国菜的菜谱检索上。因此, 我们认为未来的发展方向包括利用迁移学习实现跨区域原料识别以及菜谱检索。此外, 小样本学习也是解决大规模食物识别的一个有效途径, 因此, 小样本食物识别也是一个值得研究的方向。

责任编辑 任桐炜

参考文献

- [1] Martinel, Niki, Gian Luca Foresti, and Christian Micheloni. "Wide-slice residual networks for food recognition." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.
- [2] Chen J, Ngo C W. Deep-based ingredient recognition for cooking recipe retrieval[C]//Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016: 32-41.
- [3] Chen X, Zhu Y, Zhou H, et al. ChineseFoodNet: A large-scale image dataset for Chinese food recognition[J]. arXiv preprint arXiv:1705.02743, 2017.
- [4] Harashima J, Someya Y, Kikuta Y. Cookpad image dataset: An image collection as infrastructure for food research[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2017: 1229-1232.
- [5] Horita D, Shimoda W, Yanai K. Unseen food creation by mixing existing food images with conditional styleGAN[C]//Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management. ACM, 2019: 19-24.
- [6] Marin J, Biswas A, Ofli F, et al. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images[J]. IEEE

- [7] Bossard L, Guillaumin M, Van Gool L. Food-101—mining discriminative components with random forests[C]//European Conference on Computer Vision. Springer, Cham, 2014: 446-461.
- [8] Chen J, Ngo C W, Chua T S. Cross-modal recipe retrieval with rich food attributes[C]//Proceedings of the 25th ACM international conference on Multimedia. ACM, 2017: 1771-1779.
- [9] Chen J J, Ngo C W, Feng F L, et al. Deep understanding of cooking procedure for cross-modal recipe retrieval[C]//2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018: 1020-1028.
- [10] Chen, Jing-Jing, et al. "Deep understanding of cooking procedure for cross-modal recipe retrieval." 2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018.
- [11] Zhu B, Ngo C W, Chen J, et al. R2GAN: Cross-Modal Recipe Retrieval With Generative Adversarial Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 11477-11486.



陈静静

复旦大学青年副研究员。主要研究方向为多媒体内容分析和计算机视觉。
Email: chenjingjing@fudan.edu.cn



朱斌

香港城市大学博士研究生。主要研究方向为食物识别和跨媒体检索。
Email: andrewzhu1216@gmail.com



杨宗桦

香港城市大学教授。主要研究方向为多媒体，计算机视觉，模式识别。
Email: cwngocs.cityu.edu.hk

热点追踪

多模态医学图像合成与重建

上海交通大学 王乾

医学图像在临床诊断和治疗中发挥着基础性作用。而如何将多模态图像提供的互补信息有效融合，是医学图像领域的经典科学问题之一。针对不同来源的图像数据，通过信号处理、模式识别等手段捕捉与生理变化或疾病进展紧密相关的影像特征，建立面向特定临床应用的数据模型，是多模态融合方法在与医学交叉研究中常采用的技术路线。

近期，基于深度学习的图像合成在医学图像领域大显身手。例如，可以由磁共振(MR)图像合成获得 CT 图像，或者在对应不同成像序列或场强的 MR 图像之间进行转化。图像合成的发展，深刻改变了医学图像多模态融合。长期以来，多模态融合主要依赖于图像配准，即建立多模态图像之间的相似性测度并进行优化，从而估计图像之间的空间变换。但是，不同模态图像视觉特征迥异，跨模态的相似性测度设计是众所周知的难点。我们通过图像合成实现不同模态图像之间的相互转化，进而将多模态图像之间的配准问题分解为(多个)与合成图像相关联的单模态配准问题，重塑了多模态融合的概念。

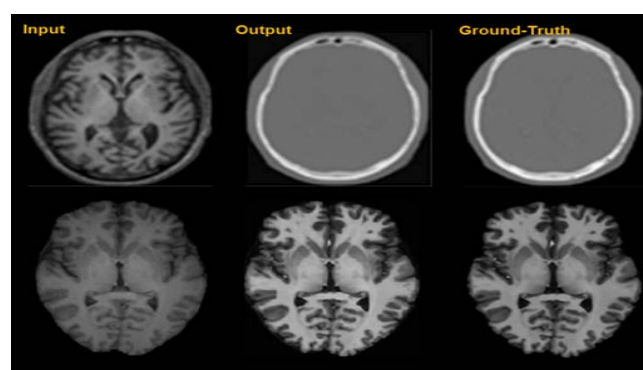


图 1 通过深度学习实现不同输入/输出之间快速合成

同时，深度学习推动多模态融合从传统图像后处理领域延伸至前端成像，奠定了图像智能重建基础。以 MR 图像中最常用 T1、T2 序列为例，我们通过深度学习融合不同序列图像互补视觉特征，将较快 T1 成像数据注入 T2 重建，使得在对 T2 数据 K 空间快速降采样的条件下，仍然可以重建出高质量 T2 图像。这些成果可以用于以新一代国产 PET-MR 设备为代表的高端医学影像设备中，从而推动我国医学影像装备产业链的发展。

以上工作已发表于国际会议 MICCAI 2017-2019。

责任编辑 王金甲



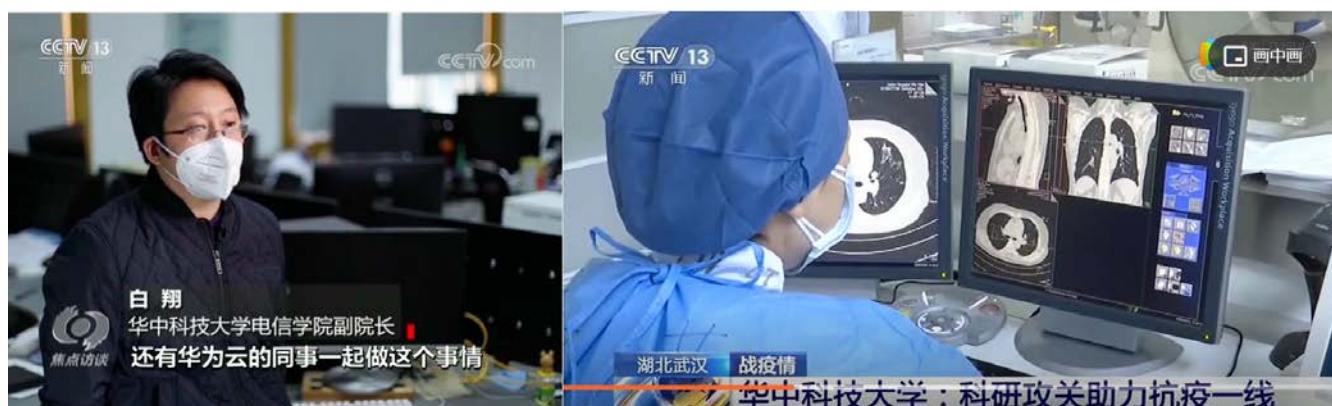
王乾

上海交通大学生物医学工程学院长聘副教授，医学影像先进技术研究院院长助理，主要研究方向为医学图像计算。

Email: wang.qian@sjtu.edu.cn

热点追踪

基于 AI 的新冠肺炎肺部 CT 影像量化分析



华中科技大学科研助力抗疫一线

计算机视觉专委会常委委员、华中科技大学白翔教授团队联合华中科技大学附属协和医院放射科、华为云、蓝网科技等共同研发出一套基于人工智能技术的面向新冠肺炎肺部 CT 影像量化分析系统，实现 CT 影像结果秒级输出。

该系统可自动分离出毫米级微小病灶，自动计算出肺部病变面积的百分比，实现单病例全自动精准量化结果的秒级输出。可以辅助医生更高效地区分新冠肺炎的早期、进展期与重症期，有利于早期筛查与防控。同时，对于确诊病人，基于对多次复查影像数据的量化分析，医生能够有效评估病情进展及用药疗效等情况，大幅提

升诊断效率，减轻医生诊断工作负荷。目前该系统已在全国几十家医院使用。

2月29日下午，中共中央政治局委员、国务院副总理孙春兰率中央督导组到华中科技大学考察疫情防控科研攻关情况，肯定了研究团队所开发系统的精确性。同时也非常关注该系统的落地应用情况，特别是在湖北省、武汉市的应用，鼓励团队结合 CT 影像与核酸试剂等信息进行更为精准的新冠肺炎诊断。3月3日央视新闻以“科研攻关助力抗疫一线”、3月25日《焦点访谈》以“救治在一线 科研当先锋”为题分别进行了报道。

责任编辑 储珺

华中科技大学新闻：

<http://news.hust.edu.cn/info/1002/38080.htm?from=groupmessage&isappinstalled=0>

“华为云和TA的朋友们”微信公众号：<https://mp.weixin.qq.com/s/GxZ1F2qug3vUk7hbChUkQw>

中央电视台官网：<http://tv.cctv.com/2020/03/25/VIDE5rVd8uYlksbfquXqbrov200325.shtml>

顶会观察

ACM Multimedia 2019

北京电子科技学院 金鑫

ACM International Conference on Multimedia (简称 ACM MM) 是中国计算机学会推荐的计算机图形学与多媒体领域的 A 类学术会议，一年举办一次。2019 年的 ACM MM 会议已于 2019 年 10 月 21 日至 25 日在法国尼斯举行。

一、多媒体会议主题的变化

多媒体会议主要由 Systems (系统)、Experience (体验)、Engagement (参与)、Understanding (理解) 四大主题组成。2019 年的多媒体会议给出了这四大主题更详细的诠释和更新：Multimedia Systems (多媒体系统)、Multimedia Experience (多媒体体验)、Engaging Users with Multimedia (用多媒体吸引用户)、Understanding Multimedia Content (理解多媒体内容)。

Multimedia Systems (多媒体系统) 今年包含的子主题为：Systems and middleware (系统与中间件)、Transport and delivery (传输与交付)、Data Systems Management and Indexing (数据系统管理与索引)，去掉了 2018 年的 Multimedia Telepresence and VR/AR (遥现与虚拟/增强现实)、Multimedia Scalability and Management (扩展与管理)、Mobile Multimedia (移动多媒体)。

Multimedia Experience (多媒体体验) 今年包含的子主题为：Interactions and Quality of Experience (交互和体验质量)、Art and Culture (艺术与文化)、Multimedia Applications (多媒体应用)，去掉了 2018

年的 Multimedia for Collaboration in Education & Distributed Environments (教学协作与分布式环境)、Music, Speech and Audio Processing in Multimedia (音乐、语音、声音)、Multimedia Authoring and Enrichment (多媒体创作与丰富化)，新添加了多媒体艺术与文化。

Engaging Users with Multimedia (用多媒体吸引用户) 主要包含：Emotional and social signals (情感与社会信号)、Multimedia Search and Recommendation (多媒体检索与推荐)、Summarization, Analytics, and Storytelling (摘要、分析与故事讲述)，去掉了 2018 年的 Social Multimedia (社交多媒体)。

Understanding Multimedia Content (理解多媒体内容) 今年全面改版，新增了：Multimodal Fusion and Embedding (多模态融合与嵌入)、Vision and Language (视觉与语言)、Media Interpretation (媒体解译)，去掉了 2018 年全部的三个子主题：Deep Learning for Multimedia (多媒体深度学习)、Multimodal Analysis and Description (多媒体分析与描述)、Multimedia and Vision (多媒体与视觉)。

笔者认为在深度学习技术火了几年之后，现在多媒体领域已经进入了后深度学习时代，更多的注意力转向多媒体问题本身的解决，而不是只关注特殊的方法，多模态融合尤其是视觉和语言的融合成为了新主题，笔者团队发表的论文^[1]正好就属于该新主题。

二、参会与论文录用情况

本次会议的参会人员共 750 名，来自 42 个国家，收到摘要投稿共 1275 篇，有效投稿 936 篇，共有 98 名领域主席以及 758 名审稿人，最终录用了 252 篇论文，录用率为 27%。

在被录用的论文中，占比最多的是 Media interpretation (媒体解译)，共 89 篇，占比 35.3%，其次是 Vision and Language (视觉与语言)，共 31 篇，占比 12.3%，排名第三的是 Multimodal Fusion & Embedding (多模态融合与嵌入)，共 25 篇，占比 9.3%。占比最少的主题是 Art and Culture，共 6 篇，占比 2.4%，其次的两个主题是 Data Systems Management and Indexing (数据系统管理与索引) 与 Summarization, Analytics, and Storytelling (摘要、分析与故事讲述)，各 7 篇，占比都是 2.8%。

与往年不同的是，今年第一次增加了一个代码可重复 Track: ACM Multimedia Reproducibility Track，收录的是 2018 年 ACM MM 录用论文中一些可重复代码的论文，以鼓励作者提供自己论文的相关代码，增加论文实验的可信度及对比实验的便利性。笔者认为这个学术氛围非常好，这两年在生物学领域出现了一些实验不可重复的丑闻，而在计算机领域，不应该出现这个现象，今年的这个 Track 为计算机领域的学术界注入了一股清风，为今后学术界的规范开了一个好头。

三、Keynote Speak

笔者今年听了两个主旨报告，第一个是来自法国国家视听研究所 (The French National Audiovisual Institute) 的 Jean Carrive，题目是：“Using Artificial Intelligence to Preserve Audiovisual Archives: New Horizons, More Questions”。法国国家视听研究所成立于 1974 年，是一个保存国家视听资料遗产的公共研究所，集科研、培训、教育、产品、咨询、编辑于一体的大型综合研究所。该研究所拥有长达 123 公里的模拟视听材料，大于 1900 万小时的电视与广播，预计在 2021 年能够完成 100% 的数字化。该研究所正在利用语音翻译、图像识别、人脸和人声识别等人工智能技术

建立视听资料的索引，同时指出了在如此庞大的视听资料中进行人工智能分析面临的两大困难：(1) 没有现成的数据集可以用于训练，(2) 资料中概念和词汇过于复杂，期望能做成一个公开项目，以得到学术界更多的技术支持。

笔者今年听的第二个主旨报告是来自比利时布鲁塞尔自由大学 (Free University Brussels) 和荷兰拉德堡德大学 (Radboud University) 的 Mireille Hidebrandt 教授，题目是：“EU Data Protection Law: An Ally for Scientific Integrity?” Hidebrandt 教授介绍了欧洲的 GDPR 组织 (General Data Protection Regulation)，是一个欧洲专门从事数据保护的公共组织，并且介绍了欧洲数据保护法的相关情况。

四、Best Paper Session

与计算机视觉的会议不同，ACM MM 会议的 Best Paper Session 需要进行现场的报告和投票，因此笔者听到了 5 篇最佳论文提名的精彩报告，分别来自中科院自动化所^[2]、山东大学^[3]、美国约翰霍普金斯大学/Snap 研究院^[4]、新加坡国立大学^[5]、山东师范大学^[6]。内容涉及基于多模态知识网络的医学问答^[2]、多模态对话系统^[3]、视听缩放^[4]、隐私保护^[5]、基于多模态哈希的大规模多媒体检索^[6]。最终，美国约翰霍普金斯大学/Snap 研究院的 “Audiovisual Zooming: What You See is What You Hear” 胜出获得了最佳论文奖，在该篇论文中，作者创造性的提出了 Audiovisual Zooming 问题，能够在视频进行缩放的同时，缩放画面所对应的声音，用户可以只听见所看见的人或物体。

五、Brave New Ideas

这个 Session 是笔者所感兴趣的版块之一，共包含了四篇论文，分别来自中国人民大学^[7]、荷兰阿姆斯特丹大学^[8]、德国马克思普朗克研究所^[9]、新加坡国立大学^[10]。内容涉及故事可视化^[7]、多模态数据超学习 (HyperLearn)^[8]、基于人眼聚散行为的内部思维检测^[9]、图像主观属性学习^[10]，其中中国人民大学的 Shizhe Chen 以画饼充饥这个故事所对应的图像序列生成为例子，生动的讲解了内容涉及故事可视化方法。

六、总结与展望

ACM MM 会议还有多个丰富多彩的 sessions, 例如 workshops、Art & Culture、艺术展、Best Demo、Open Source Competition 等等, 会议时间有限, 笔者并没有在上文中介绍全部的内容, 笔者的理解仅代表个人观点, 如有偏差还请各位读者指出。在最后的 SIGMM (国际多媒体特别专家组) 会议上, 组委会介绍

了 2020 年的一些多媒体相关会议安排, 包括 ICMR、ACM IH & MMsec、ACM Multimedia Asia (由 PCM 和 ICIMCS 合并) 等。2020 年的 ACM MM 会议将在 2020 年 10 月 12 日至 16 日在美国西雅图举行, 期待与 MM 的下一次相遇。

责编委 王金甲

参考文献

- [1] Jin X, Wu L, Zhao G, et al. Aesthetic Attributes Assessment of Images. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [2] Zhang Y, Qian S, Fang Q, et al. Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable Medical Question Answering [C]. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [3] Nie L, Wang W, Hong R, et al. Multimodal Dialog System: Generating Responses via Adaptive Decoders. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [4] Nair A, Reiter A, Zheng C, et al. Audiovisual Zooming: What You See is What You Hear. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [5] Shen Z, Fan S, Wong Y, et al. Human-imperceptible Privacy Protection Against Machines. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [6] Lu X, Zhu L, Cheng Z, et al. Flexible Online Multi-modal Hashing for Large-scale Multimedia Retrieval. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [7] Chen S, Liu B, Fu J, et al. Neural Storyboard Artist: Visualizing Stories with Coherent Image Sequences. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [8] Ayra D, Rudinac S, Worring M. HyperLearn: A Distributed Approach for Representation Learning in Datasets with Many Modalities. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [9] Huang X, Li J, Ngai G, et al. Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [10] Gelli F, Uricchio T, He X, et al. Learning Subjective Attributes of Images from Auxiliary Sources. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.



金鑫

北京电子科技学院副教授, 主要研究方向为计算美学与多媒体信息处理, 主页: jinxin.me。
Email: jinxin@besti.edu.cn

北京航空航天大学李甲教授访谈

2020年2月15日,《CCF-CV专委简报》在线采访了2019年国家优秀青年科学基金获得者、北京航空航天大学博士生导师李甲教授。下面是采访实录。

李老师,您好!首先,请您介绍一下您的个人学习和研究经历。

我是2001年从山东考到北京上大学,先后在清华、中科院计算所、新加坡南洋理工、北大、北航等五家高校和科研机构学习和工作,中间还去盛大公司创新院做了一年多的“盛斗士”,体验了工业界的生活。在多个单位的学习和工作经历,使我得到了很多锻炼的机会,也促使我习惯于从多个不同的视角思考问题。

您在视觉注意、显著性、图像视频计算与解析方面有很多的研究成果,能否跟大家分析一下这个领域目前的研究现状和未来发展趋势?

我从06年开始接触视觉注意和显著性计算这个领域,当时大家期望通过模拟人类视觉认知机理来突破部分计算机视觉任务的瓶颈,涌现了很多对眼动预测和显著物体分割的研究。但12年后深度学习的爆发,使得大家可以通过端到端的方式绕过对视觉认知机理的简单模拟,因此这个领域的发展逐渐出现了分化。

一方面,借助眼动仪开展眼动和视觉注意预测的研究有所减少,最主要的问题是说不明白通过这种方式预测出的视觉注意有什么用,而对视觉注意本身的研究也在逐渐开始探索人类视觉认知的底层生理机理。另一方

面,得益于大量的代码和数据集开源,显著物体分割目前成为了研究热点,目前的趋势包括研究弱监督和无监督环境下的显著物体分割和共分割。这个领域现在的问题在于,缺少“非它不可”的杀手级应用。

除了注意力和视觉显著以外,我们也在积极探索新的道路,尝试了很多图像视频计算与解析的新任务。目前来看,“精益求精”和“见微知著”是两条可行的发展道路。其中“精益求精”是指视觉认知粒度越来越细,细到样例级和部件级,“见微知著”是视觉认知既要适应充满噪声的真实大数据环境,又要理解万事万物的细微差别。这也是我们课题组未来将重点关注的两个方向。

您的所有研究中,您认为最有应用价值、最有前途的一项研究成果是什么,能否跟大家分享一下?

我们近期取得了一些与应用场景结合比较紧密的研究成果,包括近似车辆重识别、全景图像拼接、互联网图像视频分类、机载目标检测等,每项成果都或多或少的具有一定应用价值和发展前景。其中,与无人机应用场景相关的一些工作可能具有更好的应用价值和发展前景,因为这些工作既能体现出北航的航空航天特色,又能增强我们与其他课题组工作的区分度。

您申请了多项国内和国际专利,能否跟大家分享一下您对申请专利的认识和建议?

我在读博期间就养成了申请专利的习惯。当时高文老师团队对申请专利的要求很高,需要经过多位老师会

议评估创新性和实用性才允许申请。因此当时申请的专利多数是以论文算法为基础进一步开发的应用系统，比如视频广告插入系统等，其中少数与华为公司合作研发的系统申请了国际专利。后来到北航之后，赵沁平老师团队非常鼓励将高水平工作申请专利，因此我们课题组内部制定了一个标准：值得投稿高水平刊物和会议的工作可以申请国内专利，其中部分开创性的高质量工作可以申请国际专利。个人认为，申请专利要变成一种科学研究的习惯动作，不要仅仅只对少数重点工作申请专利，而是要在自己的主战场上，对高水平的工作持续申请专利，形成系列的专利壁垒，方便在合适时机开展转化。

您曾于新加坡南洋理工大学、盛大创新院、北京大学等机构从事科研工作，请问最后为什么选择了北京航空航天大学呢？能否介绍下您作为“卓越百人”被引进到北京航空航天大学工作的经历？

选择北航的两个主要原因是：学校好，上升趋势明显，青蝇之飞，不过数武，附之骥尾，可致千里；离家近，只隔一条马路，以每天通勤节省1小时计，每年将多出1.5个月的8小时工作日。

做为卓越百人副教授进入北航后，有三件事情对我触动很大。第一，14年刚进校时，怀校长安排了一个项目申请工作，计算机学院抽调多名教师在会议室连续奋战24小时，从第一天中午12点干到第二天中午12点，从无到有地完成了申请书初稿和PPT初稿。并且，夜里2点时，学院领导给大家送来了KFC，说是刚开完一个会议，陪大家工作一小时，然后稍微休息下再去参加早上8点的另一个会议。第二，15年初参与重点项目申请，给一位青年老师和一位博士生打电话，问可否来开会讨论，答曰“刚到食堂打了饭”，于是就没有催促，没想到十分钟后两人就出现了，说是听到会议通知，没有吃饭就跑步回来。第三，17年给一位学生通过QQ发了个信息“在么”，发完后发现忙得忘记了时间，已经接近晚上12点。本来没期望得到回复，没想到学生秒回三句话：“我在”，“刚回到寝室，什么事”，“如果需要，我马上赶回来”。从这三件事情，才切身感受到了北航的

魅力与特色，实实在在的归心于北航，认可自己是一名北航人。这种归属感，也是我从工业界返回学术界、从国外返回国内、从多个科研机构中选择北航时的最大收获。

您于2018年12月曾获得“北京航空航天大学优秀教学奖二等奖”，请问能否跟大家分享一下您的获奖经历，以及您的教学经验？

首先声明，这个奖并不难，我们学院的老师在2018年获得了一个一等奖和四个二等奖。不过当时我只是第二年从事教学工作，能获奖还是有点意外。究其原因，还是在于“感同身受”四个字。我在上课时，经常会想，如果是我坐在下面，凭什么不让我玩手机而是去听课呢？答案很简单，把课程讲得比微信朋友圈的段子更有趣，比王者荣耀参与度更强，比网络小说的包袱更多，自然就能吸引学生。当然，做到这几点很难，我也在不停的尝试和改进。另外，还有很重要的一点是，把每个听课的学生，都当成自己“名下”的学生，希望他们能真的学到东西，而不是走形式凑课时，学生自然能感受得到，也会给出相应的反馈。由于我还是教学领域的新兵，很多事情还在摸索，这些经验未必正确，仅供大家参考。

您于2018年6月曾获得“图像识别挑战赛WebVision2018冠军”，能跟大家分享一下您的挑战历程么？

当时我们和百度联合组队参加了这项比赛，其目的是尝试一种新的校企合作模式。大致流程是，鼓励完成基础科研训练（例如：投出一篇CCF A类论文）的优秀硕士生做为实习生，安排到企业参加比赛，并且做为高校团队和企业团队相互沟通的桥梁。参加比赛时，由企业团队负责提供大规模GPU资源和工程调参，高校团队辅助提出算法设计和技术路线建议；比赛后，高校团队负责撰写论文和投稿，企业团队辅助提供补充实验。打完比赛后，优秀的硕士生就地转正至企业工作，实现“三赢”。WebVision 2018时，我们使用这种策略获得了冠军，发表了一篇IEEE TPAMI论文，参与比赛的硕

士生获评为北京市优秀毕业生，并为百度提供了一名已经熟悉了相关工作内容的高素质员工。

能否介绍一下您现在的研究团队，您团队目前研究工作的重点和特色，以及您对团队的管理和规划？

课题组现在除了我之外，还有 1 名助理研究员、6 名博士生、8 名硕士生，以及多名已保研直博和硕士的本科实习生。课题组目前正处于转型期，从早期以视觉显著为单一研究重心，转向检测、分割、识别三路齐头并进。近期的主要发展思路是：维持视觉显著计算，深耕物体精细识别，开拓微小目标检测，强化航空航天应用。从长远来看，课题组希望在当前扩张的研究领域基础上，凝聚出 1-2 个具有较高创新潜力和应用价值、能够支撑重要行业领域发展的研究方向，并引进具有共同学术理想和共同发展目标的青年教师组成学术团队，通过开展“更高端”、“更大气”、“更有用”的研究，使课题组能够逐步发展为有个性、有能力、有担当的国际高水平学术团队。

您在 TPAMI、CVPR 等国际知名期刊/会议上发表了很多高水平的论文，能否分享一下您发表高水平论文的经验？

课题组的论文发表主要包括两个阶段。第一阶段是 2017 年及以前，以我为主力输出，既要自己做实验和写论文，又要帮学生想 idea，设计实验和修改/重写论文。这一阶段主要经验为：瞄准顶级，多写多练，唯手熟尔。第二阶段是 2018 年起至今，以高标准重点培养部分有潜质学生的基础科研能力，并辐射带动所有学生

的科研工作。对每个学生，制定高水平的投稿目标和有潜力的研究方向，明确时间结点，定期检查进展。在学生开展研究时，具体论文选题、实验设置、论文撰写等以学生为主，并且明确导师给出的是仅供参考的 advice，而非强制性的 supervision。近两年的实践表明，通过充分调动学生的主观能动性，鼓励学生主动开展探索和积极试错，保障好学生日常生活和开展实验的软硬件环境，能够很好地促进高水平论文的发表。当然，这些经验仅为一家之言，是否普适仍需要继续进行测试。

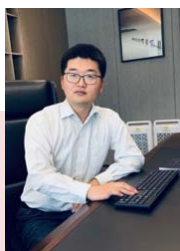
请问您是如何处理“科研与教学”、“工作与家庭”这些复杂关系的？

这些关系之所以显得复杂，主要是因为什么都想要，什么都急着要。以科研与教学为例，我觉得自己更擅长科研，那么就专心把最擅长的事情做好，努力把不擅长的事情做完。工作与家庭亦然，一方面分配的精力多些，另一方面就需要别人帮你分担份外的责任，比如工作方面的其他青年教师和高年级博士生，以及家庭方面的配偶和父母。为此，需要维持时间和精力分配的动态平衡，如需取舍，可致良知。

如果吐露研究者心声，您最想说的是什么？

年方少仗剑入京，十九载诸事未成。
庸庸碌碌行矩步，战战兢兢履薄冰。
心茫然拔剑四顾，忆往昔岁月峥嵘。
重收拾竹杖芒鞋，愿他日手挽天倾。

责任编辑 余焯 赵振兵



李甲

博士，教授，博士生导师，中国民主同盟盟员，国家自然科学基金优秀青年基金获得者。2005年7月于清华大学电子工程系取得学士学位，2011年1月于中国科学院计算技术研究所取得博士学位。此后，曾于新加坡南洋理工大学、北京大学、盛大创新院等机构从事科研工作。2014年6月被聘为北京航空航天大学计算机学院副教授，现任北京航空航天大学计算机学院教授。目前是 IEEE/CCF/CIE 高级会员和 ACM 会员，担任第十二届民盟中央青年工作委员会委员、中国计算机学会的计算机视觉专委会委员、虚拟现实与可视化技术专委会委员，以及中国图象图形学学会视觉大数据专委会委员，兼任国际期刊 IET Computer Vision 编委，曾任 Frontier of Computer Science 青年编委。研究领域包括计算机视觉、虚拟现实和人工智能技术，主要研究兴趣为视觉场景智能理解与内容生成。先后主持国家自然科学基金面上项目、国家 863 计划课题（军口部分）、国防科技创新特区课题等纵向与横向项目 12 项；参与国家自然科学基金重点项目（排名 2）、国家 242 信息安全计划课题（排名 2，副组长）等纵、横向项目 10 项。在 IEEE TPAMI/TIP、IJCV 等著名国际期刊以及 ICCV、CVPR 等著名国际会议发表论文 74 篇，包括中国计算机学会推荐的 A 类论文 33 篇；中国美国专利申请受理 57 项（美国 10 项），授权 32 项（美国 8 项）；出版 Springer 英文学术专著 1 部（第一作者）。2019 年获国家自然科学基金优秀青年科学基金资助，2017 年入选北京市科技新星计划，2016 年入选北航“校青年拔尖人才支持计划”，2014 年入选北航“卓越百人计划”，2012 年分别获评为北京市优秀博士学位论文和中国科学院优秀博士学位论文；获 2018 年中国电子学会自然科学二等奖（排名 1）、2010 年度教育部科学技术进步奖一等奖（排名 9）、IEEE 国际会议 BigMM 2015 唯一最佳论文奖、图像识别国际挑战赛 WebVision 2018 冠军。

华南理工大学许勇教授访谈

2020年3月10日,《CCF-CV专委简报》在线采访了华南理工大学“视觉与服务计算”学术团队带头人,博士生导师许勇教授。下面是采访实录。

许老师,您好!首先,请您介绍一下您的个人学习和研究经历。

我从本科到博士都是在南京大学度过的,当时的专业是数学。1999年博士毕业之后来到华南理工大学,先后从事计算机网络、计算机视觉和模式识别、大数据等科研与应用工作,如今已经在华南理工大学工作二十余年了。

您在人工智能和大数据方面有很深入的研究,能否介绍一下你们的研究情况、相关研究成果,以及这个领域目前的研究现状和未来发展趋势?

近年来我们团队的主要研究内容是稀疏表达、深度学习相关理论、技术及应用,取得了一些研究成果,相关学术论文发表在高水平期刊和学术会议上。我们团队注重产学研用,团队和多家知名企业成立了联合实验室,团队科研成果在合作企业取得了良好的效果。目前深度学习的效果比较好,但不足也非常明显,如依赖大规模标注训练集及大量计算资源等。以深度学习为代表的人工智能下一步可能会往理论上的解释性和可靠性、应用场景上的精细化和多样化发展。类脑计算等新的计算模型是未来人工智能发展的重要方向。

在科研方面,您获得了广东省科技进步二等奖等众多奖项,能跟大家分享一下令您印象最深刻的一次获奖经历,及获奖的相关科研成果么?

我们团队的研究成果涉及去雨、去噪、去反光等底层视觉任务和纹理识别、动作识别、物体跟踪、自动图像描述等高层视觉任务,这些成果在人机交互、公共安全等领域具有重要的科学价值和广泛的应用前景。通过和产业界的合作,我们将这些成果应用到基于虚拟机器人的政务服务平台,实现了政务服务平台的智慧化赋能和产业化应用,为企业和个人用户提供了高效率、智能化和个性化的政务服务体系,并取得了显著的社会经济效益。广东省科技进步奖是对我们这些工作的重要肯定,

您有去很多国外大学进行过学术访问,能跟大家分享一下您的访学经历,及您对不同国家不同大学学术环境的看法么?

也增强了我们团队产学研用结合的信心和决心。

我04-05年在美国马里兰大学学术访问,当时国内的学术环境和国外名校有一定差距,特别是美国高校的研究工作原创性高,美国访学的研究经历对我以后的科研发展影响很大。个人感觉欧洲部分大学的科研激情和创新性相对弱一些。近年来,我们国家的学术环境在不断完善,政府出台了一系列政策来规范人才培养和科研管理,逐渐和国际接轨,相信国内学术环境会越来越越好。

您不仅是 CCF-CV 的专委，还是 CCF 广州分部主席，更是中国计算机学会 CCF 杰出会员，能介绍一下您作为 CCF 广州分部主席，在组织 CCF 活动时的一些主旨么？您认为 CCF 对学术、企业界最大的贡献在哪里？

CCF 是国内计算机领域最大的学术团体，我们的主旨是把 CCF 打造成一个学术界和产业界等所有计算机领域同行之间相互交流、相互促进的平台。CCF 广州分部希望通过举办各种线上线下的学术研讨会、参观交流会等，邀请来自高校和研究所的科研工作者分享最前沿的学术成果，也请在产业界的同行来分享工业上的需求和技术难点，以促进产学研结合。CCF 广州分部做了很多努力，得到了大家的认可，也荣幸地被评为 2019 年度全国优秀分部第一名。

能介绍一下华南理工大学“视觉与服务计算”学术团队的研究情况么？

我们团队是一个跨多学科的科研团队，专注于图像/视频处理与增强、图像/视频的不变性特征、图像/视频的识别和分类、图像/视频的质量评估、稀疏表示及其应用、深度学习及应用、半监督/弱监督学习、大数据技术及应用等研究领域。团队中还有 4 名副教授，8 名博士研究生，已在高水平期刊和学术会议发表论文一百多篇。

作为华南理工大学“视觉与服务计算”学术团队带头人，能介绍一下您是如何带领及管理此团队的么？

我们学术团队几位教授之间分工明确，我们都有比较一致的目标，一是进行国际前沿的研究、同时注重产学研结合；二是培养具有科学素养、专业能力强、有责任心有担当的高水平人才。除了频繁的工作学习进展交流以外，我们团队师生经常组织体育活动，定期开展轻松愉快的聚会，彼此之间交流比较充分，氛围融洽。

您作为华南理工大学研究生院副院长，能对目前研究生的现状进行一些分析，并分享一下您的看法么？

近年来，我国研究生培养取得了很大的成效，为各行各业培养了大批高水平人才，促进了我国社会、经济和科学发展。但目前研究生的培养工作中还存在一些不足，就我的实际工作中我认为比较突出的两点，一是目前研究生的教学内容、教学方法和实践能力的培养和社会的实际需求还存在一定差距，另一方面是优秀博士生生源不足。在实践能力培养方面，我们积极地和企业建立产学研联合培养基地，争取让研究生尽早进入产业界进行实践和锻炼。在吸引优秀博士生生源方面，我们学校设立了新的博士生奖助金体系，加大硕博贯通式的人

请问下您是怎么管理自己的研究生的？您对他们的要求是什么？

才选拔力度，吸引更多的优秀人才攻读博士学位。

我比较注重因材施教，根据学生自己的职业规划和能力所长来制定他们的培养计划。对于硕士生，我希望他们能紧跟人工智能和大数据的技术前沿，侧重培养提高他们的应用和工程能力。对于博士生，我鼓励他们去开展原创性高、理论性强的科学研究，力争高水平的研究成果。此外，我要求学生平时进行体育锻炼，保持身体健康。

如果吐露研究工作者的心声，您最想说的是么？

希望科研工作者在宽松的科研环境中，专注于开展原创性研究，挑战计算机领域核心问题，共同推进我国科技创新和产业发展。

责任编辑 余焯 赵振兵



许勇

教授，博士生导师，现任 CCF 广州分部主席，华南理工大学研究生院副院长，“广东省大数据分析处理工程技术研究中心”主任。1993、1996 和 1999 年分获南京大学学士、硕士和博士学位，1999 年至 2001 年在华南理工大学从事博士后研究，2001 年至今在华南理工大学工作。历任计算机科学与工程学院系主任、副院长。教育部“新世纪优秀人才”，国际 IEEE 高级会员，ACM 高级会员，中国计算机学会 CCF 杰出会员，CCF 计算机视觉、人工智能与模式识别专委会委员，广东省“千百十”省级培养对象，华南理工大学“视觉与服务计算”学术团队带头人。主要研究方向是：人工智能和大数据等。近年来，主持了 27 项国家自然科学基金和省部级科研项目，发表科研论文 124 篇。获广东省科技进步二等奖，广东省计算机学会科学技术奖一等奖及广东省计算机学会优秀论文奖一等奖等奖项。

委员好消息

✦ 2019 年 11 月 11 日，CNCC2019 优秀论坛评选结果揭晓，CCF-CV 专委会顾委委员、中科院计算所陈熙霖研究员等组织的论坛“计算机视觉技术赋能智慧城市”获评优秀论坛。

✦ 2019 年 11 月 27 日，IEEE Fellow 2020 名单揭晓，今年一共新增 282 位，其中华人 Fellow 83 位（占比 23%），中国大陆地区 31 位入围。CCF-CV 专委会顾委委员、北京航空航天大学王蕴红教授因在虹膜与人脸识别领域的贡献入选。

✦ 2019 年 12 月 2 日，中国图象图形学学会首批会士名单公布，CCF-CV 专委会顾委主任、中科院自动化所谭铁牛院士和 CCF-CV 专委会常务委员、中山大学赖剑煌教授入选。

✦ 2019 年 12 月 3 日，CCF-CV 专委会委员、北京邮电大学李春光副教授指导的论文 Local Convex Representation with Pruning for Manifold Clustering 在 VCIP 2019 上获最佳学生论文奖。

✦ 2019 年 12 月 3 日，2019 中国自动化学会科普奖评审结果公布，CCF-CV 专委会委员、复旦大学张军平

教授的《爱犯错的智能体》一书榜上有名，荣获 2019 中国自动化学会科普奖。

✦ 2019 年 12 月 9 日，中国图象图形学学会 2019 年度科学技术奖评选结果公布，共 4 个项目获奖。由 CCF-CV 专委会委员、中山大学林惊教授和李冠彬副教授等完成的“认知启发的视觉语义理解及应用”、由 CCF-CV 专委会常务委员、南开大学程明明等完成的“弱监督条件下的图像语义分割的研究”获一等奖。由 CCF-CV 专委会委员复旦大学颜波教授等完成的“大数据环境下的网络媒体智能分享理论与应用研究”、CCF-CV 专委会委员中科院信工所张晓宇副研究员等完成的“基于协同融合的媒体智能感知”获二等奖。

✦ 2019 年 12 月 11 日，ACM 公布 2019 年最新当选的 ACM Fellow 名单，一共 58 位科学家当选（包括 7 位华人），他们在人工智能、云计算、打击网络犯罪、量子计算和无线网络等领域做出了广泛而重要的贡献。CCF-CV 专委会顾委委员、中科院计算所陈熙霖研究员因在人脸识别、手语识别和多媒体系统领域的突出贡献而当选。ACM 现有会员近 10 万人，ACM Fellow 是一

个精英团体,在ACM的全球会员中所占比例不足1%。

✪ 2019年12月23日,中国图象图形学学会发布了2019年石青云女科学家奖评选结果,共5人入选,CCF-CV专委会委员、北京航空航天大学刘偲教授获青年组石青云女科学家奖。

✪ 2019年12月23日,中国计算机学会公布了2019年CCF优秀博士学位论文奖评选结果,共10位博士获奖,另有4位博士获得提名。CCF-CV专委会委员、中科院信工所操晓春研究员指导的博士生任文琦获得了2019年CCF优秀博士学位论文奖,其博士论文题目为《融合结构先验的图像及视频去模糊研究》。

✪ 2020年1月6日,CCF公布了2019年度杰出演讲者名单。CCF-CV专委会3位委员上榜,他们是:清华大学鲁继文副教授、中国科学院大学吕科教授、北京大学彭宇新教授。

✪ 2020年1月14日,北京智源人工智能研究院公示了2019年度智源青年科学家候选人名单,CCF-CV专委会4位委员入选:中科院计算所王瑞平研究员、清华大学鲁继文副教授、北京大学施柏鑫研究员和北京邮电大学马占宇教授。

✪ 2020年3月8日,由北京智源人工智能研究院、清华大学-中国工程院知识智能联合研究中心、清华大学人工智能研究院,共同发布了Women in AI 2020榜单,共179人上榜,CCF-CV专委会委员、北京航空航天大学刘偲副教授入选。

✪ 2020年3月25日,广东省科技创新大会上颁发了2019年度广东省科学技术奖。由CCF-CV专委会委员、中科院深圳先进院乔宇研究员联合商汤科技、南京大学研发完成的“视频的深度表征与识别技术及应用”项目荣获技术发明一等奖。由CCF-CV专委会委员、华南理工大学许勇教授等完成的“基于虚拟机器人的政务服务平台及应用”项目荣获科技进步二等奖。

✪ 2020年3月27日,2019年度北京市科学技术奖评审结果公示,CCF-CV专委会委员、北京科技大学殷绪成教授牵头完成的“网络图像视频大数据的智能识别关键技术及应用”项目荣获科技进步一等奖,CCF-CV专委会委员、北京理工大学刘越教授参与完成的“虚拟现实与增强现实头戴显示关键技术及应用”项目荣获技术发明一等奖。

责任编辑 刘海波

场景图生成开源代码

西安电子科技大学 张亮 姜有亮

场景图指用图的形式对非结构化场景进行结构化表示的一种形式，具体是指利用图像或者视频信息建立以物体为节点、物体之间关系为边的图表示，其中物体可带有类别等属性，边表示物体之间的空间物理关系。场景图是机器人等智能体实现高级别推理任务的知识基础。

场景图生成任务的核心是检测视觉关系，从而形成一个三元组集合。视觉关系检测的方法主要可以分为消息传递、注意力机制、视觉翻译嵌入和语言先验等，本文从这几种方法切入介绍几个关系检测和场景图生成的开源代码以及数据集。

1、Scene Graph Generation by Iterative Message Passing

工作：该论文采用消息传递机制，通过分析场景图的结构特点，将物体（结点）和关系（边）分别构建在一对对偶的子图中，如图 1，对于关系，将其 subject 和 object 的特征（从 GRU 的隐藏状态得到），分别与该 edge 的 GRU 当前迭代的输出拼接作为新的特征送入下一次迭代。

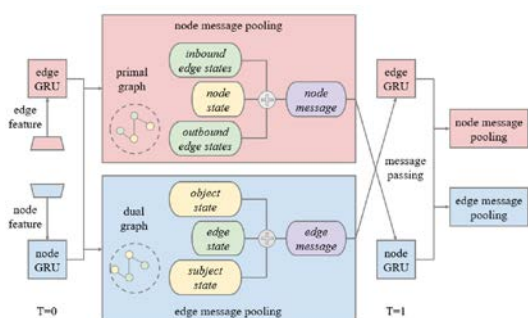


图 1 迭代的消息传递机制

论文：<https://arxiv.org/abs/1701.02426>

代码：<https://github.com/danfeiX/scene-graph-TF-release>

2、Visual Translation Embedding Network for Visual Relation Detection

工作：该论文采用视觉嵌入的方法，将 object 特征转换到一个低维的 relation 空间，在这个空间里，relation 可以被建模成简单的转换向量（translation vector），比如： $\text{subject} + \text{predicate} \approx \text{object}$ 。VTransE 结构的特征提取模块如图 2。

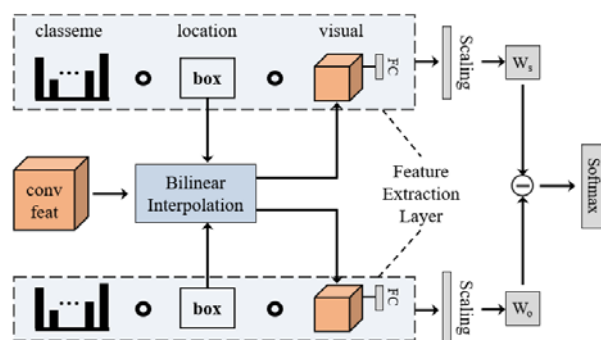


图 2 VTransE 结构的特征提取模块

论文：<https://arxiv.org/abs/1702.08319v1>

代码：<https://github.com/YANYANYEAH/vtranse>

3、Graph R-CNN for Scene Graph Generation

工作：该论文基于图卷积网络建模，并引入注意力机制进行关系检测，给定一张图像，首先抓取出一组可见的物体，假定所有物体两两间都有联系，接着用学习出来

的“relatedness”将那些不太可能的关系去掉，使得 graph 变得稀疏一些。最后，一个图卷积网络被用来整合全局上下文信息，并更新物体和关系结点的标签。图 3 给出了带注意力的图卷积网络的结点特征更新方式。

论文: <https://arxiv.org/abs/1808.00191>

代码: <https://github.com/jwyang/graph-rcnn.pytorch>

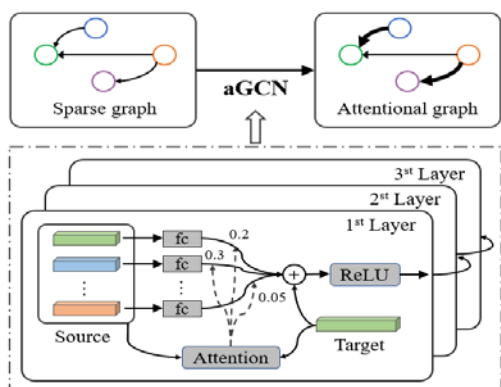


图 3 Graph R-CNN 框架的图卷积架构

4、Scene Graph Generation with External Knowledge and Image Reconstruction

工作: 为了解决现有可视化关系数据集的关系分布偏差, 该论文提出了一种新的特征细化网络, 如图 4 所示, 利用外部知识库中的常识关系进一步改进特征表示, 并使用基于注意力的 GRU 对物体特征进行迭代更新。

论文: <https://arxiv.org/abs/1904.00560>

代码: https://github.com/arxrean/SGG_Ex_RC



张亮

博士生导师, 西安电子科技大学计算机科学与技术学院从事教学与科研工作, 嵌入式技术与视觉处理研究中心主任。重点关注机器人复杂场景中环境感知与理解技术、手势/识别、图网络等深度学习热点研究方向。个人主页: <http://web.xidian.edu.cn/zhangliang/>



姜有亮

硕士研究生, 西安电子科技大学计算机科学与技术学院, 研究方向为计算机视觉。

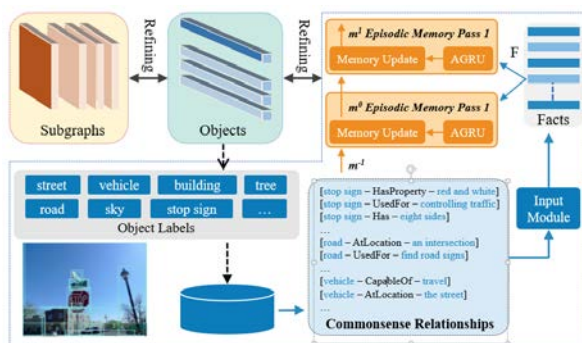


图 4 基于外部知识库的特征增强模块

开源的关系检测数据集有 VRD, Visual Genome 等, 在表 1 中给出了常见的 5 个数据集的统计信息。

表 1. 典型开源的关系检测数据集*

数据集	物体	关系	三元组	图像
Visual Phrase	8	9	1796	2769
Scene Graph	266	68	109535	5000
VRD	100	70	37993	5000
Visual Genome	33877	40480	2347187	108007
VrR-VG	1600	117	203375	58983

*注: 物体/关系分别表示对应数据集中物体的类别和关系类别数目, 三元组表示对应数据集中《主-谓-宾》三元组的数目。

责任编辑 樊鑫

迁移学习数据集

大连理工大学 付陈平 樊鑫

迁移学习是机器学习的一个重要分支，其主要目的在于将从原问题中学到的知识迁移应用到新问题中，希望模型学习具有“举一反三”的能力。目前，迁移学习被广泛应用于计算机视觉领域，例如图像分类、图像哈希、视频监控、姿态识别、人脸检测等。迁移学习的发展持续推动机器学习摆脱对大规模标记数据的依赖，同时改变机器学习总是“从头训练”模型的训练模式，因此迁移学习极大降低了人工标注数据的成本，并很大程度上节约了模型的训练时间，减少了模型的计算消耗。迁移学习的研究有助于机器学习模型泛化性、灵活性的建设，对机器学习“智能化”发展具有重要的意义。

在计算机视觉领域，从不同拍摄角度、不同光照、不同背景下捕获的场景，会造成同一目标特征分布的变化，但依然存在大量可迁移应用的相似性知识，因此，使用迁移学习构建可迁移相似知识、鲁棒的视觉模型以此解决计算机视觉任务是十分重要的。特别地，处理图像相关分类、识别等任务是计算机视觉领域应用迁移学习的一个基础性研究。研究具有更好迁移效果，构建更鲁棒的迁移学习算法一直是研究人员努力的目标。本文重点介绍几个在计算机视觉领域迁移学习常用、基础的数据集，具体包括 USPS、MNIST、COIL20、Office、Caltech-256 和 PIE 数据集。

1、USPS 和 MNIST

介绍：USPS (United States Postal Service) 数据集和 MNIST (Mini National Institute of Standards and Technology) 数据集是两个通用的手写体识别数据集，他们常应用于机器学习算法评测。USPS 是由美国邮政

署提供的手写字符数据集。MNIST 来自美国国家标准与技术研究所，由 500 个不同人手写的数字构成，其中人员组成包括 50% 的高中学生和 50% 的美国人口普查局工作人员。

USPS 数据集每张图尺寸为 16×16 ，共有 7,291 张训练图片和 2,007 张测试图片，MNIST 数据集每张图尺寸为 28×28 ，共有 60,000 张训练图片和 10,000 张测试图片。两个数据集均有 10 个类别，每个类别均是 0~9 间的某字符。图 1 是 USPS 数据集图片示例，图 2 是 MNIST 数据集图片示例，显然，两个数据集虽然字符类别相同但服从显著不同的概率分布。在迁移学习中，往往同时采用这两个数据集进行模型训练和测试，从这两个数据集中抽取数据，在进行一定的预处理（包括统一图片尺寸、编码图片像素灰度值信息等操作）后，分别组成辅助数据集和目标数据集，在辅助数据集和目标数据集间进行迁移学习的研究。图 1 和图 2 分别展示了 USPS 数据集和 MNIST 数据集的部分对象。

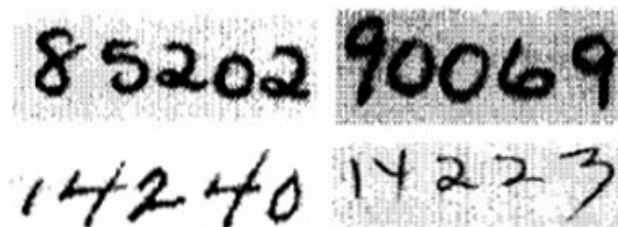


图 1 USPS 数据集

数据集地址

<https://www.kaggle.com/bistaumanga/usps-dataset>

<http://yann.lecun.com/exdb/mnist/>



图 2 MNIST 数据集

2、COIL20

介绍: COIL20 (Columbia University Image Library) 是一个拥有 20 种物体的彩色图片数据集，其常用于评测对象识别算法的识别效果，其中数据图片由哥伦比亚大学提供。

COIL20 数据集每张图片尺寸为 32×32 ，共计 1,440 张图片，20 个对象类别。图片采集拍摄对象时每隔 5 度 ($0^\circ \square 360^\circ$) 拍摄一张，每个物体共有 72 张图片。在迁移学习工作中，按旋转角度区间将数据集划分成不相交的两个数据集子集，令这两个数据集子集分别作为辅助数据集和目标数据集。例如将旋转角度为 $[0^\circ \square 85^\circ]$ 和 $[180^\circ \square 265^\circ]$ 的所有图片作为一个子集 COIL20-1，旋转角度为 $[90^\circ \square 175^\circ]$ 和 $[270^\circ \square 355^\circ]$ 的所有图片作为另一个子集 COIL20-2，将 COIL20-1 和 COIL20-2 分别作为辅助数据/目标数据集和目标数据/辅助数据集。图 3 展示了 COIL 数据集中部分对象。



图 3 COIL20 数据集

数据集地址

<http://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>

3、Office

介绍: Office 数据集是一个拥有 31 种类别，4,652 张图片的彩色图片数据集，是视觉迁移学习的主流基准数据集，其中图片的收集工作由伯克利大学完成。

Office 数据集中图片数据来自 3 个领域 (每个域均有 31 类物体)，分别是 Amazon, Webcam 和 Dslr。Amazon 是指从亚马逊在线电商获取的图片, Webcam 和 Dslr 是指使用网络摄像头或单反相机在不同照明和姿势变化下拍摄获得的办公环境图片。在迁移学习工作中，对每张图片抽取 SURF 特征，并向量化为 800 维的直方图表征，所有直方向量都进行减均值除方差的归一化处理，直方图码表由 K 均值聚类算法在 Amazon 子集上生成。从 3 个域 A(Amazon), W(Webcam)和 D(DSLR)中随机选取 2 个不同的域作为辅助数据集和目标数据集。图 4 展示了 Office 数据集中部分对象。



图 4 Office 数据集

数据集地址

<http://people.eecs.berkeley.edu/~jhoffman/domainadapt/>

4、Caltech-256

介绍: Caltech-256 (California Institute of Technology) 数据集是一个拥有 256 种类别 (每个类别的图片超过 80 张)，4,652 张图片的彩色图片数据集，是对象识别的基准数据集，其中数据图片由加利福尼亚理工学院收集整理。

Caltech-256 数据集中图片数据来自域 Google Image，并手工去除了不符合其类别的图片。在迁移学

习工作中, Caltech-256 数据集常与 Office 数据集一同使用, 具体做法为, 首先从 Office 数据集中选取任意一个域的数据集, 其次将 Caltech-256 和从 Office 中选取的数据集进行相同的预处理操作(提取特征、向量化表征、归一化处理等)后, 最后将处理后的两个数据集分别作为辅助数据/目标数据集和目标数据/辅助数据集。图 5 展示了 Caltech-256 数据集中部分对象图片。



图 5 Caltech-256 数据集

数据集地址

<https://www.kaggle.com/jessicali9530/caltech256>

5、PIE Face

介绍: PIE (Pose, Illumination, Expression) 是人脸识别的基准数据集, 现有多姿态人脸识别算法大多都在

PIE 人脸库上进行测试, 其中数据图片由卡内基梅隆大学提供。

PIE 数据集每张图片尺寸为 32×32 , 共有 68 个不同人物的 41,368 张人脸照片。PIE 名称是 3 个拍摄条件“朝向、光照、表情”英文首字母的缩写, 每张人脸有 13 个不同朝向, 43 个不同曝光程度和 4 种表情下的拍摄。在迁移学习中, 辅助数据集和目标数据集的建立主要有两种方法, 一种是固定朝向, 选择不同曝光程度的图片组成两个不相交的子集, 由此组成辅助数据集和目标数据集。另一种是固定曝光程度, 选择不同朝向的图片组成两个不相交的子集, 由此组成辅助数据集和目标数据集。图 6 展示了 PIE 数据集中部分人脸图像。



图 6 PIE 数据集

数据集地址

<https://pgram.com/dataset/cmuvasc-pie-face-dataset>

责任编辑 沈沛意



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

复旦大学团队“行人重识别”最新成果发表在 IEEE TPAMI -2020。

论文: Xuelin Qian, Yanwei Fu, Tao Xiang, Yu-Gang Jiang, Xiangyang Xue, Leader-based multi-scale attention deep architecture for person re-identification, IEEE TPAMI, vol. 42, no. 2, pp. 371-385, Feb. 2020

随着深度学习热潮地掀起,越来越多的计算机视觉任务开始使用深度学习方法进行解决,包括行人重识别任务。近几年提出的解决行人重识别任务的深度学习方法由于借鉴了目标分类的结构思想,通常只利用了单一尺度的特征。一方面,这些特征经过较深网络地提取,具有强大的抽象能力和表征能力。但是由于该特征更关注全局信息,并且不像多尺度特征能够进行反复推敲,因此对于行人重识别问题,尤其是针对一些难例,单一尺度的特征能力仍尤显不足。另一方面,利用多尺度学习的思想来进行匹配学习的过程和人类的自我感知方式非常相似。给定两张行人图片,人类通常会先粗略地捕捉图片中的显著特征或较大区域的信息,例如整体的衣着颜色、纹理等;如果这些信息不足以判断行人身份相同与否,那么人们会进一步仔细地有条理地观察一些细节信息,例如发型,鞋子款式等,直至获取足够的线索得出结论。因此如何设计一个有效的多尺度特征学习模型或许是帮助解决行人重识别问题的关键。

为此,复旦大学团队提出了一个新的基于领导的多尺度注意力行人重识别模型,如图1所示。该模型主要包含三个模块:多尺度数据流层模块、基于领导的注意力学习层模块和全局与局部分支层模块。具体地,对于输入的行人图片,多尺度数据流层对行人图片特征以不同尺度的感受野进行分析挖掘,从而提取由粗到细的不同尺度的视觉特征;同时通过基于领导的注意力学习层以全尺度信息作为指导,全方位地对各个尺度特征中具

有识别力的关键/显著性区域进行指导学习和突出强调,并对冗余干扰信息进行抑制;最后全局与局部分支层进一步从行人身体结构出发,将全局特征分割成若干等份局部特征,以此刺激局部特征的学习,并增强全局与局部特征之间的互补性。

该团队在四个常用的行人重识别任务公开数据集上对模型进行了测试。实验结果表明新模型能有效地提升行人重识别任务的精度,在各个测试数据集上均达到了世界领先的水平;同时,在迁移学习(无监督学习)的环境下,该团队也对新模型的性能进行了验证,结果表明相比于大部分工作,基于领导的多尺度注意力行人重识别新模型具有更好的可扩展性和泛化能力。

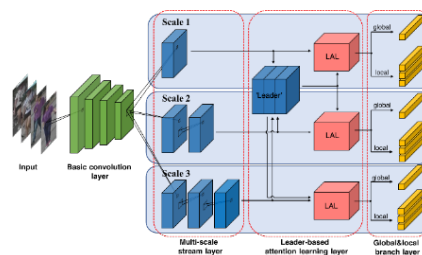


图1 基于领导的多尺度注意力行人重识别算法流程图

华中科技大学团队“基于顶点滑动的多方向目标检测”最新成果发表在 IEEE TPAMI-2020。

论文: Yongchao Xu, Mingtao Fu, Qimeng Wang, Yukang Wang, Kai Chen, Gui-Song Xia, Xiang Bai, Gliding vertex on the horizontal bounding box for multi-oriented object detection, IEEE TPAMI, Early Access, Feb. 2020

在多方向目标检测领域,广泛使用的基于顶点回归和角度回归的方法都存在各自的问题。具体而言,基于顶点回归的方法在物体倾斜接近 45° 存在标注的歧义问题;而基于角度回归的方法会因为角度预测的不准确性导致包围盒预测不准,同时该方法在角度临界点也会出现标注歧义性问题。

为此，本论文提出了一种基于顶点滑动的多方向目标表示方法，如图 2 所示。将倾斜物体的顶点表示为物体外接水平矩形的顶点在其四条边上的滑动顶点。这一做法避免了角度预测方法预测单一变量不稳定的问题，也将标注临界点歧义的问题由其他特殊角度带至易于解决的水平方向。为了进一步缓解水平物体表示的标注歧义性问题，本文提出了一种基于物体倾斜程度的包围盒选择机制。通过学习物体的倾斜程度判断物体是否接近于水平，对于预测为接近于水平的物体，该机制直接选择外接水平包围盒作为最终的输出，从而缓解了临界点预测混乱的问题。

本文的方法仅通过对多方向表示的简单修改与选择机制的配合，无需其他技术补充，在遥感目标检测 (DOTA, HRSC2016)、多方向文本检测 (MSRA-TD500, RCTW17) 以及鱼眼图像中的行人检测 (MW-18Mar) 中都取得了良好效果。

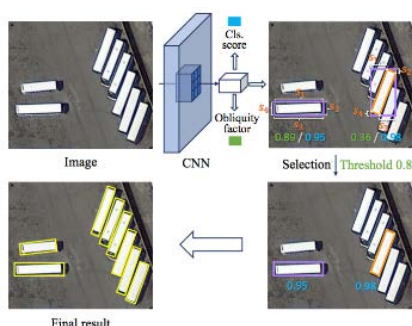


图 2 基于顶点滑动的多方向目标检测算法流程图

中国科学院自动化研究所模式识别国家重点实验室团队“基于斜率场恢复的曲面文档图像基线提取”最新成果发表在 IEEE TPAMI-2020。

论文: Gaofeng Meng, Chunhong Pan, Shiming Xiang, Ying Wu, Baselines extraction from curved document images via slope fields recovery, IEEE TPAMI, vol. 42, no. 4, pp.793-808, Apr. 2020

文档基线是用于文档内容排版的一族虚拟的平行线。基线估计在文档图像处理和分析，如文档几何畸变矫正、版面分析和文字识别中有着重要的应用。针对相机拍摄的弯曲文档图像，本文提出了一种基于斜率场恢

复的方法来提取图像中畸变的文档基线。相较于传统的基于文本行的估计方法，该方法的重大优势在于无需检测、分割文本行，同时也不受文档版面、文档语言和图像质量降质类型的影响，具有良好的鲁棒性和普适性。

本文假设弯曲的文档页面为一个可展曲面，于是文档图像的畸变基线可看作定义在文档图像斜率场上的常微分方程的一族解曲线 $dy/dx = \tan\theta(x, y)$ 。其中， $\theta(x, y)$ 表示文档基线上点 (x, y) 的切线角度。于是，文档基线的估计问题就转化为图像斜率场的估计问题。利用可展曲面的性质，跨越同一直母线的文档基线的切向量经透视变换后会聚于一点，如图 3 所示。从而图像斜率场的估计问题最终可转化为球面曲线的鲁棒回归问题。

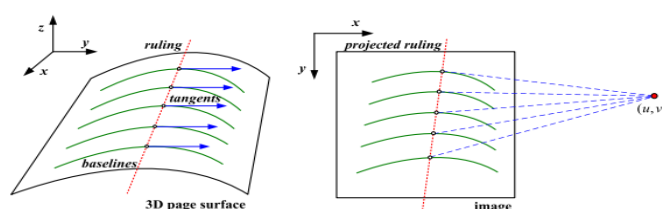


图 3 文档基线转换图

跨越同一直母线的文档基线具有相同的切向量。因此，经透视投影后，这些切线将相交于图像上一点。本文通过四个步骤来计算和提取畸变文档图像的基线：

1. 引入投影立方体作为图像边缘图的中间表示，用于快速计算图像局部方向角；
2. 利用条件随机场估计文档曲面上的直母线在图像上的投影，从而将斜率场的恢复问题转化为球面曲线的参数拟合问题；
3. 基于曲面的平滑性假设，通过球面曲线的鲁棒参数拟合，计算直母线对应的基线切向量的交点；
4. 利用欧拉法求解公式 $dy/dx = \tan\theta(x, y)$ 对应的常微分方程，得到文档基线。

在合成和真实采集的文档图像上的大量实验表明，本文提出的算法可以处理不同版面布局、不同文档类型和内容的文档图像。同时，对具有不同页面形状、不同拍摄视角和模糊的文档图像也能取得鲁棒的基线提取效果。

责任编辑 李策 贾同

美国纽约州立大学石溪分校凌海滨教授团队

CVLab@Stony Brook
石溪计算机视觉实验室

美国纽约州立大学石溪分校计算机视觉实验室由凌海滨教授和 Dimitris

Samaras 教授, Miachael Ryoo 教授, Minh Hoai Nguyen 教授共同领导。实验室的研究涉及计算机视觉研究的各个领域, 在相关领域也有着非常广泛的合作。其中, 凌海滨教授团队近年来在视觉跟踪, 视觉匹配, 航拍视频理解, 智能投影等领域取得了重要进展, 也在医学影像, 增强现实等应用领域开展了有效的合作。团队的主要成果、开源代码、以及数据集可以从网页 <https://www3.cs.stonybrook.edu/~hling/publication-selected.htm> 获取。

研究方向 1: 视觉跟踪 (Visual Tracking)

一般来说, 视觉跟踪的目的是从时序图像序列中动态地找出目标物体的状态。根据目标物体, 先验知识, 以及目标状态定义的不同, 视觉跟踪研究包括单目标跟踪, 多目标跟踪, 模型跟踪等课题。下面介绍团队在这些课题上的一些成果。

单目标跟踪的任务是在时序图像序列中自动地连续定位感兴趣的视觉目标, 比如车辆和行人。课题组潜心研究 10 余年, 在传统算法, 深度学习算法, 算法评测等方面都有重要影响。课题组提出将稀疏表达用于目标跟踪的算法是这个方向上最早的工作之一^[1], 直接或间

接启发了很多后续相关工作。如图 1 所示, 这个工作的原理在于, 跟踪目标 (图 1 中红色框) 的表观信息在时序序列中体现出非常高的冗余性, 可以使用稀疏的方式进行更高效的表达。相比之下, 误检框 (蓝色) 一般不具备这个性质。

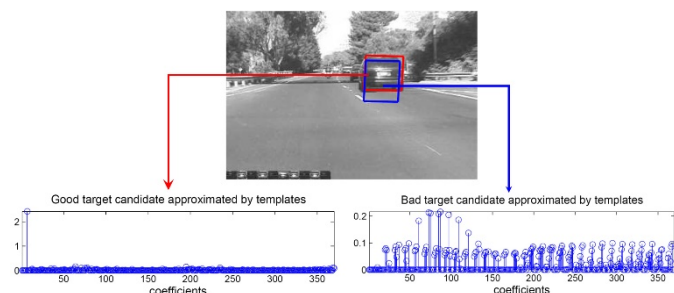


图 1 基于稀疏表示的目标跟踪原理

最近, 随着深度学习在视觉跟踪领域的兴起, 课题组也提出了基于深度学习框架的跟踪算法, 并且在主流评测集上取得了领先的结果。2019 年提出了 Cascaded-RPN 跟踪算法^[2] (如图 2 所示), 对 RPN 孪生跟踪网络进行了级联改进, 同时缓和了尺度变化和样本不均衡对跟踪算法的影响。目前, 课题组进一步研究如何有效地建模跟踪中的时序信息, 并取得了积极的初步结果。

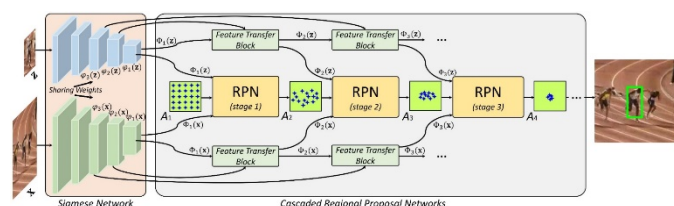


图 2 Cascaded-RPN 跟踪算法

由于数据标注的困难，视觉跟踪在数据评测集方面一直处于短板地位。课题组在这个方面做出了一系列重要贡献。2015年，课题组提出的彩色视频评测集 TC-128 已经被广泛采用^[3]。2019年，课题组提出了目前国际上最大（按照帧数）单目标跟踪库，LaSOT^[4]（图3），该库在标注质量、类别平衡、长序列跟踪、属性标注、语言描述等方面都给予了充分考虑，已经产生了重要影响。



图3 LaSOT 单目标跟踪评测集，包含 1400 个长视频序列，超过 350 万张手工标注的图像帧

在多目标跟踪方面的，课题组自 2013 起开始提出一系列基于低秩张量近似的跟踪方法^[5]。ICCV 2019 提出的 FAMNet^[6]，则将整个多目标跟踪系统以端到端的方式实现。如图 4 所示，FAMNet 的核心创新在于可微分的秩一张量近似模块以及由此串联的特征相似性学习和目标管理模块。

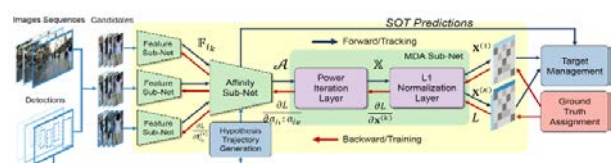


图4 FAMNet 多目标跟踪算法

此外，在基于模型的跟踪方面，课题组在手势跟踪^[7]，平面跟踪^[8]，线性结构跟踪^[9]，以及形变表面跟踪^[10]等方面都取得了相应的成果。

研究方向 2：视觉匹配 (Visual Matching)

匹配问题是计算机视觉的一个核心基础问题，在很多子领域中都扮演着重要角色。课题组早期在匹配方面

的工作主要集中于形状^[11]，特征^[12]，或图像的非刚性匹配方面^[13]。

近年来，由于结构化信息在视觉研究中的重要性，图匹配算法引起了越来越多研究学者的注意。2016年课题组开始这方面的研究，取得了一系列创新成果。其中有代表性的工作包括基于张量近似的多集合匹配以及最新的基于图网络的匹配优化算法。

多个集合之间的联合匹配在多目标跟踪和点云建模等问题中有着广泛的应用。通过深入研究经典的多维分配 (Multi-dimensional Assignment, MDA) 表达后，发现多集匹配和秩一张量近似 (Rank 1 Tensor Approximation, R1TA) 之间的等价关系，从而开辟了在多目标跟踪中使用张量分解的新思路^[5]。如图 5 所示，通过使用局部二集匹配作为下标对传统 MDA 进行改写，多集匹配可以由秩一张量近似过程来实现。这样一来，可以利用低秩张量分析中已经成熟的工具和理论来有效地解决多集匹配的问题，包括拓展到多个图之间的匹配^[14]。

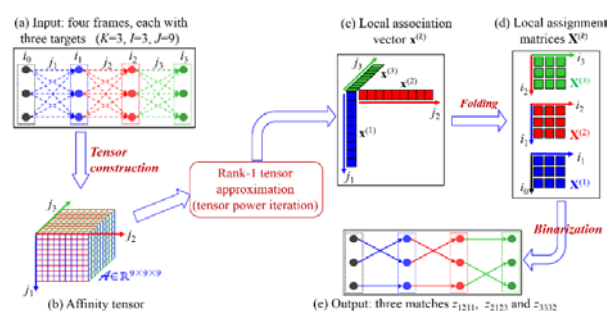


图5 基于秩一张量近似的多集匹配框架

除了传统的图匹配算法，课题组最新提出了基于图神经网络 (GNN) 的算法^[15]在流行的评测集上均取得了优异的结果。近来，深度学习算法已经开始应用于解决图匹配的问题，但之前的方案主要着眼于将匹配过程可微分化，从而使得特征和特征相似性可以通过学习的方法得到。不同于这些方案，基于 GNN 的算法^[15]之间对匹配过程进行学习。核心思想如图 6 所示，两个待匹配的图先联合构造出一个分配图 (assignment graph)，

从而将图匹配问题转换成分配图的节点标注问题，并且可以进一步使用图神经网络中边和节点的信息聚合来学习大尺度的图上下文关系。除了使匹配过程可学习化以外，该方法可以自然地将特征提取和特征相似性串联成一个端到端的可学习系统，从而进一步提高匹配精度。

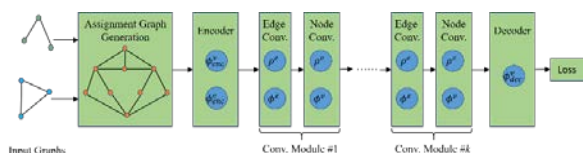


图6 基于图神经网络的图匹配算法

研究方向 3: 自适应投影补偿算法 (Projection Compensation)

随着硬件和算法的快速发展，投影技术越来越多应用于各行各业的很多应用场景中。然而，传统的投影方式极大地受限于投影平面的要求。一般来说，投影需要一个相对平坦且没有纹理的表面呈现投影效果。为了解决这个问题，投影补偿算法通过智能的方法对于待投影内容进行变换，使得投影的结果能够尽量抵消不完美的投影面(例如图7中带有图案的曲面(b))带来的影响，

达到视觉补偿的效果(例如图7中的投影结果(f))。

课题组首次提出基于深度学习的端到端补偿算法 CompenNet^[16]和 CompenNet++^[17]。其基本思想是在像素到像素的 U-Net 类似框架基础上，添加了一个额外的隐式处理成像环境的分支，同时引入了一个非线性几何矫正自网络模块，从而构建了一个端到端的补偿算法框架。这样的框架有效地避免了对于复杂光照成像过程的建模，同时隐式地利用了高层的语义信息。在大量的真实数据实验中，取得了明显优于传统方法的结果。

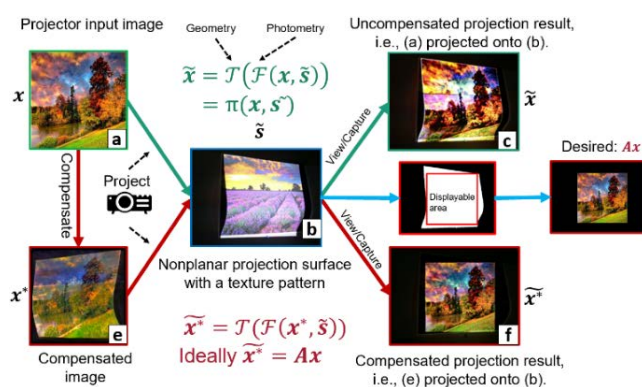


图7 投影补偿算法 CompenNet++

责任编辑 张汗灵

参考文献

- [1] Mei X, Ling H. Robust Visual Tracking using L1 Minimization. Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2009.
- [2] Fan H., Ling H. Siamese Cascaded Region Proposal Networks for Real-Time Visual Tracking. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.
- [3] Liang P., Blasch E., Ling H. Encoding Color Information for Visual Tracking: Algorithms and Benchmark. IEEE Trans. on Image Processing, 2015, 24(12):5630-5644.
- [4] Fan H., Lin L., Yang F., Chu P., Deng G., Yu S., Bai H., Xu Y., Liao C., Ling H. LaSOT: A High-quality Benchmark for Large-scale Single Object Tracking", in Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.
- [5] Shi X., Ling H., Xing J., Hu W. Multi-target Tracking by Rank-1 Tensor Approximation. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2013.
- [6] Chu P., Ling H. FAMNet: Learning Feature, Affinity and Multi-dimensional Assignment for Online Multiple Object Tracking. Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2019.
- [7] Li P., Li X., Liao C., Ling H. 3D Hand Pose Estimation Using Randomized Decision Forest with Segmentation Index Points. Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2015.
- [8] Wang T. Ling H. Gracker: A Graph-based Planar Object Tracker. IEEE Trans. on Pattern Analysis and Machine Intelligence

- (PAMI), 2018, 40(6):1494-1501.
- [9] Chu P., Pang Y., Cheng E., Zhu Y., Zheng Y., Ling H. Structure-Aware Rank-1 Tensor Approximation for Curvilinear Structure Tracking Using Learned Hierarchical Features. Proc. of the Int'l Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI), 2016.
- [10] Wang T., Ling H., Lang C., Feng S., Hou X. Deformable Surface Tracking by Graph Matching. Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2019.
- [11] Ling H., Jacobs D. Shape Classification Using the Inner-Distance. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2007,29(2):286-299.
- [12] Ling H., Okada K. An Efficient Earth Mover's Distance Algorithm for Robust Histogram Comparison. IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), 2007,29(5):840-853.
- [13] Ling H., Yang X., Latecki L. Balancing Deformability and Discriminability for Shape Matching, Proc. Of European Conf. on Computer Vision (ECCV), 2010.
- [14] Shi X., Ling H., Pang Y., Hu W., Chu P., Xing J. Rank-1 Tensor Approximation for High-Order Association in Multi-Target Tracking, International Journal of Computer Vision (IJCV), 2019,127(8):1063-1083.
- [15] Wang T., Liu H., Li Y., Jin Y., Hou X., Ling H. Learning Combinatorial Solver for Graph Matching. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2020.
- [16] Huang B., Ling H. End-to-end Projector Photometric Compensation. Proc. of IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), 2019.
- [17] Huang B., Ling H. CompenNet++: End-to-end Full Projector Compensation. In Proc. of the IEEE Int'l Conf. on Computer Vision (ICCV), 2019.



凌海滨

1997 年和 2000 年于北京大学分别获得学士和硕士学位，2006 年在美国马里兰大学获博士学位，2006 至 2007 年在加州大学洛杉矶分校从事博士后研究。2000 至 2001 年任微软亚洲研究院助理研究员，2007 至 2008 年任西门子研究院研究员，2008-2019 年任职于美国天普大学，2019 年加入纽约州立大学计算机系任 SUNY Empire Innovation 教授。主要研究领域为计算机视觉、增强现实、医学图像分析和人机交互。获 2003 年度 ACM UIST 最佳学生论文奖，2014 年度美国自然科学基金 CAREER Award。担任期刊 IEEE Trans. on Pattern Analysis and Machine Intelligence (PAMI), Computer Vision and Image Understanding (CVIU), 和 Pattern Recognition 的编委，多次担任 CVPR、ECCV 等会议的领域主席。

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 和 Pattern Recognition。

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) 由中国模式识别学术会议 (CCPR)

和中国计算机视觉大会 (CCCV) 合并而来，由中国人工智能学会 (CAAI)、中国计算机学会 (CCF)、中国自动化学会 (CAA) 和中国图象图形学学会 (CSIG) 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。

第三届 PRCV 将于 2020 年 10 月 16-18 日在南京举行，由 CCF-CV 专委会、南京理工大学、南京信息工程大学、东南大学承办。本届会议将主要汇聚国内从事 PRCV 理论与应用研究的广大科研工作者及工业界同仁，共同分享我国 PRCV 领域的最新理论和技术成果，为大家提供精彩的学术盛宴。现向广大科技工作者公开征集高质量、原创性的优秀论文。会议论文集将由 Springer 出版社 LNCS 系列出版，并被 EI 和 CPCI-S 检索。

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
ICPR 2020	2021.01.10-15	Milan, Italy	2020.04.15	https://www.icpr2020.it/
ACM MM 2020	2020.10.12-16	Seattle, USA	2020.04.25	https://2020.acmmm.org
PRCV 2020	2020.10.16-18	中国 南京	2020.04.30	http://www.prcv.cn/2020/
BMVC 2020	2020.09.07-11	Manchester, UK	2020.05.01	http://www.bmvc2020.com/
ISAIR 2020	2020.08.08-10	Kitakyushu, JP	2020.05.01	https://isair.site/
NeurIPS 2020	2020.12.06-12	Vancouver, CA	2020.05.13	https://nips.cc/Conferences/2020

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	Deep Learning for Panoramic Vision on Mobile Devices	https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/deep-learning-for-panoramic-vision-on-mobile-devices	2020.05.31
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	2020.05.31
PR	Graph-based Methods for Large Scale Fin. & Bus. Data Analysis	https://www.journals.elsevier.com/pattern-recognition/call-for-papers/graph-based-methods	2020.06.30

COMPUTER VISION NEWSLETTER

01 2020
总第 23 期



计算机视觉专委会简报



CCF 计算机视觉
专委会