

CCF

计算机视觉 专委会简报

COMPUTER VISION NEWSLETTER

2019/05 期
总第 21 期

专委动态

走进高校系列活动

科技前沿

研究热点追踪

专题综述



主 办： CCF 计算机视觉专业委员会

主 编： 王 亮

执行主编： 李实英

网 址： <http://ccfcv.ccf.org.cn>

E m a i l： ccfcvn@gmail.com

COMPUTER VISION NEWSLETTER

计算机视觉

专委简报编委会

主 编 王 亮 中国科学院自动化研究所

执行主编 李实英 上海科技大学

专委动态

主 编 毋立芳 北京工业大学

编 委 马占宇 北京邮电大学

王瑞平 中国科学院计算技术研究所

虞晶怡 上海科技大学

科技前沿

主 编 王金甲 燕山大学

编 委 崔海楠 中国科学院自动化研究所

邓 成 西安电子科技大学

任传贤 中山大学

任桐炜 南京大学

苏 航 清华大学

杨巨峰 南开大学

委员风采

主 编 余 焯 合肥工业大学

编 委 黄 岩 中国科学院自动化研究所

刘海波 哈尔滨工程大学

张汗灵 湖南大学

资源平台

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

蹇木伟 山东财经大学

金 鑫 北京电子科技学院

刘 丽 国防科学技术大学

沈沛意 西安电子科技大学

COMPUTER VISION NEWSLETTER

CONTENTS

目录

专委动态

| | | |
|------|--------------------|----|
| 走进高校 | 走进高校系列活动 | 04 |
| 走进企业 | 走进企业系列交流会 | 05 |
| 专委通知 | PRCV 2021 征集承办单位 | 06 |
| 会议推介 | PRCV 2019 博士生论坛报名 | 08 |
| | PRCV 2019 诚邀申请成果展示 | 10 |
| | PRCV 2019 开始注册 | 11 |

科技前沿

| | | |
|------|---------------|----|
| 专题综述 | 深度学习与模型优化研究介绍 | 14 |
| 热点追击 | 面向视频数据的对抗攻击研究 | 18 |
| | 行人再识别相关研究进展 | 19 |

委员风采

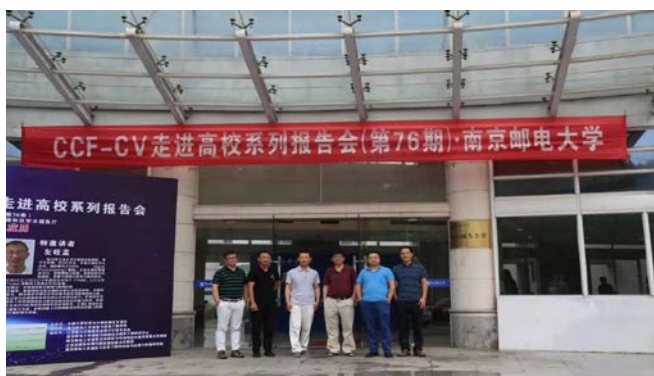
| | | |
|-------|-------------|----|
| 委员访谈 | 武汉大学夏桂松教授访谈 | 21 |
| 委员好消息 | | 24 |

资源平台

| | | |
|------|--------|----|
| 开源代码 | 视频描述 | 25 |
| 数据集 | 医学影像分割 | 31 |
| 招聘信息 | | 33 |
| 征文通知 | | 39 |

CCF-CV 走进高校系列报告会

第 76 期 南京邮电大学



2019年7月12日，第76期CCF-CV走进高校系列报告会在南京邮电大学成功召开。本次研讨会由中国计算机学会计算机视觉专委会主办，南京邮电大学通信与信息工程学院承办，江苏省人工智能学会模式识别专委会、南京邮电大学通信与网络技术国家工程研究中心、南京邮电大学宽带无线通信与传感网技术教育部重点实验室、江苏省图象处理与图象通信重点实验室以及南京邮电大学通信与信息工程学院信号处理与传输研究院协办。本次研讨会邀请到华中科技大学刘文予教授、南京理工大学杨健教授以及哈尔滨工业大学左旺孟教授做特邀报告，并由南京邮电大学通信与信息工程学院周全副教授担任执行主席。

此次CCF-CV走进高校系列报告会持续了三个多小时，三位专家学术功底深厚、研究成果丰硕，为本期“CCF-CV走进高校系列报告会活动”奉献了三场高质量的专业学术报告。每轮报告结束后，专家与听众亲切互动和交流，并对大家提出的问题一一解答，现场听众受益匪浅。

第 77 期 澳门大学



2019年8月14日上午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、澳门大学科技学院承办的第77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”在澳门大学E11-G015报告厅举行。本期报告会邀请了上海科技大学虞晶怡教授、上海交通大学严骏驰研究员、中国科学院自动化研究所兴军亮研究员、北京邮电大学薛哲博士等四位专家学者做特邀报告，澳门大学科技学院 Dr. Bob Zhang 担任本次报告会的执行主席。

此次报告会是“CCF-CV走进高校系列报告会”首次走进港澳台地区的大学，提升了CCF-CV走进高校系列活动的影响力，增进了CCF-CV与粤港澳大湾区的大学之间的友谊，同时为CCF与粤港澳大湾区的大学进一步合作起到了积极的促进作用。

（责任编辑：马占宇）

CCF-CV 走进企业系列交流会

第 18 期 慧眼科技

时间：2019 年 7 月 11 日

2019 年 7 月 11 日，中国计算机学会计算机视觉专委会（CCF-CV）走进企业系列交流会第十八期活动——走进慧眼科技，在广州市南沙区慧眼智能科技研究院成功举办。这是 CCF-CV 专委会在华南开展的第二次走进企业系列交流会，也是首次进入工业制造领域，多位著名学者齐聚一堂，针对计算机视觉前沿技术及其在工业制造中的应用展开了深度交流与探讨。



在活动开始前，慧眼科技集团董事长吴凯在致辞中首先向 CCF-CV 以及与会的业界专家和嘉宾表达了诚挚的感谢和欢迎。吴凯表示，慧眼科技是拥有 15 年历史的工业领域视觉系统及设备提供商，具备成熟稳定的应用开发团队和完备的商业管理体系。但在核心视觉算法研发方面基础相对薄弱，希望能与学术端深度融合，密切合作，使产学研转换更加迅速有效，让科研工作者的灵感和心血能真正服务行业、服务社会、创造价值。在热烈的掌声中，交流会正式开始。

首先，上海科技大学虞晶怡教授以 Learning

to Build a New Reality 为主题给大家带来了基于深度表面光场实现三维物体及场景重建的技术分享报告。虞教授为大家科普了基于四维光场数据实现三维物体重建的原理和发展历程，介绍超逼真物体实现及场景三维重建的近期工作并为大家展示了应用成果及移动端应用案例。

随后，深圳职业技术学院人工智能学院院长杨金锋教授为大家带来了基于视觉的缺陷检测及工业应用的技术报告。杨教授梳理了机器视觉应用类型及挑战，揭示了工业制造环境下缺陷检测的技术难点及独特性，围绕以上问题介绍了其课题组的解决方案以及应用实例。

接下来，北京航空航天大学祝世平副教授为大家分享了图像显著性检测的报告。祝教授围绕该问题介绍了课题组的相关成果和进展，包括基于输出引导注意力机制的显著性目标检测等。

最后，由慧眼智能科技研究院副院长朱俊勇博士对慧眼科技的公司概况、业务范围、研究内容、行业影响等作了全面介绍。

在茶歇环节及报告结束后，参会人员围绕主讲嘉宾的报告内容，就彼此感兴趣的话题开展了交流和讨论，与主讲嘉宾进行深入研讨。

主题报告结束后，吴凯董事长以及赖剑煌教授带领各位嘉宾参观了公司展厅，大家一边交流讨论，一边观摩体验慧眼科技的视觉检测技术成果和智能化设备，对慧眼科技取得的成绩表示认可并给予了高度评价。活动结束后大家一同合影留念，祝贺本次交流活动圆满成功。

（责任编辑：王瑞平）

PRCV2021 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、中国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, 2019 年由西北工业大学 (西安) 等承办, 2020 年由南京理工大学 (南京) 等承办, 现征集 2021 年承办单位。

1. 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。2021 年的举办地应该是在广州、西安、南京以外的地区。

2. 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。2021 年 PRCV 应该避开领域内主要的相关国际学术会议的时间。

3. 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。

4. 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。

5. 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。对中国自动化学会、中国计算机学会、中国人工智能学会、中国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。

6. 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 2019 年 10 月 15 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 2019 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室
王亮

邮编：100190

电话：010-82544643

邮箱：prcv@nlpr.ia.ac.cn

（责任编辑：马占宇）

PRCV2019 博士生论坛报名

中国模式识别与计算机视觉大会 (Chinese Conference on Pattern Recognition and Computer Vision) 是由中国模式识别学术大会 (CCPR) 和中国计算机视觉大会 (CCCV) 合并而来, 定位国内顶级的模式识别和计算机视觉领域学术盛会。首届中国模式识别与计算机视觉大会 (PRCV2018) 于 2018 年 11 月 23-26 日在广州成功召开。

第二届中国模式识别与计算机视觉大会 (PRCV2019) 将于 2019 年 11 月 8-11 日在西安举行。PRCV2019 由中国计算机学会 (CCF)、中国自动化学会 (CAA)、中国图象图形学学会 (CSIG) 和中国人工智能学会 (CAAI) 联合主办; 由西北工业大学、西安交通大学、西安电子科技大学以及陕西师范大学等单位联合承办。

论坛概况

本论坛为在读博士生提供了与计算机视觉和模式识别领域内经验丰富的研究者沟通交流的机会。每位学生都会被分配一位领域内的资深成员作为导师, 具体分配方案会基于学生的个人偏好或两者研究兴趣的相似程度而制定。被录用的学生可以获得跟导师面对面一对一交流的机会。导师团队包括但不限于如下著名专家学者:

谭铁牛 中科院自动化所研究员
王 亮 中科院自动化所研究员
陈熙霖 中科院计算所研究员
赖剑煌 中山大学教授
王 涛 爱奇艺公司首席科学家
查红彬 北京大学教授
白 翔 华中科技大学教授

纪荣嵘 厦门大学教授
贾云得 北京理工大学教授
林宙辰 北京大学教授
刘青山 南京信息工程大学教授
卢湖川 大连理工大学教授
鲁继文 清华大学副教授
苗启广 西安电子科技大学教授
山世光 中科院计算所研究员
王蕴红 北京航空航天大学教授
虞晶怡 上海科技大学教授
张艳宁 西北工业大学教授
高新波 西安电子科技大学教授
黄庆明 中国科学院大学教授
赵 耀 北京交通大学教授
薛建儒 西安交通大学教授
杨小康 上海交通大学教授
刘成林 中科院自动化所研究员
王井东 微软亚洲研究院研究员
姜育刚 复旦大学教授
程明 明 南开大学教授
高盛华 上海科技大学副教授、研究员
程 健 中科院自动化所研究员
彭宇新 北京大学教授
孟德宇 西安交通大学教授
左旺孟 哈尔滨工业大学教授
欧阳万里 悉尼大学副教授
董伟生 西安电子科技大学教授
方玉明 江西财经大学教授
丛 杨 中科院沈阳自动化所研究员
毋立芳 北京工业大学教授
韩军伟 西北工业大学教授

耿 新 东南大学教授

刘 偲 北京航空航天大学副教授

所有学生和导师都会被邀请参加博士生论坛，使学生有机会与导师探讨其目前的研究进展和未来的职业规划。同时每位学生都需要向其他参会者和导师展示一张海报（由会务统一打印），内容为介绍博士学位论文研究成果或最近发表的一篇论文成果。2019 年度 CCF-CV 学术新锐奖获得者和正式候选人将被直接邀请参加博士生论坛。

论坛同时欢迎近两年内在相关领域顶级会议和顶级期刊发表论文的博士生报名进行论文分享，报名成功的博士生同样可与上述专家学者进行沟通交流。具体要求如下：

1. 论文发表的期刊和会议

期刊：IEEE T-PAMI、IJCV、IEEE TIP、IEEE TNNLS、IEEE TMM、JMLR、IEEE TCSVT

会议：CVPR、NeurIPS、ICCV、ICML、ECCV、AAAI、IJCAI、ACM MM、ICLR

以及同等质量的期刊或会议，同一会议论文，oral 优先；

2. 最近年份优先，如 2019>2018；

3. 每位学生限报名一篇论文进行分享；

4. 报名材料等要求与博士生论坛相同。

参会资格

参会学生投稿时须为在读博士研究生，且研究方向为计算机视觉或模式识别领域。

申请步骤

满足参会资格的学生需把申请材料发送至论坛主席邮箱，申请材料须为单个 pdf 文件或单个 zip/rar 压缩文件（内含多个 pdf 文件），材

料中须包括内容如下：

1. 个人简历；

2. 海报对应论文或顶会顶刊论文题目以及作者列表；

3. 5 位候选导师的姓名和邮箱列表，按照偏好程度降序排列。

注：请不要列出自己的博士导师作为论坛导师。

请确保申请材料中包含以上信息，不完整的材料将会被拒绝。

审核流程

如果因条件所限无法全部邀请满足要求的申请者，论坛组委会将基于上述申请材料及申请者的毕业时间进行选择，且会考虑覆盖尽量多的研究机构。

重要时间节点

申请材料提交截止日期：2019 年 9 月 15 日

审核结果公布日期：2019 年 10 月 8 日

论坛主席

有任何问题，请联系以下论坛主席：

韩军伟, junweihan2010@gmail.com, 西北工业大学

耿 新, xgeng@seu.edu.cn, 东南大学

刘 偲, liusi@buaa.edu.cn, 北京航空航天大学

（责任编辑：毋立芳）

PRCV2019 诚邀申请成果展示

中国模式识别与计算机视觉大会 (Chinese Conference on Pattern Recognition and Computer Vision) 是由中国模式识别学术大会 (CCPR) 和中国计算机视觉大会 (CCCV) 合并而来, 定位国内顶级的模式识别和计算机视觉领域学术盛会。首届中国模式识别与计算机视觉大会 (PRCV2018) 于 2018 年 11 月 23-26 日在广州成功召开。

第二届中国模式识别与计算机视觉大会 (PRCV2019) 将于 2019 年 11 月 8-11 日在西安举行。PRCV2019 由中国计算机学会 (CCF)、中国自动化学会 (CAA)、中国图象图形学学会 (CSIG) 和中国人工智能学会 (CAAI) 联合主办; 由西北工业大学、西安交通大学、西安电子科技大学以及陕西师范大学等单位联合承办。

为促进模式识别与计算机视觉领域先进科研成果转化, 打造实用性创新型研究成果的展示平台与窗口, 加强学术界与工业界的沟通与交流, 现面向广大高校、科研院所和企业公开征集模式识别与计算机视觉领域的优秀成果展示项目。展示内容包括但不限于模式识别与计算机视觉领域非商用的新型算法、模型、技术、系统原型、产品等, 欢迎推选相关优秀 PR&CV 成果到大会进行宣传 and 展示。

有意参与 PRCV2019 成果展示者需填写申请表并发邮件至 xirunping@126.com 提供展示内容名称、作者、单位、成果简介、图片示例等信息。申请展示成果将由组委会组织专家进行评选, 大会将为每个入选的展示成果提供展位一个, 每个展位需要至少非学生注册两名参会人员 (不得与参会论文作者注册共享)。企业产品展览需另缴

纳展览费用 8000 元, 且会议期间不得私自撤展。

展示环境及要求

- 展位 (会务提供): 标准展位 (含一张展台、两把凳子和电源等)
- 展示设备 (自备): 笔记本、显示器及其它设备等
- 成果简介: 制作展示期间用海报 1 页, 推荐尺寸: 宽 80cm, 高 120cm。(参展单位也可以根据展位大小结合需求自备展架或自行布展)

重要日期

- 申请截止: 2019 年 9 月 30 日
- 评选结果: 2019 年 10 月 8 日
- 会期展示: 2019 年 11 月 8-11 日

联系方式

马惠敏, mls18@mails.tsinghua.edu.cn, 清华大学

郝润平, xirunping@126.com, 西北工业大学

附件

PRCV2019 成果展览展示申请表

| | |
|----------|-------------------|
| 成果名称 | |
| 参展单位 | |
| 参展人姓名、职称 | |
| 联系人邮箱 | |
| 联系人手机 | |
| 成果简介 | |
| 备注 | (如展示设备环境、条件等特殊需求) |

(责任编辑: 毋立芳)

PRCV2019 开始注册

中国模式识别与计算机视觉大会 (Chinese Conference on Pattern Recognition and Computer Vision) 是由中国模式识别学术大会 (CCPR) 和中国计算机视觉大会 (CCCV) 合并而来, 定位国内顶级的模式识别和计算机视觉领域学术盛会。首届中国模式识别与计算机视觉大会 (PRCV2018) 于 2018 年 11 月 23-26 日在广州成功召开, 详情参看 PRCV2018 会议总结。

第二届中国模式识别与计算机视觉大会 (PRCV2019) 将于 2019 年 11 月 8-11 日在西安举行。PRCV2019 由中国计算机学会 (CCF)、中国自动化学会 (CAA)、中国图象图形学学会 (CSIG) 和中国人工智能学会 (CAAI) 联合主办; 由西北工业大学、西安交通大学、西安电子科技大学以及陕西师范大学等单位共同承办。

本届大会将秉承提升我国模式识别与计算机视觉科学研究水平的发展理念, 着眼国内、面向国际, 为业内人士倾力奉献精彩的学术盛宴, 共同分享我国模式识别与计算机视觉领域的最新理论和技术成果。会议邀请到 Kyros Kutulakos、Tinne Tuytelaars、Kyoungmu Lee、焦李成四位专家做大会报告, 另外还有 6 个邀请报告、40 个应用展示、4 个前沿讲习班、7 项竞赛获奖交流、3 个 CCFCV 学术新锐奖交流、顶会顶刊交流、录用论文口头报告和墙报等精彩内容, 期待您共享计算机视觉与模式识别学术盛宴!

注册入口

官方网站网址: <http://www.prcv.cn/>

注册入口:

<http://event.3lhuiyi.com/1666654090/SelectCate>

手机端注册请扫如下二维码进行注册:



注册费标准

注册类型: 前期优惠注册 (2019 年 10 月 8 日前 (含))、标准注册 (2019 年 10 月 8 日后)、现场注册 (2019 年 11 月 8 日开始)

全额注册: ¥2300、¥2800、¥3200

会员注册: ¥1800、¥2200、¥2500

学生注册: ¥900、¥1200、¥1500

注:

1. 会员指中国计算机学会 (CCF) 会员、中国自动化学会 (CAA) 会员、中国图象图形学学会 (CSIG) 会员或者中国人工智能学会 (CAAI) 会员。
2. 注册费到款方可视为注册成功。
3. 学生注册不包含颁奖典礼环节。
4. 如有问题或需要帮助, 请联系我们。邮箱:

prcv19registration@163.com

支付方式

1. 支付宝付款

点击相应注册类型填写信息后选择在线支付宝付款, 可利用支付宝绑定公务卡付款

2. 微信支付

点击相应注册类型填写信息后选择在线微信支付，可利用微信绑定公务卡付款

3. 线下支付

点击相应注册类型填写信息后选择线下支付银行汇款转账支付方式，银行汇款转账信息如下：

户名：湖南星光会议会展有限公司

帐号：6606 0155 2000 03166

开户行：上海浦东发展银行长沙侯家塘支行
转账汇款信息备注为”姓名+PRCV2019“

4. 会议报到现场付款

现场填写个人信息后进行注册付款。

已获赞助

金牌赞助



银牌赞助



铜牌赞助





(责任编辑：毋立芳)

深度学习与模型优化研究介绍

大连理工大学 刘日升 程世超 马龙 付陈平 樊鑫 罗钟铉

一、引言

2012年，深度学习算法在ImageNet图像识别大赛中脱颖而出。随后深度学习在各个领域的应用出现井喷式增长。特别是2016年，基于深度学习开发的AlphaGo战胜了国际顶尖围棋高手李世石，深度学习的热度一时无两，其发展也进入突飞猛进的时代。相较于传统机器学习，深度学习可在大规模数据集上自动拟合出理想的数据分布，进而在目标明确的任务上取得成功。然而，深度学习是一个众所周知的黑箱，尽管表达能力很强，但可解释性不足。这意味着网络在使用过程中变量传播是不可控的，停止点无法为人所知。因此研究人员希望探究一套深度学习的支撑理论，使其使用过程可在人为控制下进行。

提及理论，模型优化是理论分析方面较为全面、系统的解决问题的手段。这类方法往往通过对问题的先验知识建立目标能量函数并使用各种优化策略对目标函数进行求解。然而，由于其依靠手工设计的先验知识，在数据先验分布较为复杂时无法充分发挥自身的优势。

深度学习与模型优化本质上都在试图刻画问题背后的规律，拟合输入、输出间的数据分布。在问题解决上，深度学习能很好地自动进行数据拟合但缺乏理论分析，而模型优化有充分的理论分析但在数据复杂时难以拟合数据分布。因此，可探索它们之间的联系，将两者结合以弥补各自在理论分析和数据拟合上的缺陷。

二、深度学习与模型优化研究介绍

1. 相关工作

近年来，大多数学习和视觉任务可表述为下列正则化优化模型：

$$\min_{\mathbf{x}} \Phi(\mathbf{x}) := f(\mathbf{x}) + g(\mathbf{x}) \quad (1)$$

式中 f 为损失项， g 为正则化项， \mathbf{x} 为模型输入。

不同于经典数值求解，它们纯粹基于数学推导设计迭代过程，而现阶段有许多研究试图通过大量数据训练网络以实现模型优化目标。现有研究工作大致可分为两类。一类方法的核心思想是抛弃公式(1)中正则化项即放弃使用先验知识，仅使用深度学习网络完成任务。文献[1-2]在每次迭代过程中直接使用网络结构取代先验相关的数值计算。最近，循环^[3]、展开^[4]和强化^[5]等学习策略均被引入网络训练中以用于模型优化。然而这类方法忽略使用正则项，导致先验知识难以运用到迭代过程中。此外其过于依赖网络结构，缺乏理论支撑，难以保证收敛性。另一类方法的核心思想是为特定的先验公式引入超参数，并展开公式更新过程获得优化方案。例如[6-7]对L1正则化项引入超参数并采用一阶方法对公式进行推导。这类方法主要基于特定的先验知识建立模型，因而无法应用到一般学习和视觉任务。此外，其引入的超参数太过简单难以拟合复杂数据分布。

以上工作就性能而言均取得了进展，但深度学习与模型优化理论仍处于割裂状态，我们希望将两者结合起来，以实现两者优势互补的目的。

2. 深度学习与模型优化研究工作介绍

为实现两者结合的目的，本团队主要从优化角度出发，建立模型优化和深度学习之间强有力的联系。具体而言，在已建立模型的基础上设计一个最优性模块，通过一阶最优性条件，给出一个判断准则来决定深度学习模块(即可学习的网络结构)的结果是否可以接受。若结果可接受，模型就采纳当前迭代中网络的结果，若不可接受，模型就放弃网络的结果，采用保守的近端梯度法给当前的解一些提示和推动，进而在下一次迭代中促使网络找到可靠的下降方向。下面将介绍团队基于这一思想所作的几项工作。

工作 1：为弥合深度传播和模型优化间的差

距，本团队提出了基于传播和优化的深度模型 PODM(Propagation and Optimization based Deep Model, PODM)^[8]。PODM 主要包括两个模块：传播模块 P 和优化模块 O。传播模块是基于深度学习网络设计的，其要求具有足够的灵活性以在每次迭代中能寻找到被理论支持的下降方向。优化模块是基于最优化理论建立的模块，用于评判传播模块的迭代结果是否可接受，给与传播模块下次迭代的依据。然后，PODM 将传播模块和优化模块级联到一起实现整个迭代过程。虽然该方法加入了可学习的深度网络结构，但仍然可以证明在迭代过程中变量的传播序列可以收敛到对应模型的临界点。这意味着经过有限次迭代，该方法总是可以求解到一个用户想要的局部最优解的较小邻域范围内。图 1 说明了 PODM 迭代机制。

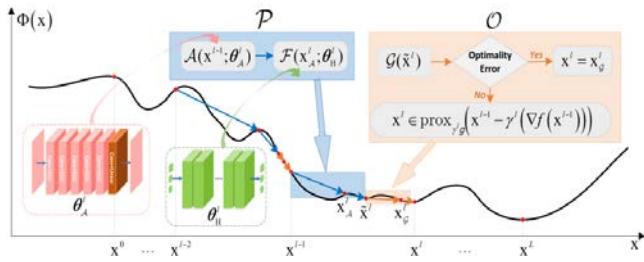


图 1 PODM 运行机制图

在此基础上，团队进一步设计了端到端的可学习网络。为方便训练，去掉了 PODM 的选择机制，直接将传播模块和近端梯度运算结合在一起进行联合训练。这样可有效避免训练数据和测试数据分布不一致时，判断准则带来的参数失效问题。部分实验结果见图 2，图 3。

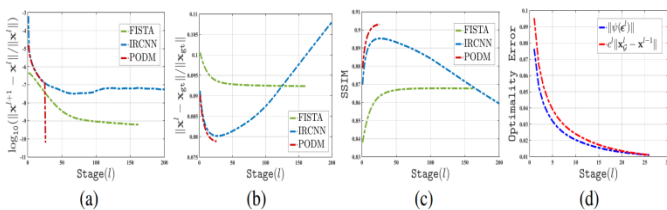


图 2 FISTA, IRCNN 和 PODM 收敛效果对比

图 2 展示了 PODM 与其他方法收敛效果的对比。首先，实验通过稀疏编码问题给出方法收敛性分析。其次，通过与传统方法 FISTA、可学习方法 IRCNN 做对比，给出相对误差、重构误差的结果对比。收敛结果显示由于引入网络，PODM 的迭代次数明显减少。此外，团队还给出了迭代过

程中网络被采纳的情况说明，事实上只要网络对问题而言是适合的，网络的结果均会被采纳。

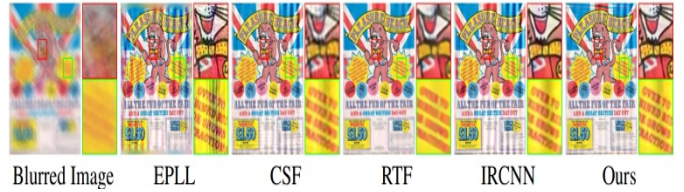


图 3 真实图像上去模糊效果对比

图 3 展示了 PODM 与其他方法在图像恢复上的效果对比。将 PODM 与传统优化方法、深度学习方法相比较，会发现 PODM 结合了二者优势，在纹理细节和文字恢复上都取得了较好的效果。

工作 2：针对图像增强问题，本团队从互补传播角度提出了一种名为深度先验集成 DPE(Deep Prior Ensemble, DPE)的框架^[9]。DPE 首先依据图像模型建立基本传播方案，然后引入残差网络帮助预测每个阶段的传播方向。DPE 可面向不同的图像增强任务集成与任务相应的知识和数据，是一种通用集成框架。DPE 从数据集中学习梯度下降方向，相较于传统基于优化的最大后验方法，这种方式不受局部最小值的影响因此更加稳健。此外，DPE 可充分利用图像模型邻域知识获得更多任务信息，而单纯深度网络只能依赖大规模的数据，无法使用额外的任务信息。

DPE 主要包括三部分：热启动、残差网络和先验投影。热启动用于刻画图像增强任务的先验知识，残差网络用于从数据集中学习图像的传播方向，基于误差反馈的先验投影则用于进一步保证传播方向是朝向理想输出方向的。最后，DPE 引入了一项次梯度约束项，将这三部分通过反馈控制策略级联起来。关于 DPE 详细说明可见图 4。

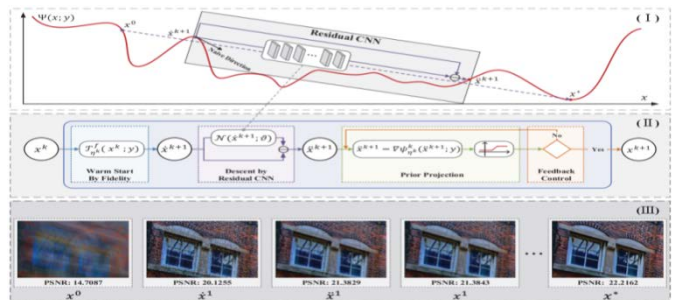


图 4DPE 说明

最后，团队将 DPE 应用于各种图像增强任务

(例如图像反卷积, 插值, 超分辨率, 雾霾去除和 水下增强), 以检验其有效性。实验证明 DPE 确实 优于现阶段主流的图像增强技术。部分实验结果 见图 5。

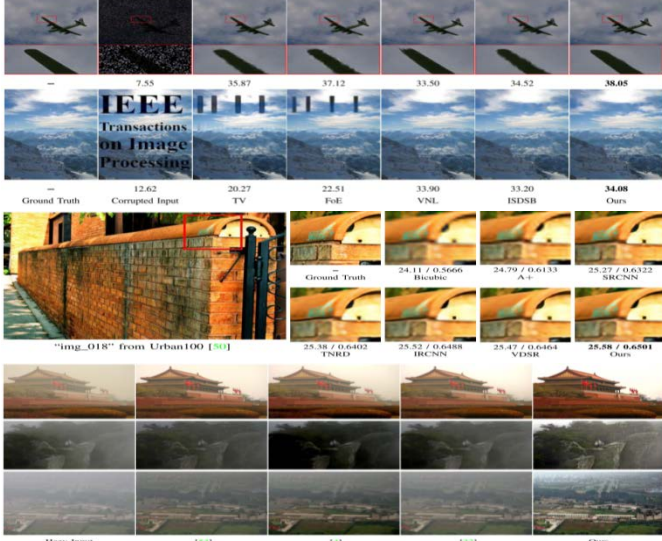


图 5 在不同图像增强任务上的效果对比

工作 3: 针对大多数深度模型缺乏理论支持, 先验信息加入会影响网络收敛的问题, 本团队提出一种通用的基于非凸优化展开的深度模型方法^[10]。该模型可以从全局收敛到原始优化模型的关键点。此外, 即使任务信息仅部分可用(例如没有先验正则化), 该模型依然可以保证收敛。

首先对能量模型 (2) 式的保真项引入 Moreau-Yosida 正则化项, 参数 μ^k (k 是模型迭代次数) 和辅助变量 u , 再引入辅助变量 v 和拉格朗日算子 $\tilde{\lambda}$, 从而获得新的正则化能量模型。

$$\min_x \{F(x) : f(x, y) + r(x, y)\} \quad (2)$$

式中 x 和 y 分别是观察和预测的变量, f 是保真项, r 是先验项。公式 (3)-(5) 是重写之后的能量模型公式。

$$M_F^{\mu^k}(x^k) = \min_u \{f_{x^k}^{\mu^k}(u) + r(u)\} \quad (3)$$

式中 $f_{x^k}^{\mu^k}(u) := f(u, y) + \frac{\mu^k}{2} \|u - x^k\|^2$, 接下来, 令 $\lambda^k = \frac{1}{\rho^k} (\tilde{\lambda})^k$, ρ^k 是惩罚项。

$$u^{k+1} = \arg \min_u f_{x^k}^{\mu^k}(u) + \frac{\rho^k}{2} \|u - (v^k - \lambda^k)\|^2$$

$$v^{k+1} = \arg \min_v r(v) + \frac{\rho^k}{2} \|v - (u^{k+1} + \lambda^k)\|^2$$

$$\lambda^{k+1} = \lambda^k + (u^{k+1} - v^{k+1}) \quad (4)$$

接下来, 我们使用残差公式替换公式(4), 将深度体系结构融合到上面的迭代中去, 得到

$$v_T = N_\alpha(v_0; W_T) := v_0 - \alpha(\sum_{t=0}^{T-1} G(v_t; W_t)) \quad (5)$$

式中 $W_T = \{W_t\}_{t=0}^{T-1}$ 是可学习参数集合, α 是步长。从优化角度而言, $G(v_t; W_t)$ 是基本网络单元, $(v_0; v_t)$ 分别是 N_α 的输入和输出。

最后, 为控制传播的每步迭代, 让 v^{k+1} 作为内置网络在第 k 次的输出, 并使用类似近端梯度的方式更新 x^{k+1} , x^{k+1} 的更新详情见公式(6)。

$$x^{k+1} = \arg \min_x r(x) + \frac{1}{2} \|x - (v^{k+1} - \nabla f_{x^k}^{\mu^k}(v^{k+1}))\|^2 := \text{prox}_r(v^{k+1} - \nabla f_{x^k}^{\mu^k}(v^{k+1})) \quad (6)$$

式中 prox_r 是 r 的近端算子。

团队将上述框架应用于具有单个变量的两个视觉应用(非盲反卷积和去雾)中, 以及具有两个变量的更复杂的视觉任务(低光图像增强)中, 以证明该深度模型在视觉应用领域中的有效性。部分实验结果可见图 6。

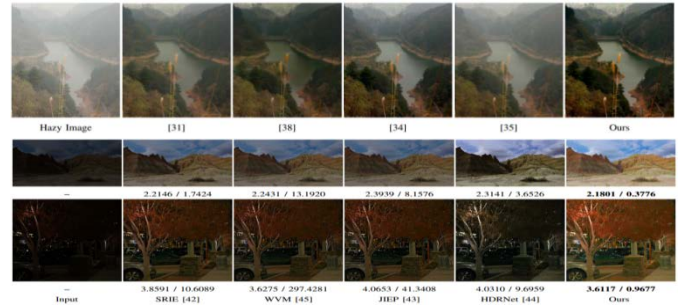


图 6 各方法在去雾、低光增强任务上实验效果对比

三、 总结

深度学习凭借大规模训练数据和五花八门的网络结构, 可以很好拟合数据分布进而在一些目标明确的任务上取得成功。但与此同时也存在着不容忽视的弊端, 其背后缺乏理论支持, 变量传播过程不可控, 此外, 与任务先验知识融合时难以保证收敛效果。模型优化是一套全面、系统的理论体系, 相较深度学习模型可充分利用任务先验知识进行数据分布拟合, 然而模型优化过于依赖手工设计先验, 因此在数据分布复杂时, 模型优化难以发挥自身优势。但两者均用于拟合数

据分布且优势互补，近年来出现很多试图将两者结合起来的的研究，介绍人团队在该面也做了较为全面、系统的探索，上述介绍的工作以及相关的工作已发表于 NIPS2018^[8]、TIP2019^[9-10]。

对于深度学习与模型优化的结合，今后可从两个方向出发。首先，可以从深度学习网络向优化方向靠拢的角度出发以拉近二者间的距离。其

次，可以从优化模型角度出发，建立特定的可学习结构以实现二者的有效结合。将深度学习与模型优化相结合，并给予收敛效果相应的理论支持，不但有利于“打开”网络黑箱，且有利于相关任务先验知识的充分利用。因此，将两者结合是一个很有必要且充满前景的研究方向。

(责任编辑：王金甲)

参考文献

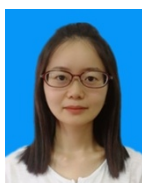
- [1] Steven Diamond, Vincent Sitzmann, Felix Heide, Gordon Wetzstein. Unrolled Optimization with Deep Priors. In CVPR, 2018.
- [2] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In CVPR, 2017.
- [3] Marcin Andrychowicz, Misha Denil, Sergio Gomez, Matthew W Hoffman, David Pfau, Tom Schaul, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In NIPS, 2016.
- [4] Risheng Liu, Xin Fan, Shichao Cheng, Xiangyu Wang, and Zhongxuan Luo. Proximal alternating direction network: A globally converged deep unrolling framework. In AAAI, 2018.
- [5] Ke Li and Jitendra Malik. Learning to optimize. In ICLR, 2017.
- [6] Karol Gregor and Yann LeCun. Learning fast approximations of sparse coding. In ICML, 2010.
- [7] Pablo Sprechmann, Alexander M Bronstein, and Guillermo Sapiro. Learning efficient sparse and low rank models. In TPAMI, 2015.
- [8] Risheng Liu, Shichao Cheng, Xiaokun Liu, Long Ma, Xin Fan, Zhongxuan Luo. A bridging framework for model optimization and deep propagation. In NIPS, 2018.
- [9] Risheng Liu, Long Ma, Yiyang Wang, Lei Zhang. Learning converged propagations with deep prior ensemble for image enhancement. In TIP, 2019.
- [10] Risheng Liu, Shichao Cheng, Long Ma, Xin Fan and Zhongxuan Luo. Deep Proximal Unrolling: Algorithmic Framework, Convergence Analysis and Applications. In TIP, 2019.



刘日升

大连理工大学副教授，主要研究方向为机器学习、深度学习、优化方法、计算

机视觉。Email: rslu@dlut.edu.cn



程世超

大连理工大学博士生，主要研究方向为计算机视觉、图像处理。

Email: shichao.cheng@outlook.com



马龙

大连理工大学博士生，主要研究方向为计算机视觉、优化方法。

Email: malone94319@gmail.com



付陈平

大连理工大学博士生，主要研究方向为计算机视觉、图像处理。

Email: fu.2019@outlook.com



樊鑫

大连理工大学教授，主要研究方向为计算机视觉、图像处理、医学影像分析。

Email: xin.fan@dlut.edu.cn



罗钟铤

大连理工大学教授，主要研究方向为计算几何与图形 / 图像、计算机视觉。

Email: zxluo@dlut.edu.cn

面向视频数据的对抗攻击研究

北京航空航天大学 韦星星

深度学习已经在很多领域取得了巨大成功，但是最新的研究成果表明深度学习非常容易受到对抗样本的干扰。所谓对抗样本是指在干净图像上添加一些肉眼不可见噪声，原本会被深度学习模型分类正确的图像就会被预测出错误结果。原本可被深度学习算法以 57.7% 置信度正确分类为“panda”的图片，在加上噪声后，却被深度学习算法以 99.3% 的置信度错误分类为了“gibbon”。

由于对抗样本的巨大危害，很多研究者从不同角度对此进行了研究，如不同对抗样本生成方法、各式各样对抗样本防御方法以及对抗样本产生内在机理等。但这些研究基本都集中在图像数据上面，面向视频数据的对抗攻击与防御并未涉及。相比图像数据，视频因为其特有时序信息，必然导致其与面向图像数据的对抗样本研究有巨大不同。下面介绍我们组两个相关工作。

首先是面向视频分类任务的对抗样本生成方法，图像数据由于其天然的时序结构信息，导致在生成对抗视频的时候不能对其每一帧按照图像的方法来进行添加噪声。相反，我们认为视频对抗样本应该具有传播性和稀疏性两个特点，稀疏性是指只需要在视频中的一些关键帧上添加噪声即可，传播性是指这些稀疏的对抗噪声可以利用帧间的时序关系来传播到未添加噪声的帧上面，从而污染整个视频片段，进而达到欺骗深度学习模型的目的。如图 1 所示，蓝实线是对视频中每帧添加噪声的度量，可以看到随着视频帧的变化，噪声幅度呈现显著下降趋势，最终趋于 0，但是却成功欺骗神经网络输出了错误的标签（见红色字）。该成果发表在 AAI2019 上面。

第二个工作是使用对抗本来攻击图像和视频的物体检测算法。这个方法具有两个特点，

一是迁移性好，我们的对抗样本可以有效攻击 Faster-RCNN 等基于分类机理的物体检测算法，还可以攻击 SSD, YOLO 等基于回归机理的物体检测算法。第二个特点是高效，不同于以往的基于优化方式来获得对抗样本，我们采用生成式对抗网络，由于在测试过程只涉及前馈网络，因此其生成速度非常快，因此可以处理图像数据还可以快速处理视频每一帧。图 2 展示了该方法的结构框图，其中的高迁移性是通过一个特征损失函数来破坏物体检测器的特征图，由于 Faster-RCNN 和 SSD 都基于特征图来进行物体检测，因此将特征图破坏后将同时使得这两类物体检测器失效。

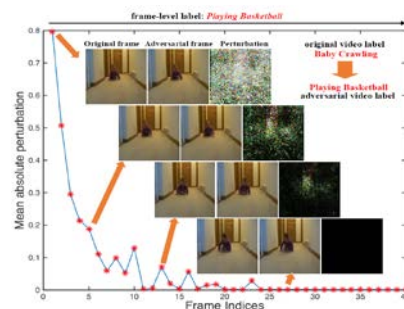


图 1 面向视频数据的可传播的稀疏对抗噪声

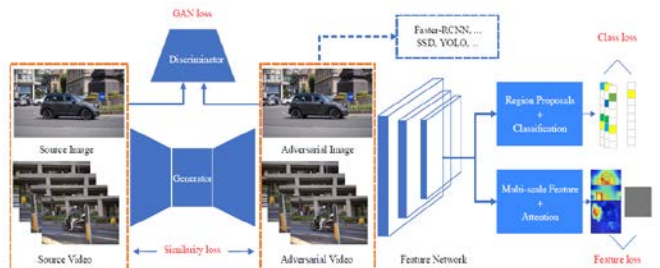


图 2 基于特征图篡改的高迁移性视频对抗样本生成方法

(责任编辑：苏航)



韦星星

北京航空航天大学计算机学院副教授，主要研究为人工智能安全，计算机视觉等。

Email: xxwei@buaa.edu.cn

行人再识别相关研究进展

中科院自动化所模式识别国家重点实验室 杨阳 吴锦林 雷震

行人再识别起源于多摄像机跟踪任务，解决不同摄像机下行人之间相互关联的问题。在 2006 年，研究人员把它作为一个特定的任务并以图像形式进行探讨和研究。随着时间推移，研究对象从单纯基于图像扩展到基于视频，以及跨模态的图像与文本和可见光与近红外。目前，基于图像的行人再识别仍是研究主流。它所面临的挑战包括：(1) 单个摄像机下行人的图片分辨率较低、遮挡以及背景干扰；(2) 不同摄像机下获取的图片存在光照变化、行人姿态各异以及摄像机视角不相同等问题。针对这些问题，早期研究人员主要通过图像特征的提取(颜色、纹理、属性语义、高阶混合、多层描述子)、特征学习(特征编码、特征降维、子空间学习)、相似度学习(距离度量、双线性度量、联合度量)和重排序。自深度学习的兴起，研究人员着手于设计端到端的网络，研究不同的损失函数(分类损失、三元组和四元组)、风格迁移，小样本学习，神经网络架构搜索，以及两幅图像的配准(姿态矫正)、分辨率低和背景干扰等问题。公开数据库(Market1501)的 Rank-1 指标一再被刷新，从最初 25% 上升到 96%。考虑到行人数据涉及隐私以及数据内容难以包含各种场景，有部分研究人员在近期通过游戏引擎，建立新的虚拟数据。一方面可以增加训练数据，辅助行人再识别性能的提升；另一方面通过限定训练和测试条件，研究不同因素对最终性能的影响，为实际场景的应用设计提供参考。

在实际应用中，有监督的行人再识别模型需要耗费大量的人力和物力对不同的应用场景重新采集、标注数据进行训练。因此无需新标注数据训练的无监督行人再识别算法，逐渐引起研究人员的关注。在 ECCV2018 和 PAMI2019 上，龚少刚团队基于现有跟踪算法和稀疏时间采样自动得到单个摄像机下的行人轨迹图像，对每个摄像机的所有行人轨迹随机打上伪标签(见图 1)，并

提出了多任务训练的框架和潜在轨迹关联的方法，通过学习潜在的正样本来提升模型的跨视角检索能力。但稍显不足之处在于它们仅仅在训练时输入的小批量数据中挖掘潜在正样本，效率较低，并且为了在输入的小批量数据中采样到潜在正样本，每一步训练需要采样比较多的数据输入到网络，导致训练时占用特别大的显存。在 ECCV2018 另一篇工作中，阮邦志团队提出一种行人轨迹渐进合并的方法，将潜在的属于同一个人轨迹的图像进行合并，使用自训练的方法训练行人再识别模型。但是，如果将不属于同一个人的轨迹合并会误导模型，且这种合并是不能撤回的，限制了无监督行人再识别的性能提升。

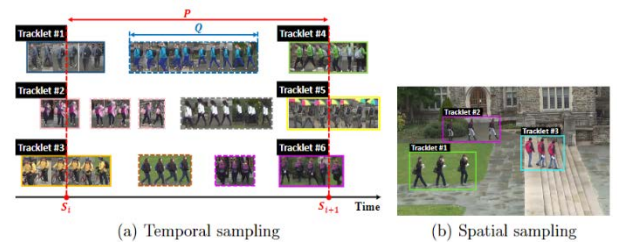


图 1 轨迹伪标签打印示意图

针对上述问题，我们提出了一种无监督图关联的方法和二阶段训练的策略(见图 2)。

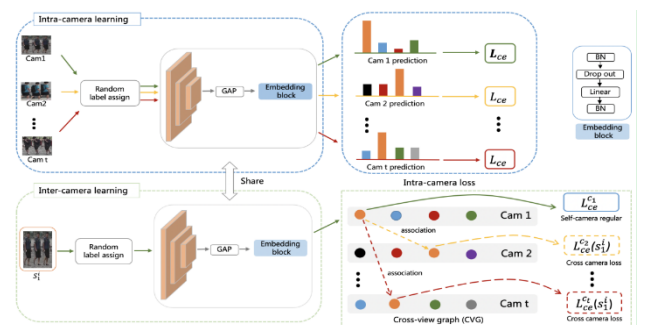


图 2 无监督图关联和二阶段训练策略

第一阶段采用单个摄像机中行人轨迹的伪标签和多任务训练的策略，使模型专注于学习单个视角内的行人图像的特征表示；第二阶段使用图关联的方法挖掘跨视角的潜在正样本，使模型专注于学习跨视角行人图像的特征表示。两个阶

段训练的策略将视角内训练和跨视角训练分开，减轻跨视角学习时噪声样本带来的干扰。图关联的方法主要是通过引入最近邻，跨视角和边对称的约束，建立一个较为精确的跨视角行人轨迹关联图。相比于已有工作，能够过滤掉较多的噪声样本对。对于跨视角行人轨迹关联图中挖掘到的潜在正样本对，我们提出了一种图加权的跨视角损失函数来提升模型的跨视角检索能力。

表 1 基于图像的无监督行人再识别算法对比

| Datasets | Reference | PRID2011 | | | iLIDS-VID | | | MARS | | | |
|----------------------|-----------|----------|------|------|-----------|------|------|------|------|------|------|
| | | R1 | R5 | R20 | R1 | R5 | R20 | R1 | R5 | R20 | mAP |
| SMP [24] | ICCV'17 | 80.9 | 95.6 | 99.4 | 41.7 | 66.3 | 80.7 | 23.9 | 35.8 | 44.9 | 10.5 |
| DGM+MLAPG [41] | ICCV'17 | 73.5 | 92.6 | 99.0 | 37.1 | 61.3 | 82.0 | 24.6 | 42.6 | 57.2 | 11.8 |
| DGM+IDE [41] | ICCV'17 | 56.4 | 81.3 | 96.4 | 36.2 | 62.8 | 82.7 | 36.8 | 54.0 | 68.5 | 21.3 |
| DASy [1] | ECCV'18 | 43.0 | - | - | 56.5 | - | - | - | - | - | - |
| GRDL [13] | ECCV'16 | 41.6 | 76.4 | 89.9 | 25.7 | 49.9 | 77.6 | 19.3 | 33.2 | 46.5 | 9.56 |
| DTW [26] | PR'17 | 41.7 | 67.1 | 90.1 | 31.5 | 62.1 | 82.4 | - | - | - | - |
| BUC [23] | AAAI'19 | - | - | - | - | - | - | 61.1 | 75.1 | 80.0 | 38.0 |
| EUG($p=0.05$) [39] | CVPR'18 | - | - | - | - | - | - | 62.7 | 74.9 | 82.6 | 42.5 |
| RACE [40] | ECCV'18 | 50.6 | 79.4 | 91.8 | 19.3 | 39.3 | 68.7 | 43.2 | 57.1 | 67.6 | 24.5 |
| TAUDL [17] | ECCV'18 | 49.4 | 78.7 | 98.9 | 26.7 | 51.3 | 82.0 | 43.8 | 59.9 | 72.8 | 29.1 |
| UTAL [18] | TPAMI'19 | 54.7 | 83.1 | 96.2 | 35.1 | 59.0 | 83.8 | 49.9 | 66.4 | 77.8 | 35.2 |
| UGA(ours) | This work | 80.9 | 94.4 | 100 | 57.3 | 72.0 | 87.3 | 58.1 | 73.4 | 81.4 | 39.3 |

1-st and 2-nd best results are in red/blue respectively.

我们在 7 个公开的行人重识别数据库进行了实验，均取得较好的结果；并且和现有轨迹关联方法对比，我们方法在基于图像和基于视频的无监督行人再识别任务上取得了非常大的提升(见表 1 和 2)。工作发表于 ICCV 2019。

表 2 基于视频的无监督行人再识别算法对比

| Dataset | Reference | Method | Market1501 | | DukeMTMC-ReID | | CUHK03 | | MSMT17 | |
|---------------|-----------|------------|------------|--------|---------------|--------|--------|--------|--------|--------|
| | | | mAP | Rank 1 | mAP | Rank 1 | mAP | Rank 1 | mAP | Rank 1 |
| HHL [49] | ECCV'18 | GAN | 31.4 | 62.2 | 27.2 | 46.9 | - | - | - | - |
| SPGAN [4] | CVPR'18 | GAN | 22.8 | 51.5 | 22.3 | 41.1 | - | - | - | - |
| SPGAN+LMP [4] | CVPR'18 | GAN | 26.7 | 57.7 | 26.2 | 46.4 | - | - | - | - |
| TJ-AIDL [36] | CVPR'17 | adaptation | 26.5 | 58.2 | 23.0 | 44.3 | - | - | - | - |
| BUC [23] | AAAI'19 | cluster | 38.3 | 66.2 | 27.5 | 47.4 | - | - | - | - |
| CAMEL [43] | ICCV'17 | cluster | 26.3 | 54.5 | - | - | - | 39.4 | - | - |
| PUL [5] | ToMM'18 | cluster | 20.1 | 44.7 | 16.4 | 30.4 | - | - | - | - |
| CDS [16] | ICME'19 | cluster | 39.9 | 71.6 | 42.7 | 67.2 | - | - | - | - |
| TAUDL [17] | ECCV'18 | tracklet | 41.2 | 63.7 | 43.5 | 61.7 | 31.2 | 44.7 | 12.5 | 28.4 |
| UTAL [18] | TPAMI'19 | tracklet | 46.2 | 69.2 | 44.6 | 62.3 | 42.3 | 56.3 | 13.1 | 31.4 |
| ECN [50] | CVPR'19 | adaptation | 43.0 | 75.1 | 40.4 | 63.3 | - | - | 10.2 | 30.2 |
| UGA(ours) | This work | tracklet | 70.3 | 87.2 | 53.3 | 75.0 | 68.2 | 56.5 | 21.7 | 49.5 |

1-st and 2-nd best results are in red/blue respectively.

(责任编辑：任传贤)



杨阳

中国科学院自动化研究所助理研究员。主要研究领域为计算机视觉，目

前侧重于行人再识别、属性分析及人脸识别的相关应用研究。

Email: yang.yang@nlpr.ia.ac.cn



吴锦林

中国科学院自动化研究所在读博士。主要研究领域包括半监督学习算法，

自动机器学习。

Email: jinlin.wu@nlpr.ia.ac.cn



雷震

中国科学院自动化研究所研究员。主要研究领域为生物特征识别和智能视

频分析。Email: zlei@nlpr.ia.ac.cn

武汉大学夏桂松教授访谈

2019年9月3日，专委秘书处采访了武汉大学测绘遥感信息工程国家重点实验室博士生导师，2019年国家优秀青年科学基金获得者夏桂松教授。下面是采访实录。

夏老师，您好！您在计算机视觉、遥感图像分析与模式识别领域做出了很多高水平的研究工作，发表了一系列的高水平论文，并获得了2019年国家优秀青年基金资助。您能跟我们详细介绍一下您的研究经历，分享一下获得成功的经验，以及您取得这些成就的动力吗？

非常感谢专委会提供这样一个平台，让我们有机会分享自己的研究经历。我本科念的是电子信息、硕士做的是遥感图像处理，到巴黎读博士学的是应用数学和图像理解，目前主要从事计算机视觉和遥感领域的交叉研究，而综合来看，我更偏向定位自己是一个计算机视觉的从业人员。回顾个人的求学和工作经历，我其实一直聚焦在信息处理的交叉研究领域。回国工作以后，我加入了武汉大学测绘遥感信息工程国家重点实验室。武汉大学的测绘遥感有非常深厚的积累，在世界上处于领先地位。可能很多计算机视觉的学者不了解，我们核心方向之一的摄影测量和遥感，研究的就是如何通过远距离获取物体影像来量测物体几何和物理属性。从核心算法来看，多视几何与计算机视觉是我们的基础。两个领域从某种程度上可以说是“同出而异名”。特别是，最近十几年，随着对地观测技术的快速发展，成像分辨率迅速提高、遥感平台越来越多样、数据量呈指数增长，遥感信息处理的传统理论和方法受到很大的挑战，和计算机视觉、数据挖掘等领域的交叉也越来越深。由于之前在武汉大学学习6年，耳濡目染，我对这个交叉领域的发展可能更敏感。2013年起，通过逐步理清遥感图像解译和计算机视觉的交叉研究前沿和突破点，目前我们已经形

成了计算视觉和摄影测量研究组（CAPTAIN），在图像结构建模、高分辨率遥感图像解译（主要包括目标检测、语义分割、变化分析）、无人系统信息处理等领域已经取得了一些成绩，初步解决了这个方向的一些问题。动力主要来自两方面：一方面，通过向不同领域的顶尖专家学习，尝试从不同的角度去思考、分析和解决问题，这本身就是一件比较好玩的事情；另一方面，我比较喜欢探索未知的过程，能在解决新问题的过程中找到成就感和归属感，同时也比较享受分享自己研究成果的喜悦。

您在结构纹理统计建模和分析方面研究的比较深，关于这一块您能谈谈该领域的一些研究现状和未来发展的一些方向和趋势么？

我很喜欢 low-level 视觉模式的建模和计算问题，一直坚持在做这方面的研究。在您说的这个方向上，我主要感兴趣的是如何从统计的角度处理、分析和理解在视觉上具有复杂几何结构模式的纹理现象，寻求和设计有效的数学工具为这一过程建模。在深度学习爆发以前，这方面的工作主要集中在对图像的局部结构统计建模上；深度学习出现以后，大家发现非线性的深度描述模型在这个问题上能取得前所未有的效果，特别是有 GAN、BigGAN 这些算法。个人认为未来在这个问题的研究上，大家还应该做更多的可解释性的工作，从原理上或统计上去理解这些视觉现象。当然，在应用上我们也面临很多挑战，比如怎么建模在空间上不平稳的多尺度纹理现象。

请问基于什么样的出发点，您决定采集和公布 DOTA 数据集呢？能跟大家分享一下这个数据库的采集过程，以及采集经验么？

首先我个人比较关注 reproducible research。

最初涉入遥感图像信息处理的时候，我发现一方面通用的测试数据集非常少，算法泛化能力没有办法得到充分的测试；另一方面这个任务上的很多工作没有公开测试的源码，如果想和别的算法比较需要大家自己实现，难度也不小。特别是当深度学习出现以后，遥感算法的研究明显缺乏大规模的公共标注数据。因此，2014年起，我们就决定建立一个公开数据集，并且提供相应的公共测试平台，其中就包括您说的DOTA，其实我们还有AID、GID等其他数据集。要构建这样的数据集和平台对我们小的研究团队来说并非易事，首先要深入了解问题需求，协调遥感数据资源，然后研发标注工具，同时也要有专业经验的标注人员；在财力、物力、人力上都是不小的挑战。但是，这其实是一个公益事业，做好了有利于整个行业的发展，我们希望也呼吁更多的企业或者行业单位能参与进来一起完成这个事情，推动遥感对地观测研究社区的发展。

您从2012年底至今一直工作于武汉大学测绘遥感信息工程国家重点实验室，能介绍一下您在此期间的工作经历吗，有哪些经历是您特别难忘的？

武汉大学测绘遥感信息工程国家重点实验室连续20年、4次科技部考评优秀。入职以来，我从副教授到教授，从硕导到博导，一晃就是7年，感受最深刻的应该是两方面：一是实验室浓厚的学术氛围和开放自由的学术环境。为开拓研究生的学术视野，实验室创立了GeoCaffee，坚持每个月邀请各个领域的专家来讲座，涉及的范围除了相关专业领域还包括文学、哲学等，覆盖范围非常广，目前应该已经举办了40多期。二是，实验室从上到下洋溢的国家情怀和奋斗精神。实验室院士、长江杰青等学者带头，日以继夜地奋斗在科研的第一线，让人特别受感染，激励每个人不断前行。

能分析一下武汉大学的研究氛围，和您之前在法国的研究氛围的异同吗？

中国和法国都有很深的文化积淀。武汉大学和我在法国学习和工作过的地方，在研究氛围和学术环境上最大的共同点，我觉得一是开放精神、能够包容和孕育不一样的学术思想；二是浪漫情怀，一个踞珞珈之秀、一个聚欧洲之美，周围的人似乎天生都有一股浪漫和洒脱的气质。当然，我们的同学更努力、更勤奋。

能介绍一下您现在的研究团队，以及您如何管理这个研究团队吗？

借助武汉大学测绘遥感信息工程国家重点实验室这个平台，从2012年起，我们逐步建立了计算视觉和摄影测量研究组，简称CAPTAIN。研究组主要聚焦在计算机视觉、摄影测量与遥感的交叉研究领域，为测量、分析和理解我们所处的视觉世界研究相应的理论、模型与算法。CAPTAIN的成员主要来自数学、电信、测绘遥感等不同领域，大家背景各异，相互学习，围绕同一目标从不同角度去分析和解决问题。和很多研究组一样，CAPTAIN也有详细的管理制度，主要目标是保持组内浓厚的学习氛围，调动同学们的研究积极性，让大家学有所成。我们每周都有学术例会，同时结合同学们的研究进展不定期一对一讨论，争取让每个人都能找到自己的位置和发展方向。

现在很多学生做研究比较盲然，也比较浮躁，您能给学生们一些建议吗？

的确存在您说的这个现象，但是这很大程度上也不能全怪我们同学。在大数据时代，我们每天都面临着来自各种媒介的海量信息轰炸，要迅速辨别这些信息的虚实，从里面找到方向，并且坚守下来，不是一件容易的事情。我的建议是同学们瞄准自己感兴趣的研究问题和目标，在关注和吸收这些海量信息的同时尽量做减法，聚焦自己的研究课题，静下心来，踏实认真地把手上的每一个研究做好、做透。弱水三千、只取一瓢。

对于研究工作者，事业和家庭都要兼顾是有

一定难度的，能否跟大家介绍一下您是如何处理事业与家庭之间的关系？

研究工作需要我们持续的投入、家人也需要我们的陪伴，怎么样平衡这两方面，我也一直在学习和摸索。我个人的做法是在安排自己工作的时候和家人保持密切的沟通，尽可能地做好分工，努力去承担自己在家庭中的责任，相互理解、互相支持、互相配合，和所有家人一起把家庭经营好。



夏桂松

男，1983年生，武汉大学测绘遥感信息工程国家重点实验室教授，博士生导师。2005年和2007年先后在武汉大学获得学士和硕士学位，2011年3月获得法国巴黎高科电信学院

(Telecom ParisTech) 博士学位。2011年4月至2012年12月先后在法国巴黎高科电信学院 LTCI 实验室、法国国家科学研究中心 (CNRS) -决策数学研究所 (CEREMADE) 从事博士后研究。2012年12月起在武汉大学测绘遥感信息工程国家重点实验室工作至今。长期从事图像分析和理解以及模式识别等领域的研究，在包括 IJCV, IEEE TPAMI/TIP/TMM/TGRS, PR, CVIU 等国际权威期刊和 CVPR、ECCV、ICIP、ICPR 等相关国际会议上发表学术论文 100 余篇。现担任国际期刊 Pattern Recognition, EURASIP Journal on Image and Video Processing 和 Signal Processing: Image Communications 等 3 个国际期刊副主编 (Associate Editor)。入选湖北省楚天学者计划，获得国家自然科学基金优青、湖北省自然科学基金杰青等人才项目资助，获湖北省自然科学二等奖（排名 1）、“第二届中国科协优秀科技论文”奖等奖励。

（责任编辑：余焯 黄岩 张汗灵）

委员好消息

✪ 2019年7月23日, CCF-CV 专委会委员、复旦大学张军平教授所著的《爱犯错的智能体》由清华大学出版社付梓。该书从开脑到烧脑, 别出心裁的从犯错的角度科普了人工智能概念, 详细介绍了视听觉感知下智能体犯错机制, 并强调了犯错机制对智能体发育的重要性。书稿内容此前发表在张军平教授的博客上, 在科学网连续 20 余次头条推荐, 在中国工程院院刊、自动化学会、新智元等微信公众号和搜狐等被广泛转载。

✪ 2019年7月25日, 150 位在各自领域做出一定成就或对 CCF 服务方面有突出表现的专业会员晋升为高级会员, 其中包括 CCF-CV 专委会 8 位委员: 北京航空航天大学黄迪副教授、中国石油大学(华东)刘伟锋教授、清华大学鲁继文副教授、中科院计算所王瑞平研究员、北京工业大学毋立芳教授、中科院自动化所严冬明研究员、南开大学杨巨峰副教授、复旦大学张文强教授。

✪ 2019年8月16日, 2019年度国家优秀青年

科学基金项目资助名单出炉, 本年度共资助 600 项, CCF-CV 专委会共 8 位委员获得优青项目资助: 南开大学程明明教授(图像场景理解)、北京航空航天大学李甲副教授(视觉显著内容计算与解析)、中科院自动化所刘静研究员(图像语义理解)、大连理工大学刘日升副教授(面向视觉与学习的高维最优化)、北京邮电大学马占宇副教授(数据的非高斯建模与分析)、南京理工大学潘金山教授(图像复原与增强)、中科院计算所王瑞平研究员(视觉数据非线性建模与度量学习)、武汉大学夏桂松教授(高分辨率遥感图像智能分析和解译)。

✪ 2019年8月16日, 2019年国家自然科学基金评审结果出炉, 其中 57 家单位 82 个国家重大科研仪器项目得到资助, CCF-CV 专委会委员、中国海洋大学董军宇教授的“水下光学高分辨率三维成像仪研究”项目获得 620 万元的资助。

(责任编辑: 刘海波)

视频描述开源代码

北京电子科技学院 李熹桥 金鑫

视频描述在人机交互、视频索引等方面有着重要的应用。视频描述任务可以理解为一个将视频图像序列转换到文本序列的任务。

这项任务目前的数据库大都是由剪辑后的视频和相对应的描述构成的。主要的难点在于理解各种各样的对象、场景、行为和属性以及结合上下文恰当地描述视频。早先的研究一般由两个阶段组成，首先是识别主语、动词、宾语等语义内容，然后基于模板生成相应的描述。这类方法生成的描述与其使用的语言模型有很大的关系，不具备自然语言的丰富性。

Input video:



Our output: A cat is playing with a toy.

Humans: A Ferret and cat fighting with each other. / A cat and a ferret are playing. / A kitten is playing with a ferret. / A kitten and a ferret are playfully wrestling.

图 1 视频描述任务示意图

近年来视频描述的网络大都采用 LSTM 来构造编码器-解码器结构。网络的输入主要包括视频图像的特征、视频声音的特征、视频先验特征等等。输出描述主要包括使用词向量的方法或者直接使用词袋两种方法。

视频描述任务在近年来又延伸出密集事件描述任务，提出的主要原因是虽然视频描述方法在描述短视频时候效果还不错，但对于更多的对象动作以及场景事件，单句无法表达长视频的全部信息。

本文着重介绍几个基于深度学习的视频描述技术的开源代码，包括训练深度学习模型的常用数据集和几个经典的视频描述深度神经网络的模型。

1. MSR-VTT 数据集

介绍: 该数据集包含 20 类，共 10000 个视频片段，分为训练，验证和测试集三部分。每个视频片段对应约 20 条英文句子标注以及该视频类别信息。

主页: <http://ms-multimedia-challenge.com/2017/challenge>

2. MSVD dataset 数据集

介绍: 该数据集包含 1970 段 YouTube 视频片段（时长在 10-25s 之间），每段视频标注了大概 40 条英文句子。

主页: <https://www.microsoft.com/en-us/download/details.aspx?id=52422>

3. Charades 数据集

介绍: 该数据集包含 9848 个视频，平均长度为 30 秒。其中每个视频都由多个文本描述、动作标签、动作间隔和交互对象的类进行注释。总共，数据集提供 27847 个视频描述，157 个动作类和 41104 个标签，46 个对象类。

主页: <http://allenai.org/plato/charades>

4. VATEX 数据集

介绍: 该数据集包含针对 41,300 个视频片段，每个片段包括 10 个英文描述和 10 个中文描述。这些视频共涵盖 600 种人类活动。VATEX-en、VATEX-zhMSR-VTT 的平均句子长度分别为 15.23、13.95 和 9.28。可以看出相比 MSR-VTT，VATEX-en 和 VATEX-zh 中的描述句子都更长，具备更广泛的描述风格，覆盖更多的动作、物体和视觉场景。

主页: <https://vatex.org/main/index.html>

5. ActivityNet Captions 数据集

介绍: 该数据集包含 20000 个视频，每个视频平均含有 3.65 个时序定位的描述语句，一共

有 100000 条描述。数据集中每个视频的语句数量相对服从正态分布。句子的平均长度为 13.48 个词，也符合正态分布。平均每个句子描述了 36 秒的事件，大约是相应视频的 31% 的内容。每个视频的完整语句描述了视频中大概 94.6% 的内容，每一段标注基本都能覆盖视频内的主要活动。

主页：<https://cs.stanford.edu/people/ranjaykrishna/densevid>

6. ActivityNet Entities 数据集

介绍：该数据集是基于 ActivityNet Captions 数据集而建立的，在此基础上增加了 15.8 万个边框标注，并且每个边界框都有对应的名词短语。



图 2 数据集标注示例

地址：<https://github.com/facebookresearch/ActivityNet-Entities>

7. MPII-MD 数据集

介绍：该数据集包括 94 部高清电影的剪辑片段和 68375 条相对应的描述。

主页：<https://www.mpi-inf.mpg.de/departments/computer-vision-and-multimodal-computing/research/vision-and-language/mpii-movie-description-dataset>

8. M-VAD 数据集

介绍：该数据集包括 63000 条剪辑片段和对应的 34000 条描述。数据集中的描述使用适当的字符替换了“某人”，使得描述与角色身份相关联。

地址：<https://github.com/aimagelab/mvad-names-dataset>

9. Sequence to Sequence - Video to Text

论文：Venugopalan S, Rohrbach M, Donahue J, *et al.* Sequence to sequence-video to text[C]. IEEE International Conference on Computer Vision, 2015: 4534-4542.

工作：该论文采用长短时记忆网络 LSTM 从视频片段和与之相对应的文字描述中学习显著的内容并输出相应的描述。论文提出了一种 S2VT 模型，模型是由两个 LSTM 网络叠加而成。第一个 LSTM 将通过卷积神经网络提取到的帧特征向量作为输入进行逐个编码。一旦读取完所有的帧，模型会逐个单词的生成一个句子。帧和单词表示的编码和解码工作由平行语料库(微软视频描述语料库(MSVD), MPII 电影描述语料库(MPII-MD), 蒙尔特特诗品注释数据集(M-VAD))中学习得到。为了更好地表示视频中活动的时序特点，论文中还计算了连续帧之间的光学流，流图像也是先通过 CNN 网络并作为输入提供给 LSTM 网络。相比之前视频描述任务，本文的主要贡献是 S2VT 结构是在整个视频的帧完整编码之后才产生对应的描述。图 3 为 S2VT 模型示意图。

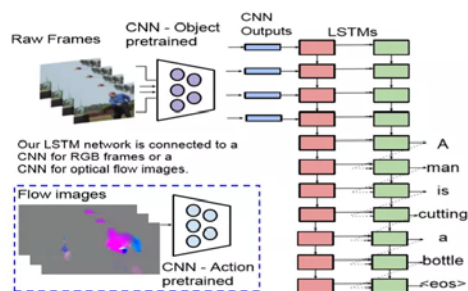


图 3 S2VT 模型示意图

论文：https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Venugopalan_Sequence_to_Sequence_ICCV_2015_paper.pdf

代码：<https://github.com/Rangozhang/VideoCaption>

10. Describing Videos by Exploiting Temporal Structure

论文：Yao L, Torabi A, Cho K, *et al.* Describing videos by exploiting temporal structure[C]. IEEE International Conference on Computer Vision, 2015: 4507-4515.

工作：论文的主要贡献是利用视频的全局与局部结构来产生描述。视频的局部结构指的是动

作或事件仅在视频几个连续帧变化，比如接听电话，站起来等动作；视频的全局结构是指对象，动作，场景在视频中出现的顺序。具体实现该论文先对视频不同帧分别提取特征，由此组成一个特征向量。在每次预测单词的时候使用注意力机制，计算特征向量中每个特征的权重，进而预测当前单词。论文中提取特征采用的是二维特征与三维特征的融合。三维特征是通过通过对视频提取三种人工设计特征（HOG、HOF、MBH），再将得到的三种人工特征块输入三维卷积网络得到的。图 4 为局部结构示意图，图 5 为全局结构示意图。

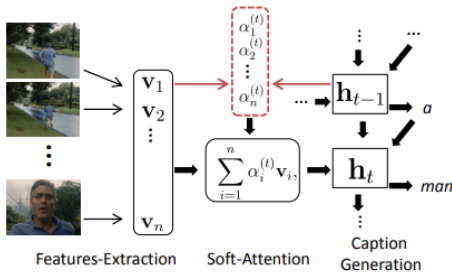


图 4 局部结构示意图

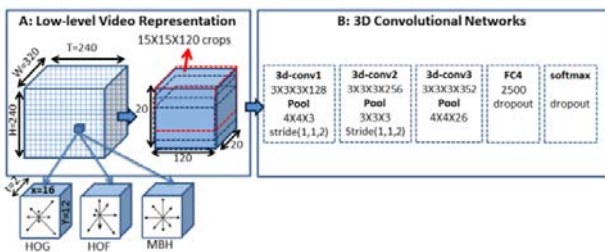


图 5 全局结构示意图

论文: https://www.cv-foundation.org/openaccess/content_iccv_2015/papers/Yao_Describing_Videos_by_ICCV_2015_paper.pdf

代码: <https://github.com/yaoli/arctic-capgen-vid>

11. Reinforced video captioning with entailment rewards

论文: Pasunuru, Ramakanth, Bansal. Reinforced video captioning with entailment rewards[J]. Conference on Empirical Methods in Natural Language Processing, 2017:979-985.

工作: 论文的主要贡献是使用强化学习方法解决交叉熵损失中对数似然方法的局限性这一问题。图 6 为该方法模型示意图，该方法使用损失函数为混合损失函数，如公式 1 所示：

$$L_{MIXED} = (1 - \gamma)L_{XE} + \gamma L_{RL} \quad (1)$$

式中参数 γ 是一个用来平衡两种损失的超参数。论文中还提出了对 CIDEr 评价标准进行提升的 CIDEr 奖励机制，最终提升了模型效果。

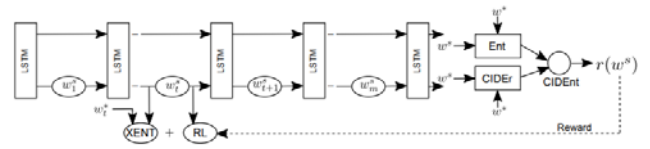


图 6 算法示意图

论文: <https://www.aclweb.org/anthology/D17-1103>

代码: https://github.com/ramakanth-pasunuru/video_captioning_rl

12. Consensus-based Sequence Training for Video Captioning

论文: Phan S, Henter G E, Miyao Y, et al. Consensus-based sequence training for video captioning[J]. arXiv preprint arXiv:1712.09532, 2017.

工作: 论文指出当前有的视频描述模型是用交叉熵损失进行训练的，而实际评估的时候用的是其他指标比如 BLEU, METEOR 等。由此论文提出基于一致性的序列训练来针对这一问题。图 7 为算法总体示意图。

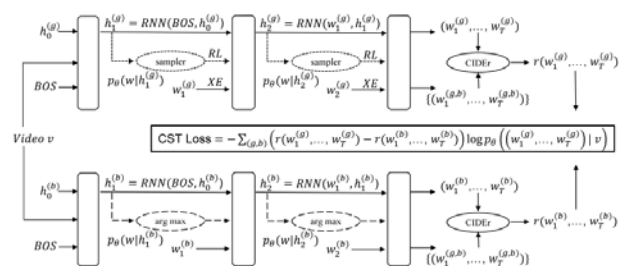


图 7 CST 总体示意图

论文: <https://arxiv.org/abs/1712.09532>

代码: https://github.com/mylnp/cst_captioning

13. Grounded Video Description

论文: Zhou L, Kalantidis Y, Chen X, et al. Grounded video description[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 6578-6587.

工作: 论文指出当前视频描述方法常常使用

先验信息而导致生成出的描述不一定与视频相匹配。在论文中提出了一个新的密集事件描述数据集 activitynet-entities 数据集。其中包含 15.8 万个边框标注以及相对应的描述。使用这个数据集进行训练可以使生成的描述可以对应视频中的物体。具体而言，论文中模型分为三个部分：基础部分，区域注意部分以及描述生成部分。在基础部分，先提取视频的视觉特征生成特征图，然后在特征图上预测边框位置和对边框内物体分类。最后将边框位置信息和类别信息融合作为区域特征输入注意力部分。在区域注意部分，使用自注意力机制筛选区域特征，然后将选出的区域特征输入描述生成部分。描述生成部分是通过 LSTM 生成最终的描述。图 8 为模型示意图。损失函数如公式 2 所示：

$$L = L_{sent} + \lambda_c L_{cls} + \lambda_\beta L_{grd} + \lambda_\alpha L_{attn} \quad (2)$$

式中整个网络的损失函数包括四个部分，生成描述的损失，边框内目标分类的损失，回归区域位置的损失以及自注意力的损失。

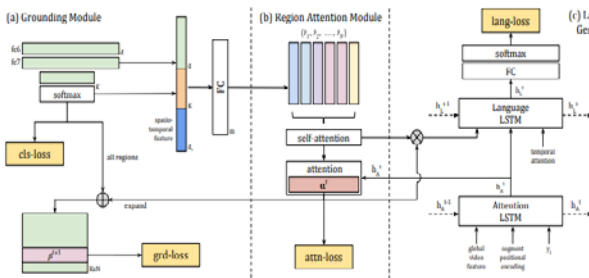


图 8 Video Description 模型示意图

论文: http://openaccess.thecvf.com/content_CVPR_2019/papers/Zhou_Grounded_Video_Description_CVPR_2019_paper.pdf

代码: <https://github.com/facebookresearch/grounded-video-description>

14. Dense-Captioning Events in Videos

论文: Krishna R, Hata K, Ren F, *et al.* Dense-captioning events in videos[C]. IEEE International Conference on Computer Vision, 2017: 706-715.

工作: 该论文指出当前的视频描述工作对于

一个视频只生成一句话来描述，缺少了视频中的一些细节。基于这一点，论文提出了密集事件描述任务，该方法能够生成一系列对于视频中出现的多个事件的描述，并且在时间上对他们进行定位。方法主要分成三部分：视频特征提取，视频事件推选模块和描述模块。视频特征提取是从输入视频中提取 C3D 特征。这些特征随后输入视频事件推选模块，进行不同步长的采样，然后将它们输入到 LSTM 单元。步长越长，能捕捉到的事件就越长。最后，结合视频上下文的描述模型可以产生每个事件的描述。图 9 为该算法的完整框图。

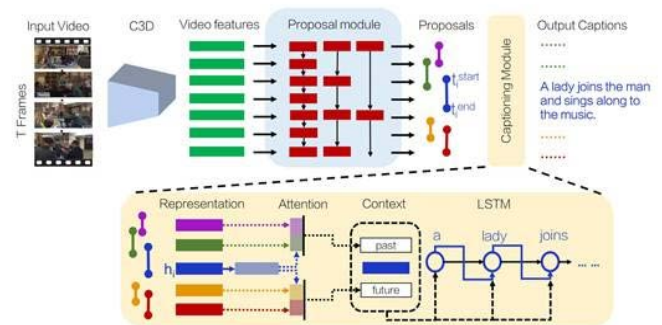


图 9 视频密集事件描述算法框图

论文: http://openaccess.thecvf.com/content_ICCV_2017/papers/Krishna_Dense_Captioning_Events_in_ICCV_2017_paper.pdf

代码: https://github.com/ranjaykrishna/densevid_eval

15. Bidirectional Attentive Fusion with Context Gating for Dense Video Captioning

论文: Wang J, Jiang W, Ma L, *et al.* Bidirectional attentive fusion with context gating for dense video captioning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7190-7198.

工作: 该论文指出了视频描述的两个主要挑战：如何利用视频时序信息更好定位事件以及如何给解码器提供有效的视觉信息来生成特定的语言描述。论文作者提出了一种双向提取事件候选区间的方法。具体来说，是通过注意力机制将事件定位模块中的隐状态与视频原始内容（例如，视频 C3D 特征）结合起来共同表征当前的事件。论文中作者还提出一种新颖的上下文门控机制

来平衡当前事件内容和它的上下文对生成文字描述的比重。图 10 为所提算法模型示意图。

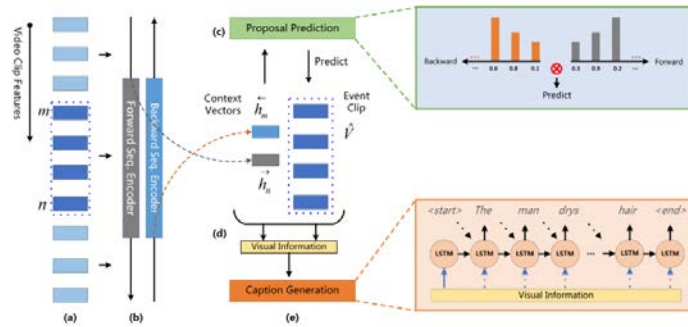


图 10 Dense Video Captioning 模型示意图

论文：http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang_Bidirectional_Attractive_Fusion_CVPR_2018_paper.pdf
 代码：<https://github.com/JaywongWang/DenseVideoCaptioning>

16. End-to-End Dense Video Captioning with Masked Transformer

论文：Zhou L, Zhou Y, Corso J J, *et al.* End-to-end dense video captioning with masked transformer[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8739-8748.

工作：该论文指出了密集事件描述任务存在的两个问题：（1）在视频编码的过程中使用 LSTM 网络难以解决长时间依赖问题。（2）无法实现端到端的训练，只能单独训练推选模块和描述模块。论文针对这两个问题提出了解决方法，使用堆叠起来的自注意力机制和全连接层组成的转换器代替 LSTM 网络来编码视频。具体而言，方法分为三个模块：视频编码模块，推选模块和描述生成模块。在视频编码模块使用转换器来编码视频。在推选模块，使用了针对该任务改良的 ProcNets 方法。在描述生成模块，同样使用了自注意力机制，值得注意的是，在这个阶段为了实现端到端的训练，通过 fM 函数对每个推举块生成掩层，并将其与 video 特征进行融合输入到 multi-head attention 中，得到对预测单词的概率，从而实现端到端的训练。图 11 为转换器示意图，图 12 为网络结构图，图 13 为 fM 函数公式。

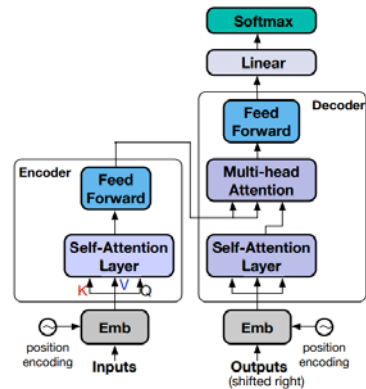


图 11 转换器示意图

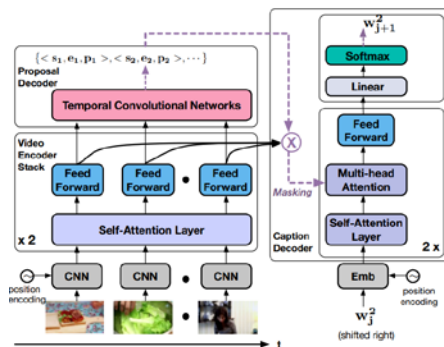


图 12 网络结构图

$$Y_{\leq t}^{l+1} = C(Y_{\leq t}^l) = \Psi(\text{PF}(\Phi(Y_{\leq t}^l)), \Phi(Y_{\leq t}^l))$$

$$\Phi(Y_{\leq t}^l) = \begin{pmatrix} \Psi(\text{MA}(\Omega(Y_{\leq t}^l)_1, \hat{F}^l, \hat{F}^l), \Omega(Y_{\leq t}^l)_1)) \\ \dots \\ \Psi(\text{MA}(\Omega(Y_{\leq t}^l)_t, \hat{F}^l, \hat{F}^l), \Omega(Y_{\leq t}^l)_t)) \end{pmatrix}$$

$$\Omega(Y_{\leq t}^l) = \begin{pmatrix} \Psi(\text{MA}(y_1^l, Y^l, Y^l), y_1^l)^\top \\ \dots \\ \Psi(\text{MA}(y_t^l, Y^l, Y^l), y_t^l)^\top \end{pmatrix}$$

$$\hat{F}^l = f_M(S_p, E_p) \odot F^l$$

$$p(w_{t+1}|X, Y_{\leq t}^l) = \text{softmax}(W^V y_{t+1}^l)$$

图 13 fM 函数公式

论文：http://openaccess.thecvf.com/content_cvpr_2018/papers/Zhou_End-to-End_Dense_Video_CVPR_2018_paper.pdf

代码：<https://github.com/salesforce/densecap>

17. Weakly Supervised Dense Video Captioning

论文：Shen Z, Li J, Su Z, *et al.* Weakly supervised dense video captioning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 1916-1924.

工作：该论文的主要贡献是为保证视频生成描述的多样性，对于同一个视频首先提取出不同

的区域序列，然后对每一个区域序列生成相应的描述。具体分为三个部分，在 Lexical-FCN 模型部分是得到视频每帧与词汇库之间的对应关系。在区域生成部分是使用子模块最大化数学方法从视频中提取出区域连续、信息连贯多样的区域序列。最后是描述生成部分，使用双向 LSTM 模型并额外添加了类别作为先验信息，使得生成效果更好。网络结构图如图 14 所示。

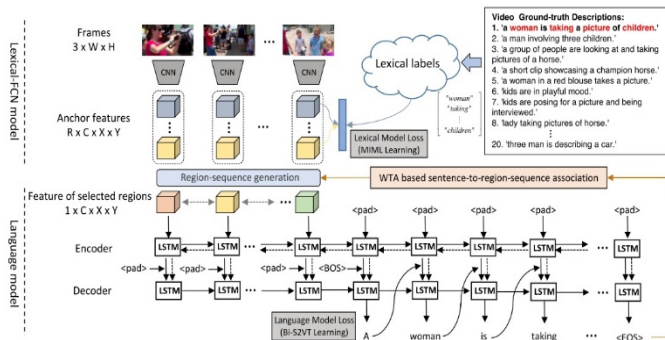


图 14 网络结构图

论文: http://openaccess.thecvf.com/content_cvpr_2017/papers/Shen Weakly Supervised Dense CVPR 2017 paper.pdf
 代码: <https://github.com/SCLinDennis/Weakly-Supervised-Dense-Video-Captioning>

18. Reconstruction network for video captioning

论文: Wang B, Ma L, Zhang W, et al. Reconstruction network for video captioning[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7622-7631.

工作: 该论文的主要贡献是通过重建视频视

觉特征来生成更好的描述。在网络结构上，编码器-解码器都是采用 LSTM 模型，然后利用解码器的状态重建视频的视频特征，以此来促进对视频语义的学习，从而为给定的视频生成更好的描述。文章提出了两种重建视频特征的结构，一种是重建视频的总特征，在重建总特征时利用编码器的隐状态和所有隐状态的平均值，其中平均值代表整个句子的语义。一种是重建每个的局部特征，是对编码器的隐状态采用注意力机制。图 15 为重建视觉特征示意图。

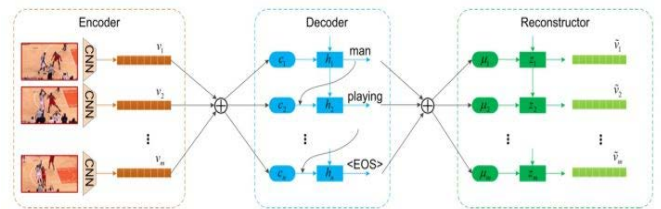


图 15 重建视频视觉特征示意图

论文: http://openaccess.thecvf.com/content_cvpr_2018/papers/Wang Reconstruction Network for CVPR 2018 paper.pdf
 代码: <https://github.com/sususushi/reconstruction-network-for-video-captioning>

(责任编辑: 蹇木伟, 樊鑫)



李熹桥

北京电子科技学院在读研究生，研究方向为计算机视觉。
 Email: azxcvgy@outlook.com



金鑫

北京电子科技学院从事教学与科研工作，北京电子科技学院可视计算与安全实验室负责人。研究方向为计算机视觉、虚拟/增强现实、人工智能安全。
 个人主页: www.jinxin.me

医学影像分割数据集

北京工业大学 弓照鹏 李晓光

随着临床需求的不断增加和医学成像技术的不断发展，医疗数据迅速增长。目前，医疗数据中有超过 90%来自于医学影像。医学影像在临床诊断和教学科研方面发挥着重要的作用。临床上一般由有经验的放射科医师对医学影像数据进行分析。人工处理的方式不仅费时费力，而且由于人主观意识的差异性容易造成漏诊甚至误诊。近年来深度学习技术在计算机视觉和图像处理领域取得了突破性的研究成果。将计算机视觉前沿技术应用到医学影像的处理中，不仅能够提高处理的效率，减轻医生工作负担，而且对降低漏诊和误诊有重要的意义。

医学影像分割是医学影像处理与分析领域关键的一步。根据图像的灰度、纹理等特征将 2D 或 3D 医学图像进行分割。分割结果主要用于感兴趣区域的提取，便于后续分析；计算人体关键解剖结构器官或病灶的容积、体积等便于临床统计分析研究；图像的三维重建和可视化。受成像时噪声、伪影、容积效应、参数设置、患者体位等的影响，实际获取到的图像具有模糊、灰度不均匀等特点。此外，人体的解剖结构具有一定的复杂性和差异性。这些复杂的条件都给医学图像的自动分割带来了极大的挑战。

近年来，以 U-net 网络结构为代表的深度学习分割方法已在医学图像领域取得了显著的成就。但是，目前全自动的分割算法在很多情况下分割结果尚难以满足要求。研究更高准确率和更加鲁棒的分割模型一直是研究人员的努力目标。本文重点介绍几个医学影像分割数据集涉及人体主要组织器官胰腺、心脏、大脑和肝脏，具体包括 Pancreas-CT, BraTS, 3Dircadb 和 ACDC。

1. Pancreas-CT

Pancreas-CT 数据集是一个胰腺分割任务的公开数据集。由美国国立卫生研究院临床中心对 53 名男性和 27 名女性受试者进行 82 次腹部增

强三维 CT 扫描得到。共有 82 例数据，19328 幅 CT 图像。其中 17 名受试者是肾脏切除术前接受扫描的健康肾脏捐赠者，其余 65 例患者由放射科医生从无重大腹部病变和胰腺癌病变的患者中挑选。受试者的年龄从 18 岁到 76 岁，年龄为 46.8 ± 16.7 。CT 成像数据的分辨率为 512×512 体素，切片厚度介于 1.5-2.5 毫米之间。数据由飞利浦和西门子的 MDCT 扫描仪获得(120 千伏峰值管电压)，以 DICOM 格式存储。数据库由影像学的学生手工标注胰腺切片，并由经验丰富的放射科医生审核和修改。该数据库的示例如图 1 所示（从左到右依次为冠状位、矢状位和轴位的标注示意图）。

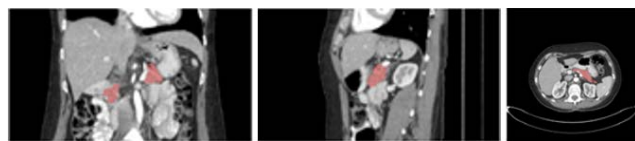


图 1 Pancreas-CT 数据集

数据集地址：<https://wiki.cancerimagingarchive.net/display/Public/Pancreas-CT>

2. BraTS

BraTS (Brain Tumor Segmentation) 数据集是一个脑肿瘤分割、患者整体生存预测的竞赛数据集，由美国宾夕法尼亚大学生物医学图像计算和分析中心等 19 个机构整理提供。

BraTS 2018 数据集共包含 542 例术前 MRI 数据，其中训练数据 285 例，验证数据 66 例，测试数据 191 例，数据格式为 NIFTI。所有数据是在标准临床条件下从多个不同中心获得的一系列脑肿瘤 MRI 扫描，但使用的设备和成像协议不同，导致图像质量极不均匀，反映了不同机构的临床实践差异。为了确保数据的规范化，所有数据都配准到相同的解剖模板，并且重采样为 1mm^3 。数据按照相同的标注协议由 1 到 4 名标注人员标注，并由经验丰富的放射科医师审核。肿瘤边界

的放射学定义，特别是在像胶质瘤这样的浸润性肿瘤中，是一个众所周知的难点问题。为了提供一种评估和评估各种肿瘤亚区的标准化方法，BraTS 在咨询了国际公认的神经放射专家后，定义了以下类型的肿瘤亚区 1) 坏死区；2) 水肿区；3) 非增强肿瘤异常区；4) 增强肿瘤区。该数据集的示例如图 2 所示(黄色区域表示瘤旁水肿区，红色区域为增强肿瘤异常区，蓝色区域为增强肿瘤区，绿色区域为坏死区)。

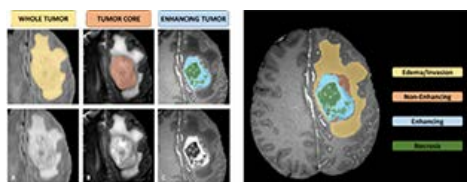


图 2 肿瘤分区示意图

数据集地址: <https://www.med.upenn.edu/sbia/brats2018.html>

3. 3Dircadb

3Dircadb(3D Image Reconstruction for Comparison of Algorithm Database)数据集，用于评估肝脏分割算法的网格生成算法。由法国抗癌研究所(IRCAD)提供。

数据集由 10 名女性和 10 名男性的 CT 扫描组成，其中 75% 的患者患有肝肿瘤，如图 3 所示。数据集下 20 个不同的文件夹对应 20 个不同的患者，每个文件夹下面包含 4 个子文件夹分别是 PATIENT_DICOM, LABELLED_DICOM, MASKS_DICOM 和 MESHES_VTK，分别包含了病人的原始 CT 扫描数据；所有感兴趣区域的人工标注结果；各个感兴趣区域 mask 子文件夹集；感兴趣区域以 VTK 格式的网格文件存储，示例图像如图 3 所示。

数据集地址: <https://www.ircad.fr/research/3d-ircadb-01>

01

4. ACDC

ACDC (Automated Cardiac Diagnosis Challenge) 是 MICCAI 2017 心脏分割和诊断竞赛数据集。由法国勃艮第大学附属医院提供。

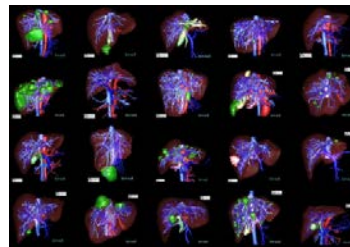


图 3 3Dircadb 数据集

数据集包含 150 名患者的 MRI 扫描数据，100 名患者数据作为训练集，50 名患者数据作为测试集。其中有 30 名健康者，30 名既往心肌梗死患者，30 名扩展型心肌病患者，30 名肥厚性心肌病患者，30 名右心室异常患者。所有数据是通过两台磁强度不同的西门子 MRI 扫描仪获取。在获得长轴切片后，获得从基底到顶点覆盖左心室的短轴切片，切片厚度为 5mm 到 10mm 空间分辨率从 1.34 到 1.68mm²/体素。

标注由经验丰富的放射科医生手工分割完成，给出了左心室、右心室和心肌结果，如图 4 所示(左侧是原始数据，右侧是标注数据；红色表示右心室，蓝色表示左心室，绿色表示心肌)。

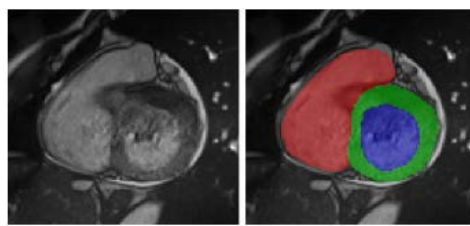


图 4 ACDC 数据集

数据集地址: <https://www.creatis.insa-lyon.fr/Challenge/acdc/databases.html>

(责任编辑: 金鑫, 贾同)



弓照鹏

北京工业大学计算机技术专业在读研究生，研究方向为医学影像智能处理。

Email: gzp@emails.bjut.edu.cn



李晓光

北京工业大学信息学部副教授。研究方向为图像超分辨率重建、人脸图像分析、医学影像处理等。

Email: lxxg@bjut.edu.cn

招聘信息

一、 旷世 2020 校招

1、后端开发工程师：

1) 工作职责：

- 负责 ToB 场景，私有云平台（定制 OS，应用部署，容器编排与离线场景的自动化运维等）的研发；
- 推动业务在可自动运维性，可用性和性能上持续改善和优化；

2) 任职资格：

- 计算机或相关专业本科以上学历；
- 热爱互联网，对互联网产品和技术有浓厚的兴趣，热衷追求技术极致和创新；
- 熟练掌握数据结构与算法、计算机网络、操作系统、编译原理等课程；
- 良好的开发语言基础，并且至少熟练如下一一种研发语言，比如 Go、Python、Java、C、C++等，并乐于转向 Go/Python 开发；
- 充满技术热情，有较强的自驱力和学习能力；
- 有项目经验优先。

2、深度学习架构师：

1) 工作职责：

- 负责旷视核心深度框架 MegBrain 的设计，演进，实现，维护和优化；
- 基于 MegBrai 平台，在异构计算设备（CUDA/Arm/x86 等）上优化深度学习算法的性能
- 持续探索改进深度学习平台的新方法（例如分布式/并行训练、模型加速、自动模型训练等）。

2) 任职资格：

- 编程能力强，有良好的数据结构与算法功底，能够熟练使用 C+语言编写较复杂的算法；
- 有丰富的编程经验，能够调试解决在 Linux 或 Android 平台的常见的问题；
- 对计算机体系有比较深入的了解，能够基于

系统特性进行性能调优；

- 对深度学习和深度学习框架（Caffe、Tensorflow、PyTorch 等）有基本了解者优先；
- 有相关工作经验或研究经验者优先。

3、测试开发工程师：

1) 工作职责：

- 根据产品特点，开发合适测试工具或自动化方案，提高测试效率；
- 参与测试平台设计和搭建；
- 设计执行测试用例，跟踪产品缺陷或问题；
- 与 PM，开发和交付人员沟通产品设计和架构定位项目问题；

2) 任职资格：

- 具有设计和开发测试工具和测试平台经验；
- 精通 Python/Perl/Shell，有一定 C++开发经验；
- 熟悉 jenkins/git，熟悉 CICD 常用方法和工具；
- 有一定数据统计分析和建模能力，和复杂分布式系统测试经验；
- 热爱人工智能行业应用和测试技术，具有快速学习能力；
- 有人工智能算法测试背景优先。

4、核心算法引擎开发工程师：

1) 工作职责：

- 负责核心引擎的架构设计及研发
- 对系统进行优化，提升可用性性能以及可运维性等；
- 为业务提供支持，解决各类技术难点；
- 负责 AI 计算芯片优化，算法模型 pipeline 设计，以适应不同场景 AI 需求。

2) 任职资格：

- 优秀的编码能力，熟练掌握 Golang/Python/C/C++/Java 等至少一门语

言；

- 熟悉常用算法和数据结构；
- 精通操作系统，网络等相关底层系统知识；
- 熟悉分布式系统设计，有丰富的研发经验；
- 富有技术理想和激情，拥有优秀的学习能力，善于分析解决问题；
- 具备良好的团队协作意识。

5、算法研究员：

1) 工作职责：

- 推进计算机视觉和深度学习领域的核心算法；
- 构建计算机视觉或深度学习领域的关键应用；
- 将最好的算法在有趣有用的商业场景中落地。

2) 任职资格：

- 过硬的科研背景：深度学习/机器学习；计算机视觉（包括分类、检测、分割、跟踪、SLAM 和三维重建）；图像或信号处理；计算摄影学和计算机图形学；
- 扎实的编程基础；
- 熟悉本研究领域的最新研究成果，公开数据集，和相关的开源系统；
- 有很强的自学能力和独立思考能力，善于思考和表达自己的想法；同时又具备良好的团队合作精神；
- 熟悉 Python 等至少一门脚本语言，使用过 Theano、Caffe、Torch、TensorFlow 等开源深度学习框架优先；
- 有深度模型训练、图像分类、物体检测与分割、视频分析、三维建模、计算机图形学等相关科研经历者（例如顶级会议第一作者）优先。

6、激光 SLAM/多传感器融合算法研究员：

1) 工作职责：

- 负责单线/多线激光 SLAM 算法开发与优化；
- 负责 lidar laser、IMU、camera、odometer 等多传感器融合算法研发。

2) 任职资格：

- 计算机、自动化、软件、机械、电子等相关专业本科或以上学历；
- 熟练掌握 C/C++ 编程语言，熟练掌握 ROS 编程，拥有扎实的 SLAM 基础理论知识以及较好的数学基础；
- 熟练掌握以下：IMU/odometer 惯导算法、激光 SLAM、Particle Filter、EKF 等一项或多项相关技术和开发经验；
- 熟悉任意一种激光 SLAM（如 cartographer、gmapping、loam）算法原理或应用；
- 有强烈的上进心和求知欲，善于学习新事物，对技术充满激情；
- 责任心强，具备很高的沟通表达能力和团队合作意识。

7、视觉感知算法(3D 方向)：

1) 工作职责：

- 负责机器人场景中图像相关机器学习算法研究与应用（车道线定位、检测分割、图像检索、特征匹配、Stereo 等）。

2) 任职资格：

- 具有计算机视觉、图形学、增强/虚拟现实、机器人等相关专业在读本科或以上学历；
- 熟悉 2D 点云或 3D 点云上常用检测、分割算法；了解深度估计、视觉定位、特征匹配等相关算法；
- 熟悉机器学习/深度学习基础知识，了解立体视觉常用知识和算法；
- 熟悉经典 CV 中的常见算法，熟悉 OpenCV；了解图像处理、光流跟踪、图像匹配、描述子等基础知识；
- Python/C++ 编程基础扎实，熟悉 Tensorflow/Pytorch/MXNet 等常用框架。
- 加分项（非必要）：
 - 具备 ACM 等大赛奖牌者优先；
 - 具有无人驾驶，机器人等场景项目经验者优先；
 - 在 CVPR、ECCV、ICCV 或 ICRA、IROS 等相关

会议取得过优异成绩或发表论文或参加竞赛者优先；

- 了解 PointNet、PointCNN 等常见深度学习方法优先；
- 了解 3D 视觉/SLAM 等算法者优先。

8、投递方式：

官网投递简历：

https://www.megvii.com/join_us/campus

二、快手 2020 校招

1、AI 平台架构算法工程师（C++）：

1) 职位描述：

- 整合业务机器学习训练和预测需求；
- 集成并优化目前主流机器学习框架，包括但不限于：TensorFlow/MxNet/Caffe/PyTorch 等；
- 根据业务需求对相关工具、算法进行稳定性、性能以及扩展性等优化；
- 研发支持超大规模训练的 AI 平台。

2) 任职要求：

- 优秀的逻辑思维和学习能力；
- 优秀的设计与编码能力，工程质量自我要求高，针对业务需求与问题，可快速设计与实现解决方案；
- 具有扎实的编程基础，熟练掌握 C++、Python；
- 熟悉 TensorFlow/PyTorch/XGBoost 等开源机器学习工具者优先；
- 熟悉训练流程的 pipeline 者优先；
- 熟悉常见机器学习算法，有分布式机器学习平台项目经验优先。

2、视频算法工程师

1) 职位描述：

- 针对移动端拍摄生成视频和直播内容的特点，持续优化视频清晰度和流畅度；
- 实现自主的视频编解码算法框架，针对不同应用场景进行质量优化，针对移动端、服务端运行环境进行速度优化。

2) 任职要求：

- 计算机、多媒体通信、信号处理等相关专业本科及以上学历；
- 精通视频编解码标准（H.264/265, VVC/H.266, VPX/AV1 或 AVS 系列），熟悉精通算法框架细节；
- 熟悉 x264/x265 等开源项目者优先；
- 深入理解数字图像处理、机器学习的业界发展方向，有深度学习应用经验者优先；
- 要求有较强的沟通表达能力和团队意识。

3、数据挖掘算法工程师

1) 职位描述：

- 对海量的业务数据、用户数据进行挖掘分析，挖掘用户特征，完成对用户兴趣和人口属性的建模；
- 发掘数据和业务背后的规律，利用用户画像数据，提升广告与用户匹配的效率；
- 针对业务流程进行分析调研，探索提升转化率效果的思路 and 方案并推动转化。

2) 任职要求：

- 熟悉大规模数据挖掘、机器学习、自然语言处理、分布式计算等相关技术，能熟练使用聚类、回归、分类等算法并调优；
- 熟悉 Linux 环境开发，至少熟悉 Python/awk/Go/C/C++ 等语言中一种或一种以上；
- 熟悉基于 Hive、Spark、ElasticSearch、Mongodb 等大数据平台的相关开发。

4、算法/系统研究员

1) 职位描述：

- 支持商业化广告：算法模型；
- 支持商业化广告：大规模推荐算法平台；
- 支持游戏业务：游戏 AI；
- 支持游戏业务：游戏数据分析，智能运营；
- 支持商业化以及 Y-Tech：线上模型压缩/手机端模型压缩，Automl；
- 支持 Y-Tech 基础研究：小样本学习。

2) 任职要求：

- 硕士及以上学历，机器学习，计算机视觉相关专业；
- 熟悉机器学习，深度学习算法；
- 熟练使用 Python、Pytorch 等工具。
- 加分项：
- 顶会论文，如 NIPS、ICML、CVPR、ICCV、KID 等；
- 竞赛获奖者。

5、图像/视频分析算法工程师

1) 职位描述：

- 针对图像和视频的中低层语义（清晰度、色彩、构图、美学评估、质量、场景等），设计图像和视频的分析理解算法；
- 在不同平台实现视频图像分析理解算法，并对其进行优化。

2) 任职要求：

- 图像/视频处理，计算机视觉，机器学习等相关专业硕士及以上学历；
- 设计开发过基于传统机器学习或深度学习方法的视频分析算法，如：图像/视频场景识别、图像/视频质量评估、美学评估、清晰度评估、显著性区域检测、运动分析、运动物体分割和检测、或者运动分类等；
- 精通 Python、C/C++ 等编程语言；
- 期望要求：
- 有在国际会议发表过图像/视频理解算法者优先；
- 有 OpenGL/Metal 等开发经验者优先。

6、推荐算法工程师

1) 职位描述：

- 负责快手推荐算法、策略机制的研究与开发；
- 帮助用户在快手发现和探索感兴趣的内容，形成基于兴趣的社群，基于深度学习模型进行召回、排序等工作，持续优化提升用户价值和黏性；
- 设计广告用户体验指标体系，优化广告排序与竞价逻辑，降低商业内容对用户体验的影响，提升广告的投放效率，深入理解关系链

和用户行为，通过数据分析、挖掘，改进算法策略和目标，并推动产品迭代；

- 研究广告用户体验与社区生态的影响因素，通过产品和策略来指导广告主优化用户体验；
- 挖掘并监控社区内影响用户体验的商业内容，通过产品和技术手段，及时干预可能的用户体验和安全风险。

2) 任职要求：

- 有扎实的编程、机器学习，对数据有较强的敏感度，具有大规模数据分析能力；
- 熟悉常见的深度学习模型，熟悉至少一种常见的深度学习平台，TensorFlow/MXNet/PyTorch 等；
- 目标导向，善于结合具体业务场景，分析与解决有挑战性的问题；
- 有良好的逻辑思维能力，善于沟通和表达，对短视频/内容类产品感兴趣。

7、音频算法工程师

1) 职位描述：

- 负责语音连麦、回声消除、噪声消除等算法和产品的研发；
- 负责音频特效、音频分类、音频检测等算法和产品的研发；
- 语音、音乐、音频领域新技术研究探索。

2) 任职要求：

- 熟悉基本声学知识或乐理知识；
- 有良好的数字信号处理基础；
- 掌握 Matlab, C/C++ 或 Python，能够快速实现音频算法；
- 熟悉常用的机器学习算法；
- 扎实的中英文文献阅读能力；
- 较强的学习能力和算法研究功底。

8、机器学习算法工程师

1) 职位描述：

- 强化学习等前沿技术在快手核心业务场景的应用和落地（推荐/精准营销/广告/游戏）。

2) 任职要求：

- 机器学习、深度学习相关方向背景扎实，有实际应用经验；
- 优秀的编程能力 C++/Python，熟悉常见的数据结构和算法；
- 熟悉 Linux 开发环境和常用开发工具，熟悉 Tensorflow/PyTorch/MxNet 等至少一种常用深度学习框架；
- 较强的文献阅读和理解能力，良好的逻辑思维、沟通表达能力；
- 机器学习竞赛成绩优异者，顶级会议论文发表者优先；
- 在强化学习、推荐/广告系统、大型分布式系统等相关领域有研究或实践经验者优先。

9、计算机视觉算法研究员

- 1) 职位描述：
 - 负责人脸 Landmark、3D 人脸相关的 CV 算法研发，推进 AI 技术在图像，视频领域的研究和应用；
 - 人脸真实感建模和渲染；
 - 结合实际进行人脸相关算法的设计和优化，推动算法产品落地；
 - 负责语义分割、目标检测、GAN 等深度学习算法优化；
 - 结合产品需求进行算法设计与优化，并推动 AI 算法落地。
- 2) 任职要求：
 - 硕士及以上学历，计算机视觉，图形图像相关专业；
 - 熟悉机器学习，深度学习在图像，视频领域的相关算法；
 - 能够熟练使用 Python 进行算法开发，具有 C/C++ 程序算法开发能力；
 - 有深度学习模型压缩，量化经验者优先；
 - 具有良好的团队意识和合作精神，善于分析、简化和解决问题。
- 加分项：
 - 顶会论文，如 CVPR、ICCV、NIPS 等，参加过有影响力竞赛并取得良好名次。

10、简历投递

快手招聘官网：

<https://campus.kuaishou.cn/#/campus/jobs>

三、去哪儿网 2020 秋招

工作地点：北京

1、开发工程师：

- 1) 职位描述：
 - 工作方向：后端/客户端/算法/大数据/搜索/安全/运维等相关开发方向；
- 2) 工作职责：
 - 按照需求，负责技术平台或业务支持系统的设计、维护，以及代码开发工作；
 - 根据开发规范编写各种开发文档及项目文档；
 - 协助需求方进行技术调研；
 - 不断提高产品代码质量，参与团队技术交流。
- 3) 任职资格：
 - 2020 年 1 月-2020 年 8 月毕业的应届生、全日制本科及以上学历，计算机相关专业；
 - 具有一定数据结构、面向对象编程、设计模式基础；
 - 至少熟悉 c/c++/java/python 等一种编程语言；
 - 熟悉 linux 操作系统者优先；
 - 有互联网开发经验者优先。

2、测试开发工程师：

- 1) 工作职责：
 - 参与去哪儿网全系列产品的研发与测试过程：产品需求、功能设计、产品可测性设计与反馈、单元测试、系统测试、性能测试、产品效果评估评测等多方面，全流程的保证产品质量；
 - 参与构建去哪儿研发体系的持续集成环境构建及自动化测试平台的研发，构建和模拟线上大规模和分布式的多样化运维场景，参与产品线特色化的测试框架，测试工具研发；

- 参与去哪儿网产品整个发布过程中的技术创新，在保证高质量交付的同时加速软件的发布周期，与业内优秀的项目流程管理人员一起，对去哪儿的核心产品的发布过程进行技术革新和流程优化，包括解决编译依赖，提升测试效率，开发创新的性能测试和监控工具来精确预估线上系统表现。
- 2) 任职资格：
- 2020年1月-2020年8月毕业的应届生、全日制本科及以上学历，计算机相关专业，优秀的代码能力和编码习惯（Java 优先）；
 - 性格外向，活泼开朗、善于交流，能积极融入团队并与其他部门同事进行良好的合作；
 - 热爱互联网，对互联网相关业务或技术充满好奇及热情；
 - 有责任心，敢于担当，工作积极主动，不断追求卓越；
 - 对软件测试领域，对发现、分析及解决问题的工有浓厚兴趣；
 - 快速的产品及业务学习能力，敏捷而周到的逻辑思维能力。
- 3、前端开发工程师：
- 1) 工作职责：
- 负责去哪儿网各产品线 Web 前端研发；
 - 负责去哪儿网各产品线易用性改进和界面技术优化；
 - Web 前沿技术研究和新技术调研。
- 2) 任职资格：
- 2020年1月-2020年8月毕业的应届生、全日制本科及以上学历，计算机相关专业；
 - 熟悉 JavaScript、Ajax 等 Web 开发技术；
 - 熟悉 HTML/XHTML、CSS 等网页制作技术，熟悉页面架构和布局；
- 4、产品经理：
- 1) 工作职责：
- 负责用户需求的调研、分析、评估和原型设计；
 - 推动、协调产品功能的开发、测试和发布；
 - 协助产品上线以后的 BUG 跟踪、收集改进意见、提供改进方案；
 - 定期对行业相关产品进行评估，并提出功能优化、用户体验升级等方面建议。
- 2) 任职资格：
- 2020年1月-2020年8月毕业的应届生、全日制本科及以上学历，专业不限；
 - 热爱互联网和旅游行业；
 - 有较强逻辑思维能力，以及较强的沟通协调能力和团队合作精神；
 - 了解用户体验的相关知识，会使用 Axure、Visio、Photoshop 等原型设计工具者优先；
 - 有电子商务、在线旅行等相关实习经验者优先。
- 5、投递方式：
- 官网投递简历：<http://campus.gunar.com>
- (责任编辑：刘丽)

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 Computer Vision and Image Understanding 特刊。

3 会议简介 (ICAIT)

ICAIT 2020 (International Conference on Advanced Computer Science and Information

Technology) 将于 2020 年 5 月 30 日至 5 月 31 日在加拿大温哥华举办，论文截止时间为 2019 年 9 月 21 日。其为研究人员交流先进信息技术的理论、方法和应用方面的知识和成果提供了一个极好的国际会议论坛。会议期待在理论和实践方面对计算机人工智能和信息技术的的所有主要领域作出重大贡献。会议的目的是为学术界和工业界的研究人员和从业人员提供一个平台，以满足和分享该领域的尖端发展。

(责任编辑：李策)

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 |
|------------|-----------------|--------------------|--------------|---|
| ICAIT 2020 | 2020. 05. 30-31 | 加拿大 Vancouver | 2019. 09. 21 | https://itcse2020.org/icait/index.html |
| AI3D 2019 | 2019. 12. 09-11 | 美国 California | 2019. 09. 22 | https://sites.google.com/view/ai3d2019/home |
| ISCFI 2019 | 2019. 11. 19-20 | 南非 Johannesburg | 2019. 09. 25 | http://www.iscfi.us |
| ACMLC 2019 | 2019. 12. 07-09 | 中国 Hong Kong | 2019. 09. 25 | http://www.acmlc.org |
| CIIS 2019 | 2019. 11. 23-25 | 泰国 Bangkok | 2019. 09. 25 | http://www.ciis.net |

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 | 截稿日期 |
|---|--|--------------|
| Computer Vision and Image Understanding | Special Issue on “Adversarial Deep Learning in Biometrics & Forensics” | 2019. 10. 01 |
| Computer Vision and Image Understanding | Special Issue on Deep Learning for Image Restoration | 2019. 12. 15 |