

CCF

计算机视觉 专委会简报

COMPUTER VISION NEWSLETTER

2019/02 期
总第 18 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
视界无限	视界无限系列研讨会	05
会议推介	PRCV 2019 征文通知	06
	PRCV 2019 诚邀申请讲习班	09
	PRCV 2019 专题论坛征集通知	10
	PRCV 2019 博士生论坛征集通知	10
	PRCV 2019 诚邀申请成果展示	12

科技前沿

热点追击	基于空间特征选择的自适应跟踪算法	13
	融合单阶段和两阶段检测思想的物体检测算法	14
	基于双链区分性 CNN 的多视角 3D 模型检索算法	15

委员风采

委员访谈	西安电子科技大学高新波教授访谈	16
委员好消息		20

资源平台

开源代码	灰度图像自动着色技术	21
数据集	唇度（视觉语音识别）	23
招聘信息		26
征文通知		28

CCF-CV 走进高校系列报告会

第 68 期 北京信息科技大学

时间：2019 年 1 月 16 日

第六十八期“CCF-CV”走进高校系列报告会活动在北京信息科技大学大学（小营校区）学术交流报告厅成功举行。本期报告会邀请了中国科学院自动化所王亮研究员，华中科技大学白翔教授，中科院计算所韩琥副研究员三位专家学者做特邀报告。北京信息科技大学自动化学院教授陈雯柏、北京石油化工学院信息工程学院副教授刘学君、北京信息科技大学智能科学与技术系副教授刘琼担任执行主席。来自视觉领域专业人士及师生共 100 余人聆听了三位专家的报告。



报告会由陈雯柏教授主持并致欢迎辞。陈雯柏教授对各位专家学者的到来表示热烈的欢迎和衷心的感谢，希望通过此次中国计算机学会计算机视觉专委会走进北京信息科技大学的活动，让相关师生进一步了解智能科学前沿及应用的相关情况，提高在人工智能领域的研究与学习热情，推动智能科学与技术学科专业建设。

王亮研究员的报告题目是 Bridging Visual-Semantic Gap via Deep Multimodal Learning。重点回顾了深度学习历史及其在视觉与文本多模态大数据分析中的应用进展。王亮重点介绍了其所在的中科院自动化所模式识别国家重点实验室团队在近两年所取得的最新的研究成果，如多模态注意网络用于图文匹配、面向多模态匹配的语义概念与序列学习方法、多模态记忆网络用于视频标注等。

白翔教授的报告题目是“任意形状文本检测与识别”。图像中的文本检测对于图像理解具有十分重要的作用与意义，白翔老师一直以来都在文本检测与识别领域深耕。报告重点介绍了团队如何实现从规则形状文本到任意形状文件的检测与识别对其所提出的研究方法进行了重点介绍。报告最后，白翔老师给大家讲出了非常宝贵的研究心得——知难而进，持之以恒。

韩琥副研究员的报告题目是“由表及里智能生物感知”。生物感知是计算科学、认知科学、心理学以及医学科学融合的前沿交叉研究领域，其研究目的是让机器既能知面又能懂心。报告重点介绍了表观属性分析研究进展与成果，同时又介绍了非穿戴的基于外部表观特征的心率检测最新研究成果，引发了与会师生极大的研究兴趣。

报告激发了与会的校内外众多领域研究者与师生的研究兴趣，专家们深入浅出的讲解，让互动环节热烈而高潮迭起。最后报告会在热烈的掌声中圆满结束。

（责任编辑：马占宇）

CCF-CV

视界无限系列研讨会

第 1 期 哈尔滨工业大学

2019 年 1 月 10 日，由中国计算机学会计算机视觉专委会主办、哈尔滨工业大学承办的第 1 期 CCF-CV “视界无限”系列活动——“生成式对抗网络 (GAN) 的前沿进展与未来趋势”研讨会在哈尔滨华融饭店成功举行。研讨会由计算机视觉专委会委员、哈尔滨工业大学左旺孟教授组织和主持。研讨会邀请了中科院自动化所赫然研究员、北京航空航天大学刘偲副教授、大连理工大学雷娜教授三位专家做主题报告。来自全国各高校和企业的 50 名参会者汇聚冰城，就 GAN 的理论与应用进行了热烈的交流和深入细致的探讨。

首先，哈尔滨工业大学姚鸿勋教授代表研讨会承办单位欢迎大家来到冰城哈尔滨参加“视界无限”活动。姚鸿勋教授建议，参会者一方面要从各位讲者的报告中有所受益，同时更要借鉴他人对问题的捕捉和思考，加深对问题和方向的认识、理解，深入和持续地研究问题。最后，姚鸿勋教授祝愿研讨会圆满成功。

计算机视觉专委会秘书长、中科院自动化所王亮研究员代表专委会领导致辞并介绍了“视界无限”系列活动的缘起、目的、思路和预期成果。王亮秘书长指出，“视界无限”是计算机视觉专委会继“走进高校”和“走进企业”之后，缜密筹划和精心组织的又一全新品牌活动，旨在围绕特定主题，组织资深研究者与活跃的青年学者进行全面的深入研讨，总结和梳理该主题前沿进展与未来趋势，为相关领域的一线研究者提供面对面密切交流的机会，推动该领域研究和技术的持续、深入、快速发展。

左旺孟教授的引导发言题目是“生成式对抗网络 (GAN) 简介”。GAN 作为一种有效的图像生成与编辑模型，近年来在计算机视觉领域获得了

广泛关注。左旺孟教授在介绍 GAN 的基本思想和模型的基础上，回顾了 GAN 的主要发展历程、典型应用、以及值得关注的挑战性问题和发展方向。

在主题报告环节，赫然研究员的报告题目是“GAN for Image Synthesis”。赫然研究员在生成式深度学习理论和方法综述的基础上，介绍了近期开展的图像合成方法和应用，包括身份保持损失和人脸图像旋转、人脸超分、表情合成等。

刘偲副教授的报告题目是“基于生成对抗网络的人脸编辑”，介绍了课题组最新的三个基于生成对抗网络的工作。

雷娜教授的报告题目是“A Geometric View of Optimal Transportation and Generative Adversarial Networks”。从传统方法与机器学习方法的范式差异出发，给出了 GAN 的几何解释，论证了 WGAN 的数学基础特别是与最优传输理论的关系，说明了最优的生成器本质上是计算最优传输变换。

讨论环节分为深度解读、漫谈、预言三部分。与会嘉宾与现场师生就诸多问题展开了深入的讨论。最后，第 1 期“视界无限”研讨会在热烈的掌声中圆满结束。



(责任编辑：王瑞平)

PRCV2019 征文通知

第二届中国模式识别与计算机视觉大会
PRCV 2019
2019年11月8-11日
曲江国际会议中心，西安

征文通知

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、西安交通大学以及西安电子科技大学等单位共同承办。

本届会议将主要汇聚国内外从事模式识别和计算机视觉理论与应用研究的广大科研工作者及工业界同僚，共同分享国内外模式识别与计算机视觉领域的最新理论和技术成果，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、原创性的优秀学术论文（英文）。大会录用的稿件将在会上展示，会议论文集将由 Springer 出版社出版，并被 EI 和 ISTP 检索。优秀的论文将推荐到国内外高质量期刊的特刊（目前已申请到专刊的期刊：Pattern

Recognition、Neurocomputing）。

会议投稿系统已开放：

<https://cmt3.research.microsoft.com/prcv2019>

重要日期

- 投稿截止日期：2019年4月15日
- 录用通知日期：2019年6月15日
- 终稿提交日期：2019年7月15日
- 会议举办日期：2019年11月8-11日

征文范围（包括但不限于）

- 模式分类与聚类分析
- 机器学习
- 神经网络与深度学习
- 特征提取与特征选择
- 结构模式识别
- 文档分析
- 手写字符识别
- 人脸识别与姿态识别
- 生物特征识别
- 语音识别与说话人识别
- 信息检索
- 性能评测和基准数据库
- 物理模型驱动视觉分析及三维重构
- 计算机视觉基础理论
- 底层视觉理解、图像处理
- 三维视觉
- 目标检测与识别
- 动作识别
- 计算摄像学、传感与显示技术

- 运动目标检测及跟踪
- 多媒体数据分析
- 生物医学图像分析
- 遥感图像
- 最优化方法
- 机器人视觉
- 模式识别与计算机视觉应用

组委会

指导委员会

主 席：谭铁牛，中科院自动化所
 委 员：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

大会主席

谭铁牛，中科院自动化所
 郑南宁，西安交通大学
 陈熙霖，中科院计算所
 张艳宁，西北工业大学

程序委员会主席

林宙辰，北京大学
 王 亮，中科院自动化所
 杨 健，南京理工大学
 石光明，西安电子科技大学

组织委员会主席

薛建儒，西安交通大学
 王 鹏，西北工业大学
 魏 巍，西北工业大学

宣传主席

山世光，中科院计算所
 苗启广，西安电子科技大学
 张兆翔，中科院自动化所

国际联络主席

虞晶怡，上海科技大学
 鲁继文，清华大学
 马占宇，北京邮电大学

出版主席

白 翔，华中科技大学
 杨 涛，西北工业大学

专刊主席

程明明，南开大学
 郑伟诗，中山大学

讲习班主席

孟德宇，西安交通大学
 彭宇新，北京大学
 聂飞平，西北工业大学

专题论坛主席

卢湖川，大连理工大学
 王蕴红，北京航空航天大学
 刘青山，南京信息工程大学

赞助主席

王 涛，爱奇艺
 杨金锋，中国民航大学
 赵歆波，西北工业大学

展示主席

马惠敏，清华大学
 郝润平，西北工业大学

竞赛主席

桑 农， 华中科技大学
 左旺孟，哈尔滨工业大学
 尹翰林，西北工业大学

博士生论坛主席

韩军伟，西北工业大学

耿 新, 东南大学

刘 偲, 北京航空航天大学

网站主席

章国锋, 浙江大学

徐 迪, 西北工业大学

财务主席

孙瑾秋, 西北工业大学

毋立芳, 北京工业大学

注册主席

朱 宇, 西北工业大学

张世周, 西北工业大学

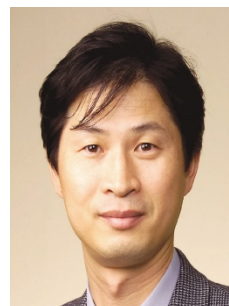
已定主题报告讲者



Tinne Tuytelaars

比利时鲁汶大学教授

IEEE TPAMI 副主编



Lee Kyoung Mu

首尔国立大学教授

IEEE TPAMI 副主编



Kyros Kutulakos

加拿大多伦多大学教授

Marr 奖获得者



焦李成

西安电子科技大学教授

IEEE Fellow、IET Fellow

已获赞助

•金牌赞助



•银牌赞助



•铜牌赞助



•赞助级别待定



会议邮箱

prcv2019@nwpu.edu.cn

会议网址

<http://www.prcv2019.com>

(责任编辑: 王瑞平)

PRCV2019 诚邀申请讲习班

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、西安交通大学以及西安电子科技大学等单位共同承办。

现诚挚邀请广大模式识别与计算机视觉领域的从业者在 PRCV 2019 会议上组织与承办计算机视觉、模式识别与机器学习方面的各种讲习班活动。讲习班针对模式识别、计算机视觉与机器学习领域的某一特定方向 (举例: 稀疏编码、行为识别、目标检测、图像细分类、跨媒体分析、生成对抗学习等), 由若干名讲者从不同维度、不同视角或者不同侧面等对特定方向的理论、方法与应用进行详细讲解。

有意愿在 PRCV 2019 组织或承办讲习班活动者, 请于 2019 年 8 月 31 日前将讲习班申请书

以电子邮件的形式提交给讲习班主席。申请结果将会在 2019 年 9 月 15 日通知组织者申请是否被大会采纳。所有被采纳的讲习班将会在 PRCV 2019 网站上公布。

讲习班申请书须包括以下信息

1. 讲习班主题
2. 讲习班简介、目的、意义
3. 拟邀请讲习班讲者及个人简介

讲习班主席

- 彭宇新, 北京大学,
pengyuxin@pku.edu.cn
- 聂飞平, 西北工业大学,
feipingnie@gmail.com
- 孟德宇, 西安交通大学,
dymeng@mail.xjtu.edu.cn

(责任编辑: 王瑞平)

PRCV2019 专题论坛征集通知

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、西安交通大学以及西安电子科技大学等单位共同承办。

现面向广大科技工作者公开征集专题论坛的组织者和论坛题目, 论坛范围包括 (但不局限于):

1. 模式识别与计算机视觉前沿论坛
2. 顶会、顶刊交流论坛

3. 高水平学术论文写作交流论坛

论坛申请应包括: 论坛题目, 论坛简介、目的与意义, 组织者个人简介, 拟邀请的讲者题目与个人简介等。

重要日期

- 论坛申请截止日期: 2019 年 8 月 31 日
- 论坛录用通知日期: 2019 年 9 月 15 日
- 论坛完整信息提交日期: 2019 年 10 月 15 日

请将论坛申请于截止日期前发送给论坛单元主席:

刘青山

(南京信息工程大学, qslu@nuist.edu.cn)

王蕴红

(北京航空航天大学, yhwang@buaa.edu.cn)

卢湖川

(大连理工大学, lhchuan@dlut.edu.cn)

(责任编辑: 马占宇)

PRCV2019 博士生论坛征集通知

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、西安交通大学以及西安电子科技大学等单位共同承办。

论坛概况

本论坛为最近已获得或即将获得博士学位的学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。所有的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，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。除海报展示外，组委会还将综合申请材料择优评选出部分参会学生进行口头报告。2019 年度评选出的 CCF-CV 学术新锐奖获得者将被直接邀请发表口头报告。

参会资格

参会学生的研究方向须为计算机视觉或模式识别领域，且已经获得博士学位或者能够在 6 个月内获得博士学位。

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），材料中须包括的内容如下：

1. 个人简历；
2. 两页纸的研究说明，总结申请者目前的研究进展；
3. 所展示海报的题目和作者列表；
4. 申请者导师的签名文件，确认该申请者的毕业时间及导师本人作为论坛导师的可能性；
5. 7 位候选导师的姓名和邮箱列表，按照偏好程度降序排列；

好程度降序排列；

注：本论坛不提供任何导师列表。申请者需根据研究方向的契合程度或其他考虑自行选择在学术界或工业界感兴趣的研究者列为导师。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。

海报格式

海报格式为宽 240cm，高 120cm 的矩形。参会时请在指定房间中自行选择可用的展示板进行张贴，论坛组委会不指定海报序号。

重要时间节点

- 申请材料提交截止日期：2019 年 8 月 31 日
- 审核结果公布日期：2019 年 9 月 15 日

论坛主席

有任何问题，请联系以下论坛主席：

- 韩军伟
(西北工业大学, junweihan2010@gmail.com)
- 耿新
(东南大学, xgeng@seu.edu.cn)
- 刘偲
(北京航空航天大学, liusi@buaa.edu.cn)

(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, 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、西安交通大学以及西安电子科技大学等单位共同承办。

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 (见附件) 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。申请展示的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, 每个展位需要至少全注册两名参会人员 (不得与参会论文作者注册共享)。

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 15 日
- 评选结果: 2019 年 9 月 30 日
- 会期展示: 2019 年 11 月 8-11 日

联系方式

- 郝润平
(西北工业大学, xirunping@126.com)
(责任编辑: 毋立芳)

基于空间特征选择的自适应跟踪算法

江南大学 徐天阳 吴小俊 萨里大学 冯振华 Josef Kittler

复杂环境下视频目标跟踪旨在自动建立目标及其周围背景的动态表观模型，实现自适应的目标定位，对于视觉环境理解、行为分析、视频检索等问题具有重要意义。近年来，通过深度神经网络得到的物体表观描述信息具有更好的鉴别性。基于相关滤波算法的跟踪器通过循环矩阵实现样本扩展和频率域内的高效求解，在目标跟踪领域内得到广泛研究并取得领先效果。但现有学习算法仅考虑不同空间位置固定权重，并没有将自适应空间特征选择考虑在内，往往容易受到干扰信息影响，对最终目标位置的推断产生影响。

基于以上考虑，提出了一种基于空间特征选择的自适应跟踪算法。该算法强调空间特征选择对目标跟踪的重要性，尤其是背景中部分高对比空间信息可以帮助实现鲁棒的目标跟踪。将自适应空间特征选择与相关滤波算法框架结合的优点如下。1) 通过对高维特征进行自适应空间特征选择，可实现对滤波器的降维学习，使其在跟踪过程中处于低维流形上；2) 通过对高维视觉特征进行空间区隔，可以实现目标和周围背景的互补学习，支撑当前鉴别性空间位置将被激活，形成稳定空间支撑模式，充分利用了前景背景中互补信息；3) 通过对不相关空间位置抑制，大大减少干扰信息，增强了算法鲁棒性。

如图 1 所示，在目标和背景都存在复杂变化

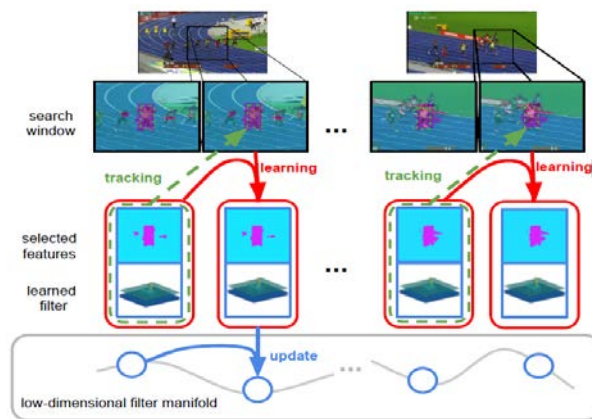


图 1 基于空间特征选择的自适应跟踪算法

的情况下，基于空间特征选择的自适应跟踪算法能够学习到合适的空间支撑模式，将前景和背景信息中的互补信息进行协同考虑。通过将原本的背景干扰物和目标进行协同考虑，增强了对目标本身的支持，同时也能够在目标变化过大的情况下依靠背景信息对目标位置进行合理推断。

上述算法在诸多目标跟踪数据集如 VOT2018, OTB100, TC128 和 UAV123 都取得了较高的跟踪精度和鲁棒性。该研究的相关成果在 VOT2018 竞赛中取得了公开数据库第一名的成绩（代码下载：<https://github.com/XU-TIANYANG/LADCF>），并被国家自然科学基金委报导“视觉跟踪重要进展”（<http://www.nsf.gov.cn/publish/portal0/tab434/info74717.htm>）。（责任编辑：王金甲）



徐天阳

江南大学在读博士研究生，目前在英国萨里大学计算机视觉，语音和信号处理中心进行为期一年的联合培养。主要研究领域包括视觉目标跟踪和机器学习。

Email: tianyang.xu@surrey.ac.uk



吴小俊

江南大学二级教授，博导，江南大学科学技术研究院院长，江苏省人工智能学会副理事长。主要研究方向为模式识别、计算机视觉、智能计算和机器学习。

Email: wu_xiaojun@jiangnan.edu

融合单阶段和两阶段检测思想的物体检测算法

中科院自动化所 张士峰 雷震 李子青

物体检测是视觉感知的第一步，也是计算机视觉的一个重要分支。物体检测的目标是用框去标出物体的位置，并给出物体的类别。如图 1 所示，目前基于深度学习的物体检测算法主要分为单阶段方法和两阶段方法。通常情况下，单阶段方法检测速度更快，而两阶段方法检测精度更高。

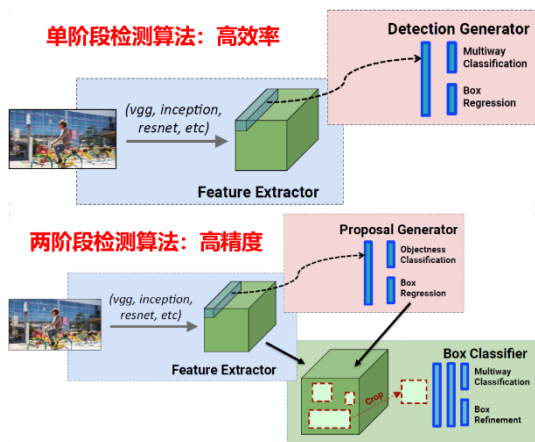


图 1 单阶段和两阶段检测算法的对比

为了继承两种方法各自优点并克服它们缺点，我们提出了 RefineDet 通用物体检测算法。该算法融合了单阶段和两阶段方法的检测思想，在准确率和速度上达到了很好的均衡。RefineDet 算法是在网络结构上对单阶段方法进行改进，以模拟两阶段方法的检测过程，从而具备两阶段方法优势并能够保持单阶段方法速度。

如图 2 所示，RefineDet 算法由锚点框校正模块和物体检测模块组成，它们之间由传输连接

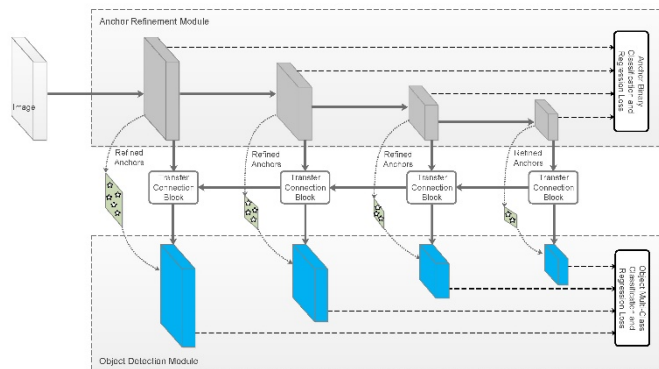
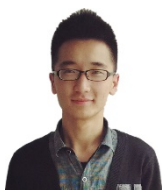


图 2 RefineDet 算法的网络架构示意图

块连接。锚点框校正模块旨在滤除简单的负样本以减少后续分类器的搜索空间，同时粗略地调整锚点框的位置和大小为后续回归器提供更好的初始化。物体检测模块把经过前者矫正过的锚点框作为输入，以进一步提高回归精度并预测多类别标签。由于锚点框校正模块中的特征被用于二分类，已具备一定鉴别能力，传输连接块把锚点框校正模块中这些有鉴别力的特征，传递给物体检测模块，以更好地预测物体位置，大小和类别标签。通过这种网络结构，RefineDet 能够在保持单阶段方法的速度前提下，具备两阶段方法的二阶段分类、二阶段回归、二阶段特征这三个优点。因此该算法能够在常用的通用物体检测数据库上取得最好检测精度的同时，保持着较快检测速度，在准确率和速度上达到了很好均衡。该研究成果发表在计算机视觉国际会议 CVPR 2018。

(责任编辑：崔海楠)



张士峰

中科院自动化所博士。主要研究方向为物体检测。

Email: shifeng.zhang@nlpr.ia.ac.cn



雷震

中科院自动化所研究员。主要研究方向为生物特征识别，计算机视觉。

Email: zlei@nlpr.ia.ac.cn



李子青

中科院自动化所研究员，IEEE Fellow。主要研究方向为生物特征识别，计算机视觉。

Email: szli@nlpr.ia.ac.cn

基于双链区分性卷积神经网络的多视角 3D 模型检索算法

山东省科学院 山东省计算中心（国家超级计算中心） 高赞

随着传感器技术、存储技术和网络传输技术的快速发展，人们获得 3D 模型的途径也越来越多。随着 3D 模型应用领域的增广，通用模型数据库中 3D 模型的种类及数目也随之增长，如何快速精确查找所需的 3D 模型也逐渐成为一个热门研究话题。然而，3D 模型不仅受到光照、尺度、物体形变和对象之间差异等因素影响，而且其本身复杂性也给 3D 模型特征描述带来了很大困难，这些困难给 3D 模型检索带来了巨大挑战。

近年来，研究人员提出了很多不同的 3D 模型检索算法，其中基于多视角的 3D 模型算法的性能表现尤为突出（将 3D 模型虚拟投影为多视角图像），该类算法大致可以分为两类：基于手工设计特征的算法和基于特征学习的算法。基于手工设计特征的算法首先通过手工设计特征对每张图像进行特征提取，然后，通过不同的检索算法计算图像之间的相似程度。由于特征提取和检索过程不统一，特征的区分性和检索性能都有待提高。基于特征学习的算法将特征提取和检索过程统一到同一个框架中，其所学习的特征更适合于后续的检索过程。这类方法的成功关键在于：1) 如何自动地挖掘多视角样本潜在关联特性。2) 如何提高所提取特征的区分性。3) 如何高效地对模型进行训练。

针对这些问题，我们提出了基于双链区分性卷积神经网络的多视角 3D 模型检索算法。该方法通过结合对比损失和对比中心损失函数，从训

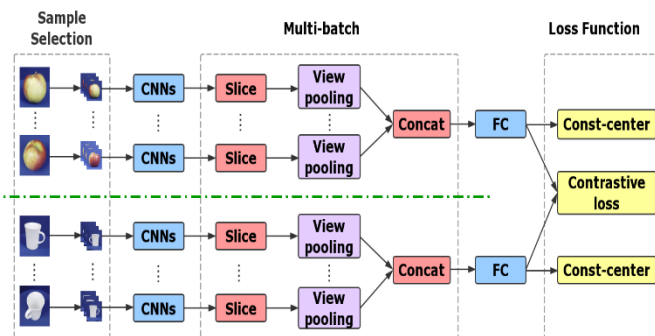


图 1 双链区分性卷积神经网络的框图

练样本中，自动选择性地蒸馏出最富信息的训练样本，高效地对网络进行训练，提高所提取特征的区分性。同时，通过加入切分层和连接层，使得网络能够同时输入多视图和多个 Batch 的训练样本，从而挖掘了多视角样本的关联特性，加快网络的收敛速度和提高模型的检索性能。如图 1 所示，该算法一共包含两条链，分别对应的是 Query 链和 Gallery 链。在训练过程中，两条链的输入样本均来自于训练样本，且两条链的输入均为经过蒸馏的多视角样本对。通过该端到端的双链网络，不仅能够减少对原始数据集中带标签训练样本的要求，而且能够自动挖掘多视角样本的潜在关联特性，同时，使得所学习特征具有更好的类内和类间的特性，从而增强了所提取特征的区分性，提高了 3D 模型检索的性能。

该研究相关成果发表在 IEEE CVPR、AAAI 和 ACM MM 上。

（责任编辑：任桐炜）



高赞

山东省科学院 山东省计算中心（国家超级计算济南中心），研究员，博士生导师，主要研究方向为人工智能、图像/视频分析、信息检索和机器学习及其应用。Email: gaoz@sdas.org

西安电子科技大学高新波教授访谈

2019年3月12日，专委秘书处采访了西安电子科技大学副校长，国家万人计划科技创新领军人才，新世纪百千万人才工程国家级人选，国家杰出青年科学基金获得者，教育部长江学者特聘教授高新波教授。下面是采访实录。

高教授，您在视觉计算和协同认知领域取得了杰出的成就，获得国家“万人计划”，新世纪百千万人才工程国家级人选，长江学者和国家杰青等人才项目，在自己的研究领域里，请问您是如何做出自己的特色和成就的？

视觉计算和协同认知领域发展非常快，研究热点层出不穷，新的成果不断涌现，没有哪个研究者能够在某个方向上一直保持领先，可以说“江山代有人才出，各领风骚三五年”。即便如此，要做出自己的特色和成就，首先需要拥有自己稳定的研究方向，切不可做跟风的研究。当然，稳定的研究方向也并不是一成不变的，还是要根据问题导向和需求牵引做到与时俱进才行。同时，在选定的研究方向上，必须要勇于解决难题。正如丘成桐先生说的“只有选择一流的问题去研究，才可能取得一流的成果”。其次，需要有“十年磨一剑”的定力和耐力，要像历史学家范文澜先生说的那样“板凳要坐十年冷，文章不写一句空”才行。否则，急于求成往往是欲速则不达。第三，还要依靠团队的力量，今天的世界已不再是牛顿时代，靠单枪匹马很难解决大的问题。尤其是涉及到工程应用或者大系统工程的科研攻关，需要发挥团队的优势。因为没有完美的个人，但可以有完美的团队。这时，团队建设就变得尤为重要，团队成员必须拥有共同的理想和价值观，只有如此大家才能心往一处使，劲往一处用，才能咬定青山不放松。如果说我在过去几年取得了一点成果的话，首先，应该感谢我的三个导师。谢维信教授把我带入无监督机器学习这个领域，饭田弘



之教授教给了博弈论和搜索策略，汤晓鸥教授帮我确立异质图像合成及其质量评价这个具体的研究方向。其次，要感谢我的团队，他们与我一起二十几年如一日，做出了系统深入的工作，取得较好的应用效果。

作为一个科研工作者，既要从事基础研究，又要从事工程技术应用，请问您在这个方面是如何考虑的？特别是在指导研究生时，您是如何给学生选择课题的，是纯科学还是纯应用？

做研究既要有长远目标，也要有近期目标，两者兼顾，才能保持一个人的研究定力和研究热情。基础研究很难，但是会有长久的生命力，如果能取得突破会对科学进步作出积极贡献，但是必须要有坚定信念、坚强意志和毅力。工程技术研究相对而言短平快，而且容易产生成就感和满足感，但是应用技术发展很快，必须不断创新才能保持领先。因此，在选择科研方向时，我喜欢选择理论和实践相结合的方向。在给研究生选题时，我总是根据学生的基础和个人天赋来选择课题，但基本上都是基础研究与应用研究相结合的问题，只不过根据不同人的特点会有所侧重。希望先让学生在应用上取得一点成就感后，增强科研攻关的动力和热情，这样就可以做到“无需扬鞭自奋蹄”了。内生动力是学生取得无论是基础研究还是应用研究成果的核心和关键。

十五年磨一剑，您和团队一起用胸中的“执着”与手中的“极致”，诠释了千锤百炼之钢背后的科研初心，获得了国家自然科学奖二等奖，分享一下您的成功经验？



2016年我团队的研究成果“图像结构建模与视觉表观重构理论方法研究”获得了国家自然科学二等奖。该成果提出了对图像结构基本规律的描述和建模表示方法，以及人眼对图像结构的感知特性和基于该特性的图像重建方法。如果存在所谓经验的话，我想主要有如下三点：一是发扬团队合作精神，十年磨一剑；二是重视规律的发现和形式化的描述；三是成果的凝练和提升具有画龙点睛的作用。我们这项成果通过国际合作从2001年开始历时15年，揭示了图像结构在时空与尺度变化上的一致性规律、图像结构显著度与视觉感知灵敏度之间存在同向激励关系，以及图像结构统计分布与视觉感知之间的内在规律。基于这些规律，分别提出了图像全局-局部结构一体化建模方法，建立了局部-非局部结构感知的表观重构模型，设计了视觉感知融合滤波的质量评价框架。最后，将这三方面的成果凝练为基于三元空间融合的视觉计算理论与方法这一框架，使成果更为系统，而且上升到了一定的理论高度。

您治学严谨，主张打破“天花板效应”，将一件事做到极致，请问您是如何坚持的？当您处于

科研瓶颈期的时候，您是如何渡过难关，最后取得成功的？

从事科学研究与工人做工和农民种地是一样的道理，都得摸索规律、积累经验、不断实践。更重要的是热爱自己的工作，因为只有热爱，才能坚持；只有热爱，在遇到困难的时候才不会退缩和逃避；只有热爱，才会“为伊消得人憔悴，衣带渐宽终不悔”。在研究过程中，感到最艰难和最痛苦的是由于能力不足而使研究陷入困境，且经过长时间攻关而毫无进展的时候。在这个过程中，时而信心丧失，准备放弃；时而灵光闪现，兴奋激动；时而试验败北，倔强坚持。研究的过程就是各种情感遍历的过程，在这个过程中磨砺了自己，提升了境界，虽未必成功却一定有所成长。因此，要想坚持初心攻克难关，就必须培养自己宁静的心态和坚忍不拔的意志、毅力和信念。要具备这些品质，除了拥有良好的专业素质以外，还必须努力提高自己的人文素养和人文精神。

您入选教育部教学创新团队，并荣获陕西省优秀青年教师标兵、校内教学名师和“师德标兵”称号，为高校教师竖立了一个典范。分享一下您在教学研究方面的心得？

作为高校教师，立德树人是最根本的任务。因此，从参加工作以来，我坚持为本科生上课。在课堂上除讲授教学大纲内容外，我喜欢跟大家分享课本上每个知识点背后的故事。这往往能让同学们了解每个理论的形成过程，了解每位学者的成长经历，提高学习兴趣，并树立自己的学习榜样和奋斗目标。这就要求我们每位教师加强教学研究，一方面扩大自己的知识面，对每个知识点做到“知其然”，“知其所以然”，能告诉学生这个知识点的推广应用情况；另一方面，要研究教育学、心理学，设计教学环节加强互动以激发大家的学习热情。之所以说，教学是一门艺术，就是因为教学有法，但无定法，必须根据不同的场景、不同的学生，因材施教才行；第三，教学和科研必须相互融合才能具有生命力，科研反哺教

学,就会为教学提供鲜活的案例,将最新研究成果作为教学内容能激发学生的学习和研究热情。教学不仅仅是传授知识,更重要的是“唤醒”。正如德国著名教育学家斯普朗格所说:“教育的最终目的不是传授已有的东西,而是要把人的创造力诱导出来,将生命感、价值感唤醒”。因此,教师是一个伟大的职业,我们必须尽职尽责。

十多年来,您和团队成员都取得了特别优秀的成绩,能否介绍一下您的科研团队?

2001年我从香港中文大学多媒体实验室博士后出站回到西电,在学校和学院的大力支持下组建了“影像处理系统留学归国人员实验室”,并与香港中文大学多媒体实验室走出的很多学者保持科研上的密切合作关系。在这十余年中,我们研究团队坚持异质图像合成和质量评价这一研究方向,先后获得陕西省重点科技创新团队、教育部长江学者创新团队,以及科技部重点领域创新团队支持。我们团队的主要特点有三:一是坚持应用基础研究,十几年如一日,不因理论研究枯燥而放弃。坚持探索图像结构建模与视觉表观重构中蕴含的规律,并对这些规律建模、设计新算法,进而研究模型的复杂性、算法的收敛性等基础问题。二是加强国际合作,团队成员均有海外留学经历,并与国际同行保持密切合作关系。通过海外留学和工作经历,一方面开阔了眼界,另一方面积累了丰富的学术资源,为科学研究的深入开展奠定了基础。三是团队成员年富力强正处于创新的旺盛期,目前我们团队的平均年龄不足四十岁,正处于人生的黄金期,还可以在本领域拼搏20年,希望还能创造出更多更好的成果。

全世界的高等学校都面对着同一个问题,到底是教学重要还是科研重要,请问您是如何看待这个问题的?您是如何将科研与教学进行完美结合的?

教学与科研之间的辩证关系是“只教不研,越教越死;只研不教,越研越窄”。因此,如果能

正确处理好两者之间的关系就会使之相互促进,反之,则会相互制约。作为高等院校的老师,立德树人是第一位的,教学是我们的立身之本;而科研是为了发现和创造新知识,深化和推广知识的应用,从而提升教师的创造力和业务素质,更好的带领学生进入知识的海洋遨游。因此,我们决不能舍本逐末,重科研而轻教学。我本人在近二十年的从教生涯中,坚持教研相长。一方面,不断把最新的科研成果转化为结构化的知识融入课堂,使课堂教学的知识性、趣味性、时代性相结合。这一过程不是简单的引入,而是要将新成果与课本知识之间建立联系,保持知识结构的连贯性和可扩展性,这需要深入的思考和细致的工作。另一方面,吸引更多的同学参与到我的科研之中,发挥他们初生牛犊不怕虎的敢想敢干的优势,为科研团队增添活力,带来更多的新思维。多年的经验告诉我们,一个人接受教育的时间越长,知识积累越丰富,创造力和冒险精神就会有所下降。而对于我们的本科生,虽然科研积累不足,但是勇于尝试、敢于冒险,往往会取得意想不到的新发现。也正是基于以上认识,多年来我坚持教学科研并重,努力做到教研相长。

所指导的博士生三人获得陕西省优秀博士学位论文、四人获得一级学会优博论文奖,一人入选中国科协青年人才托举计划,一人获得2018年第五届“工程硕士实习实践优秀成果获得者”。能否分享一下您在研究生培养方面的经验?

近年来,我的研究生取得了较好的成绩,这首先是他们自己努力的结果。其次,我们实验室的学风比较好,在这样的氛围中,每位同学都争先恐后,形成一种良性循环。在研究生培养方面,我基本上是采取宽松平等的交流和指导方式。首先为每位同学指定研究的大方向,请学生在了解该方向的研究前沿基础上自行确定拟解决的问题。当然我会根据对学生的了解给出相应的建议,比如数学基础好的同学,可以偏重理论研究类的工作,动手能力强的,可以偏重工程应用型的工作。但是,不论哪一类型,均要求理论创新与工

程应用相结合。在日常学习生活中，我注意加强学生的自我管理和相互学习能力。实践证明，在研究生学习阶段，同学们的基本技能有 60% 左右来自同学之间的相互学习。因此，导师需要做的是建立和维护一个良好的学习共同体。这就要求，一方面要有好的生源，让同学们与“高人”为伍；另一方面指定合适的研究方向，以保证学生的目标与能力相适应，让学生的成就感化作前进的动力；第三方面就是学习风气的养成和科研氛围的营造。做到这三点，学生们就会受到感染，学习有榜样，奋斗有目标，个人有追求。

您主持包括国家杰出青年科学基金项目、国家自然科学基金重点项目、面上项目、教育部重点项目、陕西省重点科技攻关项目在内的纵向研究课题 20 余项，能否分享一下您在科研项目的选题和申报方面的经验？

科研选题得结合自己的研究兴趣和研究积累。一个人要有自己稳定的研究方向，切忌经常更换方向的跟风式研究。在自己研究方向的基础上，要申报课题就得与学科前沿、国家需求和指南方向相结合。只做自己感兴趣的研究，不结合问题导向和需求牵引，很难获得立项支持。要使课题获得支持就必须能解决国家和社会面临的真问题，否则闭门造车式、修修补补的研究是没有前途的。近年来，我们团队在异质图像合成和质量评价基础理论研究的基础上，与公共安全相结合，在视频监控、画像追凶、消费电子等方面取得了多项课题的立项。一方面，我们通过国家自然科学基金的支持做自由研究，深化应用基础成果；另一方面不断创新应用领域，加强交叉研究，同时与实际应用结合，将理论成果进行推广应用，以提高研究成果的应用价值和社会效益。

您经常参加各类国际学术会议并多次担任大会程序委员会主席，能否结合您的经历，谈一下您认为做好学术报告中需要注意什么问题？

作为一名学者，首先要做好自己的研究，其

次还要加强学术交流，这是学者的任务也是重要的使命。参加国际学术会议是加强学术交流的重要途径，一方面通过这样的平台可以与国际同行面对面同台交流；另一方面也是一个展示自我、宣讲自己科研成果的舞台。这样看来，在学术会议上做好学术报告就显得非常重要了。如何才能做好自己的学术报告呢？首先，需要准备一个图文并茂的演讲提纲，根据自己的学术成果设计一些直观新颖的示意图非常关键。这需要换位思考，要把自己置于听众的角度来看需要怎样的讲解；其次，要注意把演讲的重点放在创新大思路的讲解上，在学术会议上不宜把重点放在技术细节上，交流思想更为重要；第三，互动环节也非常重要，这个阶段的收获往往最大。通过听众的提问，既可以了解到自己在哪些方面没有阐述清楚，又可以通过这些问题找到自己研究的薄弱环节，而且可能有助于找到下一步改进的方向和创新的突破口。因此，参加学术会议千万不要只把重点放在社交上，还应该把学术报告做好，这是个人研究的二次提升。所以，每一位研究者在做好研究的同时还要做好学术报告才行。这既能扩大一个人的学术影响力，也能换一个角度审视自己的工作，找到提升和深化已有研究的新方向。



高新波

博士，教授，西安电子科技大学副校长，国家万人计划科技创新领军人才，新世纪百千万人才工程国家级人选，国家杰出青年科学基金获得者，教育部长江学者特聘教授，科技部重点领域创新团队负责人、教育部创新团队负责人，教育部高等学校计算机类专业教学指导委员会委员。IET Fellow、CIE Fellow、IEEE 高级会员、中国图象图形学学会常务理事、中国计算机学会理事、中国电子学会理事。主要从事机器学习、计算机视觉、模式识别等领域的研究和教学工作，发表 SCI 论文 200 余篇，引用 12000 余次，获国家自然科学基金二等奖 1 项、省部级科学技术一等奖 3 项。

（责任编辑：张汗灵 余焯 黄岩）

委员好消息

✪ 2018年12月19日，在AI Challenger 2018总决赛上，CCF-CV专委会委员、中科院自动化所张一帆副研究员指导的CASIA-AIRIA团队在“短视频实时分类”赛道中获得该赛道亚军。本次比赛共有来自81个国家、1100所高校、990家公司的上万支参赛团队参赛。

✪ 2019年1月4日，CCF共26位在各自领域取得突出成就或对CCF服务方面有卓越表现的会员晋升为杰出会员，其中包括CCF-CV专委会3位委员：大连理工大学卢湖川教授、西南科技大学吴亚东教授、山东大学尹义龙教授。

✪ 2019年1月17日，Elsevier发布2018年中国高被引学者榜单，本次国内共1899位学者入选。CCF-CV专委会18位委员入选：华中科技大学白翔教授、浙江工业大学陈胜勇教授、南京航空航天大学陈松灿教授、南开大学程明明教授、西安电子科技大学高新波教授、清华大学郭振华副研究员、中国科学院自动化研究所刘成林研究员、清华大学鲁继文副教授、深圳大学沈琳琳教授、中国科学院自动化研究所谭铁牛研究员、南京航空航天大学谭晓阳教授、同济大学王瀚漓教授、中国科学院自动化研究所王亮研究员、中国科学院自动化研究所王鹏副研究员、哈尔滨工业大学邬向前教授、哈尔滨工业大学徐勇教授、北京理工大学杨健教授、上海交通大学杨小康教授。

✪ 2019年1月18日，中国科协公布了第四届（2018-2020年度）中国科协青年人才托举工程入选者名单，共335名青年科技工作者入选。CCF-CV专委会2位委员、南京理工大学张珊珊教授和中科院自动化所樊彬副研究员入选。

✪ 2019年1月19日，CCF-CV专委会荣获CCF 2018年度特别奖。

✪ 2019年1月27日，AAAI 2019杰出程序委员会成员名单公布，共10人获奖。CCF-CV专委

会常务委员、华中科技大学白翔教授被授予杰出高级程序委员会奖。

✪ 2019年2月16日，CCF广州会员活动中心成功举行了换届选举会议，CCF-CV专委会委员、华南理工大学许勇教授当选新一届主席，CCF-CV专委会委员、中山大学王昌栋副教授当选副主席。

✪ 2019年2月27日获悉，第四批国家“万人计划”名单公布，CCF-CV专委会8位委员入选：北京大学彭宇新教授、西北工业大学韩军伟教授、中科院自动化所王亮研究员入选科技创新领军人才，南京理工大学李泽超教授、西安交通大学孟德宇教授、西北工业大学王琦教授、中国科学院自动化研究所张兆翔研究员、银河水滴科技（北京）有限公司创始人兼CEO黄永祯博士入选青年拔尖人才。第四批国家“万人计划”入选人员名单包括两个层次，共计1419人，其中科技创新领军人才645人，科技创业领军人才397人，教学名师103人；青年拔尖人才274人。

✪ 2019年3月1日，2018年度北京市科学技术奖励大会颁奖，由CCF-CV专委会委员、中国科学院自动化研究所王亮研究员、银河水滴科技（北京）有限公司创始人兼CEO黄永祯博士、中国科学院自动化研究所谭铁牛院士完成的“远距离步态识别系统研究与应用”获二等奖。2018年度共有212项成果荣获北京市科学技术奖，包括一等奖24项，二等奖58项，三等奖130项。

✪ 2019年3月6日获悉，由CCF-CV专委会委员、南开大学程明明教授和北京理工大学沈建冰教授等完成的 *Shifting More Attention to Video Salient Object Detection* 论文高分被CVPV2019收录为Oral，从5165篇论文中脱颖而出。该文工作使得AI的注意力转移机制更加逼近人类。

（责任编辑：刘海波）

灰度图像自动着色技术开源代码

北京电子科技学院 于浩洋 金鑫

灰度图像 (gray image) 是每个像素只有一个采样颜色的图像, 即单通道图像, 这类图像通常显示为从最暗黑色到最亮白色的灰度, 理论上这个采样可以任何颜色的不同深浅, 甚至可以是不同亮度上的不同颜色。

灰度图像着色是将单通道的灰度图像, 通过一定的操作, 使其成为三通道的彩色图像的方法, 在影视处理、数字娱乐和动漫制作等方面有着广泛的应用前景。

在灰度图像着色领域, 传统方法信息提取率不高, 着色效果不理想, 但随着人工智能的不断发展, 灰度图像着色的效果越来越让人满意。目前灰度图像着色一般模式为: 首先利用神经网络的信息提取高效性, 对图像中的各类信息及特征进行提取, 构建并训练深度学习模型。训练网络时与原图像进行对比, 逐渐减小网络输出结果的信息、分类等各类型的损失。训练完成后, 只需向网络输入一张灰度图片, 即可生成一张颜色饱满、鲜明逼真的彩色图片。图 1 为灰度图像着色示意图。



图 1 灰度图像着色示意图

本文着重介绍几个基于深度学习的灰度图像着色技术的开源代码, 包括训练深度学习模型的常用数据集和几个经典的图像着色深度神经网络的模型。

1. ImageNet 数据集

介绍: ImageNet 是一个计算机视觉系统识别项目, 是目前世界上图像识别最大的数据库, 是美国斯坦福的计算机科学家, 模拟人类的识别系统建立的, 能够从图片识别物体。ImageNet 是非常有前景的研究项目, 未来用在机器人身上, 就可以直接辨认物品和人。超过 1400 万的图像

URL 被 ImageNet 手动注释, 以指示图片中的对象; 在至少一百万张图像中, 还提供了边界框。ImageNet 包含 2 万多个类别; 一个典型的类别, 如“气球”或“草莓”, 包含数百个图像。

主页: <http://www.image-net.org>

2. Place365 数据集

介绍: Places365 是 Places2 数据库的最新子集。Places365 有两个版本: Places365-Standard 和 Places365-Challenge。Places365-Standard 系列列车拥有来自 365 个场景类别的约 180 万张图像, 每张图像最多可存储 5000 张图像类别。Places365-Challenge 的数据集还有 620 万张图片以及 Places365-Standard 的所有图片(共 800 万张图片), 每个类别最多有 40,000 张图片。

主页: <http://places2.csail.mit.edu>

3. Colorful image colorization(CNN 模型)

论文: Zhang R, Isola P, Efros A A. Colorful image colorization[C]. European conference on computer vision, 2016: 649-666.

工作: 该方案在自动图像着色的图形学领域取得了进步: 设计了一个合适的损失函数来处理着色问题中的多模不确定性, 维持了颜色的多样性; 介绍了一种新型的着色算法评估框架, 而且这种评估框架有应用到其他图像合成任务的潜力; 通过在百万数量级的彩色图片上训练在这类任务上抵达了一个新的水准。该方法将图像着色任务转化为一个自监督表达学习的任务, 并且在一些基准上获得了最好的效果。图 2 显示网络结构。

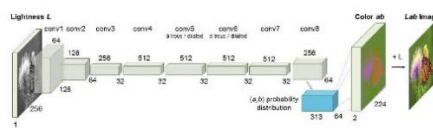


图 2 Colorful image colorization 网络结构图

代码: <http://richzhang.github.io/colorization>

4. Image Colorization with Generative Adversarial Networks (GAN 模型)

论文: Nazeri K, Ng E. Image Colorization with Generative Adversarial Networks. arXiv preprint arXiv:1803.05400,2018.

工作: 该方法使用 GAN 自动将灰度图像着色到可接受的视觉程度的彩色图像。将当前方法扩展到高分辨率图像。网络是针对公开可用的数据集（如 CIFAR-10 和 Places365）进行训练。图 3 为该网络结构图:

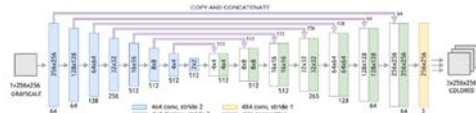


图 3 网络结构图

5. Interactive Deep Colorization(CNN 模型)

论文: Zhang R, Zhu J Y, Isola P, et al. Real-time user-guided image colorization with learned deep priors. arXiv preprint arXiv:1705.02999, 2017.

工作: 该方法是一种用于用户引导的图像着色的深度学习方法。系统将灰度图像以及本地用户的提示共同映射到卷积神经网络(CNN)中去,最终输出彩色图像。该方法通过模拟用户输入训练一百万张图像。为了引导用户进行有效的输入选择,系统根据输入图像和当前用户输入推荐可能的颜色作为备选颜色。着色在单个前馈过程中执行,可实时使用。图 4 为该方法网络结构图:

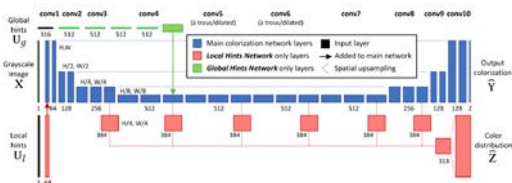


图 4 Interactive Deep Colorization 网络结构图

代码: <https://richzhang.github.io/ideepcolor>

6. Automatic Colorization of Grayscale Images(CNN 模型)

论文: Lizuka S, Simo-Serra E, Ishikawa H. Let there be color!: joint end-to-end learning of global and local image priors for automatic image colorization with simultaneous classification[J]. ACM Transactions on Graphics, 2016, 35(4):

110-120.

工作: 该方法结合了全局先验和局部图像特征。基于卷积神经网络,加入深层网络具有融合层,允许网络合并小图像块的局部信息和整个图像的全局信息,以端到端的方式进行训练。此外,与大多数基于 CNN 的现有方法不同,该架构可以处理任何分辨率的图像。利用现有的大型场景分类数据库来训练模型,利用数据集的类标签来更有效地和区别地学习局部信息和全局信息。图 5 为模型示意图:

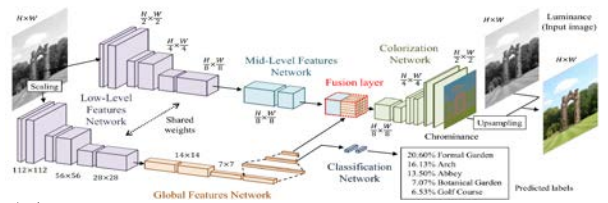


图 5 Automatic Colorization of Grayscale Images 模型示意图

代码: <https://github.com/williamFalcon/pix2pix-keras>

7. Deep Exemplar-based Colorization(CNN 模型)

论文: He M, Chen D, Liao J, et al. Deep exemplar-based colorization[J]. ACM Transactions on Graphics, 2018, 37(4): 47-62..

工作: 该方法是一种基于样本的局部着色的深度学习方法。该方法使用的端到端着色网络不是像传统的基于示例的方法那样使用手工制作的规则,而是学习如何从大规模数据中选择传播方式和预测颜色。即使在使用与输入灰度图像无关的参考图像时,该方法也可以很好地执行并且很好地推广。该网络允许用户通过简单地提供不同的参考来实现可定制的结果,只需选择顶部参考建议,就可以完全自动执行着色。

代码: <https://github.com/msracver/Deep-Exemplar-based-Colorization>

(责任编辑:刘丽,樊鑫)



于浩洋

北京电子科技学院在读研究生,研究方向为同态加密,图像生成。

Email:pec7@163.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

唇读（视觉语音识别）数据集

国防科技大学 陈小鼎 盛常冲 刘丽

唇读 (Lip Reading), 也称视觉语音识别 (Visual Speech Recognition), 通过说话者口型变化信息推断其所说的内容, 旨在利用视觉信道信息补充听觉信道信息, 在现实生活中有重要应用。例如, 应用在医疗领域辅助听力受损的病人提高沟通交流能力, 在军事领域提高情报获取和处理能力, 在多媒体领域提高人机交互的多样性和鲁棒性等。随着深度学习技术的发展, 以及数据集规模的不断完善, 基于深度学习的框架方法已经逐渐取代传统方法, 成为唇读的主流方法。

本文对构建自动唇读系统过程中常用到的数据集进行总结介绍。根据自动唇读系统解决的目标任务不同, 可将现有数据集分为字母、数字数据集, 单词、短语数据集和语句数据集三大类。

一、字母、数字数据集

1. AVLetters 数据集

数据集地址: <http://www2.cmp.uea.ac.uk/~bit/avletters>

AVLetters 数据集由英国东英吉利大学、曼彻斯特大学团队于 1998 年创建, 是第一个视听语音数据集。数据集初始包含 10 个说话人, 每个人分别 3 次独立的陈述 26 个英文字母, 总计 780 个话语实例。由于提出年份较早, 拍摄条件有限, 视频分辨率较低, 仅为 376 像素*288 像素, 每秒 25 帧。在手动定位每张图像中嘴唇的位置之后, 将整个图像裁剪至 80 像素*60 像素, 形成最终的数据集。



图 1 AVLetters 数据集样本

2. XM2VTS 数据集

数据集地址: <http://www.ee.surrey.ac.uk/CVSSP/xm2vtsdb>

XM2VTS 数据集是从 M2VTS 项目 (Multi Modal Verification for Teleservices and Security applications) 中获得, 由英国萨里大学、瑞士 IDIAP 研究所团队于 1999 年提出, 该数据集

创建的初衷是为研究团队提供高质量数据来测试多模态人脸验证算法。共 295 名志愿者参加了该数据集的记录, 每个志愿者以正常语速读两个数字序列和一个语音平衡的句子 (10 个数字、7 个单词), 这个过程进行两次。此外, 考虑到志愿者自然变化的影响, 在五个月时间内平均进行 4 次记录, 总计 7080 个话语实例。值得一提的是, 该数据集考虑了头部姿态变化因素, 并记录了志愿者在不同角度的头部图像 (头部姿态变化的过程中未说话)。



图 2 XM2VTS 数据集样本

3. BANCA 数据集

数据集地址: <http://www.ee.surrey.ac.uk/CVSSP/banca>

BANCA 数据集是由瑞士 IDIAP 研究所、西班牙卡洛斯三世大学、英国萨里大学团队于 2003 年创建, 旨在训练和测试多模态身份验证系统。该数据集由四种不同的语言 (英语、法语、意大利语、西班牙语) 进行记录, 并且在三个环境条件下 (controlled, degraded and adverse) 进行拍摄。数据量也有了一个大的提升, 总共有 208 个参与者, 将近 30000 个话语实例。参与者除了读一个数字序列之外, 还需要说出自己的姓名、住址、生日等信息。BANCA 数据集为不仅研究团体提供了在具有挑战性的数据集上测试多模式身份验证算法的机会, 也同时推动了唇读研究的发展。

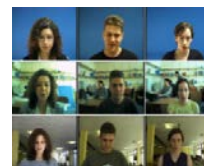


图 3 BANCA 数据集样本

二、单词、短语数据集

4. GRID 数据集

数据集地址: <http://spandh.dcs.shef.ac.uk/gridcorpus>

GRID 数据集是由美国谢菲尔德大学团队于 2006 年提出,旨在为语音感知和自动语音识别研究提供实验数据。该数据集在实验室环境下录制,只有 34 个志愿者,这在大型数据集中人数算比较少的,但每个志愿者说 1000 个短语,共 34000 个话语实例。该数据集短语构成符合一定的规律,每个短语包含 6 个单词,不是常见的短语,而是在 6 类单词中每类随机挑选一个组成随机短语。这 6 类单词分别是“命令”、“颜色”、“介词”、“字母”、“数字”和“副词”,例如: Bin blue at A 1 again。每类单词规定了数量,单词总数共 51 个。数据集是完全公开的,不需要与发布者联系,也不需要填写保密协议即可在网上下载使用。

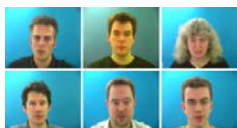


图 4 GRID 数据集样本

5. OuluVS 数据集

数据集地址: <https://www oulu.fi/cmvs/node/4131>

5

OuluVS 数据集是由芬兰奥卢大学团队于 2009 年发布,旨在为视听语音识别系统进行性能评估提供一个统一的标准。该数据集包含 20 个参与者,每名参与者陈述 10 个日常问候短语 5 次,一共 1000 个话语实例。OuluVS 数据集是最早几个针对短语任务而构建的数据集之一,将唇读(自动语音识别)系统的发展推向了一个新的起点。



图 5 OuluVS 数据集样本

6. LRW 数据集

数据集地址: http://www robots.ox.ac.uk/%7Eevgg/data/lip_reading/lrw1.html

LRW 数据集是由牛津大学视觉几何团队于 2016 年提出。因深度学习的兴起,大规模数据集的需求越来越大,LRW 数据集应运而生。不同于以往数据集,LRW 数据集数据来源于 BBC 广播电视节目而不是由志愿者或实验人员录制,使得该数据集数据量有了质的飞跃。数据集选择了 500

最常出现的单词,截取说话人说这些单词的镜头,因此说话人超过 1000 个,话语实例超过 550000 万个,一定程度上满足了深度学习对于数据量的需求。



图 6 LRW 数据集样本

7. LRW-1000 数据集

数据集地址: http://vipl.ict.ac.cn/view_database.php?id=14

LRW-1000 数据集是由中科院计算所、中国科学院大学和华中科技大学团队于 2018 年提出,旨在建立一个在室外环境下并且图像尺寸不一的大规模基准。该数据集涵盖了不同语音模式和成像条件下的自然变化,以应对实际应用中遇到的挑战。该数据集来源于中文电视节目,包含 1000 个类,每一个类对应由一个或几个汉字组成的汉语单词。该数据集是规模最大的中文单词唇读数据集,截取的镜头包括超过 2000 个说话人,将近 720000 个话语实例。该数据集数据的丰富性保证了深度学习模型得到充分的训练。同时,该数据集也是唯一一个公开的中文普通话唇读数据集。



图 7 LRW-1000 数据集样本

三、语句数据集

8. AVICAR 数据集

数据集地址: <http://www.isle.illinois.edu/sst/AVICAR/#information>

AVICAR 数据集是由美国伊利诺伊大学的研究团队于 2004 年提出,该数据集全部在汽车中进行拍摄,旨在创建一个带有环境噪声的数据集,为噪声条件下的唇读系统提供数据支持。AVICAR 数据集包括 100 名参与者,数据包括独立字母、独立数字、数字序列(10 个)和语音平衡的句子,一共 59000 个话语实例。为了模拟在车内的实际噪声情况,该数据集在 5 中不同条件下进行拍摄,分别是:时速 35 公里(开、关窗)、时速 55 公

里（开、关窗）和静止（引擎空转）。



图 8 AVICAR 数据集样本

9. MOBIO 数据集

数据集地址：<https://www.idiap.ch/dataset/mobio>

MOBIO 数据集由瑞士 IDIAP 研究所、芬兰奥卢大学、捷克布尔诺理工大学、英国萨里大学和法国阿维尼翁大学团队于 2012 年发布，该数据集基本上全部由记录者手持手机进行拍摄记录，旨在对手机上人脸语音自动识别系统进行评估，改进在移动设备上应用于生物识别技术的研究。该数据集是在不受控的条件下拍摄的，因为拍摄设备在记录人自己手中而不是固定在某一个特定的位置，记录者头部位置、背景、光照等因素都在发生变化。有 150 人参加了数据集构建，得到将近 31000 个话语实例，其语料可以分为三大类：五个提前定义好的问题的答案、对一个随机问题的长约 5 秒的回答以及提前定义好的一段文本。



图 9 MOBIO 数据集样本

10. OuluVS2 数据集

数据集地址：<http://www.ee.oulu.fi/research/imag/OuluVS2>

OuluVS2 数据集是由芬兰奥卢大学团队于 2015 年发布，是继 2009 年提出的短语数据集 OuluVS 之后在语句识别领域中的又一高质量数据集，旨在促进唇读的进一步研究。该数据集包含 53 名参与者，记录了将近 4000 个话语实例。同样包括三大类语料：连续数字序列、日常短语

和 TIMIT 语句。值得一提的是，OuluVS2 是为为数不多的多视角唇读数据集之一，采用 5 个高清摄像头同时拍摄，分别位于说话人的 0、30、45、60、90 度方向，分辨率也达到了很高的 1920 像素*1080 像素。



图 10 OuluVS2 数据集样本

11. LRS 数据集

数据集地址：http://www.robots.ox.ac.uk/~7Evagg/data/lip_reading/lrs2.html

LRS 数据集是由牛津大学视觉几何团队于 2017 年提出，是继大规模单词数据集 LRW 发布之后，针对句子任务构建的另一大规模唇读数据集。和 LRW 类似，该数据集也来源于 BBC 广播电视节目，同样处理开放世界（句子无任何限制）的问题。数据集包含超过 1000 个说话人，截取将近 150000 个话语实例，由于截取时对句子无限制，不同的单词数有将近 63000 个，数据丰富性极高，更好的适用于基于深度学习进行唇读模型的研究。随后于同年，为了研究侧面图像数据对唇读系统的影响，基于 LRS 的多视角数据集 MV-LRS 被提出，在 LRS 的基础上加入了 0~90 度所有角度的镜头，同时数据量也有所保证，达到将近 75000 个话语实例。



图 11 LRS 数据集样本
(责任编辑：李策，蹇木伟)



陈小鼎

硕士研究生，国防科技大学系统工程学院，主要研究方向为唇读。本科毕业于国防科技大学系统工程学院。



盛常冲

博士研究生，国防科技大学电子科学学院，主要研究方向为计算机视觉、深度学习。



刘丽

副教授，国防科技大学系统工程学院，主要研究方向为图像理解，计算机视觉，模式识别。

招聘信息

一、美国密苏里大学堪萨斯分校招收多媒体/计算机视觉方向博士/博士后

最近, 在 NSF、AFOSR 和工业界的资助下, UMKC 多媒体计算与通信 (MC2) 实验室在 2019 年秋季开设两个博士职位, 并获得全额资助 (包括学费、医疗保险和助学金)。

1. 研究课题:

- 未来视频编解码器研究, 新的编码工具, 尤其是新的帧内预测, 超分辨率预测和去块滤波解决方案中的深度学习技术;
- 沉浸式视觉传播研究, 点云, 光场和 360 视频捕获, 压缩和极低延迟通信。点云压缩中的计算离散几何和图形信号处理技术; 深度和传统的光场压缩和超分辨率方法;
- 深度卷积网络压缩和加速。用于视觉信号进行压缩、识别和再识别的嵌入式系统友好的深度学习性能;
- 针对大规模数据库的强大对象重新识别, 深度信息的新关键点特征, 高光谱成像, 高效的局部特征聚合和哈希。用于低分辨率/高质量的图像识别的差分关键点深度学习。

2. 应聘基本条件:

- 具有 CS / EE 的硕士学位;
- 具有 C / C ++, Python 和 Matlab 的扎实编程技能;
- 要求掌握的课程包括线性代数, 概率与统计, 数值优化;
- 有强烈的愿望从事于前沿研究。

3. 联系方式:

有意者请寄一份简短的简历、一份成绩单、发表过的文章 (如果有)、GRE/托福/雅思成绩 (特别优秀的考生可以放弃 GRE 考试的要求) 至:

Zhu Li

Associate Professor & Director

NSF I/UCRC Center for Big Learning at UMKC

Multimedia Computing & Communication (MC2) Lab

School of Computing & Engineering

University of Missouri, Kansas City, USA

Email: lizhu@umkc.edu or zhu.li@ieee.org

更多导师的相关背景及研究详情见: <http://l.web.umkc.edu/lizhu>

二、沙特阿卜杜拉国王科技大学 IVUL 组招收计算机视觉 VSRP 访问学生/硕博/博后

最近沙特阿卜杜拉国王科技大学 (KAUST, <https://www.kaust.edu.sa/en>) 针对国内 Top universities 的本科生和硕士生开放了 VSRP 访问学生项目申请。IVUL 组申请到若干名额 (PI 为 Prof. Bernard Ghanem), 欢迎热爱科研且有 Deep learning 基础的学生在以下 VSRP 网站尽快申请 (表现优秀可申请留下读研/博), 到校截止日期为 2019 年 6 月 30 日。

1. VSRP 项目福利

- 每月生活津贴 (1000 美元);
- 往返吉达市机票 (KAUST);
- 健康保险;
- 含私人卧室和浴室的住宅套房;
- 签证费 (学生必须持有有效护照);
- 社区娱乐资源;
- 社会及文化活动。

2. IVUL 组计算机视觉方向 VSRP 相关项目申请网址:

- <https://vsrp.kaust.edu.sa/pages/Large-Scale%20Human%20Activity%20Recognition%20in%20the%20Wild.aspx>
- <https://vsrp.kaust.edu.sa/pages/Teaching%20Cars%20to%20Drive%20and%20UAVs%20to%20Race.aspx>

3. IVUL 组主页

<https://ivul.kaust.edu.sa/Pages/Home.aspx>

三、阿里面向 2020 届毕业生招聘实习生

1. 职位名称: 图像算法工程师

2. 职位描述:

- CV 和深度学习相关研究;
- 人脸、人体分析, 图像或视频增强等。

3. 岗位要求:

- 计算机、通信、自动化或相关专业硕士以上学历, 图形图像处理相关专业背景;
- 熟悉常用的图像处理算法和深度学习算法;
- 动手能力强, 熟悉常用 DL 框架;
- 有顶会期刊论文发表者优先。

4. 联系方式:

简历请发送到: baichao.wbc@alibaba-inc.com, 或直接进入内推链接填写 http://alibaba.tupu360.com/campusActivity/getActivityInfo?activityCode=14pC0GibU7rFE97itRaAgoY6CoAZ6jgVcygP_UGLgKE%3D&enter=menu。

四、顺丰科技(北京)有限公司招聘计算机视觉算法工程师/实习生

1. 职位描述(以下方向但不限于此):

- 面向高精度地图的场景理解
- 面向安全驾驶的物体检测, 识别, 跟踪等
- 面向嵌入式开发的, 模型压缩, 优化, 移植
- 基于激光雷达的环境感知和定位
- 基于视频的行为识别

2. 能力要求:

- 硕士及以上学历, 计算机、电子、自动化、数学等相关专业;
- 熟悉深度学习理论和算法;
- 熟悉至少一种深度学习框架(TensorFlow/PyTorch/Caffe/MXNet/Keras 等)。

3. 有以下经验者优先:

- 熟悉深度学习模型优化和加速算法, 有实际项目经验;
- 熟悉深度学习嵌入式移植和优化, 有实际项

目经验;

- 熟悉语义分割/实例分割/场景分割等, 有实际项目经验 ;
- 发表过相关领域高质量论文或参加过视觉相关竞赛。

4. 联系方式

欢迎全职/实习的同学投简历。

邮箱: LinLi234@sf-express.com 或 Junjun.Xiong@sf-express.com。

五、美国加州圣塔克拉拉大学招收计算机工程博士研究生

圣塔克拉拉大学是美国的一所著名的私立天主教大学, 坐落于美国加州硅谷(Silicon Valley)心脏地带。福布斯 2018 年全美大学排名第 64 名。学校网页: <https://www.scu.edu>。

圣塔克拉拉大学计算机工程系刘莹教授招收博士研究生, 有奖学金机会。

1. 研究方向:

人工智能, 视频图像处理, 深度学习, 机器学习。

2. 招生要求:

- 计算机、电子信息、自动化等相关专业硕士毕业生或即将获得硕士文凭;
- 有扎实数学功底;
- Python, Matlab, C++编程能力强;
- 有视频图像处理, 视频编码和深度学习经验者优先考虑。

3. 联系方式:

导师个人网页: <https://www.cse.scu.edu/~yliu1>
电子邮箱: yliu15@scu.edu

(责任编辑: 贾同)

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 Pattern Recognition Letters, IEEE Transactions on Pattern Analysis and Machine Intelligence 特刊等。

3. 会议简介 (ICML)

ICML (International Conference on Machine Learning) 是计算机视觉领域最权威的学术会议之一，为全球模式识别领域的研究人员

及参与者提供一个交流平台，其内容涉及近期在计算机视觉、机器学习、模式识别方面的最新进展，会议论文集代表了上述领域最新的发展方向和研究水平。ICML 始于 1981 年，其后两年举办一届。

ICML-2020 将于 2020 年 7 月 12 日至 7 月 17 日在奥地利维也纳举办，论文截止时间为 2019 年 12 月 6 日。征稿范围涉及三维计算机视觉、动作识别与跟踪、大数据、医学图像分析、深度学习、图像处理、优化方法、机器人视觉、识别、检测、分割等方向的研究进展。

(责任编辑：沈沛意)

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
CAIP 2019	2019.09.02-05	意大利 Salerno	2019.04.01	http://caip2019.unisa.it
SIGGRAPH Asia 2019	2019.09.17 - 09.20	澳大利亚 Brisbane	2019.05.20	https://sa2019.siggraph.org/
ACMMM 2019	2019.10.21-25	法国 Nice	2019.04.01	https://www.acmmm.org/2019
BMVC 2019	2019.09.09-12	英国 Cardiff	2019.04.29	https://bmvc2019.org
PRCV 2019	2019.11.08-11	中国 西安	2019.04.15	http://www.prcv2019.com/
ICML 2020	2020.07.12-17	奥地利 Vienna	2019.12.06	http://conference.researchbib.com/view/event/92810

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
Pattern Recognition Letters	Pattern Recognition and Artificial Intelligence Techniques for Cultural Heritage special issue	2019.06.30
IEEE Transactions on Pattern Analysis and Machine Intelligence	Vision Computing in Environmental Perception of Intelligent Vehicles	2020.01.01