

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

02 2021

总第 28 期



CCCF 计算机视觉  
专委会

# COMPUTER VISION NEWSLETTER



## 计算机视觉专委会 简报

2021 年第 02 期

总第 28 期

### 主 办 编委会

CCF 计算机视觉专业委员会



CCF 计算机视觉  
专 委 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

#### /科技前沿/

**任传贤** 中山大学  
**杨巨峰** 南开大学  
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  
**魏秀参** 南京理工大学

#### /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

#### /视界专访/

主 编 **张军平** 复旦大学  
编 委 **贾熹滨** 北京工业大学  
**明 悦** 北京邮电大学

# CONTENTS

## 简报目录

### | 专委动态

- 04 CCF-CV 走进高校系列报告会
- 07 CCF-CV 视界无限系列研讨会
- 10 CCF-CV 第一届计算机视觉前沿讲习班成功举办
- 13 CCF-CV 常务委员会 2021 年度第一次工作会议顺利召开
- 14 2021 年度“CCF-CV 奖”评选活动正式启动
- 15 CCF-CV 专委会 2021 年执行委员增选申请开始

### | 科技前沿

- 16 基于对比关系网络的图像补全
- 19 虚拟服装试穿前沿
- 22 非接触可撤销掌纹识别研究进展
- 25 ICLR 2021

### | 委员风采

- 28 南开大学程明明教授访谈
- 30 委员好消息

### | 学术资源

- 31 无参考图像质量评价开源代码
- 34 人脸活体检测数据集
- 40 好文推荐

### | 海外学者

- 41 新加坡国立大学冯佳时教授团队
- 48 征文通知

### | 视界专访

- 49 复旦大学吴立德教授专访

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 97 期 厦门大学



2021年3月31日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，厦门大学媒体分析与计算实验室承办的第97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”在厦门大学海韵校区信息学院 A201 报告厅举行。本期报告会邀请了南京大学吴建鑫教授、南京信息工程大学刘青山教授、北京大学彭宇新教授、上海交通大学熊红凯教授、中国科学院信息工程研究所操晓春研究员五位专家学者做特邀报告。厦门大学信息学院媒体分析与计算实验室纪荣嵘教授和孙晓帅副教授担任本次报告会的执行主席。

此次 CCF-CV 走进高校系列报告会聚焦领域前沿，报告内容精彩，会现场气氛热烈。五位专家和与会听众针对计算机视觉领域的前沿问题进行了深入地交流探讨，详尽的解答了听众提出的问题。最后，参加报告会的现场听众对五位特邀讲者表示由衷的欢迎与感谢，报告会在掌声中圆满结束。

### 第 98 期 中山大学



2021年4月9日，由中国计算机学会计算机视觉专委会主办、中山大学计算机学院承办的第98期 CCF-CV 走进高校系列报告会在中山大学东校区成功召开。会议邀请了南京信息工程大学刘青山教授、中科院计算所山世光研究员、爱奇艺资深科学家王涛博士、华中科技大学白翔教授、微软亚洲研究院王井东研究员、中科院自动化所王亮研究员、北京大学林宙辰教授、东南大学耿新教授以及大连理工大学卢湖川教授等九位专家做特邀报告。中山大学赖剑煌教授、郑伟诗教授、任传贤副教授和谢晓华副教授担任会议执行主席。师生们聆听了专家们的报告，反响强烈。

此次 CCF-CV 走进高校报告会安排了两个半天时间。每位专家作完报告后，在场的师生们就科研的知识储备、切入点、方法细节、投稿回复等相关方面纷纷提出自己感兴趣的问题，专家们也给出了详尽而又可靠的解答。报告会在热烈的掌声中圆满结束。

## CCF-CV 走进高校系列报告会

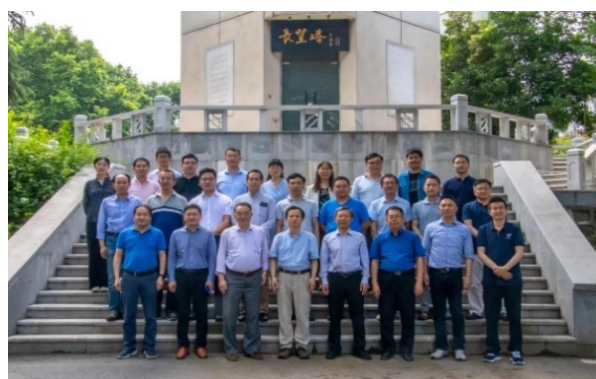
### 第 99 期 山东建筑大学



2021 年 4 月 24 日，由中国计算机学会计算机视觉专委会 (CCF-CV) 主办、山东建筑大学承办、山东省人工智能协会协办的第 99 期走进高校系列报告会在山东建筑大学成功举办。本次活动邀请了中国科学技术大学刘利刚教授、北京航空航天大学王莉莉教授、清华大学刘焯斌副教授、中国科学院自动化研究所申抒含研究员、以及北京大学连宙辉副教授五位专家学者做特邀报告。北京大学信息科学技术学院陈宝权教授和山东建筑大学计算机科学与技术学院刘萌教授担任本次会议的执行主席。本次活动中，专家们围绕“虚拟现实与三维智能”这一前沿主题做了精彩报告，报告会反响强烈，引发了广泛的共鸣。报告会由山东建筑大学计算机科学与技术学院李晓峰院长、刘萌教授和裘肖明副教授主持。

最后，北京大学陈宝权教授对报告会进行总结，再次向各位特邀专家和现场参会人员表示感谢，并欢迎计算机视觉领域的同行专家学者来山东建筑大学指导和交流工作，报告会在全场热烈的掌声中圆满结束！

### 第 100 期 南京信息工程大学



2021 年 5 月 28 日，由中国计算机学会计算机视觉专委会 (CCF-CV) 主办、南京信息工程大学计算机与软件学院承办的第 100 期 CCF-CV 走进高校系列报告会在南京信息工程大学长望楼报告厅成功召开。本次会议邀请了清华大学胡事民教授、北京大学查红彬教授、中山大学赖剑煌教授三位专家做特邀报告，南京信息工程大学计算机与软件学院刘青山教授担任本次会议的执行主席。报告会由 CCF-CV 专委会副主任南京信息工程大学刘青山教授、中国科学院自动化研究所王亮研究员、上海科技大学虞晶怡教授主持。在本次报告会上，专家们围绕“计算机视觉前沿技术及应用”做了精彩报告。

此次 CCF-CV 走进高校报告会聚焦计算机视觉领域前沿，专家们就时下热点技术展开了详细介绍，与会师生就前沿问题展开了深入的交流。通过本次报告会，师生们从多维度学习到专家们开展科研的思路和技巧，受益匪浅。至此，本次 CCF-CV 走进高校报告会在热烈的掌声中圆满落下帷幕。

## CCF-CV 走进高校系列报告会

## 第 101 期 广东工业大学



2021年6月4日(星期五)下午,由中国计算机学会计算机视觉专委会主办,广东工业大学计算机学院承办的 CCF-CV 走进高校系列报告会第 101 期活动以云直播形式成功举办。本次报告会邀请到 4 位专家做特邀报告,他们是:中山大学赖剑煌教授、厦门大学纪荣嵘教授、哈尔滨工业大学(深圳)徐勇教授和上海交通大学严骏驰副教授。广东工业大学计算机学院曾安教授和费伦科博士担任本次报告会执行主席和主持人。广东工业大学计算机学院院长程良伦教授致欢迎辞。他首先代表学院对四位演讲嘉宾表示热烈的欢迎和衷心的感谢。随后,他介绍了广东工业大学计算机学院概况,并希望借助这次报告会能和各位专家与同行加强学术交流与合作。最后,程院长预祝本次报告会取得圆满成功。

报告结束后,主持人费伦科老师再次对四位报告专家以及 CCF-CV 专委会表示衷心的感谢!受广州近期疫情影响,本次报告会临时由线下改为线上如期举办,讲者的报告质量不打折扣,B 站人气峰值也达 5000+!也为下次再次邀请各位专家到广东工业大学进行线下交流和指导工作做好充分准备!

## 第 102 期 广西师范大学



2021年6月11日上午,由中国计算机学会计算机视觉专委会主办、广西师范大学计算机科学与工程学院承办的第 102 期 CCF-CV 走进高校系列报告会在广西师范大学育才校区成功召开。会议邀请了中国科学院自动化所徐常胜研究员、中国科学院自动化研究所胡卫明研究员、北京交通大学赵耀教授、厦门大学孙晓帅副教授四位专家做特邀报告。广西师范大学苏桂发副校长出席活动并致辞。广西师范大学钟必能教授、唐振军教授、李先贤教授担任会议执行主席。师生们聆听了专家们的报告,反响强烈。

此次 CCF-CV 走进高校系列报告会持续近四个小时,专家们的讲解深入浅出,报告内容精彩。现场专家们与老师、同学们之间进行了酣畅淋漓的学术交流,让所有参加会人员享受了一场学术盛宴。最后,以本次报告为契机,热烈欢迎全国计算机视觉领域的同行们来广西师范大学交流指导,拉近在校学生和专家学者之间的距离,为计算机视觉的发展持续地注入新鲜活力!

责任编辑 毋立芳

第 9 期 手绘草图研究的前沿进展与未来趋势

## CCF-CV 视界无限系列研讨会



2021年4月28日，由中国计算机学会计算机视觉专委会主办的第9期CCF-CV“视界无限”系列活动——“手绘草图研究的前沿进展与未来趋势”研讨会在北京邮电大学成功举办。研讨会邀请了英国萨里大学宋一哲教授、香港城市大学傅红波教授、西安电子科技大学邓成教授、中山大学郑伟诗教授、西安电子科技大学王楠楠教授、北京航空航天大学于茜副教授、香港中文大学韩晓光助理教授和英国萨里大学逢开岳高级研究员做主题报告并参与圆桌讨论。北京邮电大学人工智能学院党委书记杨洁教授、中国计算机学会计算机视觉专委会主任、北京大学查红彬教授，中国计算机学会计算机视觉专委会秘书长、北京邮电大学马占宇教授，中国计算机学会计算机视觉专委会副秘书长、南开大学杨巨峰教授出席活动。本期研讨会由北京邮电大学人工智能学院承办，英国萨里大学宋一哲教授任主席、北京邮电大学齐勇刚助理教授任执行主席。研讨会由北京邮电大学人工智能学院齐勇刚老师主持。

北京邮电大学人工智能学院党委书记杨洁教授首先致辞，对各位专家学者和来宾表示热烈欢迎。她表示北京邮电大学是国内最早从事人工智能人才培养和科

学研究的单位之一，本次研讨会是一次难得的学术交流机会，希望以此为契机广交学术人才，增进友谊，激励创新，为拓展未来人工智能领域的国际合作与交流寻求新思路，与同行专家学者、海内外朋友一道，为国家科技事业发展贡献力量！



中国计算机学会计算机视觉专委会主任、北京大学查红彬教授致辞。查教授表示很高兴能够与线上线下的各位一起探讨计算机视觉的前沿问题，他指出CCF-CV专委会的视界无限活动已经办了很多期，是围绕一个主题进行深入讨论的研讨会，“草图”这个主题第一次举办，很有意义，手绘草图是人与计算机交互的重要方式，可以提高交互效率。此外，草图是多学科交叉的领域，涉

及包括视觉、图形学、艺术等。希望本次研讨会能对从事这一领域研究的老师和同学们有所启发。最后代表专委会感谢主办单位各位老师同学为筹办本次研讨会作出的努力，预祝研讨会圆满成功！



英国萨里大学 SketchX 实验室负责人宋一哲教授的引导发言题目是“Vision != Photo”。宋教授希望报告能够抛砖引玉，大家一起交流碰撞出一些好想法。报告指出视觉应该包含“看”和“理解”，但图片只有原始的像素信息，并没有任何语义信息，而草图是一种蕴含了人为处理的图像，具备了图片所没有的信息，这是草图研究最大的意义所在——帮助探索人类的视觉理解机制。宋教授表示，近些年来草图的研究因为上述原因更多的受到关注，他在报告中简要介绍了课题组近年来的 Sketch 相关工作，已及近期 Sketch 和传统视觉领域相融合热点问题，包括细粒度的视觉分析、三维视觉以及 OCR 等。

香港城市大学傅红波教授的引导发言题目是“Data-Driven Sketch Interpretation”。傅教授的发言主要围绕如何用数据驱动的手段来帮助解决草图理解的各种任务展开。傅教授首先介绍了目前处理草图的主要技术手段，包括 CNN、RNN 以及 attention 等，然后从草图的不同任务出发介绍了现有技术如何解决美化、分类、分割和 3D 建模等问题。最后介绍了现有技术手段中存在的问题，比如由于抽象程度、绘图样式和绘图错误等原因，导致机器无法正确理解草图的语义。

西安电子科技大学邓成教授的报告题目是“手绘草图视觉生成及检索研究进展”。邓教授首先总结梳理了在草图图像的生成和检索领域中，现有经典方法及其存在的不足。邓教授指出目前草图图像生成和检索领域面

临的主要挑战包括：草图与真实图片的失配以及成对数据集的稀缺；手绘图像中颜色和线条信息的缺失；手绘图像抽象程度的差异性。最后，邓成教授认为未来的发展方向主要分为三个层面：1. 在数据层面，解决数据不平衡和缺失的问题。2. 在特征层面，进一步提升草图分辨率。3. 在应用层面，提升草图检索的交互实时性。

中山大学郑伟诗教授的报告题目是“面向换装的行人重识别问题研究”。郑伟诗教授指出以往的行人重识别问题多数依赖于行人的纹理特征，但在面临行人换装的问题时，基于纹理特征的重识别方法将失去可靠性。针对上述问题，郑教授介绍了自己团队设计的基于人物素描轮廓特征的建模方案，有效提升了行人换装情况下的重识别准确率。最后，郑教授介绍了其团队对于获取人体的精准轮廓特征的最新进展，结合对人体行走姿态的相关研究，提出了人体轮廓 3D 建模的方案。



西安电子科技大学王楠楠教授的报告题目是“异质人脸图像合成与识别”。首先，王楠楠教授指出异质人脸图像合成主要是指可见光图像与这些不同模态和形态图像之间的相互转换，多用于公共安全领域进行高清晰的人脸图像快速重建。随后，王教授介绍了异质人脸图像合成的典型方法，并以人脸素描画像和照片的合成为例，讲解了异质人脸合成中的关键原理和其团队的工作进展。包括基于人脸素描的剪纸合成，公共安全领域的人像应用，人像数据集的扩充，人像合成的多任务框架。最后，王教授提出了基于现有工作的未来扩展方向，包括风格化的迁移，长依赖关系的嵌入等。

北京航空航天大学于茜副教授的报告题目是“基于手绘的自然图像和三维物体的生成”。于老师认为基于



手绘的生成任务是一个需要研究者们引起关注的领域，并指出基于手绘的自然图像生成可以看作“image to image translation”的任务来处理。随后，于老师指出目前的工作大多数关注于基于从自然图像中提取的具有规则线条的边缘图，而忽略了不具有对应的自然图像并且有不同程度形变的手绘图。于老师以其团队近期的研究为例，详细解析了“从手绘图到自然图像的翻译”这项工作的技术细节，包括分阶段引入灰度图作为中间态、自监督去噪损失等技术手段。

香港中文大学（深圳）韩晓光助理教授的报告题目是“基于深度学习的智能草图三维建模”。首先，韩老师指出简笔草图往往是用户表达三维形状最简便的方式，但构建三维几何的专业性和复杂性与之存在巨大的矛盾。随后，韩老师介绍了其团队针对三种不同类型的3D模型开展的研究工作，包括三维人脸、人造物体、动物角色。韩老师特别强调了分割生成、组合方式、细节渲染和实时交互问题的重要性。最后，韩老师认为 sketch 生成 3D 模型并不是单纯的二维图片生成三维模型问题，保证输入 sketch 和输出 model 的一致性以及生成过程的实时交互性都尤为重要。



英国萨里大学 SketchX 实验室逢开岳研究员的报告题目是“基于手绘的细粒度图像检索”。首先，逢老师指出在图像检索中，手绘是最直接的视觉输入，有效缓解了形容细粒度视觉特征时进行语义诠释的笨重性。逢老师认为由于绘画技能的差异性，收集可信赖的绘画数据集是目前面临的巨大挑战。针对这一问题，逢老师以其最新的研究为例，介绍了一系列利用跨域数据的预训练优化方法。最后，逢老师强调了 sketch 对于图像检索领域的重要意义：1. 以人类手绘的图像为基准更有助于划分人类能区分的粒度极限。2. 对于包含多种元素的图像，sketch 可以有效传达人类对图像中元素的偏好，有助于解决视觉显著性的相关问题。

紧接着是 panel 环节，由北京邮电大学齐勇刚老师主持，与八位讲者探讨了手绘草图研究的前沿进展与未来趋势。整个研讨会在下午 5 点 40 分圆满结束。



责任编辑 杨巨峰

## 第一届 CCF 计算机视觉前沿讲习班成功举办



2021年5月15-16日，第一届 CCF 计算机视觉前沿讲习班在英雄的城市武汉成功举办，活动共吸引了计算机视觉领域的研究人员、学生、工程师等 200 余人报名参加。与会讲者和学员纷纷表示，在抗击新冠肺炎疫情的严峻斗争中，武汉人民用自己的实际行动，彰显了中华民族同舟共济、守望相助的家国情怀。这次来到武汉参加讲习班，既要学习前沿知识，更要学习武汉人民识大体、顾大局，不畏艰险、顽强不屈的英雄精神，亲身感受武汉人民坚韧不拔、高风亮节的英雄气息。

本次活动由中国计算机学会 (CCF) 主办，中国计算机学会计算机视觉专委会 (CCF-CV) 与武汉理工大学联合承办，武汉理工大学计算机科学与技术学院院长熊盛武教授和华中科技大学自动化与人工智能学院白翔教授担任执行主席。讲习班旨在促进计算机视觉领域的学术交流与高级人才培养，帮助广大师生和青年从业者提升技术水平，开拓实践视野，全面学习并系统掌握计算机视觉前沿技术和应用工具。

### 开幕仪式

讲习班开幕式首先由熊盛武院长致辞，熊院长介绍了本次活动的目的、武汉理工大学计算机科学与技术学院的基本情况、本次活动组织和报名情况，对讲者和学员表示热烈欢迎。随后，CCF 计算机视觉专委会主任、北京大学查红彬教授介绍了专委会的宗旨和讲习班的意义，发表感谢致辞，并为本次活动的承办方武汉理工大学计算机科学与技术学院、以及组织活动的主席和组织委员颁发了感谢证书。



### 专家报告

北京大学查红彬教授为学员讲授第一课。查教授以“视觉 SLAM：在线学习的途径”为题，首先回顾了 SLAM 的基本概念和发展历史，然后分享了在线学习 SLAM 的相关研究工作，包括引入数据流计算的基本概念，充分挖掘密集采样传感数据内在的时空连续性，来加强 SLAM 算法的预测能力；构建基于时域变化的增量算法，并利用地图全局特征的约束以及传感数据的实时反馈作用，实现传感器轨迹的高效计算与三维地图的递进式构建；探讨建立自监督 SLAM 在线学习技术的新途径，包括：面向自监督视觉里程计的序列对抗学习方法、



具有在线自适应能力的自监督 SLAM 学习等。

中科院计算所陈熙霖研究员由于武汉雷暴天气航班取消无法到达现场，所以课程在线上进行。陈老师非常想亲临现场给学员授课，在北京首都机场两次登机两次下机，直到凌晨两点多被通知飞机取消才返回。陈熙霖研究员讲授的线上课程以“场景的层次表达与推理”为题，首先介绍了视觉智能的发展趋势以及物体识别背后的局限性。然后分享了以“万物关联、属性为桥”为基本思想的场景理解方法，利用视觉、语言以及知识综合判定检测结果；介绍了以属性为纽带的物体识别方法，不仅满足了最基本的检索、分类需求，还扩展至更丰富多样的视觉描述。最后详细讲述了以语义图为桥梁的场景理解，借鉴人类感知机理，充分利用周边场景与物体间关系信息、建立图结构，设计消息传递机制进行全局关联推理。



中科院自动化所刘成林研究员以“图神经网络与文档图像分析”为题，首先介绍了文档分析领域的相关流程及难点问题。然后讲述了结构化预测主要的方法，包括隐马尔可夫模型、循环神经网络、马尔可夫随机场等。其中，详细讲解了图神经网络的理论知识。最后，与学员分享了基于图神经网络的文档图像版面分析研究进

展，包括基于图注意网络的联机手写文档版面分析、流程图识别，复杂文档图像区域分割和二维结构分析等。



西北工业大学韩军伟教授以“遥感影像‘深度’解译”为题，介绍深度学习在高分遥感影像解译方面的一些典型应用。首先，针对高分遥感图像大规模数据库相对稀少的问题，介绍了课题组发布的两个公开的高分图像数据库：NWPU-RESISC45 场景分类数据库和 DIOR 目标检测数据库。然后，针对高分遥感影像解译存在的一些难点问题和不同的应用需求，介绍了课题组提出的一些弱监督和旋转不变目标检测算法、型号识别算法以及结合深度学习和度量学习的场景分类算法，最后总结了对未来研究方向的展望。



中科院信工所操晓春研究员以“人工智能驱动的网络空间内容安全”为题，从一整套视觉大数据价值挖掘系统出发，首先介绍其团队在网络空间安全应用方向学术探索的最新进展，然后介绍了在系统中其他方面开展的相关研究包括：开源数据的云弹性采集与强智能搜索、数据驱动的自底向上深度抽象归纳、知识驱动的自顶向下演绎推理、人工智能深度攻防对抗等。

百度飞桨高级技术经理刘其文讲者以“深度学习产

业落地的四大挑战”为题，结合飞桨的实际案例，从四个方面介绍深度学习产业落地的四大挑战与解决方法，包括 1. 业务挑战：什么样的产业场景适合人工智能落地？2. 人才挑战：人工智能落地，需要什么样的复合型人才？3. 技术挑战：从深度学习框架的角度，还有哪些核心技术有待突破？4. 生态挑战：需要政府、科研、教育和产业怎样相互协同，才能构建良好的人工智能产业生态？



阿里巴巴资深算法专家、达摩院 OCR&智慧教育方向负责人王永攀讲者以“OCR&智慧教育”为题，首先介绍了阿里云读光产品在 OCR 研究方面的应用和功能，演示了其中用于古籍 OCR 的产品——汉典重光，然后介绍了 OCR 算法演化和场景延伸，最后引出智慧教育话题，详细介绍了利用 AI 算法助力教学的难点和思路。

南方科技大学于仕琪副教授以“步态识别技术和最新进展”为题，首先介绍了步态识别的研究背景，然后结合自己在该方向研究的心路历程，以幽默诙谐的方式向大家讲述了其研究成果和最近进展，以及在研究过程中遇到的问题和解决问题的思路。



## 活动总结

第一届 CCF 计算机视觉前沿讲习班为计算机视觉领域的广大师生和工程师提供了一个与专家学者近距离交流学习的宝贵机会，大家对专家所讲的课程内容产生了极大兴趣，现场互动非常活跃，学术氛围浓厚。



在结业典礼上，执行主席白翔教授对本次活动进行了总结。首先祝贺本次活动在武汉理工大学举办取得了圆满成功，然后征集学员意见，请大家为下一届讲习班的组织工作提供建议。最后，为学员颁发学员证书。全体学员合影留念，记录下了自己在第一届 CCF 计算机视觉讲习班顺利结业的瞬间。



责任编辑 杨巨峰

# CCF-CV 常务委员会 2021 年度第一次工作会议 顺利召开

## 中国计算机学会计算机视觉专委会 常务委员会工作会议

广州 2021年4月10日

中国计算机学会计算机视觉专委会 (CCF-CV) 常务委员会 2021 年度第一次工作会议 4 月 10 日在广州中山大学顺利召开!

会议由专委会主任查红彬教授主持, 中国计算机学会秘书长唐卫清研究员和中国计算机学会常务理事臧根林教授应邀参加, 秘书处全体成员列席。

本次常委会工作会议系统地梳理了专委会各项工作的进展和不足, 明确了专委会发展的工作重点, 特别是对专委会开创的各项品牌学术活动的继续创新突破指明了方向。本次会议的召开对于专委会的持续蓬勃发展具有承前启后的重大意义。



会议首先请学会秘书长唐卫清研究员致辞。唐秘书长介绍了换届后学会运营发展的新理念以及促进学会专业化、国际化、行业化等方面的新举措, 肯定了 CCF-CV 专委会开展的各项特色活动和取得的工作进展, 鼓励专委会继续引领活动创新、展现专委会活力和影响力。

随后, 查红彬主任做主题发言。查主任充分肯定了专委会过去一年取得的工作成果, 尤其肯定了疫情期间采用线上方式持续推进专委会工作的努力, 并对后续工作提出了切实可行的发展建议。

接下来, 常委会委员听取了专委会近期活动规划, 并针对专委会发展相关议题展开了热烈讨论, 逐一形成了具体可行的指导性建议, 为专委会后续发展明确了方向和重点。专委会发展相关议题来自专委会委员们的建言献策, 秘书处进行分类汇总, 涉及专委会日常工作、学术活动、组织建设、领域发展等方面。

最后, 查红彬主任作了总结发言。会议在紧张热烈而有序的讨论氛围中结束。



责任编辑 黄岩

## 2021 年度 "CCF-CV 奖" 评选活动正式启动

**为**发挥中国计算机学会计算机视觉专委会（简称 CCF-CV）的学术评价作用，激励、调动计算机视觉及相关领域专业工作者的积极性，促进计算机视觉及相关领域的创新和进步，CCF-CV 自 2017 年起设立中国计算机学会计算机视觉专委会奖（简称 CCF-CV 奖），旨在奖励计算机视觉及相关领域中为领域发展做出突出贡献的老前辈、取得公认成就的科研人员、意义重大的科技成果、影响深远的学术成绩、具备潜力的学术新锐，以及为专委会工作提供卓越服务的优秀个人。

### 评选要求

#### 一、候选人提名

CCF-CV 奖项实行委员提名制，一名以上 CCF-CV 专委会委员即可（联合）提名推荐。每个奖项的评选标准等事宜如下：

##### 1. CCF-CV 终身学术贡献奖

授予在计算机视觉及相关领域做出了突出贡献的中国计算机科技工作者（年龄在 70 岁以上）。每年评选一次，获奖人数不超过二名。获奖者可获得奖励证书、奖杯及奖金。

##### 2. CCF-CV 杰出成就奖

最高标准评选出计算机视觉领域取得了公认的杰出成就的计算机科技工作者，其获奖主要工作应在中国境内完成，国籍不限。每年评选一次，获奖人数不超过一名。获奖者可获得奖励证书、奖杯及奖金。

##### 3. CCF-CV “中科视拓 Seeta 服务贡献奖”

授予专委会活动、会议组织、秘书处工作、工委小组工作有重要贡献者。每年评选一次，获奖人数不超过六名。获奖者可获得奖励证书、奖杯和奖金。

##### 4. CCF-CV “中科视拓 Seeta 学术新锐奖”

授予在计算机视觉及相关领域从事相关研究的研究生（含硕士研究生与博士研究生），尤其是具备一定科研潜力的低年级博士生。所有申请参评人员需要提供支持同一主题的不超过 3 篇代表作。本奖项每年评选一次，获奖人数不超过三名。获奖者可获得奖励证书、奖杯及奖金。

##### 5. CCF-CV 持久影响力工作奖

鼓励原创性、基础性并能够产生持久影响力的创新性研究，其获奖工作的主要支撑应为 2011 年 1 月 1 日至 2011 年 12 月 31 日期间在计算机视觉主流会议和期刊上正式发表的论文，且第一单位为中华人民共和国境内（含港澳台地区）的学术机构。本奖项每年评选一次，获奖工作不超过一项。获奖者可获得奖励证书、奖杯及奖金。

#### 二、推荐方法

由推荐人联合填写推荐表 1 份，将其电子版扫描件于截止日期之前提交至电子邮箱 [ccfcvaward@sina.com](mailto:ccfcvaward@sina.com)。

### 评奖流程

#### 一、评奖办法

在 CCF-CV 常委会领导下，由 CCF-CV 提名与奖励工作组负责评奖工作，具体包括：工作组负责征集提名、

初筛与资格审查（“持久影响力工作奖”还将提名评审委员会参与评审）。所有奖项候选名单提交常委会表决，过半数通过。

## 二、报名截止时间

即日起—2021 年 7 月 15 日。

## 三、其他未尽事宜

详见《中国计算机学会计算机视觉专委会奖励条例》。

责任编辑 任传贤

# CCF-CV 专委会 2021 年执行委员增选申请开始

自 2013 年 10 月成立以来，中国计算机学会 (CCF) 计算机视觉专业委员会 (ccfcv.ccf.org.cn) 发展迅速，举办了很多有影响力的活动，如计算机视觉前沿进展研讨会 (RACV)、CCF-CV 走进高校系列报告会、CCF-CV 走进企业系列交流会、CCF-CV 视界无限系列研讨会，与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、中国人工智能学会模式识别专委会共同举办中国模式识别与计算机视觉大会 (PRCV)，定期出版专委简报，建设专委中英文网站，专委微信公众号文章平均阅读上千次，专委活动视频在专委 Bilibili 账号发布。搭建了全方位、高水平、大规模的计算机视觉领域交流平台。专委会成立八年以来，已经发展执行委员 354 人，在 CCF 专委评估中获得“特色活动奖”、“综合进步奖”、“优秀专委奖”、“年度特别奖”等 6 个奖项。为了保持专委会的活力、促进国内外视觉领域人员的交流和合作，专委会现开放 2021 年计算机视觉专委会的执行委员增选工作。

**申请时间：**2021.6.15—2021.10.15

**申请流程：**填写申请表，发送给秘书处 (ccfcv@139.com)，主题“2021 新执行委员申请-姓名-单位”。（注：推荐人必须是现任专委执行委员，名单可以从专委网站查询。电子版申请表中需填写推荐人姓名和意见，执行委员增选成功后可以补签名）。

**申请资格：**任职国内外学术界或企业界副教授或等同级别以上的人员，拥有计算机视觉相关领域的高水平研究成果，是 CCF 计算机视觉专委委员，且积极参加计算机学会计算机视觉专委会的各项活动。特别优秀的讲师、企业人士亦可考虑。

**申请需知：**现任专委执行委员每人可推荐最多 3 名候选人。本次申请结果将在“2021 年中国模式识别与计算机视觉大会”期间（2021 年 10 月 29 日-11 月 1 日）举行的专委工作年会上投票确定（申请者届时必须“注册参会”）。

**特别说明：**按照 CCF 的新规定，CCF 专业会员通过 CCF 会员系统关注相关专委后即加入专委并成为其委员，其后每年可以更改一次关注的专委。委员在专委中无选举权和被选举权，但具有对专委的评价权。原来的专委委员自动升级为专委执行委员，享有选举权、被选举权以及对专委的评价权。

责任编辑 毋立芳

热点追踪

# 基于对比关系网络的图像补全

中科院自动化研究所 周晓强 赫然

图像补全任务的难点之一在于对补全结果结构合理性和纹理协调性的要求。在本文中，我们提出了一个两阶段的补全网络来解决此问题。我们首先利用分割重建网络预测残缺图像的完整语义分割图，之后在第二阶段通过图像生成器恢复细粒度的图像细节。在第二阶段的图像生成器中，我们新提出了一种基于图的关系网络，对残缺图像中存在的关系进行建模。此外，为了促使关系网络中的图网络部分学习到更好的节点特征表示，我们提出了对比损失函数，对图网络中的节点特征表示进行优化。相关成果获 ICPR 2020 最佳科学论文奖。

图像补全是指对图像中缺失的区域进行内容补全，要求补全后的图像看起来自然逼真。图像补全技术在图像恢复、图像编辑等领域具有广泛应用。传统图像补全算法<sup>[1,2]</sup>使用残缺图片背景区域中的像素或图像块对缺失区域进行补全。当图片残缺区域较大，或待补全内容具有较复杂的语义结构时，这类传统算法的补全结果往往有明显瑕疵。

近年来蓬勃发展的生成对抗网络给图像补全带来了新的解决思路。研究人员利用卷积神经网络和生成对抗网络强大的表征学习和图像重建能力，提出了若干基



图 1 图像补全示意图

于生成对抗网络的图像补全算法<sup>[3,4,5]</sup>。然而，如何同时生成结构合理且纹理逼真自然的补全结果，仍是图像补全任务面临的一大挑战。为了解决上述问题，我们提出了基于对比关系网络的图像补全算法。整体框架包含两个阶段：分割重建网络和图像生成网络。补全过程示意图如图 1 所示，整体网络框架图如图 2 所示。

我们首先利用分割重建网络预测残缺图像的完整语义分割图，之后在第二阶段通过图像生成器恢复细粒

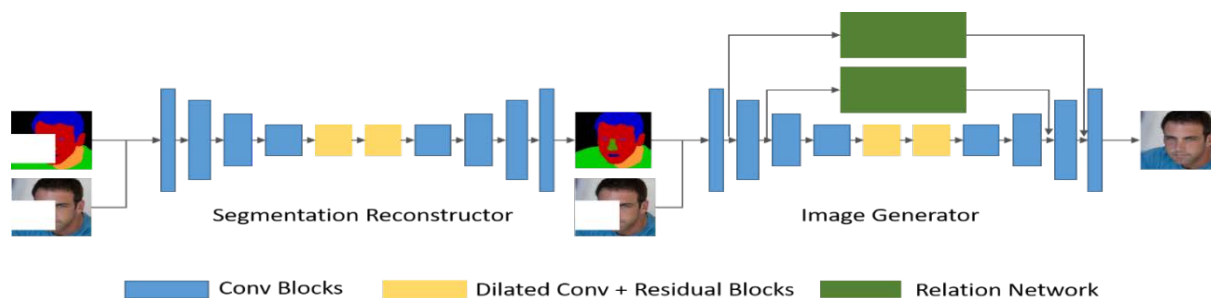


图 2 整体网络框架

度的图像细节。第一阶段中补全的完整语义分割图会对第二阶段的图像细节生成提供指导。

在第二阶段的图像生成器中，我们新提出了一种基于图的关系网络，来对残缺图像中存在的关系进行建模。关系网络的具体结构如图 3 所示。

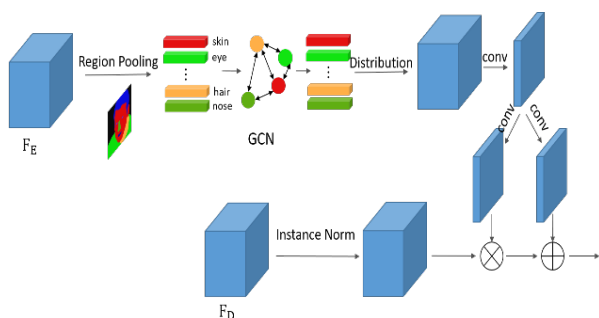


图 3 关系网络

在关系网络中，我们既考虑了同一语义区域内像素之间的内部关系，也考虑了不同语义部分之间的相互关系，以此来提高图像纹理的一致性和协调性。此外，为了促使关系网络中的图网络部分学习到更好的节点特

征表示，我们提出了对比损失函数。对比损失函数通过扩大负样本对之间的特征差异，同时提高正样本对的特征相似度，来对图网络中的节点特征表示进行优化。

为了测试算法的有效性，我们在 CelebA-HQ 和 DeepFashion 数据集上分别进行图像补全实验。定量结果和定性结果对比如表 1 和图 4 所示。与最新相关算法相比，提出的方法不仅在定量实验中取得了更好的性能，同时在定性分析实验中实现了对残缺图像更逼真、自然的补全结果。

表 1 不同算法定量指标对比

		PSNR	SSIM	FID
CelebA-HQ	GMCNN <sup>[3]</sup>	27.35	0.8839	8.81
	EdgeConnect <sup>[4]</sup>	26.60	0.8724	8.38
	GatedConv <sup>[5]</sup>	27.12	0.8814	7.74
	StructureFlow <sup>[6]</sup>	27.48	0.8885	7.21
	Ours	<b>28.61</b>	<b>0.9013</b>	<b>5.94</b>
DeepFashion	GMCNN <sup>[3]</sup>	22.40	0.8074	9.41
	EdgeConnect <sup>[4]</sup>	22.79	0.8162	10.56
	GatedConv <sup>[5]</sup>	23.32	0.8175	8.24
	StructureFlow <sup>[6]</sup>	23.20	0.8166	8.00
	Ours	<b>24.16</b>	<b>0.8197</b>	<b>7.37</b>



图 4 不同算法补全结果对比

责任编辑 崔海楠

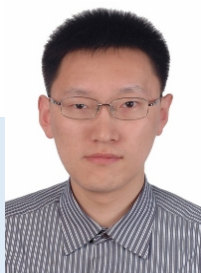
## 参考文献

- [1] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, “Image inpainting,” in Proceedings of the Conference on Computer Graphics and Interactive Techniques, pp. 417–424, 2000.
- [2] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, “Patchmatch: A randomized correspondence algorithm for structural image editing,” ACM Transactions on Graphics, vol. 28, no. 3, pp. 24–33, 2009.
- [3] Y. Wang, X. Tao, X. Qi, X. Shen, and J. Jia, “Image inpainting via generative multi-column convolutional neural networks,” in Advances in Neural Information Processing Systems, pp. 331–340, 2018.
- [4] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, “Edgeconnect: Generative image inpainting with adversarial edge learning,” arXiv preprint arXiv:1901.00212, 2019.
- [5] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, “Free-form image inpainting with gated convolution,” in Proceedings of the IEEE International Conference on Computer Vision, pp. 4471–4480, 2019.
- [6] Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and G. Li, “Structureflow: Image inpainting via structure-aware appearance flow,” in Proceedings of the IEEE International Conference on Computer Vision, pp. 181–190, 2019.



周晓强

中科院自动化研究所博士生，导师为谭铁牛院士。主要研究方向为计算机视觉、图像生成。  
Email: xiaoqiang.zhou@cripac.ia.ac.cn



赫然

中科院自动化研究所研究员，博导。主要研究方向为计算机视觉和模式识别。  
Email: ran.he@ia.ac.cn

热点追踪

## 虚拟服装试穿前沿

台湾阳明交通大学电子研究所 陈婕云 帅宏翰 郑文皇

时尚与我们的生活密不可分，例如：妆容以及穿搭，都充分展现个人特质。随着深度学习与计算机视觉的发展，有许多前瞻性的技术能够帮助我们找到适合自己的时尚风格，例如：虚拟上妆、虚拟服装试穿、服装搭配推荐等。本文聚焦于介绍虚拟服装试穿的技术发展，效果展示如图1。试想我们购买新衣服不外乎两种方式：(i) 上街亲自去试穿购买，以及 (ii) 上网看模特的试穿效果购买。上街买衣服可以很直接地判断新衣服适不适合自己，但过程（逛街->挑选到喜欢的衣服->进一步试穿确认是自己喜欢的版型）却非常耗费时间。而近年逐渐流行网购，却存在许多问题，例如：商品的实际色泽与网络商品图不同；衣服穿在模特身上的感受度，和在消费者身上感受度不同；网络上的商品图无法呈现不同姿势角度的试穿效果；衣服的尺寸不合适等。因为这些问题导致网购的退货率极高，这一来一往对于消费者以及店商，都是成本。

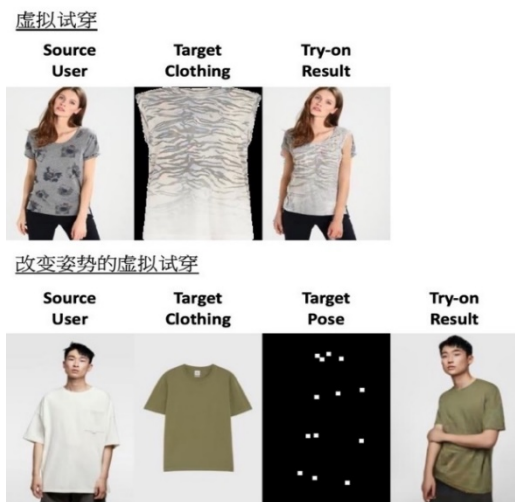


图1 虚拟试穿效果展示

因此我们基于深度学习中生成对抗网络 (Generative Adversarial Network, GAN) 的概念，设计了多姿势虚拟试穿模型<sup>[1]</sup>，如图2所示。不仅达到虚拟试穿的效果，更提供多姿势角度的呈现效果。我们基于渐进式学习，将模型分为三个阶段，(i) 姿势引导语义分割转换 (Pose-guided Parsing Translator)，(ii) 语义分割区块上色 (Segmentation Region Coloring)，以及 (iii) 重点区块优化 (Salient Region Refinement)。在阶段 (i)，我们先将要试穿的使用者照片，简化为语义分割图，再基于 ResNet，根据要试穿的衣服掩膜 (mask) 以及目标姿势，来生成变换姿势且试穿后的人体语义分割结果。接着在阶段 (ii)，借由前一个阶段生成的语义分割图，来引导真实色彩的上色区域，帮助模型认识人体的不同部位，例如：脸、头发、手臂、衣服区块等。此阶段使用了条件式生成对抗网络 (cGAN)，特别设计了判别器 (Discriminator) 的目标，并非单纯以生成的图片以及其对应的 groundtruth 作为目标，我们将目标设计为成对的形式：生成的图片 - 真实图片，groundtruth - 真实图片。这样的设计能够帮助生成器不仅生得像 groundtruth，也生得像真实的图片。最后，在阶段 (iii)，我们针对试穿衣服时，消费者最在乎的两个区块：脸以及衣服，分别使用 cGAN 以及 UNet 来进行优化，生成更多的细节，例如：发丝、五官轮廓、衣服褶皱等。优化脸部的 cGAN 架构。与阶段 (ii) 使用的架构相似，我们将阶段 (ii) 的 cGAN 架构去掉全连接层，以保存更多细节，用于优化脸部。此方法<sup>[1]</sup> 发表于 ACM Multimedia 会议上，并由下方链接观看更多结果。  
<https://github.com/fashion-on/FashionOn.github.io>

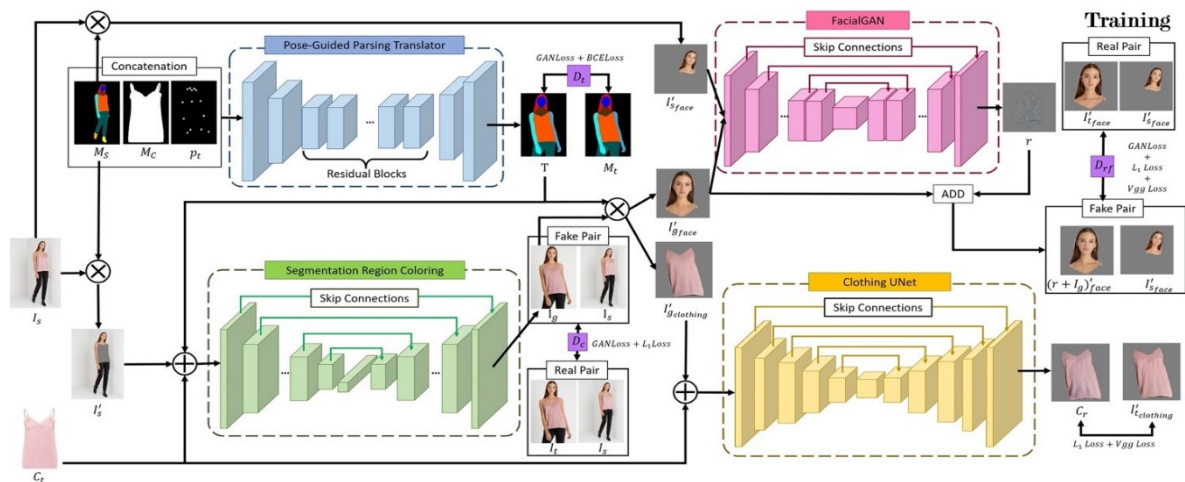


图2 多姿势虚拟试穿架构图

由于衣服多样性非常复杂，例如：有没有衣领、没有图案、有没有钮扣等诸多细节，我们强化方法<sup>[1]</sup>的优化模块，针对有穿着服饰的区域进行更细致的优化，使得细节保存更加拟真，并以此方法获得2020 CVPR Look Into Person Challenge冠军（多姿势虚拟服装试穿任务）。由于改变姿势会使得原本的下半身信息损失，因此针对下半身的裤子/裙子做优化，同时基于注意力机制 (attention)改进衣服优化模块。下身优化使用两个编码器 (encoder)，以及一个解码器 (decoder)，分别对转换姿势前的下身以及转换姿势后的下身抽取特征，藉由转换姿势前的下身信息，来强化转换姿势过程中丢失的信息，将两者融合后，解码出优化后的下身。上衣使用三个编码器，分别对下列三种衣服图片抽取特征值：(i)目标试穿服饰-代表全局信息、(ii)转换姿势后的衣服区块-为主要的信息来源，以及 (iii)前两者与转换姿势前的衣服区块的串接 (concatenation)-这边会侦测转换姿势前到转换姿势后的衣服信息间的光流 (flow)，并基于 (ii)的信息，使用注意力机制来将 (i)全局的信息以及 (iii)姿势间衣服流动信息，将细节生得更丰富。效果如图3。更详细的方法细节请参考下方公开视频。

<https://www.bilibili.com/video/BV1f54y1B7VJ/>

除此之外，我们发现特定的姿势，能够更好地呈现特定衣服的优点，例如：素色的T-shirt，搭配侧身的姿势，就可以呈现肌肉线条，效果如图4所示。因此我们延伸<sup>[1]</sup>，分析衣服特征与姿势间的关联性，基于衣服来生成适合的姿势，并进一步进行虚拟试穿。同时，我们设计一个新的架构来优化衣服，使用两个编码器，分别对

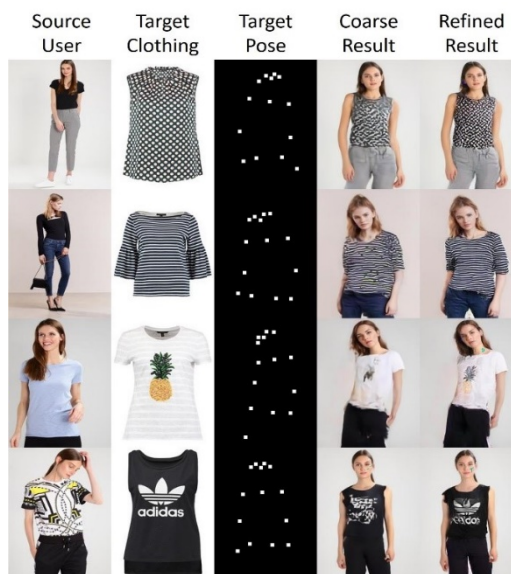


图3 2020 CVPR LIP Challenge 冠军方法试穿效果图

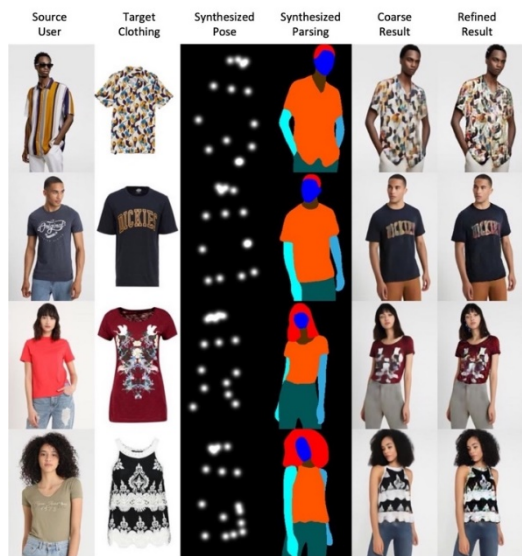


图4 根据衣服生成适当姿势的虚拟试穿效果展示

输入的衣服以及阶段 (ii)生成的衣服区块抽取特征, 再通过解码器, 生成优化后的衣服, 接着使用全局 (global) 以及区域性 (local) 判别器, 不仅判断生得像不像衣服, 更关注细节, 帮助生成细节含量更丰富的衣服。相关方法<sup>[2]</sup>已发表至IEEE Transactions on Neural Networks and Learning Systems (2021)。

针对时尚与计算机视觉的全面性发展, 我们与北京大学王选计算机研究所刘家瑛老师团队合作, 整理了近两百篇相关的论文、对应的数据集以及开源代码, 发表在ACM Computing Surveys (2021)<sup>[3]</sup>。

责任编辑 魏秀参

## 参考文献

- [1] C.-W. Hsieh, C.-Y. Chen, C.-L. Chou, H.-H. Shuai, J. Liu, and W.-H. Cheng, "FashionOn: Semantic-guided Image-based Virtual Try-on with Detailed Human and Clothing Information," The 27th ACM International Conference on Multimedia (MM 2019). (Oral, acceptance rate=26.5%).
- [2] C.-L. Chou, C.-Y. Chen, C.-W. Hsieh, H.-H. Shuai, J. Liu, and W.-H. Cheng, "Template-Free Try-on Image Synthesis via Semantic-guided Optimization," To appear in IEEE Transactions on Neural Networks and Learning Systems (TNNLS), 2021.
- [3] W.-H. Cheng, S. Song, C.-Y. Chen, S. C. Hidayati, and J. Liu, "Fashion Meets Computer Vision: A Survey," To appear in ACM Computing Surveys (CSUR), 2021.



### 陈婕云

台湾阳明交通大学电子研究所硕士生。主要从事计算机视觉、虚拟服装试穿相关研究, 获得 2020 CVPR LIP Challenge 冠军 (多姿态虚拟服装试穿任务)。  
Email: cychen.ee09g@nctu.edu.tw



### 帅宏翰

台湾阳明交通大学电机系副教授。主要从事机器学习、多媒体处理、计算机视觉相关研究, 指导学生获得 2020 CVPR LIP Challenge 冠军 (多姿态虚拟服装试穿任务)。  
Email: hhshuai@nctu.edu.tw



### 郑文皇

台湾阳明交通大学电子研究所特聘教授。主要从事多媒体信息处理、计算机视觉、人工智能相关研究, 指导学生获得 2020 CVPR LIP Challenge 冠军 (多姿态虚拟服装试穿任务)。中国计算机学会杰出会员, 并担任计算机视觉专委会委员。  
Email: whcheng@nctu.edu.tw

热点追踪

# 非接触可撤销掌纹识别研究进展

南昌航空大学 冷璐 南昌大学 闵卫东

## 一、引言

掌纹 (Palmprint) 识别利用手掌的线、点和纹理等特征鉴别身份, 具有多方面优势, 包括丰富的鉴别信息 (区域面积大)、较少的限制条件 (低分辨率和少量磨损情况下仍然可保持较高的识别精度)、较高的用户接受度、较强的隐蔽性 (不易泄漏) 等优势。因此, 掌纹迅速发展成为生物特征家族中的后起之秀<sup>[1]</sup>。

### 1.1. 非接触掌纹识别的意义和技术挑战

接触式掌纹识别时, 用户手掌需要接触采集装置表面, 导致以下问题: 大量用户触摸装置表面, 容易导致疾病传染, 疫情时期更为明显; 容易导致采集设备表面污染, 影响后续样本采集质量; 生物特征可能残留装置表面, 被不法分子获取, 增加信息泄露风险; 部分传统国家或民族禁止异性用户接触同一物体表面; 自由性和灵活性较低, 降低用户舒适度。

因此, 需要非接触式 (Contactless) 掌纹识别技术。然而, 非接触掌纹识别具有背景复杂多变、光照变化明显、手掌姿态和位置难以控制等挑战, 必须设计解决这些严峻挑战的实用化技术方案。

### 1.2. 生物特征模板保护的迫切需求

已有生物特征识别技术主要关注识别精度, 未充分考虑安全性以及隐私性。生物特征具有不变性, 不能像密码那样, 为了保证安全性而定期灵活更改。生物特征具有唯一性, 若多个数据库保存相同的生物特征模板, 可能导致跨数据库攻击。若某一数据库中的生物特征模板泄露, 则其他存有相同生物特征模板的数据库也会受

到威胁。生物特征还可能含有疾病、基因缺陷等信息, 直接存储会造成严重的隐私泄露风险。

## 二、研究现状

### 2.1. 非接触掌纹识别

本部分介绍非接触掌纹识别中的四项技术, 包括辅助拍摄、特征级优选及融合、数据压缩降维、下采样。

#### (1) 掌纹辅助拍摄技术

完全自动化的识别过程虽然可以提高用户的体验感和舒适度, 但是对算法精度、复杂度等提出了较高要求。在拍摄预览界面绘制辅助图形, 引导用户正确摆放手部位置和姿态, 从而在不明显降低用户舒适度的同时, 降低算法复杂度, 提高抵御各类干扰的鲁棒性以及复杂环境的识别精度<sup>[2]</sup>。

#### (2) 鉴别能量分析及多源空间拓展

通过鉴别能量分析 (Discrimination Power Analysis, DPA) 可以对每维特征进行鉴别性统计和排序, 优选高鉴别性的特征。特征优选不仅可以提高精度, 还降低了模板保存和匹配计算时的数据量, 从而降低了原始特征的隐私泄露。

以左右手的掌纹作为两个实例, 构建双源特征空间, 并在此空间进行动态加权 DPA, 实现多实例掌纹的特征级融合<sup>[3]</sup>。由于多源空间比单源空间具有更大的特征搜索范围, 可选择更多鉴别性分量, 提升了识别精度。此技术解决了特征级融合的融合源之间的特征尺寸和类型的非一致问题, 可实现不同维度和异质型特征的特征级融合和特征优选, 具有广泛通用性。

### (3) 基于二维双向随机投影的数据压缩降维

随机投影 (Random Projection, RP) 通过随机矩阵直接对原始高维数据进行压缩降维。与传统子空间算法相比, RP 可以完全无需训练, 具有极高的速度。通过构造稀疏 (Sparse) RP、二维 (2D) RP、双向二维 (Two-directional 2D) RP, 可明显降低 RP 的计算量和存储开销<sup>[4]</sup>。此技术可广泛应用于包括掌纹识别在内的各类数据压缩和降维场景。

### (4) 掌纹编码中的下采样策略

掌纹纹理特征编码可以不需要训练, 存储开销低, 匹配速度快。传统编码直接保留每个尺寸为  $4 \times 4$  的块内左上角的点。然而每个块内左上角的点不一定能准确表达整个块的信息, 并且对平移错位等干扰比较敏感。为此, 提出极限下采样保留每个块内具有最明显鉴别性特征的点<sup>[5]</sup>; 提出民主投票下采样通过块内所有的点参与投票, 共同决策下采样结果<sup>[6]</sup>。

## 2.2. 掌纹特征模板保护

通过特定双因变换, 将原始生物特征转变为扭曲或映射的可撤销形式。若可撤销模板受到威胁, 重新调整控制扭曲或映射的可变因子, 生成新的可撤销生物特征, 保证受保护模板具备替换和重新发布能力。两个因子缺一不可, 和单因认证相比, 具有更高的安全性<sup>[7]</sup>。

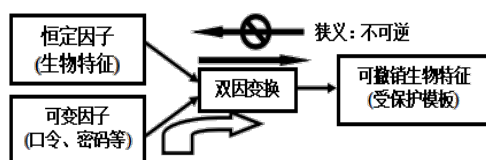


图 1 可撤销生物特征双因认证框架

构造了多样化、可撤销、可更新、可复用的掌纹特征模板, 包括二维掌纹哈希码 (2DPalmHash Code, 2DPHC)<sup>[8]</sup>、二维掌纹相位码 (2DPalmPhasor Code, 2DPPC)<sup>[9]</sup>, 避免了跨数据库的交叉匹配, 保证每个数据库中掌纹模板的安全性和隐私性, 也避免了直接使用原始掌纹特征引发的安全隐患和隐私风险。此框架可推广至其他生物特征模态, 解决推广应用的瓶颈问题。

## 2.3. 掌纹密码系统

若能够提取可靠稳定的生物特征直接作为密钥, 代替传统密钥并输入现有密码认证系统, 构造生物特征密

码系统 (Biometric Cryptosystem), 可充分发挥生物特征和密码学两种认证方式的双重优势。

然而生物特征比对和密钥比对具有明显差异。生物特征比对具有模糊性, 当注册和认证模板的差异度小于一定阈值, 即认定为合法用户。而传统系统的认证密钥要求绝对精确匹配。

模糊承诺 (Fuzzy Commitment) 和模糊保险箱 (Fuzzy Vault) 是两种典型的生物特征密码系统框架。基于 2DPHC 构造出可撤销掌纹模糊保险箱<sup>[10]</sup>, 有效抵御多种攻击。鉴于 2DPHC 的水平错位容忍性, 提取免于对齐的特征集合。此外, 将行单生模式扩展为行共生模式, 明显提高了重构精度。

## 三、未来发展方向

本文论述了非接触可撤销掌纹识别中三项技术的研究进展, 综合考虑了精度、效率、安全、隐私等多方面性能。未来发展方向包括以下几个方面。

### (1) 拓展至其他生物特征模态

其他生物特征模态的常见方法通常可以在掌纹识别中使用, 因此掌纹的相关技术框架为其他模态应用提供了重要的参考和借鉴, 理论上可以移植到其他模态。

### (2) 完备和严谨的定量化理论

由于生物特征模板保护和生物特征密码系统涉及模式识别、计算机视觉、网络空间安全等多个学科, 需要有机融合并充分借鉴各学科领域的研究方法、思路和评价标准, 形成更为完备、严谨的定量化理论。

### (3) 远程网络环境的多因子认证

远程网络环境通常面对无处不在的安全威胁以及有限的计算、存储、通信等资源束缚。需要根据具体的安全等级和实用性要求, 设计在精度、效率、安全等多方面综合权衡的多因子 (身份证等物品、密码等知识、掌纹等生物特征) 融合认证算法和远程协议<sup>[11]</sup>。

随着生物特征模板保护等研究的不断深入和完善, 生物特征识别技术的应用规模和范围将不断扩大, 为人们提供更为广泛的便捷认证服务。

责任编辑 储璐

## 参考文献

- [1] Fei L K, Lu G M, Jia W, Teng S H and Zhang D. 2019. Feature extraction methods for palmprint recognition: a survey and evaluation. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 49(2): 346-363 [DOI: 10.1109/TSMC.2018.2795609]
- [2] Leng L, Gao F M, Chen Q and Kim C S. 2018. Palmprint recognition system on mobile devices with double-line-single-point assistance. *Personal and Ubiquitous Computing*, 22(1): 93-104 [DOI: 10.1007/s00779-017-1105-2]
- [3] Leng L, Li M, Kim C S and Bi X. 2017. Dual-source discrimination power analysis for multi-instance contactless palmprint recognition. *Multimedia Tools and Application*, 76(1): 333-354 [DOI: 10.1007/s11042-015-3058-7]
- [4] Leng L, Zhang J S, Chen G, Khan M K and Alghathbar K. 2011. Two-directional two-dimensional random projection and its variations for face and palmprint recognition//*Proceedings of International Conference Computational Science and Its Applications*. Santander, Spain: 458-470 [DOI: 10.1007/978-3-642-21934-4\_37]
- [5] Yang Z Y, Leng L and Min W D. 2020. Extreme downsampling and joint feature for coding-based palmprint recognition. *IEEE Transactions on Instrumentation & Measurement*, 70: 1-12 [DOI: 10.1109/TIM.2020.3038229]
- [6] Leng L, Yang Z Y and Min W D. 2020. Democratic voting downsampling for coding-based palmprint recognition. *IET Biometrics*, 9(6): 290-296 [DOI: 10.1049/iet-bmt.2020.0106]
- [7] Patel V M, Ratha N K and Chellappa R. 2015. Cancelable biometrics: A review. *IEEE Signal Processing Magazine*. 32(5): 54-65 [DOI: 10.1109/MSP.2015.2434151]
- [8] Leng L and Zhang J S. 2013. PalmHash Code vs. PalmPhasor Code. *Neurocomputing*, 108: 1-12 [DOI: 10.1016/j.neucom.2012.08.028]
- [9] 冷璐,黎明,张明仁,金千植. 2015. PalmPhasor 算法性能的理论分析. *软件学报*. 26(5): 1237-1250 [DOI: 10.13328/j.cnki.jos.004594]
- [10] Leng L and Teoh A B J. 2015. Alignment-free row-co-occurrence cancelable palmprint Fuzzy Vault. *Pattern Recognition*, 48(7): 2290-2303 [DOI: 10.1016/j.patcog.2015.01.021]
- [11] Amin R, Islam S K H, Biswas G P, Khan M K, Leng L and Kumar N. 2016. Design of anonymity preserving three-factor authenticated key exchange protocol for wireless sensor network. *Computer Networks*, 101: 42-62 [DOI: 10.1016/j.comnet.2016.01.006]



## 冷璐

南昌航空大学副教授。主要研究方向为生物特征识别、生物特征模板保护、计算机视觉。  
Email: leng@nchu.edu.cn



## 闵卫东

南昌大学教授。主要研究方向为生物特征识别、计算机图形图像处理、计算机图形学、云计算分布式系统和智慧城市信息技术。  
Email: minweidong@ncu.edu.cn

顶会观察

# ICLR 2021

斯坦福大学 宋超

**国**际表征学习大会 (International Conference on Learning Representations, ICLR) 是机器学习领域的顶级会议, 与神经信息处理系统大会 (NeurIPS) 和国际机器学习大会 (ICML) 并称为机器学习三大顶级会议。最初由深度学习先驱, 图灵奖得主 Yoshua Bengio 和 Yann LeCun 于 2013 年创办。位列清华大学推荐学术会议和期刊列表 A 类会议。在 2021 年 Google Scholar Metrics 榜单中, h5 指数为 201, 高居所有出版物第 17 位, 「工程和计算机科学」领域第 3 位。

由于疫情原因, ICLR 2020 举办线上虚拟会议, 第九届年会 ICLR 2021 再次在线上举办虚拟会议。今年的会议日期是 5 月 3 日到 7 日。

## 一、ICLR 2021 的亮点

相比去年, 今年的会议组织汲取了经验, 大幅提升了线上会议的参会体验。例如, 海报 (Poster) 展示不再采取去年 Zoom 房间的形式, 而是代之以可以虚拟游览和自由走动的 GatherTown。口头报告 (Oral) 和邀请讲座 (Invited Talk) 由 SlidesLive 提前录制且与幻灯片同步, 并采用 RocketChat 让观众与讲者实时互动问答。会议也考虑到了世界各地研究者的时差问题, 例如海报展示和各类报告的时间都被分成两个及以上的不同时区让参会者灵活选择。本次会议由于在线上举行, 注册费用比疫情之前大幅下降。学生注册只需要 50 美元, 其他参会者为 100 美元。另外, 6 月 8 日以后本次会议各类报告的视频已经全部向公众免费开放。

除了论文成果展示之外, ICLR 2021 的一大特色是精心组织的社交活动。这些社交活动有着丰富多彩的目

标人群, 例如女性研究者、拉丁人群、韩国人、物理学背景的研究者、医疗健康从业人群等。这些社交活动提供了一些非正式的场所让各种身份的研究者们互相了解, 分享经验, 增进友谊。

## 二、论文录用和评审

ICLR 2021 总计收到了 2997 篇投稿, 录用了其中的 860 篇 (录取率 28.7%), 包括 53 篇 Oral (录取率 1.8%), 以及 114 篇 Spotlight (录取率 3.8%)。全球录取论文数量最多的机构是 Google, 达到 197 篇。斯坦福大学和加州大学伯克利分校分别以 124 篇和 94 篇文章紧随其后。国内论文录取数量最多的是清华大学、北京大学和上海交通大学。

共有 4086 名审稿人和 450 名领域主席参与到了论文评审之中, 贡献了超过 11500 条评审意见。ICLR 是论文评审平台 OpenReview 的创立者, 一直以来秉持着让论文评审过程更加开放透明的理念。所有 ICLR 论文的评审过程, 包括审稿人意见、作者回应, 以及领域主席的评论都是公开可见的。除此之外, 任何人也都可以在平台上对任何一篇论文公开发表意见, 从而影响评审过程。

## 三、邀请报告

本次会议共有 8 个邀请报告, 涉及机器学习领域最新的技术和理论进步、实际应用与落地, 以及可能导致的社会伦理问题等。前 Google Brain 研究科学家 Timnit Gebru 介绍了她因为揭露大规模语言模型的负面社会影响而被 Google 解雇的经历, 并结合埃塞俄比

亚内战等时事，深入浅出地分析了人工智能系统和社交网络平台对边缘人群（例如黑人女性）所带来的潜在压迫。华盛顿大学副教授 Yejin Choi 讨论了如何让自然语言处理系统具备常识，涉及到符号系统和神经网络的整合，以及用神经网络学到的知识表示进行逻辑推理。帝国理工学院教授、推特图学习研究中心主任 Michael Bronstein 用电影级别的视频论述了几何深度学习的宏大愿景，讨论了如何通过几何学中的不变性为深度神经网络自动引入归纳偏置。摩根大通人工智能研究部门主管 Manuela Veloso 介绍了人工智能在金融界的应用和前景。伦敦大学学院教授 Lourdes Agapito 总结了从二维的图片序列和视频中获取三维信息的最新进展。波士顿大学副教授 Kate Saenko 讨论了数据集的偏置问题，并描述了如何用迁移学习的域适应（domain adaptation）、半监督学习（semi-supervised learning）和生成式模型（generative modeling）来减少它的危害。首尔国立大学教授 Kyu Jin Cho 讲述了软体机器人的设计和实现，以及如何与人类友好交互。最后，加州大学伯克利分校教授 Alexei (Alyosha) Efros 从哲学和心理学的角度出发，论述了自监督学习（self-supervised learning）除了减少训练所需的数据标注量以外，还具有可能变革当前人工智能研究既有框架的更为深远的意义。

#### 四、会议奖项与热点

ICLR 2021 共有 8 篇文章获得杰出论文奖（Outstanding paper award）。值得注意的是，并非所有的杰出论文都是从 Oral 文章里筛选出来的——其中有 2 篇属于 Spotlight。这表明杰出论文委员会的评选并没有拘泥于审稿人和领域主席对文章的打分。获得杰出论文奖的文章分别为：

**Score-Based Generative Modeling through Stochastic Differential Equations**<sup>[1]</sup>，来自斯坦福大学和 Google Brain。文章提出了一个全新的生成式模型的框架：利用随机微分方程可以很容易地把数据变成噪声；学习该随机微分方程的逆过程就可以从噪声中生成数据。这个框架统一了计分匹配生成式模型（score-based generative models, SGM）以及降噪扩散模型

（denoising diffusion probabilistic models, DDPM），同时能够精确计算似然函数（likelihood）和求解逆问题（inverse problems）。

**Rethinking Architecture Selection in Differentiable NAS**<sup>[2]</sup>，来自加州大学洛杉矶分校和滴滴 AI 实验室。文章指出了可微神经网络结构搜索（Differentiable NAS）研究中一个长久被忽视的重要问题。作者提出不应该按照权值的大小来决定是否选择一个神经网络链接，而应该采用微扰的方法判断它对损失函数的影响程度。

**Optimal Rates for Averaged Stochastic Gradient Descent under Neural Tangent Kernel Regime**<sup>[3]</sup>，来自日本东京大学和理化学研究所。文章使用神经切向核（Neural Tangent Kernel）来研究训练双层过参数化（over-parameterized）神经网络的最优收敛速率，并证明平均随机梯度下降法（averaged stochastic gradient descent）在一定条件下可以达到这个最佳收敛速率。

**Neural Synthesis of Binaural Speech from Mono Audio**<sup>[4]</sup>，来自 Facebook Reality Labs。文章提出了一种端到端的方法把单声道音频转化成能够随位置信息而变化的双声道音频，获得了超越以往所有方法的表现，并给出了一个在无回声室采集到的双声道音频数据集。

**Learning Mesh-Based Simulation with Graph Networks**<sup>[5]</sup>，来自 DeepMind。文章给出了一个图神经网络模型来学习基于多边形网格的计算机模拟，速度上获得了比传统有限元方法一到两个数量级的提升，并展示了在流体力学、结构力学、电磁学和声学等各方面的应用。

**EigenGame: PCA as a Nash Equilibrium**<sup>[6]</sup>，来自 DeepMind。文章指出传统的主成分分析（Principal Component Analysis, PCA）可以转化为一个博弈论问题。利用梯度优化求解这个博弈论问题，可以大幅提升主成分分析算法的并行度，从而能够用来分析超大规模的数据集。

**Complex Query Answering with Neural Link Predictors**<sup>[7]</sup>, 来自伦敦大学学院以及阿姆斯特丹自由大学。文章研究如何用神经网络学习从知识图谱中提取信息来回答一阶逻辑命题。相比现有的方法,大幅减少了训练数据集的规模,提升了泛化性能,并且具有更好的可解释性。

**Beyond Fully-Connected Layers with Quaternions: Parameterization of Hypercomplex Multiplications with  $1/n$  Parameters**<sup>[8]</sup>, 来自亚马逊 AWS、Google Research、苏黎世联邦理工等。文章推广了之前四元数的方法,提出用 Kronecker 乘积来表示神经网络全连接层中的权值矩阵,从而减小参数数量。

除了 8 篇杰出论文之外,本次 ICLR 2021 还收录了 53 篇 Oral 论文和 114 篇 Spotlight 论文,这些论文也有很大的影响力。例如来自 Google Brain 的 Vision Transformer,首次表明在计算机视觉领域

Transformer 有着取代卷积神经网络 CNN 的潜力;来自 Google、剑桥大学等机构的研究者提出了 Performer,实现了线性复杂度的自注意力计算等。

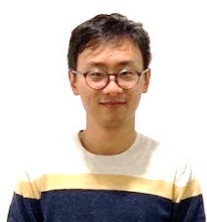
## 五、总结与展望

参加完今年的 ICLR 会议,可以看到如下几个领域正在迎来飞速的发展:(1)半监督学习(Self-supervised learning);(2)Transformer 的加速以及在计算机视觉中的应用;(3)计分匹配生成式模型(Score-Based Generative Models)和降噪扩散模型(Diffusion Models);以及(4)图神经网络(Graph neural networks)。此外,受疫情影响,线上会议仍然可能是明年 ICLR 的主要形式。随着线上会议的组织变得愈加完善,疫情之后的 ICLR 有可能采取线上和线下混合的方式,降低注册费用,方便因为各种原因无法到现场的研究者参与交流。

责任编辑 王金甲

## 参考文献

- [1] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling for Stochastic Differential Equations, ICLR2021.
- [2] Ruochen Wang, Minhao Cheng, Xiangning Chen, Xiaocheng Tang, and Cho-Jui Hsieh. Rethinking Architecture Selection in Differentiable NAS, ICLR2021.
- [3] Atsushi Nitanda, and Taiji Suzuki. Optimal Rates for Averaged Stochastic Gradient Descent under Neural Tangent Kernel Regime, ICLR2021.
- [4] Alexander Richard, Dejan Markovic, Israel D. Gebru, Steven Krenn, Gladstone Alexander Butler, Fernando Torre, and Yaser Sheikh. Neural Synthesis of Binaural Speech from Mono Audio, ICLR2021.
- [5] Tobias Pfaff, Meire Fortunato, Alvaro Sanchez-Gonzalez, and Peter Battaglia. Learning Mesh-Based Simulation with Graph Networks, ICLR2021.
- [6] Ian Gemp, Brian McWilliams, Claire Vernade, and Thore Graepel. EigenGame: PCA as a Nash Equilibrium, ICLR2021.
- [7] Erik Arakelyan, Daniel Daza, Pasquale Minervini, and Michael Cochez. Complex Query Answering with Neural Link Predictors, ICLR2021.
- [8] Aston Zhang, Yi Tay, Shuai Zhang, Alvin Chan, Anh Tuan Luu, Siu Hui, and Jie Fu. Beyond Fully-Connected Layers with Quaternions: Parameterization of Hypercomplex Multiplications with  $1/n$  Parameters, ICLR2021.



## 宋颺

斯坦福大学人工智能实验室在读博士,师从 Stefano Ermon。本科毕业于清华大学物理系。研究方向为生成式模型和机器学习鲁棒性。苹果奖学金和摩根大通奖学金得主。ICLR 2021 杰出论文奖得主。

Email: yangsong@cs.stanford.edu

## 南开大学程明明教授访谈

2021年3月9日,《CCF-CV专委简报》在线采访了南开大学计算机系主任、博士生导师程明明教授。下面是采访实录。

程老师,您好!首先,请您分享一下您的个人学习和研究经历。

我于2007年本科毕业于西安电子科技大学,之后免试进入清华大学并师从胡事民教授攻读博士学位,主要研究计算机图形学。2012年博士毕业之后,前往英国牛津大学去Philip Torr教授课题组做了2年博士后,主要研究计算机视觉问题。2014年回国,在南开大学计算机学院工作至今。

您在计算机视觉和计算机图形学等领域内颇有成就,能否介绍一下您在这些领域中最突出的几项研究成果?

研究成果方面,我一直认为好的研究成果应该可以通过创造性的思维去改变大家的固有认知,同时成果本身对实际应用具有明显的促进和推动作用。虽然这些年做了很多研究工作,真正能够做到这一点的也不是特别多。第一个是显著性物体检测领域的研究工作,通过这些年的不断努力,我们的研究成果多次刷新了大家对显著性物体区域怎样检测,显著性检测方法的能力边界等问题的一些固有认知。第二个是关于多粒度图像场景理解方面的研究,视觉场景中非常广泛的存在多粒度的信息,如何充分利用这些信息成为设计高效鲁棒算法的关

键,我们的研究在这方面也取得了一些有意义的进展。

您近年发表了多篇CCF A类顶刊或顶会学术论文,而且您的论文具有很高的引用次数,能跟大家分享一下您是如何做到持续产出高水平论文的呢?

在开展科研工作的过程中,我一直反复地问自己,这个问题到底值不值得投入很大的精力,我们的努力是否能给同行更好地解决类似问题提供有意义的帮助。对第一个问题的思考推动我不断地跟工业界同仁进行沟通,寻找真正重要的问题。对于第二个问题的思考始终提醒我,重要的发现,往往是简单深刻的,经得起反复验证的。对这两个问题的关注让我在研究工作中受益颇丰,也分享给大家。

在CCF A类国际期刊会议为代表的国际同行评审中,大家对于能够帮助同行们更好解决问题的简单深刻的发现非常看重。通过持续不断地观察和总结这些有意义的经验,我相信发表论文和论文引用次数的提高都是水到渠成的事情。学术影响力的提高同样只是结果,它的因同样在于给学术界和工业界的同行提供有意义的思路、方案、和支持。

能分享一下提高学术影响力有哪些好的容易操作的做法吗?

我的经验首先是关注和解决给大家科研和工作中造成瓶颈的卡脖子问题,并积极主动地将有用的思路和成果跟同行们进行分享。选题方面适合每个人的经验不

同。但是知识分享和传播方面有很多共性。我曾经提出过一个“开放共享科研记录行动倡议 (DOCX)”。包括：

Demo: 尽可能为自己科研论文中涉及的每个问题做一个在线的 demo, 方便自己随时使用, 方便同领域老师的课堂教学实验, 公众科普。

Open source: 为了避免大家不必要地反复低水平重复工作, 我们应尽量开源自己每篇论文的代码。这里的开源, 除了开源代码, 也包括开源实验数据。

Chinese version: 在英文顶级会议或者期刊发表论文的同时, 可以在自己主页上共享一个中文翻译版, 方便国内读者阅读。

eXplain: 大家在阅读论文时, 经常碰到疑惑, 希望能够和作者及同行交流。作者很多时候也希望听到来自同行的反馈。因此, 我建议大家尽量为自己的每篇论文建立一个项目主页, 并积极回答读者留言提问。

您很注重技术成果转化, 可否举例说明一下您的研究成果是如何成功落地应用的呢? 在成果转化中应该注意什么呢?

正如上面提到的, 我们很多科研成果的出发点是充分了解工业界需求。在长期开源共享的过程中, 我们也和工业界建立了很多信任。因此, 我们有幸能够接触到很多成果转化落地的机会。我感觉转化的过程中最应该注意的是开源共享和企业诉求之间的平衡。一方面, 通用的知识可以通过论文的形式共享出来。另一方面, 针对企业产品具体需求的成果需要注意保护企业的商业利益。只有互利共赢的合作才能长期为双方带来价值。

您获批了多项国家级项目, 能跟大家分享一下您成功申请的经验和体会吗?

我认为项目和科研是一样的。不能从要发表论文,

或者要获得项目作为出发点。而应该从要解决那些重要的问题, 怎么样做才能更好地解决问题出发。如果能够长时间地坚持去解决同行们关注的重要问题, 并且乐于分享, 我相信在申请国家级项目的同行评议中, 我们就能够得到同行们的支持。

您领导着非常优秀的团队, 请问您是如何管理和运作您的团队的? 您是如何管理研究生的? 您对他们的要求是什么?

我曾经看过这么一个说法, 三流的团队靠打卡, 二流的团队靠绩效, 一流的团队靠理想。我很认同这个说法。作为大学教授, 我们有义务去帮助自己的研究生实现他们的职业理想。我经常跟学生聊他们的职业规划, 并提醒和帮助他们做一些具体的长期的准备。我相信, 哪怕只有一半的学生对自己的职业发展很满意, 团队就已经在非常好的状态了。此外, 我们成果转化过程中对应的收益很大部分都给了学生本人。让研究生们在取得学术成果的同时能够收获自己的劳动果实。对于研究生来讲, 导师能够从自己的理想出发帮助自己, 并充分尊重学生的劳动成果, 这种情况下不需要太多额外的管理。

如果吐露研究工作者的心声, 您最想说的是什么呢?

研究工作无疑是辛苦的。我们所有研究的意义, 都是试图去超越人类对重要问题的认知极限。哪怕只是超越了一点点, 它的背后离不开艰辛的努力。但是科研不仅仅是辛苦, 更包含取得突破时的喜悦。研究不是百米赛跑, 需要我们去享受这个过程, 才能走得更长久。我最想说的是, 科研工作要不忘初心, 知道自己要做什么, 应该以什么样的心态去面对困难。只有享受探索的过程, 而不是探索的结果, 才能收获更多。

责任编辑 赵振兵 余烨



## 程明明

程明明，南开大学教授，计算机系主任，国家“万人计划”青拔(2016年度)、“优青”(2019年)。他的主要研究方向是计算机视觉和计算机图形学，在 CCF A 类顶级国际期刊和会议上发表学术论文 60 余篇，论文 Google 学术引用 2 万余次，一作论文单篇最高引用 3000 余次，连续 4 年入选 Elsevier 中国高被引学者榜单。技术成果被应用于华为手机智能拍照、推想科技 CT 影像智能分析、金风科技风电设备运行监控、和中化农业病虫害识别等领域。获得 ACM 中国新星奖、天津市青年科技奖、吴文俊人工智能自然科学二等奖、中国图象图形学学会自然科学一等奖、教育部自然科学一等奖等奖项。现为中国图象图形学学会副秘书长、中国计算机学会计算机视觉专委会常务委员，并担任 SCI 一区期刊 IEEE TIP 编委。

## 委员好消息

2021 年 4 月 22 日，爱思唯尔正式发布了 2020 年中国高被引学者榜单，CCF-CV 专委会共有 34 位委员入选，他们是华中科技大学白翔教授、北京大学林宙辰教授、彭宇新教授、北京邮电大学邓伟洪教授、大连理工大学卢湖川教授、哈尔滨工业大学邬向前教授、徐勇教授、左旺孟教授、杭州电子科技大学俞俊教授、华中科技大学尤新革教授、江西财经大学方玉明教授、南京航空航天大学谭晓阳教授、南京理工大学李泽超教授、唐金辉教授、南开大学程明明教授、

清华大学郭振华副研究员、鲁继文副教授、厦门大学纪荣嵘教授、天津理工大学陈胜勇教授、同济大学张林教授、武汉大学荆晓远教授、西安电子科技大学邓成教授、董伟生教授、高新波教授（现在重庆邮电大学）、西北工业大学程焱研究员、韩军伟教授、聂飞平教授、王琦教授、中科院深圳先进院乔宇研究员、中科院西安光机所卢孝强研究员、中科院自动化所雷震研究员、王亮研究员、中山大学林惊教授、郑伟诗教授。

责任编辑 刘海波

# 无参考图像质量评价开源代码

兰州理工大学 刘昊 李策

**无**参考图像质量评价 (Blind image quality assessment) 是指在没有参考图像和信息的前提下, 对图像质量进行准确的评判。随着图像数据的广泛应用, 有效的图像质量评价不仅可以作为独立的工作提高人们在图像中获取信息的效率, 而且可以作为辅助任务提高其它计算机视觉任务的精度, 例如在图像增强、去噪、缩放等任务中, 常常通过准确比较处理前后图像的质量, 对相关图像处理算法的性能做出合理的判断。此外, 远程图像或视频传输、图像筛选等方面, 也需要通过评价图像的质量来调整传输过程或筛选算法的参数和模型。同时, 用来评价的参考图像及信息难以获取, 故研究 BIQA 有着极为重要的现实意义和广泛的应用前景。

随着 BIQA 在现实中的广泛应用, 学者们不再局限于合成失真图像的质量评价, 开始更多的关注真实失真图像的质量评价, 并结合其它的视觉任务对图像质量进行全新的定义, 使得 BIQA 越来越靠近我们的生活。自 2015 年真实失真图像质量评价数据库 LIVE Challenge 建立, 许多学者开始研究能够同时解决真实和合成失真图像质量评价问题的模型。本文将重点介绍几个能够解决多样失真图像质量评价的开源模型, 以及与其它计算机视觉任务紧密相关的图像质量评价数据集。

## 1、Blind Image Quality Assessment Using A Deep Bilinear Convolutional Neural Network

**工作:** 该论文设计了深度双线性模型来实现图像质量评价, 如图 1 所示。该模型用两个卷积神经网络(CNN)来同时处理合成失真和真实失真两类不同失真的图像。对于合成失真图像, 作者用经过大量数据预训练的 CNN

对图像失真类型和级别进行分类。对于真实失真图像, 采用预训练的 CNN 进行图像分类。对两个 CNN 的特征进行双线性池化, 得到一个统一的特征表示, 并进行最终的质量预测。

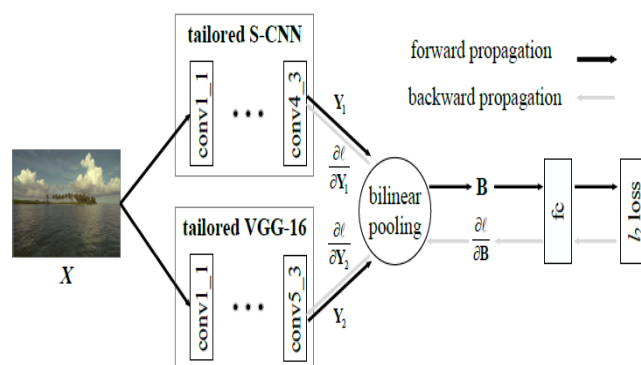


图 1 基于双线性池化的 BIQA

**论文:** <https://ieeexplore.ieee.org/document/8576582>

**代码:** <https://github.com/zwx8981/DBCNN>

## 2、Multiple Level Feature Based Universal Blind Image Quality Assessment Model

**工作:** 该论文提出了一个考虑多层特征的 BIQA 模型, 来确保迁移模型在多样的失真图像质量评价中有较好的鲁棒性。作者考虑到 CNN 网络的低层特征有利于评价合成失真图像, 高层特征有利于评价真实失真图像。因此, 通过定义编码层, 在深度 CNN 模型的每层中提取有意义的特征, 然后将这些特征串联并回归到主观评分中, 对模型进行训练。通过这种方式得到的 BIQA 模

型能够有效评价多样的失真图像质量。

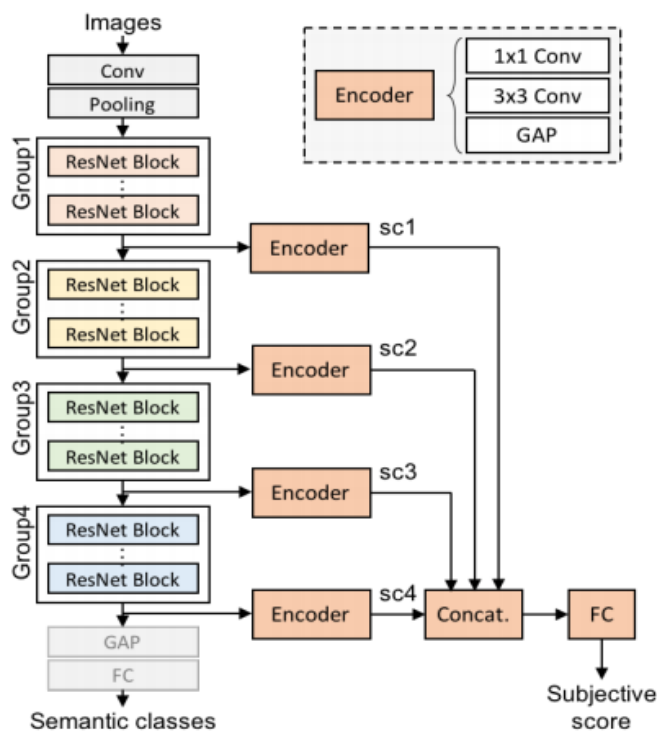


图2 基于多级特征的 BIQA

论文: <https://arxiv.org/abs/1712.00516>

代码: <https://github.com/zer01ike/multiple-level-feature-BIQA>

### 3、Continual Learning for Blind Image Quality Assessment

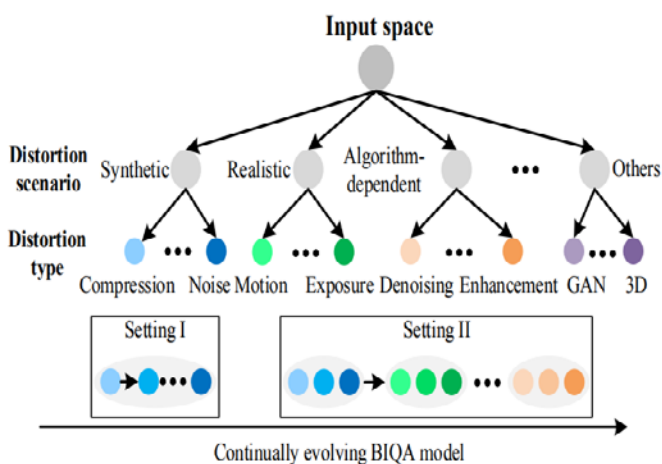


图3 BIQA 的持续学习

工作: 该论文的侧重点在于利用持续学习的方法来解决

新加入的失真图像会导致数据遗忘的问题,如图3所示。所提算法针对不同的数据库设计不同的预测头,同时这些预测头共享主干网络。然后利用正则化,使得所有的预测头都能够随着新的数据集不断进化,并防止对旧数据的遗忘,从而得到较好的图像质量评价结果。

论文: <https://arxiv.org/pdf/2102.09717v1.pdf>

代码: [https://github.com/zwx8981/BIQA\\_CL](https://github.com/zwx8981/BIQA_CL)

### 4、Blind Image Quality Assessment Based on Multi-scale KLT

工作: 该论文主要研究感知相关的 BIQA 特征。首先利用去均值对比度归一化系数对图像进行归一化,然后利用 Karhunen-Loève transform(KLT)以数据驱动的特征提取方式,提取图像的结构特征,利用不同大小的核进行多尺度分析。最后利用广义高斯分布对 KLT 系数在不同光谱分量中的分布进行建模。所提算法在常见和不常见的失真类型上,都能够得到与人类主观得分高度一致的结果。

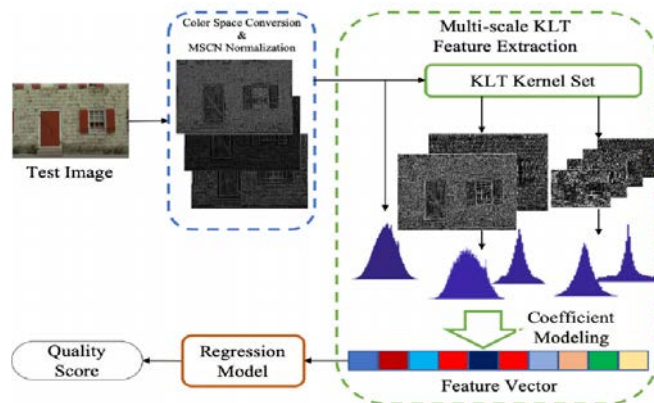


图4 基于 KLT 学习的 BIQA

论文: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9113755>

代码: <https://github.com/charlesyangshu/MsKLT>

### 5、Assessing Image Quality Issues for Real-World Problems

工作: 该论文将图像质量评价与实际视觉任务相结合,

提出了一个新的大型图像质量评价数据库 VizWiz-QualityIssues。不同于传统的图像质量评价数据库，VizWiz-QualityIssues 中的 39181 张图像均由盲人拍摄，示例图像和标签如图 5 所示。

在这个工作中，作者认为实际的视觉任务可以分为可识别性分类任务和可回答性分类任务两类。因此，两类任务对图像质量的要求就在于图片的内容是否因质量缺陷而不可识别或不可描述。该数据集包括模糊、过亮、过暗、遮挡、旋转、其它等 7 类由盲人拍摄时造成的图像质量缺陷，所有图片的标签在众包平台 Amazon Mechanical Turk 上获得。图 6 是图像质量缺陷和可识别性间的百分比。

论文: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9157810>

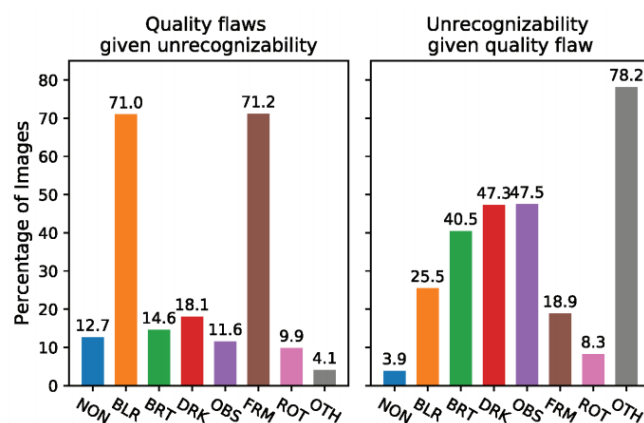


图 6 左图为无法识别的图像中，不同质量缺陷的百分比，右图为存在质量缺陷的图像中，不可识别的百分比。

代码: <https://github.com/chiutaiyin/VizWiz-QualityIssues>



图 5 VizWiz-QualityIssues 数据库示例图像和标签

责任编辑 沈沛意



刘 昊

兰州理工大学电气工程与信息工程学院博士研究生，研究方向为图像质量评价、计算机视觉。



李 策

兰州理工大学电气工程与信息工程学院教授，博士生导师，主要研究方向为计算机视觉、智能信息处理、机器学习。

# 人脸活体检测数据集

中国科学院自动化研究所 万军

人脸识别技术的应用面临着很多安全隐患。不法分子在社交媒体上获取合法用户的打印人脸、视频回放或 3D 面具，通过向人脸识别系统展示攻击样本以窃取用户的数据与财产。因此，人脸真伪判断是人脸识别系统的一道重要的安全防线。

面对如此多种且不可预知的人脸攻击，设计一个鲁棒且检测准确率高的算法是学术界与工业界共同关注的研究方向。同时，训练模型所采用的数据集也是影响模型性能的一个重要因素。其中，数据集中样本的质量、攻击种类、场景环境设置、采集设备种类等是一个高质量数据集必须考虑的重要因素。本文详细介绍学术界人脸活体检测任务中主流的数据集，包括 NUAA, Replay-Attack, CASIA-FASD, MSU MFSD, MSU USSA, OULU-NPU, SiW, CASIA-SURF, CASIA-SURF CeFA, CASIA-SURF HiFiMask, CASIA-SURF 3DMask。

## 1、NUAA 数据集

**介绍:** NUAA(NUAA photograph imposter database) 数据集由南京航空航天大学于 2010 年发布，是学术上发布较早的数据库之一。如图 1 所示，该数据集包含 15 名被采集者，共收集约 12614 张人脸图像，包含真实人脸与照片攻击人脸两类，且所有图像的分辨率为  $640 \times 480$  像素。整个数据集分三个阶段完成采集，每次间隔两周且设置不同的光照条件。为了提高算法在实际应用中的鲁棒性，数据集采集过程中设置了多种人脸外观变化，包括性别、光照环境、是否带眼镜等因素。同时，伪造照片分别进行了移动、旋转、弯折等操作，以模拟

实际中照片攻击的真实情况与真实真脸的深度结构。

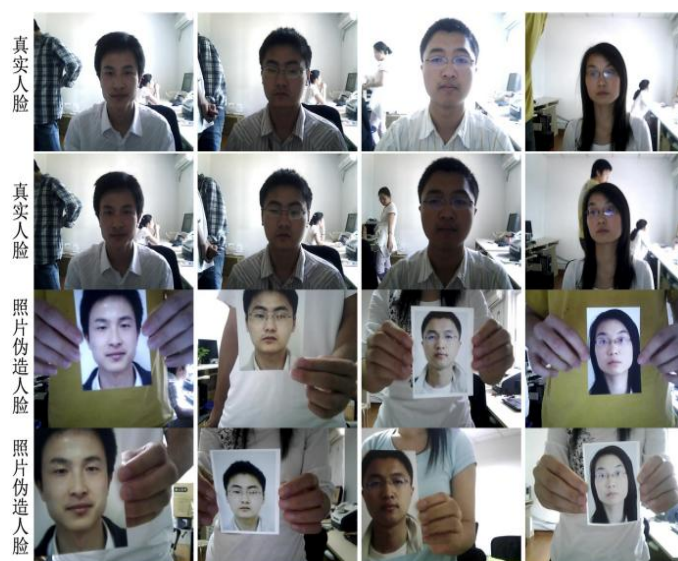


图 1 NUAA 数据集示例

NUAA 作为本领域较早提出的数据集，存在一定的不足，例如仅包含照片伪造人脸，照片的尺寸和分辨率比较单一，并与真实人脸尺寸相差较大。参与的被采集者和每个被采集者拍摄的照片数目都较为有限。

**数据集地址:** [http://parnec.nuaa.edu.cn/\\_upload/tp/02/db/731/template731/pages/xtan/NUAAImposterDB\\_download.html](http://parnec.nuaa.edu.cn/_upload/tp/02/db/731/template731/pages/xtan/NUAAImposterDB_download.html)

## 2、Replay-Attack 数据集

**介绍:** Replay-Attack 是瑞士戴尔莫尔感知人工智能研究所 (Idiap research institute) 于 2012 年提出的人脸活体检测数据集，并在 2013 年 ICB 组织的第二次人脸

活体检测竞赛用到该数据库。该数据集包含来自 50 名被采集者在不同光照条件和采集规则下拍摄的共 1300 段视频组成, 其中每帧图像的分辨率均为  $320 \times 240$  像素。

Replay-Attack 数据集在“受控环境”和“复杂环境”中进行采集。在“受控环境”中, 开启室内灯光并放下百叶窗以保证光照均衡, 背景环境也较为单一; 在“复杂环境”中, 打开百叶窗并关闭室内灯光, 同时在更加复杂的背景中采集数据, 以增加活体检测的难度。此外, 对于照片和视频伪造人脸, 分别在“手持状态”和“固定条件”下进行了拍摄。在“固定状态”下, 照片伪造人脸在视频片段中不会有任何位置变化, 而屏幕伪造人脸的位置变化与录制的攻击视频中完全一致; “手持状态”下拍摄的伪造人脸视频片段中, 人脸区域均会由于手持攻击媒介自身的运动而产生额外的位置变化。图 2 所示为部分样本, 其中第一行样本在“受控环境”中采集, 第二行样本在“复杂环境”中采集, 从左到右依次为真实人脸, 打印照片攻击, 手机电子屏幕伪造攻击, 笔记本电脑屏幕伪造攻击。



图 2 Replay-Attack 数据集中的部分样本

**数据集地址:** <https://www.idiap.ch/en/dataset/replayattack>

### 3、CASIA-FASD 数据集

**介绍:** CASIA-FASD 数据集由中国科学院自动化研究所 (Institute of Automation, Chinese Academy of Sciences) 生物识别与安全技术研究中心在 2012 年提出。该数据集含有来自 50 名用户的共计 600 段视频, 涉及三个不同类别的攻击样本, (1)传统照片攻击: 攻击者打印出合法用户的高清照片, 并进行旋转、扭曲或弯

折操作, 以模仿真实人脸的运动情况; (2)剪裁照片攻击: 对照片的双眼区域进行抠图, 并隐藏在照片后眨眼, 以模拟真实人脸的眨眼运动; (3)视频攻击: 在 iPad 等电子设备屏幕上, 展示事先录制好的人脸视频片段。

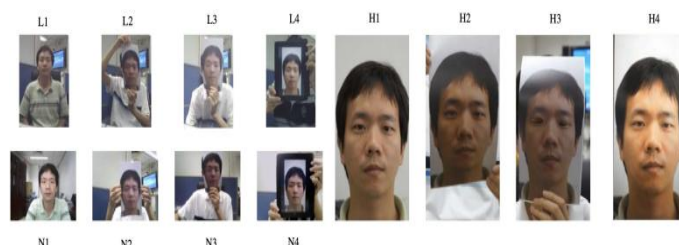


图 3 CASIA-FASD 数据集中的部分样本

该数据集要求每个被采集者录制 3 段真实人脸视频和 9 段攻击人脸视频。如图 3 所示, 每种类型的样本包含 3 种不同分辨率的视频片段, 分别为 Sony 相机采集的  $1280 \times 720$  像素高画质视频片段(H1-H4), 新购置 USB 摄像头采集的  $640 \times 480$  像素中等画质视频片段(N1-N4)和旧摄像头采集的  $640 \times 480$  像素低画质视频片段(L1-L4)。该数据集在不同背景、不可控的光照条件下采集以模拟真实的部署场景, 并使用去除双眼区域的照片以模仿真实人脸的眨眼运动, 同时提供的不同分辨率图像可以更好地对活体检测算法的通用性和实用性进行评估, 它是人脸活体检测领域典型数据集之一。

**数据集地址:** <http://www.cbsr.ia.ac.cn/china/Databases%20CH.asp>

### 4、MSU MFSD 数据集

**介绍:** MSU MFSD (Mobile Face Spoofing Database) 数据集由来自密歇根州立大学 (Michigan State University) 模式识别与图像处理实验室于 2015 年采集并公布。包含 35 名被采集者的共 280 段视频。如图 4 所示, 使用 Google Nexus 5 智能手机摄像头(第一行)和 MacBook Air 13 笔记本电脑摄像头(第二行)采集的一些样本。其中(a)为真实人脸; (b)为 iPad 生成的视频回放攻击人脸; (c)为 iPhone 生成的视频回放攻击人脸; (d)为打印照片攻击。该数据集主要应用于使用多

种电子设备进行人脸活体检测的场景。

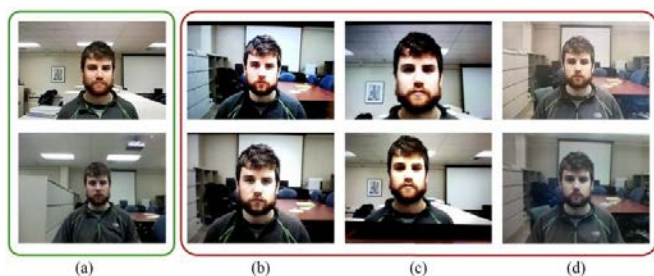


图 4 MSU MFSD 数据集中的部分样本

**数据集地址:** <http://biometrics.cse.msu.edu/Publications/Databases/MSUMobileFaceSpoofing/>

### 5、MSU USSA 数据集

**介绍:** MSU USSA(Unconstrained Smartphone Spoof Attack database)数据集由密歇根州立大学模式识别与图像处理实验室于 2016 年提出, 包含来自 1040 名被采集者的 9360 张照片。与 MSU MFSD 数据集类似, 主要用于基于电子设备屏幕的伪造人脸的检测。主要区别是, 本数据集中所有数据均为单帧图像, 不含有时间信息。一些图像样本如图 5 所示。



图 5 MSU USSA 数据集中的部分样本

**数据集地址:** [http://biometrics.cse.msu.edu/Publications/Databases/MSU\\_USSA/](http://biometrics.cse.msu.edu/Publications/Databases/MSU_USSA/)

### 6、OULU-NPU 数据集

**介绍:** OULU-NPU 数据集由芬兰奥卢大学(University of Oulu)于 2017 年采集, 它同时用于 2017 年 IJCB 人脸活体检测竞赛的数据集。该数据集包含来自 55 名被采集者的共 4950 段视频, 每帧图像的分辨率为 1920×1080 像素。收集者分别使用 6 个不同品牌手机的前置摄像头在三种不同的光照和背景环境中拍摄(即分三个时期采集完成), 并基于两种不同材质的照片和两种不同品牌的电子屏幕设备作为攻击媒介进行伪造人脸攻击, 部分样本如图 6 所示。



图 6 OULU-NPU 数据集中的部分样本

另外, 该数据集定义了四个不同的测试协议, 分别评估活体检测算法在不同光照背景(时期)、不同攻击媒介(打印攻击与视频回放攻击)、不同采集设备(6 种不同品牌的手机)等设定下的检测准确率与泛化性。

**数据集地址:** <https://sites.google.com/site/oulunpudatabase/>

### 7、SiW 数据集

**介绍:** SiW(Spoofing in the Wild Database)数据集由密歇根州立大学(Michigan State University)于 2018 年收集并发布。随着传感器技术的进步, 现有的人脸活体检测系统很容易受到新兴的高质量媒介的攻击。为了使系统能够抵御这些攻击, 该课题组收集了 SiW 数据集, 包含来自 165 个被采集者的真实人脸与多种攻击视

频，总共 4478 段。对于每个被采集者，采集 8 段真实人脸视频和最多 20 段攻击视频，每段视频均为 30fps，约 15 秒。如图 7 所示，为了提升算法对无关活体线索的抵抗能力，采集过程中分四个时期完成，以丰富采集环境与光照变化，并要求被采集者做出摇头、表情变化、人脸姿势变化与距离变化等操作。攻击视频由真实人脸视频通过打印纸张与电子屏幕等载体进行展示并采集。为了进一步探究该数据集的性能，定义了三个不同的协议：第一个协议旨在评估不同面部姿势和表情下活体检测算法的泛化性能。第二个协议评估算法对同一种攻击类型在不同攻击媒介上的泛化能力。第三个协议旨在评估算法对未知攻击类型的泛化性能。



图 7 SiW 数据集中的部分样本

**数据集地址:** <http://cvlab.cse.msu.edu/siw-spoof-in-the-wild-database.html>

## 8、CASIA-SURF 数据集

**介绍:** CASIA-SURF 数据集由中国科学院自动化研究所 (Institute of Automation, Chinese Academy of Sciences) 生物识别与安全技术研究中心在 2019 年提出。同时，该数据集用于 CVPR2019 上的人脸活体检测竞赛 - ChaLearn Face Spoofing Attack Detection Challenge@CVPR2019，吸引了国内外学术界与工业界共 300 多支队伍参与。该数据集是一个大规模公开数据集，包含 1000 个被采集者共 21000 段视频。对于每一位采集者，包含一个真实人脸样本与 6 种打印攻击，分别包括弯曲与多种抠图方式的组合，部分样本如图 8 所示。为了探究多模态信息对人脸活体检测性能的提升，该数据集同时为每一个样本提供 3 种模态的视频，分别为 RGB，Depth 与 IR 模态。

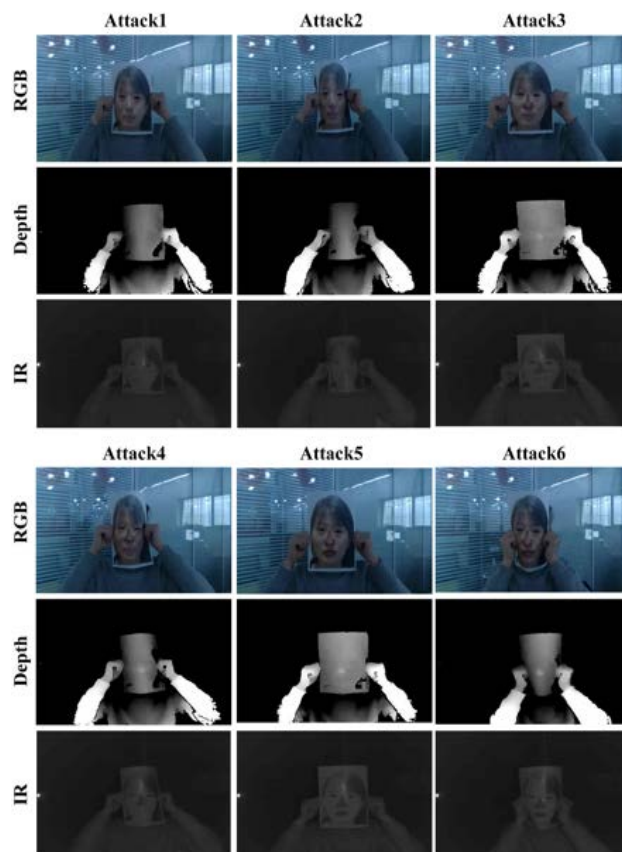


图 8 CASIA-SURF 数据集中的部分样本

**数据集地址:** [http://www.cbsr.ia.ac.cn/users/jwan/database/CASIA-SURF\\_agreement.pdf](http://www.cbsr.ia.ac.cn/users/jwan/database/CASIA-SURF_agreement.pdf)

## 9、CASIA-SURF CeFA 数据集

**介绍:** CASIA-SURF CeFA 数据集由中国科学院自动化研究所 (Institute of Automation, Chinese Academy of Sciences) 生物识别与安全技术研究中心在 2020 年提出。同时，该数据集用于 CVPR2020 上的人脸活体检测竞赛 - Chalearn Multi-modal Cross-ethnicity Face anti-spoofing Recognition Challenge@CVPR2020，吸引了国内外学术界与工业界共 340 多支队伍参与。该数据集是迄今为止最大的跨种族人脸活体检测数据集，涵盖 3 个种族、3 种模态、1607 个被采集者、同时包含 2D 和 3D 攻击类型。

如图 9 所示，前三行为原始视频帧样本，后三行为处理后的样本，处理过程包含人脸检测，去除背景区域，采样等操作。被采集者涵盖 3 个种族，分别为亚洲人、

中亚人与非洲人。每个样本包含 3 个模态，分别为 RGB, Depth 与 IR 模态。为了充分探究该数据集的性能，该课题组定义了 5 种测试协议：第一种协议测试算法在不同种族上的泛化性能(亚洲，中亚与非洲)；第二种协议测试算法对不同攻击类型的泛化性能(打印与视频回放攻击)；第三种协议测试算法对不同模态的泛化性能(RGB, Depth 与 IR 模态)；第四种测试协议测试算法对多种因素同时变化时的泛化性能；第五种协议测试算法对种族因素的无偏性。

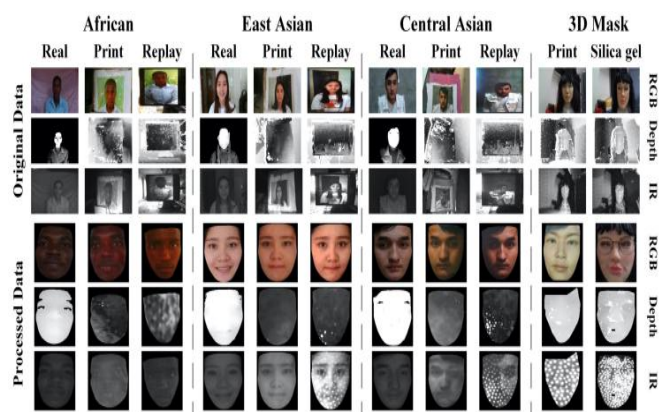


图 9 CASIA-SURF CeFA 数据集中的部分样本

**数据集地址:** [http://www.cbsr.ia.ac.cn/users/jwan/database/CeFA\\_agreement.pdf](http://www.cbsr.ia.ac.cn/users/jwan/database/CeFA_agreement.pdf)

## 10、CASIA-SURF HiFiMask 数据集

**介绍:** CASIA-SURF HiFiMask 数据集由中国科学院自动化研究所(Institute of Automation, Chinese Academy of Sciences)生物识别与安全技术研究中心在 2021 年提出。同时，该数据集用于 ICCV2021 上的人脸活体检测竞赛 – Chalearn 3D High-Fidelity Mask Face Presentation Attack Detection Challenge@ICCV2021，吸引了国内外学术界与工业界共 180 多支队伍参与。

随着目前人脸活体检测算法对 2D 攻击的应对能力，以及 3D 打印技术的成熟，攻击者将目光转向颜色纹理、几何形状与真实人脸更为相近的人脸面具。如攻击者仅需向面具制作公司提供 2 张不同角度的人脸平面图像，

则可获得逼真的人脸面具。因此，基于 3D 面具攻击的人脸活体检测算法越来越受到研究者关注。目前公开的 3D 面具数据集存在以下几点不足，参与采集的人数量较少，面具 ID 与真人 ID 不对应，参与采集的人种族单一，面具材质单一，拍摄场景与光照环境受控，采集设备单一等。因此，收集一个大规模、能弥补上述不足的人脸面具数据集是研究 3D 面具攻击人脸活体检测算法的首要任务。



图 10 CASIA-SURF HiFiMask 数据集中的部分样本

CASIA-SURF HiFiMask 数据集是迄今为止最大的人脸 3D 面具攻击检测数据集，包含 75 个被采集者，其中肤色为黄色、黑色与白色的各 25 人。对于每一位被采集者，共有 3 种材质的面具与其身份对应，分别为塑料透明、树脂与石膏材质。在采集过程中，该课题组设置了 6 种场景以模拟真实的应用场合，分别为白光闪光环境、绿色闪光环境、红绿蓝闪光环境、室外阳光处与阴影处环境、运动模糊环境。在每一种场景中，考虑 6 种不同的光线以提升算法对光线变化的鲁棒性，分别为正常光、暗光、强光、背光、侧光与顶光。在每一种光照方向下，同时采集 7 段视频，分别由 iPhone11, iPhoneX, MI10, P40, S20, Vivo, HJIM 7 个设备完成采集，部分样本如图 10 所示。经过上述步骤，共收集 54600 段视频，其中真实人脸 13650 段，面具人脸 40950 段。基于数据集，该课题组定义了 3 个协议，第一个协议测试算法在“看得见的”因素上的性能，即面具类型、场景种类、光照方向、采集设备在训练集、验证集与测试集中都出现。协议 2 测试算法对“看不见”

的攻击类型的泛化性能，即测试集中的面具材质在训练集与验证集中不存在。第三个协议测试算法对多种不可见因素的泛化性能，即更符合实际的部署环境。

**数据集地址:** <https://sites.google.com/qq.com/face-anti-spoofing/welcome/challengeiccv2021?autouser=0>

### 11、CASIA-SURF 3DMask 数据集

**介绍:** CASIA-SURF 3DMask 数据集由中国科学院自动化研究所 (Institute of Automation, Chinese Academy of Sciences) 生物识别与安全技术研究中心在 2021 年提出。为了评估跨数据集和未知攻击类型的 NAS 可传输性，该课题组发布了一个大规模的 3D mask 数据集，即 CASIA-SURF 3DMask，以支持新的“跨数据集跨类型”测试协议。

具体来说，对于真实人脸视频，包含来自 48 个被采集者的 288 段视频(每个真人 6 段视频)。考虑到现实环境的变化，数据采集采用了正常、背光、前照、侧光、室外阴影和阳光六种条件。为了收集攻击视频，该课题组采用 3D 打印收集了 48 个被采集者的 3D 面具。除了只使用原始的面具，还考虑了两个更真实的装饰情况(即面具具有/没有头发和眼镜)。总共采集了 864 个面具视频(48 名被采集者，有 3 个面具装饰和 6 个环境条件)。

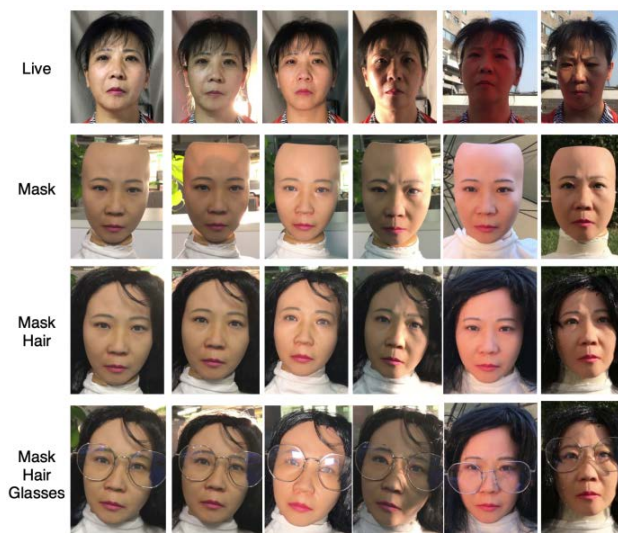


图 11 CASIA-SURF 3DMask 数据集中的部分样本

在 3DMask 数据集中，视频由多个品牌(如苹果、华为和三星)的最新移动设备拍摄而来。每个视频序列持续约 10 秒，帧速率为 30fps，分辨率为 1080p。就年龄分布而言，大多数受试者都在[20,30]和[50,60]岁之间。年龄最小的 23 岁，最大的 62 岁。图 11 显示了一些真实人脸和攻击样本。

**数据集地址:** <http://www.cbsr.ia.ac.cn/users/jwan/database/3DMask.pdf>

责任编辑 李策



## 万 军

副研究员，硕士生导师，在模式识别国家重点实验室生物识别与安全技术研究中心从事科研工作，IEEE Senior Member, Chalearn 理事会成员。研究方向为智能视频分析与理解、智能交互技术(手势/手语)。担任人工智能顶级期刊 IEEE TPAMI 和生物特征识别权威期刊 IEEE TBIOM 客座编辑、SCI 期刊 IET Biometrics 副主编(2020-2022)和 ICME 2021 领域主席，在计算机视觉国际顶级会议 CVPR 2019、CVPR 2020、ICCV 2021 上连续三届举办人脸防伪检测国际研讨会并主导相关防伪竞赛，共吸引超过 800 余支队伍参赛，获得国内外广泛关注。个人主页：<http://www.cbsr.ia.ac.cn/users/jwan/>

## 好文推荐

佛罗里达大学、西澳大利亚大学克劳利分校和长安大学团队的“Deep Affinity Network for Multiple Object Tracking”最新成果发表在 IEEE TPAMI 2021。

论文：Shijie Sun, Naveed Akhtar, HuanSheng Song, Ajmal Mian, Mubarak Shah. Deep Affinity Network for Multiple Object Tracking, IEEE TPAMI, 43(1): 104-119, 2021

多目标跟踪(Multiple Object Tracking, MOT)在视频分析等计算机视觉相关的基础问题中起着重要的作用。大多数 MOT 方法主要两个步骤：目标检测和数据关联。第一步是在视频的每一帧中检测出感兴趣的目标，第二步是在不同帧中建立检测到的目标间的对应关系以获得目标轨迹。近几年，基于深度学习的目标检测方法得到了极大的发展。然而，用于跟踪的数据关联仍然依赖于手工制作的约束条件(如外观、运动、空间邻近性、分组等)，按此计算不同帧中对象间的亲和性。在本文中，团队成员利用深度学习在数据关联跟踪方面强大的特征表达能力，以端到端的方式联合建模不同帧间的目标对象外观及亲和力。具体而言，团队提出了一个深度亲和网络(DeepAffinity Network, DAN)，该网络首先在多个抽象层次上学习预先检测到的目标的特征，然后在任何两帧中对这些特征进行详尽的配对排列以此推断目标间的亲和关系。此外，DAN 还进一步解释了在视频帧间出现和消失的多个物体。团队利用该网络产生的亲和关系，将当前帧中的目标对象深入到前一帧中，最终实现可靠的在线跟踪。

如图 1 所示，团队对视频帧中目标的外观进行建模，并使用所提 DAN 计算它们的跨帧亲和性。DAN 是一个端到端训练的网络结构，主要包括两个组件，即特征提取器 (Feature extractor) 和亲和估计器 (Affinity

estimator)。DAN 训练输入包括视频当前帧 $I_t$ ，当前帧目标中心 $C_t$ ，以及视频前帧 $I_{t-n}$ ，视频前帧目标中心 $C_{t-n}$ 。该网络被用于连续视频帧中跟踪目标，但使用有利于整体方法的非连续帧对其进行训练，从而实现可靠地将给定帧中的目标与之前多个帧中的目标关联起来。DAN 的输入在图 1 中用红色表示，下面将描述该网络的内部组件的结构以及相关的工作、实现细节。

特征提取器：DAN 的第一个主要组件称为特征提取器，该子网络模型用于全面、紧凑的提取视频中目标特征。如图 1 所示，特征提取将视频帧对和目标中心输入通过两个卷积层进行。该网络以常用 VGG 网络结构为基础，并进一步出于对目标空间特征的考量，团队将原本 VGG 的全连接层和 softmax 层替换为卷积层，剩余结构可参考 VGG16 网络。亲和估计器：DAN 这个组件的目标是利用特征提取模块提取的特征对目标对象间的亲和度进行编码。该模块将两个目标的特征矩阵的列沿张量的深度维数相连(涵盖所有可能的排列)，然后利用一个压缩网络将这个张量映射到一个矩阵上。该压缩网络将一个对目标对象特征组合进行编码的张量映射到一个对特征(即对象)之间的相似性进行编码的矩阵。因此，它通过卷积核沿着输入张量的深度逐步降维，不允许特征映射的相邻元素相互影响。

DAN 的部署：DAN 的特征提取器组件是一个双流网络，然而该组件双流部分间的参数是共享，因此团队将其部署成一个单流模型。具体而言，团队先将视频当前帧 $I_t$ 和当前帧目标中心 $C_t$ 输入到网络中，以计算该帧及当中目标中心特征，然后将其传递给亲和估计器。亲和估计器使用前一帧的特征矩阵，计算每帧视频帧对的变异张量。接下来，通过一个简单的前向传播网络和一个连接操作将所得张量映射到一个亲和矩阵，如上所述。每一帧只通过一次目标检测器和特征提取器，但特征被

循环多次应用于与其他多帧成对的亲和力的计算。DAN 多目标跟踪：团队存储帧的特征矩阵及其时间戳以将当前帧中的目标对象与多个先前帧中的相关目标对象关联起来。第 0 帧之后，团队初始化 DAN 网络并初次得到计算结果，之后即可按照 DAN 的部署中的描述使用特征提取器提取的特征矩阵计算当前帧和任意先前帧中目标对象间的亲和关系。

团队使用 Pytorch 框架实现 DAN 网络结构，并在 NVIDIA GeForce GTX Titan GPU 上执行。利用 MOT17 数据集对 DAN 的超参数进行了优化，超参数具体设定为，batch size 设为 8，训练次数为 120，每帧最大目标数设为 80，图片输入尺寸为  $900 \times 900$ 。团队使用 SGD 优化器训练 DAN，动量和权重衰减参数分别设置为 0.9 和  $5e^{-4}$ ，初始学习率设为 0.01，在第 50、80 和 100 训练轮次时，学习速率下降到先前值的 1/10。

团队使用的 MOT17 数据集共包括 7 个场景视频。这些视频数据集场景在背景、照明条件和摄像机视角方面存在很大的差异，这些差异变化在很大程度上增加了检测的挑战性。团队进一步分析了这 7 个场景数据集的主要属性，除场景变化外，这些数据集在分辨率、摄像机帧率、每帧对象的平均数量(即密度)和轨迹总数也存

在明显的变化。例如：scene 05 具有与其他场景明显不同的分辨率，场景 07 使用较低角度拍摄，这导致了明显的遮挡。这 7 个场景数据集的编号名称分别为：02, 04, 05, 09, 10, 11, 13。此外，团队为保证实验结果的可靠、完善性，还选取了 MOT15 和 UA-DETRAC 数据集。其中，MOT15 数据集与 MOT17 数据集类似，但 MOT15 发布时间较早且具有更多的视频场景数据。UA-DETRAC 数据集包含 100 个视频，记录了大约 10 小时的车辆交通情况。这段视频是在 24 个不同的地点录制的，录制内容包括各种常见的车辆类型和交通状况，录制场景包括城市公路、十字路口、丁字路口等。简而言之，该数据集包含约 140k 左右的视频帧、8250 辆汽车和 1210k 边界框。与 MOT 挑战类似，UA-DETRAC 挑战接受跟踪方法的提交，主机服务器在单独的测试数据上评估它们的性能。

团队共选取 12 个评价指标，分别为：MOTA、MOTAL、MOTP、Rcll、IDF1、MT、ML、FP、FN、ID\_Sw、Frag、Hz。大量实验结果表明，本文算法在大部分指标上的表现均优于现阶段主流的多目标跟踪算法，此外，团队还提供了可视化的检测对比结果，可视化结果进一步表明了本文所提算法的有效性。

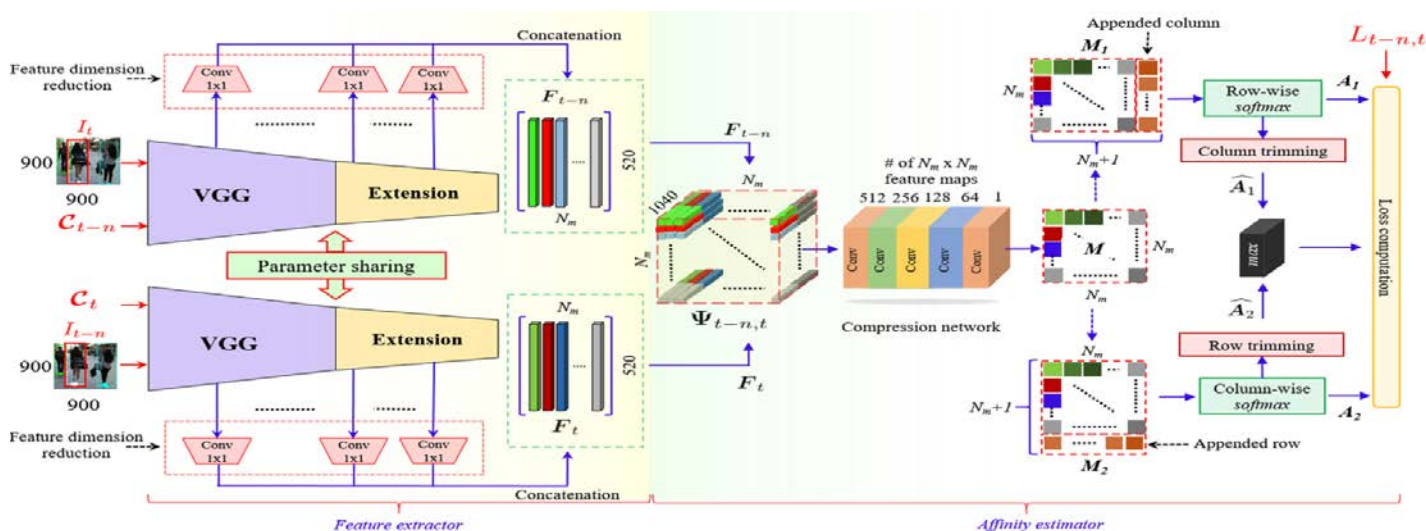


图 1 DeepAffinity Network 方法流程图

责任编辑 贾同 樊鑫

## 好文推荐

**密** 歇根州立大学团队的“On Learning 3D Face Morphable Model from In-the-Wild Images”最新成果发表在 IEEE TPAMI2021。

论文: Luan Tran, Xiaoming Liu. On Learning 3D Face Morphable Model from In-the-Wild Images, IEEE TPAMI, 43(1): 157-171, 2021

三维形变模型(3D Morphable model, 3DMM)是一种经典的三维面部形状和反照率统计模型,广泛应用于面部分析等领域(例如:模型拟合、图像合成)。传统3DMM模型是通过学习一组三维人脸扫描和相关的具有良好特性的二维人脸图像,并借助两组PCA基函数获得。然而,传统3DMM的表达能力受限于训练数据的类型、数量和线性基。针对上述问题,本文提出一个新框架,该框架将从大量野外人脸图像中学习非线性的3DMM模型,同时不需要收集3D人脸扫描。具体来说,给定一张人脸图像作为输入,网络编码器估计投影、光照、形状和反照率参数。两个解码器作为非线性3DMM,分别将形状和反照率参数映射到三维形状和反照率。整个网络结构是端到端且为弱监督训练的网络模型。网络实现代码及其补充实验结果可见网络链接:<http://cvlab.cse.msu.edu/project-nonlinear-3dmm.html>。

如图1所示,团队使用两个深度网络将形状、反照

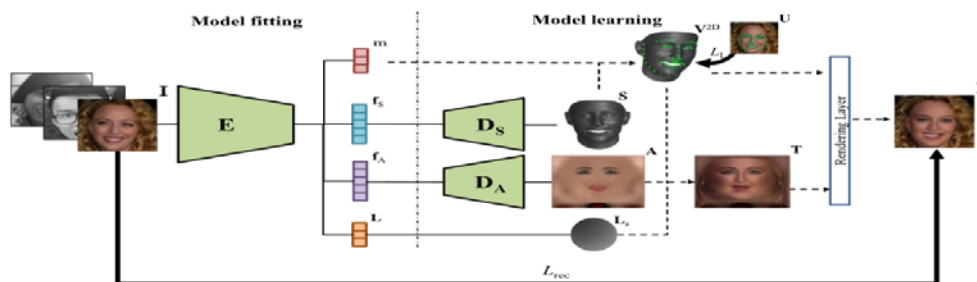


图1 DECAMEL 算法流程图

率参数分别解码为三维面部形状和反照率。为保证框架端到端可训练,上述参数由编码器网络估计,本质上是拟合所提3DMM模型的模型参数。基于物理的渲染层,模型将一个编码器和两个解码器这三个深度网络联合起来,最终实现重建输入人脸图像的目标。具体而言,给定一张2D人脸图像,模型中的编码器 $E$ 旨在估计映射 $m$ 、光线参数 $L$ 、形状参数 $f_s$ 以及反照率参数 $f_A$ 。3D形状解码器 $D_S$ 用于将形状参数 $f_s$ 解码到3D形状 $S$ ,反照率解码器则用于将反照率参数 $f_A$ 解码到真实三维世界的反照参数 $A$ 。所提3DMM模型旨在利用 $m$ 、 $L$ 、 $S$ 和 $A$ 所绘制图像 $\hat{I}$ 近似于原始图像 $I$ ,即遵循公式:

$$\operatorname{argmin}_{E, D_S, D_A} \sum_{i=1}^K \|\hat{I}_i - I_i\|_1,$$

其中绘制图像可表达为:

$$\hat{I} = R(E_m(I), E_L(I), D_S(E_S(I)), D_A(E_A(I))),$$

公式中,  $R(m, L, S, A)$ 是渲染层网络操作。

实验从模型表达能力、表征能力以及面部分析三个方面对所提出的非线性3DMM模型进行了研究。大量实验结果表明本文所提算法具有良好的表达、表征及分析性能。

责任编辑 樊鑫 贾同

## 好文推荐

香港中文大学和思谋科技团队在“面向三维目标检测的深度立体网络”方面的最新研究成果发表在 IEEE/CVF CVPR 2020。

论文: Yilun Chen, Shu Liu, Xiaoyong Shen, Jiaya Jia, DSGN: Deep Stereo Geometry Network for 3D Object Detection, IEEE/CVF CVPR, 2020, pp. 12536-12545

在 3D 目标检测领域, 根据特征的代表方法不同, 3D 目标检测器主要分为基于图像的 3D 检测器和基于激光雷达的 3D 检测器。当前大多数最先进的 3D 目标探测器严重依赖于激光雷达传感器。激光雷达传感器通过发射多束激光捕捉准确的 3D 信息, 具有强大的 3D 信息准确检索能力, 由此出现了各种利用激光雷达生成点云的 3D 目标探测器, 其不足之处在于雷达设备笨重且价格昂贵, 且得到的是带有激光束的稀疏分辨率的数据。相比之下, 视频摄像机要便宜很多, 并且能够产生更加稠密的分辨率, 这种基于摄像机拍摄图像的 3D 目标检测方法, 通过立体对应估计预测视差从而在立体图像上计算场景深度。虽然目前有几种基于单目和立体视觉的 3D 检测方法都在突破基于图像的 3D 目标检测的极限, 但与基于激光雷达的方法相比, 精度仍有很大差距。这种基于图像的方法最大的挑战是如何为 3D 对象提供适当和有效的表示。

为此, 香港中文大学和思谋科技团队提出了一种端到端的 3D 物体检测深度立体几何网络(DSGN), 该检测框架的核心在于通过空间变换将 2D 特征转换成有效的 3D 结构, 该结构被称为 3D 几何体(3DGV), 如图 1 所示。该模型在平面扫描体中建立立体对应关系, 然后将其转换为 3DGV, 从而能够对 3D 几何和语义线索进行编码, 以便在 3D 规则空间中进行预测。3D 几何体积定义于 3D 世界空间中, 由在摄像机截锥中构造的平面扫描体积(PSV)转换而来。网络在 PSV 中可以很好

地学习像素对应约束, 而在 3DGV 中可以学习真实物体的 3D 特征。这种体积构造是完全可微的, 因此可以共同优化以学习立体匹配和目标检测。该团队提出的 3DGV 面向 3D 真实对象很容易施加像素对应约束以及编码完整的深度信息, 同时提供了具有几何信息的 3D 表示, 这使得学习现实世界对象的 3D 几何特征成为可能。作者设计了一个端到端的框架, 以提取用于立体匹配的像素级特征和用于目标识别的高级特征, 如图 2 所示。所提出的 DSGN 能同时估计场景深度并检测 3D 目标, 从而实现多种实际应用。

该方法可用于 3D 场景深度估计和 3D 物体检测, 具有广泛的实际应用价值。在排除背景中花哨装饰的条件下, 这种简单且完全可微的网络优于过去基于立体视觉的 3D 探测器, 在 KITTI 3D 物体检测排行榜上与几种基于激光雷达的方法具有相当的性能。

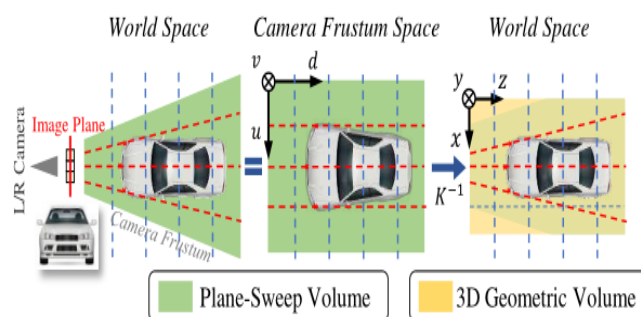


图 1 空间变换示意图

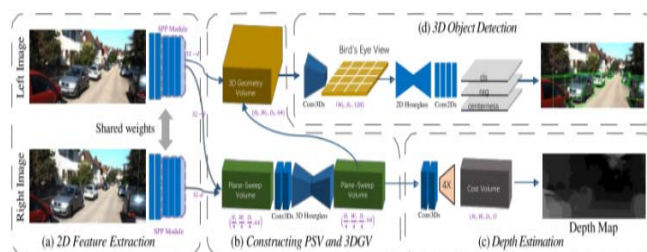


图 2 深度立体几何网络 DSGN 结构图

责任编辑 沈沛意

## 新加坡国立大学冯佳时教授团队



Learning & Vision Lab

**新加坡国立大学** LV 实验室现由冯佳时教授领导, 聚焦于机器学习相关的重要

问题, 包括自监督学习、新型网络架构设计, 以及多模态机器学习等。

### 团队发展历史

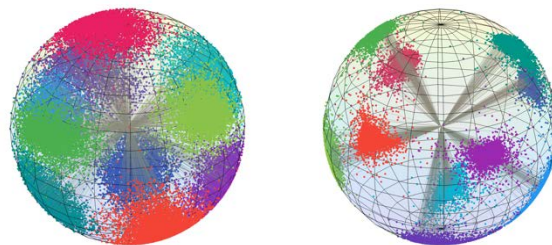
LV 实验室由颜水成教授于 2007 年创立。自 2015 年起, 实验室由冯佳时教授领导, 颜水成教授兼任顾问。自成立以来, LV 团队一直致力于超越已有的知识边界, 并鼓励针对计算机视觉、机器学习以及相关技术在现实生活中的应用开展最前沿、最具原创性的科学研究。团队的研究项目范围从机器学习 (包括深度学习) 的基础研究到尖端的人脸/人和图像/视频分析技术, 再到智能搜索和推荐系统。

在过去的十年中, LV 团队不断壮大, 取得了令人瞩目的成就。迄今为止, LV 实验室的成员们在国际顶级学术期刊和会议上发表了 600 多篇学术论文, 其中 10 余篇获得了最佳论文奖或者最佳学生论文奖。他们在计算机视觉核心竞赛 (即 Pascal VOC 和 ImageNet) 上获得了十多次获胜奖或者荣誉奖。他们还获得了 3 个工业许可, 充分展示了实验室研究的巨大商业潜力。特别的是, 这支优秀的团队在 ACM MM (多媒体最佳会议) 上赢得了大满贯, 包括最佳论文奖, 最佳学生论文奖和

最佳演示奖。目前, LV 组有 24 名博士研究生, 3 名博士后 (统计信息可能会随时间而变化)。截至 2021 年 3 月, 已有超过 110 位校友在 LV 实验室工作, 包括 21 位博士生, 43 位博士后/科研助理和 30 多位访问教授和学生。他们目前在高校/公司任职, 继续为学术界和社会做出贡献。一直以来, LV 团队致力于将目光投向那些尚未被探索或尚未被解决的问题, 这些问题的解决可以从根本上推动计算机视觉和机器学习技术的发展, 从而可以进一步改善人们的生活方式。

### 研究方向 1: 自监督学习

近年来, 自监督学习在学术界和工业界获得了广泛关注。相比于传统的有监督学习, 自监督学习的优势在于可以利用海量的无标记数据学到具有好的特征表达的模型。



(a) Training with  $\mathcal{L}_{ce}$  (b) Training with  $\mathcal{L}_{ce} + \mathcal{L}_{con}$

图 1 在 CIFAR10 验证集上基于 ResNet-18 的特征可视化。(a). 交叉熵损失; (b). 对比正则化损失。

现存的基于自监督学习研究通常聚焦于如何提升模

型的表征能力，却忽略了如何在下游任务中更好地利用这些预训练模型。当前的普遍做法是直接利用交叉熵损失函数对模型进行微调，但这样微调得到的模型无法解决预训练模型所带来的类内特征分散使得模型判别能力有限的问题（如图 1 所示）。针对这一问题，冯佳时教授团队发现在模型微调阶段引入对比学习，能够进一步正则化特征空间并促进模型优化，因此提出了基于对比正则化的微调算法<sup>[1]</sup>。实验结果表明，所提出的微调算法可以在图像分类、语义分割等任务上有效提升自监督模型的迁移性能，并在领域泛化和对抗鲁棒等任务上充分提升了模型的泛化能力和鲁棒性。

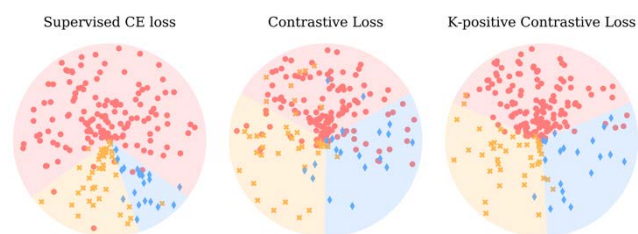


图 2 在类别不均衡数据上用不同损失函数学到的特征对比（示意图）。左：有监督的交叉熵损失；中：无监督的对比损失；右：基于 k-正样本的对比损失。

除了探索如何使用自监督模型，冯佳时教授团队进一步从特征空间的角度研究了基于对比的自监督学习。通过比较有监督的交叉熵损失和无监督的对比损失在类别不均衡数据上的表现，冯佳时教授团队发现：优化有监督的交叉熵损失虽然可以得到语义判别性足够的特征空间，但同时会导致特征空间偏向于样本数量占主导的类别；而优化无监督的对比损失可以得到一个均衡但判别性很弱的特征空间。基于以上发现，冯佳时教授团队提出了有监督的基于 k-正样本的对比损失<sup>[5]</sup>。通过优化提出的损失函数，模型可以学到均衡且判别性良好的特征空间（如图 2 所示）。实验结果表明，得益于特征空间同时具有足够的均衡性和判别性，所提出的损失函数在长尾识别任务上取得了领先的结果，并且可以作为新的表示学习方法，提供泛化能力更强的特征表达。

除了表示学习，冯佳时教授团队还在其他领域中探

索了自监督学习。最近的几个成果包括：用自监督学习帮助无监督域适应中目标域的特征学习<sup>[4]</sup>；在半监督 3D 行为识别中，用自监督学习来保证有标签数据和无标签数据的特征一致性<sup>[3]</sup>；将自监督学习用于 3D 人体姿态识别任务的测试阶段优化，从而提高模型对测试样本的表现<sup>[2]</sup>。

## 研究方向 2：新型网络架构设计

近年来，由于结构强大的神经网络是深度学习能取得巨大成功的重要因素。基础网络结构的设计一直是 LV 团队重点聚焦的方向。早在 2013 年，LV 团队就提出了  $1 \times 1$  卷积操作<sup>[15]</sup>，后来被研究者们广泛应用，成为神经网络的标准算子之一。冯佳时教授团队对基础网络结构的多个部件都进行了探索，设计了多种高效的新型网络结构。其中，DPN<sup>[16]</sup>兼具 ResNet 和 DenseNet 的优势，帮助 LV 团队取得了最后一届 ILSVRC 比赛物体定位任务的冠军； $A^2$ -Net<sup>[17]</sup>和 CoordAttention<sup>[9]</sup>提出了更高效的注意力机制；OctConv<sup>[14]</sup>提出了更高效的卷积算子，SCNet<sup>[12]</sup>提出了自我校正的卷积操作；SPNet<sup>[13]</sup>提出了独特的特征池化方法；MobileNeXt<sup>[6]</sup>重新思考了针对移动端的卷积网络设计，并且提出一种更高效的瓶颈结构。

除了对人工设计网络的探索，冯佳时教授团队还对自动设计神经网络结构的技术即神经架构搜索（NAS）开展了广泛的研究。其中，WSNet<sup>[18]</sup>提出一种新的基于权值采样的网络训练方法，天然地保证了网络内部的参数共享，从而得到紧凑而高效的小网络；团队还首次提出网络结构和网络参数分离设计的概念<sup>[7]</sup>。在神经网络图拓扑固定的情况下，NES<sup>[7]</sup>算法通过自动学习一个网络参数的缩影以及相应的映射函数，可以使用少数参数充填一个大网络结构，以得到更高效的神经网络结构。此外，冯佳时教授团队发现，近期 NAS 工作的搜索空间大多基于 MobileNet-V2 里面提出的逆残差模块的衍生结构，为此需要较强的人为干预，在一定程度上限

制了 NAS 算法发现新模块结构的能力。基于以上观察，冯佳时教授团队提出了 AutoSpace<sup>[10]</sup>，利用改进后的评估和进化算法，在最小化人为干预的情况下，自动生成神经网络算法所需要的搜索空间（如图 3 所示）。

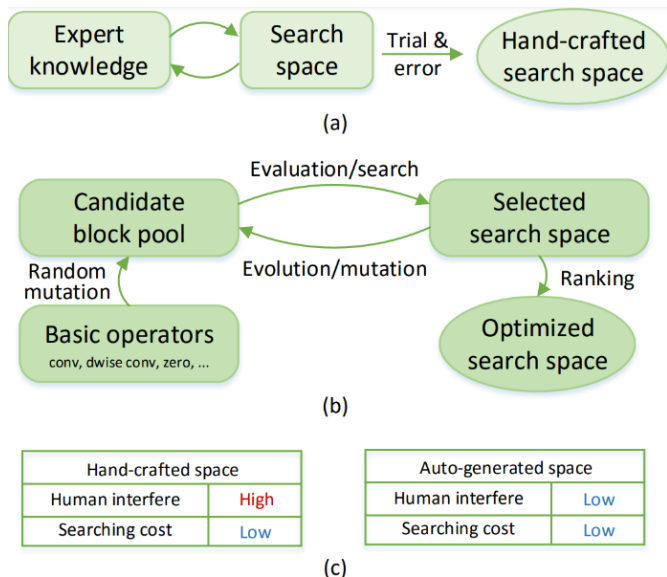


图 3 传统 NAS 搜索算法和 AutoSpace 搜索算法流程对比。(a)表示传统 NAS 算法的流程。由相关专家手工设计搜索空间。(b)表示 AutoSpace 的算法流程。搜索算法所需的搜索空间将基于提前定义好的基本操作算子，自动生成。

除了对于卷积网络架构的研究，冯佳时教授团队也对 Transformer 模型的设计开展了多项研究。其中，ConvBERT<sup>[8]</sup>提出用跨度动态卷积以改进现有注意力机制，从而提升 BERT 模型的性能。近来，Transformer 模型被引入到计算机视觉领域，在多个视觉任务上取得了媲美甚至超过卷积神经网络的表现。冯佳时教授团队也对视觉 Transformer 模型的结构设计开展了研究。Vision Transformer (ViT) 缺少类似卷积网络中的归纳偏置，需要依赖超大规模数据才能取得合理的性能，而直接在 ImageNet 上训练时，其性能仍然差于同级别的卷积神经网络，极大限制了 Transformer 在视觉任务中的应用。为此，冯佳时教授团队提出了 T2T-ViT 模型，该模型中的 T2T 在对图像结构信息建模的同时也可以逐步减少图像 tokens 的长度以节省计算量。另一方面，

T2T-ViT 模型融合了通用的卷积网络中的结构设计，主体框架采用了深窄结构，在减少冗余的注意力结构的同时提高了模型的性能。T2T-ViT 模型可以直接在 ImageNet 上训练，同时超过通用的卷积神经网络如 ResNet 的效果（如图 4 所示），或者和轻量级卷积神经网络如 MobileNet 达到同等效果。此工作在不使用超大规模额外数据的前提下，第一次真正意义上使用纯 Transformer 网络超过了卷积神经网络。

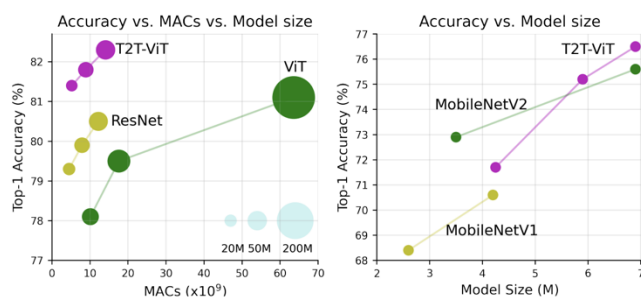


图 4 T2T-ViT 与 ResNet、MobileNet 的性能对比

与此同时，冯佳时教授团队发现，现有的 Transformer 模块设计并不能很好地拓展到更深的网络。所提出的 DeepViT<sup>[11]</sup> 用实验表明，与卷积神经网络不同，在训练超参相同的情况下，当 Transformer 的模块超 24 个的时候，网络的分类准确度不再随着网络深度的增加而增加。冯佳时教授团队提出了 Re-attention 机制，通过增加不同注意力头之间的信息交流，重新生成每个模块的注意力特征图，进而保证了 ViT 随着网络深度增加，性能也能提升（如图 5 所示）。

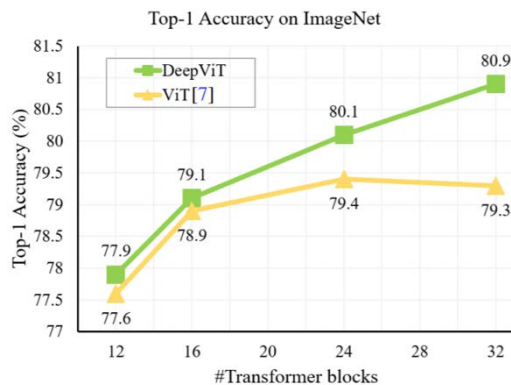


图 5 Vision Transformer 性能随网络深度的变化

责任编辑 张汗灵

## 参考文献

- [1] Zhang Y, Hooi B, Hu D, Liang J., Feng J. Unleashing the Power of Contrastive Self-Supervised Visual Models via Contrast-Regularized Fine-Tuning. arXiv preprint arXiv:2102.06605, 2021.
- [2] Zhang J., Nie X., Feng J. Inference Stage Optimization for Cross-scenario 3D Human Pose Estimation. Advances in Neural Information Processing Systems (NeurIPS), 2020.
- [3] Si C., Nie X., Wang W., Wang L., Tan T., Feng J. Adversarial Self-Supervised Learning for Semi-Supervised 3D Action Recognition. European Conference on Computer Vision (ECCV), 2020.
- [4] Liang J., Hu D., Wang Y., He R., Feng J. Source Data-absent Unsupervised Domain Adaptation through Hypothesis Transfer and Labeling Transfer. arXiv preprint arXiv:2012.07297, 2020.
- [5] Kang B., Li Y., Xie S., Yuan Z., Feng J. Exploring Balanced Feature Spaces for Representation Learning. International Conference on Learning Representations (ICLR), 2021.
- [6] Zhou D, Hou Q, Chen Y, Feng J, Yan S. Rethinking Bottleneck Structure for Efficient Mobile Network Design. European Conference on Computer Vision (ECCV), 2020.
- [7] Zhou D, Jin X, Hou Q, Wang K, Yang J, Feng J. Neural Epitome Search for Architecture-agnostic Network Compression. International Conference on Learning Representations (ICLR), 2020.
- [8] Jiang Z, Yu W, Zhou D, Chen Y, Feng J, Yan S. ConvBERT: Improving Bert with Span-based Dynamic Convolution. Advances in Neural Information Processing Systems (NeurIPS), 2020.
- [9] Hou Q, Zhou D, Feng J. Coordinate Attention for Efficient Mobile Network Design. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [10] Zhou D, Jin X, Lian X, Yang L, Xue Y, Hou Q, Feng J. AutoSpace: Neural Architecture Search with Less Human Interference. arXiv preprint arXiv:2103.11833, 2021.
- [11] Zhou D, Kang B, Jin X, Yang L, Lian X, Hou Q, Feng J. DeepViT: Towards Deeper Vision Transformer. arXiv preprint arXiv:2103.11886, 2021.
- [12] Liu J, Hou Q, Cheng M, Wang C, Feng J. Improving Convolutional Networks with Self-calibrated Convolutions. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [13] Hou Q, Zhang L, Cheng M, Feng J. Strip Pooling: Rethinking Spatial Pooling for Scene Parsing. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [14] Chen Y, Fan H, Xu B, Yan Z, Kalantidis Y, Rohrbach M, Yan S, Feng J. Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution. IEEE International Conference on Computer Vision (ICCV), 2019.
- [15] Lin M, Chen Q, Yan S. Network in Network. International Conference on Learning Representations (ICLR), 2014.
- [16] Chen Y, Li J, Xiao H, Jin X, Yan S, Feng J. Dual Path Networks. NeurIPS, 2017.
- [17] Chen Y, Kalantidis Y, Li J, Yan S, Feng J.  $A^2$ -Nets: Double Attention Networks. NeurIPS, 2018.
- [18] Jin X, Yang Y, Xu N, Yang J, Jojic N, Feng J, Yan S. WSNet: Compact and Efficient Networks through Weight Sampling. International Conference on Machine Learning (ICML), 2018.
- [19] Yuan L, Chen Y, Wang T, Yu W, Shi Y, Jiang Z, Tay F, Feng J, Yan S. Tokens-to-Token ViT: Training Vision Transformers from Scratch on Imagenet. arXiv preprint arXiv:2101.11986, 2021.



## 冯佳时

2007年在中国科学技术大学获得学士学位，2014年在新加坡国立大学获博士学位。2014至2015年在加州大学伯克利分校从事博士后研究。2015年加入新加坡国立大学电子与计算机工程学院担任助理教授。主要研究领域为计算机视觉和机器学习。获2018年度ACM MM最佳学生论文奖。多次担任ACM MM、NeurIPS、ICLR、ICML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 Pattern Recognition Letters, Image and Vision Computing, World Wide Web, IEEE Transactions on Intelligent Transportation Systems, Pattern Recognition Letters 和 IEEE Transactions on Multimedia。

## 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。

第四届 PRCV 将于 2021 年 10 月 29 日至 11 月 1 日在北京国际会议中心举行，由北京科技大学、北京交通大学和北京邮电大学共同承办，中山大学、清华大学协办。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
MM ASIA 2021	2021.12.1-3	Gold Coast, Australia	2021.07.19	<a href="https://mmasia2021.uqcloud.net/important-dates/">https://mmasia2021.uqcloud.net/important-dates/</a>
AAAI 2022	2022.2.22-3.1	Vancouver, Canada	2021.09.09	<a href="https://aaai.org/Conferences/AAAI-22/">https://aaai.org/Conferences/AAAI-22/</a>

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
PRL	Few-shot Learning for Human-machine Interactions (FSL-HMI)	<a href="https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/few-shot-learning-for-human-machine-interactions-fsl-hmi">https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/few-shot-learning-for-human-machine-interactions-fsl-hmi</a>	2021.07.20
IVC	Deep Learning for Panoramic Vision on Mobile Devices	<a href="https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/si-deep-learning-for-panoramic-vision-on-mobile-devices">https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/si-deep-learning-for-panoramic-vision-on-mobile-devices</a>	2021.08.01
IEEE T-ITS	Responsible Artificial Intelligence for Autonomous Driving	<a href="https://mc.manuscriptcentral.com/t-its">https://mc.manuscriptcentral.com/t-its</a>	2021.08.15
WWW	Synthetic Media on the Web	<a href="https://www.springer.com/journal/11280/updates/18575890">https://www.springer.com/journal/11280/updates/18575890</a>	2021.09.01
IEEE T-MM	Weakly Supervised Learning for Image and Video Understanding	<a href="http://mc.manuscriptcentral.com/tmm-ieee">http://mc.manuscriptcentral.com/tmm-ieee</a>	2021.09.20

## 心底无私视界宽 ∞ 吴立德教授专访

自 50 年代以来,我国在计算机视觉领域展开了相关的科研工作。而今,我国已经拥有了一支庞大的、在这一领域辛勤耕耘且能与世界一流水平并驾齐驱的科研队伍。在这一过程中,有一批见证了视觉领域的发展,为我国计算机视觉领域的奠基做出了重大贡献的先驱。

《视界专访》栏目希望通过对计算机视觉研究历史、进展的见证者作一个系列专访,以帮助从事计算机视觉及相关领域的科研工作者或爱好者,全方面地了解 50 年代以来信息技术、信号处理技术以及计算机视觉相关的一些历史发展及进步,也希望能帮助我们在见证这段历史的同时,展望计算机视觉领域的未来。

本期专访的是复旦大学计算机科学技术学院的吴立德教授。复旦大学计算机科学技术学院曾于 1956 年研制成功我国第一台电子模拟计算机“复旦 601 型电子积分机”,并于 1964 年研制成功我国第一批真正意义上的数字电子计算机“602 电子计算机”。吴立德教授 1982 年就在模式识别顶级期刊 IEEE TPAMI 上发表了中国大陆第一篇 TPAMI 论文,并在 1984、1990 又相继在此顶级期刊上发表论文两篇,期后在计算机视觉领域也做出了大量重要的研究成果。

2021 年 2 月 26 日元宵节,《CCF-CV 专委简报》委托复旦大学黄萱菁教授、吴祖煊副研究员等专访了复旦大学计算机科学技术学院吴立德教授。为能保持本次访谈的**原汁原味**,我们采用问答的形式,整



图 1 复旦大学吴立德教授

理了吴教授的谈话实录。

**问:** 吴老师好!我们了解到,吴老师您从 1958 年复旦毕业之后,就直接留校任教了,我们想请您和我们分享一下您的求学历程,包括在复旦求学的时候,那时计算机可能还不是很普及,那当时您是怎么会决定转去计算

机学院的呢?

**吴老师:** 我是 1955 年进复旦的, 60 年毕业。那个时候需要一部分学生提前毕业留校, 所以我留校了, 研究概率统计。开始最先接触的项目来自上海水文地质大队, 他们有些数学问题, 就找我们数学专业的解决。一接触到实际, 就发现项目需求跟我们学校里的研究确实是不一样的, 需要面对大量数据, 然后根据大量数据做一些统计和分析, 就在那个时候我接触了计算机。1975 年, 学校成立计算机系, 那么我们就从数学系过来。整个计算机系大概三分之二的教授是从数学系过来的, 三分之一是从物理系过来的。所以, 真正有计算机系大概 75 年, 但是在这之前我们做了一些数学的应用, 都要用到计算机的。我们比较成功的就是做了这两件事情, 一件是上海的地面沉降问题, 还有一件是用地震方法探测石油的问题。这个就比 CV 要早了, 当时还算是属于信号处理, 还没到图像 (image)。

**问:** 吴老师您 07 年就退休了, 但是一直工作到了 2016 年。请问是哪些原因使您一直工作下去?

**吴老师:** 一个就是一种习惯。07 年以前一直都有事情做, 到 07 年虽然已经退休, 但是学生还没有完全毕业, 还要等两年才完全毕业。另外一方面, 我对研究还是很感兴趣的, 一下子停下来很不习惯, 还想参加一些工作。所以实验室研究生的讨论班我也一直在参加。我觉得就是一种惯性, 也有很浓的兴趣。当然, 我也想退休以后还能做一些力所能及的事情。

**问:** 下一个问题就是您除了参与科学工作还长期从事教学工作, 能不能请您谈一谈您在教学方面的心得体会?

**吴老师:** 教学和科研实际上是相辅相成的。我们的老师, 苏步青先生曾说过: 科研和教学一起做是最好的。如果专门做科研, 科研有时候做得顺利, 有时候做得不顺利, 那您可能很长一段时间做不出结果出来, 人就会很懊丧。教学的话, 你就是在讲课。所以你两方面做, 既做教学



图 2 复旦大学吴立德教授及夫人吴蔼成教授

又做科研, 科研如果有点心得的话也可以放到教学里面。然后做教学的话, 你也会对你过去学的东西有些更深入的了解。有时候你读书, 对于一些问题并不是很了解, 或者说不是很透, 然后你教的时候就能弄得清楚点, 不然学生一问就把你问倒了。所以说教学和科研实际上摆在一起是最好的。

**问:** 您在计算机视觉和自然语言处理两个领域都颇有建树, 而且我们现在学术领域的话有很多工作, 两个领域之间的方法都可以互相借鉴。然后您是如何看待 CV 和自然语言处理这两个领域的关系, 你认为他们两个领域之间有没有什么共通之处?

**吴老师:** 最初的时候是没有这个感觉的, 我最初在数学系的时候做的是信号处理 (signal processing), 就是一维的信号, 但是就已经是搞信号处理了。就是我把信号变成一个序列, 然后用计算机来处理。那么后来就从一维做到二维, 就变成图像 (image)。自 70 年代开始, 我就从信号处理角度来做图像处理了。图像处理就做到了视觉, 和图像有关的还做了模式识别。

后来就是 MIT 有个比较有名的教授叫马尔 (D. Marr), 他出了一本《视觉计算理论》, 那么那个时候就看到他那个东西。当时他这个东西还是比较前沿的, 也觉得比较好, 就转到了计算机视觉去。所以本来是做

信号处理，然后做图像处理，之后再是模式识别。那做模式识别时碰到的问题是，把一个图像变成特征向量，即所谓特征提取。第二部分就根据这个特征向量来把它分类，一般都这样做。后来，做模式识别的科研人员都发觉，实际上难是难在特征。特征 (feature) 很难处理，feature 对后面的分类非常重要。那么马尔的那一套视觉的东西，他提出了一个比较统一的框架。他那个体系也有一些生物学背景，当时就感觉他那个理论很新鲜。

后来这个视觉做起来也挺难的，因为当时做视觉就有一个设备的问题，当时的计算机还是很差劲，就 286 之类的，存储一个图像进去都很难弄。稍微分辨率高的话，那个图像都很难处理。那个时候从人工智能的角度来想，觉得这个自然语言处理比视觉更靠近智能。因为实际上，好像某种意义上讲，人的智能最基本的就是语言。从人工智能的角度讲，这个东西也更本质一点。另外，当时觉得语言是一维信号，对计算机算力的要求也更低一点，那么我们学校的情况好像也好一点，所以也开始做这个。而图像和语言这两者之间好像是最近深度学习出来以后才有联系，以前的话这两个都是分开的，各自都有自己的一套方法在做。当然深度学习 (deep learning) 是 CV 方面先出来。之前有一篇文章我在讨论班里有讲过，NLP from Scratch。就是把自然语言处理里面的分词也好，词性标注也好，完全用深度学习的方法重做一遍，而且做出来比过去用自然语言方法做，出来的结果更好。从那以后，这两者就都从深度学习来做。你从数学上来看其实就更加简单，实际上就是做最小二乘方程的曲线拟合，或者说是做一个曲面拟合，多维的曲面拟合。我有一大批输入数据和输出数据，然后要拟合一个函数出来。无非就是这个曲面的话，不是过去的幂级数或者傅里叶三级级数。所以你如果从数学角度看，那就是曲面拟合的事情了，但是你用的基函数，是一个神经网络，所以这个联系是从方法上引起的。

我有一年去 MIT 开会，大概在二十年前，那时候参

观他们的一个项目。那个项目是打造智能会议室，那个会议室装了很多麦克风和摄像头，比如说五六个人在里面开会，随便讨论。然后里面的摄像头就记录了好多的视频，同时有很多录音机，录了好多音频。那么他就是搞一个系统，能够自动分析，把会议总结出一些东西。我看他们后来也没做成，因为那个东西太复杂了。那个很超前，他们也雄心勃勃地想做这么一个东西。现在应该这个条件越来越好了，算力越来越好，将来可以做成这个。能够在二十年前提出这个，已经很了不起。



图3 专访合影。吴祖焯、黄萱菁、吴立德、吴蔼成、周雅倩、刘勤

**问：**现在就是高校和企业都在人工智能上面展开了很多的研究，而且企业他们的计算资源可能比高校更丰富，（那么对于高校研究人员而言，如何应对这一状况呢）？

**吴老师：**对，现在腾讯、阿里、讯飞，他们的算力都比我们强。不过我们学校也不错。确实是存在这个问题的。但是像 MIT 之类的学校，在美国很早就出现这种情况了。大公司一定比高校强，但是 Deep learning 图灵奖的三位得主都是来自高校的，Hinton 是多伦多大学的，Bengio 是蒙特利尔大学的，LeCun 是纽约大学的。这些学校的设备一定也不如美国大公司，他们也不容易，这三位搞神经网络几十年了，九十年代就在研究，还是要思考怎么做研究，要比企业更超前。2000 年我在 MIT

开学术会议的时候，他们就在研究这些了。学校怎么能够探索更前一步的地方，这要仔细的思考。企业用的算法没什么厉害的，只要算力大就能有不错的效果。Deep learning 就是曲面逼近，变成优化问题，用的就只是最速下降法，搞优化理论的人看起来那个是最基本的算法，都是一阶算法。谷歌里有一群人认为，只要计算能力越来越强，数据越来越多，很多问题就都可以解决。自然语言处理(Natural Language Processing, 简称NLP)方面，就有人分析过这个工作，几十年NLP技术的进步最主要的两个因素，一个是算力提高了，还有一个是标注数据。计算机能力强了，标注的规模大了，这是主要的因素。至于分词算法，词性标注的算法，进展没有很大。CV方面也有类似的观点。现在增加了无监督学习，不要标注的语料，不仅仅标注语料很重要，没有标注的语料也很重要。CV中的ImageNet, 李飞飞成名作，到目前还是最大的数据集。

**问：**目前学界的工作成果越来越丰富，也吸引人越来越多的研究者参与，但是要做出真正有价值的工作是很难的，您认为要做出有价值、影响力的工作，需要做出哪些努力呢？

**吴老师：**很多有价值的工作都是下了几十年的功夫，要坐得了冷板凳，能够坚持，是很重要的。得诺贝尔奖的人都是对这方面感兴趣，对研究感到乐趣，自己做这件事本身就很高兴。有些孩子，家长要他去弹钢琴，他觉得很痛苦，这就不行了。能做出好工作的人，他本身就是感到做这件事情本身的乐趣。好的工作肯定是要积累的，当然要靠本人的坚持，也靠社会的环境能够允许这个人去坚持。我们就坚持不了，我们刚进学校的时候念数学，后来说数学没用，然后就去搞应用，后来又宣传陈景润，就又回来念数学，老是折腾，这是不行的。无论是搞理论也好，搞应用也好，要做出很好的工作哪有那么容易，肯定要长期坚持，在做的时候自己也感到蛮开心，这样才行。有些人觉得看电影是开心的，有些人

解出一个问题就高兴得要命。

**问：**近年来研究领域吸收了很多新鲜血液，吴老师对从事科研的年轻人有什么建议？

**吴老师：**巴浦洛夫对青年人搞科研提了三条，我觉得很好，一个是循序渐进，搞新东西前要把基础打扎实；一个是要热情，要做好的东西确实不容易，会碰到很多挫折，花费很多时间，没有热情也不行的；还有一个就是要谦虚，不然稍微做出一点成果就满足了。这三点我觉得确实是很重要的。巴浦洛夫的文章很短，网上都能找到。最近一个报道说，Nature上发的，中国、美国、俄罗斯、印度，针对这四个国家的计算机专业和工程专业的大学生做了大规模的测试。分成精英大学，相当于中国的985、211，一般大学，用同样的题目对这四个国家的学生进行测试，包括数学能力、物理能力等等，测试三次，一次是大学入学的时候，第二是两年以后的测试，最后是四年以后的测试。结果发现，美国大学生的批判性思维比其他三国都好，中国学生的测试结果就蛮不好的，刚入学的时候，中国学生比俄罗斯、印度的学生数学、物理都好，越到后来领先的程度越少，差别越来越小，有一部分俄罗斯学生还反超中国学生。特别是批判性思维，对一个意见提得出不同意见，提得出质疑的意见，这种能力比较欠缺。所以要真正做出好的工作，很需要培养质疑的能力，听到一个意见以后，不要马上接受它，要去想想和你过去学到的东西、和你相信的东西是不是一致的。中国文化在这方面也有所限制，比如年纪大的人说话，年纪轻的就不太好意思说话。不要说上下级，就同事之间，好像对他的说法提出不同意见就是不尊重他，或者对他有什么意见，这个就很阻碍批判思维的发展。小学也有这样的现象。我的外孙问我数学题，我用一个方法做，他回来告诉我老师说这种方法不对，我问怎么不对，他说这和标准答案的解法不一样。本来一个问题可以有不同解法，中国如果这样培养学生的话，怎么有发散的思维？如果不能发现问题，

怎么去做改进，怎么去做创新？如果觉得现有的都已经完美无缺了，就根本没办法去创新。

**问：**对 CCF-CV 刊物的读者寄语。

**吴老师：**十九世纪末二十世纪初的时候，搞物理的人说牛顿力学完美得很，但是在远处似乎有两朵乌云，实际上就是后来的相对论和量子力学。有作为的人就抓住这些漏洞做出新的工作，一定要有发现问题和批判的能力，

比具体解问题更重要。华罗庚先生说，解决数学问题就像是抓住一个兔子，最大的本事是知道哪里有兔子，然后去把兔子抓住。发现兔子，发现问题，比你去抓兔子还难。如果不知道哪里有兔子，要怎么抓呢？要找到一个好题目还是不容易的。社会环境也要允许年轻人讲出一些自己的想法。

**责任编辑** 张军平 明悦 贾嘉滨



## 吴立德

吴立德教授生于 1937 年，江苏宜兴人，1955 年考进复旦，1958 年毕业于复旦大学数学系数学专业并留校任教，在校任复旦大学首席教授、计算机软件博士生导师，主要从事图像处理、模式识别、计算机视觉和自然语言处理等方面的研究。1988 年获国家人事部授予“中青年有突出贡献专家”称号，1990 年起获国务院政府特殊津贴，1998 年获上海市优秀教育工作者称号。在上海辞书出版社和北京外文出版社出版的《中国人名大词典·当代人物卷》等名录中，吴立德教授均以学者身份名列其中。2018 年，吴立德教授获得中国计算机学会计算机视觉专委会（CCF-CV）终身学术贡

# COMPUTER VISION NEWSLETTER

02 2021  
总第 28 期



## 计算机视觉专委会简报



CCF 计算机视觉  
专委会