

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

04 2020

总第 26 期



CCCF 计算机视觉
专委会

COMPUTER VISION NEWSLETTER



计算机视觉专委会 简报

2020 年第 04 期

总第 26 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

/委员风采/

任桐炜 南京大学
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 常务委员会会议顺利召开
- 11 2020 年度 CCF-CV 专委工作会议顺利举办
- 13 CCF-CV 2021 年度系列活动申请开始征集
- 15 CCF-CV 秘书处召开 2020 年度第二次工作会议
- 16 PRCV 2021 邀您投稿参会
- 18 CCF-CV 成立教育工作组

| 科技前沿

- 19 图像美学质量评价进展
- 22 双路对抗自编码与聚类应用
- 24 面向自动驾驶场景的注意力引导实时语义分割
- 26 ECCV 2020
- 29 NeurIPS 2020

| 委员风采

- 32 南京理工大学张姗姗教授访谈
- 36 委员好消息

| 学术资源

- 38 双目立体匹配网络开源代码
- 41 人脸检测数据集
- 44 好文推荐

| 海外学者

- 47 美国加州大学洛杉矶分校谢建文博士团队
- 57 征文通知

CCF 计算机视觉
专委会

 CCFCV.CCF.ORG.CN

 CCFCVN@GMail.com

CCF-CV 走进高校系列报告会

第 90 期 华北电力大学



2020年7月28日上午，中国计算机学会计算机视觉专委会（CCF-CV）走进高校系列报告会第90期活动——“计算机视觉前沿技术及应用”，通过线上直播的形式在华北电力大学成功举行。本期报告会由华北电力大学承办，邀请了北京大学林宙辰教授、上海交通大学林巍峭教授、清华大学鲁继文副教授以及中国科学院计算技术研究所韩琥副研究员四位专家做特邀报告，由华北电力大学电子与通信工程系戚银城教授、张珂副教授和赵振兵副教授担任本次报告会的执行主席。在本次报告会上，专家们围绕“计算机视觉领域的前沿技术及应用”做了精彩报告，并在圆桌讨论环节就“电力计算机视觉技术”进行了深入的交流，Bilibili直播平台人气峰值达2.2万人，报告会反响热烈。

最后，张珂副教授进行活动总结，再次向与会的各位专家学者表示衷心的感谢，并期待更多的专家、老师和同学能关注电力视觉技术，共同努力推动计算机视觉在电力领域中的研究和应用，报告会取得了圆满成功！

第 91 期 武汉大学



2020年9月19日下午，中国计算机学会计算机视觉专委会（CCF-CV）走进高校系列报告会第91期活动——“计算机视觉前沿技术及应用”通过线上直播的形式在武汉大学成功举行。本期报告会由武汉大学测绘遥感信息工程国家重点实验室·空间智能研究所承办，邀请了中科院自动化所王亮研究员，北京大学黄铁军教授，北京大学林宙辰教授，电子科技大学李宏亮教授四位专家做特邀报告，由武汉大学测绘遥感信息工程国家重点实验室杨必胜教授和涂志刚研究员担任本次报告会的执行主席。在本次报告会上，专家们围绕“计算机视觉领域的前沿技术及应用”做了精彩报告，并在圆桌论坛环节就计算机视觉领域的多个前沿学术问题、热点应用问题进行了深入的探讨，引起了广泛的共鸣。

最后，杨必胜教授进行活动总结，再次向与会的各位专家学者表示衷心的感谢，并期待更多的专家、老师和同学能关注武汉大学计算机视觉技术的发展，推动计算机视觉的研究和应用，祝贺报告会取得了圆满成功！

CCF-CV 走进高校系列报告会

第 92 期 江西财经大学



计算机视觉前沿技术及应用

中国计算机学会计算机视觉专委会走进高校系列报告会
CCF-CV Series Lectures
江西财经大学 (第92期)



2020年9月26日上午,由CCF计算机视觉专委会(CCF-CV)主办、江西财经大学信息管理学院承办的第92期CCF-CV走进高校系列报告会通过线上直播进行。国家优秀青年基金获得者、CCF南昌执行委员会主席、江西财经大学信息管理学院院长方玉明教授出席并致辞。本期报告会主题为“计算机视觉前沿技术及应用”,邀请了中国图象图形学学会副理事长、会士、中山大学赖剑煌教授,国家杰出青年基金获得者、科技部中青年创新领军人才、大连理工大学创新学院院长卢湖川教授,国家百千万人才、西安电子科技大学电子工程学院副院长邓成教授,北京交通大学计算机科学系副主任桑基韬教授在线出席活动并作精彩报告。本次报告会上,专家们就计算机视觉的前沿技术和应用进行了深入的交流。

最后,江西财经大学姜文晖博士进行活动总结,向各位特邀讲者、与会的各位专家学者和中国计算机学会计算机视觉专委会表示感谢。以本次报告会为契机,欢迎全国的同行学者来江西财经大学指导和交流工作,并期待更多的专家学者关注江西财经大学计算机视觉技术的发展,共同努力推动计算机视觉的研究和应用!

第 93 期 南京邮电大学

CCF计算机视觉专委会 职业技能

CCF计算机视觉... 3349 No. >1000

CCF-CV Series Lectures

特邀讲者

- 13:30-14:00 开场和致欢迎辞
- 14:00-14:45 杨健 教授
· 以人为本的视觉感知
- 14:50-15:35 赫然 教授
· 生成式图像分析与理解
- 15:40-16:25 耿新 教授
· 标记增强——释放标记空间的威力
- 16:30-17:15 王楠楠 教授
· 图像跨域重建与识别
- 17:20-17:30 总结

执行主席

高广谓

2020年10月23日下午,由中国计算机学会计算机视觉专委会(CCF-CV)主办、南京邮电大学承办的第93期CCF-CV走进高校系列报告会活动通过线上直播的形式在南京邮电大学成功举行。南京邮电大学先进技术研究院高广谓副研究员担任本次报告会的执行主席。本期报告会邀请了南京理工大学杨健教授、中科院自动化所赫然研究员、东南大学耿新教授以及西安电子科技大学王楠楠教授四位专家做特邀报告。在本次报告会上,专家们围绕“视觉信息智能处理前沿技术”做了精彩报告,报告会反响热烈。

最后,南京邮电大学高广谓副研究员对本次活动进行了总结,向各位特邀讲者、与会的专家学者和中国计算机学会计算机视觉专委会再次表示感谢。希望通过此次线上活动进一步加强与国内科研院校和企业的领域专家的交流与合作。欢迎全国的同行学者来南京邮电大学交流合作,推动视觉信息智能处理的研究和应用。

CCF-CV 走进高校系列报告会

第 94 期 山东省人工智能研究院



2020年11月13日下午，中国计算机学会计算机视觉专委会（CCF-CV）走进高校系列报告会第94期活动——“计算机视觉前沿技术及应用”在济南珍珠泉宾馆成功举行。本期报告会由山东省人工智能研究院承办，邀请了中科院自动化所胡卫明研究员，哈尔滨工业大学（深圳）刘洪海教授，清华大学刘永进教授，南京大学曹汛教授，浙江大学冯结青教授，北京航空航天大学徐迈教授，南京信息工程大学郑钰辉教授等七位专家做特邀报告，由天津理工大学陈胜勇教授担任本次报告会的大会主席，山东省人工智能研究院高赞研究员担任执行主席。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，同时，也顺利完成了山东省人工智能协会的成立仪式。山东省人工智能相关企业技术人员、山东省人工智能研究院和齐鲁工业大学的师生共计200余人参加了会议，现场气氛热烈，引起了广泛的共鸣。

最后，高赞研究员对山东省人工智能相关企业技术人员、与会专家学者的到来，以及中国计算机学会计算机视觉专委会为本次活动的辛勤付出，表示衷心的感谢。

第 95 期 济南大学



2020年11月28日下午，中国计算机学会计算机视觉专委会（CCF-CV）走进高校系列报告会第95期活动——“计算机视觉前沿技术及应用”在济南大学第一报告厅成功举行。本期报告会由济南大学承办，邀请了北京大学田永鸿教授，电子科技大学王正宁副教授，中国科学院信息工程研究所葛仕明研究员和南京理工大学的张姗姗教授等四位专家做特邀报告，由济南大学信息科学与工程学院彭京亮教授担任本次报告会的执行主席。在报告会上，专家们围绕“计算机视觉前沿及应用”做了报告，报告会反响强烈，引发了广泛的共鸣。

最后，济南大学彭京亮教授进行活动总结，再次向各位特邀专家与会的各位专家学者和中国计算机学会计算机视觉专委会表示感谢。以本次报告为契机，热烈欢迎全国的同行业学者来济南大学指导和交流工作，并希望更多的专家学者关注济南大学计算机视觉技术的发展，共同努力推动计算机视觉的研究和应用，报告会在雷鸣般的掌声中圆满结束。

责任编辑 毋立芳

第 7 期 视觉行为理解的前沿进展与未来趋势

CCF-CV 视界无限系列研讨会



2020年11月8日，由中国计算机学会计算机视觉专委会主办的第7期CCF-CV“视界无限”系列活动——“视觉行为理解（VAU）的前沿进展与未来趋势”研讨会在北京工业大学成功举办。研讨会邀请了浙江大学杨易教授，中国人民大学金琴教授，上海交通大学林巍晓教授，南京大学王利民教授，北京工业大学胡永利教授，中科院自动化所黄岩副研究员，南京理工大学舒祥波教授做主题报告并参与圆桌讨论。北京工业大学科学技术发展院陈国英副院长、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副主任、中科院自动化所王亮研究员，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出席活动。本期研讨会由北京人工智能研究院、北京工业大学信息学部承办，北京工业大学科学技术发展院协办，北京工业大学尹宝才教授、毋立芳教授任执行主席。CCF计算机视觉专委会B站公众号进行了全程直播。

研讨会由北京人工智能研究院副院长冀俊忠教授、北京工业大学信息学部顾镛教授、毋立芳教授主持。

北京工业大学科学技术发展研究院陈国英副院长致欢迎辞，陈院长首先对各位专家的到来表示热烈欢迎和衷心感谢，她指出人工智能是当今非常热门的研究领域，很难得的机会可以与各位专家交流，北京工业大学为了推动首都人工智能的技术进步和发展，专门成立了北京人工智能研究院，希望北京工业大学能够在专委会的支持和帮助下，同各位专家为人工智能腾飞共同发力。希望同学们在专家的带领下，不断学习，将来成为行业内的新生力量，栋梁之才！

中国计算机学会计算机视觉专委会主任查红彬教授致辞，查教授指出北京工业大学作为市属高校成立北京人工智能研究院，CCF-CV专委的视界无限活动能在这里举办很有意义，很高兴能够一起探讨计算机视觉、人工智能方面的前沿问题。CCF-CV有丰富的系列活动包括走进高校、走进企业、视界无限等。其中视界无限活动定位围绕一个主题进行深入讨论，希望通过本期活动，对于从事“视觉行为理解”领域研究的老师和同学有所启发。也感谢北京工业大学相关的老师、同学们为举办本次活动做出的努力。最后预祝活动圆满成功。



毋立芳教授引导发言的题目是“视觉行为理解的一点思考与实践”。毋教授从看图说话的基本范式抽象出视觉理解任务包括对象识别、行为识别、场景识别、行为描述等，并进一步分析了视觉行为识别以及群体行为识别面临的挑战，如相同对象的不同关系表达不同行为、相同关系的不同组合顺序表达不同行为、群体行为和个体行为的复杂关系，以及行为描述面临的语义鸿沟、多模态对齐等问题，相信通过今天的活动，一线专家的经验分享能够给大家带来启示。最后简单介绍了他们组的基于运动模式的群体行为识别工作。

杨易教授的报告题目是“Towards efficient and effective video analysis”。杨教授指出视频比图像包含更多的信息，但对视频的行为理解分析，却比处理图像的难度大很多。因此对于视频的行为分析理解，要做到兼顾效率与准确性。针对上述问题，杨教授介绍了 feature aggregation 网络、multi-rate modeling 方法、Faster recurrent networks for efficient video classification 论文中实现的网络框架、专门针对视频分析如何提高标注效率的 SF-Net。最后介绍了第一视角的视频的应用场景，在 object-centric feature alignment 上提出了一种新的 attention 机制。



金琴教授的报告题目是“Visual activity understanding: my two cents”。金教授认为视觉行为理解是很宽的一个主题，我们可以从不同方面、不同的层次、不同的数据集、不同的领域等细分视觉行为理解的工作。主要讲解：1.视觉关系检测，以及其中重要的 human object interaction (HOI)检测的相关工作。2.针对更为复杂的用自然语言描述图像/视频行为的相关工作。3.特定领域、细粒度行为的一些总结，最后介绍了未来的研究方向。

林巍峭教授的报告题目是“以人为中心的复杂视频分析挑战”。重点介绍了自己团队在数据集方面的工作。在数据标注的过程中，针对目标检测、姿态估计、姿态跟踪、行为分析等任务对数据进行了不同级别的标注，对现有主流的 MSCOCO、MPII、CrowdPose 等数据集进行了对比。介绍了举办的 ACM MM Grand Challenge，最后介绍了未来的研究方向，讲解了论文 Multi-source sounding object localization 如何通过音频信息将视频中的目标定位出来。

王利民教授的报告题目是视频动作理解：识别、检测与跟踪。王教授首先介绍了近年来视频理解领域的相关进展，之后深入浅出地讲解了他们组在视频人体动作理解方面的系列工作，包括：1.视频运动的表征方法。2. 视频动作的检测框架。3. 视频目标的跟踪技术。针对视频短时运动的表征与建模，提出了简单高效的时序建模模块 (TEINet and TAM)，在速度效率和建模精度方面取得较好的效果；针对动作时空检测，提出了一种免锚框设计的时空管道检测器 (MOC)；针对视频目标跟踪，提出了一种简单的在线跟踪框架 (FCOT)，通过设计全卷积操作同时实现物体中心和大小的估计。

胡永利教授的报告题目是“流形上的稀疏低秩表示模型及应用”。胡教授首先从如何无监督地获取视频信息这个问题切入，引出了稀疏低秩表示模型这个方法，并介绍了几个经典模型。随后介绍了他们团队提出的流形上的稀疏低秩表示模型，在高维数据流形表示的基础上，建立了单一流形、乘积流形和异构流形上的低秩表示模型，给出了保持数据流形结构和非线性度量的模型约束，并解决了流形上稀疏低秩模型的复杂优化问题。

将流形上的稀疏低秩表示模型应用于图像视频等高维数据的聚类，所提出的流形空间方法对比欧氏空间方法性能得到大幅提升。



黄岩副研究员的报告题目是“弱监督和语言驱动的视频行为定位”。黄老师首先介绍了视频行为定位的定义，又举例说明了弱监督行为定位如何进行标签标注。黄老师认为弱监督标签可以减少标注的代价，并结合两个工作说明了弱监督方法的可行性：1. 如何建模不同行为标签之间的关系以缓解行为不均衡问题。2. 如何高效地进行跨模态行为搜索。最后黄老师简要展望了未来可能的发展方向。



舒祥波教授的报告题目是“个体-交互-群体”多粒度下的视频行为解析进展。舒老师从三个方面进行此次汇报：1、骨骼点动作预测合成，2、交互动作分析，3、群体活动分析。在骨骼点动作预测方面，他们团队使用了时空协同注意力 RNN 的方法，利用注意力机制在时空两个维度捕捉关键性骨骼点。在交互动作识别方面，

舒老师团队提出了一种新的长短时同现记忆网络，利用该网络来直接学习人与人的关联动作特征，而不是个体动作特征。在群体活动分析方面，舒老师认为个体“有用”动作与自身前后的动作具有相似的特征表达，个体“有用”动作与其他个体的大部分动作具有相似的特征表达。因此，制定了一种基于一致性约束图 LSTM 的个体-群体行为识别方法。最后简要讨论了未来可能的重点研究方向。



紧接着是 panel 环节，毋立芳教授和七位讲者一起探讨了视觉行为理解领域前沿问题和发展趋势。



最后，由北京工业大学信息学部副主任杨震教授进行活动总结。杨主任感谢计算机学会计算机视觉专委会将每季度一次的重大活动“视界无限”放在北工大来举行，同时感谢讲者们为在场的师生以及线上的同学们送上了精彩的学术盛宴。

责任编辑 杨巨峰

CCF-CV 常务委员会会议顺利召开

2020年10月16日，中国计算机学会计算机视觉专委会（CCF-CV）常务委员会会议于南京顺利召开。

本次常委会工作会议距离上一次召开已过十个月，尤其在专委会工作经受住了新冠疫情考验之后，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。会议由专委会主任查红彬教授主持，秘书处全体成员列席。

会议首先由专委会主任查红彬教授讲话。查主任充分肯定了专委会过去一年取得的各项进展，尤其肯定了疫情期间采用线上方式持续推进专委会各项工作的努力，并对接下来的工作提出了多项切实可行的发展建议。



随后，常委会委员对今年的新申请委员进行了投票评选。值得指出的是，今年新申请委员人数较多、

竞争激烈，只有33名申请者幸运当选。接下来，常委会委员针对专委会发展相关议题展开了高效深入的讨论，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。

责编黄岩

2020 年度 CCF-CV 专委工作会议顺利举办



中国计算机学会计算机视觉专委会 (CCF-CV) 年度工作会议于 2020 年 10 月 16 日在南京国际会展中心顺利举办, 超 400 名来自全国高校、科研院所、企业的现任委员和新申请委员在线上或现场参加了工作会议。会议由专委会秘书长、北京邮电大学马占宇教授主持。

专委领导致辞

会议首先由专委会主任、北京大学查红彬教授致辞。查主任感谢从祖国各地到场参会的新老委员, 感谢因现场人数限制而在线参加会议的委员, 他积极肯定了专委会在过去一年中各方面取得的丰硕成果, 鼓励委员们在当前人工智能热潮中聚焦学术前沿、做出高质量研究工作, 希望专委会大家庭一起精诚合作, 各项活动进一步提升品牌质量, 打造精品学术交流活动, 为推进计算机视觉学术研究与产业发展继续发挥积极的引领作用。

CCF 代表讲话

随后, 中国计算机学会 (CCF) 常务理事、浙江大学卜佳俊教授代表学会致辞。卜佳俊教授对专委会工作会议的召开表示祝贺, 对专委会开展的各项活动给予了积极评价, 特别肯定了专委会建立的一系列学术交流品牌活动。

CCF 专委工委委员、中科院计算技术研究所蒋树强

研究员代表专委工委致辞。蒋树强研究员热情洋溢地回顾了几年来连续参加 CCF-CV 专委工作会议的体会, 积极肯定了专委会取得的各项丰硕成果, 并祝专委会在各位主任和常委们的带领下齐心协力、越办越好。

年度工作报告

接下来, 专委会秘书长马占宇教授向与会委员做了年度工作报告。报告简要介绍了专委会组织结构, 包括专委会新增加的顾问委员会和国际顾问委员会; 通报了 1 月份专委会年度常委会议和 3 月份线上秘书处工作会议所作出的若干新举措, 全面回顾了专委会过去一年的学术交流活动, 指出了工作中存在的问题和改进方案; 总结了过去一年专委会工作的重要成就, 特别强调 ICCV 2025 申办成功、专委会委员在助力抗击疫情中的贡献, 以及专委会委员获得的各项奖励和荣誉, 最后介绍了下一年度专委会工作计划。

CCF-CV 专委会为了更好地发挥资深委员的作用, 成立了顾问委员会, 旨在进一步推动专委会服务学术、联系业界等工作开展。专委会邀请顾问委员会委员出席专委会常委会议、秘书处工作会议等各项活动, 并就专委会的工作请求指导建议。

CCF-CV 专委会为了推进专委会的国际化建设、促进海内外计算机视觉领域的交流合作, 成立了国际顾问委员会。专委会邀请国际顾问委员会委员参加专委会的学术活动, 定期向顾问委员会汇报专委会的发展情况, 并就专委会开展的重要活动进行咨询。

常委和委员增选

根据日程, 工作会议还进行了新增常委和委员的选

举，由专委副主任、中科院自动化研究所王亮研究员主持。本年度共收到 4 位委员申请常委委员，75 位候选人申请成为委员。4 位常委候选人逐一上台进行了申请陈述，经过参会的现任委员投票表决，共有 2 名候选人当选为新任常务委员。他们是北京工业大学毋立芳教授、南京理工大学唐金辉教授。新委员的选举由现任常务委员通过投票表决，最终共有 33 位候选人当选为新任委员。

CCF-CV 奖项颁奖

接下来，进入激动人心的 CCF-CV 颁奖环节。颁奖仪式由专委会副主任、提名与奖励工作组组长、南京信息工程大学刘青山教授主持。刘青山教授详细介绍了本年度设立的奖项评选范围与评选规则，包括：终身学术贡献奖、杰出成就奖、服务贡献奖、学术新锐奖和新增设的持久影响力论文奖。5 个奖项的获奖人既有早年投入我国计算机视觉奠基性研究的老一代科学家，又有新一代崭露头角的学术新星，也有为专委发展尽心尽力、无私奉献的中生代科研工作者，充分展示了计算机视觉专委会大家庭的繁荣兴盛。

本年度的 CCF-CV 终身学术贡献奖授予南京理工大学杨静宇教授，由专委会主任查红彬教授颁奖。杨静宇教授现场发表了质朴亲切的获奖感言，感谢专委会的认可和颁奖，勉励青年研究者继续做出更大成绩。杨静宇教授的发言赢得了代表们的热烈掌声，他的鼓励与期许极大振奋了委员们的科研工作热情。

CCF-CV 杰出成就奖本年度授予专委会主任查红彬教授，以表彰他在三维视觉计算、环境几何建模等方面做出的系统的创造性贡献，由专委顾问委员会委员陈熙霖研究员颁发奖励证书。查红彬教授发表了热情风趣的获奖感言，衷心感谢专委会同仁的支持与认可，并表示继续致力于开展 SLAM 研究。

CCF-CV 专委首次颁发持久影响力论文奖，奖励在计算机视觉领域中产生持久影响力的创新性研究论文。Kaiming He, Jian Sun, Xiaoou Tang 发表于 ECCV 2010 (Oral) 的《Guided Image Filtering》获得此殊

荣，由专委副主任王亮研究员颁奖。

本年度 CCF-CV 专委的中科视拓 Seeta 学术新锐奖授予香港科技大学李铎、南开大学高尚华、浙江大学彭思达三位博士研究生，由专委副主任虞晶怡教授颁奖。中科视拓 Seeta 服务贡献奖授予哈尔滨工业大学左旺孟教授、南京理工大学潘金山教授、合肥工业大学余烨副教授等 3 位委员，由专委秘书长马占宇教授颁奖，感谢他们在专委会活动、会议组织、专委简报/网站工作等方面作出的重要贡献。

委员建言献策

委员建言献策环节由专委副主任虞晶怡教授主持。虞晶怡教授介绍了 CVPR 2021 采用基于 AI 的领域选择机制、将由投稿者担任审稿工作，欢迎更多专委委员担任 CVPR、ICCV 等国际顶级会议的领域主席和审稿专家，提升中国计算机视觉研究者在国际学术界的影响力。南开大学程明明教授倡议计算机视觉领域能够在新形势下继续推进开放共享，以引领技术发展和生态建设。合肥工业大学贾伟教授提出希望通过创办高水平的期刊和会议建立以中国为中心的计算机视觉研究。查红彬主任针对厦门大学杨晨晖教授关于中美脱钩可能对学术界带来影响的提问，指出保持国际交流合作是专委会要坚持的一个方向，专委为此成立了国际顾问委员会，希望通过海外华人学者带动更大范围的国际交流合作。

最后，专委会 2020 年度工作会议在热烈的掌声中圆满结束，期待明年北京再聚！



责任编辑 黄岩

CCF-CV 2021 年度系列活动申请开始征集



中国计算机学会计算机视觉专委会 (CCF-CV) 是国内计算机视觉领域最有影响力的专委会之一，为了促进本领域研究人员更好地开展学术/技术交流和发展战略研究，宣传高水平研究成果，推动国内计算机视觉学科发展，提升我国计算机视觉研究在国际上的影响力，专委会开展了 CCF-CV 走进高校系列活动、CCF-CV 走进企业系列交流会、CCF-CV 视界无限系列研讨会，上述活动类型丰富多样，尽可能地覆盖了不同类型专业人员的需求。各项活动开展以来，得到同行的大力支持，在国内外计算机视觉领域引起了强烈反响。在这里，向对活动的顺利开展提供过支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

现开始征集 2021 年度各类活动申请，欢迎高校、科研院所、企业从事计算机视觉领域研究与应用的人员积极申请！

CCF-CV 走进高校系列活动

活动宗旨

为了更好地推动计算机视觉领域的学术与技术交流，促进国内外学者间的了解与合作，全面推动国内计

算机视觉学科发展，提升我国计算机视觉研究在国际上的影响力，中国计算机学会计算机视觉专委会拟在全国范围的高校和科研院所等开展 CCF-CV 走进高校系列报告会活动。活动采取请进来的方式邀请高水平学者以学术报告等形式宣传高水平研究，同时对承办单位青年研究人员等进行面对面指导，帮助青年学者快速成长。

活动开展

自 2015 年 11 月开展以来，CCF-CV 走进高校系列报告会活动得到了高校和研究所、讲者、听众的大力支持。截至 2020 年 11 月底，CF-CV 走进高校系列报告会活动已成功举办了 95 期，遍及祖国大陆所有省市自治区直辖市，并在澳门、悉尼成功举办。目前，活动已邀请讲者做分享专题报告 370 余场，活动现场人均听众 200 人次，自 2020 年以来开展线上活动近 10 场，B 站人气峰值达 2.2 万，微信公众号平均阅读 1000 余次。同时在征得讲者同意的前提下，于专委 B 站账号分享部分讲者的报告视频，并于专委会网站分享讲者的报告 PPT。在这里，向对活动的顺利开展提供支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

活动申请

活动面向计算机视觉领域研究者开放申请，如您有意申请，请与 CCF-CV 秘书处联系。联系方式：

毋立芳：lifu@bjut.edu.cn

杨巨峰：yangjufeng@nankai.edu.cn

CCF-CV 走进企业系列交流会**活动宗旨**

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

活动开展

从 2015 年底至今，CCF-CV 企业交流活动得到了企业、讲者、听众的大力支持。截至目前，CCF-CV 企业交流活动已成功举办了 21 期，相继走进了多家国内外知名企业。在这里，向对活动的顺利开展提供支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

活动申请

活动采取自愿申请的方式，有意承办活动的企业请下载并填写申请表并发至 CCF-CV 秘书处。联系方式：

黄 岩：yhuang@nlpr.ia.ac.cn

杨巨峰：yangjufeng@nankai.edu.cn

CCF-CV 视界无限系列研讨会**活动宗旨**

“视界无限”是由中国计算机学会计算机视觉专委会 (CCF-CV) 新启动的品牌学术活动。每季度举办一次，全年四次。每次活动由 1-2 位计算机视觉专委会委员负责

召集和组织，围绕计算机视觉领域某一主题，组织资深的研究者与优秀青年学者进行全方位的深入研讨，总结该主题前沿进展与未来趋势。

活动开展

从 2019 年 1 月至今，CCF-CV 视界无限系列研讨会得到了国内外视觉领域同行、讲者、听众的大力支持。截至目前，已成功举办了 7 期，围绕“生成式对抗网络”、“同时定位与地图构建”、“行人重识别”、“视觉目标跟踪”、“虚拟现实/增强现实”、“新冠肺炎影像诊断”、“视觉内容理解”等主题开展深入探讨，有力地推动了我国计算机视觉领域的研究和发展。在这里，向对活动的顺利开展提供支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

活动申请

活动面向计算机视觉领域研究者开放申请，如您有意组织“视界无限”活动，请与 CCF-CV 秘书处联系。联系方式：

杨巨峰：yangjufeng@nankai.edu.cn

黄 岩：yhuang@nlpr.ia.ac.cn

责任编辑 毋立芳

CCF-CV 秘书处召开 2020 年度第二次工作会议



2020 年 11 月 8 日，中国计算机学会计算机视觉专委会 (CCF-CV) 秘书处本年度第二次工作会议在北京召开，专委副主任王亮研究员出席会议并指导工作，秘书长马占宇教授主持会议。

本次会议主要讨论如何落实专委会常务委员会第六次会议的各项决议，并制定秘书处下一阶段的工作计划。2020 年，专委各项学术活动和组织建设稳步推进，走进高校、走进企业、视界无限、专委简报与网站的活

动形式与内容不断推陈出新，继续展示 CCF-CV 专委的组织活力与特色引领性。新一年度，各项活动如何更有实效地开展，如何更加广泛地激发委员们参与专委建设的积极性，如何办好专委年度学术论坛 RACV 2021 和讲习班，大家围绕这些议题展开了热烈讨论，并形成了具体可行的执行方案。



责任编辑 黄岩

PRCV 2021 邀您投稿参会

第四届中国模式识别与计算机视觉大会 (The 4th Chinese Conference on Pattern Recognition and Computer Vision, PRCV 2021) 将于 2021 年 10 月 29 日至 11 月 1 日在北京举行。PRCV2021 由中国图象图形学学会 (CSIG)、中国人工智能学会 (CAAI)、中国计算机学会 (CCF) 和中国自动化学会 (CAA) 联合主办; 由北京科技大学、北京交通大学和北京邮电大学共同承办, 中山大学协办, 是国内顶级的模式识别和计算机视觉领域学术盛会。

本届会议将主要汇聚国内国外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行, 共同分享我国模式识别与计算机视觉领域的最新理论和技术成果, 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、原创性的优秀英文学术论文。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。

征文范围 (包括但不限于)

模式分类与聚类分析
结构模式识别
机器学习
神经网络与深度学习
特征提取与特征选择
计算机视觉基础理论
底层视觉理解、图像处理

三维视觉与重构
文档分析与识别
字符识别
人脸识别与姿态识别
目标检测、跟踪与识别
行为识别
多媒体分析与推理
医学图像处理与分析
生物特征识别
遥感影像解译
优化及学习方法
多模态信息处理
性能评测和基准数据库
视频分析与理解
视觉应用与系统
机器人、自动驾驶中的视觉问题

重要日期

投稿截止日期: 2021 年 4 月 30 日
录用通知日期: 2021 年 6 月 30 日
终稿提交日期: 2021 年 8 月 15 日
会议举办日期: 2021 年 10 月 29 日 至 2021 年
11 月 1 日

会议投稿

投稿系统:

<https://cmt3.research.microsoft.com/PRCV2021/>

论文模板:

<https://www.springer.com/cn/computer-science/Incs/conference-proceedings-guidelines>

相关要求:

论文必须未公开发表过, 文章最长 12 页 (含参考文献), 需为英文论文;

论文应包括题目、摘要、关键字、正文和参考文献。

大会组委会

指导委员会:

主席: 谭铁牛

委员: 陈熙霖, 刘成林, 芮勇, 查红彬, 郑南宁, 周杰

秘书长: 王亮

大会主席:

谭铁牛 中科院自动化所

王耀南 湖南大学

赖剑煌 中山大学

赵 耀 北京交通大学

程序主席:

马惠敏 北京科技大学

王 亮 中科院自动化所

张长水 清华大学

吴 飞 浙江大学

组织主席:

殷绪成 北京科技大学

马占宇 北京邮电大学

朱振峰 北京交通大学

王瑞平 中科院计算所

赞助主席:

俞能海 中国科技大学

白 翔 华中科技大学

刘 越 北京理工大学

杨金锋 深圳职业技术学院

宣传主席:

孔祥维 浙江大学

梅 涛 京东 AI 研究院

刘家瑛 北京大学

曾 丹 上海大学

国际联络主席:

虞晶怡 上海科技大学

李学龙 西北工业大学

阮邦志 香港浸会大学

讲习班主席:

郑伟诗 中山大学

程明明 南开大学

韦世奎 北京交通大学

专题论坛主席:

黄 华 北京师范大学

彭宇新 北京大学

王楠楠 西安电子科大

博士生论坛主席:

彭 玺 四川大学

苏 航 清华大学

白慧慧 北京交通大学

竞赛主席:

桑 农 华中科技大学

左旺孟 哈尔滨工业大学

谢晓华 中山大学

专刊主席:

鲁继文 清华大学

向世明 中科院自动化所

吴建鑫 南京大学

出版主席:

林宙辰 北京大学

林春雨 北京交通大学

田华伟 中国人民公安大学

注册主席：	殷君君 北京科技大学	网站主席：	朱 超 北京科技大学
	明 悦 北京邮电大学		何召锋 北京邮电大学
	肖继民 西交利物浦大学		丛润民 北京交通大学
展示主席：	杨小康 上海交通大学	财务主席：	王卫苹 北京科技大学
	祝晓斌 北京科技大学		毋立芳 北京工业大学
	张淳杰 北京交通大学		刘美琴 北京交通大学

责任编辑 黄岩

CCF-CV 成立教育工作组

在学会的统一部署下，为推动计算机视觉相关专业的教育发展和人才培养，进一步发挥计算机视觉专委会（CCF-CV）在自身专业领域的优势，专委会设立了教育工作组。

经专委会常务委员会讨论决定，成立的教育工作组成员如下：

组长：刘青山 南京信息工程大学

成员：林宙辰 北京大学

耿 新 东南大学

毋立芳 北京工业大学

张军平 复旦大学

责任编辑 黄岩

专题概述

图像美学质量评价进展

山东财经大学 崔超然

一、引言

传统的图像质量评价研究主要关注在图像的获取、传输和存储等过程中引入的不同类型的失真(如噪声、扭曲、压缩等),以及由此带来的图像质量下降问题。随着图像采集设备的不断升级和编解码技术的快速进步,普通图像的失真问题目前已经得到一定程度的缓解。与此同时,一些研究者开始尝试从美学的角度对图像质量进行客观评价,并提出了图像美学质量评价(image aesthetics assessment)的概念。

长久以来,美学(美感)一直作为哲学问题来讨论。美的作品由人来评判,并使人们获得审美的愉悦。但是,面对日益激增的海量图像资源,人们已经无法胜任繁重的评判工作,希望计算机能够进行辅助美学分析。图像美学质量评价旨在模拟人类视觉及审美思维,进而对图像进行美学建模,使计算机能够自动地对图像的美学价值进行定量的评价。

目前,图像美学质量评价技术已经被应用到多个领域。例如,针对用户拍摄的关于同一场景的多张候选照片,筛选最具美感的作品保存和展示,合理地降低数据的存储开销;在图像作品的创作和编辑中,分析对比候选方案的美学质量,提升作品的视觉美感;在图像检索系统中,考虑返回图像的美学质量,为用户提供准确且更有吸引力的检索结果。因此,在图像的存储、编辑、检索等方面,图像美学质量评价技术都具有重要的研究价值和广阔的应用前景。

二、研究现状

图像美学质量评价研究的核心在于建立图像内容

表达和图像美学判定之间的关联。伴随计算机视觉技术的快速发展,现有的图像美学质量评价方法主要关注于提取和构造有效的图像视觉特征,并结合视觉特征利用分类方法对高美感和低美感的图像进行区分,或采用回归方法来评估图像美感分值。

早期的图像美学质量评价研究主要以摄影中的美学规则为先验知识,如三分构图法、景深和视觉平衡原则等^[1],人为地设计提取图像的视觉特征。一些高层次视觉属性特征也相继被提出以建模图像的色度、亮度、内容布局等信息^[2]。尽管这些早期工作取得了一定进展,但人工设计图像美学特征往往需要较强的领域知识和一定的摄影经验。此外,在部分图像美学质量评价任务中,一些通用的图像视觉特征(如颜色直方图、SIFT等)也被证明要优于这些定制的基于摄影规则的特征^[3]。

伴随深度学习的兴起,卷积神经网络(Convolutional Neural Networks, CNN)被广泛用于自动学习图像美学特征,并取得了较好的效果。[4]是第一个将CNN用于图像美学质量评价的工作,该方法分别抽取图像的局部特征和全局特征,并利用图像的风格和语义标签进一步提升美学质量评价的性能。[5]提出了一种深度图像块融合网络,从单幅图像中随机裁剪多个图像块并将它们的特征进行融合。在此基础上,[6]提出了一种启发式的图像块选择策略,在训练阶段识别最具判别性的图像块特征。[7]引入注意力机制来自适应地调整不同图像块特征在融合时的权重以提高学习效率。[8]证明将人工设计的特征与预训练CNN提取的特征相融合,可以有效提高图像美学质量评价的性能。为了规避CNN对输入图像的固定尺寸要求,[9]在卷积层后插入

空间金字塔池化层,从而使模型可以接收任意尺寸和长宽比的图像。

另一方面,考虑到美感本质上是人的一种主观感受,不同的用户面对同一幅图像时往往会得出不同的美学评价结果,即所谓“各花入各眼”。从这点出发,[10]提出了一种面向质量等级分布的图像美学质量评价方法,直接预测一幅图像处在不同质量等级上的分布,以分布的形式反映用户关于图像审美感受的差异性。[11]和[12]分别通过挖掘用户对图像的文本评论和点赞信息,来建模用户的审美偏好,进而构建个性化的图像美学质量评价系统。[13]收集了一个带有打分者身份信息的图像美学质量评价数据集,并利用主动学习算法来预测指定打分者对图像的美学评分。[14]提出了一个具有共享权重的多任务学习网络,以预测图像的美学分布和喜欢该图像的用户的人格特质,最终通过考虑用户和图像之间的人格特质匹配程度来给出个性化的美学评价结果。[15]

提出了一种基于元学习的个性化图像美学质量评价方法,利用双层梯度优化算法仅根据少量样本就可以得到目标用户的个性化美学评价模型。

三、未来发展方向

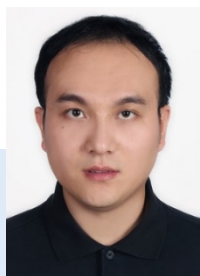
目前,个性化图像美学质量评价逐渐受到研究人员关注,如何理解用户的审美偏好是一个亟待解决的问题。然而,用户的审美偏好是相对抽象的,且难以直接获取和表示。在社交媒体中,用户与图像之间存在多种交互行为。考虑到人的认知和行为是彼此相互影响的,可以借助用户的社交行为来感知用户的审美偏好。因此,我们认为未来可以从“社会感知”的角度尝试解决个性化图像美学质量评价问题^[16]。此外,个性化图像美学质量评价中的小样本学习和冷启动问题也是值得研究的方向。

责任编辑 任桐炜

参考文献

- [1] Ritendra Datta, Dhiraj Joshi, Jia Li, James Z Wang. Studying aesthetics in photographic images using a computational approach. In the 8th European Conference on Computer Vision, 288–301, 2006.
- [2] Sagnik Dhar, Vicente Ordonez, Tamara L Berg. High level describable attributes for predicting aesthetics and interestingness. In the IEEE Conference on Computer Vision and Pattern Recognition, 1657–1664, 2011.
- [3] Luca Marchesotti, Florent Perronnin, Diane Larlus, Gabriela Csurka. Assessing the aesthetic quality of photographs using generic image descriptors. In the IEEE International Conference on Computer Vision, 1784–1791, 2011.
- [4] Xin Lu, Zhe Lin, Hailin Jin, Jianchao Yang, James Z Wang, Rating image aesthetics using deep learning. IEEE Transactions on Multimedia, 17(11): 2021–2034, 2015.
- [5] Xin Lu, Zhe Lin, Xiaohui Shen, Radomir Mech, James Z Wang. Deep multi-patch aggregation network for image style, aesthetics, and quality estimation. In the IEEE International Conference on Computer Vision, 990–998, 2015.
- [6] Shuang Ma, Jing Liu, Chang Wen Chen. A-lamp: Adaptive layout-aware multi-patch deep convolutional neural network for photo aesthetic assessment. In the IEEE Conference on Computer Vision and Pattern Recognition, 4535–4544, 2015.
- [7] Kekai Sheng, Weiming Dong, Chongyang Ma, Xing Mei, Feiyue Huang, Bao-Gang Hu. Attention-based multi-Patch aggregation for image aesthetic assessment. In the 26th ACM International Conference on Multimedia, 879–886, 2018.
- [8] Michal Kucer, Alexander C Loui, David W Messinger. Leveraging expert feature knowledge for predicting image aesthetics. IEEE Transactions on Image Processing, 27(10): 5100–5112, 2018.
- [9] Long Mai, Hailin Jin, Feng Liu. Composition-preserving deep photo aesthetics assessment. In the IEEE Conference on Computer Vision and Pattern Recognition, 497–506, 2016.

- [10] Chaoran Cui, Huiui Liu, Tao Lian, Liqiang Nie, Lei Zhu, Yilong Yin. Distribution-oriented aesthetics assessment with semantic-aware hybrid network. *IEEE Transactions on Multimedia*, 21(5): 1209–1220, 2019.
- [11] Guolong Wang, Junchi Yan, Qin Zheng. Collaborative and attentive learning for personalized image aesthetic assessment. In the 27th International Joint Conference on Artificial Intelligence, 957-963, 2018.
- [12] Chaoran Cui, Wenya Yang, Cheng Shi, Meng Wang, Xiushan Nie, Yilong Yin. Personalized image quality assessment with social-sensed aesthetic preference. *Information Sciences*, 512: 780-794, 2020.
- [13] Jian Ren, Xiaohui Shen, Zhe Lin, Radomir Mech, David J Foran. Personalized image aesthetics. In the IEEE Conference on Computer Vision and Pattern Recognition, 638-647, 2020.
- [14] Leida Li, Hancheng Zhu, Sicheng Zhao, Guiguang Ding, Weisi Lin. Personality-assisted multi-task learning for generic and personalized image aesthetics assessment. *IEEE Transactions on Image Processing*, 29: 3898-3910, 2020.



崔超然

山东财经大学计算机科学与技术学院教授，主要研究方向为数据挖掘与多媒体信息处理。
Email: crcui@sdufe.edu.cn。

热点追踪

双路对抗自编码与聚类应用

中山大学 任传贤

无监督聚类作为计算机视觉和模式识别领域的基本任务，一直被视为数据分析中探索无标签数据结构和提取分类信息的重要途径。经典的聚类算法如K-means、高斯混合模型(GMM)旨在通过数据分布提取分类信息，然而该信息一般与风格、背景等干扰信息耦合在一起，这类算法又大多需计算原始数据空间中样本间的距离，致高维情况下处理结构复杂数据的效果不尽人意。另一类以谱聚类为代表的聚类算法旨在直接识别类别，但其效率与样本数量息息相关，因而难以处理大规模数据集。

近年来，深度学习促进了聚类方法在高维空间的发展并建立了各种具有显著性能模型。以深度生成式模型为例，它们可在某些有关潜在变量的假设下学习潜在变量的数据分布，进而编码得到丰富的潜在结构。变分自编码器(VAE)是训练生成式模型的一个重要框架，其利用随机梯度变分贝叶斯和重参数化技巧来最大化似然函数，但也可能出现过度正则化而导致聚类效果退化。对抗自编码(AAE)在AE基础上引入对抗式训练机制，将隐层编码的聚合后验分布与一任意先验分布匹配，保证模型尽可能接近复杂的先验分布。AAE在图像类型和内容的分离、降维、数据可视化等应用中具有良好的性能，然而其最大化似然函数这一目标与聚类目标不全相符，同时对对抗式机制容易产生模式崩塌。因此，如何解决相关问题以获得更准确、更稳定的聚类效果仍是研究热点。

基于以上目标我们提出了双路对抗自编码器模型(Dual-AAE)，该模型在最大化似然函数的同时还最大化观测样本及潜变量子集之间的互信息，从而捕获到更多数据结构信息，似然函数和互信息的变分下界通过相同

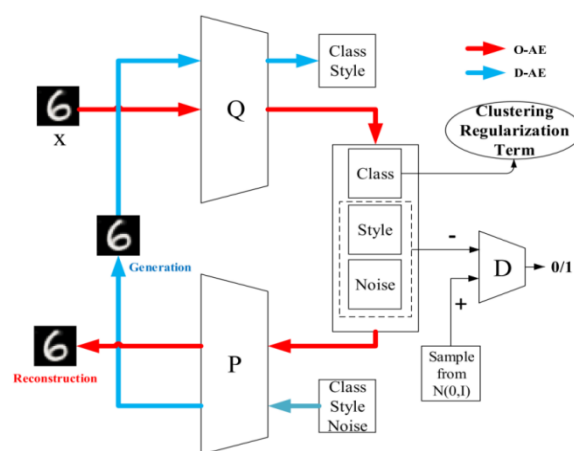


图 1 Dual-AAE 主要网络结构

的变分分布推断得到。Dual-AAE由用于重构输入数据的自编码O-AE和重构潜在变量的自编码D-AE构成，两个AE共享网络参数。我们对O-AE中的潜在变量采用对抗机制以获得更好的数据流形，并引入基于后验分布和边际分布熵的聚类正则项(CR)代替对抗式训练。该模型不仅能有效提取分类信息，而且不易出现模式崩塌的风险。

Dual-AAE的主要网络架构如图1所示。红色箭头表示D-AE的数据流方向，观测样本通过编码器Q编码为类别、风格和噪声变量，再经由解码器P重构，该过程最小化观测样本的重构误差以确保潜在变量包含全部信息。在D-AE的编码层通过CR项获得分类变量的更多信息，同时将风格变量和随机噪声输入判别器D，对聚合后验正则化以使其与高斯先验匹配。蓝色箭头表示O-AE的数据流方向，从先验分布提取出的类别、风格和噪声变量输入编码器P生成新样本，再由解码器Q重构，最小化重构误差来减少信息丢失。为进一步提高聚类效果，基于大多数错分样本的最大聚类概率远小于1的经验观察，可以引入拒绝

项来避免对困难样本做决策，利用更有效的方法如人工监督单独解决。

Dual-AAE模型的目标函数由多部分构成：1) 重构损失。O-AE的重构损失鼓励更好地解释输入数据，D-AE的重构损失鼓励结构化潜变量从数据中提取尽可能多的信息；2) 对抗损失。利用对抗式训练过程正则化风格变量和随机噪声的后验以匹配高斯先验；3) 聚类正则项，促使类别变量从输入数据中捕获分类信息。利用反向传播算法对网络参数进行优化后，Dual-AAE将目标数据分配给后验概率最高的类别作为聚类结果。一种难以避免的情况是，一些容易混淆的样本即使通过有监督的分类算法也无法有效处理。我们这里通过添加拒绝项，设置阈值并拒绝最大后验分布不超过该值的输入样本，用人工注释等更有效的监督方法来处理。

我们的新型生成式聚类算法Dual-AAE为聚类问题提供了训练简化且稳定的模型，其有效性在4个公开的数据集上进行了实验验证，表现普遍优于现有多种较为流行的聚类算法。详细结果如表1所示。在MNIST数据集上对带有拒绝项的Dual-AAE进行评估，其在少数样本被拒

表 1 不同算法的聚类准确率

Method	MNIST	HHAR	STL-10	REUTERS
K-means	53.55	60.08	64.86	54.04
N-Cuts	44.83	53.60	64.23	N/A
GMM	53.73	60.34	72.44	55.81
VAE+GMM	72.94	68.02	78.86	70.98
DEC	84.30	79.86	80.62	75.63
GMVAE	88.54	-	-	-
DEPICT	96.50	-	-	-
AAF	84.10	83.77	81.24	75.12
CatGAN	84.61	80.77	83.08	75.9
InfoGAN	96.33	83.37	82.14	75.33
VaDE	94.46	84.45	84.46	79.38
Dual-AAE(without CR)	97.04	85.94	86.85	79.56
Dual-AAE	97.86	86.79	89.15	81.45

绝的情况下几乎可以正确聚类所有样本。特别的，在无监督条件下该模型还能区分图像的风格内容。

除了聚类任务，由于对抗式自编码模块的存在，Dual-AAE还可以从特定群集中生成高质量的样本，即实现图像生成，这对完成图像自动生成(采样)、图像信息补全等工作具有一定的借鉴意义。图2可见，(a)VAE和(b)GAN无法生成指定类别的样本，而(d)Dual-AAE不仅克服了该缺点，而且得到比生成模型(c)InfoGAN更平滑真实的图像。

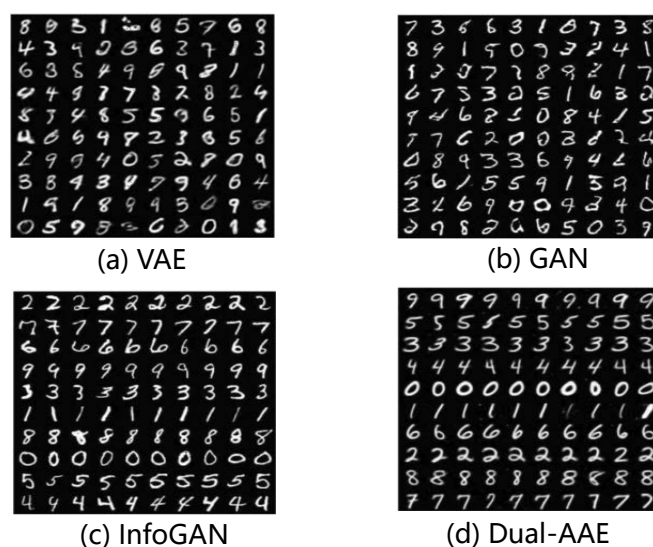
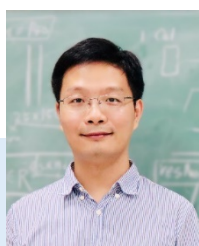


图 2 不同模型的生成图像

整体来说，相比VAE，虽然Dual-AAE也利用变分推断学习生成模型，但其同时最大化似然函数和互信息，并且引入聚类正则项，不仅提高了聚类效果，还摆脱模型崩塌的潜在风险，这为处理半监督学习和图像风格迁移问题也提供了可能的指导意义。以上工作于2020年发表在IEEE TNNLS。

责任编辑 储璟



任传贤

中山大学数学学院副教授，研究方向为统计模式识别。

Email: rchuanx@mail.sysu.edu.cn

热点追踪

面向自动驾驶场景的注意力引导实时语义分割

南京邮电大学 王林杰 周全

语义分割是计算机视觉中发展最快的领域之一，因为它提供了必要的基于对场景理解的像素级上下文，因此在自动驾驶中扮演了重要的角色。近年来，构建更大更深的卷积神经网络成为实现语义分割的主要趋势。大多数性能较好的 CNNs，如 ResNet 系列^[1,2]，都有成百上千个卷积层和特征通道来提高特征的表达能力。尽管这一方式可以使网络的性能得到提升，但会增加网络运行时间。而像无人机、机器人等这些资源受限的平台，不适合部署此类网络。

针对在资源受限环境中追求最佳分割结果这一热点问题，AGLNet (Attention-guided Lightweight Network)这一轻量化语义分割网络被提出来。它主要有三个模块(1)SS-nbt 单元；(2)GAUM 模块；(3)FAPM 模块。AGLNet 的整体架构如图 1 所示。分离重排非瓶颈块(split-shuffle-non-bottleneck, SS-nbt)通过通道分离和卷积分解来降低参数量，借助残差连接降低训练难度。分解注意力金字塔模块 (FAPM, factorized attention pyramid module)，由金字塔特征注意力

(PFA, pyramid feature attention)和全局池化注意力(GPA, global pooling attention)两部分构成。PFA 采用层次化 U 形结构，融合上下文信息；GPA 模块用于提取图像的全局先验注意力。得益于 PFA 的层次化结构，FAPM 可以捕获多尺度的上下文信息，并对特征产生像素级注意力。受文献[3]启发，全局注意力上采样模块(GAUM, global attention upsample module)，它通过对空间注意力和通道注意力进行编码，提高网络的表示能力，同时保持较低的计算成本。GAUM 和 FAPM 的具体结构如图 2 所示。

表 1 AGLNet 与其他模型在 Cityscapes 测试集上的对比

Method	Input size	Extra data	#Params	FLOPs	FPS	mIoU
ERFNet	512 × 1024	no	2.1M	26.86G	49	68.0
CGNet	512 × 1024	no	0.5M	7.01G	64	64.8
ENet	512 × 1024	no	0.36M	4.4G	65	59.3
EDANet	512 × 1024	no	0.68M	8.95G	102	67.3
ESPNetv2	512 × 1024	ImN	1.25M	5.85G	65	66.2
FSCNN	512 × 1024	Coa.	1.14M	1.76G	230	62.8
DABNet	512 × 1024	no	0.76M	10.46G	99	68.1
FPENet	512 × 1024	no	0.12M	1.58G	110	55.2
Ours	512 × 1024	no	1.12M	13.88G	52	70.1
Ours†	512 × 1024	Coa.	1.12M	13.88G	52	71.3

ImN 和 Coa 分别表示用 ImageNet 或 Cityscapes 粗糙标注数据集预训练；-表示该方法未提供相应参数

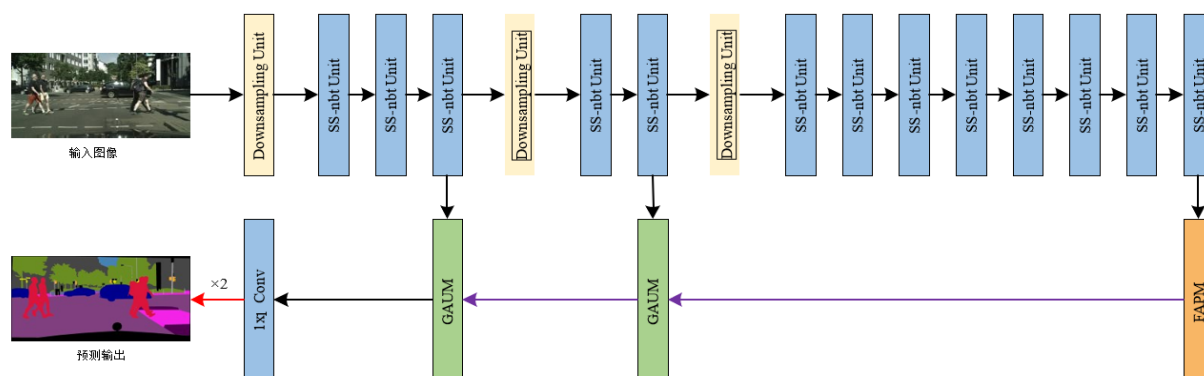


图 1 AGLNet 网络结构图

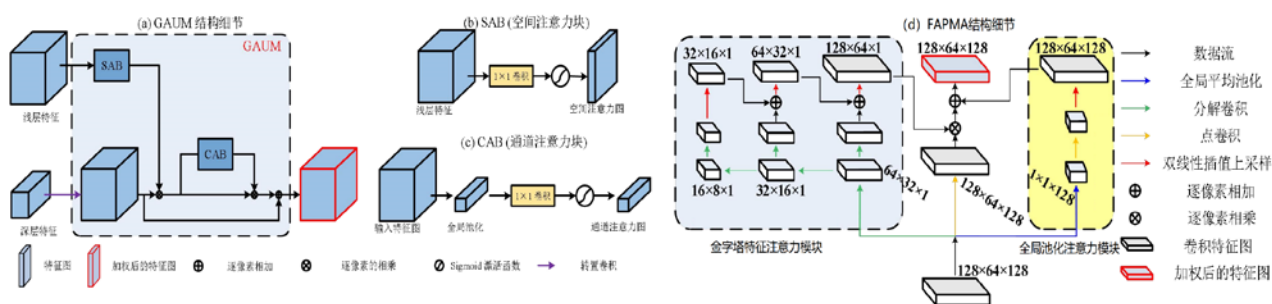


图 2 GAUM 和 FAPM 结构图

AGLNet 在 Cityscapes 测试集的实验结果如表 1 所示, 可以看出, AGLNet 的分割结果优于这些方法, 并能达到实时分割的效果。在所有方法中, AGLNet 模型大小仅 1.12M; 仅使用额外粗糙训练数据时, 分割精度提高到 71.3%。实验结果表明, AGLNet 模型在 CityScapes 数据集上实现了分割精度和运行效率间的

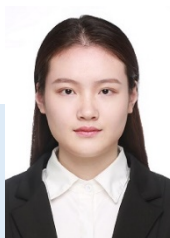
最佳平衡。

本文^[4]于 2020 年 3 月发表在人工智能权威期刊 Applied Soft Computing。

责任编辑 任桐炜

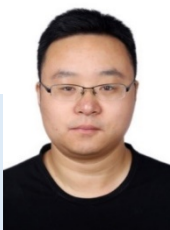
参考文献

- [1] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778.
- [2] C. Szegedy, V. Vanhoucke, S. Ioffffe, J. Shlens, Z. Wojna, Rethinking the inception architecture for computer vision, in: IEEE International Conference on Computer Vision and Pattern Recognition, 2016, pp. 2818–2826.
- [3] Z. L. Zhang, X. Y. Zhang, C. Peng, X. Y. Xue, J. Sun, Exfuse: Enhancing feature fusion for semantic segmentation, in: European Conference on Computer Vision, 2018, pp. 269–284.
- [4] Q. Zhou et al., “AGLNet: Towards real-time semantic segmentation of self-driving images via attention-guided lightweight network,” Appl. Soft Comput. J., vol. 96, p. 106682, 2020, doi: 10.1016/j.asoc.2020.106682



王林杰

南京邮电大学硕士研究生。主要研究方向为计算机视觉。
Email: linjiew678@163.com



周全

南京邮电大学副教授。主要研究方向为计算机数视觉, 深度学习, 模式识别。
Email: quan.zhou@njupt.edu.cn

顶会观察

ECCV 2020

中国科学院自动化研究所 朱翔昱

欧洲计算机视觉国际会议(European Conference on Computer Vision, ECCV)是计算机视觉领域的顶级会议,该会议每两年举办一次,与ICCV、CVPR并称为计算机视觉领域三大国际顶会。在2020谷歌学术指标的榜单上高居「工程和计算机科学」领域排名第二。ECCV的高知名度与认可度使其吸引了大量顶尖实验室、研究所以及企业机构参与其中。

ECCV 2020 原定于2020年8月23日至28日在英国格拉斯哥举行,但由于新冠肺炎的影响,会议全部转移到线上进行,给予了全新的参会体验。

一、欧洲计算机视觉国际会议的亮点

由于今年年初新冠肺炎的爆发,ECCV 2020 采用了数字虚拟会议的方式,会议主办方提前进行协调,并建立了虚拟会议的网址,让世界各地的参会者通过互联网参加虚拟会议。会议采用摘要、图片、论文和讲解视频等作为虚拟会议的交流媒介。

本次会议具有直播和预先录制两种类型,并拥有向演示者发布问题等交互手段。大多数现场活动分两次进行,相互间隔12小时,以便所有时区的参会者可以在方便的时间参加。所有的现场活动都被记录下来,在2021年6月之前可以随时查看,为不能及时参会或有时差的研究者们提供了极大的便利。本次会议在保证安全的情况下最大限度地保障了参会者的自由度和便利程度,让参与者获得了良好的参会体验。

二、论文录用情况

ECCV 2020 共收到有效投稿 5025 篇,接收 1361

篇(占比为 27%),其中 oral 论文 104 篇,仅占 2%。从总体数据来看,今年的有效投稿是 ECCV 2018(2439 篇)的两倍以上,接受率和 oral 率占比与上届差别不大。这些数据反映出了该会议的认可度迅速提升,也可看出被接受论文的质量之高。

本次 ECCV 2020 收录论文中,中国学者的表现相当出色。其中来自香港中文大学的作者人数最多,排名第二,仅次于谷歌。其他贡献突出的中国高校还包括北京大学、清华大学、上海交通大学等。除了学术界的科研力量外,来自中国工业界的各大互联网企业如百度、腾讯、旷视等也获得了非常不错的成绩。其中腾讯 AI Lab 入选 16 篇,旷视研究院入选 15 篇,百度 AI 入选 10 篇。这些被录用的论文在许多重要工业应用领域有着重大的突破,包括多模态学习、视频内容理解、自动驾驶识别及定位、声源定位及场景识别等领域。此外,谷歌在本次会议表现十分抢眼,共有 50 篇论文入选,其中 10 篇入选 oral,涵盖了诸多热门领域包括视角合成、3D 人体重建、6D 位姿估计等。

ECCV 2020 会议涵盖的方向包括图像分类、目标跟踪、人脸识别、姿态估计、3D 点云、图像/视频恢复、模型加速、GAN 和 NAS 等,论文关键词包括语义分割(Semantic Segmentation)、弱监督(Weakly Supervised)、深度估计(Depth Estimation)、对抗攻击(adversarial attack)和领域自适应(Domain Adaptation)等。

三、会议奖项

除了常规的 Best Paper 奖项,ECCV 2020 也颁发

了其它的奖项来鼓励或嘉奖对计算机视觉领域具有影响力的人物或工作。

Koenderink 奖用于表彰发表时间超过十年并经受时间检验的研究。今年 Koenderink 奖授予以下两篇论文: 1) Improving the Fisher Kernel for Large-Scale Image Classification。该论文提出的 Fisher 内核在图像分类的背景下, 被证明超越了计数统计, 扩展了流行的视觉单词包。2) BRIEF: Binary Robust Independent Elementary Features。该论文提出了一种有效的特征描述子, 能在取得比 SIFT 更好的结果前提下, 速度更快。

Mark Everingham 奖的设立是为了纪念在 2012 年去世的 Mark Everingham。他在 The PASCAL Visual Object Classes (VOC) 数据集做出了主要贡献, 也是该比赛项目的发起人。今年第一位获得 Mark Everingham 奖的是 Antonio Torralba。他持续十多年定期发布新的数据集, 包括 Tiny images, SUN / SUN - 3D, MIT - Places, 并发布了创建数据集新工具包括 LabelMe。第二位获 Mark Everingham 奖的是 COLMAP SFM 和 MVS 软件库的提出人 Johannes Schonberger, 以表彰他为 3D 图像重建提出的一个开源的端到端 pipeline, 并为之提供的支持、开发与编写文档, 这些软件库已经变成了运动结构和立体多视图的参考软件。

最佳 Demo 奖颁给了题目为 Inter-Homines: Distanc-Based Risk Estimation for Human Safety 的工作。该工作能够基于视频流实时评估监控区域内的传染风险。通过在三维空间中定位人物, 计算人际距离, 并通过构建监控区域的动态地图来预测风险水平。同时提出了一个时空动态风险估计的参数模型, 可以通过预测现场的风险水平, 对社会距离预防措施的接受程度进行安全监测, 对当前疫情的管控具有重大的实际意义。

四、会议主题及热点

本次会议热点主题包括: 1) 识别、检测、分割和姿态估计; 2) 半监督、无监督、迁移、表征和小样本学习; 3) 3D 计算机视觉与机器人; 4) 图像和视频合成; 5) 视觉

和语言。此外, 一些新兴领域, 如标签高效方法(label-efficient methods)和低级视觉(low-level vision)等, 也在逐渐引起学术界的关注。

在这些主题中, 来自学术界和工业界的多篇论文引起了广泛的讨论。如 FaceBook 提出的 DETR^[1], 是一种端到端的基于编码器-的 Transformer 架构。DETR 使用更少的先验, 如 Anchor、NMS 等, 在准确率、运行效率上与高度优化的 Faster RCNN 基本持平, 且容易复现。但缺陷是训练时间长, 在小目标上效果不是很好。Google 提出的 BiT^[2], 重新探讨了迁移学习的简单范例: 对大量带标签的源数据(例如, JFT-300M 和 ImageNet-21k 数据集)进行预训练, 然后针对不同任务对这些训练权重进行微调, 这样既减少了目标任务所需的数据量, 又减少了微调时间。香港理工和阿里达摩院提出了一种新型的梯度中心化 (gradient centralization, GC) 优化技术^[3]: 梯度中心化, 该方法通过中心化梯度向量使其达到零均值, 从而直接在梯度上执行。GC 方法可以看作是对权重空间和输出特征空间做了正则化, 从而提升了 DNN 的泛化性能。此外, GC 还能提升损失函数和梯度的 Lipschitz 属性, 从而使训练过程更加高效和稳定。东京大学提出了一种更好的增广搜索方法 Faster AutoAugment^[4]。该论文提出将原始图像和基于梯度优化的增强图像进行密度匹配, 以找到最佳策略。该研究认为数据增广可以视为填充原始数据缺点的一种方式, 并将增强数据的分布与使用了对抗学习的原始数据之间的距离最小化。加拿大高等研究院提出了一种二步法的策略^[5]来解决小样本学习微调中过度优化的问题。该论文第一步利用隐空间的相似性度量将每个样本映射到已知的基类, 第二步对特征进行关联对齐, 微调特征提取器。

此外, 最受大家关注的则是荣获最佳论文和最佳论文提名奖的论文。荣获最佳论文奖的是 Zachary Teed 等人的 RAFT^[6], 作者是普林斯顿大学视觉与学习实验室邓嘉门下的一名博士生。RAFT 提出了一种新的光流预测结构: 递归全对场变换, 通过提取每个像素的特征, 为所有像素对构建多尺度 4D 相关体, 并通过一个循环单元迭代更新光流场。RAFT 算法具有较强的跨数据集

泛化能力,在推理时间、训练速度、参数计数等方面具有较高的效率。荣获最佳提名奖的论文之一是 NeRF^[7]。NeRF 提出一种在合成复杂场景新视图任务中实现 SOTA 结果的新方法,该方法通过使用稀疏的输入视图集来优化基础的连续体场景函数。NeRF 使用了一个 9 层、256 个通道的全连接网络将场景表示为一个连续的 5D 函数,其输入是一个单一连续的 5D 坐标,即 3D 空间位置和观看方向,其输出是 RGB 颜色和不透明度(输出密度)。为了合成一个给定的视图,渲染过程包括沿摄像机射线查询 5D 坐标,并使用经典的体积渲染技术将输出颜色和密度投射到图像中。另外一篇荣获最佳提名奖的论文是[8]。该论文第一作者 Mengtian Li 是来自卡内基梅隆大学机器人研究所的博士生,曾获得南京大学学士学位。该工作提出一种明确的指标来对比不同方法的最优延迟-准确率曲线,指出标准离线评估和实时应用之间的差异:算法处理完特定图像帧时,周围环境已经发生改变,还提出将延迟和准确率协调地集成到一个度量指标中,用于实时在线感知,即流准确率。

ECCV 2020 也举办了很多高质量的研讨会与讲习班。涉及的主题包括通用物体识别分类、对抗样本、深度估计、弱监督学习、域自适应、自动化深度学习、微软的 Hololens2 产品介绍等。这些研讨会和讲习班给学者和学生提供了一个很好的学习平台。

五、总结与展望

回顾 ECCV 以及 CVPR、ICCV 近几年的论文,可以发现视觉领域的几大趋势:1)自主学习,包括网络结构、数据增广、损失函数、超参的搜索等;2)无监督/半监督/自监督学习,一大关注点是如何利用大量无标签数据学习一个好的表征;3)3D 物体感知和识别,落地场景有无人驾驶、虚拟现实等。此外,受到疫情的影响,线上虚拟会议也可能是以后几年的趋势,此次 ECCV 2020 虽然搭建了虚拟的在线平台来提升体验,但依然存在很多非热门的论文的会议室无人问津的冷场情形。如何提高线上会议的体验,促进学者的交流热情,是线上会议的下一个挑战。

责任编辑 崔海楠

参考文献

- [1] Carion, Nicolas, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. "End-to-End Object Detection with Transformers." In ECCV, 2020.
- [2] Kolesnikov, Alexander, Lucas Beyer, Xiaohua Zhai, Joan Puigcerver, Jessica Yung, Sylvain Gelly, and Neil Houlsby. "Big transfer (BiT): General visual representation learning." In ECCV, 2020.
- [3] Yong, Hongwei, Jianqiang Huang, Xiansheng Hua, and Lei Zhang. "Gradient Centralization: A New Optimization Technique for Deep Neural Networks." In ECCV, 2020.
- [4] Teed, Zachary, and Jia Deng. "RAFT: Recurrent All-Pairs Field Transforms for Optical Flow." In ECCV, 2020.
- [5] Hataya, Ryuichiro, Jan Zdenek, Kazuki Yoshizoe, and Hideki Nakayama. "Faster autoaugment: Learning augmentation strategies using backpropagation." In ECCV, 2020.
- [6] Afrasiyabi, Arman, Jean-François Lalonde, and Christian Gagné. "Associative Alignment for Few-shot Image Classification." In ECCV, 2020.
- [7] Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis." In ECCV, 2020.
- [8] Li, Mengtian, Yu-Xiong Wang, and Deva Ramanan. "Towards Streaming Image Understanding." In ECCV, 2020.



朱翔昱

中国科学院自动化研究所副研究员。主要研究方向为计算机视觉和三维重建。
Email: xiangyu.zhu@nlpr.ia.ac.cn

顶会观察

NeurIPS 2020

北京大学 ZERO Lab 林宙辰 王奕森

神经信息处理系统大会 (Conference on Neural Information Processing Systems, NeurIPS) 是机器学习领域的顶级会议，是神经计算方面最好的会议之一，在中国计算机学会推荐国际学术会议中被评为人工智能领域的 A 类会议。在 Google Scholar 发布的 2020 年学术指标中，h5 指数高达 198，位于计算机领域的第 3 位、位列所有领域出版物的第 21 位。

今年第 34 届 NeurIPS 原定于 12 月 6 日到 12 日在加拿大温哥华举办。因为疫情原因，NeurIPS 2020 变成了一届纯线上会议。

一、NeurIPS 2020 的亮点

由于新冠疫情的持续蔓延，NeurIPS 2020 完全采取了虚拟线上会议的形式，让世界各地的参会者通过互联网参会交流。为了提升线上会议的参与体验，本次 NeurIPS 大会组委会专程邀请了两名在线体验主席，并在 poster 环节使用了名为 GatherTown 的交互网站，在模拟线下交流上十分逼真。除此之外，会议注册费用大幅下降。学生注册费用为 25 美元，其他参会者则为 100 美元，而去年价格分别为 420 和 750 美元。值得注意的是，所有人都可以在无需注册的情况下访问 Tutorial、Keynote、Oral 论文和 Spotlight 论文的宣讲内容^[1]。在会议结束后，NeurIPS 2020 预计还会将 Poster 的录像公开。

除了涵盖人工智能领域丰富的主题，NeurIPS 2020 还开始探讨机器学习的社会影响，例如投稿论文需要添加 Broad Impact、审稿流程中引入道德审查机制。不过仅有 290 篇投稿论文被审稿人提出了道德担忧，

主要集中在虚假人脸的生成、新数据集的隐私泄露风险等。同时，大会邀请报告讨论了如何避免机器学习系统的不良社会影响和关注数据标注员的职业发展问题。今年还特意举办了关于新冠病毒 COVID-19 的特别研讨会。这体现了人工智能社区和科技工作者的社会责任感。

二、论文录用情况

NeurIPS 2020 的投稿量再创新高，共有 9467 篇论文被提交，相比 2019 年增长了 40%。经过 7062 名审稿人的同行评议，最终有 1898 篇论文入选，录取率仅为 20.0%，低于去年的 21.6%，达到历史最低。其中 105 篇被录用为 Oral，280 篇被录用为 Spotlight，Oral 和 Spotlight 的录取率仅有 1.11% 和 2.96%。

在本次 NeurIPS 2020 中，谷歌位于所有高校/机构的榜首，入选论文高达 202 篇。麻省理工学院排名第二，共有 109 篇论文入选。斯坦福大学位居第三，共有 106 篇论文入选。国内高校/机构中，清华大学排名第一，共有 63 篇论文入选，排在国内外所有高校/机构的第 7 位；北京大学第二，共有 37 篇论文入选。从作者角度，有华人参与的论文共有 568 篇，占总论文数的 29.91%。其中人为第一作者的论文共有 321 篇，占华人参与论文数的 56.51%，占总论文数的 16.90%。华人在 NeurIPS 2020 中贡献了重要力量。

NeurIPS 作为机器学习领域的顶级会议，收录的论文包罗了人工智能领域的各种主题，包括深度学习及其应用、强化学习和规划、纯理论研究、概率方法、优化、机器学习和社会、神经科学和认知科学等方方面面。

三、邀请报告

今年的 NeurIPS 2020 共有七个邀请报告 (Invited Talk), 主题丰富, 涵盖了密码学、反馈控制理论、因果推断、生物学等多个学科, 反映了机器学习正与其他学科不断影响并相互交融; 同时探讨了人工智能大潮下的多个社会问题, 展现了科研人员的社会责任感。2012 年的图灵奖得主、MIT 的密码学教授 Shafi Goldwasser 展示了在密码学的意义下, 机器学习正面临的三大挑战。阿卜杜拉国王科技大学(KAUST)的教授 Jeff Shamma 重点介绍了一些反馈控制概念, 特别是鲁棒性、被动性、跟踪和稳定性, 因为它们与进化博弈论、无遗憾学习和多智能体学习中的特定问题相关。冷泉港实验室的神经科学家 Anthony M. Zado 教授报告了从生物学中借鉴本能瓶颈的想法。微软研究院的 Chris Bishop 回顾具有真实应用场景的微软研究院的研究项目, 分享了他对于 AI 未来发展趋势和现实世界影响的思考。卡内基梅隆大学的客座教授 Saiph Savage 指出人工智能产业在不断创造出新的工作的同时, 也给人们带来了许多冲击和挑战。美国国家科学院 Kavli Fellow 获得者、乔治亚理工学院计算机学院 Charles Isbell 教授提到社会需要采用系统的方法来研发鲁棒性的对人们生活具有重大影响的产品。

四、会议热点论文

本次会议涌现了许多优秀的工作, 它们具有非常高的学术价值与应用价值。NeurIPS 2020 的最佳论文奖由三篇论文摘得, 分别是自米兰理工大学的《No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium》、UC 伯克利的《Improved Guarantees and a Multiple-Descent Curve for Column Subset Selection and the Nyström Method》, 以及 OpenAI 的《Language Models are Few-Shot Learners》(即著名的 GPT-3)。

No-Regret Learning Dynamics for Extensive-Form Correlated Equilibrium. 相关平衡(CE)易于计算, 并且可以获得比众所周知的纳什平衡要高得多的社会福利。在普通形式的游戏中, CE 的一个功能是可以简单且分散的算法将特定遗憾的概念

降到最低, 从而找到 CE。论文说明了在大型游戏(即扩展型(或树型)游戏)中收敛到 CE 的遗憾最小化算法的存在。该结果解决了博弈论、计算机科学和经济学界的一个长期存在的开放性问题, 并且可能对涉及调解员的游戏产生重大影响, 例如, 通过导航应用程序对有效的交通路线产生影响。

Improved Guarantees and a Multiple-Descent Curve for Column Subset Selection and the Nyström Method. 从大矩阵中选择列向量的一个小但有代表性的子集是一个困难的组合问题, 并且已知一种基于基数受限的行列式点过程的方法可以提供一种实用的近似解决方案。论文针对最佳可能的低秩近似, 推导了近似解的近似因子的新上限和下限, 该上限甚至可以捕获子集大小的多重下降行为。论文将分析进一步扩展, 获得了 Nyström 方法的理论保证。由于这些逼近技术已在机器学习中得到广泛应用, 因此这篇论文有望产生重大影响, 并为核方法、特征选择和神经网络的双下降行为等理论和现象提供新的见解。

Language Models are Few-Shot Learners. AI 大厂向来奉行大力出奇迹, OpenAI 的 GPT-3 工作更是将这个理念推到极致, 坚持 Bengio 2003 年提出的简简单单的 Neural Language Model, 加上海量的数据和算力, 实现了前所未有的表示能力。论文向我们清晰地展示了, 只要算力足够, 神经网络还有相当大的空间可以开发, 而我们通常讨论的数据集和模型大小, 离达到神经网络的极限仍然很远。

此次 NeurIPS 2020 一共收录了 105 篇 Oral 论文和 280 篇 Spotlight 论文, 大会将它们分成了 35 个 Track, 包括: Representation / Relational, Language / Audio Applications, Reinforcement Learning, Clustering / Ranking, Dynamical Sys / Density / Sparsity, Deep Learning, Vision Applications, Deep Learning / Theory, Continual / Meta / Misc Learning, Social / Adversarial Learning, Graph / Meta Learning / Software, Graph / Relational / Theory, Unsupervised / Probabilistic.

Oral 论文仅仅是录用论文的冰山一角。其他录用论文中还有大量值得探索和讨论的内容。

除此之外，大会还举办了 30 个教程(Tutorials)和多个研讨会(Workshops)，包括离线强化学习、生物学与 AI、鲁棒机器学习、深度隐私模型等丰富的主题。这些教程和研讨会，为科研人员扩展视野和交流互动提供了便利的渠道和平台，利于机器学习领域的长期发展。

五、总结与展望

在 NeurIPS 2020 的投稿论文中，算法(占投稿总量的 29%)、深度学习(19%)及其应用(18%)、强化学习(9%)等依旧是机器学习和计算神经科学的热门主题。值得注意的是，探讨机器学习的社会层面问题的论文增长迅速，由去年的 2% 上升到今年的 5%。机器学习得到了越来越多的应用，如何避免模型可能带来的不良社会影响(公平、隐私等)成为不可忽视的问题。

在新冠病毒全球肆虐的大背景下，各类人工智能会议不得不转到线上举行。从最初 ICLR 2020 每篇论文分配独立的 Zoom 会议室，到 NeurIPS 2020 使用 GatherTown 举办 Poster Session，线上虚拟会议的体验持续提升。随着虚拟会议技术的持续提升，虚拟会议的体验有可能逼近线下会议，并且注册费低得多，有利于扩大参会人数，或许会成为未来科研交流的另一种重要形式。

致谢：本文由林宙辰、王奕森组稿，凌泽南、陈勇、王一飞、何翎申、李明杰、徐大鹏、孔浩、耿正阳、杨一博、沈铮阳、徐鑫参与写稿。关于 ZERO Lab 的更多信息见 <https://zero-lab-pku.github.io/>。如需了解更多关于 NeurIPS 2020 的论文介绍，见微信公众号链接 <https://mp.weixin.qq.com/s/BaUIZEuuTCjFGqg5sa3x A>。

责任编辑 王金甲

参考文献

[1] <https://proceedings.neurips.cc/paper/2020>



林宙辰

北京大学信息科学技术学院智能科学系教授，国家杰青，IAPR/IEEE Fellow。研究方向为机器学习、计算机视觉。担任中国图象图形学学会机器视觉专委会主任、中国自动化学会模式识别与机器智能专委会副主任、中国计算机学会计算机视觉专委会常务委员、中国人工智能学会模式识别专委会常务委员。

Email: zlin@pku.edu.cn



王奕森

北京大学信息科学技术学院智能科学系助理教授。主要研究方向为机器学习，特别是对抗学习、图学习。

Email: yisen.wang@pku.edu.cn

南京理工大学张姗姗教授访谈

2020年11月11日,《CCF-CV专委简报》在线采访了南京理工大学计算机科学与工程学院博士生导师张姗姗教授。下面是采访实录。

张老师,您好!我们注意到您的经历非常丰富,在国内读的本科和硕士,德国读的博士,然后又分别在德国、日本和美国从事了研究工作。能否首先请您介绍一下您的个人学习和研究经历?

我本科、硕士曾就读于同济大学电信学院,硕士期间曾在导师的推荐下前往日本国立情报学研究所学习访问。硕士期间的学习经历,特别是在日本的访学经历,激发了我继续从事科学研究,以及走出国门多看看的强烈愿望,于是硕士毕业后我便前往德国知名学府波恩大学计算机系攻读博士学位。博士毕业后有幸进入德国顶尖的科研机构马普所从事博士后研究。回国后,我曾在学校的支持下前往美国加州大学伯克利分校访问。在多个国家的学习和工作经历使我受益良多,得到了多位德高望重的老师的指导,认识了很多优秀的朋友,我感到自己的各方面的能力都得到很大提高。

您是1987年出生的,非常的年轻,现在就已经是博导、教授,有丰富的科研经历且取得了很多科研成果,能否跟大家分享一下您的快速成长历程,以及您的成长感悟?

我出生于江西的一个小县城,在我上大学之前都没有高速公路,从我们县城沿着盘山公路开车需要五六个

小时才能到达省城南昌,2004年考上大学我才第一次走出大山,来到繁华的大上海。来自偏远山区的我特别珍惜这来之不易的受教育的机会,深知这是自己唯一的出路。在后续的学习和工作中,我一直十分努力,不敢懈怠。同时,我觉得自己特别幸运,一路走来得到多位师长和领导的关心、帮助和支持,包括我的硕士导师刘富强教授、博士生导师 Armin B. Cremers 教授、博士后合作导师 Bernt Schiele 教授以及我目前所在团队的负责人杨健教授。他们为我提供了良好的科研环境,并且不断地鼓励我、帮助我、支持我,使得我能够沉下心来专注地做好自己的科研课题。我非常感谢他们!

在我的科研道路上,我最深切的体会是,作为年轻人,首先要立足于一个小的领域,深耕其中,不甘寂寞,踏实前行。正如,沈向洋老师最近在他的报告中分享的“Before you go broad, go deep”。从博士以来,我一直从事行人检测方面的研究,至今已有十年之久。其实,在我博士后工作初期,我曾经考虑要换一个课题,可是当时我的导师 Bernt Schiele 教授鼓励我说,任何一个课题,你起码要研究十年以上才可能成为这方面的专家。在他的鼓励下,我继续深入这方面的研究,并且通过实验分析发现现有的算法在很多困难场景下,检测性能还远远不能满足实际应用的需求。在接下来的几年里,我致力于解决困难场景下的行人检测问题,如遮挡、夜间环境、跨域检测等,取得了一系列新的成果。功夫不负有心人,我在微软学术统计的全球学者近五年排位中,位列行人检测领域第一位。

您在行人检测方面做了非常多的研究工作，尤其是抗遮挡方面，能否跟大家分享下这方面的研究及应用现状？您认为未来几十年行人检测会朝着什么方面发展呢？

行人检测是一个传统的研究课题，自深度学习兴起之前就已经被广泛研究。深度学习兴起之后，行人检测依然独立于通用目标检测，受到持续关注，比如今年 CVPR 投稿系统中将其单列为一个子领域。一方面，行人检测在自动驾驶、视频监控、智能机器人等诸多领域有着极其广泛的应用背景；另一方面，行人作为一种典型的目标类别，有很多区别于其他物体的显著特征，使得我们在设计行人检测算法的时候可以利用很多先验知识来提高精度。

目前我们研发的行人检测算法已经被丰田汽车公司用于无人驾驶实验车上，我们看到基于视觉的行人检测方法在一般理想环境下能够取得不错的检测效果。然而，在一些困难场景下，如遮挡、恶劣天气、夜间等情况下的检测精度还有待提高。

近年来，我们在抗遮挡方面做了一系列工作。2017 年我们提出了一个新的行人检测数据集 CityPersons，该数据集具有较强的多样性和丰富的遮挡实例，我们利用数据驱动的方法提高了现有模型应对遮挡场景的能力。该数据集目前已成为行人检测领域的标准、权威数据集，在领域内被广泛采用。最近两年，我们研究了如何自适应地对人体遮挡模式进行建模，从而提高被遮挡人体特征的辨识度，如卷积通道间注意力机制、图卷积网络信息传递等。但是我们感到仅利用单帧图像处理遮挡问题具有一定的局限性，因为被遮挡部分信息缺失难以得到补偿。未来我们将利用多视角图像对室外人体进行三维重建，并利用全景理解以及视频帧协同决策等技术借助上下文信息来进一步解决遮挡问题。

除了行人检测，您目前的研究聚焦点还有哪些？在研究方面，您未来的规划是怎样的？

回国工作后，我开始逐渐扩展自己的研究方向。由于我多年来从事行人检测方面的研究，我对图像视频中的人体分析技术十分感兴趣。目前我的研究课题较广泛地涉及到行人检测、行人重识别、行人搜寻、人体语义解析、人体姿态估计等，但我始终坚持“广而不散”的原则，即我们围绕着“以人为中心的视觉感知”这一主题开展研究，所有的研究课题都与人有关。未来随着我们小组规模的逐步扩大，我们将开展更多的与人有关的视觉感知方面的研究。

您目前的这些研究工作中，请问哪一项是您认为最自豪的？能否具体介绍一下？

我们 2018 年发表在 T-PAMI 上的工作《Towards Reaching Human Performance in Pedestrian Detection》，是我目前为止最喜欢的一项工作。在这个工作中，我们极大地提高了行人检测的性能，在很多场景下使其接近人体视觉系统所能达到的精度。和以往工作不一样的是，我们提出的新算法是依据误差分析而提出的。首先，我们提出了对行人检测器的误差进行定性-定量层次化分析的方法，并根据定量分析的结论，找到影响性能的关键因素，从而提出有针对性的解决方案。在这个工作中，我们所提出的分析方法被后续工作沿用，同时也在工业界得到推广。此外，在完成这个工作过程中，我深切地体会到从误差源出发寻找解决方案是一种很好的思路，可以让我们在设计算法的过程中少走很多弯路，这对我们后续的其他工作都具有指导意义。

您带领团队参加了一些挑战赛并获得了奖项，如：2018 年中国智能车未来挑战赛“前方车辆位置检测”第一名和“跨越险阻 2018”陆上无人系统挑战赛环境感知组第二名，能跟大家分享一下参加这些挑战赛的经验、经历和心得么？

我个人认为，参加比赛是将我们在实验室里研究的算法落地的重要实践机会。理论算法研究通常只在有限的公共数据集上进行测试，而实际场景复杂多变，我们需要具体问题具体分析具体解决，这绝不只是一个简单的工程化的过程。更重要的是，在实际场景中遇到的困难可以帮助我们提炼出新的科学问题，找到新的突破点，做出更多有影响力的工作。

您担任 Pattern Recognition (PR) 期刊的编委，PAMI、IJCV、CVPR 等多个知名期刊和会议的审稿人，请问您是如何评价学术论文的？您能否对大家投稿这些顶尖期刊及会议论文提供一些建议？

我们 VALSE 曾在今年七月份举办过一期《如何写/审 AI 领域的论文》的专题论坛，当时罗杰波老师和杨明玄 (Ming-Hsuan Yang) 老师等多位资深学者给大家分享了很多关于撰写论文和评价论文方面的宝贵经验，我也从中受益良多。在我看来，首先，学术论文是研究者之间互相交流的一个重要工具，因此我们有必要重视交流的方法和技巧，我们要站在一个完全不了解自己工作的读者的角度来审视自己的工作，学会用最简洁的语言和形象的图表来传递自己的研究思想；同时，学术论文也是我们作者和自己的研究工作对话的一个窗口。我常常有这样的感受，在初期的实验过程中觉得自己的 idea 很赞很妙，可是一下笔，总能发现其中的不足之处，于是再改进，这样不断改进和行文的迭代过程能够帮助我们将研究工作做得更加完备。

能否介绍下您的团队及您对团队研究生的管理情况？

我的研究小组隶属于杨健教授的团队 PCALab，目前我的小组共有 7 名博士生，6 名硕士生，以及若干本科实习生。得益于我们整个 PCALab 良好的科研氛围，不少硕士生都在二年级的时候转博了，所以目前博士生

数量比硕士生还要多。

我按照学生的研究课题，将所有学生分成三个小组，每个小组每周一次组会，我和学生们一起讨论最近的研究进展。我鼓励学生们提出不同的看法，时常会有激烈的争论，当我听到他们独特的见解时，总是感到很欣慰。同时，我也鼓励同学们跨小组参加组会，通过了解其他课题的研究近况，常常也会对自己的课题有所启发。

大家对您研究之外的生活也比较感兴趣，比如您的兴趣爱好之类的，能分享下么？

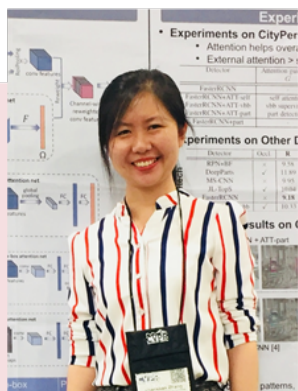
科研之余，我们小组也经常组织团建活动，如爬山等。我和学生们年龄差距比较小，他们就像我的弟弟妹妹一样，我们相处起来十分融洽。

我一直是一个热爱生活的人。在研究工作之余，我积极参加户外运动，多多呼吸新鲜空气，不仅心情舒畅，有的时候还会突然有好的 idea 蹦出来。除此之外，我还爱好瑜伽、舞蹈、阅读和旅行。我在博士期间曾经选修过拉丁舞的课程，回国工作后也定期在学校工会参加形体和舞蹈方面的培训。我想只有保持身心健康，才能更高效地工作。

如果吐露研究工作者的的心声，您最想说的的是什么？

回国工作四年来，我深切地感受到，我们国家对青年科研工作者的支持力度非常大，为我们创造了良好的科研环境，同时，业界的很多前辈也对我们年轻人非常关心和支持，在此，我深表感谢。我很庆幸自己当年选择回国工作，这几年来，虽然工作辛苦，但我很欣喜地感受到自己一步一个脚印不断进步，希望自己将来在“以人为本的视觉感知”方面做出更多更好的工作。

责任编辑 余焯 赵振兵



张姗姗

南京理工大学计算机学院教授，博士生导师，江苏省“社会安全图像与视频理解”重点实验室副主任。研究方向为计算机视觉，特别关注视频中人体检测和分析。曾于2015年2月在德国波恩大学获得计算机博士学位，博士论文获大德语区（德、瑞、奥、三）计算机学会最佳博士论文提名；博士毕业后在德国马普计算机研究所从事为期两年的博士后研究。2018年入选中国科协“青年人才托举工程”、江苏省“双创博士”、微软“铸星学者”计划等。目前以第一/通讯作者身份在 CVPR、ECCV、T-PAMI、T-IP 等顶级国际会议和国际期刊上发表论文四十余篇，谷歌学术引用1900余次；在微软学术统计的全球学者近五年排位中，位列行人检测领域第一位。目前担任模式识别权威期刊 Pattern Recognition 编委、中国人工智能学会模式识别专委会副秘书长、VALSE 资深领域主席。

委员好消息

❖ 2020年9月19日，第22届国际工业博览会落幕，CCF-CV专委会委员、南开大学程明明教授主持研制的基于可解释人工智能技术的新冠肺炎肺部CT影响筛查系统获优秀展品奖。

❖ 2020年9月27日，CCF-CV专委会共21位委员晋升为CCF高级会员，他们是：江苏大学苟建平副教授、天津大学冀中副教授、复旦大学金城教授、北京电子科技学院金鑫副教授、中科院自动化所雷震研究员和申抒含研究员、兰州理工大学李策教授、电子科技大学李永杰教授、上海交通大学林巍峭教授、北京大学林宙辰教授、北京理工大学刘越教授、北京航空航天大学陆峰教授、贵州师范大学欧卫华副教授、云南大学钱文华教授、中山大学王昌栋副教授、微软亚洲研究院王井东首席研究员、中国海洋大学王胜科副教授、武汉大学夏桂松教授、大连理工大学杨鑫教授、中国矿业大学姚睿副教授、南京邮电大学周全副教授。

❖ 2020年10月16日，CCF-CV专委会会议举办颁奖仪式。CCF-CV专委会主任、北京大学查红彬教授获得CCF-CV杰出成就奖。CCV-CF专委会委员、哈尔滨工业大学左旺孟教授、合肥工业大学余烨副教授和南京理工大学潘金山教授获得CCF-CV中科视拓Seeta服务贡献奖。

❖ 2020年10月20日，2020年度CCF科学技术奖评选结果公布，CCF-CV专委会常务委员、北京大学林宙辰教授主持的“高维复杂数据的低秩模型理论与方法”项目获自然科学一等奖，CCF-CV专委会副主任、南京信息工程大学刘青山教授参与的“高效视觉目标跟踪理论与方法”项目获自然科学二等奖。

❖ 2020年11月13日，第六届“中国科普作家协会优秀科普作品奖”获奖作品名单正式发布，CCF-CV专委会委员、复旦大学张军平教授所著的《爱犯错的智能体》获科普图书类金奖。

❖ 2020年11月24日，IEEE2021年度新增会士名单出炉，CCF-CV专委会副主任、上海科技大学虞晶怡教授因在计算摄影学和计算机视觉理论、算法与系统领域的突出贡献当选。

❖ 2020年11月24日，教育部公布了首批国家级一流本科课程认定结果。CCF-CV专委会委员、哈尔滨工程大学刘海波教授负责的《物联网工程导论》获评线上一流课程，安徽大学罗斌教授负责和汤进等参与的《数字图像处理(双语)》、湘潭大学欧阳建权教授参与的《信息安全》和同济大学赵才荣副研究员参与的《人工智能原理》获评线下一流课程，西安电子科技大学邓成教授参与的《数字电路与系统设计》获评线上线下混合式一流课程，中北大学秦品乐教授负责的《“智能移动互联网+”创新创业实践课》获评社会实践一流课程。

❖ 2020年11月26日，2020年度黑龙江省科学技术奖励结果公布，CCF-CV专委会委员、哈尔滨工业大学左旺孟教授、中山大学林惊教授等完成的“基于模型引导的深度视觉计算理论与学习”获自然科学一等奖。

❖ 2020年12月3日，2020年中国图象图形学学会石青云女科学家奖评审结果公布，CCF-CV专委会委员、中科院自动化所董晶副研究员荣获青年组石青云女科学家奖。

❖ 2020年12月3日，2020年度中国图象图形学学会优秀博士学位论文奖评选结果公布，CCF-CV专委会委员、厦门大学纪荣嵘教授指导的《面向视觉搜索的紧凑视觉表征方法与研究》和西北工业大学韩军伟教授指导的《基于深度神经网络的图像显著性检测关键技术研究》获奖。

❖ 2020年12月4日，2020年度中国图象图形学学会科技奖励评选结果公布。CCF-CV专委会委员、中山大学郑伟诗教授、赖剑煌教授等完成的“跨时空移动目标动态追踪基础理论与方法研究”荣获自然科学一等奖，

中科院计算所蒋树强研究员、中国科学院大学黄庆明教授等完成的“视觉媒体的局部关联分析与表示”、重庆邮电大学高新波教授（在西安电子科技大学）参与完成的“属性监督的异质人脸图像合成与识别理论与方法”、山东大学张伟教授参与完成的“视觉信息质量评价理论与方法”荣获自然科学二等奖，北京科技大学马惠敏教授等完成的“无人系统复杂环境目标认知关键技术及应用”荣获技术发明一等奖，北京工业大学毋立芳教授、简萌老师等完成的“面向3D打印的多光源自主融合和模型适配控制方法”荣获技术发明二等奖。

❖ 2020年12月7日，陕西省科技奖励评审结果公示，CCF-CV专委会委员、西安电子科技大学苗启广教授等完成的“彩色地形图智能解译模型和优化理论”、西安交通大学孙剑教授等完成的“基于模型驱动与先验发现的视觉计算新理论与新方法”、西安电子科技大学董伟生教授等参与完成的“图像质量评价及内容重建方法与技术”获得自然科学一等奖。

❖ 2020年12月8日，2020年度中国电子学会科学技术奖名单公布，CCF-CV专委会委员、清华大学鲁继文副教授、中科院自动化所樊彬副研究员、国防科大刘丽副教授和清华大学郭振华副研究员等完成的“视觉信息的紧致特征表示理论与方法”获自然科学一等奖，北京大学彭宇新教授等完成的“网络多媒体内容理解关键技术及应用”获科技进步一等奖，上海海事大学周日贵教授等完成的“复杂地铁供电系统安全高效智能运维关键技术研发及应用”获科技进步三等奖。

❖ 2020年12月9日，北京市2020年度“杰青”名单正式公布，41项获资助。CCF-CV专委会委员、北京交通大学张淳杰教授的“基于深度图卷积网络的少样本细粒度图像分类”获得资助。

❖ 2020年12月10日，第十届吴文俊人工智能科学技术奖获奖名单公布，CCF-CV专委会委员共13人获奖。中国科学院大学黄庆明教授和中国科学院计算技术研究所蒋树强研究员等完成的“图像视频的多尺度表征与语义映射”和北京师范大学鄂霞教授等完成的“基于智能计算的脑机制研究”获自然科学一等奖，南开大学程明明教授和杨巨峰教授等完成的“图像场景理解与内容敏感图像处理”获自然科学二等奖，电子科技大学李永杰教授和四川大学高绍兵副研究员等完成的“受生物视觉启发的智能感知理论与算法”获自然科学三等奖，清华大学深圳国际研究生院郭振华副研究员等完成的“基于时序知识图谱的金融风控与经济量化决策平台产业化”获科技进步一等奖，上海交通大学林巍峭教授等完成的“语义驱动的视频智能分析及适配传输关键技术与应用”和复旦大学金城教授等完成的“智能视频计算关键技术及其在公共安全中的应用”获科技进步二等奖，复旦大学张军平教授的著作《爱犯错的智能体》获科技进步奖（科普项目），上海交通大学严骏驰研究员和中山大学李冠彬副教授获优秀青年奖。

责任编辑 刘海波

双目立体匹配网络开源代码

东北大学 贾 同

计算机视觉的研究让计算机可以通过二维图像信息认知三维环境信息的能力，双目立体视觉是计算机视觉的一种重要形式，能够以低成本从双目图像中便捷恢复场景深度，在自动驾驶、人脸识别、物体检测等方面都有广泛的应用。人们可以使用被测物体的两幅图像，利用立体匹配算法找出图像中相同像素的几何关系，计算图像对应点间的位置偏差来获取物体三维几何信息，因此设计一种精准获取 3D 信息的双目立体匹配算法十分重要。

基于深度学习的立体匹配网络是双目立体视觉中的一个热门方向。相较于手工设计函数的传统立体匹配方法，基于卷积神经网络的端到端立体匹配网络在弱纹理和遮挡区域的表现更佳，能够在学习大量数据的基础上得到更准确的视差结果。自 2015 年提出的立体匹配网络 FlowNet，立体匹配网络已经发展出各种高精度的模型。本文将重点介绍几个开源的双目立体匹配网络模型。

1、End-to-End Learning of Geometry and Context for Deep Stereo Regression

工作：该论文提出的深度学习体系结构——一个端到端的立体匹配网络 GC-Net，用于从一对经校正的立体图像回归视差。如图 1，该网络利用卷积神经网络提取图像对的深层特征表示，根据极线约束原则生成代价体，并在此代价体上使用 3D 卷积，利用编码器-解码器架构结合上下文信息。使用一种可微的 Soft Argmin 操作从代价体回归视差值，这保证了端到端网络训练时的亚像

素精度，并且不需要任何额外的后处理或正则化。

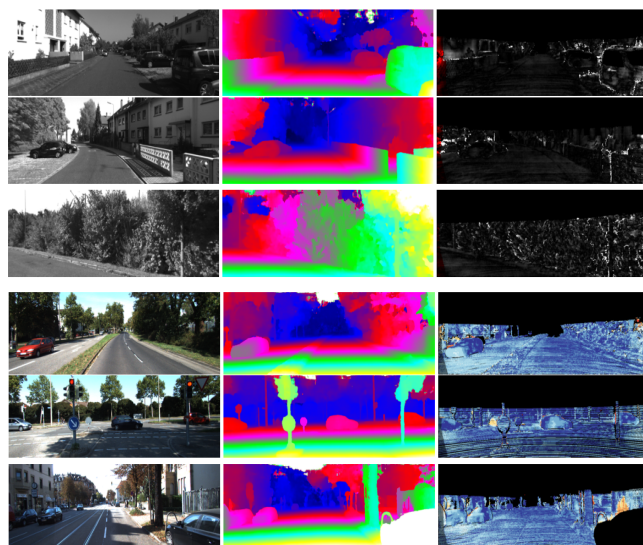
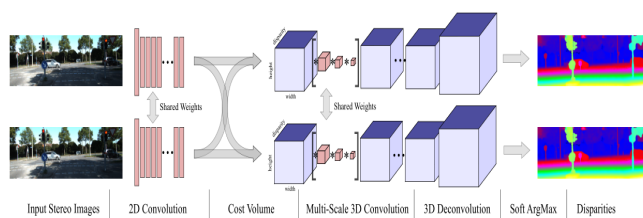


图 1 GC-Net 网络结构及网络预测的深度图和误差图

论文：<https://arxiv.org/abs/1703.04309>

代码：<https://github.com/kelkelcheng/GC-Net-Tensorflow>

2、StereoNet: Guided Hierarchical Refinement for Real-Time Edge-Aware Depth Prediction

工作：该论文提出了一个实时立体匹配的端到端深度神

神经网络模型 StereoNet，其能够通过使用一个非常低分辨率的代价体来实现实时性能。如图 2 所示，模型使用孪生网络从左右图像中提取特征，利用降采样大大降低特征图的分辨率，通过生成的代价体预测视差值，然后利用分层模型学习一个上采样函数，该函数以颜色输入为指导，分层次地恢复高频细节，能够产生高质量的边缘感知输出。

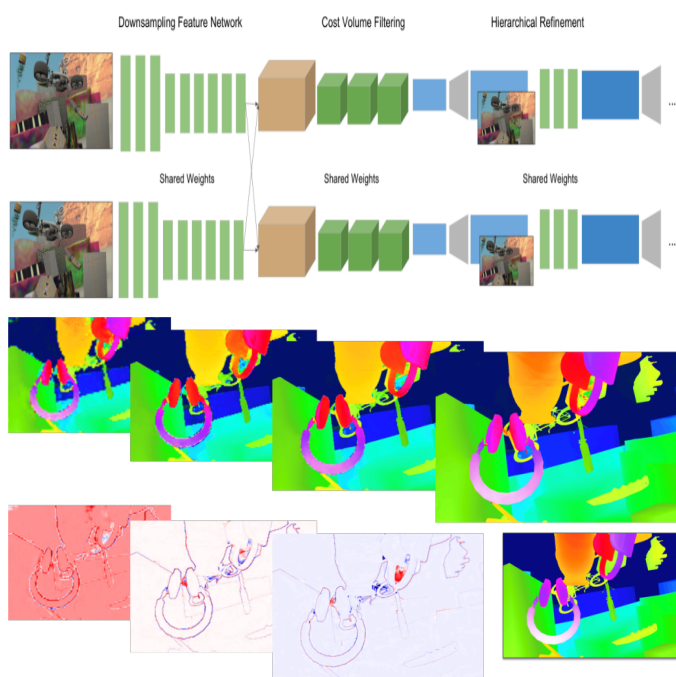


图 2 StereoNet 的网络结构和深度图的生成效果

论文: <https://arxiv.org/abs/1807.08865>

代码: <https://github.com/meteorshowers/StereoNet>

3、Real-time Self-adaptive Deep Stereo

工作: 针对当前端对端有监督训练的立体匹配模型在与训练集明显不同的场景中准确性显著下降的问题，该论文研究对深度立体网络进行无监督和连续在线适应，让模型在任何环境下都保持准确性。该论文提出了模块化自适应网络 MADNet，并且开发了一种模块化自适应 (MAD) 算法，该算法独立地训练网络的子部分。图 3 展示了训练全局网络与训练网络子部分的自适应网络结构以及不同结构预测的深度图效果，结果表明自适应网

络 MADNet 能够在异构数据集上具有竞争性的性能，而将 MADNet 与 MAD 结合实现了新的轻量级、高效的深度立体结构。

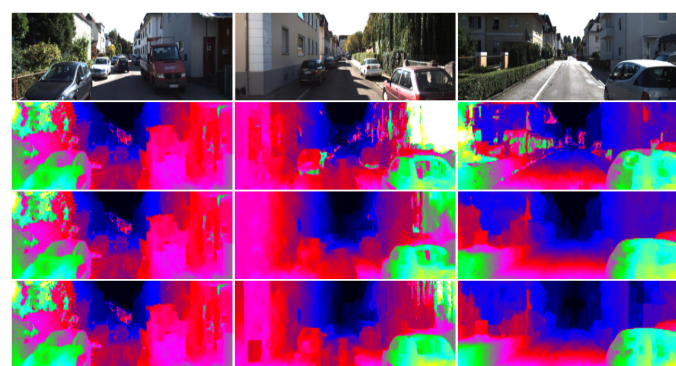
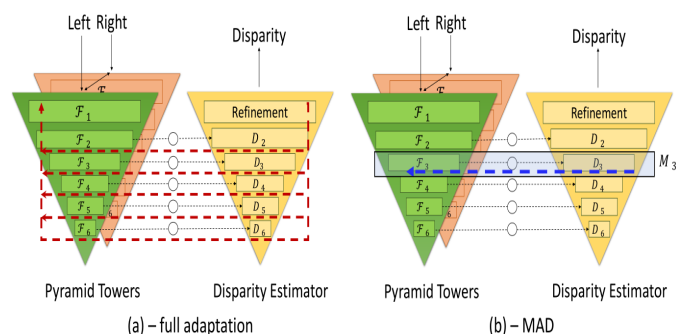


图 3 MADNet 不同算法下的网络结构和深度图效果

论文: <https://arxiv.org/abs/1810.05424>

代码: <https://github.com/CVLAB-Unibo/Real-time-self-adaptive-deep-stereo>

4、Pyramid Stereo Matching Network

工作: 该论文研究另一种用端到端的卷积神经网络解决深度估计任务：在采用编码-解码架构聚合全局上下文信息的同时，在立体匹配过程中聚合不同尺度、不同位置的上下文信息。论文提出了一种金字塔立体匹配网络 PSMNet，如图 4 所示，它由两个主要模块组成：空间金字塔池和 3D 卷积。空间金字塔汇聚模块将像素级特征扩展为拥有不同尺度感受野的区域级特征，结合全局和局部特征线索形成可靠视差估计的代价量。3D 卷积堆叠多个沙漏网络，采用中间监督方法，以自顶向下或自底向上的方式重复处理代价体，进一步提高全局上下文信息的利用率。图 4 中展示了该网络估计的视差图，

PSMNet 显著减少了不适定区。

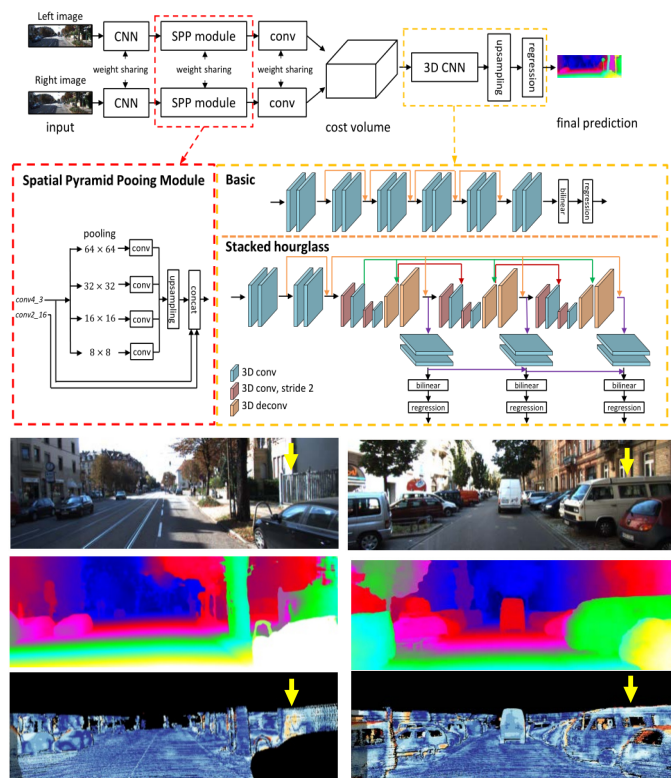


图 4 PSMNet 网络结构及网络预测的深度图和误差图

论文: <https://arxiv.org/abs/1803.08669>

代码: <https://github.com/JiaRenChang/PSMNet>

5、Learning Depth with Convolutional Spatial Propagation Network

工作: 该论文对 PSMNet 网络的空间金字塔池化和沙漏 3D 卷积进行改进, 提出了一种简单高效的线性传播模型卷积空间传播网络(CSPN), 传播采用一种循环卷积操作的方式来执行, 通过深度卷积神经网络学习相

邻像素之间的亲和力。如图 5 所示, 论文基于空间传播网络(SPN)通过四个方向的空间传播收集周围点的相似信息, 设计了 CSPN 和 3D CSPN 网络, 用卷积核的方式直接得到 4 个方向上的信息, 另外通过循环迭代, 得到更大相邻半径的信息。论文在 CSPN 网络基础上设计了卷积空间金字塔融合(CSPF), CSPF 模块和 CSPN 整体网络结构如图 6 所示。

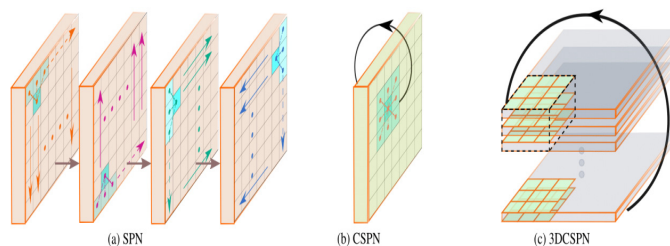


图 5 卷积空间传播网络

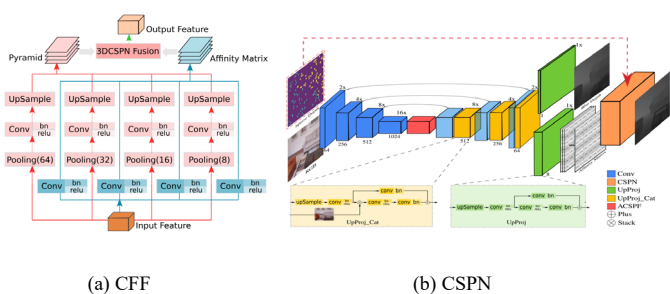
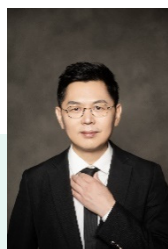


图 6 CSPF 模块和 CSPN 网络结构

论文: <https://arxiv.org/abs/1810.02695>

代码: <https://github.com/XinJCheng/CSPN>

责任编辑 李策



贾同

东北大学智能感知与机器人研究所所长, 教授, 博士生导师, 研究方向为三维视觉感知、图像识别等。
个人主页: <http://faculty.neu.edu.cn/ise/jiatong/>

人脸检测数据集

大连理工大学 付陈平 樊鑫

人脸检测是一个复杂且具有挑战性的模式检测问题，其主要目的在于定位出图像中的人脸位置，在身份认证与安全防护、图像检索、媒体与娱乐等应用领域十分重要。简单来说，人脸检测技术的发展主要经历四个阶段，第一个阶段是初期阶段，该阶段的特点是检测背景简单，人脸由人脸器官的局部特征进行描述；第二个阶段是人机交互阶段，该阶段将统计检测理论融入到人脸描述过程中，但依然需要人工先验知识的指导；第三个阶段是机器自动识别阶段，由于计算机硬件的发展，该阶段人脸特征的提取和利用得到进一步提升和革新；第四个阶段是深度学习阶段，该阶段借助深度学习检测技术的发展人脸检测效果取得了大幅度提升，展现出巨大的商业前景，代表性商用方法有 Facebook 的 DeepFace、Google 的 FaceNet、百度的 EasyDL 等。

人脸图像数据集是计算机视觉领域中研究历史最久，也是应用最广泛的图像数据集，其采集基本来源于静态的单张照片或者动态的视频帧。从人脸检测、人脸姿态与 3D、人脸活体与伪造到人脸风格化等不同人脸任务，人脸图像数据集得到了全方位、多方向、多类型发展，本文将立足于人脸检测任务，详细介绍人脸检测数据集，包括：UFDD、MAFA、MALF、以及 WIDER Face 数据集。

1、UFDD 数据集

介绍：UFDD(Unconstrained Face Detection Dataset)数据集发布于 2018 年，由日本神奈川县富士通实验室有限公司联合约翰·霍普金斯大学和罗格斯大学共同创建。数据集收集自各大浏览器网站，包括谷歌、雅虎、

百度、微软等网站，通过“雨+人脸”、“雪+人脸”、“雪+人群”、“黑夜+人群”等关键字进行搜索整理。通过该种搜索方式获得共计 6425 张图像，并标记 10897 张人脸。该数据集是一个场景随机、背景复杂的人脸检测数据集，包括雨天(Rain)、雪天(Snow)、雾天(Haze)、模糊(Blur)、光照(Illumination)、晶体障碍(Lens Impediments)和干扰物(Distractors)等 7 个场景。其中雨天图像 628 张、雪天图像 680 张、雾天 442 张、模糊图像 517 张、光照图像 612 张、晶体障碍图像 612 张、干扰物图像 3450 张。UFDD 是第一个明确捕捉不同天气条件下的变化和其他退化的数据集，如运动模糊(Motion Blur)和焦点模糊(Focus Blur)等。图 1 举例了 UFDD 数据集中的图像，从左到右依次显示了雨、雪、雾、模糊、光照、晶体障碍、干扰物场景中的人脸图像。



图 1 UFDD 数据集示例

数据集地址：<http://www.ufdd.info/>

2、MAFA 数据集

介绍: MAFA(Masked Faces)数据集发布于 2017 年,由中国科学院信息工程研究所、北京航空航天大学和中国科学院大学联合创建。这是一个遮挡人脸检测数据集,共计包含 30811 张图像,并标记 35806 张遮挡的人脸。该数据基础图像来源于各大浏览器网站(谷歌、微软等),通过“脸”、“遮挡”、“覆盖”、“面具”等关键词进行搜索,此后,筛选出尺寸、遮挡等均符合条件的图片 30811 张,接下来,人工对这些图像进行进一步的遮挡修改。数据集组建人员为每张脸定义了 6 个需要手动标注的属性,包括脸的位置、眼睛的位置、面罩的位置、脸的朝向、遮挡等级和遮挡类型。其中脸的位置用方框标记出来,眼睛的中心点位置由人工标记出来,面罩的位置需要标记出遮挡物在脸上的具体位置,脸的朝向包括 5 个(左向、右向、前向、左前向、右前向),遮挡等级按照遮挡区域的数量分为 3 级(低度遮挡、中度遮挡、重度遮挡),遮挡类型分为 4 种(人造的纯色遮挡物、人造的复杂纹理遮挡物、手/头发等身体造成的自遮挡以及复杂类型)。图 2 详细说明了 6 个需要手动标记属性的位置,图 3 按照 5 个遮挡等级、5 个人脸朝向展示了数据集中的一些图像。

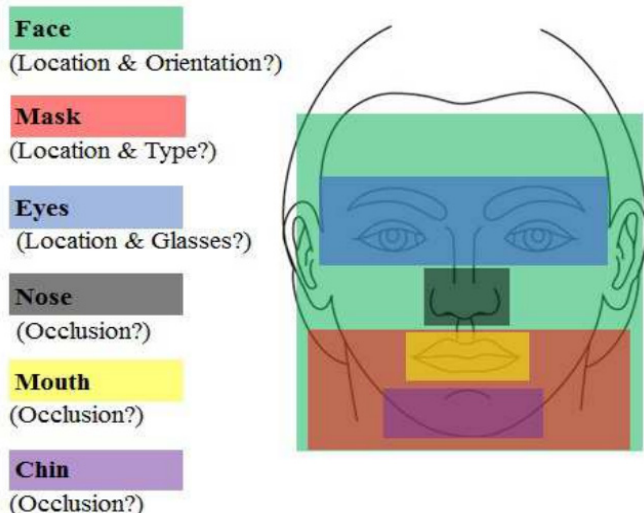


图 2 手动标记属性



图 3 MAFA 数据集示例

数据集地址: <http://www.escience.cn/people/geshiming/mafa.html>

3、MALF 数据集

介绍: MALF(Multi-Attribute Labelled Faces)数据集发布于 2015 年,由中国科学院自动化研究所创建。这是一个更加细粒度地评估野外环境中人脸检测模型的数据库,共计包含 5250 张图像、并标注 11931 张人脸。数据集中图像主要来源于互联网,在保证样貌多样性的前提下,由人工手动挑选符合数据集要求的图像。所有采集的图像均为 RGB 和 JPEG 格式,平均图像大小是 573×638,每张图片平均包含 2.27 张脸,其中 46.97%的图片包含一张脸,43.41%的图片包含 2~4 张脸,8.30%的图片包含 5~9 张脸,1.31%的图片包含超过 10 张脸。在人脸尺度上,所有人脸的平均尺寸为 83×83,中值人脸尺寸为 64×64 大,数据集组建人员按照 60×60 和 90×90 的尺寸将人脸划分为小、中、大 3 个尺寸,同时将尺寸小于 20×20 的人脸忽略掉。与此同时,该数据集还提供了性别,是否带眼镜、是否遮挡、是否是夸张的表情等辅助信息。图 4 展示了该数据集中的一些图像。



图 4 MAF 数据集示例

数据集地址: <http://www.cbsr.ia.ac.cn/faceevaluation/>

4、Wider-Face 数据集

介绍: Wider-Face 数据集发布于 2015 年, 发布机构为香港中文大学。该数据集是一个规模很大的数据集, 在很长一段时间里, 大型互联网公司和科研机构都在 Wider-Face 上做人脸检测算法竞赛。该数据集中图像来源于互联网, 由人工筛选掉没有人脸的图像, 共计 32203 张图片, 标记 393703 张人脸。图像种的人脸具有多种尺寸、姿势、遮挡、表情、妆容和光照。此外, Wider-Face 数据集中图像分辨率较高, 所有图像的宽都缩放到 1024 像素, 最小标注的人脸大小为 10×10 ,



图 5 Wider-Face 数据集示例

数据集地址: <http://mmlab.ie.cuhk.edu.hk/projects/WIDERFace/index.html>

责任编辑 沈沛意



付陈平

博士研究生, 大连理工大学国际信息与软件学院, 研究方向为计算机视觉。



樊鑫

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

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

好文推荐

西安电子科技大学团队“基于结构和特征的图自适应池化算法”最新成果发表在 WWW 2020。

论文: Liang Zhang, Xudong Wang, Hongsheng Li, Peiyi Shen, Guangming Zhu. Structure-Feature based Graph-Adaptive Pooling, WWW, April. 2020.

图池化(Graph Pooling)是对图数据进行下采样, 去除冗余节点, 保留重要节点, 来达到减少计算量, 增加计算结果准确性的目的。被广泛应用于药物分子识别、社群分类、编队类型分类等领域。传统的图池化方法为了构建池化拓扑, 仅从单个角度评估节点的重要性, 过于简单和客观。其次, 在池化过程中, 未选择节点的特征信息直接丢失, 不可避免地导致大量图特征信息丢失。该论文提出融合节点选择机制与节点特征融合方法, 并应用于药物和蛋白质分子分类问题。该类网络框架被称为 GSAPool, 网络具体构建过程如下。首先, 图中包含大量的节点和边, 这些节点和边表示丰富的结构和特征信息, 需要通过图卷积网络和多层感知器分别计算节点的结构信息得分和特征信息得分, 以此判断候选的被保留节点。然后, 由于部分携带有效特征信息的节点可能也被删除了, 所以在删除节点之前, 对要保留的节点以及其一阶邻域的节点做一次特征信息聚合, 这样可以使被保留的节点携带更多的信息, 减少特征信息损失的程度。最后, 将池化后的图输出, 参与下一层运算。

GSAPool 的基础结构图如图 1 所示, 具体的求解步骤如下:

- 1、根据节点的信息得分计算被保留节点。
 - 1) 使用图卷积网络计算节点的结构信息得分;
 - 2) 是用多层感知器计算节点的特征信息得分。
- 2、对于被保留节点, 使用图注意力网络进行特征信息聚合, 这样在丢弃节点之前使用未选择节点的信息对其特征进行聚合。可以更有效地利用节点的特征信息, 使最终的图嵌入向量更具代表性。

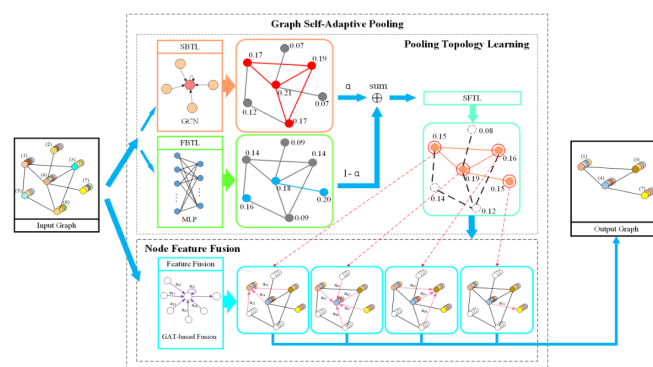


图 1 自适应图池化算法结构展示

图 1 中的 SBTL 为根据结构信息学习拓扑模块, FBTL 为根据特征信息学习拓扑模块, SFTL 为二者加权求和后的模块, Node Feature Fusion 为特征融合模块。实验结果表明, GSAPool 在药物和蛋白质分子分类任务中有效提高了分类精度。

表 1 GSAPool 与目前主流图池化算法在图分类任务上的对比结果

Methods	DD	NCI-1	NCI-109	Mutagenicity	Average
gPool[9]	75.01 ± 0.86	67.02 ± 2.25	66.12 ± 1.60	67.44 ± 2.78	68.90
SAGPool[20]	76.45 ± 0.97	67.45 ± 1.11	67.86 ± 1.41	76.89 ± 1.12	72.16
SET2SET[25]	74.50	71.50	68.60	76.40	72.75
DiffPool[31]	78.00	76.00	74.10	80.60	77.18
EigenPooling[22]	78.60	77.00	74.90	79.50	77.50
GSAPoolGCNConv&MLP	78.27 ± 0.96	72.84 ± 1.15	77.07 ± 1.89	79.90 ± 0.80	77.02
GSAPoolGCNConv+Fusion _{GAT}	77.31 ± 2.00	69.14 ± 1.45	76.31 ± 1.22	78.28 ± 1.35	75.26
GSAPoolGCNConv&MLP+Fusion _{GAT}	80.84 ± 1.17	75.20 ± 1.55	77.43 ± 1.35	81.99 ± 1.20	78.87

责任编辑 贾同, 樊鑫

好文推荐

南京理工大学和大连理工大学团队的“基于物理模型约束的生成对抗模型的图像恢复”最新成果发表在 IEEE TPAMI-2020。

论文: Jinshan Pan; Jiangxin Dong; Yang Liu; Jiawei Zhang; Jimmy Ren; Jinhui Tang; Yu Wing Tai; Ming-Hsuan Yang. Physics-Based Generative Adversarial Models for Image Restoration and Beyond, IEEE TPAMI, January. 2020.

图像恢复和底层视觉问题(例如图像去模糊、去雾和去雨)是指从给定的输入中估计出清晰的图像。然而, 这些问题都是不适定的, 基于启发式图像先验的方法很难获得问题的最优结果。因此, 本文提出了一个端到端的、基于物理模型约束的生成对抗模型 (Physics Constrained GAN)。其中, 物理模型用于指导模型估计结果的生成, 以保证估计结果与观测图像的一致性。实验结果表明, 所提方法可有效应用于图像恢复和底层视觉问题。

如图 1 所示, Physics Constrained GAN 主要包括 4 个模块, 分别是: 2 个判别模块 \mathcal{D}_g 、 \mathcal{D}_h , 1 个生成器 \mathcal{G} 和一个图像重建模型。其中, \mathcal{G} 用于生成估计图像,

\mathcal{D}_g 用于判别 \mathcal{G} 输出的图像分布是否与真实图像分布接近, \mathcal{D}_h 用于判别重建图像是否与观测图像相一致, 图像重建模块用于重建 \mathcal{G} 的输出图像。以图像去模糊问题为例, 假设 $\{x_i\}_{i=1}^N$ 和 $\{y_i\}_{i=1}^N$ 分别表示为清晰图像和对应的模糊图像, 方法流程可描述如下:

1、 \mathcal{G} 学习清晰图像 x_i 和模糊图像 y_i 间的映射关系, 并生成输入图像 y_i 对应的估计图像 $\mathcal{G}(y_i)$;

2、图像重建模块按照所提物理模型将估计图像 $\mathcal{G}(y_i)$ 进行重建, 以获得 $\mathcal{G}(y_i)$ 对应的再生成图像 \tilde{y}_i ,

$$\tilde{y}_i = \mathcal{K} \otimes \mathcal{G}(y_i), \quad (1)$$

式(1)中, \mathcal{K} 代表模糊核, \otimes 代表卷积操作;

3、 \mathcal{D}_h 判断 \tilde{y}_i 的模糊程度是否与 y_i 的模糊程度相一致, 同时, \mathcal{D}_g 判断 $\mathcal{G}(y_i)$ 的清晰度是否达到 x_i 的清晰程度。

不同视觉问题对应的物理模型的具体形式有所不同, 本文分别具象了图像去模糊、去雾和超分辨率等问题的物理模型。随后在图像去模糊、去雾、超分辨率和去噪等任务上进行了大量实验。实验结果表明, 所提方法的表现均优于其他方法, 表明 Physics Constrained GAN 的有效性。

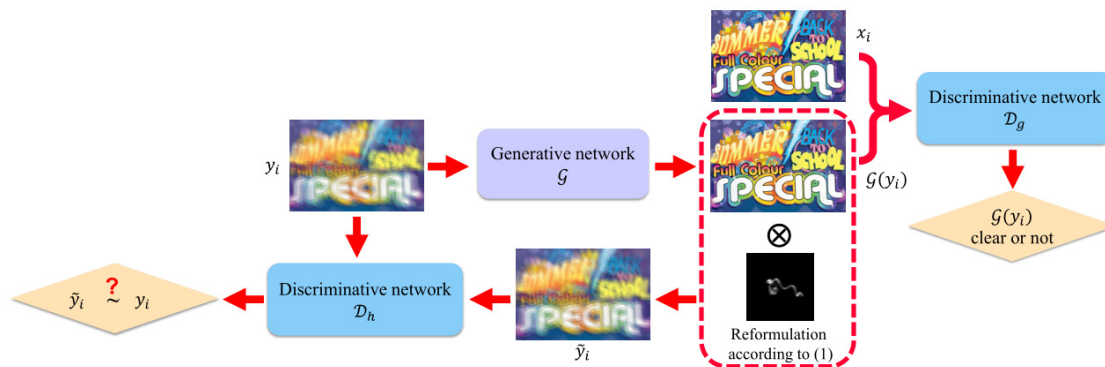


图 1 DECAMEL 算法流程图

责任编辑 贾同, 樊鑫

好文推荐

北京大学和耶鲁大学团队合作的“基于上下文信息的单幅图像联合雨检测和去除”最新成果发表在 IEEE TPAMI-2020。

论文: Wenhan Yang; Robby T. Tan; Jiashi Feng; Zongming Guo; Shuicheng Yan; Jiaying Liu. Joint Rain Detection and Removal from a Single-Image with Contextualized Deep Networks, IEEE TPAMI, vol. 42, no. 6, pp. 1377-1393, June. 2020.

降雨会显著降低图像可见度, 会损害许多计算机视觉系统。雨水带来的可见度下降主要有两种类型: 远处的雨线积累, 在视觉上表现出类似于雾或雾的大气遮蔽效果。雨水的堆积使光线散射并进入视线, 严重降低了可见度。近处的雨线产生镜面高光和场景遮挡, 雨线的形状、大小和方向变化多样, 特别是在大雨时, 会导致可见度得严重下降。当前大部分算法会平滑没有雨区域的纹理细节, 且当前常用的去雨模型不能很好覆盖真实雨水图像中的一些重要因素(例如水汽、雨线形状、方向), 此外, 现有方法往往没有考虑空间上下文信息。针对这些问题, 本文设计了一个降雨模型和一个深度网络架构。其中, 降雨模型包含一个显示雨线区域的二值雨水映射

图, 其包含重叠雨线的各种形状、方向和大小, 以及用于模拟大雨场景的降雨积累。在此基础上, 构建了一个多任务深度学习网络, 联合学习二值雨水映射图、雨线层和无雨场景。同时, 本文引入可利用区域上下文信息的深度扩展网络, 用于生成不受雨线变化干扰的特征。采用循环逐步去除雨线策略, 用以处理各种形状和方向重叠的雨线。此外, 利用二值映射图提供的约束作为额外信息来训练此网络。

所提方法的流程图如图 1 所示, 具体的流程描述如下: 所提除雨方法的架构, 包括周期性联合雨检测和除雨模块, 以及放置在两个递归之间去除积累雨线并保留细节的模块。每个递归是一个多任务网络联合执行雨检测和去除的过程(在蓝色破折号框)。此外, 一个利用上下文信息的深度扩展网络(在灰色区域)从输入的雨图 O_t 中提取降雨特征 F_t , 然后依次预测 R_t 、 S_t 和 B_t 进行联合雨线检测、估计和去除。在相邻两次递归中, 利用保留细节模块去除积累的雨线以提高图像可见度

本文方法模型在 Rain12、Rain20L 和 Rain100H 这 3 种基准数据集上进行了大量实验。实验结果表明, 本文方法的表现均优于其他主流的去雨方法模型, 显示了所提方法的有效性。

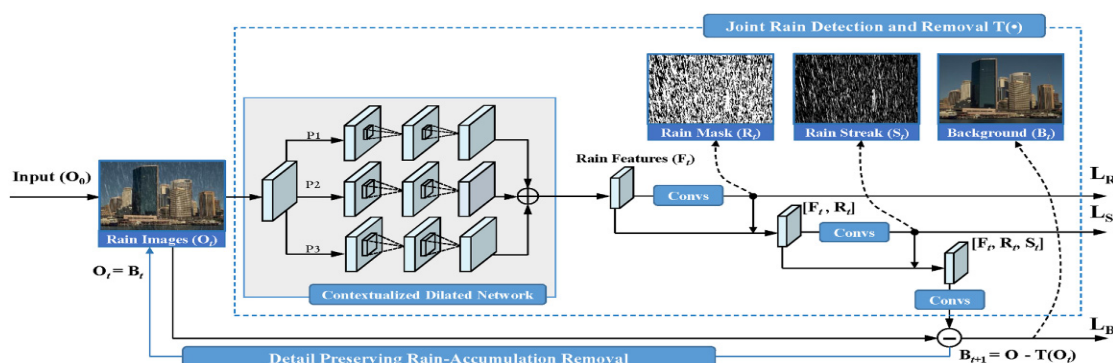


图 1 所提方法流程图

责任编辑 樊鑫, 贾同

美国加州大学洛杉矶分校谢建文博士团队

美国加州大学洛杉矶分校谢建文博士的科研一直聚焦于人工智能、计算机视觉和机器学习的基础性问题，其中包括视觉统计建模与计算 (statistical modeling and computing)、生成式学习 (generative learning)、表征学习^[1] (representation learning) 以及无监督学习 (unsupervised learning) 等。谢博士与其科研合作者近年来在基于能量的深度视觉生成模型、协作式生成学习、可解释性神经细胞表达模型方面取得了重要进展。下面简单聚焦谢博士目前三个重要的科研方向。

研究方向 1: 基于能量的视觉生成模型

广义模式理论 (pattern theory)^[2]，最初由统计学家 Ulf Grenander 于 60 年代提出，是一套针对自然界各种不同模式进行统一建模的理论与计算框架，并由数学家 David Mumford^[3] 倡导推广至计算机视觉与模式识别领域。在广义模式理论的框架下，所有视觉模式都由统计生成模型 (statistical generative models) 来表示，其数学形式为视觉信号 (例如图片、视频等) 的概率密度分布，这样的概率模型能够通过马尔科夫链蒙特卡罗采样 (Markov chain Monte Carlo/MCMC)^[4] 的方式来可视化模型所学到的模式样例。

谢建文博士在美国加州大学洛杉矶分校攻读博士学位期间，与其导师朱松纯教授、吴英年教授一直致力于寻找一套统一的统计理论与计算框架，用于表示计算机

视觉先驱 David Marrs 在其名为《Vision》^[5] 一书中所倡议的早期视觉 (early vision) 概念。他们遵循“广义模式理论”的框架，于 2016 年提出了一类全新的视觉生成模型，称之为“Deep FRAME”模型^[6,7]，该模型是 1996 年 FRAME (Filters, Random Fields and Maximum Entropy)^[8] 模型往深度学习方向的拓展，首次把深度学习的神经网络结构^[9]与统计物理学的吉布斯分布 (Gibbs distribution) 或马尔科夫随机场模型 (MRF)^[10,11] 进行结合。

如图 1 所示，Deep FRAME 模型在视觉信号空间直接定义了一个显式的概率密度函数，其数学形式是一个基于能量的生成模型 (Energy-based model)，能量函数由一个现代的神经网络进行参数化，并通过从观测到的训练数据中 (譬如，一万张人脸图片或单张纹理图片) 学习该密度分布函数。更准确地说，它通过学习一个由神经网络参数化的能量函数把较低的能量 (即较高的概率) 赋值给观测到的样本，而把较高的能量 (即较低的概率) 赋值给未观测到的其他样本。模型的训练使用统计学中最严谨的最大似然估计 (maximum likelihood)，学习算法遵循“合成分析法 (analysis by synthesis)”，即不停交替以下两个步骤：步骤一、使用蒙特卡罗采样，例如朗之万动力学 (Langevin dynamics)，从模型中采样生成新的样本；步骤二、通过对比生成样本和训练样本的统计量来更新模型的参数 (见图 1)。两个步骤的关键在于计算神经网络的梯度：步骤一是对数据变量求梯度，用于数据生成；步骤二是对参数变量求梯度，用于

模型更新，两者均可以通过反向传播（Back propagation）轻易获得。图 2 展示了 Deep FRAME 模型通过蒙特卡罗采样所生成的图片例子。

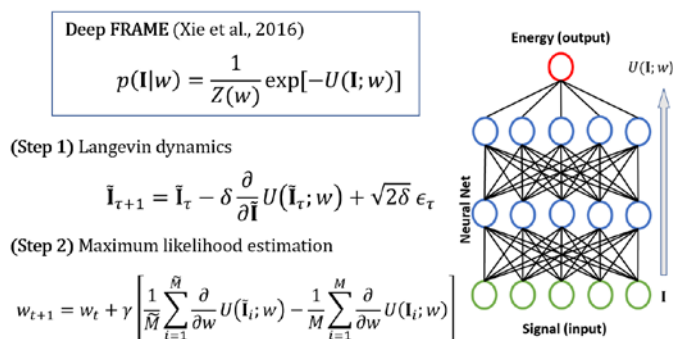


图 1 Deep FRAME 模型

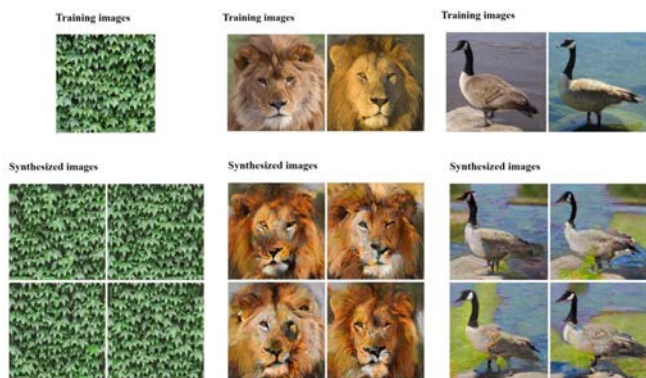


图 2 纹理与物体图片的生成

与目前流行的深度生成模型，譬如对抗生成模型 [12] (Generative Adversarial Net 或 GAN) 和变分自编码器 (Variational Auto-Encoder 或 VAE) [13] 相比，Deep FRAME 模型并不需要引入额外的辅助性模型，这意味着其自底向上 (bottom-up) 的表达过程与自顶向下 (top-down) 的生成过程被统一在同一组模型参数里，即所谓“一个大脑”的原则。这大大减轻了神经网络结构设计的负担，另外，由于模型的训练使用统计上严谨的最大似然估计法，其学习过程较对抗式学习来说更为稳定，并不会出现“模式崩塌 (mode collapse)”等问题。

谢建文博士等人指出，Deep FRAME 模型的数学形式能直接从判别式神经网络 (ConvNet) 中推导得来 (见

图 3)，因此 Deep FRAME 模型亦被称作为“基于能量的生成式神经网络 (Energy-based Generative ConvNet)” [6]。这一发现直接导致可以通过训练 Deep FRAME 模型来学习分类器，或者通过训练神经网络分类器的方式来训练生成模型 Deep FRAME。

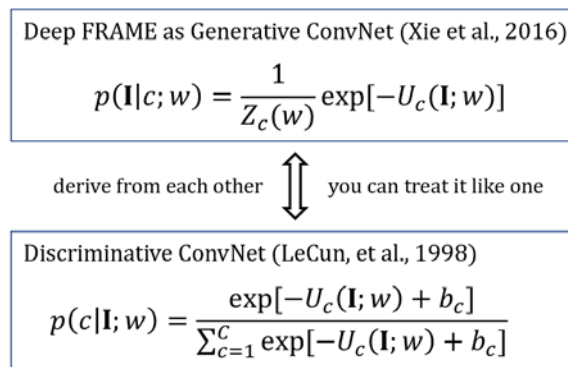


图 3 Deep FRAME 模型与判别式神经网络的相互转换

谢建文博士等人在论文[14, 15]中指出，基于蒙特卡罗的 Deep FRAME 模型参数的最大似然估计可理解为一个“自我对抗”的学习过程。他们先定义一个价值函数 (如图 4)，该函数是一个以生成图片集和模型参数为自变量的方程。学习算法在使用朗之万采样的阶段通过最小化能量函数来生成图片，因此采样的过程 (即步骤一) 是在最小化该价值函数；而算法在模型更新阶段 (即步骤二)，则通过更新参数来最大化价值函数。因此模型的学习过程是一个“极小化极大 (Mini-max) 算法”。

Gradient of Likelihood:

$$L'(w) = \frac{\partial}{\partial w} \left[\frac{1}{\tilde{M}} \sum_{i=1}^{\tilde{M}} U(\tilde{\mathbf{I}}_i; w) - \frac{1}{M} \sum_{i=1}^M U(\mathbf{I}_i; w) \right]$$

Value function: $V(\{\tilde{\mathbf{I}}_i\}, w)$

Mini-max game: $\min_{\{\tilde{\mathbf{I}}_i\}} \max_w V(\{\tilde{\mathbf{I}}_i\}, w)$

图 4 Deep FRAME 模型的自我对抗解释

另外，Deep RAME 模型的学习过程可以看作是一

个交替“寻找局部最小值区域 (mode seeking)”和“移动局部最小值区域 (mode shifting)”的过程。图 5 展示了该过程。其中横坐标代表图片空间，纵坐标代表能量值。红色曲线是真实分布函数的能量方程，观测到的数据 (红色叉) 便位于其最小值区域。蓝色曲线代表了模型的能量函数，蓝色圆圈则代表 MCMC 合成的样本。如图 5(1) 所示，采样的过程是郎之万动力学通过随机梯度下降在寻找能量函数最低点 (高概率区域)，并把合成的图片安置在该低能量区的周围 (高密度)。图 5(2) 展示了更新参数的过程实质是调整能量函数的形状，其包括把低能量区域 (即高密度区域) 从生成图片的位置往训练图片的位置移动，此过程一般伴随能量函数的低能量区会变得越来越陡峭。图 5(3) 是当能量函数发生改变之后，采样过程会在新的能量函数上重新寻找低能量区。图 5(2)和(3)两个过程不断交替，直到模型与真实数据分布的低能量区域匹配上为止，如图 5(4)。

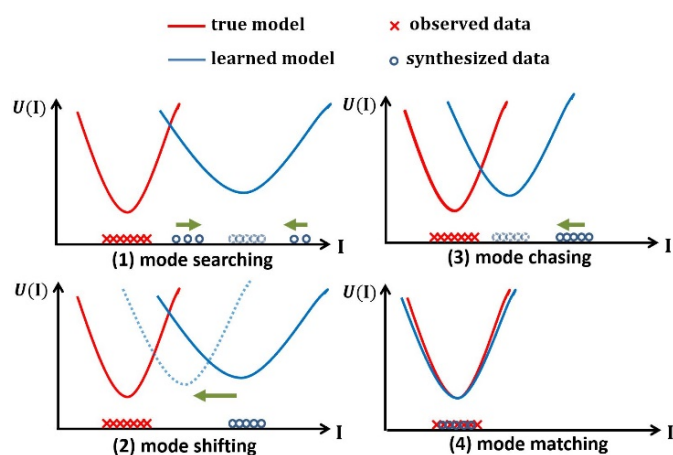


图 5 Deep RAME 模型的训练过程示意图

谢建文博士等人于 2017 年通过使用自底向上的时空卷积网络来参数化能量函数，从而把 Deep FRAME 模型推广至视频建模^[14, 15]。图 6 展示了使用时空 Deep FRAME 模型生成视频的两个例子。在每个例子当中，模型从一段长达 70 帧的动态纹理 (dynamic texture) 视频中进行学习，并使用蒙特卡罗采样的方法生成两段不一样的动态纹理。(注：图 6 的每段视频只随机展示了其中六帧图片。)

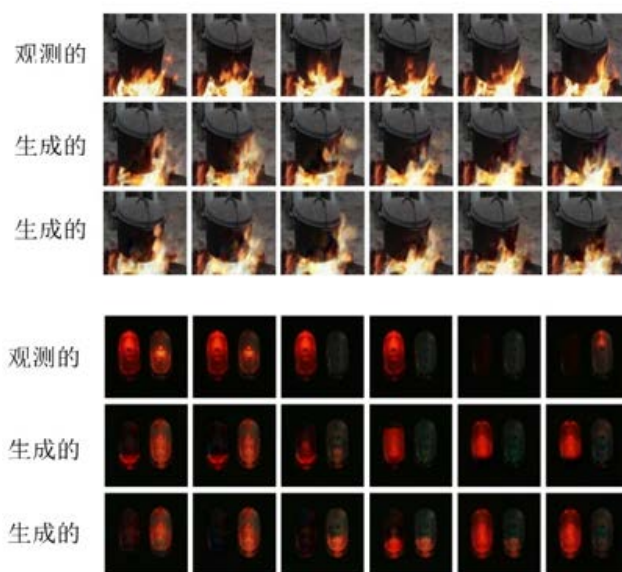


图 6 时空 Deep FRAME 模型的视频生成

时空 Deep FRAME 模型还具备从带有缺失或被遮挡像素、不完整的视频中进行学习的能力，这对于对抗生成模型 (GAN) 来说是不可能的。原学习算法只需要做一个小改动，即在每次迭代过程中添加一个郎之万蒙特卡罗采样，用以修复训练样本中的缺失部分，便能够实现从不完整的数据中进行学习。以数据恢复为目的的采样过程使用当前的模型参数，并基于训练样本中可见部分的像素来生成未知部分的像素，属于对一个条件概率分布进行采样。其具体操作为，在每次迭代中，郎之万采样以带有缺失数据的训练样本进行初始化，通过有限步郎之万更新来实现训练样本中被遮挡区域像素的修复。

被修复的训练样本将重新用于模型的学习，因为更新模型参数的梯度依赖于新生成的样本以及被恢复的训练样本；反过来，更新后的模型会在下次迭代中用于修复训练样本。因此，模型同时实现了学习、生成以及修复三个任务，三个任务彼此帮助、互相促进形成闭环。图 7 展示了三个例子。它们分别使用了不同类型的像素遮挡方法，分别有 50% 的椒盐噪声遮挡、方块状局部遮挡以及缺失帧。模型在三个例子中均能获得良好的数据恢复效果。

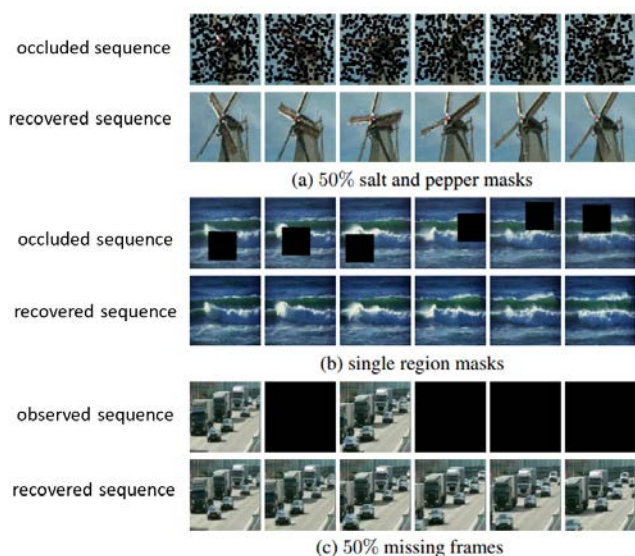


图 7 时空 Deep FRAME 模型的无监督视频恢复

有趣的是，上述的修复算法还可以用于移除视频中不需要的移动物体。假设给定一段视频，如果要移除当中的某个移动物体，可以先手动遮挡视频里每一帧出现该物体的区域，然后直接使用上述的修复算法从该被遮挡的视频进行学习，由于被遮挡的物体在当前的训练视频中自始至终都没有出现，因此算法并不能把它还原出来。结果，视频中被遮挡的区域会被背景内容所填充，从而实现了移除物体的效果。图 8 为使用了时空 Deep FRAME 模型进行视频背景填充的例子。

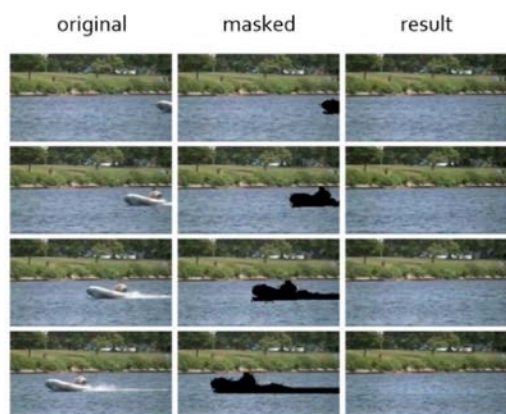


图 8 时空 Deep FRAME 模型的视频背景填充

谢建文博士等人于 2018 年进一步把 Deep FRAME 模型拓展至三维体素物体的建模^[16, 17]。三维 Deep FRAME 模型使用三维卷积神经网络作为能量函数，该

网络映射以三维体素作为输入，通过多层三维卷积操作最终输出数据的特征值作为能量。图 9 展示了使用三维 Deep FRAME 模型对三维物体进行建模的例子。图中每一行代表一个物体类别。左边三列随机展示了每类训练样本中的三个例子，而右边则为 Deep FRAME 模型通过蒙特卡罗采样生成的三维物体。

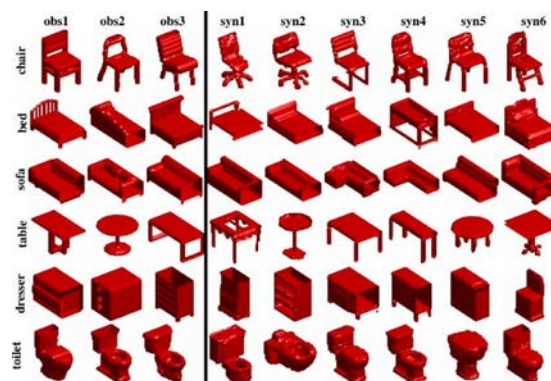


图 9 三维 Deep FRAME 模型的体素生成

自 2019 年起，DeepFRAME 模型已经开始引起计算机视觉和机器学习社区的注意，科研人员对生成模型的研究，逐渐从“对抗式框架”向谢博士等人于 2016 年提出的“基于能量的 DeepFRAME 框架”进行迁移。

研究方向 2：协作式生成学习

谢建文博士等人于 2018 年提出“协作式学习 (Cooperative Learning)”概念以及“生成式协作网络 (Generative Cooperative Networks 或 CoopNets)^[18, 19]”。它是一个通过让两类概率生成模型以师生互助的方式进行学习的生成模型。生成式协作网络由一个描述网络 (Descriptor Network) 和一个生成网络 (Generator Network) 组成 (见图 10)。描述网络是一个基于能量的模型 (energy-based model)，又被称之为马尔科夫随机场 (Markov random field) 或指数族分布 (exponential family model)，其起源于统计物理学，概率密度分布由一个能量函数所定义，该能量函数由一个从信号到能量的自底向上 (bottom-up) 的神经网络所参数化。生成网络是一个隐变量模型 (latent

variable model), 又被称之为有向图模型 (directed graphical model)。它假设数据 (如图片) 由一个服从已知先验分布的隐变量经过非线性变换得到, 该变换由一个自顶向下 (top-down) 的神经网络所参数化。

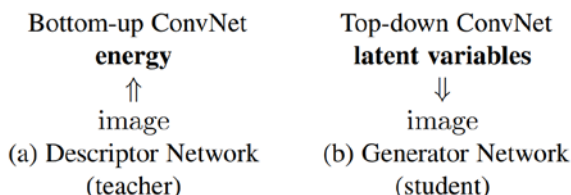


图 10 构成“生成式协作网络”的两大组成部分

由于两类模型的似然梯度均包含棘手的期望运算, 一般来说需要借助蒙特卡罗采样 MCMC 来近似, 然而 MCMC 非常耗时并且难以收敛, 因此使用最大似然估计的办法单独训练两类模型难度极大。生成式协作网络的提出使得两类模型的最大似然估计算法能够互相交织在一个“协作式学习”的框架当中, 并起到相互促进的作用。其基本思想是: 在训练基于能量的模型 (即描述网络) 的过程中, 通过同时学习一个隐变量模型 (即生成网络) 作为近似的采样器来快速初始化描述网络的蒙特卡罗采样过程。

如图 11 所示, 协作式学习算法的基本流程为: 在每一次迭代中, 算法先从先验概率分布 (譬如高斯分布) 随机生成一定数量的隐性变量, 并通过当前的生成网络以非线性变换的方式生成图片。接着, 描述网络的蒙特卡罗采样以生成网络提供的初始生成图片作为起始点, 走有限步 MCMC 以获得最终的生成图片。描述网络的学习通过比较训练样本和最终生成样本来更新模型, 而生成网络的学习则通过在隐变量和最终生成图片之间进行回归。换句话说, 生成网络从 MCMC 修改过的图片中进行学习, 由于该样本所对应的隐性变量在每次迭代中是已知的, 所以算法并不需要计算后验概率来推断隐性变量, 从而使得生成网络的学习变成更为简单的有监督学习。图 12 为使用协作式学习算法生成不同类型的视觉样式的例子, 其中包括纹理、人脸图片、视频以

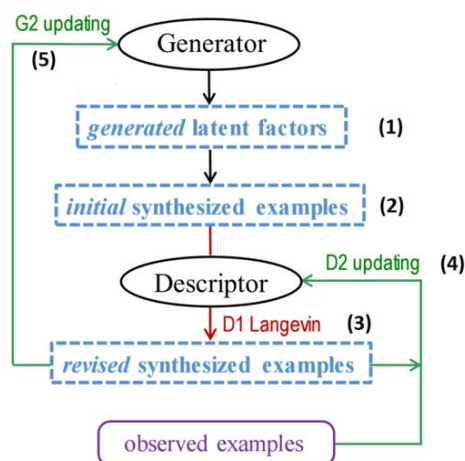


图 11 协作式学习算法的描述

及三维体素物体等。

不妨把上述的算法作一个类比, 在协作式学习的框架中, 描述网络和生成网络的关系如同师生关系一般, 描述网络扮演着导师的角色, 而生成网络则充当学生的身份。学生快速起草论文的初稿, 导师接着修改学生的稿件, 然后投稿接受评审。学生从老师的修改中学习, 而导师则通过外部审稿意见进行学习。在这个任务中, 导师指导了学生, 但学生完成了大部分的写作工作。

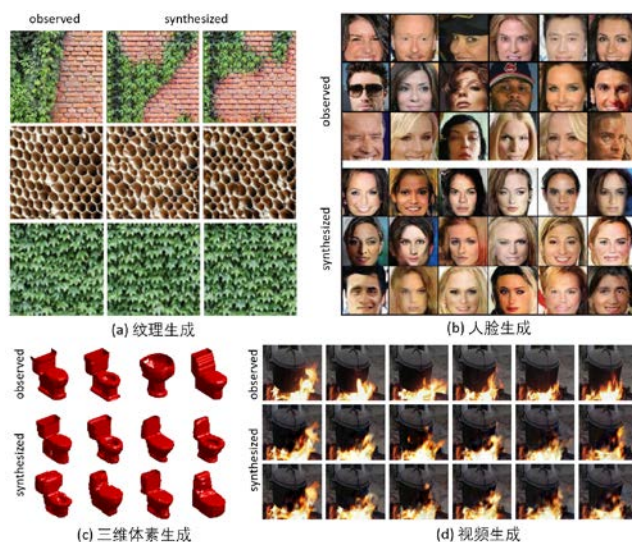


图 12 “生成式协作网络”的视觉数据建模

首先, 描述网络的训练依然遵循最大似然下的“成分分析法”模式, 即使用蒙特卡罗采样从模型中生成新

的样本，并通过比较生成样本和训练样本来更新模型，但其每次的郎之万采样过程将由生成网络的输出进行初始化，其原因是：生成网络能够使用直接映射的办法简单快速生成样本，好的初始化起点能够使得描述网络以更少的迭代次数生成符合模型分布的样本。其次，生成网络把描述网络的生成样本当作训练数据，学习蒙特卡罗如何修改初始生成样本。换句话说，基于能量的描述网络把学到的知识蒸馏到了生成网络中，该过程被称之为“马尔科夫链蒙特卡罗教学” (MCMC teaching) [18, 19]。

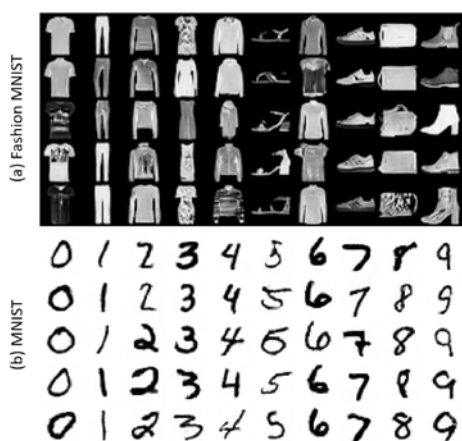


图 13 通过“生成式协作网络”进行基于类别的图片生成

上述的模型为非条件概率模型，谢博士等人进一步把协作式学习的框架推广至有监督的条件学习 (supervised conditional learning) 场景 [20]。假设给定样本格式为“(输入, 输出)”的训练数据，其中，“输入”和“输出”一般属于不同的数据形态并且维度极高。例如，输入是人脸素描图，而输出是彩色照片。条件学习的目的是要建立起“输入”和“输出”之间的映射关系，往往这种关系是“一对多”的，即给定一个输入，其输出是一个条件概率分布。谢博士与合作者通过把“非条件的协作式学习框架”里的描述网络和生成网络修改成条件概率形式 (conditional model)，并使用协作式学习的方法来联合训练两个网络，从而学习 $P(\text{输出}|\text{输入})$ 的条件概率分布。模型在基于类别条件的图片生成 (如图 13)，图片翻译 (image translation)

与图片填充 (见图 14) 等视觉任务中均取得良好效果。

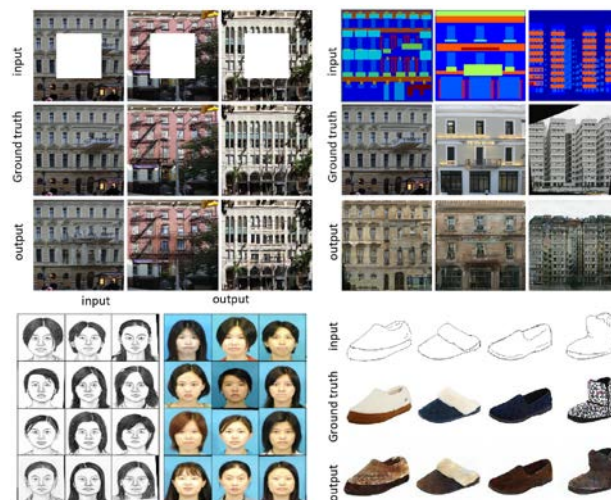


图 14 通过“生成式协作网络”进行图片到图片的翻译

不妨把上述问题中的 (输入, 输出) 称之为 (问题, 答案)，那么上述的条件学习问题可以看成是逆向强化学习 (inverse reinforcement learning) 中一个关于“学习如果解决问题”的问题 (learn to solve a problem)。其中，条件描述网络作为一个基于能量的条件模型，其能量函数定义了解决该问题的目标函数 (objective function) 或价值函数 (value function)，即给定某个问题，什么样的答案才是可取的。该答案一般需要通过迭代的优化算法或采样算法才能获得。换句话说，条件描述网络定义了一个形式为迭代算法的求解器 (solver)，因此它是一个“深思熟虑” (slow thinking) 的过程。相反，该框架的另一个组成部分，即条件生成网络则定义了一个直接从问题到答案的映射，它属于一个“快速思考” (fast thinking) 的过程。虽然快速思考模型求解效率较高，但其缺乏一个内在的价值函数去描述解决该问题的本质，因此泛化能力会较差。在谢博士等人所提出的框架中 (见图 15)，他们根据问题先利用“快速思考”的生成器产生一个初步的答案，然后再利用“深思熟虑”的求解器改善该初始的答案，改善过程是通过优化其习得的价值函数来实现的。由于初始答案为优化过程提供了良好的起始点，因此优化的步数将会

大大减少。谢博士等人已经把“快速思考-深思熟虑 (fast thinking and slow thinking)”这一框架成功应用在逆向最优控制领域 (inverse optimal control) [21]。

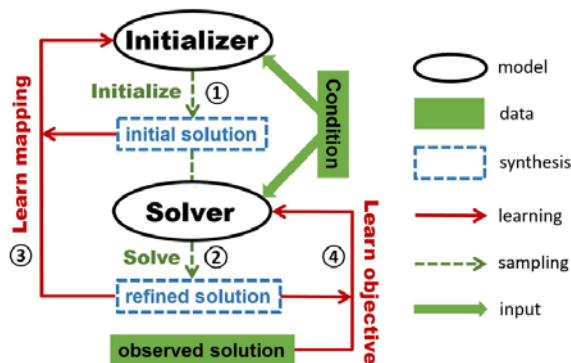


图 15 “快速思考-深思熟虑” 框架

“生成式协作网络”还能应用在无监督条件学习 (unsupervised conditional learning) 场景。假设给定两个来自不同域的数据集，譬如两个图片集：一个是真实照片集 A，而另一个是艺术家梵高的油画集 B。这两个图片集样本之间的对应关系是不知道的，因此直接学习单边条件概率分布 $P(A|B)$ 或 $P(B|A)$ 来实现跨域图片翻译 (cross-domain image translation) 是困难的。谢博士等人通过联合训练两个“生成式协作网络”来分别表示“从域 A 到域 B”和“从域 B 到域 A”的条件概率，并且同时强制两者互为对方的推断模型。该方法已成功应用在图片风格迁移 (图 16) 和动作迁移 (图 17) 等视觉应用场景中 [22]。

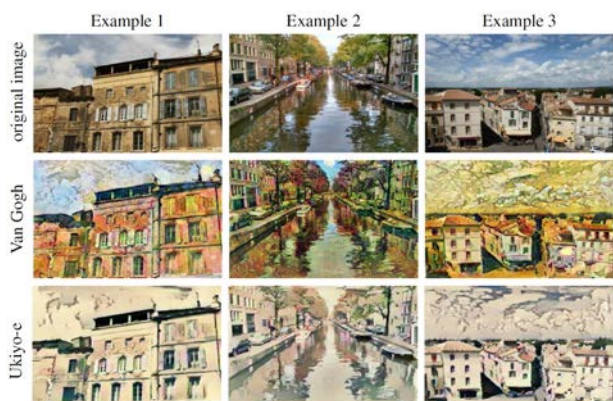


图 16 基于“生成式协作网络”的风格迁移

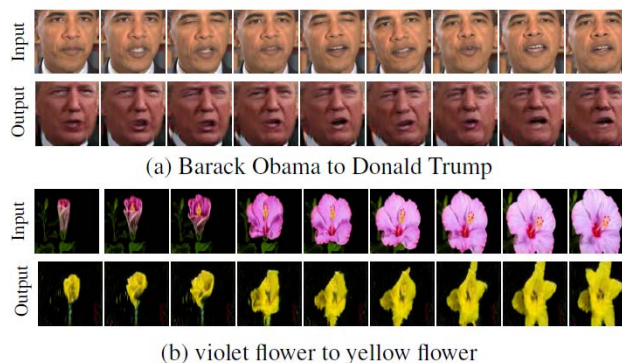


图 17 基于“生成式协作网络”的动作迁移

研究方向 3：计算神经科学

物理学通过寻找能揭示物理现象成因和机理的数学定律来理解物理世界，同样，计算神经科学试图发现能解释人类大脑中不同神经系统机制的数学定律来理解大脑。因此，寻找神经细胞的可解释数学模型不单单是计算神经科学的基础性研究，也是揭开大脑之谜、实现强人工智能的必经之路。

众所周知，存在于大脑内嗅皮质 (entorhinal cortex) 中用于指导生物体导航和识别自身位置的网格细胞 (grid cell) 以及存在于大脑视觉皮层 (visual cortex) 中用于视觉感知的 V1 细胞 (V1 cell) 是两类非常重要的神经细胞。它们的发现，分别于 2014 年和 1981 年被授予诺贝尔生理学或医学奖，可见其地位的重要性。然而，人们对于这些细胞的理论理解依然非常有限。

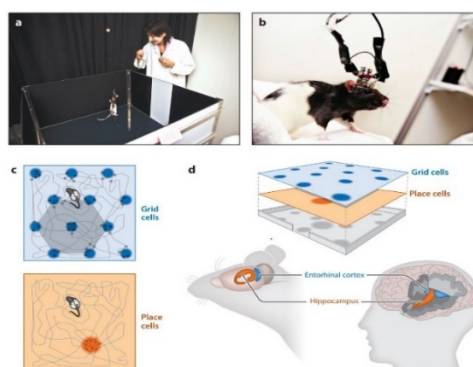


图 18 网格细胞与位置细胞

图 18 展示了老鼠在空间移动时,位置细胞与网格细胞的活跃反应。当老鼠在空间中自由走动,抵达某一特定位置时,位置细胞的神经信号开始活跃起来。同时,当老鼠到达某些位置时,则网格细胞开始活跃,每一个网格细胞对多个空间位置产生反应,这些位置点之间的排列构成规则的六边形,网格细胞也由此得名。多个网格细胞协同工作形成坐标系统,让老鼠具备识别位置的导航能力。图 19 则展示了大脑的视觉区域。初级视觉皮层 V1 是大脑表达视网膜成像信号的第一步。V1 区的神经细胞能够对图像中不同位置、方向以及尺寸的边缘作出响应。

为了解开上述这些神经细胞的谜团,谢建文博士和加州大学洛杉矶分校实验室的科研合作者提出了一种理论上非常新颖的数学模型,即“向量-矩阵模型”^[23],来分别表达和理解网格细胞与 V1 视觉细胞。这类数学模型最大的特点是,采用了基于向量和矩阵的表示。虽然向量表示的概念在深度学习的框架中已经十分常见,然而,作用在向量上的矩阵表示则非常罕见,这将是该模型中最新颖、也是最关键的元素之一。

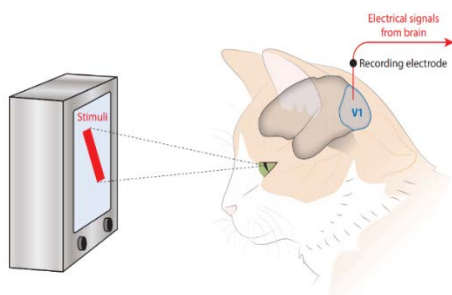


图 19 大脑的 V1 视觉区域

在使用“向量-矩阵模型”对网格细胞进行建模的工作中^[23, 24],他们大胆假设了动物自身的空间位置在大脑中是使用高维向量进行编码,而自身的空间位移则通过一个用于旋转位置向量的旋转矩阵来表示。向量的每一个维度对应一个网格细胞,即所有的网格细胞共同参与空间位置的编码,把空间坐标映射成向量。模型的核心思想是:动物从当前的位置移动到下一个位置,在大脑中是通过“向量-矩阵乘法”来实现,并且大脑中向量

旋转的角度大小和外部物理空间的位移量是等距的(isometric)。换句话说,下一个位置的向量表示,是通过位移矩阵与当前位置的向量相乘所获得的(如图 20)所示。这样简单的数学模型具备很好的可解释性,因为它具有非常明显的代数与几何。该模型能够从仿真的动物运动轨迹中学习到具有清晰六边形图案的网格细胞(见图 21),这和生物学中所观察到的网格细胞特征非常吻合。模型还能够实现生物导航中的两大基本任务:路径积分(path integral)和路径计划(path planning)。该模型是当前最简单、最具可解释性并且获得的网格样式最清晰的网格细胞数学模型。在这项工作中,谢建文博士与合作者试图用他们的模型来解答为何生物大脑中用于导航的网格细胞会具有如此奇妙的规范六边形网格样式,他们给出的答案是:大脑的网格细胞极有可能采用了“向量-矩阵乘法”这样的数学机制在运作。

$$\begin{array}{ccc}
 x_t & \xrightarrow{+\Delta x} & x_{t+1} \\
 \downarrow & & \downarrow \\
 v(x_t) & \xrightarrow{M(\Delta x) \times} & v(x_{t+1})
 \end{array}$$

图 20 向量-矩阵表达模型

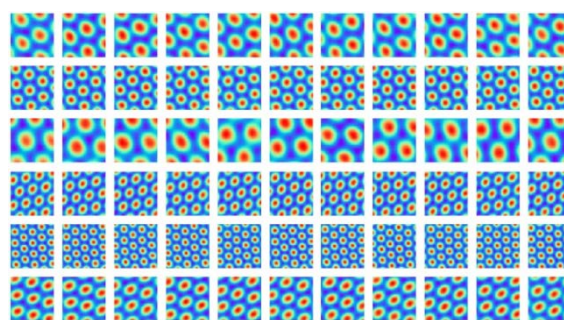


图 21 模型学习到的网格细胞呈现规则六边形样式

V1 细胞在视觉感知中扮演非常重要的角色。目前,V1 细胞的主流数学模型是稀疏编码模型(sparse coding)^[25],该模型对于 V1 细胞的理解是:V1 细胞对图像进行稀疏表达是导致 V1 细胞具有 Gabor 滤波器样式的根源。在 2019 年的一项研究工作中,谢建文博士等人提出颠覆性的 V1 数学模型^[26],他们大胆假

设 V1 细胞的直接目的是为了感知视觉的局部运动，即图像中由物体与周围环境的相对运动而造成的像素移动。他们使用了“向量-矩阵模型”的框架来重新诠释 V1 细胞，其核心思想是：每一个 V1 细胞是一个线性滤波器，图像中每个像素的局部内容被表示成高维向量，向量中的每一维对应着一个 V1 细胞的响应，而像素的移动则被表示为一个旋转矩阵。当图像内容由于像素移动而发生改变时（譬如图片中一个物体在移动），则意味着图像每个像素的向量被表示局部变化的矩阵进行旋转。见图 22 左示意图。实验表明，在没有引入稀疏这一特性下，模型依然能够学到 Gabor 滤波器的样式（如图 22 右所示，这是 V1 细胞的独特特征），从而验证了他们对 V1 视觉细胞的机理的假设，即 V1 细胞的出现是为了感知视觉的局部运动。新的模型颠覆了 V1 细胞的存在是因为大脑为了进行稀疏编码这一传统认知。

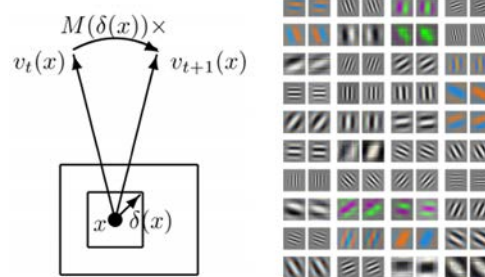


图 22 向量-矩阵模型的视觉 V1 细胞建模

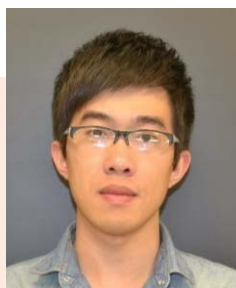
“向量-矩阵模型”和物理学中的量子理论以及数学中的李群代数 (Lie Algebra) 有着非常紧密的联系^[24]。谢建文博士和他的科研合作者相信，这样的数学表达模型可能不仅仅存在于物理世界，还同样存在于人类大脑的神经系统当中，这也许是未来解开大脑之谜的关键。

责任编辑 张汗灵

参考文献

- [1] Jianwen Xie, Ruiqi Gao, Erik Nijkamp, Song-Chun Zhu, Ying Nian Wu. Representation Learning: A Statistical Perspective. Annual Review of Statistics and Its Application (ARSIA), 2020.
- [2] Ulf Grenander, Michael I Miller. Pattern Theory: From Representation to Inference. Oxford University Press, 2007.
- [3] David Mumford, Agnes Desolneux. Pattern Theory: the Stochastic Analysis of Real-World Signals. CRC Press, 2010.
- [4] Adrian Barbu, Song-Chun Zhu. Monte Carlo Methods, Springer, 2020.
- [5] Marr, David. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. MIT press, 2010.
- [6] Jianwen Xie *, Yang Lu *, Song-Chun Zhu, Ying Nian Wu. A Theory of Generative ConvNet. Proc. of the International Conf. on Machine Learning (ICML), 2016. (*equal contribution)
- [7] Ying Nian Wu, Jianwen Xie, Yang Lu, Song-Chun Zhu. Sparse and Deep Generalizations of the FRAME Model. Annals of Mathematical Sciences and Applications (AMSA), 2018.
- [8] Song-Chun Zhu, Ying Nian Wu, David Mumford. Minimax Entropy Principle and its Application to Texture Modeling. Neural Computation, 9(8):1627–1660, 1997.
- [9] Yann LeCun, Leon Bottou, Yoshua Bengio, Patrick Haffner. Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE, 86(11):2278–2324, 1998.
- [10] Julian Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems. Journal of the Royal Statistical Society. Series B (Methodological), pages 192–236, 1974.
- [11] Stuart Geman, Christine Graffigne. Markov Random Field Image Models and Their Applications to Computer Vision. In Proceedings of the International Congress of Mathematicians, volume 1, page 2, 1986.
- [12] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua

- Bengio. Generative Adversarial Nets. In Advances in Neural Information Processing Systems, pp. 2672-2680, 2014.
- [13] Diederik P. Kingma, Max Welling. Auto-Encoding Variational Bayes. arXiv preprint arXiv:1312.6114, 2013.
- [14] Jianwen Xie, Song-Chun Zhu, Ying Nian Wu. Synthesizing Dynamic Pattern by Spatial-Temporal Generative ConvNet. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [15] Jianwen Xie, Song-Chun Zhu, Ying Nian Wu. Learning Energy-based Spatial-Temporal Generative ConvNets for Dynamic Patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2019. (*equal contribution)
- [16] Jianwen Xie*, Zilong Zheng*, Ruiqi Gao, Wenguan Wang, Song-Chun Zhu, Ying Nian Wu. Learning Descriptor Networks for 3D Shape Synthesis and Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8629-8638. (CVPR) 2018. (*equal contribution)
- [17] Jianwen Xie*, Zilong Zheng*, Ruiqi Gao, Wenguan Wang, Song-Chun Zhu, Ying Nian Wu. Generative VoxelNet: Learning Energy-Based Models for 3D Shape Synthesis and Analysis. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2020. (*equal contribution)
- [18] Jianwen Xie, Yang Lu, Ruiqi Gao, Ying Nian Wu. Cooperative Learning of Energy-Based Model and Latent Variable Model via MCMC Teaching. In AAAI, vol. 1, no. 6, p. 7. 2018.
- [19] Jianwen Xie, Yang Lu, Ruiqi Gao, Song-Chun Zhu, Ying Nian Wu. Cooperative Training of Descriptor and Generator Networks. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI), 2018.
- [20] Jianwen Xie*, Zilong Zheng*, Xiaolin Fang, Song-Chun Zhu, Ying Nian Wu. Cooperative Training of Fast Thinking Initializer and Slow Thinking Solver for Conditional Learning. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI). 2021. (*equal contribution)
- [21] Yifei Xu, Jianwen Xie, Tianyang Zhao, Chris Baker, Yibiao Zhao, Ying Nian Wu. Energy-Based Continuous Inverse Optimal Control. arXiv preprint arXiv:1904.05453.2020.
- [22] Jianwen Xie*, Zilong Zheng*, Xiaolin Fang, Song-Chun Zhu, Ying Nian Wu. Unsupervised Cross-Domain Translation via Alternating MCMC Teaching. arXiv preprint 2020. (*equal contribution)
- [23] Ruiqi Gao*, Jianwen Xie*, Song-Chun Zhu, Ying Nian Wu. Learning Grid Cells as Vector Representation of Self-Position Coupled with Matrix Representation of Self-Motion. In International Conference on Learning Representations (ICLR). 2018. (*equal contribution)
- [24] Ruiqi Gao, Jianwen Xie, Xue-Xin Wei, Song-Chun Zhu, Ying Nian Wu. A Representational Model of Grid Cells' Path Integration Based on Matrix Lie Algebras. arXiv preprint arXiv:2006.10259. 2020.
- [25] Bruno A. Olshausen, David J. Field. Sparse Coding with an Overcomplete Basis Set: A Strategy Employed by V1?. Vision Research. 1997.
- [26] Ruiqi Gao, Jianwen Xie, Siyuan Huang, Yufan Ren, Song-Chun Zhu, Ying Nian Wu. Learning Vector Representation of Local Content and Matrix Representation of Local Motion, with Implications for V1. arXiv preprint arXiv:1902.03871. 2019.



谢建文 (Jianwen Xie)

2016 年获加州大学洛杉矶分校 (UCLA) 统计学博士学位, 2016 至 2017 年在加州大学洛杉矶分校计算机视觉、认知、学习与自主机器人实验室从事博士后研究。2017 至 2020 年加入海康威视美国研究院成为创始成员之一, 并担任资深研究科学家。2020 年加入百度美国研究院认知计算实验室并担任资深研究科学家至今。主要研究领域为计算机视觉、统计建模与计算、机器学习与计算神经科学。谢建文博士的 Deep FRAME、Generative ConvNet、CoopNets 模型系列工作被视为基于能量的深度生成模型领域的开创性工作。

征文通知

1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 Image and Vision Computing, Pattern Recognition, Pattern Recognition Letters 和 International Journal of Approximate Reasoning。

3 会议简介

国际多媒体博览会 (IEEE International Conference on Multimedia and Expo) 是 IEEE 一年一度的学术性会议，会议的主要内容是多媒体信号处理技术。ICME 是多媒体研究领域旗舰类国际学术会议之一。

2021 年 ICME 在中国深圳举行，本届会议将汇聚全世界从事多媒体理论与应用研究的广大科研工作者及工业界同仁，共同分享多媒体研究领域的最新理论和技术成果，为大家提供精彩的学术盛宴。

责任编辑：刘帅奇

表 1 计算机视觉领域相关国内外会议

会议名称	会议时间	会议地点	截稿日期	会议网站
ICIP 2021	2021.09.19-22	Anchorage, USA	2021.01.14	https://2021.ieeeicip.org/
IJCAI 2021	2021.08.21-26	Montreal, Canada	2021.01.19	https://ijcai-21.org/
CHIL 2021	2021.04.08-10	Online	2021.01.12	https://www.chilconference.org/
ICML 2021	2021.07.18-24	Vienna, Austria	2021.02.05	https://icml.cc/Conferences/2021
ICCV 2021	2021.10.11-17	Montreal, Canada	2021.03.18	http://iccv2021.thecvf.com/

表 2 计算机视觉领域相关国内外期刊专刊

期刊名称	专刊题目	投稿网址	截稿日期
IVC	Learning with Manifolds in Computer Vision	https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/special-issue-on-learning-with-manifolds-in-computer-vision	2021.1.31
IJAR	Time Series Clustering and Classification	https://www.journals.elsevier.com/international-journal-of-approximate-reasoning/call-for-papers/special-issue-on-time-series-clustering-and-classification	2021.03.31
PR	Machine Learning for Combinatorial Optimization and Its Applications (ML4CO)	https://www.journals.elsevier.com/pattern-recognition/call-for-papers/machine-learning-for-combinatorial-optimization	2021.04.01
PRL	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	2021.07.20

COMPUTER VISION NEWSLETTER

04 2020
总第 26 期



计算机视觉专委会简报



CCF 计算机视觉
专委会