顶会观察

CVPR 2021

北京邮电大学 邓伟洪

国际计算机视觉与模式识别会议（IEEE Conference on Computer Vision and Pattern Recognition，CVPR）是计算机视觉及模式识别的顶级学术会议，与 ICCV、ECCV 并称为计算机视觉领域三大顶会。CVPR 有着非常高的学术影响力，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 发布的学术指标中，其 H 指数雄踞人工智能领域的榜首。引人注目的是，今年大会组委会有不少华人面孔：中国科学院院士谭铁牛担任 General Chair，上海科技大学信息科学与技术学院教授虞晶怡、肯塔基大学计算机系终身教授杨睿担任 Program Chair，中山大学智能工程学院副教授梁小丹担任 Tutorials Chair 等。

一、 国际计算机视觉与模式识别会议的亮点

CVPR 2021 于美东时间 2021 年 6 月 19 日至 25 日在线上举行。由于疫情原因，会议主办方建立了虚拟会议的网站，以供参会人员进行展示及技术交流。作者需要为每篇论文准备一个五分钟的预先录制的视频和海报的 PDF 文件来演示工作，参会者可以按需查看演示文稿和视频。同时，会议为每篇论文安排了指定的线上交流时间，允许作者与感兴趣的参会者通过文本聊天或线上会议的方式进行交流。CVPR 研讨会以及教程将通过直播视频进行，主持人和参与者之间进行现场问答。会议还包括具有视频和文本聊天元素的多个在线网络活动，为参会者提供了自由、便利的会议环境，使每位参会获得了良好的参会体验及交流经历。

除了后面将介绍的最佳（学生）论文奖外，大会还颁发了两项传统大奖。Longuet-Higgins Prize 以认知科学家 H. Christopher Longuet-Higgins 的名字命名，表彰十年前对计算机视觉研究产生重大影响的 CVPR 论文。今年的获奖论文是来自微软的 Real-time human pose recognition in parts from single depth image（从单一深度图像中实时识别人体姿势）和来自石溪大学的 Baby talk: Understanding and generating simple image descriptions（婴儿谈话：理解和生成简单的图像描述）。Young Researcher Awards 旨在表彰对计算机视觉做出杰出研究贡献的年轻研究人员，今年获奖者是 FAIR 的 Georgia Gkioxari 和 MIT 的 Phillip Isola。

今年大会的亮点是首次颁发的 Thomas S. Huang 纪念奖。该奖项为了缅怀一代 CV 宗师、华人计算机视觉泰斗 Thomas S. Huang（黄煦涛），由 PAMITC 奖励委员会选出。今年首届获奖者是 MIT 电子电气工程与计算机科学教授 Antonio Torralba。Torralba 的研究领域包括场景理解和上下文驱动的目标识别、多感官知觉整合、数据集构建以及神经网络表征的可视化和解释。

二、 论文录用情况

CVPR 2021 总共收到了 7,039 篇有效投稿，其中 1,661 篇论文被接收，接收率约为 23.5%，相比 CVPR 2020 论文接受率略有回升。其中，有 295 篇论文入选 oral presentation，oral 率约为 4.1%，低于去年的 5.7%。同时，从大会公布的数据来看，今年大会接收到的注册及有效投稿数量都有显著的提高。CVPR 2021 会议涵盖的方向包括目标检测、行为识别、对抗攻击与防

御、生物特征、计算摄影、图像和视频检索、图像和视频合成、图像分类、姿态估计、无监督学习、视频理解、多模态等方向。在 CVPR 2021 接收的论文中，3D 视觉、计算摄影学、视频图像合成三个方向的论文数量最多，无监督、半监督、自监督三个关键词的出现次数相较 CVPR 2020 上升了 50%。在 CVPR 2021 最佳论文奖的 32 篇候选论文中，有华人参与的论文高达 18 篇，华人为一作的论文共有 16 篇，且其中 6 篇的一作为国内机构学者。本次 CVPR 2021 收录论文中，来自中国工业界的各大互联网企业获得了不俗的成绩。根据公开数据，商汤及联合实验室共 66 篇论文入选，腾讯 AI 实验室与优图团队共有 33 篇论文入选，华为诺亚方舟研究团队有 30 篇论文入选，旷视有 22 篇论文入选。这些被录用的论文在很多重要工业应用领域上取得了重大突破，包括模型压缩、网络架构搜索、语义理解、底层视觉、光流估计、无监督学习、人体姿态估计、目标检测等。此外，谷歌在本次会议表现依旧亮眼，共有 70 余篇论文入选， 其中华人为第一作者的论文共有 34 篇。

三、 主题演讲

为了方便深入交流，大会将受邀演讲者的研究领域大致分为三组：AI 伦理、计算机视觉中的机器学习、人类和机器人感知。

安第斯大学的 Pablo Arbelaez 博士带来了“人工智能促进全球健康”的演讲。 Google Ghana 的 John Quinn 博士的演讲从乌干达和加纳团队的工作角度，讨论了计算机视觉如何为健康、气候和粮食安全这些联合国关注的与持续发展相关的领域的进步做出贡献，以及所面临的困难和风险。麻省理工学院的 Catherine D’ Ignazio 教授认为伴随海量数据而来的不平等的生产条件、不对称的应用方法以及它们对个人和群体的不平等影响越来越难以被数据科学家和其他在工作中依赖数据的人忽视，并讨论了如何实现合乎道德和公平的数据使用。

麻省理工学院的 Constantinos Daskalakis 博士从优化、复杂性理论和拓扑方法等方面阐述了对均衡计算与多智能体学习的见解。纽约大学的 Meredith Whittaker 教授探讨了美国军方与计算纠缠的历史，并且其认为冷战思维的延续产生了“人工智能军备竞赛”，使美国认为必须赢得这场竞赛才能维持军事和经济霸权。南京大学的周志华教授简要介绍了学习理论研究的悠久历史和关于 Boosting 的争论，并揭示了在学习过程中最大化边际均值时最小化边际方差的重要性，以及其为强大学习算法的设计所带来的灵感。

加州理工学院 Katie Bouman 讲述了如何利用计算成像管道代替传统光学成像，以获取传统光学成像无法获取到的图像。加州大学圣克鲁兹分校的 Su-hua Wang 博士分享了人类婴儿如何在动态事件中学到物体的表示以及不同的模式，图宾根大学 Matthias Bethge 展示了为了让机器像人类一样看待事物所进行的工作，百度的 Liang Huang 介绍了机器在同声翻译领域近期的突破与进展。此外，一个名为“计算机视觉遇见安全”的附加小组会议探讨了计算机视觉安全基础对技术、市场和研究的相关影响。

四、会议获奖和热点论文、教程与竞赛

最佳论文奖评审委员会由 CVPR 领域的 9 名国际权威学者组成，包括来自中科院计算所视觉信息处理与学习组的陈熙霖教授。今年大会共评选出了 1 篇最佳论文，2 篇最佳学生论文，2 篇最佳论文提名，3 篇最佳学生论文提名。

最佳论文：GIRAFFE：Representing Scenes as Compositional Generative Neural Feature Fields 的主要贡献是将组合式三维场景表示纳入生成模型，使得图像合成更加可控。该模型将图像场景表示为生成神经特征场的组合，可以在无需任何额外的监督的情况下，从非结构化和非特定视角的图像集合中学习到如何将每个对象的形状和外观从背景中分离出来。通过将场景表示与神经渲染管道（Neural Rendering Pipeline）相结合获得一个快速而逼真的图像合成模型，能够分解单个物体，并允许其在场景中平移和旋转，还可以改变摄像机的姿势。

最佳学生论文：Task Programming: Learning Data Efficient Behavior Representation 提出一种用

于减少自动行为分析任务中数据集标注工作量的方法。该论文基于多任务自监督学习，提出了一种用于行为分析的有效轨迹嵌入方法—TREBA。利用该方法专家们可以通过“任务编程”过程来有效地设计任务，即使用程序编码将领域专家的知识结构化。通过交换数据注释时间来构造少量编程任务，可以有效减少领域专家的工作量。在行为神经科学领域的数据集上，小鼠和果蝇两个领域内三个数据集的测试评估表明，TREBA 使注释负担减少到原来的十分之一。

最佳论文提名：Exploring Simple Siamese Representation Learning 发现了简单的孪生网络可以学习有意义的表示，实验验证了：即使不使用 1）负样本对，2）大 batch，3）momentum 编码器中的任何一项，也能获得很好的自监督学习效果。Learning High Fideity Depths of Dressed Humans by Wathing Socail Media Dance Videos 研究了如何高效利用社交媒体中大量的舞蹈视频进行自监督学习：将预测的局部几何体从一幅图像在不同的时刻扭曲到另一幅图像，使得自监督学习对预测实现时间的一致性。

最佳学生论文提名：Less Is More: ClipBERT for Video-and-Language Learning via Sparse Sampling 表明用少量稀疏采样片段的端到端学习会比使用从全长视频中密集提取的离线特征更加准确，验证了视觉模型训练样本中的“少即是多”原则。Binary TTC: A Temporal Geofence for Autonomous Navigation 通过一系列简单的二元分类来估计场景的相对深度 (Time-to-Contact)，对于视觉导航有重要意义。Real-Time High-Resolution Background Matting 提出了一种实时、高分辨率的背景更换技术，该技术可以在 GPU 上以 30fps 速度运行 4K 分辨率和以 60fps 的速度运行高清分辨率。

另外，来自国内的多篇论文也引起了广泛的讨论。北大的论文 Generalizing to the Open World: Deep Visual Odometry with Online Adaptation 提出了一种结合了深度学习和几何计算优点的在线自适应框架，使得深度视觉里程计网络能够以自监督的方式快速适应新的场景，在多个数据集上实现了更好的泛化性能与深度估计性能。中科院自动化所的论文 Information Bottleneck Disentanglement for Identity Swapping 提出了一种基于信息瓶颈解耦的高身份辨识度换脸方法，通过约束互信息、学习身份信息的最小充分统计量，将相互耦合的身份信息（identity）和感知信息（perception）显式地分流，生成了高身份可辨识度的图像，并根据对比学习的思想提出了一项度量身份可辨识度的统计评价指标，有力地推动了换脸技术的发展。中科院计算所的论文 FAIEr: Fidelity and Adequacy Ensured Image Caption Evaluation 针对图像描述生成任务，借鉴人类语言翻译中的“信-达-雅”分级评价思想，提出了以图像描述忠实性和充足性为核心的层次化评价指标，并通过构建视觉与语言跨模态场景图以对齐多粒度图文信息，将图像描述生成的评价问题形式化为多实例多模态场景图匹配问题，获得了与人类评价结果高度一致的图像描述评价系统。微软亚洲研究院的工作 DEKR：Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering 提出了一种基于密集关键点坐标回归的多人姿态检测模型，以解决在拥挤的人群的场景下，由于人群过于密集，重合程度太高，导致每个人的位置难以用人体检测框表示的问题，达到了目前自底向上姿态检测的最好结果。腾讯优图与南京理工大学大学合作论文 Consistent Instance False Positive Improves Fairness in Face Recognition 提出了一种基于误报率惩罚的损失函数，在无需人口统计学标注的情况下，通过增加实例误报率（FPR）的一致性来减轻人脸识别模型在基于不同属性划分的人脸组上的性能偏差，缓解了人脸识别中的不公平问题。华南理工的论文 Implicit Feature Alignment: Learn to Convert Text Recognizer to Text Spotter 提出了一种被称为隐式特征对齐的方法，该方法可以被集成到普通的文本识别器中，使其能够处理多行文本，在多个文档识别任务上获得了最先进的性能。

此外，CVPR 2021 包含精心组织举办的 83 个研讨会（Workshops）与 30 个讲习班（Tutorials），涵盖了可解释机器学习、细粒度视觉、视觉数据压缩、对抗鲁棒性、自监督学习、自动驾驶、视听场景理解、医学计算机视觉等多个领域。这些研讨会与讲习班为参会者提

供了优质的交流平台与学习资源，同时帮助参会者拓宽了学术视野。在研讨会上举办的各项挑战赛中，国内学术界和工业界都取得了突出的成绩。由欧洲科学院外籍院士焦李成领衔的西电人工智能团队在洪水中高分辨率 UAS 图像的分类与语义分割、洪水中高分辨率 UAS 图像的视觉问答、无监督二类地物分类变化检测等赛事中，获得 4 项冠军的优异成绩。百度在此次参与的 7 项挑战赛中共获得 10 项冠军，涉及自动驾驶、人体解析、智慧城市、物体检测、图像修复增强、视频目标分割、视频理解等多个方向。

五、总结与展望

回顾近几年的 CVPR 获奖和热点论文，我们可以发现，CVPR 越来越青睐致力于解决真实场景下存在的视觉问题的方法或工作，包括对真实场景的建模、实现各类视觉子任务（分类、检测、语义分割）的统一融合，从而模拟人类视觉系统对真实物理世界的认知。随着相关算法与硬件计算能力的不断升级，3D 视觉算法效果得到大幅提升，三维几何重建更加精细，表面纹理重建更加清晰，正为我们带来更加逼真的视觉观感和数据生成效果。无监督学习和弱监督学习通过不使用标签或减少对标签数量、质量的要求来迅速降低深度模型对于数据的标注需求，大幅提高了数据的利用效率，正在由量变引发质变。随着视觉认知能力的提升，多模态融合也从感知视觉内容，逐渐拓展到学习物理关系，逻辑推断，因果分析等知识，正在从感知智能迈向认知智能。

责任编委 王金甲

邓伟洪

北京邮电大学“鸿雁人才”教授，教育部青年长江学者。研究方向为生物特征识别、可信人工智能、情感计算、多模态学习。曾入选北京市优秀博士学位论文、教育部新世纪优秀人才、北京市科技新星、Elsevier 中国高被引学者等。

Email: whdeng@bupt.edu.cn