

顶会观察

ECCV 2022

东南大学 祁磊

欧洲计算机视觉国际会议 (European Conference on Computer Vision, ECCV) 是计算机视觉顶级会议之一, 与 CVPR 和 ICCV 并称为计算机视觉领域三大顶会。ECCV 会议两年召开一次, 与 ICCV 会议正好错开。今年大会的主席成员包括: 来自摩德纳大学的 Rita Cucchiara、来自捷克理工大学的 Jiri Matas、来自 Mobileye 的 Amnon Shashua 和来自以色列理工学院的 Lihi Zelnik-Manor。值得关注的是, 本届大会获得最佳论文奖的论文为 On the Versatile Uses of Partial Distance Correlation in Deep Learning。该论文的作者来自威斯康星大学麦迪逊分校和 Butlr, 其中第一、二作者皆为华人, 他们是 Xingjian Zhen 和 Zihang Meng, 本科分别毕业于清华大学和浙江大学。本届 ECCV 大会于 2022 年 10 月 23 日至 10 月 27 日在以色列特拉维夫举办, 包括 3 天的正会和 2 天的 Workshops & Tutorials。

一、会议概况

自新冠疫情流行以来, ECCV 2022 首次线下举办, 无法线下参会人员仍可选择线上参会。据主办方统计, 截至大会开幕, 有来自 76 个国家约 5000 人注册参会, 其中约 3200 人现场参会, 约 1800 人以线上方式参会。本届会议的所有与会者, 无论是线上还是线下, 都可以在会议平台上观看每篇论文时长为五分钟的演示视频。此外, 亲临现场的作者们可以现场做 presentation, 线上参会者只能观看, 不做在线工作汇报。

大会的主席团成员 (General Chairs) 介绍了会议的具体安排: 60 场 Workshops, 13 场 Tutorials、2 场主题报告、2 场现场演示、2 场 Mentoring Events 和

1 场 Industry Track。大会的程序主席们 (Program Chairs) 则对 ECCV 2022 论文的审核情况作了详细介绍: 在审稿过程中, 有 846 篇投稿因各种原因在第一阶段被拒稿 (Desk Rejected), 其中很多是因为暴露了作者的身份, 违反了双盲政策。在排除不同阶段被撤回的论文后, 大会最后收稿数量为 5804 份。所有的这些投稿都至少收到了三篇评审意见 (只有 6 篇文章只收到 2 条), 评审意见总数超过 15000 条。论文作者有机会提交 rebuttal, 然后由领域主席 (Area Chairs) 和分配给每篇论文的审稿人进行讨论。为了确保尽可能公平公正, 每篇论文的最终录用决定是由领域主席与一位协作领域主席讨论后作出的。此外, 程序委员会主席监督整个审稿过程, 尤其是在领域主席的决定与所有审稿人的意见存在明显分歧时, 程序委员会主席会增加关注。

本届 ECCV 较为特别的是首次举办 Industry Track, 其旨在通过组织一系列的特邀报告和 panel, 来突出计算机视觉技术应用性质和所产生的广泛影响, 以丰富本次大会的会议内容。Industry Track 于 10 月 26 日下午进行, 包括一系列关于创业的受邀演讲和小组讨论, 尤其是探讨如何跨越计算机视觉研究与其现实世界应用之间的界限。讲者包括以科研起家的公司创始人和 CEO, 以及相关领域的投资人和其他专家。

二、录用情况

ECCV 2022 收到的有效投稿和录用数量相比上一届都有大幅提高。大会共收到来自 18310 位作者的 6773 篇投稿, 总共有 5804 份有效投稿, 由 276 位 Area Chairs (AC) 处理, 并征求了 4719 位审稿人的意见。整个过程由 Program Chairs (PC) 监督, 并得到

General Chairs (GC) 的持续支持。最终有 1645 篇 (28%) 论文被接收, 包括 157 篇 Orals (2.7%)。相较于上一届, 今年 ECCV 的投稿量提升 15.5% (779 篇), 同时录用数量也有所上升, 录用论文数量提升为 20.9% (284 篇)。ECCV 2022 会议涵盖的方向包括: 识别 (检测、分类与检索)、3D 视觉、图像与视频的生成、表征学习和深度学习、视频分析与理解、视觉与语言、迁移学习、计算摄影、姿态估计与跟踪、多任务学习、无监督学习、行为识别、自动驾驶等方向。值得关注的是, 国内科研机构和企业在本届 ECCV 2022 上斩获颇丰。商汤共 70 篇论文入选, 包含 6 篇 Oral, 主要方向为自动驾驶、计算摄影、视频理解与分析、迁移学习、多模态等前沿研究和应用。旷视共有 20 篇论文入选, 其中 3 篇 Oral, 内容涵盖目标检测、3D 重建、图像复原等多个研究方向。腾讯则有 29 篇论文入选, 内容涵盖人脸安全、图像分割、目标检测等研究方向。国外机构方面, 谷歌在本次会议中有很不错的表现, 共有 60 多篇论文入选。苹果和亚马逊则分别有 7 篇和 12 篇论文入选。

三、 热点论文

2022 年度 ECCV 最佳论文奖评审委员会由 7 名国际权威学者组成, 其中包括两名华人学者: 宾夕法尼亚州立大学 Yanxi Liu 教授和宾夕法尼亚大学 Jianbo Shi 教授。本年度大会共评选出了 1 篇最佳论文, 2 篇最佳论文提名, 论文具体介绍如下。

最佳论文: On the Versatile Uses of Partial Distance Correlation in Deep Learning^[1], 来自威斯康星大学麦迪逊分校和 Butlr。目前神经网络模型的功能行为比较虽然取得了一些进展, 但系统的功能比较, 特别是不同网络之间的功能比较, 仍然是很困难的, 且往往是逐层进行的。本文重新审视了一种并不为人熟知的统计学的方法, 称为距离关联方法, 旨在评估不同维度特征空间之间的相关性。文章还描述了该方法适用于大规模模型的部署必要步骤, 可以为一系列新应用的开发打开大门, 比如可以调节一个深度模型与另一个模型的关系、学习不相干的表征以及优化不同的模型等。实验表明, 一个多功能的正则器 (或约束) 具有许多优点, 可以规避在这类分析中出现的一些常见困难。值得一提

的是, 该文章第一作者和第二作者皆为华人, 第一作者 Xingjian Zhen 为威斯康星大学麦迪逊分校计算机科学博士生, 曾于 2017 年获得清华大学电子工程系的学士学位; 第二作者 Zihang Meng 现为 Meta AI (原 Facebook AI) 研究科学家, 曾于 2017 年获得浙江大学电子信息工程学士学位。

最佳论文提名 1: Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields^[2], 来自图宾根大学、马克斯·普朗克计算机科学研究所和 Meta 现实实验室。众所周知, 姿态或动作先验对于生成逼真的新姿态非常重要, 对从有噪声或局部观察数据重建精确的姿态也非常重要。本文提出了一种基于神经距离场 (NDFs) 的人体姿态连续模型: Pose-NDF。该模型学习一系列各种可能的姿态作为神经隐式函数的零水平集 (zero level set), 将 3D 建模隐式曲面的思想扩展到高维域 $SO(3)^K$ 。与之前基于 VAE 的人体姿态先验 (将姿态空间转换为高斯分布) 相比, 该研究对真实姿态流形进行建模, 以保留姿态之间的距离。该方法在各种下游任务中优于 SOTA 方法, 包括对真实世界的人体动作数据进行去噪、从遮挡数据中恢复姿态以及从图像中重建 3D 姿态。此外, 与基于 VAE 的方法相比, Pose-NDF 可生成更多样化的姿态。

最佳论文提名 2: A Level Set Theory for Neural Implicit Evolution under Explicit Flows^[3], 来自加利福尼亚大学圣迭戈分校。基于坐标的神经网络参数化隐式表面已经成为几何的有效表示, 它们高效充当了参数水平集, 其中零水平集定义了感兴趣的表面。本文提出了一个新框架, 允许将三角形网格定义的变换操作应用于这类表面。这其中的部分操作中可以被视为在显式表面引起瞬时流场的能量最小化问题。该方法通过扩展水平集的经典理论, 利用流场来实现参数化隐式表面变形。此外, 通过形式化与水平集理论的关联, 研究者分析认为现有的可微表面提取和渲染方法偏离了理论, 并利用本文方法对表面平滑、平均曲率流、逆渲染和用户定义的隐式几何编辑等应用加以改进。

此外, 本次大会共有 157 篇论文入选 oral 环节, 其中华人学者为第一作者的论文数量超过六成, 多篇文章

也引起广泛关注。华中科技大学、约翰霍普金斯大学、字节跳动和牛津大学的论文 In Defense of Online Models for Video Instance Segmentation^[4]提出了一个基于对比学习的视频实例分割 online 算法: IDOL, 可学习更具有区分度的 instance embedding, 并充分利用视频的历史信息来保证算法稳定性, 将 online 模型表现提高到与 offline 模型相当甚至更高的水平。华为诺亚实验室的 CLIFF: Carrying Location Information in Full Frames into Human Pose and Shape Estimation^[5]基于 HMR 网络结构提出新的动作捕捉算法, 在网络输入和监督信号中引入裁剪框的全局位置, 同时使用新的方法构造人体动作捕捉伪标注, 降低估计误差 40%以上。谷歌、康奈尔大学和加州大学伯克利分校的 InfiniteNature-Zero: Learning Perpetual View Generation of Natural Scenes from Single Images^[6]提出一种新型混合方法, 在迭代渲染、优化和重复框架中集成了几何和图像合成功能, 从而可采用单一图像, 生成由数百个具有逼真和多样化内容的新视图组成的长相机轨迹, 对永久视图生成 (prepetual view generation) 问题进行了进一步探索。

四、大会获奖

Koenderink Prize. 也称时间检验奖 (test of time), 该奖旨在表彰计算机视觉领域的基础性贡献研究, 获奖论文均为发表时间超过十年并经受住时间检验的研究。今年获奖论文共有两篇, 第一篇是来自纽约大学、伊利诺伊大学厄巴纳-香槟分校和微软剑桥研究院 2012 年发表的论文 Indoor Segmentation and Support Inference from RGBD Images^[7]。该文章提出了一种从 RGBD 图像中将典型、通常杂乱无章的室内场景解析为地板、墙壁、支撑面和对象区域, 并恢复其支撑关系的方法, 还创建了一个新颖的整数规划公式来推断物理支撑关系, 并提供了一个包含 1449 张 RGBD 图像, 捕获了 464 个具有详细注释的不同场景的新数据集。实验证明了该方法在复杂场景中推断支撑关系的能力, 并验证了 3D 场景提示和推断支持能够实现更好的对象分割效果; 第二篇是来自华盛顿大学、马普研究所、佐治亚理工学院 2012 年发表的论文 A Naturalistic

Open Source Movie for Optical Flow Evaluation^[8]。该文章基于虚拟 3D 动画短片生成一种新的光流数据集, 以克服光流算法难以在现实数据上进行训练和测试的难题, 并基于此数据集对主流光流算法进行有效评估, 以促进光流算法的改进。文章开展的有效实验也证明了该数据集与现实场景数据集的相似性。

PAMI Everingham Prize. 该奖项旨在纪念英国计算机视觉领域专家、The PASCAL Visual Object Classes (VOC) 数据集的主要贡献者以及该比赛项目的发起人 Mark Everingham, 由 IEEE 计算机协会模式分析与机器智能 (PAMI) 技术委员会颁发, 以表彰对计算机领域社区做出无私贡献的研究者或研究团队。本届获奖者是美国圣母大学计算机科学与工程副教授 Walter J. Scheirer, 其主要研究领域包括开集识别、视觉识别的极值理论模型以及受生物启发的学习算法。此外, UCF101 (2012 年) 和 HM51 (2011 年) 动作识别数据集团队的全体成员也获得该奖项。

Young Researcher Award. 该奖项由欧洲计算机视觉联盟 (ECVA) 设立, 旨在鼓励年轻研究人员在计算机视觉方面的杰出研究成就。今年该奖项颁发给欧洲联邦理工学院 (EPFL) 计算机科学助理教授 Amir Zami。此外本次会议还补发了去年 (2021 年) 青年学者奖, 图宾根大学计算机科学教授 Zeynep Akata 获此殊荣。

ECVA PhD Award. 该奖项旨在鼓励和表彰近两年完成博士学位论文的杰出研究者, 通常来说每年会有两个获奖名额, 并在接下来的 ECCV 会议上颁发。在本次大会中牛津大学的 Triantafyllos Afouras、摩德纳雷焦艾米利亚大学的 Marcella Cornia、慕尼黑工业大学的 Iro Laina 和马克斯·普朗克计算机科学研究所的 Yongqin Xiang 共四位研究者凭借 2020 年和 2021 年的博士学位论文获得该奖项。

五、总结展望

本年度 ECCV 大会中 3D 视觉分析、目标检测 (跟踪)、图像修复、图像分割、多模态分析、视觉 Transformer、无监督方法、对比学习、小样本学习等领域保持较高热度。综合近年来 ECCV 收录论文情况来看,

会议越发重视实际应用问题的解决, 如针对数据量缺乏现实, 研究小样本学习、迁移学习方法; 针对数据标注成本高昂问题, 发展半监督、无监督领域自适应研究; 用多模态方法, 尽可能使用各种模态(视角)数据来优化提高模型性能; 3D 视觉的检测与分割、视频理解等技术的发展, 把模型应用从二维的、静止的场景拓展到更贴合实际的三维的、动态的场景; 开展深度模型泛化性研究, 以提升模型的环境适应性。笔者认为, 计算机

视觉技术发展的目的是能解决工业应用和人们生活中的实际需求, 因此如何推动模型落地应用, 如何让模型从识别到理解、从大数据驱动到小样本学习、从单一静止场景到多模态多风格场景, 是计算机视觉任务未来的发展方向, 也是开展科研工作的重要切入点。

责任编辑 魏秀参

参考文献

- [1] Xingjian Zhen, Zihang Meng, Rudrasis Chakraborty, Vikas Singh. On the Versatile Uses of Partial Distance Correlation in Deep Learning. ECCV 2022.
- [2] Garvita Tiwari, Dimitrije Antic, Jan E. Lenssen, Nikolaos Sarafianos, Tony Tung, Gerard Pons-Moll. Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields, ECCV2022.
- [3] Ishit Mehta, Manmohan Chandraker, Ravi Ramamoorthi. A Level Set Theory for Neural Implicit Evolution under Explicit Flows. ECCV2022.
- [4] Junfeng Wu, Qihao Liu, Yi Jiang, Song Bai, Alan Yuille, Xiang Bai. In Defense of Online Models for Video Instance Segmentation. ECCV 2022.
- [5] Zhihao Li, Jianzhuang Liu, Zhensong Zhang, Songcen Xu, Youliang Yan. CLIFF: Carrying Location Information in Full Frames into Human Pose and Shape Estimation. ECCV2022.
- [6] Zhengqi Li, Qianqian Wang, Noah Snavely, Angjoo Kanazawa. InfiniteNature-Zero: Learning Perpetual View Generation of Natural Scenes from Single Images. ECCV2022.
- [7] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, Rob Fergus. Indoor Segmentation and Support Inference from RGBD Images. ECCV 2012.
- [8] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, Michael J. Black. A Naturalistic Open Source Movie for Optical Flow Evaluation. ECCV 2012.



祁磊

东南大学计算机科学与工程学院 PLAM 实验室助理研究员(至善博士后), 2020 年博士毕业于南京大学计算机科学与技术系。主要研究方向为计算机视觉、模式识别。

Email: qilei@seu.edu.cn