

顶会观察

CVPR 2022

南京理工大学 李俊

国际计算机视觉与模式识别大会 (IEEE/CVF Computer Vision and Pattern Recognition Conference, CVPR) 是计算机视觉和模式识别的顶级会议之一，与 ICCV 和 ECCV 并称为计算机视觉领域三大顶会。CVPR 的学术影响力越来越大，不仅是中国计算机学会推荐的人工智能领域 A 类国际学术会议，而且还在 Google Scholar 期刊与会议影响力榜单中排名第四(仅次于 Nature、NEJM 和 Science)。相比于去年，今年大会组委会仍有不少华人学者：香港科技大学权龙教授担任 General Chair，Wormpex AI Research 副总裁华刚博士担任 Program Chair，Google 研究科学家 Boqing Gong 担任 Tutorials Chair，旷视科技首席科学家孙剑博士等百名华人担任 Area Chairs。本届 CVPR 大会于 2022 年 6 月 19 日至 24 日在美国路易斯安那州新奥尔良举办，包括 4 天的正会和 2 天的 Workshops & Tutorials。

一、会议亮点

线下线上混合形式：自新冠疫情流行以来，CVPR 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，共有 9981 人注册参会，其中 5641 人现场参会，4340 人以线上方式参会。今年与以往不同，每篇论文都要以海报形式和虚拟会议两种方式进行工作展示及技术交流。海报形式：对于线下参会的作者，与 2019 年以前一样海报张贴；对于线上参会的作者，主办方提供了海报打印服务并由志愿者张贴。虚拟会议：作者准备一个五分钟的预录制视频和一张海报的文件，在会议平台上展示工作。为了

方便交流，主办方还为每篇论文安排了线上、线下、同步和异步方式与参会者进行充分的交流。

严格的评审机制：今年会议组织方邀请了 304 位专家作为领域主席(area chair)，包括 4 位紧急领域主席。同时，组织方还邀请了 6427 位从业者作为审稿人参与论文评审，包括 1723 位紧急审稿人。在经验分布上，学生审稿人占比 31%，教职/研究人员审稿人占比为 69%。更重要的是，每篇论文会由 3 位 AC 一起处理(包括至少一位资深 AC 和至少两位领域内专家)，其中 AC 的分配是由 TPMS 和学科领域自动匹配产生。每篇论文至少分配 3 位审稿人，其中至少 2 位为主 AC 推荐。主 AC 负责主持审稿线上会议，AC 之间相互审核报告，核查错误，并详细讨论审稿意见，最终向作者发出通知。

开源代码：随着 CVPR 影响力的进一步提升，越来越多的研究工作发布了代码或数据，开源已经成为趋势。据 GitHub 项目的不完全统计，CVPR 2020 和 CVPR 2021 均有超 300 篇论文开源了代码。更惊喜的是，据 paperdigest 团队统计，今年已经超过 600 篇论文公开了源代码。代码开源的优势明显，免费透明，不仅可以增加研究者之间的协作机会，而且还能提升研究工作的影响力。当然，也存在一定的未知风险，这需要大家共同努力完善开源代码。

除了上述会议亮点外，大会现场缅怀了孙剑博士，通过一段视频带领大家一起回忆了孙剑博士科研成长之路。最佳(学生)论文、Longuet-Higgins 奖，青年研究者奖，Thomas S. Huang 纪念奖将在后面介绍。

二、录用情况

CVPR 2022 收到的有效投稿和录用数量都有显著提高,大会共收到了 8161 篇有效投稿,最终接收了 2064 篇论文,接收率约为 25.3%。相较于 2021 年,今年 CVPR 的投稿量提升 15%(1068 篇),录用率也有所上升,录用论文数量提升约为 24.3%(403 篇)。其中,有 342 篇论文录用为 Oral Presentations,比去年增加 47 篇,Oral 率约为 4.1%(与去年基本持平)。CVPR2022 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、底层视觉(超分、恢复、去雾等)、深度学习与表示学习、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、场景分析和理解、无监督学习、行为识别、数据集与评估等方向。在 CVPR 2022 录用的论文中,识别、3D 视觉、图像与视频的生成、深度学习与表示学习四个方向的论文数量最多,均超过 150 篇。值得关注的是,来自中国学术机构与工业界企业取得了相当不俗的业绩。据公开接受列表,按单位统计,清华大学共 113 篇论文入选,中国科学院共有 91 篇论文入选,腾讯有 83 篇论文入选,阿里巴巴也有 67 篇论文入选。另据报道,商汤科技及联合实验室共 71 篇论文入选、南洋理工大学与香港理工大学多媒体实验室吕健勤教授团队有 18 篇论文入选、上海交通大学马利庄教授团队有 14 篇论文入选等。此外,谷歌与脸书在本次会议中仍有不错表现,分别有 89 篇和 73 篇论文入选。

三、主题报告

本次 CVPR 2022 会议邀请了三位 Keynote 演讲者,报告内容涵盖类人智能、整合 AI 和视觉外观理解。

Learning to See the Human Way. 麻省理工学院脑与认知科学系教授 Josh Tenenbaum 报告并讨论了如何建立类人智能的计算机视觉系统。近年来,计算机视觉是已经取得了瞩目的成绩,然而我们仍然无法拥有类人智能的机器系统(观察一张图或真实世界,就能轻松认识一切事物)。报告者受人类视觉和视觉认知系统的启发,引导出一种交替方法去建构可实践的机器视觉系统。基于可微与概率编程技术,此方法利用一些经典范式(如逆图形、综合分析、最佳解释推断等)理解视觉。虽然报

告者展示了少许机器视觉成功的例子,但不幸的是,建构像人一样观察世界的机器系统,依旧困难重重。最后,报告者还指出了这一领域一些大挑战任务。

Toward Integrative AI with Computer Vision. Microsoft 技术会士、Azure AI 首席技术官黄学东描述并探讨了面向计算机视觉的整合 AI。当不相关的 AI 任务数量快速增长时,如何整合 AI 成为了一个重要的问题。报告者分享了一种多语言多模式的整合 AI 方法,即利用完整的语义表示统一在语音、语言和视觉的多种任务。特别是,当应用到计算机视觉任务时,报告者正在发展一种基础模型(Florence),通过大尺度的图像和语言预训练,提出语义层概念,并在共同视觉任务(如识别、检测、分割等)中提升零样本/少样本的学习能力。通过搭建语义表示与视觉下游任务的关系,Florence 不仅在 COCO, VQA and Kinetics-600 等数据库上取得了 state-of-the-art 实验结果,而且还发现了图形理解的新结果。最后,报告者还预见语义层将赋予计算机视觉有超越视觉感知的能力,而且能够流畅地连接像素到人类智能的核心(意图、推理、决策)。

Understanding Visual Appearance from Micron to Global Scale. 康奈尔大学 Ann S. Bowers 计算与信息科学学院院长 Kavita Bala 介绍了从微观到宏观的视觉外观理解。计算机视觉的核心是,利用视觉输入,让计算机理解和发掘我们生活的真实世界,包括形状、材质、场景、活动等。同时,我们可以在微观或宏观尺度下观察这个真实世界,例如:微米级的 CT 图、行星级的卫星图。因此,我们希望计算机既能从微观角度深层理解个体目标外观,又可以宏观地理解世界规模的事件。在本报告中,Kavita 介绍了团队在视觉理解方向的最新成果,包括真实视觉外观与渲染的图模型,形状与材质的重建,世界规模发现的视觉模式等。最后,报告者还探讨了这一领域的发展趋势。

四、热点论文

2022 年度最佳论文奖评审委员会由 10 名国际权威学者组成,其中包括两名华人学者:北京大学的林宙辰教授和宾夕法尼亚州立大学 Yanxi Liu 教授。相比 2021 年,增加了两位评审委员。本年度大会共评选出

了 1 篇最佳论文, 1 篇最佳学生论文, 1 篇最佳论文提名, 1 篇最佳学生论文提名。

最佳论文: Learning to Solve Hard Minimal Problems^[1], 来自苏黎世联邦理工学院、华盛顿大学、佐治亚理工学院和捷克理工大学。在计算机视觉中, 很多任务都面临困难的几何优化问题, 例如 3D 重建、图像匹配、视觉里程计等。这类问题通常会简化为含有许多无意义解的最小化问题。本文提出了一种学习策略去选择一个起始的(问题-解)对, 可以有效地找到问题的兴趣解, 进而避免大量无意义解的计算。本文还开发了一个 RANSAC 求解器来计算三个校准相机的相对位置, 进而证实了此学习策略的有效性。

最佳论文提名: Dual-Shutter Optical Vibration Sensing^[2], 来自卡内基梅隆大学。视觉振动测量技术在声学、材料学科和生命科学领域中有着广泛的应用, 如: 远程捕捉音频、材质的物理属性、人体心率等。基于工作频率为 130Hz 的传感器, 本文提出了一种新的视觉振动测量方法, 不仅可以感知高速(高达 63kHz)的振动, 而且还能同时感知多个场景源。此方法依赖于同时使用两个分别装有滚动和全局快门传感器的相机来捕捉场景。其中, 滚动快门相机可以捕捉反映高速物体振动的扭曲斑点图像; 全局快门相机能捕捉斑点图案的未扭曲参考图像。最后, 通过捕捉由音源(如扬声器、人声和乐器)产生的振动, 验证了这种新的测量方法。

最佳学生论文: EPro-PnP: Generalized End-to-End Probabilistic Perspective-n-Points for Monocular Object Pose Estimation^[3], 来自同济大学和阿里巴巴集团。从单图像中定位 3D 物体是计算机视觉中长期存在且非常重要的问题。本文提出了一种新颖的 Perspective-n-Points(PnP)方法, 即用于一般端到端姿势估计的概率 PnP 层, 它可以输出 SE(3)流形上的姿势分布, 本质上是将分类 Softmax 引入连续域。2D-3D 坐标和对应的权重被视为中间变量, 并通过最小化预测和目标姿势分布之间的 KL 散度来学习。此方法的效果明显优于有竞争力的基准方法, 并且在 LineMOD 6DoF 姿势估计和 nuScenes 三维物体检测基准上, 缩小了基于 PnP 的方法与特定任务最佳性能之间的差距。

最佳学生论文提名: Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields^[4], 来自哈佛大学和谷歌研究院。近年来, 神经辐射场(NeRF)是一种流行的视图合成技术。虽然它擅长表现具有平滑变化且视线依赖性外观的精细几何结构, 但往往不能准确捕捉和重现光泽表面的外观。本文提出了一种反射 NeRF(Ref-NeRF), 通过引入反射辐射的表示方法代替了 NeRF 的 MLP 网络参数, 并使用一系列空间变化的场景属性来构造这一网络函数。结果表明, 结合法线向量的正则器, Ref-NeRF 显著提升了镜面反射的真实性和准确性。

另外, 还有 29 篇论文入选最佳论文入围名单, 其中华人学者为第一作者的论文数量超过半数, 并且多篇论文也引起了广泛的讨论。脸书 AI 研究院的论文 Masked Autoencoders Are Scalable Vision Learners^[5] 利用随机遮盖输入图片的子块, 重建丢失块像素, 由此来预训练模型, 使得模型在下游任务中具有很好的泛化性能。中国科学院大学的 AnyFace: Free-style Text-to-Face Synthesis and Manipulation^[6] 提出了一种风格自由的文本到人脸方法, 去合成并编辑人脸。香港城市大学、德国马普所和斯坦福大学的 Learning to Deblur Using Light Field Generated and Real Defocus Images^[7] 提出了一种新颖的去焦距去模糊网络, 它可以利用强度并克服光照的缺点。

五、大会获奖和竞赛奖

Longuet-Higgins Prize。该奖以理论化学家和认知科学家 H. Christopher Longuet-Higgins 的名字命名, 它是由 PAMI 技术委员会颁发的计算机视觉基础贡献奖, 表彰十年前对计算机视觉研究产生了重大影响的 CVPR 论文。今年获奖论文是来自卡尔斯鲁厄理工学院和丰田工业大学芝加哥分校的论文 Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite。

Young Researcher Awards。该奖旨在表彰计算机视觉领域的青年研究人员, 鼓励年轻科学家继续做出开创性工作。今年获奖者分别是康奈尔大学计算机科学系助理教授 Bharath Hariharan 和普林斯顿大学计算机

科学系助理教授 Olga Russakovsky。

Thomas S. Huang 纪念奖。自 2021 年起，为了缅怀一代宗师、华人计算机视觉泰斗 Thomas S. Huang(黄煦涛)教授，由 PAMITC 奖励委员会选出，以表彰在计算机视觉研究、教育和服务方面被公认为楷模的研究人员。今年第二届获奖者是斯坦福大学计算机科学系教授李飞飞。研究领域包括认识启发的 AI，机器学习、深度学习、计算机视觉和 AI+医疗健康。

竞赛奖。在 CVPR 2022 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不俗的成绩。西安电子科技大学在 Woodscape 鱼眼目标检测挑战和农业视觉 CropHarvest 竞赛中，分别获得了冠军。南京理工大学在 NTIRE 的竞赛中，获得了高效超分辨率赛道和 4 倍超分辨率赛道的双冠军。腾讯太极团队在轻量化 NSA 竞赛中，获得了超网络赛道冠军。美团在届细粒度视觉分类竞赛中，获得了植物标本识别赛道和大规模跨模态

商品图像召回赛道的双冠军。联想研究院在 BDD100K MOT 挑战赛中，获得平均多目标跟踪准确度的冠军。

六、 总结展望

本年度 CVPR 大会中识别、3D 视觉、图像与视频的生成、深度学习、表示学习、Transformer 等领域依旧保持高热度。相比于 2021 年，底层视觉热度显著回升，计算摄影热度有所下降。更值得关注的是，计算机通过视觉传感器感知并认知真实物理世界，CVPR 越来越注重解决真实场景下的视觉问题，从 2D 到 3D，从微观到宏观，从识别到类人智能，甚至是超越类人智能。从第三方角度观察，计算机视觉领域还面临许多挑战：从物理世界到传感器与机器认知过程中，现有方法的能力上界是什么、人类已经发现的规律如何辅助机器认知物理世界、机器如何探知人类未曾发现的物理世界规律等。笔者认为回答好上述问题将会是计算机视觉的新机遇，从而更好地迈向更高层级的智能。

责任编辑 魏秀参

参考文献

- [1] Petr Hruby, Timothy Duff, Anton Leykin, Tomas Pajdla. Learning to Solve Hard Minimal Problems. CVPR 2022.
- [2] Mark Sheinin, Dorian Chan, Matthew O'Toole, and Srinivasa G. Narasimhan. Dual-Shutter Optical Vibration Sensing, CVPR2022.
- [3] Hansheng Chen, Pichao Wang, Fan Wang, Wei Tian, Lu Xiong, Hao Li. EPro-PnP: Generalized End-to-End Probabilistic Perspective-n-Points for Monocular Object Pose Estimation. CVPR2022.
- [4] Dor Verbin, Peter Hedman, Ben Mildenhall, Todd Zickler, Jonathan T. Barron, Pratul P. Srinivasan. Ref-NeRF: Structured View-Dependent Appearance for Neural Radiance Fields. CVPR 2022.
- [5] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollar, Ross Girshick. Masked Autoencoders Are Scalable Vision Learners. CVPR2022.
- [6] Jianxin Sun, Qiyao Deng, Qi Li, Muye Sun, Min Ren, Zhenan Sun. AnyFace: Free-style Text-to-Face Synthesis and Manipulation. CVPR2022.
- [7] Lingyan Ruan, Bin Chen, Jizhou Li, Miu-Ling Lam. Learning to Deblur Using Light Field Generated and Real Defocus Images. CVPR 2022.



李俊

南京理工大学计算机科学与工程学院教授，国家高层次青年人才计划入选者，江苏省高层次人才计划入选者。主要研究方向为计算机视觉、机器学习、深度学习与计算物理/化学交叉等。Email: junli@njust.edu.cn