

顶会观察

CVPR 2023

南京理工大学 周涛

国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 是世界顶级的计算机视觉会议 (三大顶会之一, 另外两个是 ICCV 和 ECCV)。CVPR 会议每年举办, 旨在推动计算机视觉、模式识别及相关领域的发展, 展示这些领域最新的研究成果。会议涵盖的主题包括图像处理、模式识别、计算几何、机器学习、数据挖掘以及其他相关领域。目前, 根据 Google Scholar, CVPR 的 h5 指数所有会议或出版物中位列第四, 仅次于《自然》、《科学》和《新英格兰医学杂志》。过去 5 年被引用最多的出版物中, CVPR 位列第 4。今年大会组委会仍有不少华人学者, 美国国家工程院院士、斯坦福大学李飞飞教授担任 General Chair, 南京大学吴建鑫教授和 ETH Zurich 助理教授汤思雨担任 Tutorials Chairs 等。本届 CVPR 大会于 2023 年 6 月 18 日至 22 日在加拿大温哥华国际会议中心举办, 包括 2 天的研讨会 (workshops) & 讲习班 (tutorials) 和 3 天的主会 & 博览会 (Expo)。

一、投稿与录用情况

CVPR 的论文投稿量持续增加, 从 CVPR2019 的 5160 篇有效投稿增长到 CVPR2020 年 6656 篇, 到 CVPR2021 的 7500 篇, 再到 CVPR2022 的 8161 篇。而今年又再创新高, 达到了 9155 篇, 收录 2360 篇, 接收率为 25.78%。相比 2022 年, 今年 CVPR 的投稿量提升 12.12% (994 篇); 相比 2021 年, 投稿量提升 30.06% (2116 篇)。今年投稿的论文主要研究主题包括多视图和传感器 (隐式表示, 大量 NeRF), 图像

和视频生成 (主要是扩散模型), 面部/身体/姿势估计, 持续学习, 多模态 (视觉 + 语言) 建模等。投稿量排在前十的研究方向包括: 1) 多视角和传感器下的三维重建; 2) 图像和视频的合成与生成; 3) 面部、身体姿势和手势运动; 4) 迁移学习、元学习、小样本学习、持续学习或长尾学习; 5) 识别: 分类、检测、检索; 6) 视觉、语言和推理; 7) 低层次视觉; 8) 分割、分组和形状分析; 9) 深度学习架构和技术; 10) 多模态学习。此外, 在医学和生物视觉、细胞显微镜技术领域有 420 篇投稿 (53 篇录用), 反映出 AI 医疗相关的研究人员也在关注偏向计算机视觉和人工智能的顶级会议 CVPR。

在录用的论文中, 有 12 篇入围本届最佳论文候选名单 (award candidate), 占提交论文的 0.13%, 录用论文的 0.51%; 以及 235 篇录用为 highlights 论文, 占提交论文的 2.6%, 录用论文的 9.96%。值得注意的是, 在 12 篇最佳候选论文中, 有部分来自于武汉大学、香港中文大学、香港科技大学、上海人工智能实验室、商汤科技、华为等国内大学及研究机构。

二、大会概览

会议形式: 自新冠疫情流行以来, CVPR2023 是第二次进行线下举办, 大约 8300 多人注册参会, 其中线下人数近 7000。可以看出, 今年参会人数也是自疫情以来创历史最高。与以往不同的是, 今年的 CVPR 采用单一轨道模式, 让所有参会者都能参加所有活动。每篇论文都以海报形式展示。此外, 所有最佳候选论文额外进行全体大会口头报告。每个参会者都获得个性化的数字计划, 以便轻松浏览每个海报展示单元中的约 400 个

海报。为了方便交流，虚拟平台为每篇论文提供了海报展示、视频和聊天交流等功能。另外，无法线下的参会人员可以通过在线直播观看大会活动。

评审机制：CVPR2023 的审稿过程由 400 多名领域主席 (area chair) 负责管理，而今年新增了 30 位高级领域主席。高级领域主席在许多方面提供了帮助，主要负责对于处理困难案例的裁决、应急情况的处理、选择重点论文和评选奖项候选人等。今年共有 6625 名审稿人，每篇论文至少收到 3 份评审意见。与往年一样，在收到初步评审后，作者对意见进行回复，然后 AC 负责组织线上讨论并给出初步推荐意见。最后，由领域主席三人小组共同作出每篇论文的接收/拒绝决定。

主题报告：本届 CVPR 2023 的主题演讲主要围绕人工智能领域的下一步发展方向展开讨论。麻省理工学院 Rodney Brooks 教授带来了“Recycling old vision ideas in a modern computational world”的报告。许多计算机视觉领域的想法一次又一次地被重新审视，包括基于神经计算的现代计算机视觉。特别是当前计算硬件取得了令人难以置信的发展，报告中主要探讨这样强大的计算能力是否能够被忽视的旧想法注入新的动力。华盛顿大学 Yejin Choi 教授带来了“An AI Odyssey: the Dark Matter of Intelligence”主旨报告。在追求更强大的智能系统的过程中，我们发现了一种被称为“智能的暗物质”的现象，这是指 AI 中存在的巨大而复杂的知识和能力，但我们仍然对其运作方式和内部机制知之甚少。报告主要探讨这个令人困惑且引人瞩目的领域，并探索解决这个问题的途径。此外，Meta AI 的研究科学家 Larry Zitnick 带来的“Modeling Atoms to Address Our Climate Crisis”报告。气候危机是当今世界面临的巨大挑战之一，而模拟和理解原子水平的过程对于解决这一问题至关重要。该报告主要探讨使用原子模型并结合人工智能技术来应对当前面临的气候危机的重要性。

三、最佳论文

2023 年度最佳论文奖评审委员会由计算机视觉相关领域的 7 名国际权威学者组成。本届 CVPR 大会共评选出 5 篇论文获奖，其中 2 篇最佳论文，1 篇最佳学生

论文，以及最佳学生论文提名和最佳论文提名各 1 篇。

最佳论文：Planning-oriented Autonomous Driving^[1]，来自上海人工智能实验室、武汉大学、商汤科技。论文作者首次将感知、预测和规划等 3 大类主任务、6 小类子任务（目标检测、目标跟踪、场景建图、轨迹预测、栅格预测和路径规划）整合到统一的基于 Transformer 的端到端网络框架下，实现了全栈关键任务驾驶通用模型。该论文也是首次提出感知决策一体化的自动驾驶通用大模型 (UniAD)，开创了以全局任务为目标的自动驾驶大模型架构先河，为自动驾驶技术与产业的发展提出了新的方向。值得注意的是，这是近 10 年来计算机视觉三大顶级会议中 (CVPR、ICCV、ECCV)，第一篇以中国学术机构作为第一单位的最佳论文。

最佳论文：Visual Programming: Compositional visual reasoning without training^[2]，来自艾伦人工智能研究所 (Allen Institute for AI, AI2)。通用人工智能系统旨在为用户提供简单的自然语言界面与模型进行交互。构建这些系统的主要方法是采用大规模无监督预训练和监督多任务训练。然而，这种方法需要为每个任务准备精心策划的数据集，难以扩展到复杂任务的无限长尾部分。在本文研究中，研究人员探索使用大型语言模型来处理复杂任务的长尾部分，通过将这些任务用自然语言描述并分解为较简单的步骤，可以由专门的端到端训练模型或其它程序来处理。基于此，论文提出了 VISPROG，一种根据自然语言指令解决复杂和组合视觉任务的方法。只需要给出几个自然语言指令的示例和所需的高级程序，VISPROG 使用 GPT-3 中的上下文学习为任何新指令生成一个程序，然后在输入图像上执行该程序以获得预测结果。

最佳学生论文：3D Registration with Maximal Cliques^[3]：来自西北工业大学。三维点云配准是计算机视觉的关键研究问题之一，其目的是搜索最佳位姿以对齐一对点云。本文作者提出了一种基于最大种群的 3D 配准方法，核心是放宽先前的最大种群约束，在图中挖掘更多局部一致性信息来生成姿态假设。首先构建了一个兼容性图来表示初步对应关系之间的亲和矩阵，然后

搜索最大种群，每个种群代表一个一致集。然后执行基于节点的种群选择，每个节点对应图权重最大的最大种群。对于选择的种群，通过奇异值分解算法 (SVD) 计算变换假设，并选择最佳假设进行配准。在多个数据集 (U3M、3DMatch、3DLoMatch 和 KITTI) 上进行了大量实验，结果表明该方法能够有效提高配准性能，且优于各种最先进的方法。

最佳论文提名: DynlBaR: Neural Dynamic Image-Based Rendering^[4], 来自谷歌和康奈尔大学。

从动态场景的单目视频中合成新视图是一个更具挑战性的动态场景重建问题。基于时间变化的神经辐射场 (也称为动态 NeRF) 的最先进方法在这个任务上展示了令人印象深刻的结果。然而，对于具有复杂物体运动和不受控制的相机轨迹的长视频，这些方法可能会产生模糊或不准确的渲染图像，限制了它们在实际应用中的使用。本文提出了一种新方法 (DynlBaR)，通过采用基于体素的图像渲染框架来处理这些局限性，以一种感知场景运动的方式聚合来自附近视角的特征来合成新视角。该方法在能够建模复杂场景和视角依赖效果方面保留了之前方法的优势，同时还能够从具有复杂场景动态和不受约束的相机轨迹的长视频中合成逼真的新视角。在动态场景数据集上的结果展示了该方法的先进性，并将方法应用于具有挑战性的相机和物体运动的野外视频中，在这些情况下，之前的方法无法生成高质量的渲染图像。

最佳学生论文提名: Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation^[5], 来自谷歌和波士顿大学。大型文本到图像生成模型在人工智能的发展中取得了显著的突破，能够根据给定的文本提示 (Prompt) 生成高质量且多样化的图像。然而，这些模型缺乏模仿或再现给定参考集中主体外观的能力，特别是对个人定制的物体，很难实现个人物体在风格、造型等方面的实现能力。在本研究中，研究人员提出了一种用于“个性化”的文本到图像的扩散模型。仅提供主体的几张图像作为输入，对预训练的文本到图像模型进行微调，使其学习将唯一标识符与特定主体绑定。一旦主体嵌入到模型的输出领域

中，该唯一标识符可以用于合成在不同场景中具有上下文的主体的新的逼真图像。通过利用模型中嵌入的语义先验和新的自生类别特定先验保持损失，该模型可以合成出在参考图像中不存在的多样化场景、姿势、视角和光照条件下的图像。最后，作者将该模型应用到多个挑战的任务上，包括主题重构、文本引导视图合成和艺术渲染。此外，本文作者为这个主题驱动生成的新任务提供了一个新的数据集和评估协议。

四、大会奖项

Longuet-Higgins Prize: Longuet-Higgins 奖由 IEEE 模式分析和机器智能 (PAMI) 技术委员会在每年的 CVPR 上颁发，以表彰在计算机视觉领域所作出的基础贡献。该奖项旨在表彰十年前对计算机视觉研究产生重大影响的 CVPR 论文，它是以理论化学家和认知科学家 H. Christopher Longuet-Higgins 的名字来命名的。本届 Longuet-Higgins 奖颁发给了 2013 年 CVPR 论文 “Online Object Tracking: A Benchmark”。

Young Researcher Awards: 青年研究员奖是颁发给在获得博士学位后 7 年内，对计算机视觉领域做出杰出研究贡献的一位或两位青年研究人员。本屆的青年研究员奖颁发给了 Facebook 人工智能研究院研究科学家 Christoph Feichtenhofer 和佐治亚理工学院助理教授 Judy Hoffman。

Thomas S. Huang Memorial Prize: 这个奖项从 2020 年开始设立，是为了纪念已故的黄煦涛教授。黄教授是在 CV 和图像处理领域做出大量贡献的先驱学者。本屆该奖项颁给了加利福尼亚大学伯克利分校教授 Alyosha Efros，他的研究兴趣包括计算机视觉、计算机图形、计算摄影、机器学习及人工智能等。

五、Tutorials和Workshops

Tutorials: 本屆 CVPR 包括 33 个 tutorials，例如 “Large-Scale Visual Localization”，“Vision Transformer: More is different”，“Object localization for free: Going beyond self-supervised learning”，“Denoising Diffusion Models: A Generative Learning Big Bang”，“Recent advances

in anomaly detection”，“Recent Advances in Vision Foundation Models”，“ML Systems for Large Models and Federated Learning”，“All Things ViTs: Understanding and Interpreting Attention in Vision”等。通过这些也不难发现，当前的研究热点仍然围绕视觉定位、Transformer、自监督学习、扩散模型、视觉基础模型、联邦学习、注意力机制等。因此，这些 tutorials 资料，系统地为我们梳理了相关领域的发展历史脉络、知识结构、研究进展以及未来的发展方向。

Workshops: 本届 CVPR 共包含 100 个 workshop，其内容涵盖非常丰富，既有关关注计算机视觉基础任务，也有学习范式，还有特定应用。粗略选取收录论文数量相对较多的一些 workshop，例如 1) 8th New Trends in Image Restoration and Enhancement Workshop and Challenges (77 篇)，主要聚焦图像与视频恢复与增强研究；2) 7th AI City Challenge Workshop (35 篇)，主要聚焦人工智能城市中的实体零售业务和智能交通系统；3) 4th International Workshop on Event-Based Vision (33 篇)，聚焦于智能相机和处理这些传感器数据的算法研究；4) 6th Efficient Deep Learning for Computer Vision (29 篇)，聚焦于高效深度学习算法、以及低资源学习等方面的研究；5) 2nd Workshop on Learning With Limited Labelled Data for Image and Video Understanding (27 篇)，聚焦于面向标记数据受限情况下的图像与视频理解算法研究；6) Generative Models for Computer Vision (24 篇)，聚焦生成模型在计算机视觉领域的应用和发展；7) 19th CVPR Workshop on Perception Beyond the Visible Spectrum (22 篇)，聚焦复杂场景下的目标检测、识别和跟踪等算法研究；8) Workshop on Autonomous Driving (16 篇)，聚焦自动驾驶的运动预测研究；9) EarthVision Large Scale Computer Vision for Remote Sensing Imagery (16 篇)，聚焦计算机视觉和机器学习技术在遥感数据上的研究；10) 4th Workshop on Continual Learning in Computer Vision (14 篇)，聚焦持续学习在计算机视觉中的应用。可以看出，部分 workshop

已经连续举办多届，而有些是第一届举办。此外，也不难发现这些 workshop 更关注计算机视觉中的许多应用场景，例如图像恢复与增强、智慧城市、目标检测识别以及自动驾驶中轨迹预测等问题。

此外，在 CVPR2023 研讨会上举办的各项挑战赛中，国内学术界和工业界在不同任务上取得了备受关注的业绩。西安电子科技大学在在多个赛道取得冠军 (CVPR 2023 FMDC 挑战赛中零样本/少样本图像分类、VizWiz 视觉问答、跨模态视频检索等)。合肥工业大学 CVPR 2023 国际情感行为分析挑战赛冠军。南京理工大学在细粒度视觉分类研讨会挑战赛的多个赛道多个赛道取得冠军。联想研究院 PC 创新与生态系统团队在 CVPR 2023 WAD BDD100K MOTs (多目标追踪与分割) 挑战赛中获得冠军，在 CVPR 2023 WAD ARGOVERSE 挑战赛中，斩获端到端运动预测赛道中的 3D 运动预测，3D 多目标跟踪，3D 目标检测三项子赛道冠军等。旷视研究院在 CVPR 2023 自动驾驶国际挑战赛中 OpenLane Topology 赛道夺得冠军。中国电信天翼云 AI 团队在 CVPR 2023 Workshop on Foundation Model 中多任务大模型赛道获得冠军。小视科技在第一届 CVPR 2023 WorkShop 大模型挑战赛中，取得了 A 榜第一名。腾讯优图荣获 CVPR 2023 视觉异常检测挑战赛冠军。

六、总结展望

本届 CVPR 的研究热点和关键词包括 3D 视觉、图像和视频的合成与生成 (扩散模型)、识别、迁移学习/元学习/小样本学习、视觉/语言推理、深度学习、Transformer 等。值得关注的是，神经-符号方法重新回到人们视野，它结合了神经网络和符号推理，在扩展人工智能能力方面展示了潜力。此外，数据驱动方法和迁移学习对于提高模型性能和泛化能力起着重要作用。更值得关注的是，将多个任务整合到一个框架中可以提高性能和协调能力。这种趋势强调开发统一的框架来优化多目标，实现复杂场景下不同任务之间更好的协调和性能。此外，大模型技术引起了广泛关注，数据集和基准对于推动计算机视觉研究和评估算法性能仍然至关重要。由于大模型在自然语言处理和计算机视觉等领域取

得成功，研究人员开始探索将其引入计算机视觉任务中，这也是一个令人期待的趋势。此外，通用人工智能 (AGI) 旨在模仿人类的认知能力，包括感知、理解、学习和推理，跨越更广泛的领域解决复杂问题并独立做出决策。

因此，我们期待大模型与 AGI 将推动计算机视觉技术的发展，并为解决复杂视觉任务和实际应用提供新的思路和方法。

责编委 魏秀参

参考文献

- [1] Hu, Yihan, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai et al. Planning-oriented autonomous driving. CVPR 2023.
- [2] Gupta, Tanmay, and Aniruddha Kembhavi. Visual programming: Compositional visual reasoning without training. CVPR 2023.
- [3] Zhang, Xiyu, Jiaqi Yang, Shikun Zhang, and Yann. ing Zhang. 3D Registration with Maximal Cliques. CVPR 2023.
- [4] Li, Zhengqi, Qianqian Wang, Forrester Cole, Richard Tucker, and Noah Snavely. Dynibar: Neural dynamic image-based rendering. CVPR 2023.
- [5] Ruiz, Nataniel, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. CVPR 2023.



周涛

南京理工大学计算机科学与工程学院教授、博导
主要研究方向为医学图像分析、计算机视觉、机器学习等。
Email:taozhou@njust.edu.cn