

顶会观察

## ICCV 2025

复旦大学 陈静静

**国**际计算机视觉大会 (IEEE International Conference on Computer Vision, ICCV) 是计算机视觉领域的顶级会议之一,与 CVPR 和 ECCV 并称为计算机视觉领域三大顶会。ICCV 为两年一届 (奇数年举办),其学术影响力持续攀升,是中国计算机学会 (CCF) 推荐的人工智能领域 A 类国际学术会议,也是 Google Scholar 期刊与会议影响力榜单中的常客。本届 ICCV 大会于 2025 年 10 月 19 日至 23 日在美国夏威夷州檀香山的夏威夷会议中心举办。相比往届,ICCV 2025 的组委会阵容依旧强大且具有广泛的国际代表性,其中不乏华人学者的身影:上海科技大学的虞晶怡教授担任大会主席,同其他几位主席包括 Hilde Kuehne (图宾根大学)、Gerard Medioni (亚马逊)、Dimitris Samaras (石溪大学) 和 Ramin Zabih (康奈尔大学) 共同组织了本次盛会。6 名程序委员会主席中有 2 名华人学者:Google DeepMind 资深主任研究科学家孙德庆和百度视觉技术首席架构师王井东。

## 一、会议亮点

**回归热带与线下交流:** 本届 ICCV 在风景秀丽的夏威夷檀香山举办,这也是继 CVPR 2017 之后,顶级视觉会议再次回归夏威夷。会议采用线下为主、线上结合的混合模式,为全球学者提供了绝佳的交流环境。据主办方统计,数千名研究人员亲临现场,享受了高密度的学术讨论与热带风情相结合的独特体验。

**限投令与全员评审机制:** 面对近年来 AI 会议投稿量爆炸式增长挑战,ICCV 2025 首次引入了“每位作者最多投稿 25 篇”的硬性上限 (Paper Cap),旨在抑制为了“刷量”而产生的低质投稿,鼓励研究者专注于更有深度的工作。同时,大会实施了更严格的义务评审

制度,要求所有符合资格的作者 (除非担任 AC 等职务) 必须参与审稿,且对于未能按时提交审稿意见的作者实施了严厉的“Desk Reject”连坐惩罚,这一举措显著提升了审稿效率。

**前沿技术爆发:** 随着生成式人工智能 (GenAI) 的持续演进,本届会议上关于基础模型、3D 生成、视频生成以及具身智能的论文数量占据了主导地位。特别是结合大语言模型与视觉感知的多模态研究,成为了会场讨论最热烈的焦点。

## 二、录用情况

**投稿与录用:** 大会共收到了 **11,239** 篇有效投稿,相较于 ICCV 2023 的 8,060 篇增长了约 **39.4%**。经过严格的评审,最终接收了 **2,701** 篇论文,录用率约为 **24.0%**,比上届的 26.8% 有所下降。在录用的论文中,仅有 **64** 篇被选为 Oral Presentations, **Oral 率仅为 0.57%**,不仅远低于往年,也显著低于 CVPR (通常约 3%-4%)。这使得本届 ICCV 的 Oral 资格成为极具含金量的荣誉。此外,另有 **263** 篇论文被选为 Highlights,约占投稿总数的 2.3%。

**评审团队:** 为了处理海量投稿,会议组织了庞大的评审团,包括 6 位程序主席、**510** 位领域主席以及 **11,859** 位审稿人。每篇论文至少经过了 3 位独立审稿人的评估及一轮 Rebuttal。

**中国力量:** 来自中国 (含港澳台) 的投稿量继续领跑全球,占比约为 43%。在录用论文中,中国高校及企业表现亮眼。据不完全统计,清华大学以超过 90 篇的录用数量位居高校前列,上海科技大学在虞晶怡主席的带领下也有超 30 篇入选。企业方面,百度、华为 (诺

亚方舟)、腾讯和阿里巴巴均有数十篇论文入选,显示了中国工业界在基础视觉研究上的深厚积累。

### 三、热门研究方向

根据投稿和录用分布,ICCV 2025 呈现出以下技术趋势:

**图像和视频合成和生成:** 这一方向的研究重心已不再满足于文本生成视频,而是集中在无需重训练的推理时引导与复杂的轨迹编辑上。例如, TITAN-Guide 和 TrajectoryCrafter 等工作实现了对摄像机运镜及物体运动轨迹的细粒度控制,解决了传统生成模型难以驾驭物理规律和长时连贯性的痛点。同时,动态帧率采样等高效生成技术被提出,旨在降低算力成本的同时保持视频的流畅度,标志着视频生成技术正逐步从“艺术创作”迈向可用于模拟真实物理世界的“世界模拟器”。

**基于多视角与传感器的 3D 重建:** 3DGS 凭借其高效性彻底取代了 NeRF 成为主流, FreeSplatter 和 StreamGS 等前沿工作突破了对密集视角和预先相机标定的依赖,实现了从未标定稀疏视角到流式输入的极速重建。研究前沿正迅速向 4D 动态场景和语义特征融合拓展,这不仅让重建过程更加实时,还赋予了 3D 场景语义理解能力,实现了从“静态几何复制”到机器人可用的“动态实时感知”的跨越。

**多模态大模型:** 社区开始系统性地验证原生多模态模型在参数扩展下的表现,证明了相比于简单的胶水式拼接,原生混合训练在算力充足时具有更优的扩展潜力。此外,视觉推理正从直觉式的模式识别向类似人类 System 2 的“慢思考”演进,通过引入长思维链来处理科学图表分析等复杂任务,推动模型从单纯的“看图说话”工具进化为具备深度逻辑能力的视觉智能体。

### 四、热点论文

2025 年度最佳论文奖评审委员会由 6 名国际权威学者组成,包括 1 名华人学者:中国科学院计算技术研究所陈熙霖研究员。本年度大会的竞争异常激烈,最终评选出了 1 篇最佳论文,1 篇最佳学生论文,以及 2 篇最佳论文提名。

**最佳论文 (Marr Prize):** Generating Physically

Stable and Buildable Brick Structures from Text<sup>[1]</sup>, 来自卡内基梅隆大学。在生成式 AI 从 2D 迈向 3D 物理世界的浪潮中,现有的文本生成 3D 模型往往忽略了物理约束,导致生成的物体无法在现实中制造或组装。本文提出了 BrickGPT,这是一个端到端的生成框架,能够根据文本提示生成既符合视觉语义又具备物理稳定性的乐高积木结构。作者构建了一个包含大规模物理稳定积木设计的数据集,并训练了一个自回归大语言模型来进行积木的序列预测。该研究创新性地引入了物理感知的回滚机制,在生成过程中实时修剪不稳定的结构,确保了生成的模型不仅美观,而且能够被机械臂或人类在现实中搭建出来。这一工作被认为是连接“生成式 AI”与“物理制造”的重要桥梁。

**最佳学生论文:** FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models<sup>[2]</sup>, 来自以色列理工学院。在基于文本的图像编辑领域,如何在修改图像语义的同时完美保留原图的结构细节(如物体轮廓、背景布局)一直是个棘手难题。现有的基于扩散模型的方法通常依赖于复杂 Inversion 过程来回溯噪声,这不仅计算昂贵,往往还伴随着重建伪影和由于数值误差导致的结构变形。本文基于流匹配 Flow Matching 理论,提出了一种创新的“免反演”编辑框架。该方法利用了预训练流模型天然的数学可逆性,通过求解 ODE 直接在源图像与目标文本条件之间建立精确的轨迹映射。实验表明,FlowEdit 彻底摒弃了传统方法中繁琐的优化与微调步骤,能够以极快的速度实现结构高度一致的图像编辑,为下一代交互式内容创作工具提供了基于流模型的高效新范式。

### 五、大会获奖和竞赛奖

**Helmholtz Prize:** Helmholtz Prize 由 IEEE 模式分析与机器智能 (PAMI) 技术委员会设立,每两年在 ICCV 大会上颁发一次。该奖项以 19 世纪德国物理学家赫尔曼·冯·亥姆霍兹 (Hermann von Helmholtz) 命名,旨在表彰十年前发表于 ICCV、并对计算机视觉领域产生基础性且深远影响的研究成果。本年度共有两篇论文获奖:

Delving Deep into Rectifiers: Surpassing

Human-Level Performance on ImageNet Classification<sup>[3]</sup>, 作者: 何恺明、张祥雨、任少卿、孙剑。该研究基于 PReLU 网络, 在 ImageNet 2012 分类数据集上取得了 4.94% 的 Top-5 测试错误率。相比 ILSVRC 2014 冠军模型 GoogLeNet 6.66% 的数值, 其性能提升幅度达 26%。这也标志着在视觉识别挑战中, 算法首次超越了 5.1% 的人类水平错误率。

Fast R-CNN<sup>[4]</sup>, 作者: Ross Girshick。该论文提出了多项革新, 利用深度卷积网络对候选区域进行高效分类, 在显著提升训练与推理速度的同时, 也优化了检测精度。数据显示, Fast R-CNN 在训练 VGG16 网络时, 速度是 R-CNN 的 9 倍, 测试速度更是快了 213 倍; 与 SPPnet 相比, 其训练速度快 3 倍, 测试速度快 10 倍。此外, 该模型在 PASCAL VOC 2012 数据集上也取得了更高的平均精度均值。

**Everingham Prize:** Everingham Prize 旨在表彰为计算机视觉社区做出重大且持续性贡献的个人或团队。该奖项为纪念 Mark Everingham 而设立, 意在激励后继者推动社区的整体发展。该奖项每年颁发一次, 奇数年于 ICCV、偶数年于 ECCV 颁发。今年的获奖者分别是 SMPL Body Model 团队以及 VQA 团队。

**Azriel Rosenfeld 终身成就奖:** Azriel Rosenfeld 终身成就奖表彰的是在整个职业生涯中为计算机视觉领域做出重大贡献, 并对该领域的发展产生非凡影响的研究学者。本年度获奖者 Rama Chellappa 是约翰霍普金斯大学电气与计算机工程及生物医学工程专业的彭博杰出教授, 并兼任数据科学与人工智能专项计划的临时主任。他在计算机视觉、模式识别及机器学习领域的卓越建树, 已对生物识别、智能汽车、法医学以及面部、物体和地形的二/三维建模等方向产生了深远影响。

## 六、总结展望

ICCV 2025 不仅是一次学术成果的集中展示, 更是对未来计算机视觉发展方向的风向标。从本届会议可以看出, 计算机视觉正加速从“看懂世界”向“生成世界”和“操作世界”演进。

展望未来, 随着计算能力的提升和算法的迭代, 视觉模型将更加通用化和物理化。研究者们正致力于解决大模型在真实物理环境中的落地问题, 以及如何让视觉系统具备类似人类的常识推理能力。对于中国学者而言, 在保持论文数量优势的同时, 在底层基础理论和原创性架构上的突破将是下一阶段的重要目标。

责任编辑 魏秀参

## 参考文献

- [1] Ava Pun, Kangle Deng, et al. Generating Physically Stable and Buildable Brick Structures from Text.. ICCV 2025 (Marr Prize).
- [2] Vladimir Kulikov, et al. FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models. ICCV 2025.
- [3] Kaiming He, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. ICCV 2015.
- [4] Ross Girshick, et al. Fast R-CNN. ICCV 2015.



## 陈静静

复旦大学智能机器人与先进制造创新学院副教授, 国家高层次青年人才计划入选者。主要研究方向为多媒体内容分析与理解、生成式人工智能安全、计算机视觉等。

Email: chenjingjing@fudan.edu.cn