

顶会观察

## ACM MM 2024

厦门大学 吴垚 曲延云

**国**际计算机图形学与多媒体领域会议（ACM International Conference on Multimedia，ACM MM）是计算机视觉和模式识别领域最重要的会议之一。这一会议由 ACM 主办，自 1993 年首次举办以来，已成为该领域内学术界和工业界交流的重要平台，也是中国计算机学会（CCF）推荐的 A 类国际学术会议。ACM MM 专注于多媒体技术领域的最新研究成果、技术创新和行业趋势。会议涵盖了多媒体内容的创建、处理、传输和交互等多个方面，旨在促进学术界和工业界在多媒体技术应用和产品开发方面的交流与合作。

## 一、会议概况

ACM MM 每年举办一次，此次为第 32 届多媒体国际学术会议。于 2024 年 10 月 28 日~11 月 1 日在澳大利亚墨尔本（Melbourne）市召开。本次会议的投稿数量、审稿人数和参会人数都是前所未有的。首先，ACM MM 2024 公布了今年的论文录用结果：2024 年共有 4385 篇投稿进入审稿阶段，相比去年增加了 42%，最终 1149 篇论文被录用，录取率为 26.20%。而在录用的投稿中，仅有 174 篇被进一步评选为 Oral，其接受率为 3.97%。会议议程紧凑而丰富，涵盖了多媒体技术领域的多个关键议题。会议举办了多场主题演讲、研讨会、技术教程等活动，深入探讨了机器学习与人工智能、多语言处理、云计算与虚拟化等前沿话题。

## 二、特邀报告

笔者聆听并记录了 3 个特邀报告。来自香港科技大学的 Pascale Fung 教授带来了主题为《From Assistants to Agents in the LLM Era》的演讲。自 20 世纪 90 年代以来，AI 助手已经融入我们的生活。基于

大语言模型（LLMs）构建的 AI 代理改变了我们进行研究、开发和部署的方式。这些代理是具备不同程度推理和规划能力的 AI 系统，能够预见用户需求，自主确定满足这些需求的步骤，访问大语言模型之外的知识和工具，以完成用户的任务。它们标志着大语言模型时代的新一代 AI 助手。在本次演讲中，Pascale Fung 教授概述 AI 助手向 AI 代理演进的过程，当前仍面临的挑战以及未来开发涵盖不同模态和界面的代理家族所带来的机遇。随着未来不同模型架构的出现，LLMs 将继续作为构建代理的有效工具，取代手动设计。

随后，来自悉尼大学的 Judy Kay 教授带来了主题为《Empowering People to Harness and Control their Multimodal Data》的演讲。随着技术的发展，日益丰富的数字设备生态系统渗透我们的生活中，这些设备能够捕获大量的长期个人数据。在本次演讲中，Judy Kay 教授分享一系列案例研究，他指出利用他的研究可以“创建系统和界面，使人们能够掌握和控制这些数据及其使用”，在此基础上介绍了未来计划。首先，第一组案例探讨了如何利用来自可穿戴设备（如智能手表）的数据，为个性化界面提供支持，帮助我们长期了解自身情况，包括对大型数据集（超过 14 万人）的分析，以及用于锻炼的虚拟现实游戏。其次，第二组案例来自正式教育环境，针对来自个人数据界面的数据进行学习，探讨了开放学习者模型（Open Learner Models，简称 OLMs）是如何利用学习数据的。最后，Judy Kay 教授分享一些关键见解，这些见解对于研究议程具有重要意义，包括：终身学习的 OLMs；快速思考与慢速、审慎思考所需的不同界面的本质；传达不确定性；帮助人们从多模态数据中真正了解自己；以及这些如何与人



最佳 demo 奖由延世大学的论文《DanceMimic: Awaken Your Dancing Instinct through a Real-time Dance Imitation Capture System》<sup>[4]</sup>荣获。除此之外，还有 7 篇论文获得荣誉奖。来自中国的研究者获得了其中 6 个。

#### 四、总结

在 ACM MM 2024 录用的全部论文中，第一作者

来自中国大陆的论文占到整体录取论文近 70%左右。这充分说明，我国当前计算机图形学与多媒体领域的研究已经走到了世界舞台的前列，让国际计算机学界听到了更多来自中国的声音。笔者衷心希望肩负使命的研究者在多媒体处理、机器学习、计算机视觉等多个方面做出更多出彩的成绩，对中国乃至世界的科技发展产生更加深远的影响。

责任编辑 崔海楠

#### 参考文献

- [1] Zhang, Z., Li, L., Cong, G., Yin, H., Gao, Y., Yan, C., van den Hengel, A., & Qi, Y. (2024). From Speaker to Dubber. Movie Dubbing with Prosody and Duration Consistency Learning. In MM '24 (pp. 7523-7532). ACM. <https://doi.org/10.1145/3664647.3680777>
- [2] Cai, Z., Ghosh, S., Adatia, A.P., Hayat, M., Dhall, A., Gedeon, T., & Stefanov, K. (2024). AV-Deepfake1M: A Large-Scale LLM-Driven Audio-Visual Deepfake Dataset. In MM '24 (pp. 7414-7423). ACM. <https://doi.org/10.1145/3664647.3680795>
- [3] Zejun Zhang, Xiao Zhu, Anlan Zhang, and Feng Qian. 2024. An In-depth Study of Bandwidth Allocation across Media Sources in Video Conferencing. In MM'24. (pp.7696–7704).ACM. <https://doi.org/10.1145/3664647.3681007>
- [4] Seongjean Kim, Jungwoo Huh, Yeseung Park, Jungsu Kim, and Sanghoon Lee. 2024. DanceMimic: Awaken Your Dancing Instinct through a Real-time Dance Imitation Capture System. In MM'24 .(pp.11267–11269).ACM. <https://doi.org/10.1145/3664647.3684991>



### 吴 垚

厦门大学信息学院，博士研究生。主要研究方向计算机视觉，涵盖 3D 场景理解，多模态学习等方向。

Email: wuyao@stu.xmu.edu.cn



### 曲延云

厦门大学信息学院计算机科学与技术系教授、博导，教工党支部书记（全国样板支部），中国自动化学会混合智能专委会副主任兼秘书长，CCF 高级会员，IEEE Senior Member。获得福建省自然科学二等奖，入选“2023 年度科学影响力排行榜”全球前 2% 顶尖科学家。长期从事计算机视觉、模式识别研究，发表论文 150 余篇，其中中国计算机学会推荐的 CCF-A 类期刊会议论文五十余篇，研究成果发表在计算机视觉顶刊 TPAMI 和 IJCV，图像处理、机器学习领域 JCR 主流期刊 TIP、TNNLS、TYCB、PR、TKDE 等，领域顶级会议 CVPR、ICCV、NeurIPS、AAAI、IJCAI、ECCV、ACM MM。

Email: quyanyun@gmail.com