软件学报 ISSN 1000-9825, CODEN RUXUEW
2026,37(5):1885−1886 [doi: 10.13328/j.cnki.jos.007546] [CSTR: 32375.14.jos.007546]

E-mail: jos@iscas.ac.cn
http://www.jos.org.cn
Tel: +86-10-62562563

# 多媒体智能理解与生成专题前言*

孙立峰[1], 闵巍庆[2], 马占宇[3], 蒋树强[4], 彭宇新[5], 田 丰[6], 黄庆明[4]

[1](清华大学 计算机科学与技术系, 北京 100084)

[2](中国科学院 计算技术研究所 智能信息处理重点实验室, 北京 100190)

[3](北京邮电大学 人工智能学院, 北京 100876)

[4](中国科学院大学 计算机科学与技术学院, 北京 100049)

[5](北京大学 王选计算机研究所, 北京 100080)

[6](中国科学院 软件研究所 人机交互技术与智能信息处理实验室, 北京 100190)

通信作者: 闵巍庆 E-mail: minweiqing@ict.ac.cn

中文引用格式: 孙立峰, 闵巍庆, 马占宇, 蒋树强, 彭宇新, 田丰, 黄庆明. 多媒体智能理解与生成专题前言. 软件学报, 2026, 37(5): 1885–1886. http://www.jos.org.cn/1000-9825/7546.htm

多媒体技术是对文字、图像、图形、音视频等多种媒体信息进行内容处理和交互展示的技术. 随着互联网、移动互联网、社交网络等泛在网络环境下多媒体数据的爆发式增长, 以及以大模型为代表的人工智能技术的迅猛发展, 多媒体技术正迎来新一轮的发展浪潮. 如何高效、精准地理解与生成多媒体内容, 已成为该领域的前沿热点. 认知科学的研究表明, 人脑能够通过视觉、听觉和语言等多种感官信息的关联协同认知外部世界, 因此如何借鉴人脑的跨模态处理特性和认知机理, 是实现多媒体内容智能理解与生成的关键. 大模型等人工智能技术为多媒体领域带来了方法论的变革, 推动了多媒体智能理解与生成技术的快速发展, 与此同时也催生出诸多新挑战. 例如, 如何实现开放环境下可靠的多媒体内容感知与理解? 如何平衡神经系统和符号系统, 实现神经-符号相结合的跨媒体推理? 如何基于多模态大模型实现精细可控的个性化多模态生成与编辑? 如何协同多模态环境感知、知识引导和物理约束建立世界模型, 以增强智能体的多模态自主交互能力等.

本专题公开征文, 共收到投稿 25 篇. 论文均通过了形式审查, 内容涉及多媒体检索、问答、推荐和生成等. 特约编辑先后邀请了 30 多位专家参与审稿工作, 每篇投稿至少邀请 2 位专家进行评审. 稿件经初审、复审、ChinaMM 2025 会议宣读和终审 4 个阶段, 历时 3 个多月, 最终有 8 篇论文入选本专题. 根据主题, 这些论文可以分为 3 组.

(1) 多模态学习

《融合音乐知识结构化表征的高精度符号音乐理解》提出了基于音乐理论的音乐知识与音乐序列的结构化表征方法, 结合卷积神经网络、自注意力机制以及自适应增强的特征融合方法, 实现对序列语境的感知与语义特征的增强, 提高了符号音乐理解的准确率.

《基于音频-语言模型的端到端说话人日志系统》提出了一种音频-语言多模态模型, 通过两阶段训练策略实现语音识别能力与判断说话人归属能力的协同优化, 将音频-语言模型的能力泛化到各种下游任务中, 验证了所提方法的有效性.

《交通场景多模态双阶反馈的三维目标检测方法》提出了一种创新的多传感器融合方法, 利用 RGB 图像深度补全生成伪点云, 与真实点云结合以识别感兴趣区域, 缓解了伪点云精度受限与计算量增大的矛盾, 提升了特征提取效率与检测精度.

(2) 多媒体检索、问答和推荐

《基于 CLIP 引导标签优化的弱监督图像哈希》提出了一种 CLIP 引导标签优化的弱监督哈希方法, 通过微调 CLIP 挖掘图像关联的潜在文本标签, 利用优化后的文本和图像进行跨模态全局语义交互, 并针对用户标签的

* 收稿时间: 2025-09-28; jos 在线出版时间: 2025-09-29

分布不平衡问题, 设计了一种标签平衡损失, 通过动态加权增强模型对难样本的表征学习.

《面向免训练视频问答的双重自适应冗余消除》提出了双重自适应冗余消除框架, 通过时空冗余协同优化机制, 实现免训练范式下视频语义理解精度与答案质量的系统性提升.

《基于多模态异质图表征的专利推荐算法》提出了一种基于多模态异质图网络的专利推荐方法, 利用预训练表征模型和图注意力网络学习企业在不同模态下的偏好表征, 引入适配向量和注意力机制实现节点偏好表征与多模态表征的融合, 最后基于企业和专利的特征相似度计算实现专利推荐.

(3) 多媒体生成

《姿态控制人物生成技术综述》梳理了该领域的主流方法, 归纳了常用生成模型和姿态表示方式, 分析了其可行性与局限, 并总结了常用数据集与评测基准, 还介绍了姿态控制人物生成技术在虚拟试衣、视频生成与编辑等场景的应用.

《视听协同的交互式步态干预训练》构建了多模态提示生成框架, 依据用户实时步态数据生成协同的视听提示; 搭建交互式训练系统, 根据步态表现动态调整提示内容, 形成感知与提示生成的闭环迭代; 最后在早期帕金森病康复辅助应用中验证了可行性.

本专题主要面向多媒体、计算机视觉、人机交互等多领域的研究人员和工程人员, 反映了我国学者在多媒体智能理解与生成领域最新的研究进展. 感谢《软件学报》编委会和中国计算机学会多媒体技术专业委员会对专题工作的指导和帮助, 感谢专题全体评审专家及时、耐心、细致的评审工作, 感谢踊跃投稿的所有作者. 希望本专题能够对人工智能赋能的数据管理、分析与系统相关领域的研究工作有所促进.

## 作者简介

**孙立峰**, 博士, 清华大学计算机系长聘教授, 网络多媒体北京市重点实验室主任, CCF 杰出会员. 长期从事网络多媒体、视频智能处理、多媒体边缘计算等领域的研究工作, 承担了重点基金、重点研发计划、973、863 项目. 研究成果获北京市科学技术一等奖 1 项、中国电子学会自然科学一等奖 1 项、中国电子学会技术发明一等奖 1 项.

**闵巍庆**, 博士, 中国科学院计算技术研究所副研究员, CCF 杰出会员, 主要研究方向为多媒体内容分析和食品计算. 获多媒体领域主流期刊 ACM TOMM 和 IEEE MM 的最佳论文奖. 入选北京市杰青, 获中国图象图形学学会 (CSIG) 青年科学家奖、ACM 中国 SIGMM 新星奖及北京市科技进步二等奖.

**马占宇**, 博士, 北京邮电大学教授, 博士生导师, 发展规划处处长, 国家杰出青年科学基金获得者, 国务院学位委员会学科评议组成员, CCF 杰出会员. 主要研究领域为人工智能、模式识别与机器学习基础理论与方法, 及其在计算机视觉、多媒体信号处理等领域的应用. 先后主持国家自然科学基金委“杰青”、“优青”等项目以及科技部“科技冬奥”重点研发计划课题等; 曾获中国人工智能学会“第七届吴文俊人工智能科学技术奖”一等奖, 中国图象图形学学会技术发明一等奖, 国际会议最佳论文奖等.

**蒋树强**, 博士, 中国科学院大学教授, 中国科学院计算技术研究所客座研究员, 博士生导师, 国家杰出青年科学基金获得者, CCF 杰出会员, 主要研究领域为图像/视频等多媒体内容分析、多模态具身智能, 主持承担科技创新 2030—“新一代人工智能”重大项目、国家自然科学基金等项目 20 余项, 先后获中国计算机学会科学技术奖、中国图象图形学会自然科学二等奖、吴文俊人工智能自然科学一等奖和北京市科技进步二等奖.

**彭宇新**, 博士, 北京大学二级教授, 博雅特聘教授, 国家杰出青年科学基金获得者, 国家万人计划科技创新领军人才, 科技部中青年科技创新领军人才, 863 项目首席专家, 中国人工智能产业创新联盟专家委员会主任, 中国工程院“人工智能 2.0”规划专家委员会专家, CCF 会士. 主要研究方向为跨媒体分析、计算机视觉、机器学习、人工智能. 获 2016 年北京市科学技术奖一等奖和 2020 年中国电子学会科技进步一等奖, 2008 年获北京大学宝钢奖教金优秀奖, 2017 年获北京大学教学优秀奖.

**田丰**, 博士, 中国科学院软件研究所二级研究员, CCF 会士, 《软件学报》编委, 国家“万人计划” 科技创新领军人才, 享受国务院政府特殊津贴专家, 国家科技创新 2030 重大项目首席科学家, 首批国家重点研发计划项目首席科学家. 担任中国计算机学会人机交互专委会主任, ACM SIGCHI 中国分会主席 (2011–2019) 等职. 长期从事人机交互领域中交互状态呈现、复杂笔迹结构理解、沉浸式自然交互等领域的科学研究, 获 2018 年度国家科技进步二等奖、2015 年度北京市科学技术一等奖.

**黄庆明**, 博士, 中国科学院大学讲席教授, 博士生导师, CCF 会士, 国家杰出青年科学基金获得者, 百千万人才工程国家级人选并被授予“有突出贡献中青年专家”荣誉称号, 享受国务院政府特殊津贴. 主要研究方向为多媒体计算、图像与视频分析、模式识别、机器学习、计算机视觉等, 主持承担了新一代人工智能国家科技重大专项、国家自然科学基金重点和重点国际合作项目、863 课题、973 课题、中科院前沿科学重点研究项目等国家和省部级项目的研究工作. 相关研究成果获得吴文俊人工智能自然科学一等奖、中国图象图形学学会自然科学一等奖、教育部科技进步一等奖等多项国家学会和省部级奖励.