软件学报 ISSN 1000-9825, CODEN RUXUEW
Journal of Software,2024,35(5):2099−2100 [doi: 10.13328/j.cnki.jos.007030]

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

多模态协同感知与融合技术专题前言*

孙立峰[1], 宋新航[2], 蒋树强[2], 王莉莉[3], 申恒涛[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](北京航空航天大学 计算机学院, 北京 100191)

[4](电子科技大学 计算机科学与工程学院, 四川 成都 611731)

通信作者: 宋新航, E-mail: xinhang.song@ict.ac.cn

中文引用格式: 孙立峰, 宋新航, 蒋树强, 王莉莉, 申恒涛. 多模态协同感知与融合技术专题前言. 软件学报, 2024, 35(5): 2099–2100. http://www.jos.org.cn/1000-9825/7030.htm

与人类利用视觉、听觉、触觉等多种感官信息来感知世界相似, 计算机智能系统也可通过不同的传感器, 如摄像头、雷达、麦克风、触觉传感器等, 来获取人类和物理世界中的数据与信息. 随着智能终端和多模态传感设备的普及, 可用于感知世界的数据来源、维度和数据量都在快速增长, 单独模态数据所提供的信息已经不能满足智能系统感知与理解世界的需求. 因此智能系统在感知世界时, 需要从更多模态数据的差异化获取、动态适配、互补融合、协同感知等角度开展深入研究, 这也是多媒体领域的一个非常重要和具有挑战性的问题. 本专题强调多模态的协同交互与有机融合, 研究多模态协同感知与融合技术, 重点关注视觉语言多模态交互理解技术、多模态交互生成与重建技术和多模态智能融合与协同学习技术, 旨在促进多模态特征表示、自适应融合、协同学习和交互生成等相关理论与方法的研究进展.

本专题公开征文, 共收到投稿 28 篇. 论文均通过了形式审查, 内容涉及多模态交互、理解、生成、智能融合与协同学习等. 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、ChinaMM2023 会议宣读和终审 4 个阶段, 历时 5 个月, 最终有 9 篇论文入选本专题. 根据主题, 这些论文可以分为 3 组.

(1) 视觉语言多模态交互理解技术

《视觉语言模型引导的文本知识嵌入的小样本增量学习》提出了一种基于文本知识嵌入的小样本增量学习方法, 在视觉特征中嵌入具有抗遗忘能力的文本特征, 通过视觉语言多模态特征融合, 实现小样本增量学习中新旧类别数据的有效学习.

《面向跨模态检索的查询感知双重对比学习网络》提出了一种查询感知的跨模态语义融合策略, 根据感知到的查询语义自适应地融合视频的视觉模态特征和字幕模态特征等多模态特征, 获得视频的查询感知多模态联合表示, 并利用双重对比学习机制, 以增强不同模态的语义对齐效果, 从而提高不同模态数据表示的可分辨性和语义一致性.

《面向遥感视觉问答的尺度引导融合推理网络》提出了一种遥感视觉问答方法, 通过对多尺度空间关系的建模、推理与融合, 构建了一个知识表征增强的遥感图像视觉问答模型, 在利用交叉注意力机制的基础上, 通过自监督范式、对比学习方法、图文匹配机制等方法训练目标来自适应地对齐、融合多模态特征, 并辅助预测最终答案.

(2) 多模态交互生成与重建技术

《基于条件语义增强的文本到图像生成》提出了一种基于条件语义增强的文本到图像生成方法, 将条件语义增强用于生成模型的文本嵌入表示, 通过融合图像空间特征和文本语义, 提高了生成图片的细节表达.

《分层特征编解码驱动的视觉引导立体声生成方法》提出了一种基于分层特征编解码的视觉引导的立体声生成方法, 面向听、视觉模态异构问题, 构建了一种由深到浅不同深度特征层间跳跃连接的解码器结构, 以实现对

* 收稿时间: 2023-09-07; jos 在线出版时间: 2023-09-11

视听觉模态信息的浅层细节特征与深度特征的充分利用.

《结合面部动作单元感知的三维人脸重建算法》提出了一种基于面部动作单元感知的三维人脸重建算法, 以面部动作单元和人脸关键点作为桥梁, 构建了从 2D 图像到 3D 人脸关键点的重建模型, 并发布了 300W-LP-AU 数据集.

(3) 多模态智能融合与协同学习技术

《多模态特征分析的帕金森病辅助诊断方法》提出了一种基于多模态特征分析的帕金森病辅助诊断方法, 结合步态和眼动等多模态信息, 分析了不同特征组合方式评估帕金森病的显著性, 验证了虚拟现实场景下高沉浸诱发型任务范式和多模态帕金森病辅助诊断系统的有效性.

《基于多模态关系建模的三维形状识别方法》提出了一个基于多模态关系的三维形状识别网络, 利用点云和多视图局部特征之间的关系学习 3D 特征表示, 并采用自注意力机制的门控模块来探索特征内部的关联信息, 将聚合得到的全局特征进行加权以抑制冗余信息.

《事件融合与空间注意力和时间记忆力的视频去雨网络》提出了一种事件数据与注意力机制融合的视频去雨技术, 结合事件数据与常规视频信息的互补性, 借助事件信息的高动态范围、高时间分辨率等优势, 利用三维对齐将稀疏事件流转化为图像等同维度, 用于融合去雨.

本专题主要面向多媒体、计算机视觉、音频处理、自然语言处理等多领域的研究人员和工程人员, 反映了我国学者在多模态内容分析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. 希望本专题能够对多模态协同感知与融合相关领域的研究工作有所促进.

孙立峰(1972－), 男, 博士, 清华大学计算机科学与技术系教授, 网络多媒体北京市重点实验室主任. 主要从事网络多媒体、视频高效智能处理、多媒体边缘智能等领域的研究工作, CCF 杰出会员. 曾获中国电子学会自然科学一等奖、中国电子学会技术发明一等奖、北京市科学技术一等奖、IEEE Transactions on CSVT 年度最佳论文奖、ACM Multimedia 最佳论文奖.

宋新航(1988－), 男, 博士, 中国科学院计算技术研究所副研究员, CCF 专业会员. 主要研究方向包括场景感知与机器人导航等. 曾在 IEEE TPAMI, TIP, TMM, CVPR, ICCV, NeurIPs 等高水平国际期刊和会议上发表论文 30 余篇. 曾获 ACM Multimedia 图题概括生成竞赛、CVPR 具身智能竞赛视觉导航赛道冠军, 曾获北京市科技进步二等奖、中国图象图形学学会自然科学二等奖.

蒋树强(1977－), 男, 博士, 中国科学院计算技术研究所研究员, 博士生导师, 中国科学院特聘研究员, 中国科学院大学岗位教授, 国家杰出青年科学基金获得者, CCF 杰出会员, 任国际期刊 ACM ToMM 编委、CCF 多媒体专委会秘书长、ACM SIGMM 中国分会副主席、IEEE CASS 北京分会副主席. 主要研究方向是多媒体内容分析与多模态智能技术, 先后获中国计算机学会科学技术奖、中国科学院青年科学家国际合作奖、中国图象图形学学会自然科学二等奖、吴文俊人工智能自然科学一等奖、北京市科技进步二等奖.

王莉莉(1977－), 女, 博士, 北京航空航天大学计算机学院教授, 博士生导师, CCF 杰出会员, 任虚拟现实技术与系统全国重点实验室副主任. 获得国家科技进步一等奖, 国家技术发明二等奖, 中国电子学会科技进步一等奖. 在领域顶级国际期刊和会议 IEEE TVCG、IEEE VR 等发表学术论文 80 余篇, 担任领域顶级国际会议 IEEE VR 2021–2023 程序委员会主席、IEEE ISMAR 2021–2022 程序委员会主席. 担任 IEEE TVCG 编委、中国科学信息科学中英文版青年编委, 图形学领域著名国际杂志 IEEE CG&A 编委.

申恒涛(1977－), 男, 博士, 电子科技大学计算机科学与工程学院院长, 欧洲科学院外籍院士, CCF 专业会员, ACM Fellow, IEEE Fellow, OSA Fellow, 发表了 360 余篇高水平同行评审论文, 包括 150 多篇 IEEE/ACM Transactions 和 250 多篇 CCF A 类论文, 获得了 8 个国际会议和期刊的最佳论文奖. 主要研究领域为多媒体搜索、计算机视觉、人工智能.