



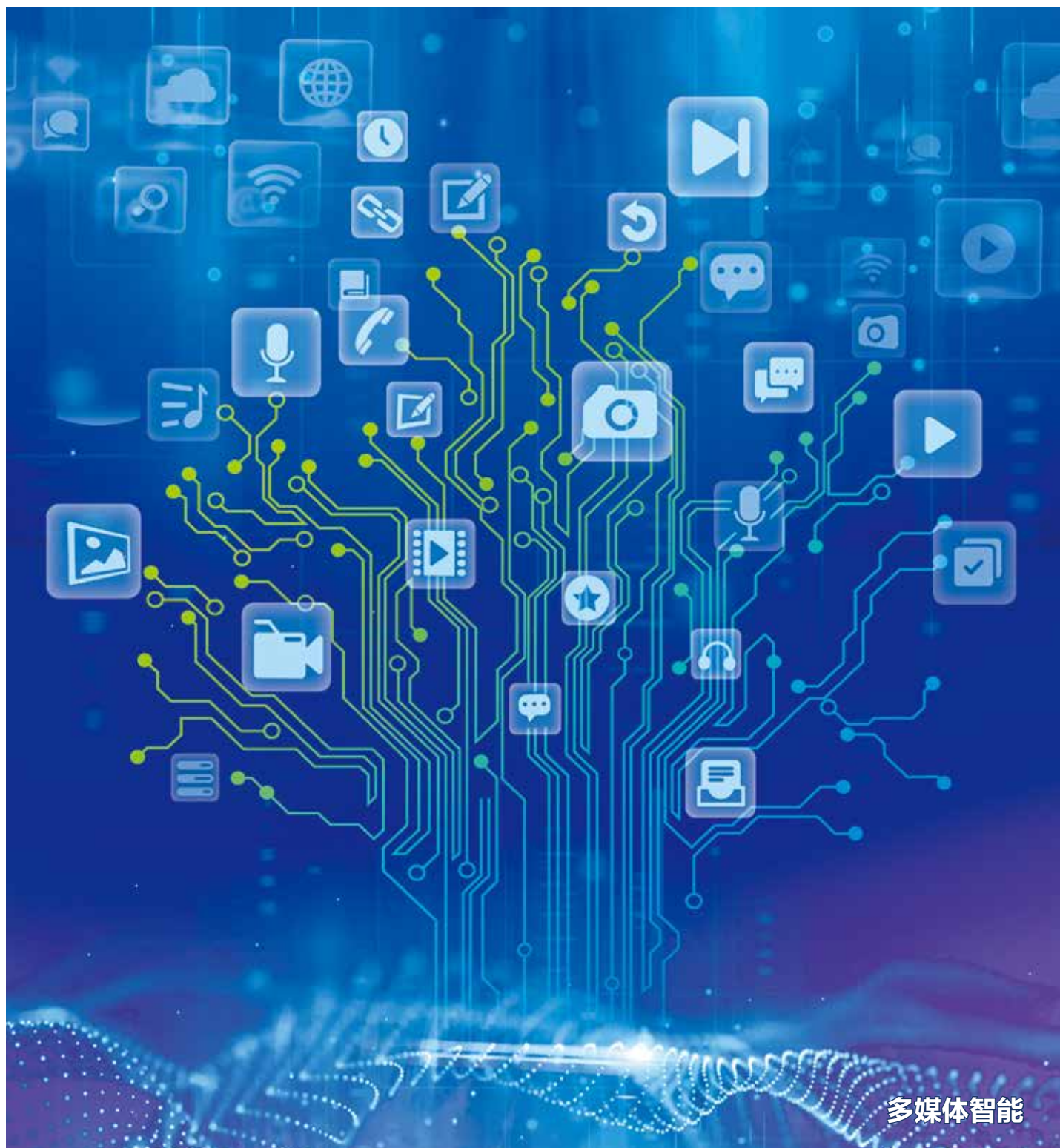
JOURNAL OF IMAGE AND GRAPHICS

主办: 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

中国图象图形学报

2022
09
VOL.27

ISSN1006-8961
CN11-3758/TB



多媒体智能



第27卷第9期（总第317期）
2022年9月16日

中国精品科技期刊
中国国际影响力优秀学术期刊
中国科技核心期刊
中文核心期刊

版权声明

凡向《中国图象图形学报》投稿，均视为同意在本刊网站及CNKI等全文数据库出版，所刊载论文已获得著作权人的授权。本刊所有图片均为非商业目的使用，所有内容，未经许可，不得转载或以其他方式使用。

Copyright

All rights reserved by Journal of Image and Graphics, Institute of Remote Sensing and Digital Earth, CAS. The content (including but not limited text, photo, etc) published in this journal is for non-commercial use.

主管单位 中国科学院
主办单位 中国科学院空天信息创新研究院
中国图象图形学学会
北京应用物理与计算数学研究所

主 编 吴一戎
编辑出版 《中国图象图形学报》编辑出版委员会
通信地址 北京市海淀区北四环西路19号
邮 编 100190
电子信箱 jig@aircas.ac.cn
电 话 010-58887035
网 址 www.cjig.cn

广告发布登记号 京朝工商广登字20170218号
总 发 行 北京报刊发行局
订 购 全国各地邮局
海外发行 中国国际图书贸易集团有限公司
(邮政信箱: 北京399信箱 邮编: 100048)
印刷装订 北京科信印刷有限公司

Journal of Image and Graphics

Title inscription: Song Jian | Monthly, Started in 1996

Superintended by Chinese Academy of Sciences
Sponsored by Aerospace Information Research Institute, CAS
China Society of Image and Graphics
Institute of Applied Physics and Computational Mathematics

Editor-in-Chief Wu Yirong
Editor, Publisher Editorial and Publishing Board of Journal of Image and Graphics
Address No. 19, North 4th Ring Road West, Haidian District, Beijing, P. R. China
Zip code 100190
E-mail jig@aircas.ac.cn
Telephone 010-58887035
Website www.cjig.cn

Distributed by Beijing Bureau for Distribution of Newspapers and Journals
Domestic All Local Post Offices in China
Overseas China International Book Trading Corporation
(P.O.Box 399, Beijing 100048, P.R.China)
Printed by Beijing Kexin Printing Co., Ltd.

CN 11-3758/TB
ISSN 1006-8961
CODEN ZTTXFZ

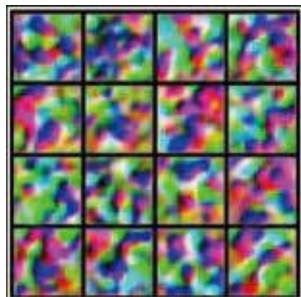
国外发行代号 M1406
国内邮发代号 82-831
国内定价 60.00元



多媒体智能：当多媒体遇到人工智能(第2551页)



面向海洋的多模态智能计算：挑战、进展和展望(第2589页)



基于真实数据感知的模型功能窃取攻击(第2721页)

《中国图象图形学报》多媒体智能专刊简介

朱文武, 黄庆明, 黄华, 蒋树强, 彭宇新, 刘青山, 王井东, 纪荣嵘, 邓伟洪, 方玉明, 刘家瑛, 韩向娣 2549

学者观点

多媒体智能：当多媒体遇到人工智能

朱文武, 王鑫, 田永鸿, 高文 2551

视觉知识：跨媒体智能进化的新支点

杨易, 庄越挺, 潘云鹤 2574

面向海洋的多模态智能计算：挑战、进展和展望

聂婕, 左子杰, 黄磊, 王志刚, 孙正雅, 仲国强, 王鑫, 王玉成, 刘安安, 张弘, 董军宇, 魏志强 2589

综述

基于深度学习的人—物交互关系检测综述

廖越, 李智敏, 刘侃 2611

人类面部重演方法综述

刘锦, 陈鹏, 王茜, 付晓蒙, 戴娇, 韩冀中 2629

视觉语言多模态预训练综述

张浩宇, 王天保, 李孟择, 赵洲, 浦世亮, 吴飞 2652

Bayer阵列图像去马赛克算法综述

魏凌云, 孙帮勇 2683

多媒体智能安全

多特征决策融合的音频copy-move篡改检测与定位

张国富, 肖锐, 苏兆品, 廉晨思, 岳峰 2697

多级特征全局一致性的伪造人脸检测

杨少聪, 王健, 孙运莲, 唐金辉 2708

基于真实数据感知的模型功能窃取攻击

李延铭, 李长升, 余佳奇, 袁野, 王国仁 2721

目标智能检测

利用时空特征编码的单目标跟踪网络

王蒙蒙, 杨小倩, 刘勇 2733

结合时空一致性的FairMOT跟踪算法优化

彭嘉淇, 王涛, 陈柯安, 林巍峭 2749

多媒体分析与理解

融合知识表征的多模态Transformer场景文本视觉问答方法

余宙, 俞俊, 朱俊杰, 匡振中 2761

结合多层级解码器和动态融合机制的图像描述

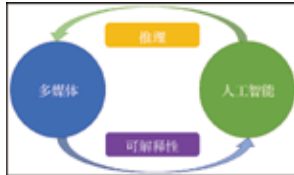
姜文晖, 占锟, 程一波, 夏雪, 方玉明 2775

面向非受控场景的人脸图像正面化重建

辛经纬, 魏子凯, 王楠楠, 李洁, 高新波 2788

CONTENTS

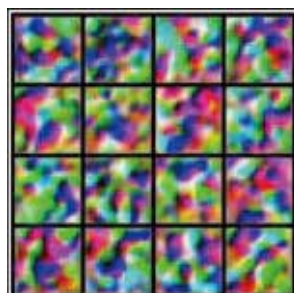
JOURNAL OF IMAGE AND GRAPHICS



Multimedia intelligence: the convergence of multimedia and artificial intelligence(P2551)



Marine oriented multimodal intelligent computing: challenges, progress and prospects(P2589)



Model functionality stealing attacks based on real data awareness(P2721)

Scholar View

Multimedia intelligence: the convergence of multimedia and artificial intelligence

Zhu Wenwu, Wang Xin, Tian Yonghong, Gao Wen 2551

The review of visual knowledge: a new pivot for cross-media intelligence evolution

Yang Yi, Zhuang Yueting, Pan Yunhe 2574

Marine oriented multimodal intelligent computing: challenges, progress and prospects

Nie Jie, Zuo Zijie, Huang Lei, Wang Zhigang, Sun Zhengya, Zhong Guoqiang, Wang Xin, Wang Yucheng, Liu An'an, Zhang Hong, Dong Junyu, Wei Zhiqiang 2589

Review

A review of deep learning based human-object interaction detection

Liao Yue, Li Zhimin, Liu Si 2611

Critical review of human face reenactment methods

Liu Jin, Chen Peng, Wang Xi, Fu Xiaomeng, Dai Jiao, Han Jizhong 2629

Comprehensive review of visual-language-oriented multimodal pre-training methods

Zhang Haoyu, Wang Tianbao, Li Mengze, Zhao Zhou, Pu Shiliang, Wu Fei 2652

The review of demosaicing methods for Bayer color filter array image

Wei Lingyun, Sun Bangyong 2683

Multimedia Intelligent Security

Multi-feature decision fused detection and localization method for copy-move forgery of digital audio clips

Zhang Guofu, Xiao Rui, Su Zhaopin, Lian Chensi, Yue Feng 2697

Multi-level features global consistency for human facial deepfake detection

Yang Shaocong, Wang Jian, Sun Yunlian, Tang Jinhui 2708

Model functionality stealing attacks based on real data awareness

Li Yanming, Li Changsheng, Yu Jiaqi, Yuan Ye, Wang Guoren 2721

Object Intelligent Detection

A spatio-temporal encoded network for single object tracking

Wang Mengmeng, Yang Xiaoqian, Liu Yong 2733

Spatio-temporal consistency based FairMOT tracking algorithm optimization

Peng Jiaqi, Wang Tao, Chen Kean, Lin Weiyao 2749

Multimedia Analysis and Understanding

Knowledge-representation-enhanced multimodal Transformer for scene text visual question answering

Yu Zhou, Yu Jun, Zhu Junjie, Kuang Zhenzhong 2761

The integrated mechanism of hierarchical decoders and dynamic fusion for image captioning

Jiang Wenhui, Zhan Kun, Cheng Yibo, Xia Xue, Fang Yuming 2775

Face frontalization for uncontrolled scenes

Xin Jingwei, Wei Zikai, Wang Nannan, Li Jie, Gao Xinbo 2788

中图法分类号: TP309; TP399 文献标识码: A 文章编号: 1006-8961(2022)09-2549-02

论文引用格式: 朱文武, 黄庆明, 黄华, 蒋树强, 彭宇新, 刘青山, 王井东, 纪荣嵘, 邓伟洪, 方玉明, 刘家瑛, 韩向娣. 2022. 《中国图象图形学报》多媒体智能专刊简介. 中国图象图形学报, 27(09): 2549-2550 [DOI: 10.11834/jig.2200009]

《中国图象图形学报》多媒体智能专刊简介

朱文武¹, 黄庆明², 黄华³, 蒋树强⁴, 彭宇新⁵,
刘青山⁶, 王井东⁷, 纪荣嵘⁸, 邓伟洪⁹, 方玉明¹⁰, 刘家瑛⁵, 韩向娣¹¹

1. 清华大学, 北京 100084; 2. 中国科学院大学, 北京 100049; 3. 北京师范大学, 北京 100091;
4. 中国科学院计算技术研究所, 北京 100094; 5. 北京大学, 北京 100091; 6. 南京信息工程大学, 南京 210044;
7. 微软亚洲研究院, 北京 100190; 8. 厦门大学, 厦门 361005; 9. 北京邮电大学, 北京 100876;
10. 江西财经大学, 南昌 330013; 11. 《中国图象图形学报》编辑部, 北京 100190

近年来, 人工智能和高性能计算快速发展, 大规模多媒体数据的智能应用需求日趋广泛, 多媒体数据的融合、转换、理解、搜索、推理与推荐等方向的新问题也不断涌现, 推动了多媒体智能处理与分析技术的迅速发展, 在学术界和产业界均引起极大关注。

为更好地推动多媒体智能处理与分析理论、技术、应用的发展, 及时记录我国多媒体智能领域的最新技术、前沿进展、学者观点、算法研究、数据集构建及应用落地等方面的优秀成果, 《中国图象图形学报》邀请国内多位一线专家共同策划推出“多媒体智能”专刊, 以期为图像视频处理与分析、多媒体信息检索和推荐、多媒体内容分发、数据挖掘与机器学习等相关领域的研究人员提供参考。

经过严格评审, “多媒体智能”专刊共收录学术论文 15 篇, 包括“学者观点”3 篇、“综述”4 篇、“多媒体智能安全”3 篇、“目标智能检测”2 篇、以及“多媒体分析与理解”3 篇。

“学者观点”栏目中, 《多媒体智能: 当多媒体遇到人工智能》围绕“大数据”时代多媒体与人工智能融合的背景, 提出了多媒体智能的新概念, 探讨了多媒体和人工智能之间的相互影响, 具体包括以下两个方向: 1) 多媒体推动人工智能向着更具可解释性的方向发展; 2) 人工智能促进多媒体推理能力的发展。这两个方向形成了一个多媒体智能循环, 其中多媒体和 AI 以交互和迭代的方式相互促进增强。该文讨论了每一循环中的研究进展, 特别是研究多媒体如何推动机器学习发展以及机器学习如何反过来推动多媒体发展。总结了循环中已经完成的工

作, 并指出了完成循环所需要做的未来工作, 然后对值得进一步深入探索的多媒体智能相关研究方向进行了思考。

《视觉知识: 跨媒体智能进化的新支点》介绍了一种新的智能表达方式——视觉知识的 3 个基本要素, 即视觉概念、视觉关系、视觉推理, 并对每个要素展开详细讨论与分析。视觉知识有助于实现数据与知识驱动的统一框架, 学习可归因可溯源的结构化表达, 推动跨媒体知识关联与智能推理。视觉知识具有强大的知识抽象表达能力和多重知识互补能力, 为跨媒体智能进化提供了新的有力支点。

《面向海洋的多模态智能计算: 挑战、进展和展望》首次从多模态数据技术的视角, 系统地介绍面向海洋现象/过程的智能感知、认知和预知的交叉研究进展。通过梳理海洋科学大数据全生命周期的阶段演进过程, 明确海洋多模态智能计算的研究对象、科学问题和典型应用场景。在海洋多模态大数据内容分析、推理预测和高性能计算 3 个典型应用场景中展开现有工作的系统性梳理和介绍。针对海洋数据分布和计算模式的差异性, 提出海洋多模态大数据表征建模、跨模态关联、推理预测、高性能计算 4 个关键科学问题中的挑战, 并提出未来展望。

“综述”栏目中, 《基于深度学习的人—物交互关系检测综述》一方面分析空域人—物交互关系检测任务, 从数据内容场景、标注粒度两方面总结和分析当下数据库和基准。然后从两阶段分段式方法和单阶段端到端式方法两个流派出发系统性地阐述当前检测方法的发展现状, 分析两个流派方法的特性

和优劣,厘清该领域方法的发展路线。另一方面,对时空域人一物交互关系检测任务进行总结,分析现有时空域交互关系数据集构造与特性和现有基线算法的优劣。最后对未来的研究方向进行展望。

《人类面部重演方法综述》对现阶段面部重演领域的发展进行梳理和总结。从面部重演模型入手,对面部重演存在的问题、模型的分类以及驱动人脸特征表达进行阐述,列举并介绍了训练面部重演模型常用的数据集及评估模型的评价指标,对面部重演近年研究工作进行归纳、分析与比较,最后对面部重演的演化趋势、当前挑战、未来发展方向、危害及应对策略进行了总结和展望。

《视觉语言多模态预训练综述》总结了视觉语言多模态预训练领域的进展,对常见的预训练数据集和预训练方法进行了汇总,然后对目前最新方法和经典方法进行系统概述,按输入来源分为图像—文本预训练模型和视频—文本多模态模型两大类,阐述了各方法之间的共性和差异,并将各模型在具体下游任务上的实验情况进行了汇总。最后总结了视觉语言预训练面临的挑战和未来发展趋势。

《Bayer 阵列图像去马赛克算法综述》对 Bayer 采样阵列原理和图像去马赛克技术进行概述;将现有方法按照传统方法和基于深度学习方法两类进行总结,根据去马赛克任务是否具有独立性,将学习方法进一步细分为独立去马赛克任务和联合去马赛克任务两类,进一步分析了不同方法的原理和优缺点,并重点阐述了基于深度学习的去马赛克方法网络结构和重建机理,介绍去马赛克领域中常用的公共数据集和性能评价指标,并对相关实验进行分析对比。最后,围绕网络深度、运算效率、实用性等方面分析了现阶段图像去马赛克技术面临的挑战及未来的发展方向。

我们期待广大读者和科技人员通过“多媒体智能”专刊,能够更深入、更全面地了解该领域的最新方法和应用,吸引更多学者从事相关研究并产生具有国际影响力的优秀成果,为本领域的发展做出新的贡献。

专刊编委会:

朱文武,清华大学计算机系教授,主要研究领域为多媒体网络、多媒体大数据智能。

黄庆明,中国科学院大学计算机科学与技术学院教授,主要研究领域为多媒体计算、图像处理、模式识别、机器学习、计算机视觉。

黄华,北京师范大学人工智能学院教授,主要研究领域为可视媒体计算。

蒋树强,中国科学院计算技术研究所研究员,主要研究领域为多媒体技术、图像处理、视频分析。

彭宇新,北京大学教授、博雅特聘教授,主要研究领域为跨媒体分析与推理、图像视频识别与理解、计算机视觉、人工智能。

刘青山,南京信息工程大学教授,主要研究领域为图像理解、视频分析、模式识别、机器学习。

王井东,微软亚洲研究院首席研究员,主要研究领域为计算机视觉、多媒体以及机器学习。

纪荣嵘,厦门大学教授,主要研究领域为计算机视觉、多媒体技术和机器学习。

邓伟洪,北京邮电大学教授,主要研究领域为计算机视觉与模式识别、情感计算、多模态学习。

方玉明,江西财经大学教授,主要研究领域为视频图像处理、计算机视觉。

刘家瑛,北京大学副教授,主要研究领域为图像/视频编码优化、增强重建、分析与理解。

专刊责编:

韩向娣,副编审,主要研究领域为学术出版和媒体传播。E-mail:hanxd@radi.ac.cn