

中图法分类号: TP309; P71 文献标识码: A 文章编号: 1006-8961(2026)08-2687-02

论文引用格式: 董军宇, 杨易, 翟广涛, 丛润民, 蹇木伟, 韩向娣. 2026. 《中国图象图形学报》具身智能与人形机器人前沿专栏简介. 中国图象图形学报, 31(8):2687-2688[DOI:10.11834/jig.2600008]

《中国图象图形学报》 具身智能与人形机器人前沿专栏简介

董军宇¹, 杨易², 翟广涛³, 丛润民⁴, 蹇木伟⁵, 韩向娣^{6*}

1. 中国海洋大学, 青岛 266000; 2. 浙江大学, 杭州 310058; 3. 上海交通大学, 上海 200240; 4. 山东大学, 济南 250061;
5. 山东财经大学, 济南 250014; 6. 《中国图象图形学报》编辑部, 北京 100190

随着人工智能和机器人技术的迅猛发展, 具身智能逐渐成为推动机器人技术跨越式发展的新引擎。具身智能不仅强调机器人在感知、认知、运动控制等多个层面的协同工作, 更注重将环境互动与自身运动、操作能力有机结合, 形成一套高效、灵活的自主学习和决策系统。构建基于具身智能的机器人系统, 有望打破传统固定算法和预设流程的限制, 实现机器人在面对未知或动态环境时的自主适应和实时调节。

深入研究具身智能与机器人前沿技术及应用, 不仅能推动多传感器融合、深度学习、强化学习等前沿技术在实际机器人系统中的整合应用, 还能通过理论创新促进机器人在复杂决策、任务协调和人机交互等多个方面的性能提升。此外, 应积极探索如何通过具身智能实现机器人自我学习和自我进化, 为机器人赋予更高层次的适应性和创造性, 以更好地满足智慧制造、医疗辅助、灾害救援、智能家居等多领域应用的迫切需求。

《中国图象图形学报》邀请业内专家共同策划推出“具身智能与人形机器人前沿”专栏, 本专栏的研究不仅具有重要的理论意义, 如推动智能机器人理论体系的完善和新型算法的突破, 更具备显著的实践意义和产业价值。通过构建开放、共享的实验平台和标准化的数据处理方法, 将有效促进科研机构与企业之间的合作, 加速成果转化, 推动机器人产业的升级和跨界融合。

专栏收到领域内相关学者积极投稿。经过严格评审, 共收录学术论文4篇, 其中综述3篇、算法论

文1篇。

综述《具身智能驱动的分心驾驶检测: 框架研究与前沿展望》(作者: 蹇木伟*, 凌钰坤, 张昊然, 张琳松, 马嘉骏)是2026年第8期的封面论文。该文从具身认知视角系统梳理了分心驾驶检测的关键技术路径, 将分心驾驶行为视为“驾驶员—车辆—环境”持续交互过程中产生的具身认知结果, 重点涵盖了多模态智能驾驶垂域大模型、领域自适应与持续学习等前沿研究方向, 并从具身语义建模、环境适应与经验累积的角度分析其在分心驾驶检测中的作用; 结合复杂交通环境的动态特征, 提出了一个具有层次化结构的具身智能驱动分心驾驶检测理论框架, 为分心驾驶检测研究提供新的参考思路和解决方案。最后, 论文总结了具身智能驱动的分心驾驶检测系统在落地部署过程中面临的技术瓶颈与挑战, 阐述了具身智能在分心驾驶检测中的应用前景, 并展望了未来潜在的发展方向, 可为具身智能驱动的智能驾驶安全系统的持续演进提供参考。

综述《具身智能模型评测: 从感知到执行》(作者: 李春一, 张建博, 肖嘉豪, 闫博闻, 郭晟毓, 叶桐瑞, 林维斯, 翟广涛*)围绕“从感知到执行”的完整链路, 对2020~2025年间百余篇文献进行系统梳理, 首次提出“静态数据集—仿真平台—真实机器人”三级金字塔式评测范式, 并从感知、认知、决策、执行4个环节拆解出20余项核心能力维度与量化指标, 分别从方法学假设、适用边界、固有局限、成本—可信度曲线4个角度进行横向对比。文章进一步汇总了46套主流基准数据集在数据规模、任务类型、评估

*通信作者

指标、开源程度、安全伦理考量等维度的差异。基于此,提出“能力导向、协议统一、三级协同”的未来框架:1)在任务层,由“功能对标”转向“能力对标”,建立可分解、可溯源、可加权的多维误差体系;2)在协议层,制定统一的场景描述、接口规范与指标定义,实现跨平台、跨任务、跨模型的可比性;3)在系统层,构建可远程接入、7×24小时运行的共享机器人集群,形成“仿真预训练—真机微调—在线更新”的闭环追踪,降低重复建设成本,打通从实验室到场景落地的最后一公里。论文最后讨论了安全伦理、文化偏见、能效评估等新维度如何纳入量化框架,并给出标准化路线图的短、中、长期目标。本文的工作为具身智能从“技术涌现”走向“科学共识”提供了可操作的评测基础设施与参考范式,具体的静态—仿真—真机榜单可在:<https://opencompass.org.cn/embodied-intelligence>中访问。

综述《面向具身智能的视觉—语言—动作模型研究进展》(作者:刘锐,王文冠,王俊,杨易)系统梳理了基于VLA的具身智能研究进展,从发展历程、模型架构、系统分类、训练与评估等方面展开综述。首先,文章追溯了视觉与语言基础模型的演进脉络,并阐述VLA概念的提出背景;随后,本文深入剖析VLA的关键技术模块,包括视觉编码、语言表征及动作词元化与解码机制;在此基础上,本文引入系统架构分类法,将现有工作归纳为单系统、双系统与层次化3类,并分析其设计权衡与适用场景;此外,论文总结了模型的预训练与后训练策略,并梳理了仿真及真实环境下的主流评测基准;最后,分析了VLA在实时推理效率、数据质量、环境泛化性与安全伦理等维度面临的挑战,并展望从被动感知到主动推理、持续学习、场景泛化与可靠部署等未来方向。旨在为相关研究者提供系统的技术参考,推动VLA在开放世界具身任务中的理论发展和应用落地。提及的算法、数据集和评估指标已汇总至<https://github.com/DefaultRui/vision-language-action-models-for-embodied-AI>。

算法论文《语义对齐与局部感知驱动的开放词汇语义分割》(作者:王馨静,高颖*,邹亚琦,朱政宇,徐春雪,赵琦)构建TG-CLIP框架,以冻结的CLIP为特征源,设计两类推理期算子以增强局部与跨尺度语义:1)文本引导的重校准机制,将文本查询作为条件信号对patch级表征进行语义性重构并回投;2)多视角一致性推理,通过重采样与镜像视角的结果融

合来放大跨视角一致的预测信号。所有操作均在前向路径上完成,无需额外标注或网络微调。在8个公开基准上的评测显示,TG-CLIP的平均mIoU达45.5%,超越多个现有方法,比排名第2的ProxyCLIP(40.1%)高出4.4%。定性对比实验进一步显示,TG-CLIP在复杂背景与细小结构处能更好地保留目标细节并减少误分割与漏分割,与定量结果相互印证。消融与超参数研究表明,两类算子可累积提升性能:文本引导重校准对温度参数表现出良好鲁棒性,多视角一致性推理在尺度与翻转策略上存在可量化的精度—效率权衡,推荐的部署配置在实用性上取得较好折中。TG-CLIP在保持视觉—语言模型零样本能力的前提下,通过两类轻量的推理算子实现了对像素级语义一致性与尺度、视角鲁棒性的显著改进,为免训练的开放词汇分割提供了一种易于部署且效果稳健的工程化方案。代码链接:<https://www.scidb.cn/s/7baeYf>。

我们期待广大读者和科技人员通过“具身智能与人形机器人前沿”专栏,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。

专栏编委会

董军宇,中国海洋大学教授,主要研究方向为计算机视觉、水下智能感知、海洋智能感知和声呐等。

杨易,浙江大学教授,主要研究方向为人工智能和跨媒体研究、物理AI具身智能等。

翟广涛,上海交通大学教授,主要研究方向为图像处理、感知建模、大模型与具身智能评测。

丛润民,山东大学教授,主要研究方向为人工智能、模式识别、机器学习和具身感知等。

蹇木伟,山东财经大学教授,主要研究方向为计算机视觉等。

专栏责编

韩向娣,编审,主要研究方向为学术出版和媒体传播等。E-mail:hanxd201310@aircas.ac.cn。