

中图法分类号: TP391 文献标识码: A 文章编号: 1006-8961(XXXX)XX-0001-16

论文引用格式: Jian Muwei, Ling Yukun, Zhang Haoran, Zhang Linsong, Ma Jiajun. XXXX. Embodied Intelligence-Driven Distracted Driving Detection: A Framework and Research Prospects. Journal of Image and Graphics, XX(XX):0001-0016(蹇木伟, 凌钰坤, 张昊然, 张琳松, 马嘉骏. XXXX. 具身智能驱动的分心驾驶检测: 框架研究与前沿展望. 中国图象图形学报, XX(XX):0001-0016)[DOI:10.11834/jig.250514]

具身智能驱动的分心驾驶检测: 框架研究与前沿展望

蹇木伟^{1,2*}, 凌钰坤¹, 张昊然¹, 张琳松¹, 马嘉骏¹

1. 山东财经大学计算机科学与人工智能学院, 济南 250014; 2. 空天信息大学计算机科学与人工智能学院, 济南 250299

摘要: 分心驾驶是导致交通事故的主要因素之一, 每年造成的巨额经济损失和人员伤亡对公共交通安全构成严重威胁。与以静态感知和离线分类为核心的传统驾驶行为识别方法不同, 具身智能强调智能体通过身体感知、环境交互与行为反馈形成闭环认知, 为分心驾驶检测走向“行为—认知—反馈”协同建模提供了新的研究范式。在分心驾驶检测场景中, 驾驶员并非被动的信息输入源, 而是与车辆、道路环境持续交互的具身主体, 其风险状态由视觉、动作、注意力与情境共同驱动。基于具身智能的核心思想, 本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, 本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。

关键词: 分心驾驶检测; 深度学习; 具身智能; 具身感知; 具身执行; 具身学习与进化; 智慧交通

Embodied Intelligence-Driven Distracted Driving Detection: A Framework and Research Prospects

Jian Muwei^{1,2*}, Ling Yukun¹, Zhang Haoran¹, Zhang Linsong¹, Ma Jiajun¹

1. School of Computing and Artificial Intelligence, Shandong University of Finance and Economics, Jinan, 250014, China; 2. School of Computing and Artificial Intelligence, Aerospace Information Technology University, Jinan, 250299, China

Abstract: According to the National Highway Traffic Safety Administration, distracted driving has become one of the leading causes of motor vehicle accidents, resulting in severe economic losses and casualties worldwide. Distracted driving refers to behaviors in which a driver's attention shifts from the primary driving task to unrelated activities, impairing visual perception, cognitive awareness, decision-making, and operational control. Common distractions such as calling, texting, eating, and conversing are often covert, dynamic, and difficult to detect using a single sensing modality. Therefore, developing a framework capable of robust perception, adaptive cognition, and real-time reasoning remains a major challenge in intelligent driving research. Embodied intelligence, as a core concept of intelligent manufacturing, emphasizes deep interaction and tight coordination between intelligent agents and the physical environment. Through continuous perception—

收稿日期: 2025-10-19; 修回日期: 2026-01-20

* 通信作者: 蹇木伟 jianmuweihk@163.com

基金项目: 山东省泰山学者特聘专家项目 (项目编号: tstp20250536)

Supported by: Taishan Scholars Program of Shandong Province (tstp20250536)

© 中国图象图形学报版权所有

action coupling, embodied intelligence enables the emergence and evolution of cognition, endowing intelligent systems with adaptive behavior generation, self-optimization, and environmental awareness. In recent years, embodied intelligence has been widely applied in bionic robotics, unmanned systems, and intelligent vehicles, driving innovation and theoretical progress in human-machine collaboration and cognitive autonomy. Building upon this theoretical foundation, this study systematically reviews the key technological pathways for distracted driving detection from the perspective of embodied cognition, focusing on frontier research trends such as large-scale multimodal models, domain adaptation, and continual learning. By integrating the cognitive logic of embodied intelligence with the dynamic characteristics of complex traffic environments, this work introduces a hierarchical theoretical framework consisting of five interrelated layers that together form a closed-loop structure of perception, cognition, and execution. The Physical Layer serves as the foundation, responsible for multimodal data acquisition through cameras, infrared sensors, microphones, and physiological sensors, ensuring comprehensive coverage of driver behaviors and environmental conditions. The Communication and Fusion Layer handles data transmission, synchronization, and multimodal fusion across heterogeneous devices, enabling reliable multi-source collaboration and adaptive information exchange under varying network and vehicular conditions. The Data and Knowledge Layer manages and semantically interprets the collected information, constructs knowledge graphs of driver states and contextual events, and supports the integration of prior driving knowledge for enhanced reasoning. The Cognition and Reasoning Layer applies advanced learning and inference mechanisms to interpret multimodal inputs, recognize driver states, predict behavioral intentions, and formulate context-aware strategies for distraction mitigation. Finally, the Capability and Application Layer functions as the executive component, connecting cognitive outputs to real-world interaction, and supports human-machine communication through digital human interfaces, semantic prompts, and feedback control mechanisms for adaptive intervention and behavioral regulation. Together, these five layers establish a theoretical structure that transitions from passive perception to active cognition and embodied execution, forming an integrated mechanism for closed-loop understanding and intervention. Unlike existing detection systems that primarily focus on recognition and alerting, the proposed embodied-intelligence-based framework highlights a conceptual transformation toward human-centered adaptability, self-evolving learning, and knowledge-driven decision-making. As a theoretical contribution, this study provides a unified and extensible foundation that bridges multimodal perception, cognitive reasoning, and control feedback under the embodied intelligence paradigm. The proposed framework is conceptual rather than engineering-oriented, serving as a blueprint for future empirical validation, experimental simulation, and real-world deployment. Meanwhile, the implementation of the proposed framework still faces considerable challenges. From the technical perspective, the constrained computational resources of in-vehicle terminals limit the deployment of complex deep learning models; insufficient multimodal sensing and synchronization hinder the scalability of perception systems; and inference hallucination in large-scale reasoning models introduces uncertainty that threatens decision reliability and system safety. From the application perspective, ethical and legal controversies related to driver monitoring, data privacy, and algorithmic transparency remain unresolved, while high economic costs continue to impede large-scale deployment and commercialization. Addressing these challenges will require future research on lightweight optimization, trustworthy and interpretable inference, distributed learning, and regulatory standardization. Such efforts are expected to ensure that embodied-intelligence-based distracted driving detection can evolve toward safer, more efficient, transparent, and socially sustainable intelligent driving ecosystems. Overall, this study presents a systematic, forward-looking, and theoretically grounded exploration of embodied-intelligence-driven distracted driving detection. By consolidating fragmented advances in multimodal perception, driver behavior modeling, and adaptive decision-making into a unified theoretical architecture, it redefines the principles of human-machine collaboration, establishes a pathway for embodied cognitive understanding, and provides theoretical guidance for the evolution of next-generation intelligent and safe driving systems.

Key words: driver distraction detection; deep learning; embodied intelligence; embodied perception; embodied execution; embodied learning and evolution; smart transportation

0 引言

交通事故已经成为全球重大的公共安全难题,不仅造成大量人员伤亡,也给社会经济带来沉重负担。根据世界卫生组织(WHO)最新的统计数据,全球有超过 119 万人因为交通事故身亡(Rosen 等, 2025)。分析事故原因显示,人为过失是主要因素,其中驾驶过程中的注意力分散问题尤为突出。所谓的分心驾驶是指驾驶员注意力转向与当前驾驶任务无关的活动,造成驾驶操作能力下降的一种现象,典型的分心驾驶的行为包括使用手机、交谈和饮食等。

如图 1 所示,从信息来源的感知角度出发,现有分心驾驶检测方法可划分为生理信号、车辆运动信息与视觉信息三大类。基于驾驶员生理信号的检测方法依靠生理信号传感器,通过监测驾驶员的心率(Freitas 等, 2024)、脑电波(Kim 等, 2025)、眼电(Zhou 等, 2025)等生理指标来分析驾驶员的注意力状态。该类方法通常需要驾驶员长时间佩戴传感器,可能导致驾驶员身体不适,进而影响正常驾驶体验。生理信号对疲劳或注意力分散的反应也存在一定的滞后性,难以准确反映驾驶员的实时状态,从而限制了其在实际驾驶场景中的应用效果。

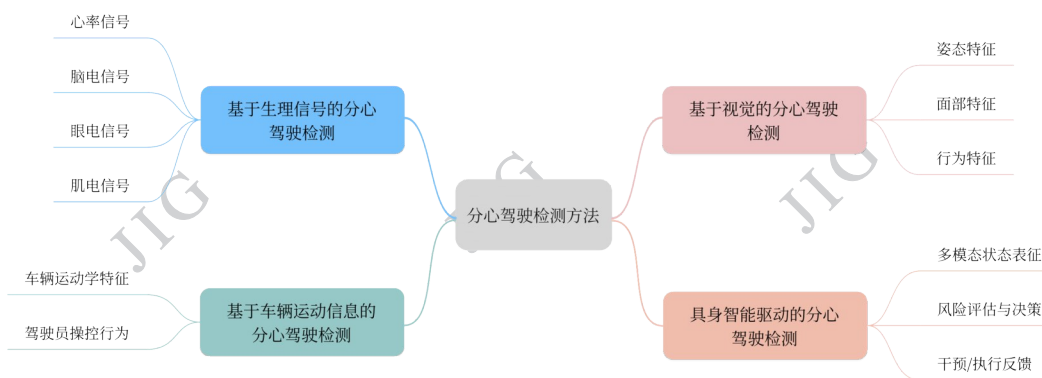


图 1 分心驾驶检测方法的类别划分与信息来源

Fig. 1 Categorization of distracted driving detection methods and their information sources

除生理信号感知方法外,基于车辆运动信息的分心驾驶检测方法主要通过分析车辆运行状态的变化来间接推断驾驶员的注意力状态。该类方法通常基于车辆动态信号对驾驶行为进行建模。例如,一类研究侧重于方向盘转角以及车辆横向和纵向加速度等运动学特征(张辉等, 2018)。另一类研究则主要关注车道偏移量以及油门和制动踏板操作等操控行为特征(张玉婷等, 2022)。基于车辆运动信息的分心驾驶检测方法具有实现简单、部署成本低、数据获取便捷的优势。然而,由于基于车辆运动信息的检测方法主要依赖车辆运行状态的外在表现,缺乏对驾驶员关键动作细节及视觉行为特征的直接刻画,检测结果容易受到车辆类型差异、道路交通环境变化以及个体驾驶习惯等因素的影响。

随着计算机视觉和深度学习的发展,基于车载摄像头的非接触式驾驶行为检测已经成为研究的主流,在特征提取和行为分析方面表现出优越的性能。Hao 等人(2025)基于轻量化卷积设计了一种具备可

解释性的驾驶员异常行为识别网络模型,该模型能够满足边缘设备的部署需求。Tang 等人(2024)结合卷积神经网络(convolutional neural networks, CNN)和Transformer研发了一种轻量级分心驾驶检测系统,并在经典的分心驾驶行为数据集 AUC 和 SFDDD 上验证了其有效性。潘健东等人(2025)则结合 YOLO 算法设计了一种驾驶员分心驾驶行为检测模型,具有部署灵活与非侵入式的特点,更适合复杂多变的驾驶环境。

然而,基于深度学习的方法在实际应用中仍面临诸多挑战。一方面,这些方法多依赖于静态图像或短视频,难以有效建模驾驶行为的时序演化特征(Yang 等, 2023)。另一方面,模型通常将视觉信息作为认知来源,在光照变化、遮挡干扰的复杂驾驶环境下容易出现性能波动,影响系统的可靠性和稳定性(Cheng 等, 2023)。此外,传统的深度学习方法普遍采用离线训练模式,难以针对驾驶员个体习惯与动态驾驶场景实现在线学习。更突出的问题是,现

有公开数据集覆盖场景有限,图像采集主要来源于少量车辆样本,驾驶情境缺乏多样性,驾驶员行为标签的维度也较为单一,缺乏对复杂行为组合及其演化过程的精细标注,制约了模型的泛化能力(胡学敏等,2024)。在系统层面,现有分心驾驶检测系统大多仍停留在识别与报警层面,缺乏与车辆控制模块的深度联动,难以在突发风险情境时充分保障交通安全(Koniakowsky等,2026)。上述研究表明,在分心驾驶检测任务中,仍有必要在状态感知、决策分析与行为反馈之间建立更加紧密的系统级协同机制。

为应对传统方法在开放环境下感知维度、适应能力和行为建模方面的局限,近年来具身智能作为人工智能的重要发展方向受到广泛关注,在智慧交通(Chen等,2025)、智能制造(Tan等,2025)、人形机器人(张伟男等,2025)等领域日益展现出显著的应用潜力。2025年《政府工作报告》首次提出要“大力培育发展以具身智能为代表的新兴未来产业”,进一步凸显了该领域在国家战略层面的前瞻价值与发展空间。在此背景下,具身智能强调智能体与现实环境之间的动态耦合和深度协同,通过物理交互实现认知的交互学习,进而赋予智能体生成智能行为与自主进化的能力,为驾驶场景中的实时决策与行为调控提供了新的理论基础和技术支撑,有效缓解现有方法在复杂驾驶环境和驾驶员个体差异下适应性

不足的问题。因此,在分心驾驶检测中引入具身智能,可为提升检测系统在真实交通场景中的驾驶智能可靠性和稳定性提供技术思路。

本文的贡献如下:

1)从具身认知视角对分心驾驶检测问题进行建模分析,将驾驶行为置于驾驶员—车辆—环境持续交互的动态语境中加以分析;

2)构建了一种以具身智能为核心驱动机制的分心驾驶检测系统框架,采用分层式系统设计,在系统层面实现多模态感知、知识建模与认知推理的闭环协同;

3)系统性梳理并对比多模态智能驾驶大模型、领域自适应与持续学习等技术在具身智能驱动的分心驾驶检测系统中的功能定位与适用场景,分析其在复杂驾驶环境下支撑系统鲁棒性与跨场景适应能力的作用,为具身智能方法在智慧交通与智能驾驶安全领域的应用研究提供参考。

论文结构安排如图2所示,第1节从具身智能视角出发,梳理相关理论基础与系统框架,为后续研究提供统一的认知与方法背景;第2节围绕分心驾驶检测任务,系统综述多模态智能驾驶垂域大模型、领域自适应与持续学习等关键技术;在此基础上,第3节构建具身智能驱动的分心驾驶检测系统架构,并对系统功能进行了分层分析;最后在第5节与第6节分别给出未来展望与总结。

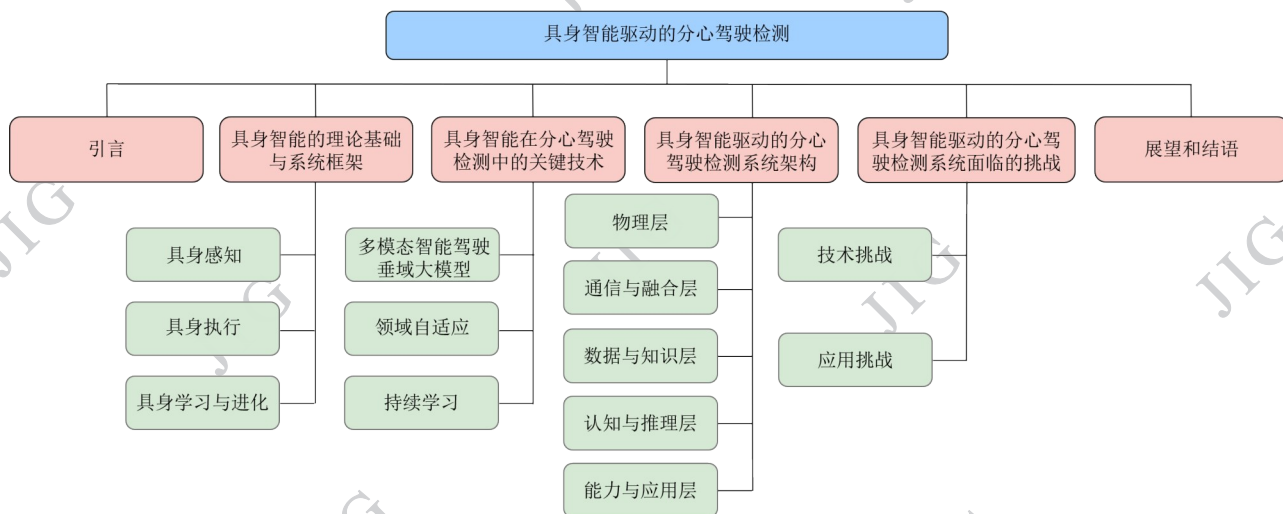


图2 本文结构框架

Fig. 2 The Designed Overall Structure

1 具身智能的理论基础与系统框架

在分心驾驶检测场景下,具身智能是一类以物理实体作为载体的智能系统(卫培刚等,2025),凭借人工智能、传感器、机器人控制和系统设计等技术原理,通过智能体与环境的实时交互,形成以行为反馈驱动动态闭环机制,使执行结果能够反向影响系统后续的感知与决策,实现“具身感知—具身执行—具身学习与进化”的连续循环。区别于传统一次性报警式检测,具身智能强调交互反馈与策略自适应,具身智能驱动的交互模型如图3所示。

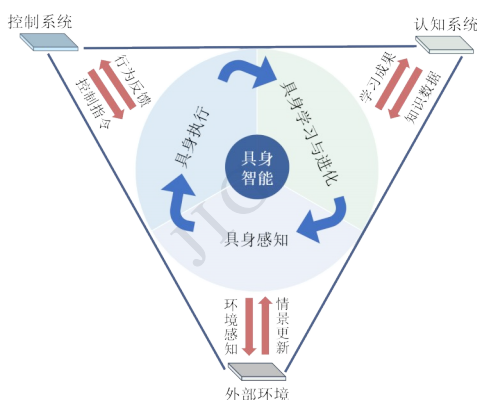


图3 具身智能驱动的三元交互模型

Fig. 3 A Triadic Interaction Model Driven by Embodied Intelligence

1.1 具身感知

具身感知是具身智能驱动的分心驾驶检测系统实现驾驶员状态识别与环境理解的基础能力。分心驾驶检测任务中,驾驶员分心行为通常在驾驶过程中随姿态变化、操作行为及注意力转移等逐步显现,使得分心状态的感知与判别具有一定难度。具身感

知通过智能体与驾驶环境的持续物理交互,对多模态感知信息进行联合采集与解析,以支持对驾驶员行为状态、注意力变化及其与车辆和道路环境之间关系的动态刻画与综合分析。其核心技术主要包括:

1)多源交互与语义理解:系统基于多源异构数据,动态识别驾驶员与环境及其行为目标之间的空间交互关系。该技术实现了对空间关系的深度理解和行为目标的精准定位,为具身智能体的决策机制提供了环境感知基础,促进了感知、认知与执行环节

高效衔接。

2)多模态融合与自主感知:系统融合图像、语音、红外等多模态信息,弥补了单一模态在复杂驾驶环境下面临的信息缺失与稳定性不足问题,从而提升分心驾驶检测在真实场景中的鲁棒性与可靠性,有助于形成对外部环境及个体状态更完整的描述。

3)动态自适应建模:系统通过整合多源知识并引入跨域适配机制,实现对感知模型的在线更新,从而有效缓解传统模型在固定训练条件下易出现的性能退化问题,有效增强了分心驾驶检测系统的跨场景适应能力与个性化建模能力。

针对具身智能在智能驾驶中的应用,如图4所示,在分心驾驶检测中,具身感知整合车内摄像头、传感器和驾驶行为数据,实时捕捉驾驶员的姿态。系统基于多源信息联合推理,刻画驾驶行为与环境的交互,为智能决策与干预提供有力支持。

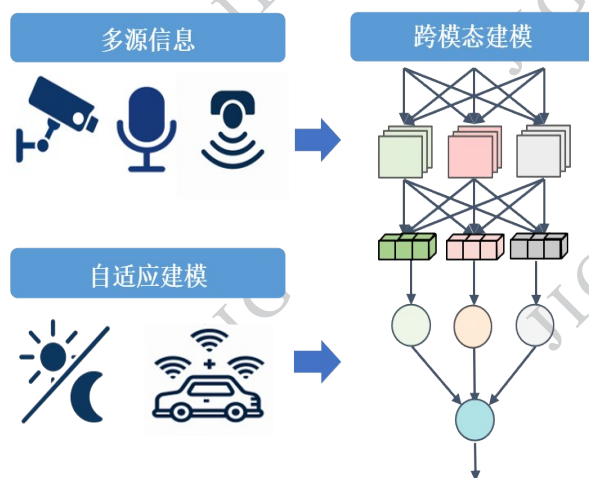


图4 具身感知系统的核心组成与建模结构

Fig. 4 Schematic Core Technologies and Modeling Architecture of Embodied Perception

1.2 具身执行

具身执行是具身智能驱动的分心驾驶检测系统将感知与认知结果转化为面向驾驶员的实际响应与交互行为的关键环节,其核心目标是在检测到驾驶员分心状态或风险趋势后,实现及时、合理且可控的反馈与干预。具身执行关联的关键技术环节如下:

1)状态判断与指令生成:状态判断模块负责整合分析各个来源的信息,识别现阶段驾驶员的行为状态和潜在目的。系统根据识别结果生成操控指令,通过策略映射驱动反馈并实施模式切换。

2)控制模块联动执行:系统基于分心风险识别结果生成决策,触发控制模块与执行单元协同响应,形成由感知状态到物理响应的闭环联动。控制模块涵盖语音、图像、触觉等提示机制,可协同车辆控制系统,实现快速反应和多模态反馈。

3)人机交互与可解释性反馈:系统通过自然语言、视觉符号、触觉提示等多模态交互方式,向人类用户传达感知结果和决策依据,增强用户对系统行为的理解和信任。可解释性反馈有助于驾驶员理解系统行为,配合系统完成干预控制,是实现人机协同与闭环控制的重要保障。

面向具身智能的驾驶安全,如图5所示,在分心驾驶检测中,具身执行主要负责驾驶状态识别结果的实时反馈、干预策略的触发控制,以及人机交互界面的协调。该过程依赖于低延迟的数据处理、决策机制与反馈机制,确保系统在驾驶员行为出现异常时,能及时发出提醒或介入控制,降低发生潜在风险的可能性。



图5 具身执行的核心机制

Fig. 5 Core Mechanism of Embodied Execution

1.3 具身学习与进化

学习与进化是推动智能系统自主操作能力发展的关键机制,是具身智能驱动的分心驾驶检测系统在长期运行过程中保持适应性与一致性的重要理论基础。具身学习与进化通过“印象-记忆-知识”模型,实现对驾驶环境感知信息的不断积累,逐步完成认知构建。该机制在结构与功能上借鉴了人脑的处理模式,有助于提升系统的环境适应性、自主规划与演化能力,是支撑高阶智能行为实现的核心技术基础(Zhou 等, 2024)。

如图6所示,具身学习与进化的印象-记忆-知识机制主要包括:

1)印象形成阶段:印象是智能体与环境交互过程中产生的初步感知与行为记录,具有即时性。在该阶段,系统通过多模态传感器获取感知数据,并结合当前行为决策,构建对外部事件的初步表

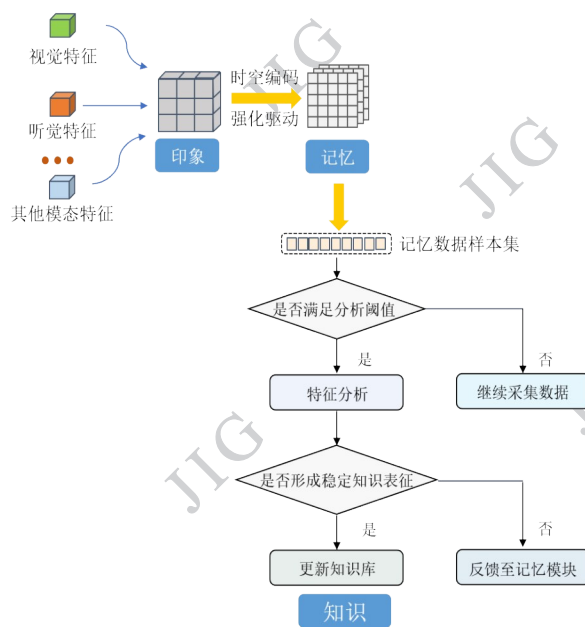


图6 具身智能中的印象-记忆-知识机制

Fig. 6 Core Mechanism of Embodied Execution

征。该阶段在分心主要用于记录驾驶员姿态变化、操作行为及环境状态的初始感知特征,为后续行为分析提供基础信息。

2)记忆存储与提取阶段:系统对交互事件进行筛选、压缩和编码,将策略信息存入记忆单元。借助强化学习及自监督机制,系统得以形成长期记忆。面对相似情境时,可迅速调取历史经验,完成策略迁移,增强模型泛化能力。在分心驾驶检测中,该阶段有助于积累不同驾驶员及不同情境下的分心行为模式,使系统能够在相似驾驶状态下快速参考历史经验。

3)知识抽象与策略优化阶段:系统通过对记忆中数据的归纳分析,提取出通用性行为规律和因果关系,完成从“经验”向“知识”的跃迁。对于分心驾驶检测任务,该阶段使系统能够形成对分心行为具有判别性的高层行为知识,并为检测策略的持续优化提供依据。

在分心驾驶检测中,具身学习与进化机制侧重于持续建模驾驶员行为,动态优化检测策略。系统通过对多模态感知数据的长期积累,构建行为表征、提取关键记忆,并在此基础上形成通用知识和个性化策略,实现对分心状态的精准识别和自适应干预。

2 具身智能在分心驾驶检测中的关键

技术

面向分心驾驶检测中多模态行为理解、跨场景适应与长期行为演化建模等关键需求, 相关研究逐步形成了支撑具身智能“感知—执行—学习”闭环能力的三类关键技术方向, 包括多模态智能驾驶垂域大模型、领域自适应与持续学习。多模态智能驾驶垂域大模型通过统一多源感知信息实现驾驶行为的语义建模, 领域自适应保障语义表示在不同场

景与条件下的稳定迁移, 而持续学习进一步从时间维度支持驾驶行为模式的长期演化建模。三类技术分别从语义建模、环境适应与经验演化等层面体现具身智能的核心能力, 共同构成具身智能驱动的分心驾驶检测体系的关键技术支撑。在系统层面, 这些技术围绕驾驶行为感知、决策输出与真实交互反馈之间形成闭环协同机制, 如图7所示。表1对具有代表性的相关研究进行了对比分析, 从关键技术、

表1 具身智能关键技术在分心驾驶检测中的代表性研究对比
Table 1 Representative Studies of Key Embodied Intelligence Technologies for Distracted Driving Detection

参考文献	关键技术	解决的分心驾驶问题	支撑的系统层级
Zhou 等 , 2025	多模态生理信号融合	自动驾驶场景下驾驶员分心状态识别	感知层; 数据与知识层
Yang 等 , 2024	视觉-姿态融合的时序行为建模	基于人体姿态与外观特征的分心驾驶行为识别	感知层; 认知与推理层
Hasan 等 , 2024	视觉-语言模型	自然驾驶视频中的分心行为语义识别	感知层; 认知推理层
Zhang 等 , 2024	基于多模态预训练大模型	复杂分心驾驶行为的语义理解与可解释风险评估	数据与知识层; 认知与推理层
Li 等, 2024	领域自适应	跨数据集与真实部署场景中的领域偏移	数据与知识层; 认知与推理层
Duan 等 , 2024	Score-Softmax 学习策略	分心驾驶检测在跨数据集场景下的识别性能	数据与知识层; 认知与推理层
Yang 等 , 2025	持续学习	分心驾驶行为的长期演化与个性化适配	数据与知识层; 认知与推理层; 能力与应用层
Ma 等, 2024	驾驶员行为的时序建模与在线更新	面向时序变化的分心驾驶行为识别与定位	认知与推理层; 能力与应用层

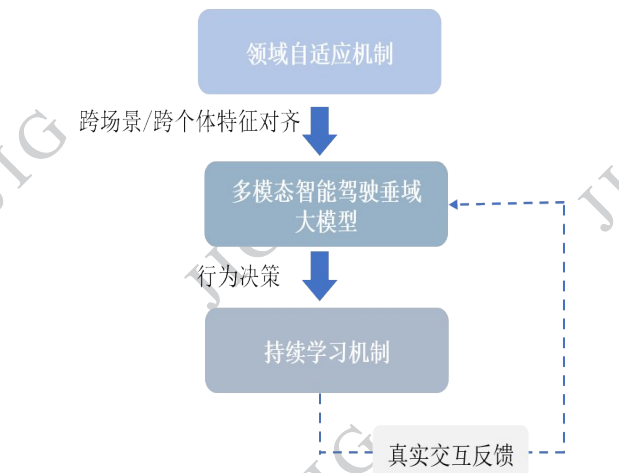


图7 具身智能驱动分心驾驶检测的关键技术闭环机制
Fig. 7 Collaborative framework of key embodied intelligence techniques for distracted driving detection

解决的分心驾驶问题以及支撑的系统层级三个维度进行总结。

2.1 多模态智能驾驶垂域大模型

如图8所示, 多模态智能驾驶垂域大模型是近年来人工智能领域的重要发展方向之一, 通过整合视频、语音、生理信号及环境感知等多源模态信息, 构建能够理解驾驶员状态、道路环境及车辆动态关系的统一模型。相较于通用多模态模型, 智能驾驶垂域大模型更注重跨模态关联与时空建模, 能够学习交通场景的动态语义结构和驾驶行为逻辑, 从而

为分心检测、疲劳识别及风险预测等任务提供统一的认知表达与决策支撑(Zhou 等, 2024)。

在垂域化多模态模型的发展过程中, 大语言模型(large language models, LLM)提供了强大的语义理解与跨模态对齐能力, 为多模态驾驶感知信息的

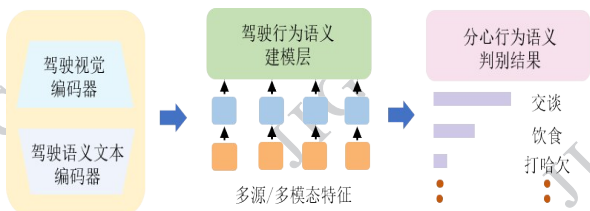


图8 多模态智能驾驶垂域大模型的语义建模框架

Fig. 8 Semantic Modeling Framework of a Multimodal Domain-Specific Driving Model

统一表示与任务迁移奠定了基础(闫啸彤等, 2024)。近年来,基于Transformer架构的大语言模型已被广泛用于驾驶场景中的语义描述生成、场景问答及语义指令理解,实现了从感知信息到语言表达的闭环映射,为构建具备驾驶场景语义理解能力的多模态垂域模型提供了可行路径。

智能驾驶场景中,驾驶环境、道路要素与驾驶员行为之间存在复杂的多模态关联,这对模型的多源信息理解与语义推理能力提出了更高要求。为此,研究者进一步探索了语言与其他模态的深度融合,其中视觉-语言大模型(vision-language models, VLM)作为多模态大模型的代表性分支,通过图像与语言的深度融合,实现跨模态的理解与描述,推动了感知与推理的深度结合。如图9所示,典型的视觉-语言大模型一般包括三个关键部分:视觉编码器负责将图像信息编码为高维特征表示;语言解码器基于视觉嵌入生成对应的自然语言描述或指令;跨模态融合模块则被用于整合视觉与语言信息,增强模型对多模态输入的整体理解能力,为智能驾驶场景中的行为识别与状态推理提供了有力支持。

多模态垂域大模型因其具备统一表征学习、上下文感知和跨模态推理能力,已成为智能驾驶研究中的核心方向之一。它能够支持智能驾驶系统从多个维度感知交通状况,实现对驾驶员意图、道路动态和环境风险的综合理解,从而为智能驾驶任务提供了全新的解决路径(Driessen等, 2024)。在应用研究方面,Xu等人(2024)以多模态大模型为核心,构建了具备可解释性的端到端自动驾驶系统,该系统支持语音交互,驾驶员可直接获取车辆状态、行驶环境与操作建议。在合法性研究中,Cui等人(2023)提出将具备法律判断能力的大模型 DriveLLM 集成至驾驶系统,使其能够识别并拒绝驾驶员的非法操作请求,从而增强了自动驾驶系统在实际应用中的合

规性。在驾驶员行为检测方面,Zhang等人(2024)训练了一种可定位驾驶员全身姿态的视觉-语言模型,用于提取驾驶员的行为线索,结合分心驾驶推理链框架,实现对驾驶员分心行为的细粒度检测。基于公开分心检测数据集 100-Driver(Wang等, 2023)的实验结果显示,该方法在各项指标均优于当前主流模型,展示出较强的鲁棒性和推广能力。基于深度学习的分心驾驶识别系统多采用串联式结构,先由图像识别模型提取静态或动态特征,再由单独设计的行为分类器进行判别与决策(Abbass等, 2024)。然而,在真实驾驶环境中,驾驶员的分心状态往往是多模态交互和连续行为演化的结果,依赖单一模态或短时片段极易造成误判(Wang等, 2024)。相比之下,具身智能驱动的分心驾驶检测系统通过引入多模态智能驾驶垂域大模型,可融合来自摄像头、麦克风、生理传感器的多源数据,进行感知与执行的一体化建模,实现驾驶员状态的实时建模和高效响应,增强系统在复杂驾驶环境下的鲁棒性与智能交互能力(陈妍妍等, 2024)。通过实时决策与自适应干预,系统能够在发现驾驶分心的同时输出语音提示、触觉反馈或紧急制动等控制命令,从而显著增强了驾驶安全性。随着模型架构的优化和算力提升,多模态垂域大模型将在具身智能系统中扮演越来越关键的角色。

2.2 智能驾驶中的领域自适应

领域自适应属于迁移学习的一种方法,常见于智能驾驶等多源异构场景,以解决源任务与目标任

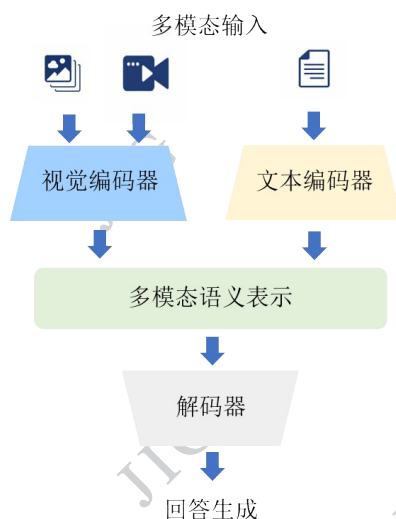


图9 典型视觉-语言生成模型

Fig. 9 Typical Vision-Language Generation Model

务一致但数据分布存在差异的问题。如图 10 所示, 在分心驾驶检测系统中, 领域自适应通过引入共享特征提取与领域对齐机制, 将来自不同场景(如白天与夜间)的多源驾驶数据映射至统一的跨域行为表示空间, 从而在保持行为语义一致性的前提下, 实现对分心驾驶行为的稳定判别。领域自适应的核心思想是通过缩小源域与目标域分布差异, 实现模型在目标域的有效迁移(崔福伟等, 2023)。该方法有利于提升模型在目标域的泛化性能, 降低数据的标注成本, 增强系统在跨域部署时的稳定性和安全性。深度学习中的领域自适应方法主要包括三类: 基于分布差异的领域自适应、基于域重构的领域自适应和基于对抗学习的领域自适应。

在多源感知任务中, 传感器类型、采集环境和行为模式差异显著, 不同场景间数据分布往往存在明显偏移。针对这一问题, 基于分布差异的领域自适应通过优化深度网络模型来减少源域与目标域分布差异, 可在有标签或无标签数据上进行训练; 基于重构的领域自适应方法利用自编码, 对源域样本进行编码器参数学习, 使解码器能重构目标域样本, 提取具有跨域一致性的特征表示; 基于对抗学习的方法则通过引入对抗训练机制, 借助对抗优化来缩

为代表性的应用对模型的泛化性能提出了较高要求, 车载摄像头的安装位置、光照条件、天气变化等因素均会影响系统对驾驶员状态判断的准确性(Liu 等, 2024)。为克服这些困难, 研究人员尝试引入领域自适应技术, 通过缩小源域与目标域之间的分布差异, 使模型在不同场景间迁移时保持稳定表现。在视觉感知方面, Wang 等人(2023)提出了一种基于对抗训练的多尺度领域自适应网络, 显著提升了模型的跨域泛化能力。Li 等人(2024)构建了一种基于部分特征对齐的领域自适应方法, 通过引入 PMMD 损失函数和分类混淆最小化策略, 有效解决了驾驶员分心检测中的域转移问题, 并减少了对大量标注数据的依赖。针对生理信号检测, Wang 等人(2024)注意到疲劳驾驶与睡眠脑电在频域特征上的相似性, 将在公开睡眠数据集 Sleep-EDF 上训练得到的模型迁移应用于疲劳驾驶检测中, 有效缓解了数据不足对模型性能的制约。

在具身智能驱动的分心驾驶检测系统中, 领域自适应不再局限于单一任务的改进手段, 而是支撑系统整体鲁棒性与跨场景适应性的关键机制。在视觉感知层面, 领域自适应能够实现跨设备与跨场景的迁移; 在生理信号层面, 则能够实现跨任务的知识共享。通过多层次的适应与迁移, 系统在复杂环境中展现出更强的鲁棒性与自适应能力, 并为自主演化与多场景推广奠定了坚实基础。

2.3 智能驾驶中的持续学习

领域自适应方法虽然在一定程度上缓解了跨域分布差异带来的性能退化, 但面对驾驶行为随时间变化导致的动态分布偏移, 仍存在适应不足的问题。因此, 持续学习(Continual Learning)机制被引入智能驾驶系统, 以应对时间维度上的知识更新与模型泛化问题。持续学习强调智能体在面对不断变化的任务或环境时, 能够在保留既有知识的基础上不断习得新知识。相较于一次性训练的静态学习方式, 持续学习有助于提升系统的适应性和泛化能力, 是具身智能系统实现自主进化与终身学习的核心机制之一(Lyu 等, 2025)。

持续学习的方法主要是: 基于正则化的方法(Rannen 等, 2017)、基于记忆重放的方法(Liu 等, 2020)和基于动态结构的方法(Yan 等, 2024)。基于正则化的方法, 是通过在损失函数中加入正则化项, 以限制模型参数在学习新任务时的更新, 降低对已

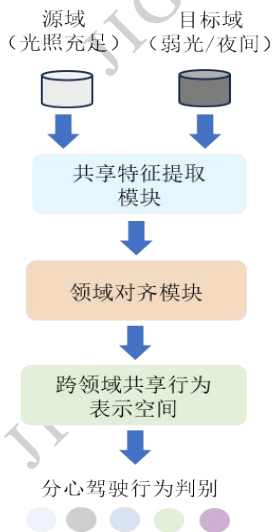


图 10 智能驾驶分心检测中的领域自适应机制示意图
Fig. 10 Schematic Illustration of Domain Adaptation Mechanism for Distracted Driving Detection

小不同域之间的特征分布差异, 使模型获得领域不变性表示。

具体到智能驾驶的任务层面, 以分心驾驶检测

有知识的遗忘;基于记忆重放的方法则是通过定期使用旧任务数据,使模型在接受新知识时仍保持对已有知识的记忆,有效缓解了灾难性遗忘;基于动态结构的方法是采用扩展模型结构的方法同时适应新旧任务,特别适用于任务间差异明显的复杂情境。随着预训练模型在人工智能中的广泛应用(Wang等,2023),基于预训练的持续学习方法展现出更强大的泛化能力(路悦等,2025),为智能系统在长期动态任务中实现自我进化提供了重要支持。

如图11所示,智能驾驶中的持续学习机制通过引入驾驶行为长期记忆模块,将历史驾驶数据与当前任务决策相结合,在系统运行过程中不断更新模型参数与行为表征,从而在时间维度上实现对驾驶行为演化的持续建模。Yuan等人(2023)针对驾驶员行为识别系统中使用车载摄像头而引发隐私泄露的问题,提出了一种点对点的持续学习框架。该框架无需中央服务器,支持用户级个性化建模,适用于隐私敏感、多终端、自主适应等智能驾驶场景。Yang等人(2025)构建了一种人类引导的持续学习框架,使自动驾驶系统能自动学习驾驶员在接管车辆时的操作行为,增强个性化决策能力,助力人-车共驾系统迈向智能化。Fanan等人则(2025)设计并评估了一种融合持续学习机制的驾驶员身份识别方法,在鲁棒性和跨任务迁移能力方面表现突出。

在具身智能驱动的分心驾驶检测系统中,持续学习机制发挥着关键作用。由于驾驶行为具有高度动态性和个体差异性,系统难以依赖一次性训练模型覆盖所有潜在情境。持续学习能够在此基础上完成模型更新与性能保持,确保系统在长期运行中维持识别结果的稳定性与准确性。

3 具身智能驱动的分心驾驶检测系统架构

分心驾驶检测涉及对驾驶员行为、车辆状态与交通环境的持续感知与综合分析,其过程不仅依赖多模态信息的同步获取,还需要在时间维度上对行为状态进行连续理解与判断,并进一步与风险预警和干预反馈形成联动。围绕上述任务特性,相关研究从多模态语义建模、跨场景适应以及长期行为演化等方面展开了系统性探索,如表1所示。这种从感知到认知、再到决策与执行的连续处理特性,使系

智能驾驶中的持续学习机制

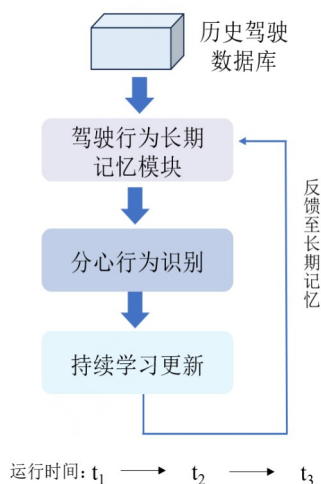


图11 智能驾驶中持续学习机制示意图

Fig. 11 Schematic Illustration of the Continual Learning Mechanism in Intelligent Driving

统在结构组织上需要同时兼顾功能解耦与信息协同。为满足上述需求,分层化系统设计可作为一种有效的系统组织方式,其通过对感知、建模、推理与执行等功能进行层级划分,有助于构建完整的信息处理闭环。相关研究已在具身智能与自主系统中探索了类似的分层化组织方式,用于支撑复杂环境下的持续感知与决策过程(Ren等,2024)。如图12所示,具身智能驱动的分心驾驶检测系统通过自

底向上的层级结构实现对驾驶行为与环境状态的持续感知,并在高层完成分心状态判断与行为干预,同时将执行结果回流至系统内部,支持长期运行中的模型更新与策略调整,从而为分心驾驶检测中复杂行为理解、跨场景适应与持续运行提供清晰的系统组织方式。

3.1 物理层

物理层作为具身智能驱动的分心驾驶检测系统与外部环境交互的起点,涵盖多源数据的采集和初步处理,为多模态语义建模与跨场景分析提供原始感知基础。该层集成了多类传感设备,例如:多模态摄像头用于获取视觉成觉信息,包括驾驶员的面部特征和手部动作;车载麦克风用于识别驾驶员说话和通话行为(Marques等,2022);方向盘上集成的嵌入式传感器用于非侵入式采集肌电等生理信号(Lu等,2021);电容传感器布设于驾驶舱中,用于提取驾

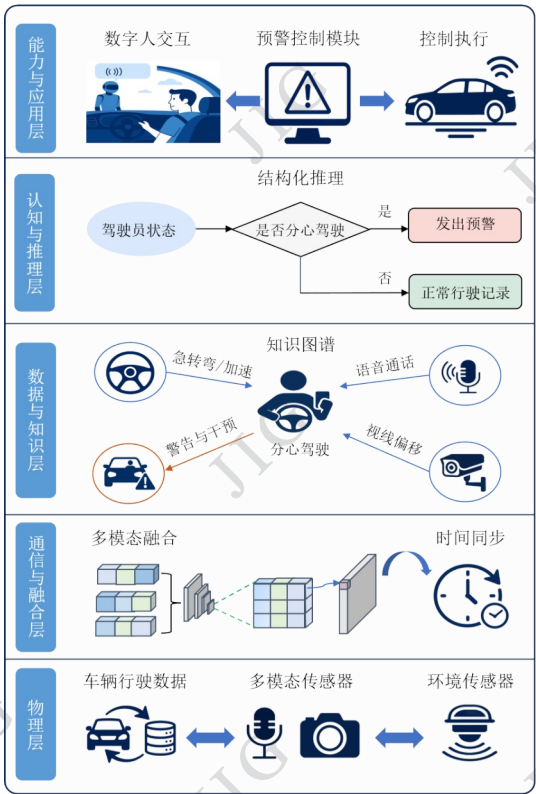


图 12 具身智能驱动的分心驾驶检测系统功能层级架构
Fig. 12 Functional Hierarchical Architecture of an Embodied Intelligence-Driven Distracted Driving Detection System

驶员的姿态特征(Sebastian 等, 2025)。车辆运行过程中的速度、加速度、车道偏移和方向盘转速等动态数据亦可作为关键补充输入,有助于提升系统对分心行为的实时感知和判别能力(Wang 等, 2022)。

3.2 通信与融合层

通信与融合层作为连接物理层与数据分析模块的桥梁,承担多源感知数据的协同传输、时间对齐和预处理任务,是系统实时响应和稳定运行的重要支撑环节。针对视觉、语音、生理等模态在数据规模和时效性方面的差异,该层在实现异构数据传输和时间同步的基础上,引入多模态融合机制以减少信息丢失,统一表征多模态信息(Liu 等, 2023)。在融合机制设计方面,通信与融合层注重多模态数据之间的结构对齐与语义整合。不同模态信息经预处理后被统一建模为共享特征表示,从而提升系统对复杂驾驶场景的理解能力(Guan 等, 2025)。通过对多模态信息的语义抽象与结构化建模,融合结果一方面为多模态垂域大模型提供可解释的高层语义支撑,另一方面为跨场景条件下的知识共享与迁移提供统一表示基础,并为后续持续学习过程中的知识更新

与积累提供载体。

3.3 数据与知识层

数据与知识层作为具身智能系统的信息中枢,发挥着多源数据的统一管理、语义抽象与知识建构的作用,是实现系统认知推理的重要基础。该层通过规范化策略,该层提取并组织驾驶相关的关键信息,构建可复用、可扩展的结构化知识模块,为上层智能单元提供稳定的数据支撑。为增强系统的语义理解能力,数据与知识层引入知识图谱方法,利用图结构显式建模驾驶行为之间的语义关联,提升模型对复杂驾驶情境的判断能力(Xu 等, 2022)。在此基础上,该层还构建了高效的访问与反馈机制,一方面支持认知模块对知识单元的快速查询和调度,另一方面通过推理结果回传实现知识库的动态更新与持续优化,从而进一步增强系统的知识表达能力与自适应决策水平。

3.4 认知与推理层

认知与推理层是具身智能系统中承担认知建模与智能推理的核心模块,通过融合感知信息和结构化知识,实现驾驶状态的理解与行为判断,是系统认知能力的核心支撑。该层引入自注意力机制实现全局依赖建模策略(Peng 等, 2022);利用图神经网络高效表达结构化关系(Liu 等, 2021);结合先验知识和实时感知数据,完成驾驶行为的语义解析和情境推理。为提升瞬时分心状态的快速识别能力,该层集成了视觉 Mamba 模型(Li 等, 2024)和时间序列建模方法(Wharton 等, 2021),增强对驾驶行为时序演化特征的建模效果。推理结果可回传至数据与知识层,促进具身智能系统持续学习与结构优化,从而构建具备自我进化能力的智能闭环。

3.5 能力与应用层

能力与应用层作为具身智能系统面向实际驾驶任务的输出终端,负责分心检测结果的应用部署、行为干预执行与系统服务调度等任务。该层通过构建多维功能模块,完成对驾驶员状态的持续感知、分心行为的智能预警,以及多任务环境下的能力动态调度,保障系统在复杂交通情境中的稳定运行和快速响应。

为增强人机交互的情境适配性和行为干预的接受度,系统进一步引入情智虚拟交互模块,可采用包括数字人(赵思成 等, 2025)在内的多种形式实现,作为人与具身智能系统之间的关键交互媒介。与传

统的语音模块相比,情智虚拟交互模块具备语义理解、情绪识别与拟人化表达的综合优势,能够通过自然流畅的语音、表情及肢体动作开展个性化提醒(王庞伟等,2025)。该能力显著增强了系统在分心检测情境中的干预效果,形成具有高接受度和引导性的行为调控机制。

在执行控制方面,能力与应用层集成驾驶安全预警模块与车辆控制接口,并与控制策略单元协同工作,以提升系统在复杂环境下对危险状态的响应效率和干预精度。系统在检测到分心行为后,可先通过语音提示、触觉反馈等柔性手段进行预警,若驾驶员未能及时响应,则进一步下达紧急制动、转向修正和车道保持等强制控制指令,实现从检测到干预的闭环控制。本层同时具备个性化学习与反馈机制(Marafie,2021),依据驾驶历史和个体行为差异自适应优化识别逻辑与干预策略,增强系统的实用性与人机协同体验。此外,该层产生的干预执行结果与驾驶员反馈可回传至系统内部,构成行为—反馈—更新的闭环,为持续学习机制在系统层面的实施提供实际运行基础。

4 具身智能驱动的分心驾驶检测系统面临的挑战

尽管具身智能驱动的分心驾驶检测系统在多模态感知融合、语义理解与行为推理等方面已展现出突出优势,但其在实际应用中所面临的挑战并非仅源于通用智能系统的共性技术瓶颈。受制于驾驶任务的安全关键属性与人—车—环境持续交互的影响,该类系统在分心状态建模、人机闭环干预、跨个体与跨设备适应以及风险可控推理等方面呈现出显著的任务独特性。

4.1 技术挑战

4.1.1 车载终端计算资源受限问题

具身智能驱动的分心驾驶检测系统必须适应复杂多变的交通场景,这就要求系统在极短的时间内融合处理图像、语音、生理信号等多源数据,以实现驾驶状态的实时分析与控制响应。但是由于车载终端计算资源的限制,目前低功耗设备在处理高频、异构的大规模感知数据时,往往无法满足系统对实时性能的要求。终端的硬件架构在算力匹配、内存管理及并行执行等方面与智能算法也存在明显的适

配障碍,这进一步限制了系统计算能力的释放和功能扩展能力的提高。

4.1.2 多模态数据支持能力不足

分心驾驶检测场景下的多模态数据,其语义往往由驾驶员姿态、操作行为与交通情境的持续交互共同决定,对数据采集、标注与对齐提出了更高要求。当前多模态建模已经取得一定进展,但大规模多模态数据的采集与构建仍面临诸多现实困难。现实场景下构建多模态分心驾驶数据集时,常面临多源数据采集链路复杂、设备部署受限、标注成本高昂等问题。虚拟仿真环境下生成的数据又难以有效映射真实行为模式,加剧了数据层面的适配难度。不同模态数据在时间粒度、空间特征和语义结构上存在显著的模态异构性,缺乏统一的对齐机制和融合规范,模态间容易产生语义冗余或融合偏差,融合策略稳定性下降,制约模型泛化性能。

4.1.3 推理幻觉问题对系统安全性的挑战

人工智能幻觉是人工智能模型在缺乏真实依据或语义支撑的条件下,生成虚假、不一致或不合理内容的现象,常见于语言生成、图像识别与认知推理等任务中(Huang等,2025)。在具身智能驱动的分心驾驶检测系统中,认知与推理模块依赖多模态输入与结构化知识对驾驶行为和状态进行综合判断。一旦模型出现认知错误或推理失真,易导致驾驶状态识别失准,干预响应偏差,严重时可能诱发系统性风险。因此,如何在具身智能架构中引入可信推理机制,构建针对模型幻觉的检测、校验与响应体系,实现对模型生成内容的校验和约束,是保障检测系统可靠性和安全性的关键挑战。

4.1.4 测试与验证体系仍待完善

对于智能驱动的分心驾驶检测系统性能的评估不仅体现在分心状态识别的准确性上,还需要综合考虑识别时机、运行稳定性及其对驾驶行为的潜在影响。然而,现有评测方法仍以离线指标和静态结果为主,难以刻画系统在真实驾驶过程中分心状态随时间演化的动态特征及其对整体驾驶行为的影响。此外,从算法研究到车载系统部署之间,尚缺乏统一的系统级测试规范与闭环验证机制,难以对不同方法在实际应用场景下的可靠性、稳定性与安全性进行系统评估,在一定程度上制约了具身智能分心驾驶检测系统的工程化验证与规模化应用(李永福等,2025)。

4.2 应用挑战

4.2.1 伦理与法律问题

尽管具身智能驱动的分心驾驶检测系统在提升驾驶安全和人机协同效率方面表现出广阔的应用前景,但在实际部署过程中仍暴露出一系列尚未充分解决的伦理与法律问题。现阶段,具身智能系统的责任归属划分缺乏清晰法律界定,相关伦理规范也缺乏统一共识(Kokciyan 等,2021)。全球范围内仍缺乏针对具身智能的统一法律框架与监管体系,监管机制的不完善对具身智能驱动的分心驾驶检测系统的部署、销售及跨区域运营构成现实的合规性挑战,其伦理后果与法律责任也因此呈现出更高的不确定性与敏感性。

4.2.2 经济成本与商业化问题

具身智能驱动的分心驾驶检测系统依赖多模态传感器、大规模计算与车载通信模块,其硬件配置与算法部署均带来较高成本。在长期运行中,传感器维护、计算资源升级与通信费用等问题进一步增加了系统的全生命周期成本。这在一定程度上限制了该类系统在不同车型和交通环境中的大规模推广与应用。

5 展望

展望未来,具身智能驱动的分心驾驶检测系统具有良好的发展前景和应用潜力。为了解决计算资源受限问题,推动模型向轻量化、低功耗方向发展;为缓解推理效率瓶颈,可通过构建“大模型-小模型”分层协同控制架构进行改善,由大模型承担任务分解与语义决策,而小模型负责底层感知与控制,从而实现高层决策和底层执行的高效互补。

在数据支撑方面,需系统性采集与标注多源异构数据,用以构建高质量具身智能分心驾驶数据集。同时结合计算机视觉和虚拟现实技术,创建高度还原的驾驶仿真环境,生成多场景、多状态的训练样本,以弥补真实数据的不足,提升数据的泛化能力和环境适应性。

平行智能理论的引入为具身智能系统的建模、验证与可解释性提供了新思路(Wang,2022)。与以数据驱动为核心的传统方法相比,平行智能通过建立可解释的人工系统并开展平行实验,实现行为推演、风险评估和任务适配,为提升系统在复杂交通环

境中的可控性与鲁棒性提供了理论依据。

为应对模型幻觉等问题导致的推理偏差,可设计多智能体协同验证机制,优化训练策略和反馈体系,提升系统对错误信息的判别能力,逐步构建具备可信推理能力与高风险防控能力的具身智能系统。随着跨模态感知、智能决策与人机交互等关键技术的不断突破,具身智能将在智能驾驶领域发挥越来越核心的功能,推动分心检测系统向更安全、更稳定、更智能化的方向持续演进。

6 结语

本文从具身智能的视角出发,介绍了具身智能的相关概念,系统性梳理了具身智能驱动的分心驾驶检测中的关键技术路径,并据此提出了一种涵盖物理层、通信与融合层、数据与知识层、认知与推理层以及能力与应用层的五层技术架构,为分心驾驶检测系统的构建提供了参考与借鉴。

同时,本文还深入分析了具身智能驱动的分心驾驶检测在实际应用中面临的技术挑战,并对未来可能的发展趋势进行了探讨。本研究为具身智能赋能的分心驾驶检测提供了系统化的研究思路和方法框架,将有助于推动具身智能驱动的智能驾驶安全系统的落地应用。

参考文献(References)

- Rosen H E, Khalaf M K, Mitra S, Neki K, Hyder A A and Paichadze N. 2025. Factors associated with mortality of hospitalized road traffic injury patients in 4 low-and middle-income countries. *Traffic injury prevention*: 1-6 [DOI: 10.1080/15389588.2025.2471558]
- Freitas A, Almeida R, Gonçalves H, Conceição G and Freitas A. 2024. Monitoring fatigue and drowsiness in motor vehicle occupants using electrocardiogram and heart rate—a systematic review. *Transportation research part F: traffic psychology and behaviour*, 103: 586-607 [DOI: 10.1016/j.trf.2024.05.008]
- Kim J H, Kim Y, Cho Y, Kim T K, Jang T, Park C and Kang S. K. 2025. Biosignal-based attention monitoring for evaluating train driver safety-relevant tasks. *Transportation Research Part F: Traffic Psychology and Behaviour*, 111: 1-13 [DOI: 10.1016/j.trf.2025.02.025]
- Zhou Y, Chen Y, Zhang Y. 2025. Driver Distraction Detection in Conditionally Automated Driving Using Multimodal Physiological and Ocular Signals. *Electronics*, 14 (19): 3811 [DOI: 10.3390/elec-

- tronics14193811]
- Zhang H, Qian D L, Shao C F, Qian Z W and Jian M Y. 2018. Identification of Driver Distraction States with Optimized Support Vector Machine Method. *Journal of Transportation Systems Engineering and Information Technology*, 18(01): 127-132 (张辉, 钱大琳, 邵春福, 钱振伟, 菅美英. 2018. 驾驶人分心状态判别支持向量机模型优化算法. *交通运输系统工程与信息*, 18(01): 127-132) [DOI: 10.16097/j.cnki.1009-6744.2018.01.019]
- Zhang Y T, Chen B Y, Zhang S Y, Yan X D and Li X M. 2022. Recognition Model of Distracted Drivers When Approaching Signalized Intersections. *Journal of Transportation Systems Engineering and Information Technology*, 22 (01): 217-224 (张玉婷, 陈波佑, 张双焱, 闫学东, 李晓梦. 2022. 抵近信控交叉口分心驾驶识别模型. *交通运输系统工程与信息*, 22 (01): 217-224) [DOI: 10.16097/j.cnki.1009-6744.2022.01.023]
- Hao J, Sun X, Liu X, Hua D and Hu J. 2025. A lightweight and explainable model for driver abnormal behavior recognition. *Engineering Applications of Artificial Intelligence*, 139: 109559 [DOI: 10.1016/j.engappai.2024.109559]
- Tang X, Chen Y, Ma Y, Yang W, Zhou H and Huang J. 2024. A lightweight model combining convolutional neural network and Transformer for driver distraction recognition. *Engineering Applications of Artificial Intelligence*, 132: 107910 [DOI: 10.1016/j.engappai.2024.107910]
- Pan J D, Fan C J, Jiang Z X, Deng H W, Wu X H, Peng Y. 2025. Real-Time Detection of Driver Distraction Behavior Based on YOLOv8. *Journal of Southwest University (Natural Science Edition)*, 47(8): 38-48 (潘健东, 范超杰, 姜卓希, 邓涵文, 伍贤辉, 彭勇. 2025. 基于YOLOv8的驾驶员分心行为实时检测研究. *西南大学学报(自然科学版)*, 47(8): 38-48) [DOI: 10.13718/j.cnki.xdsk.2025.08.004]
- Yang W, Tan C, Chen Y, Xia H, Tang X, Cao Y, Zhou W, Lin L, Dai G. 2023. BiRSwinT: Bilinear full-scale residual swin-transformer for fine-grained driver behavior recognition. *Journal of the Franklin Institute*, 360(2): 1166-1183 [DOI: 10.1016/j.jfranklin.2022.12.016]
- Cheng W, Wang X and Mao B. 2024. A multi-feature fusion algorithm for driver fatigue detection based on a lightweight convolutional neural network. *The Visual Computer*, 40(4): 2419-2441 [DOI: 10.1016/j.jfranklin.2022.12.016]
- Hu X M, Huang T Y, Yu Y L, Ren J J, Xie W and Chen L. 2024. Decision technologies of simulation to reality for autonomous driving: A survey. *Journal of Image and Graphics*, 29(11): 3173-3194 (胡学敏, 黄婷玉, 余雅澜, 任佳佳, 谢微, 陈龙. 2024. 仿真到现实环境的自动驾驶决策技术综述. *中国图象图形学报*, 29(11): 3173-3194) [DOI: 10.11834/jig.230780]
- Koniakowsky I, Forster Y, Krems J. F and Keinath A. 2026. Driver monitoring systems mitigate visual distraction only when drivers are adequately instructed and driving in assisted mode. *Transportation Research Part F: Traffic Psychology and Behaviour*, 116: 103396 [DOI: 10.1016/j.trf.2025.103396]
- Chen M, Wang C, He X, Zhu F, Wang L and Vasilakos A V. 2025. Embodied Artificial Intelligence-Enabled Internet of Vehicles: Challenges and Solutions. *IEEE Vehicular Technology Magazine*, 20 (2): 63 - 70 [DOI: 10.1109/MVT.2025.3544675]
- Tan J, Shi J, Wu L, Chen B, Tang H, Zhang C, Zhang W and Wang S. 2025. Embodied Intelligence Empowering Customized Manufacturing: Architecture, Opportunities, and Challenges. *IEEE Access*, 13: 92740 - 92755 [DOI: 10.1109/ACCESS.2025.3572778]
- Zhang W N, Liu T. 2025. Research and application of embodied intelligence. *CAAI Transactions on Intelligent Systems*, 20(1): 255-262 (张伟男, 刘挺. 2025. 具身智能的研究与应用. *智能系统学报*, 20(1): 255-262) [DOI: 10.11992/tis.202406044]
- Wei P G, Cao S S, Liu J F, Liu Z H, Sun W, Kong F T. 2025. Embodied Intelligent Agricultural Robots: Key Technologies. *Smart Agriculture (卫培刚, 曹姗姗, 刘继芳, 刘振虎, 孙伟, 孔繁涛. 2025. 具身智能农业机器人: 关键技术、应用分析、挑战与展望. 智慧农业(中英文) [DOI: 10.12133/j.smartag.SA202505008]*
- Zhou M, Dong H, Song H, Zheng N, Chen W H and Wang H. 2024. Embodied intelligence-based perception, decision-making, and control for autonomous operations of rail transportation. *IEEE Transactions on Intelligent Vehicles* [DOI: 10.1109/TIV.2024.3517335]
- Zhou Y, Chen Y and Zhang Y. 2025. Driver Distraction Detection in Conditionally Automated Driving Using Multimodal Physiological and Ocular Signals. *Electronics*, 14 (19): 3811 [DOI: 10.3390/electronics14193811]
- Yang X, Qiao Y, Han S, Feng Z and Chen Y. 2024. Appearance-posture fusion network for distracted driving behavior recognition. *Expert Systems with Applications*, 257: 124883 [DOI: 10.1016/j.eswa.2024.124883]
- Hasan M Z, Chen J, Wang J, Rahman M S, Joshi A, Velipasalar S, Hegde C, Sharma A and Sarkar, S. 2024. Vision-language models can identify distracted driver behavior from naturalistic videos. *IEEE Transactions on Intelligent Transportation Systems*, 25(9): 11602-11616 [DOI: 10.1109/TITS.2024.3381175]
- Zhang K, Wang S, Jia N, Zhao L, Han C and Li L. 2024. Integrating visual large language model and reasoning chain for driver behavior analysis and risk assessment. *Accident Analysis & Prevention*, 198: 107497 [DOI: 10.1016/j.aap.2024.107497]
- Li G, Wang G, Guo Z, Liu Q, Luo X, Yuan B, Li M and Yang L. 2024. Domain adaptive driver distraction detection based on partial feature alignment and confusion-minimized classification. *IEEE Transactions on Intelligent Transportation Systems*, 25(9): 11227-11240 [DOI: 10.1109/TITS.2024.3367665]
- Yang H, Zhou Y, Wu J, Liu H, Yang L and Lv C. 2025. Human-guided continual learning for personalized decision-making of autonomous driving. *IEEE Transactions on Intelligent Transportation Systems*

- tion Systems: 5435-5447 [DOI: 10.1109/TITS.2024.3524609]
- Ma Y, Du R, Abdelraouf A, Han K, Gupta R and Wang Z. 2024. Driver digital twin for online recognition of distracted driving behaviors. *IEEE Transactions on Intelligent Vehicles*, 9(2): 3168-3180 [DOI: 10.1109/TIV.2024.3353253]
- Zhou X, Liu M, Yurtsever E, Zagar B L, Zimmer W, Cao H and Knoll A. C. 2024. Vision language models in autonomous driving: A survey and outlook. *IEEE Transactions on Intelligent Vehicles* [DOI: 10.1109/TIV.2024.3402136]
- Yan X T, Tang X B, Shen T, Lei N. 2024. Overview of the development of large language models. *J Statistics*, 5: 13 - 18 (闫啸彤, 唐晓彬, 沈童, 雷诺. 2024. 大语言模型发展综述. *统计学报*, 5: 13 - 18) [DOI: 10.19820/j.cnki.ISSN2096-7411.2024.04.002]
- Driessen T, Dodou D, Bazilinskyy P and Winter J. 2024. Putting ChatGPT vision (GPT-4V) to the test: risk perception in traffic images. *Royal Society open science*, 11(5): 231676 [DOI: 10.1098/rsos.231676]
- Xu Z, Zhang Y, Xie E, Zhao Z, Guo Y, Wong K Y K, Li Z and Zhao H. 2024. Drivegpt4: Interpretable end-to-end autonomous driving via large language model. *IEEE Robotics and Automation Letters*, 9(10): 8186 - 8193 [DOI: 10.1109/LRA.2024.3440097]
- Cui Y, Huang S, Zhong J, Liu Z, Wang Y, Sun C, Li B, Wang X, Khajepour A. 2023. Drivellm: Charting the path toward full autonomous driving with large language models. *IEEE Transactions on Intelligent Vehicles*, 9(1): 1450-1464 [DOI: 10.1109/TIV.2023.3327715]
- Wang J, Li W, Li F, Zhang J, Wu Z, Zhong Z and Sebe N. 2023. 100-driver: A large-scale, diverse dataset for distracted driver classification. *IEEE Transactions on Intelligent Transportation Systems*, 24(7): 7061-7072 [DOI: 10.1109/TITS.2023.3255923]
- Abbass M A B and Ban Y. 2024. Mobilenet-based architecture for distracted human driver detection of autonomous cars. *Electronics*, 13(2): 365 [DOI: 10.3390/electronics13020365]
- Wang R, Wang W, Gao J, Lin, D, Yap K H and Li B. 2024. Multi-fuser: Multimodal fusion transformer for enhanced driver action recognition//2024 IEEE 26th International Workshop on Multimedia Signal Processing (MMSP). IEEE: 1-6 [DOI: 10.1109/MMSP61759.2024.10743612]
- Chen Y Y, Tian D X, Lin C M and Yin H B. 2024. Survey of end-to-end autonomous driving systems. *Journal of Image and Graphics*, 29(11): 3216-3237 (陈妍妍, 田大新, 林椿呀, 殷鸿博. 2024. 端到端自动驾驶系统研究综述. *中国图象图形学报*, 29(11): 3216-3237) [DOI: 10.11834/jig.230787]
- Cui F W, Wu X X, Chen Y F, Liu J, Xu J A. 2023. Survey of Domain Adaptive Methods with Knowledge Integrating. *Computer Science*, 50(08): 142-149 (崔福伟, 吴璇璇, 陈钰枫, 刘健, 徐金安. 2023. 融合知识的领域自适应方法综述. *计算机科学*, 50(08): 142-149) [DOI: 10.11896/jsjx.220800040]
- Liu W, Wang S, Zhu J, Xie X and Zhang L. 2024. Domain adaptation transformer for unsupervised driving-scene segmentation in adverse conditions. *IEEE Transactions on Intelligent Transportation Systems*, 25(12): 21129 - 21141 [DOI: 10.1109/TITS.2024.3461468]
- Wang J, Wu Z C. 2023. Driver distraction detection via multi-scale domain adaptation network. *IET Intelligent Transport Systems*, 17(9): 1742-1751 [DOI: 10.1049/itr2.12366]
- Wang F, Gu T, Yao W. 2024. Research on the application of the Sleep EEG Net model based on domain adaptation transfer in the detection of driving fatigue. *Biomedical Signal Processing and Control*, 90: 105832 [DOI: 10.1016/j.bspc.2023.105832]
- Lyu F, Wang L, Li X, Zheng W S, Zhang Z, Zhou T and Hu F Y. 2025. A comprehensive survey of continual learning. *Journal of image and graphics*, 1-32 (吕凡, 王亮, 李玺, 郑伟诗, 张彰, 周涛, 胡伏原. 2025. 持续学习研究进展. *中国图象图形学报*, 1-32) [DOI: 10.11834/jig.240661]
- Rannen A, Aljundi R, Blaschko M B and Tuytelaars T. 2017. Encoder based lifelong learning//Proceedings of the IEEE international conference on computer vision: 1320-1328 [DOI: 10.1109/ICCV.2017.3307819]
- Liu X, Wu C, Menta M, Herranz L, Raducanu B, Bagdanov A. D, Jui S and de Weijer, J. V. 2020. Generative feature replay for class-incremental learning//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops: 226-227 [DOI: 10.48550/arXiv.2004.09199]
- Yan H, Wang L, Ma K and Zhong Y. 2024. Orchestrate latent expertise: advancing online continual learning with multi-level supervision and reverse self-distillation//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA: IEEE: 23670-23680 [DOI: 10.1109/cvpr52733.2024.02234]
- Wang X, Chen G, Qian G, Gao P, Wei X Y, Wang Y, Tian Y and Gao W. 2023. Large-scale multi-modal pre-trained models: A comprehensive survey. *Machine Intelligence Research*, 20(4): 447-482 [DOI: 10.1007/s11633-022-1410-8]
- Lu Y, Zhou X Y, Zhang S Z, Liang G Q, Xing Y H, Chen D and Zhang Y N. 2025. A Survey of Continual Learning Methods Based on Pre-training[J/OL]. *Computer Engineering* (路悦, 周翔宇, 张世周, 梁国强, 邢颖慧, 程德, 张艳宁. 2025. 基于预训练的持续学习方法综述[J/OL]. *计算机工程* [DOI: 10.19678/j.issn.1000-3428.0070575]
- Yuan L, Ma Y, Su L and Wang Z. 2023. Peer-to-peer federated continual learning for naturalistic driving action recognition//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition: 5250-5259 [DOI: 10.48550/arXiv.2304.07421]
- Fanan M, Dalle Pezze D, Efatinasab E, Carli R, Rampazzo M and Susto G. A. 2025. Continual Learning for Behavior-based Driver Identification. *Engineering Applications of Artificial Intelligence*, 150: 110459 [DOI: 10.48550/arXiv.2412.10780]

- Ren L, Dong J, Liu S, Zhang L, and Wang L. 2024. Embodied intelligence toward future smart manufacturing in the era of AI foundation model: 1-11 [DOI: 10.1109/TMECH.2024.3456250]
- Marques I, Sousa J, Sá B, Costa D, Sousa P, Pereira S, Santos A, Lima C, Hammerschmidt N, Pinto S & Gomes T. 2022. Microphone array for speaker localization and identification in shared autonomous vehicles. *Electronics*, 11(5): 766 [DOI: 10.3390/electronics11050766]
- Lu J., Zheng X., Tang L., Zhang T., Sheng Q. Z., Wang C., Jin J, Yu S and Zhou W. 2021. Can steering wheel detect your driving fatigue? . *IEEE Transactions on Vehicular Technology*, 70 (6) : 5537-5550 [DOI: 10.1109/TVT.2021.3072936]
- Sebastian P, Narayana P P, Jamshir M and Vijayakumar S. 2025. π -Model Capacitive Sensor for Steering Wheel Hand Position Detection. *IEEE Sensors Letters*, 9 (8) : 2475-1472 [DOI: 10.1109/LSENS.2025.3583753]
- Wang X, Xu R, Zhang S, Zhuang Y and Wang Y. 2022. Driver distraction detection based on vehicle dynamics using naturalistic driving data. *Transportation research part C: emerging technologies*, 136: 103561 [DOI: 10.1016/j.trc.2022.103561]
- Liu L, Wang Z and Qiu S. 2020. Driving behavior tracking and recognition based on multisensors data fusion. *IEEE Sensors Journal*, 20 (18): 10811-10823 [DOI: 10.1109/JSEN.2020.2995401]
- Guan D, Wang Q, Wang K and Song X. 2025. Fatigue and Distracted Driving Recognition Method Based on Multimodal Information Fusion. *IEEE Access*, 13: 83815-83827 [DOI: 10.1109/ACCESS.2025.3568625]
- Xu W, Feng L, Ma J. 2022. Understanding the domain of driving distraction with knowledge graphs. *PLoS one*, 17 (12) : e0278822 [DOI: 10.1371/journal.pone.0278822]
- Peng K, Roitberg A, Yang K, Zhang J and Stiefelhagen R. 2022. TransDARC: Transformer-based driver activity recognition with latent space feature calibration//2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) : 278-285 [DOI: 10.1109/IROS47612.2022.9981445]
- Lin Z, Qiu T, Liu P, Zhang L, Zhang S and Mu Z. 2021. Fatigue driving recognition based on deep learning and graph neural network. *Biomedical Signal Processing and Control*, 68: 102598 [DOI: 10.1016/j.bspc.2021.102598]
- Li K, Li X, Wang Y, He Y, Wang Y, Wang L and Qiao Y. 2024. Vid-eomamba: State space model for efficient video understanding//European conference on computer vision: 237-255 [DOI: 10.1007/978-3-031-73347-5_14]
- Wharton Z, Behera A, Liu Y and Bessis N. 2021. Coarse temporal attention network (cta-net) for driver's activity recognition//Proceedings of the IEEE/CVF winter conference on applications of computer vision: 1279-1289 [DOI: 10.48550/arXiv.2101.06636]
- Zhao S C, Feng Y F, Zhang Z C, Sun B, Zhang S P, Gao Y, Yang J F, Liu M, Yao H X and Wang Y N. 2025. Research advance-ments on emotionally and intellectually integrated digital humans and robotics. *Journal of Image and Graphics*, 30(6): 2139-2160 (赵思成, 丰一帆, 张知诚, 孙斌, 张盛平, 高跃, 杨巨峰, 刘敏, 姚鸿勋, 王耀南. 2025. 情智兼备数字人与机器人研究进展. *中国图象图形学报*, 30(6): 2139-2160) [DOI: 10.11834/jig.240780]
- Wang P W, Xu J H, He X Z, Wang S M. 2025. AI Digital Human Experiment System for Intelligent Transportation Scenario Applications. *Research and Exploration in Laboratory*, 44(9): 1-5 (王庞伟, 徐京辉, 何昕泽, 王思淼. 2025. 面向智慧交通场景应用的AI数字人实验系统. *实验室研究与探索*, 44(9): 1-5 [EB/OL]. [2025-09-06]. <https://link.cnki.net/urlid/31.1707.T.20250901.1651.059>
- Marafie Z, Lin K J, Wang D, Lyu H, Liu Y, Meng Y and Ma J. 2021. AutoCoach: an intelligent driver behavior feedback agent with personality-based driver models. *Electronics*, 10(11): 1361 [DOI: 10.3390/electronics10111361]
- Huang L, Yu W, Ma W, Zhong W, Feng Z, Wang H, Chen Q, Peng W, Feng X, Qin B and Liu, T. 2025. A survey on hallucination in large language models: Principles, taxonomy, challenges, and open questions. *ACM Transactions on Information Systems*, 43(2): 1-55 [DOI: 10.1145/3703155]
- Kokciyan N, Srivastava B, Huhns M N and Singh M P. 2021. Sociotechnical perspectives on AI ethics and accountability. *IEEE Internet Computing*, 25(6), 5-6 [DOI: 10.1109/MIC.2021.3117611]
- Wang F Y. 2022. Parallel intelligence in metaverses: Welcome to Hanoi! . *IEEE Intelligent Systems*, 37(1): 16-20 [DOI: 10.1109/MIS.2022.3154541]
- Li Y F, Huang X, Zhang J, Li S, Yu G Z, Wang Z Y, Wang P W, Hu J, Wang J F, Duan X T, Gong S Y, Tian Y. 2025. Comprehensive review on smart highway evolution and research 2025: Autonomous driving enabled by cooperative vehicle-infrastructure, 42(10): 112-114 (李永福, 黄鑫, 张健, 李深, 余贵珍, 王章宇, 王庞伟, 胡筋, 王江锋, 段续庭, 龚思远, 田野. 2025. 智慧公路发展与研究大综述(2025): 车路协同自动驾驶. *公路交通科技*, 42(10): 112-144) [DOI: 10.3969/j. issn.1002-0268.2025.10.005]

作者简介

蹇木伟, 通信作者, 1982年生, 男, 教授, 主要研究方向为计算机视觉。E-mail: jianmuweihk@163.com

凌钰坤, 男, 硕士研究生, 主要研究方向为计算机视觉。E-mail: 232115005@mail.sdufe.edu.cn

张昊然, 女, 博士研究生, 主要研究方向计算机视觉。E-mail: haoranzhang@mail.sdufe.edu.cn

张琳松, 男, 博士研究生, 主要研究方向为计算机视觉。E-mail: 212115031@mail.sdufe.edu.cn

马嘉骏, 男, 硕士研究生, 主要研究方向为计算机视觉。E-mail: majiajun@mail.sdufe.edu.cn