

## 具身智能驾驶：概念、方法、现状与展望

沈甜雨<sup>1</sup>, 李志伟<sup>1</sup>, 范丽丽<sup>2</sup>, 张庭祯<sup>1</sup>, 唐丹丹<sup>3</sup>, 周美华<sup>4</sup>, 刘华平<sup>5</sup>, 王坤峰<sup>1</sup>

(1. 北京化工大学信息科学与技术学院, 北京 100029;

2. 北京理工大学信息与电子学院, 北京 100081;

3. 燕山大学信息科学与工程学院, 河北 秦皇岛 066000;

4. 北京同仁医院眼科研究所, 北京 100730;

5. 清华大学计算机科学与技术系, 北京 100084)

**摘要:** 具身智能突破了传统人工智能的界限, 强调了机器与物理世界交互的重要性, 旨在通过促进软硬件一体化智能体的环境适应性学习和智能行为的演化, 解决更多智能系统在现实应用中的问题。在这一理念的启发下, 提出了具身智能驾驶的概念与框架, 将具身智能思想融入自动驾驶汽车的开发与应用中, 通过物理智能体、虚拟智能体和真实交通场景之间的持续交互, 实现智能系统在复杂交通场景中的精准感知、高效执行、自主进化, 提升自动驾驶车辆对复杂开放交通环境的长期适应能力。基于提出的具身智能驾驶框架, 进一步总结了其中涉及的核心技术, 并剖析了相关技术的发展现状与存在问题。此外, 通过探索虚实互动数据智能、基础模型与基础智能、持续学习与平行智能在其中的重要作用和应用潜力, 提出了对这一领域的思考与展望。旨在推进具身智能驾驶的系列创新性研究和自动驾驶技术在更广泛场景的应用落地, 并为智能车辆等移动机器人系统的发展, 提供新的思路和解决方案。

**关键词:** 具身智能; 自动驾驶; 具身感知; 具身执行; 具身进化

**中图分类号:** TP39

**文献标志码:** A

**doi:** 10.11959/j.issn.2096-6652.202404

## Embodied intelligent driving: concept, methods, the state of the art and beyond

SHEN Tianyu<sup>1</sup>, LI Zhiwei<sup>1</sup>, FAN Lili<sup>2</sup>, ZHANG Tingzhen<sup>1</sup>, TANG Dandan<sup>3</sup>, ZHOU Meihua<sup>4</sup>,  
LIU Huaping<sup>5</sup>, WANG Kunfeng<sup>1</sup>

1. College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

2. School of Information and Electronics, Beijing Institute of Technology, Beijing 100081, China

3. School of Information Science and Engineering, Yanshan University, Qinhuangdao 066000, China

4. Beijing Institute of Ophthalmology, Beijing Tongren Eye Center, Beijing 100730, China

5. Department of Computer Science and Technology, Tsinghua University, Beijing 100084, China

**Abstract:** Embodied intelligence transcends the boundaries of traditional artificial intelligence by emphasizing the importance of interaction between machines and the physical world, facilitating the development of intelligent entities that combine hardware and software to learn from and adapt to their environments, thereby solving real-world problems. Inspired by this philosophy, the concept and framework of embodied intelligent driving are introduced, aiming at integrating the idea of embodied intelligence into the development and application of autonomous vehicles. Through the continuous interaction between physical agents, virtual agents, and real traffic scenes, intelligent driving systems can achieve precise per-

收稿日期: 2024-02-15; 修回日期: 2024-03-03

通信作者: 李志伟, lizw@buct.edu.cn; 王坤峰, wangkf@buct.edu.cn

基金项目: 国家自然科学基金 (No.62076020; No.62120106005)

**Foundation Items:** The National Natural Science Foundation of China (No.62076020, No.62120106005)

ception, efficient execution, and autonomous evolution in complex scenes, enhancing the long-term adaptability of autonomous vehicles in open traffic environments. Based on the embodied intelligent driving framework, the relevant technologies are summarize and the development status and existing problems of such technologies are analyzed. Furthermore, thoughts and prospects in this field are demonstrated by exploring the important roles and application potential of virtual-real interactive data intelligence, foundation models and foundation intelligence, continuous learning and parallel intelligence. This paper is expected to promote innovative research and the application on embodied intelligent driving in a wider range of scenarios, and provide new ideas and solutions for the development of mobile robot systems such as intelligent vehicles.

**Key words:** embodied intelligence, autonomous driving, embodied perception, embodied execution, embodied evolution

## 0 引言

具身智能旨在基于机器与物理世界的交互, 构建软硬件结合、可自主学习进化的智能体, 其智能行为可以被具有对应形态的智能体通过适应环境的方式学习, 从而帮助解决更多现实世界的问题。当前, 以 ChatGPT 为代表的大模型和基础模型的巨大成功, 标志着人工智能 (artificial intelligence, AI) 研究的一个重大转变<sup>[1-3]</sup>。然而, 智能体必须拥有“身体”才能由虚拟世界进入真实世界, 并通过与真实世界的交互进化出智能。在此背景下, “具身智能”引发了新的研究和技术浪潮, 促使了 AI 研究从纯数字世界的存在跨越为在物理环境中主动作用的新范式。

具身智能已在众多领域取得了一定应用, 涵盖了教育、医疗、交通等不同行业<sup>[4]</sup>。然而, 现阶段的应用主要集中于仿生机器人, 这些机器人通过模拟生物形态与环境互动, 能够完成特定的任务<sup>[5]</sup>。这一趋势反映出当前研究对物理形态与环境互动的深刻理解, 以及如何通过这种互动促进智能行为的学习与适应。如 Gupta 等<sup>[6]</sup>通过介绍深度进化强化学习框架, 展示了环境复杂性、形态智能及其控制学习能力之间的关系, 强调了形态对于学习新任务的重要性。Howard 等<sup>[7]</sup>提出了多级进化概念, 进一步拓展了从材料到机器人的设计过程, 强调了材料和结构在适应环境任务中的作用。Jin 等<sup>[8]</sup>通过微型机器人的设计, 展示了如何利用具身智能处理环境污染问题, 再次证明了智能体与环境互动的重要性。在当前的实际应用层面上, 具身智能主要指将传统智能技术 (如计算机视觉、自然语言处理、强化学习) 集成在智能体中, 以帮助机器人解决特定环境中的各种任务, 即人工智能与机器人技术的结合。

自动驾驶技术指通过人工智能算法和传感器设备使汽车等交通工具在无须人类干预的情况下实现自主感知、决策和执行, 是人工智能应用的重要领域之一, 也是现代交通智能化发展的重要方向。目前, 自动驾驶技术已经被成功应用于工厂内部工序间的搬运、物流系统、矿区运输等封闭和规则环境<sup>[9-11]</sup>, 在特定条件下的应用已成为现实<sup>[12-13]</sup>。然而, 在复杂的开放交通场景中, 自动驾驶系统的安全性仍面临着巨大挑战。主要原因在于开放交通场景存在数据形式多模态、数据质量不可控、数据分布长尾化、数据特性动态漂移等特点, 难以保证智能驾驶系统对各类复杂环境的长期适应能力, 自动驾驶汽车距离成为交通主流还很远<sup>[14-15]</sup>。

本文认为具身智能为以智能车为代表的移动机器人智能系统提供了新的研究思路和解决方案<sup>[7,16]</sup>。一方面, 具身智能强调智能体与复杂环境之间的动态交互, 通过场景综合感知、路径规划、实时控制与决策等能力, 精准、高效地理解周围环境, 并与之互动, 从而实现安全可靠的智能驾驶; 另一方面, 具身智能涵盖了对智能体持续自主进化能力的需求, 通过物理智能体、虚拟智能体、真实环境之间的持续交互, 不断提升智能驾驶系统对复杂环境的长期适应能力。这不仅展现了具身智能在实现高度自主和安全驾驶方面的巨大潜力, 也为交通系统的智能化和自动化开辟了新的路径。

基于以上分析, 本文立足于具身智能的核心思想和自动驾驶的技术需求, 提出具身智能驾驶的概念与框架。该概念着眼于如何将具身智能技术融入自动驾驶汽车的开发与应用中, 旨在通过技术融合, 提升自动驾驶智能系统的泛化性、鲁棒性、自主进化性。具身智能驾驶通过集成先进的传感器、执行部件和智能算法, 将图像信息、深度信息、语言信息和音频信息等多模态数据整合在一个通用智

能体中，完成综合的学习任务<sup>[1]</sup>，实现智能系统在复杂交通场景中的精准感知、高效执行、自主进化，提升自动驾驶车辆对开放环境的长期适应能力。

基于本文提出的具身智能驾驶的概念与框架，本文进一步阐述了其中涉及的核心技术，并对当前相关技术的发展现状与存在问题进行了深度剖析，随后提出了对这一领域的思考与展望，探索了虚实互动数据智能、基础智能、平行智能在其中的重要作用和应用潜力，旨在推进具身智能驾驶的系列创新性研究和自动驾驶技术在更广泛场景的应用落地，并为相关领域学者和从业人员提供一定的借鉴和思路。

## 1 具身智能驾驶概念

### 1.1 具身智能

具身智能的概念最早由图灵在其论著 *Computing Machinery and Intelligence*<sup>[17]</sup>中提出，其探索和发展跨越了一个世纪。从20世纪早期计算机科学的兴起开始，人工智能的概念逐渐形成。最初，AI的研究重点是如何通过软件程序实现逻辑推理、知识表示和问题解决等智能行为，这个阶段的AI研究试图在缺乏物理身体的情况下模拟人类的思考过程。然而，这种方法忽略了身体和感官经验在智能发展中的关键作用。

随着认知科学和机器人学的进步，20世纪的后半叶，人们开始了对智能本质的重新思考。研究者逐渐认识到智能行为并不仅源自大脑的信息处理，还植根于智能体的物理形态以及它与环境的互动中。这一时期，通过构建能够感知和在物理世界中行动的机器人，智能行为被理解为身体和环境相互作用的产物。

到了21世纪，具身认知理论在认知科学领域获得了广泛认同，智能被视为大脑、身体及环境三者相互作用的结果。这一理念促使研究者设计出能够自主学习和适应环境的高度复杂智能系统，开启了具身智能的新纪元。在2023年全球半导体大会上，NVIDIA创始人及首席执行官黄仁勋主持了一次引人深思的讨论，预测具身智能将成为人工智能领域的新前沿。此观点迅速激发了社会各界的广泛关注与科技领域的深度探讨。黄仁勋将具身智能定义为可以理解环境、进行推理并能与之互动的智能系统。斯坦福大学教授李飞飞进一步深化了这一概

念，指出具身智能的核心不仅仅体现在其物理实体上，更重要的是它与环境之间的动态交互，以及通过这种交互所展现的能力与需求。

多学科技术的快速发展，特别是在传感器技术、计算能力和机器学习算法方面的进步，极大地推动了具身智能的研究与应用。具身智能已经从实验室走向实际应用，覆盖了从自主移动机器人到智能家居系统，再到服务机器人和自动驾驶汽车等多个领域。特别是在自动驾驶领域，具身智能理念的应用不仅展示了其解决复杂实际问题的巨大潜力，也预示着它对交通生态和社会生活方式的根本性改变。

自动驾驶技术的发展，作为具身智能应用的典范，展现了智能系统如何深入理解其所处的物理世界并与之互动<sup>[7,14-15]</sup>。通过集成多种先进的传感器和智能算法，自动驾驶汽车能够理解复杂的交通环境、预测其他交通参与者的行为，并作出即时的决策，实现安全高效的行驶。这种技术进步不仅提升了道路的安全性，减少了交通事故，还为未来的城市交通系统提供了新的设计思路。

### 1.2 自动驾驶

自动驾驶汽车的发展历程是一段充满挑战和创新的历史，自动驾驶发展历史时间线如图1所示。1478年达·芬奇设计出预编程发条马车的草图<sup>[18]</sup>，预示自动驾驶概念的诞生。1939年，世界上第一辆自动驾驶概念车Futurama在纽约世博会上展出，它是一种由无线电控制的电磁场引导的电动汽车，预示了未来汽车及交通系统的发展方向。同时这一前瞻性设计还提出了自动驾驶和自动高速公路的概念，为智能汽车和智能交通系统的发展奠定了基础。

20世纪80年代，王飞跃教授开始研究移动机器人，并与NIST合作开发了Spider Robot无人车；1998年，王飞跃教授团队研发了VISTA自动驾驶车辆，在亚利桑那州的沙漠完成了VISTA Car的实验，并在凤凰城51号高速公路上进行了公众演示。在解决有限资源、满足无限性能需求这个自动驾驶难题的过程中，代理(agent)技术的出现促发了“平行驾驶”理论的雏形，并于2005年由王飞跃教授在西安举行的国际汽车电子与安全会议上正式提出。平行驾驶利用agent技术<sup>[19]</sup>，通过网络虚拟“驾驶员”辅助实际驾驶，将“简单决策留车内，复杂计算放车外”，优化资源利用。该理念在



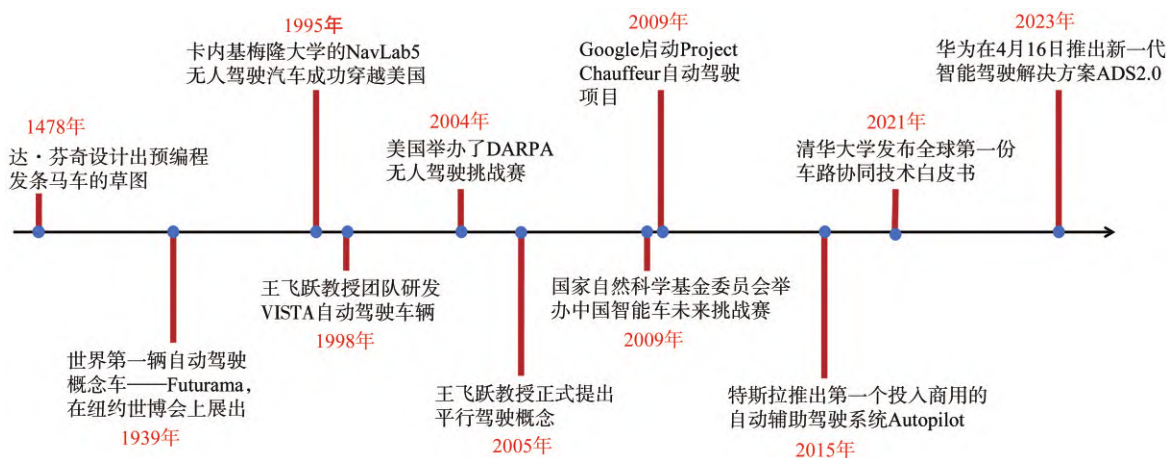


图1 自动驾驶发展历史时间线

VISTA 研制和 DVPG 设计中均有所应用，并在国家 863 计划项目中实施，促进了国内无人车技术的巨大发展。1995 年，卡内基梅隆大学的 NavLab5 无人驾驶汽车成功穿越美国，行程 160.9344 千米。这次测试展示了自动驾驶技术在复杂道路条件下的可行性，为未来智能汽车和智能交通系统的发展提供了有力支持。这一成就不仅凸显了自动驾驶技术的前景，也为汽车产业的变革拉开了序幕。

进入 21 世纪，自动驾驶技术得到了更加飞速的发展。2004 年，美国国防高级研究计划署（Defense Advanced Research Projects Agency, DARPA）举办了 DARPA 无人驾驶挑战赛，激发了全球对自动驾驶技术的关注和研究。国家自然科学基金委员会从 2009 年开始，每年举办 1 届中国智能车未来挑战赛<sup>[20]</sup>，通过设计开放环境下的无人驾驶测试道路，为无人车在真实道路环境中学习、训练提供了很好的平台。为了让无人驾驶的测试更专业、更体系化，2015 年 8 月，常熟市政府联合西安交通大学、中国科学院自动化研究所、长安大学和青岛智能产业技术研究院，在常熟高新技术产业开发区共建中国智能车综合技术研发与测试中心，由此实现了国内无人驾驶测试中心零的突破。

2009 年，Google 启动 Project Chauffeur 自动驾驶项目，利用激光雷达等技术实现完全自动驾驶。通过追踪路上的行人、车辆和自行车等物体，车辆能理解交通信号灯的变化。同年年底，改装版 Prius 成功完成了 4 828.032 千米无须人工接管的自动驾驶，为智能交通系统发展奠定了基础。2015 年，特斯拉推出第一个投入商用的自动辅助驾驶系统 Autopilot，允许车辆在特定条件下实现自动驾驶，

包括自动车道保持、自动转向和自动泊车等功能。百度也完成了开放高速公路自动驾驶测试，展示了自动驾驶技术的实际应用能力。2021 年，清华大学发布全球第一份车路协同技术白皮书《面向自动驾驶的车路协同关键技术与展望》，为自动驾驶技术的发展提供了新的思路。2023 年 4 月 16 日，华为技术有限公司推出新一代智能驾驶解决方案 ADS2.0，其先进的网络架构和工程化设计为自动驾驶领域树立了新的标杆。

衡量自动驾驶的自动化程度对于了解和管理自动驾驶技术的发展至关重要。通过对自动驾驶系统的自动化水平进行评估和分类，可以更清晰地了解车辆在不同情况下的自主性和自动化能力，从而为技术的应用和发展提供更有力的支持。自动驾驶的自动化程度可通过系统处理不确定性能力来衡量，2014 年，SAE International<sup>[18]</sup>定义了从 L0 到 L5 的 6 级自动驾驶标准，其中 L0~L5 级分别代表了不同的自动化程度。L0 级车辆完全由驾驶员手动驾驶，L1 级车辆可完成单一维度驾驶任务，L2 级车辆可同时实现多维度辅助，L3 级车辆可在特定环境下实现自动驾驶，L4 级车辆在限定条件下无须驾驶员接管方向盘，L5 级车辆可在任何条件下完全自动行驶，完全替代人类驾驶员。自动驾驶等级示意如图 2 所示。自动驾驶技术的进化，从最初的辅助驾驶迈向完全自动化的过程，体现了具身智能的核心理念。

### 1.3 具身智能驾驶

具身智能驾驶旨在实现高级别的自动驾驶。高级自动驾驶是指高级自动化的自动驾驶系统，车辆能够在不同道路和环境条件下自主行驶。同时车辆

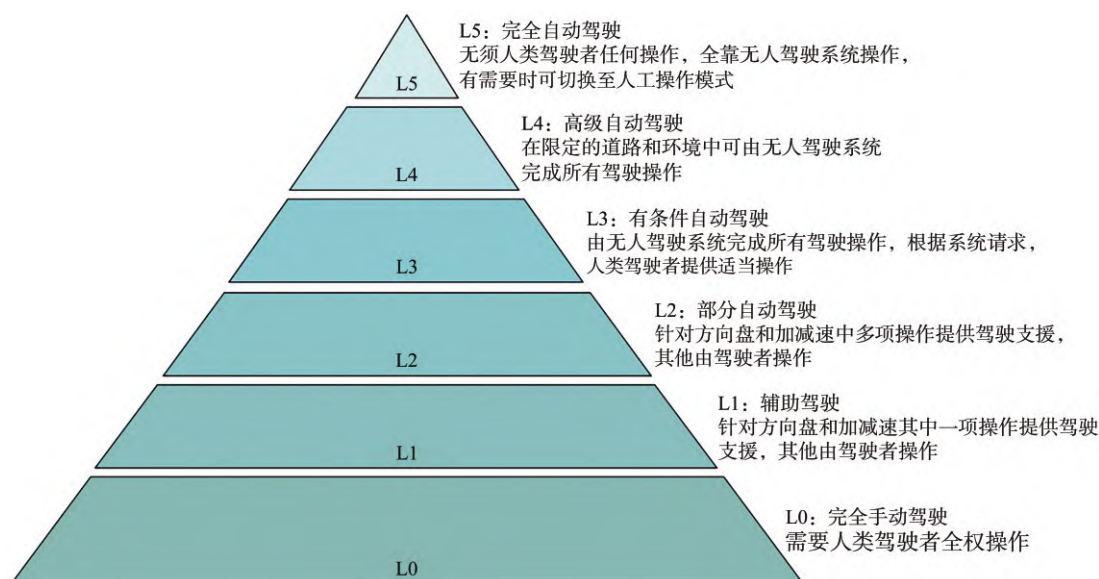


图2 自动驾驶等级示意

具备全面的感知、决策和控制能力，能够应对各种复杂的交通情况和突发事件，实现高度智能化和自主化的驾驶，具身智能驾驶分层概念如图3所示。

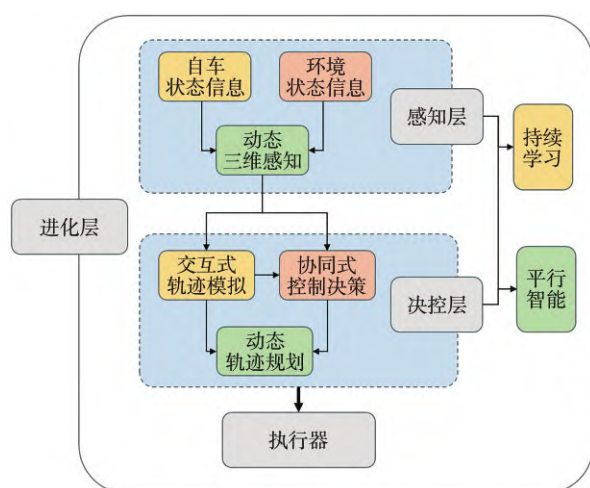


图3 具身智能驾驶分层概念

实现高级别自动驾驶需要综合运用先进的传感技术、人工智能技术等多种技术手段。实现具身自动驾驶技术可从以下3个方面考虑。

1) 高度泛化的动态环境全感知能力。车辆需要搭载多种传感器（如激光雷达、摄像头、毫米波雷达等）实时获取周围环境的信息。通过获取的数据，车辆可以对道路、车辆、行人等各种元素进行准确识别和跟踪，实现全面的环境感知。不同于传统的环境感知技术，具身智能感知技术可通过与环

境实时交互，在不同的光照条件、天气条件等情况下选择不同的感知策略，以最大化利用不同传感器和感知算法的优势。

2) 高度可靠的决策规划和控制执行能力。通过人工智能技术，车辆可以对感知到的环境信息进行分析 and 处理，作出合理的驾驶决策，并实时控制车辆的加速、转向、制动等动作，确保车辆按照规划路径安全行驶。车辆需要能够预测其他车辆和行人的行为，规划最优的行驶路径，并实时调整驾驶策略以应对不同的交通情况。同时，车辆还需要具备自主避障和紧急制动等功能，能够在遇到突发情况时，及时作出反应，保障行车安全。具身智能技术通过与多感官环境信息的交互（如视觉感知信息、震动触觉信息、语音听觉信息等）综合考虑多种信息，实现具身智能决策、规划和自主执行控制策略。

3) 高度智能的自主能力。为了实现长期自主学习，自动驾驶车辆需要具备持续进化的学习和适应能力。这可以通过增量学习和持续学习等技术实现。车辆在与环境交互的过程中，不断收集新的数据和经验，通过学习和适应，提高自身的性能和智能水平。此外，通过物理智能体、虚拟智能体和真实环境之间的持续交互，自动驾驶车辆可以不断地在各种情境下学习和实践，提高其对复杂环境的适应能力。结合平行智能与数字孪生的相关思想，可以构建一个多维度、多层次的智能驾驶系统，使其能够在虚拟环境中进行大量的学习和训练，提高其



对复杂环境的适应能力。

## 2 具身智能驾驶核心框架

不妨把具身智能驾驶技术的发展想象成一个人从婴儿到长大成人的过程。婴儿从刚开始只能感知周围的世界到可以通过自身与环境间的交互决定自己下一步的行动,再到从交互中逐渐成长,最终长大成人。因此,本文将具身智能驾驶系统核心框架分为3个核心模块,分别是具身感知、具身执行和具身进化,如图4所示。

### 2.1 智能驾驶具身感知

具身感知是实现具身智能驾驶的基础。与基于计算机视觉的感知不同,智能驾驶具身感知应具有多模态、动态和自适应的特点,以应对多样、不确定和复杂的环境。其中的核心技术主要包括以下3个方面。

#### 2.1.1 多模态信息融合

具身感知强调感知不仅依赖于视觉、听觉等传感器的输入,还依赖于运动、触觉等身体的输出,因此对于智能驾驶具身感知而言,多传感器集成、

多源信息融合与多模态学习方法是不可或缺的。目前自动驾驶领域中已经涌现了大量的多模态信息融合方法<sup>[21-23]</sup>,通过不同传感器采集的数据信息互补以实现更稳定、更安全的感知。多模态融合框架示例如图5所示。但是现有的多模态信息融合只涉及相机、激光雷达和毫米波雷达等数据,缺乏对环境状态信息与听觉信息等对齐与融合。直接将具身数据加入现有算法中,不仅会面临融合检测速率下降的问题,数据对齐的困难还可能导致融合检测精确度降低。因此,亟须探求一种新的多模态融合方法,以实现视觉、听觉、触觉等具身数据的全方位有效融合。

#### 2.1.2 动态三维感知

婴儿在探索世界时往往在不断运动中逐渐感知周围环境和细节,具身智能驾驶同样如此。由于车辆在绝大部分时间都处于行驶过程中,所以想实现智能驾驶具身感知,需要车辆具备对三维场景中的运动物体进行实时检测、跟踪、识别和重建的能力。而在传统自动驾驶领域中,实现高质量动态三维感知的难点就在于如何处理运动物体的遮挡、形

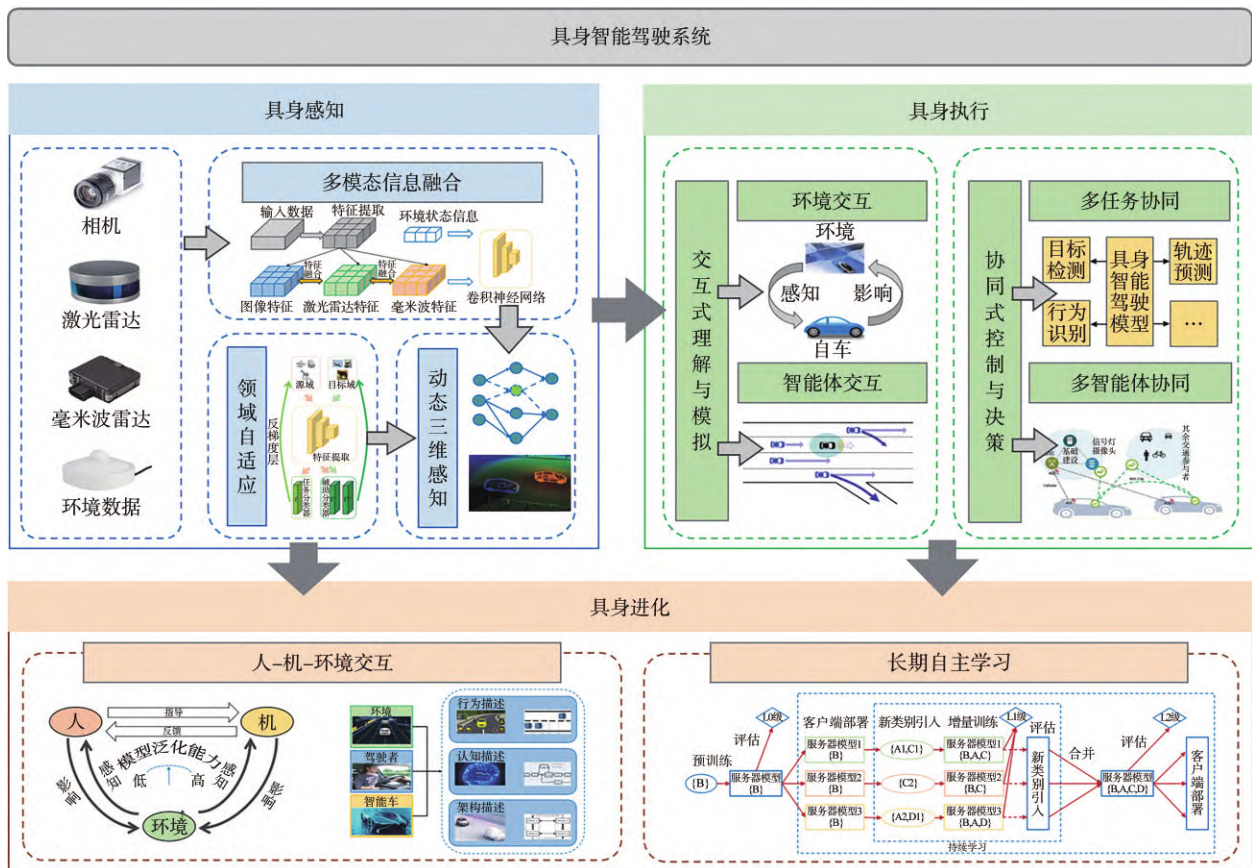
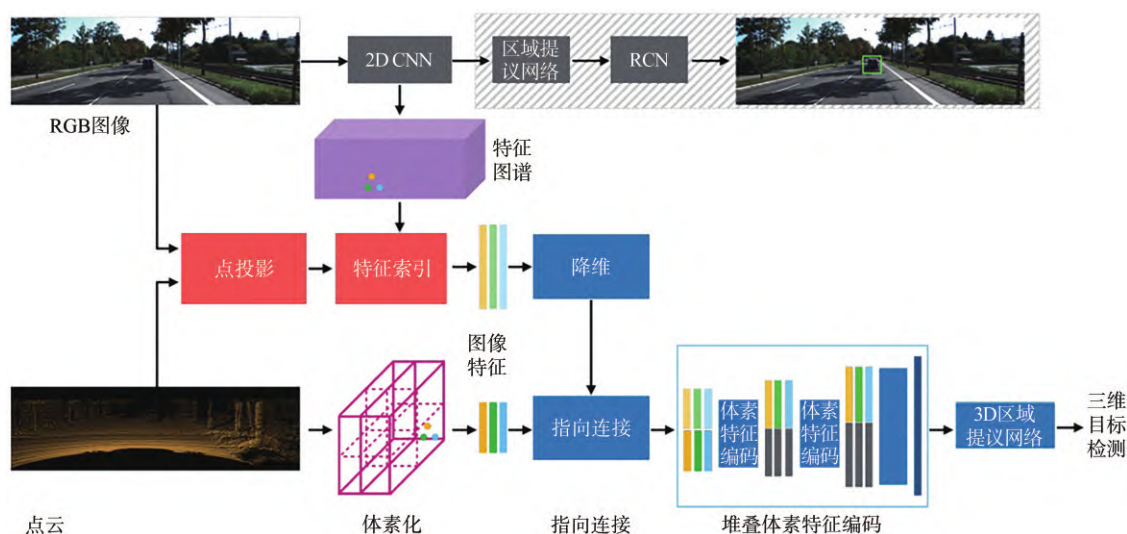


图4 具身智能驾驶系统核心框架

图5 多模态融合框架示例<sup>[24]</sup>

变、光照变化等问题<sup>[25]</sup>，而对于智能驾驶具身感知，有了光照变化、振动等环境状态数据的加入，便可以较为简单地解决这类问题。与此同时，为了保证运算速率和检测精度的前提下处理如此多模态且大量的数据，智能驾驶具身感知需要具备根据环境状态信息动态切换三维感知模式的能力，如在光照条件异常的情况下，对相机数据进行一定的滤除，更多采信雷达数据；当在高速公路上且车辆较为稀少时，对近处的识别进行一定程度的简化，只需检测是否存在障碍物等。

### 2.1.3 领域自适应

具身智能驾驶系统需要在不同光照、天气等条件下以及不同复杂度的场景中实现稳定可靠的感知。领域自适应可以帮助其利用已有的数据和模型，适应不同的域分布，提高系统泛化能力和安全性。目前，已有的领域适应方法可分为无监督领域适应<sup>[26-27]</sup>和半监督领域自适应<sup>[28]</sup>两类，其中无监督领域自适应假设目标域没有标签数据，只能通过对齐源域和目标域的特征分布或输出分布来实现迁移。半监督领域自适应假设目标域有少量的标签数据，可以通过标签数据来指导迁移过程，或者通过自学习或主动学习等技术来扩充标签数据。由于目前可用于智能驾驶具身感知的数据极其稀少且标注较为困难，所以寻找可行的领域自适应方法以提升具身智能驾驶系统的鲁棒性和场景适配性尤为重要。

## 2.2 智能驾驶具身执行

具身执行是具身智能驾驶的核心部分，也是具身智能驾驶由概念到落地所需迈出的重要一步。智

能驾驶具身执行由传统智能驾驶控制与决策演化而来，并对控制过程中的交互性和协同性提出了更高要求。

### 2.2.1 交互式理解与模拟

具身智能强调从感知周围环境对环境进行建模，进而执行系统决策以验证并调整模型。而在具身智能驾驶中，由于交通场景中不仅需要考虑路面情况和周围环境，还需要考虑其他交通参与者（如车辆、行人、骑行者等），这些交通参与者的行为不仅取决于自身意图，还会受周围环境和参与者的影响。因此，具身智能驾驶模型需要积极与周围环境或其他交通参与者进行交互式理解与模拟，并在理解与模拟中验证并输出自身下一步行动。通过深度神经网络对其他参与者的行为进行模拟，模型在获取预测结果以后便可以如同人类一样在脑海中模拟可能发生的情况，并通过基于规则或数据驱动的风险评估算法<sup>[29-30]</sup>判断是否需要切换人工驾驶模式或采取紧急措施。然而，绝大多数现有的轨迹预测与碰撞预测方法<sup>[31-32]</sup>忽略了交通参与者之间的交互行为的影响，无法满足智能驾驶具身执行的需求，因此需要从具身智能的角度开发新的算法。

### 2.2.2 协同式控制与决策

协同自动驾驶是自动驾驶领域中解决交通拥堵的重要方式，也是具身智能驾驶真正走入现实世界的重要前提。车联协同决策模型示例如图6所示，2020年5月，《道路机动车辆协同自动驾驶相关术语的分类和定义》根据M2M通信对车辆动态自动驾驶任务性能和交通管理的影响，将协同自动驾驶



的协同功能分为以下4种：状态共享、意图共享、协同决策、协同调度。其中，状态共享是指由发送实体将交通环境感知信息和环境交互信息等传递给接收实体使用，具身智能驾驶车辆可以通过状态共享的方式将自身与环境交互得到的消息发送给其他车辆，从而大幅缩减对模型算力和信息存储能力的需求。意图共享是指发送实体将计划未来行动的信息发送给接收实体使用，通过意图共享可以使具身智能车辆在行驶中共享（如车道变更的意图）以及及时提醒周围车辆作出相应反应。协同决策与协同调度则是由车辆之间将协作信息即“我希望你做什么”和“我会按照指示去做什么”相互传递，可以帮助具身智能车辆避免潜在冲突或及时为应急车辆让行。目前，绝大多数协同算法与决策算法仍面临着场景泛化能力不足、严重依赖专家经验等诸多挑战，虽然可以通过类脑学习型决策系统在一定程度上缓解或避免，但仍存在安全性与泛化性不足的问题<sup>[33]</sup>。同时，由于具身智能驾驶系统需要在行驶过程中同时处理目标检测、速度预测、行为识别、碰撞评估等多个任务，因此通过样本梯度相似性<sup>[34]</sup>等方式实现多任务协同学习处理同样十分重要。

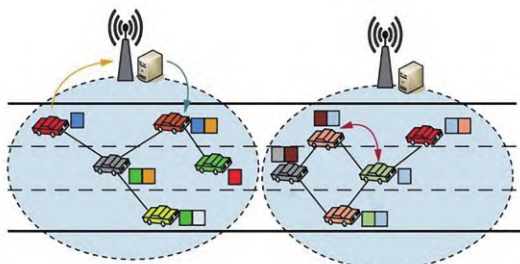


图6 车联协同决策模型示例<sup>[35]</sup>

## 2.3 智能驾驶具身进化

### 2.3.1 人—机—环境交互

人类是最突出的具身智能体，人类的学习成长方式值得我们深入思考，从而为具身智能的进化提供参照，即所谓的具身进化。驾驶车辆时，人们往往需要其他具有丰富驾驶经验的驾驶者来指导其驾驶。在遇到没有见过的地形或路面上有坑洞等路面状况不好的情况时，新手驾驶者往往来不及避让，选择牺牲一定的驾驶舒适性直接行驶过去；在经历过几次之后，大家就会熟练地绕过坑洞或采用其他合理的驾驶方法经过此路段。具身智能驾驶可以通过模仿此形式，在面临模型未训练过的路况时，通

过人为提示和环境交互信息进行判断，在下次遇到此种路况时，便可作出正确的决策，并可通过云端共享给所有具身智能车辆，实现快速进化，从而提高模型的泛化能力。蔚来公司提出的4D舒适领航便有类似的效果，只需要配备4D舒适领航功能的车辆经过一处路段4次，便可使车辆在经过此路段时，自动调节底盘和悬挂，从而提高驾驶舒适性。

### 2.3.2 长期自主学习

清华大学的朱军教授团队指出，持续学习的目的是在保障资源效率的背景下，确保适当的稳定性—可塑性权衡，以及充分的任务内/任务间泛化能力<sup>[36]</sup>。在具身智能驾驶中，由于自车物理结构及功耗的限制，智能车辆的算力极为有限，而对模型的泛化能力的需求又极高，因此在具身智能驾驶模型中引入持续学习变得尤为重要。目前，已有的持续学习方法包括以下5种：基于正则化的方法<sup>[37]</sup>、基于重放的方法<sup>[38]</sup>、基于优化的方法<sup>[39-40]</sup>、基于表征的方法<sup>[41-42]</sup>以及基于架构的方法<sup>[43]</sup>。部分持续学习方法之间的关系如图7所示，尽管这些方法从各个角度试图提升学习的可塑性和任务间的可推广性，然而目前持续学习的模型仍无可避免地面临灾难性遗忘<sup>[44]</sup>的问题，即在学习新知识时，难以保留之前的知识。因此，具身智能驾驶的长期自主学习方法仍需进一步探索与实践。

## 3 现状与问题

### 3.1 具身智能驾驶数据集

在自动驾驶多模态数据集中，KITTI<sup>[46]</sup>、NuScenes<sup>[47]</sup>、Waymo<sup>[48]</sup>和Argoverse<sup>[49]</sup>因其大规模的驾驶场景和样本，以及高质量的标注信息而被广泛使用。KITTI数据集首次采用前视相机、激光雷达和惯性传感器作为传感器，支持目标检测、目标跟踪、深度估计等多种任务。其中相机具有很高的鲁棒性，但受光线和天气影响，当光线弱、逆光、下雨时，相机图像缺乏RGB信息来观察物体。在相应的场景中，激光雷达获取的空间信息非常紧密，弥补了相机的缺点。然而，在物体较多的场景中，激光雷达无法有效区分重叠或相对较近的物体。惯性传感器主要检测和测量加速度、倾斜、冲击、振动、旋转和多自由度运动，是导航、定向和运动载体控制的重要部件。在NuScenes、Waymo、Argoverse等数据集的数据获取中，首次配备了环视相机数据。全方位视野、增强安全性、精准位置感知、



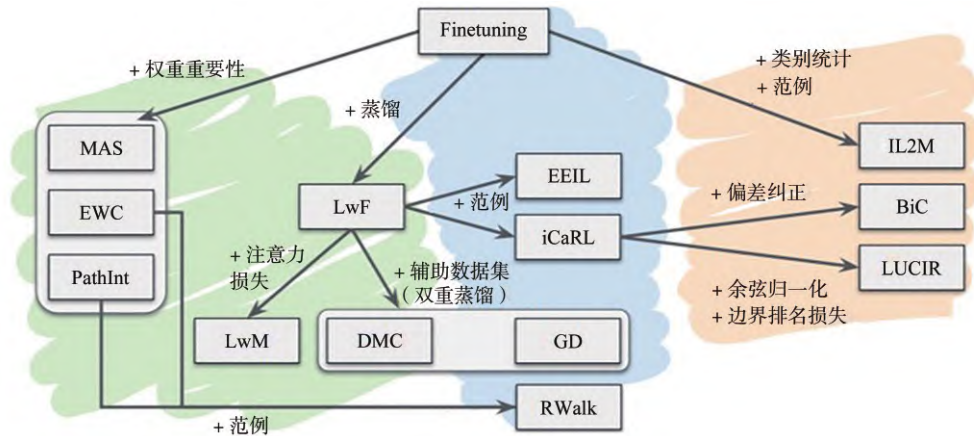


图7 部分持续学习方法之间的关系<sup>[45]</sup>

增强自动驾驶功能以及降低成本等特点，使环视相机在自动驾驶系统中扮演着不可或缺的角色，为实现更智能、安全的自动驾驶技术提供了强有力的支持。NuScenes数据集获取还首次配备了地图数据用于决策规划任务。此外车身信号、全球定位仪、惯性测量单元、红外相机、超声波雷达传感器、事件相机传感器等传感器数据也为探索不同类型传感器下感知算法研究以及多模态融合算法研究提供了丰富的数据信息。自动驾驶多模态数据集如表1所示。

以往的研究通常只利用数据集的部分类型数据进行实验，由于得到的信息具有局部性，其在模拟人类综合运用多维度信息进行预测与决策的能力上存在明显不足。人类在感知世界时，能够从多个角度获取信息，综合考虑各种因素作出决策，而传统研究往往只关注单一视角或数据源，限制了模型的全面性和准确性。因此，为了更好地模拟人类感知世界的能力，具身自动驾驶将更加注重整合多源数据，从不同角度综合考虑信息，从而提高预测和决策任务的效果，提升自动驾驶系统的智能水平，使其更加适应复杂多变的交通环境和道路情况。

在具身智能驾驶领域，端到端的学习方法直接

从传感器输入控制命令，省略了传统自动驾驶系统中的多个中间步骤。这种方法需要大量的数据集进行训练，并且数据集的多元性与质量对算法的创新、开发及其性能评估起到决定性作用。这些数据集通常包括从实车传感器（如摄像头、雷达）收集的数据以及相应的驾驶操作（如转向角度、加速度）。自动驾驶端到端数据集如表2所示，CARLA<sup>[53]</sup>、ApolloScape<sup>[54]</sup>、A2D2<sup>[55]</sup>以及NuScenes<sup>[47]</sup>等数据集属于自动驾驶领域经典的端到端数据集。这些数据集不仅涵盖了从模拟环境至真实世界的广阔场景，还囊括了从前视摄像机到360°摄像机、激光雷达、雷达等多种传感器类型，为环境感知提供了丰富的信息源泉。然而，各数据集在传感器配置、覆盖范围及标注精度等方面的差异性，对算法的研发及性能评估等环节会产生显著影响。

具体而言，CARLA主要提供了前视摄像头的视觉数据，同时也包含了基础的车辆动态数据，如速度和加速度信息，通过CAN-bus系统记录。CARLA以其虚拟环境中生成的数据而闻名，为自动驾驶算法提供了无风险的测试环境。然而，它在感知深度和环境理解方面存在一定局限性。

表1 自动驾驶多模态数据集

数据集	年份	传感器类型									
		前视相机	环视相机	激光雷达	雷达	车身信号	全球定位系统	惯性传感器	立体相机	高精度地图	高分辨率相机
KITTI <sup>[46]</sup>	2012	✓		✓			✓	✓			
NuScenes <sup>[47]</sup>	2019	✓	✓	✓	✓		✓			✓	
Argoverse <sup>[49]</sup>	2019		✓	✓						✓	
Waymo <sup>[48]</sup>	2019		✓	✓							
ONCE <sup>[50]</sup>	2021		✓	✓							
S3E <sup>[51]</sup>	2022			✓				✓	✓		
ZOD <sup>[52]</sup>	2023			✓				✓			✓

表 2 自动驾驶端到端数据集

数据集	年份	传感器类型									
		前视相机	环视相机	激光雷达	雷达	车身信号	全球定位系统	惯性传感器	立体相机	高精度地图	高分辨率相机
CARLA <sup>[53]</sup>	2017	✓		✓			✓	✓		✓	
ApolloScape <sup>[54]</sup>	2018	✓		✓			✓	✓		✓	✓
A2D2 <sup>[55]</sup>	2020	✓		✓		✓	✓				
NuScenes <sup>[47]</sup>	2019	✓	✓	✓	✓		✓	✓		✓	

另外, ApolloScape、A2D2 和 NuScenes 提供了更全面的传感器配置, 包括前视摄像头、激光雷达、雷达和全球定位系统等。特别是 NuScenes 数据集还包含了 360° 摄像头, 为环境的全方位感知提供了支持。这些数据集不仅有丰富的视觉信息, 还在物理位置和环境理解方面提供了更为详尽的数据。如 A2D2 和 NuScenes 数据集通过高精度地图支持对环境进行深层理解。然而, 这些数据集的挑战在于处理和分析大规模的多模态数据, 以及在高动态和复杂环境下保持算法的鲁棒性。有效处理和应用这些大规模、高维度的数据集, 需要投入大量的计算资源, 并需要采用先进的数据处理技术。此外, 数据集的多样性与数据丰富性直接关系到模型训练的效率和自动驾驶系统的可靠性, 这对未来数据集的发展提出了更高的要求: 不仅要扩展数据规模和覆盖范围, 更需提升数据类型与数据分布的丰富度。

综上所述, 自动驾驶研究依赖的数据集需求复杂多变, 既包含对理想化模拟环境数据的需求, 也包含真实世界复杂场景数据及高精度多模态感知数据。数据集的发展趋势应聚焦于增强数据真实性与多样性, 优化传感器配置, 并提高数据标注精度, 以促进自动驾驶技术向更高层次发展。

### 3.2 具身智能驾驶基础模型

现有的智能驾驶任务主要包括行人检测、目标跟踪、驾驶行为预测等, 这些任务的数据来源有车载相机、激光雷达以及车辆内部传感器。相机可以提供高分辨率的图像并捕捉更多的细节和目标特征, 从而识别和分类不同类型的物体。但相机的性能易受到光线和恶劣天气的影响, 在夜间或强光照下表现不佳。相较于相机, 激光雷达可以在雨雪等恶劣天气条件下工作并不受光线的影响, 但雷达的分辨率通常较低, 难以提供关于物体形状和细节的详细信息。智能驾驶车辆内部传感器可以实时采集车辆行驶状态下的各种数据, 具有高精度和准确性, 并能在不同的环境条件下工作。因此, 不同模

态的信息各有优劣, 挖掘各个模态之间的映射关系并设计相应的融合策略, 可以弥补单一模态信息的不足, 提高智能驾驶系统的鲁棒性和可靠性。

相机-雷达是一种常见的融合策略, 通常用于 3D 目标检测<sup>[56-60]</sup>和多目标跟踪<sup>[61-66]</sup>任务中。TransFusion<sup>[22]</sup>是一种基于 Transformer 的融合激光雷达和摄像机的 3D 目标检测模型, 该模型细致地进行了粗粒度融合, 并在退化的图像中显示了卓越的鲁棒性。Xie 等<sup>[67]</sup>提出了一种新颖的基于注意力的多阶段多任务融合方法, 利用点级、RoI 级和多任务融合来补充激光雷达和摄像头的缺点, 从而解决了激光雷达和 RGB 图像之间点融合的稀疏性问题。随着 3D 多模态目标检测研究的发展, 基于多模态的多目标跟踪方法也受到广泛学者的关注。Chiu 等<sup>[68]</sup>通过学习如何融合二维图像和三维 LiDAR 点云的特征, 从而捕获对象的外观和几何信息, 并在数据关联中结合马氏距离和特征距离进行跟踪框和检测框的度量。Huang 等<sup>[61]</sup>提出了用于使用 2D 图像和 3D 点云测量的联合对象检测和关联的端到端深度神经网络, 通过设计鲁棒性的相似度计算模块来计算 3D 空间中的遮挡感知外观和运动相似度, 开发了综合数据关联模块, 用于检测置信度、亲和度和开始-结束概率之间的联合优化。以上研究均从跨模态融合的角度来解决智能驾驶技术中的常见问题, 融合后的模型可以综合利用两种类型的传感器数据, 实现车辆对周围环境信息的全面感知。

与相机-雷达之间的 2D-3D 数据融合相比, 对车辆内部传感器 1D 信息的融合研究较少。内部传感器可以提供车辆的动力学信息 (如速度、加速度、转向角等), 这些信息能够实时反映车辆的运动状态且不易受外界复杂环境的干扰, 为智能驾驶系统提供了稳健可靠的先验信息, 有利于提高模型的准确性。不同的车辆动力学信息在不同的驾驶任务中作用不同, 由于速度和转向角信息可以实时反映车辆真实的驾驶状态, 因此常用于驾驶行为预测、变道预测等任务。Shahverdy 等<sup>[69]</sup>通过递推绘

图技术将加速度等车辆信息转化为图像数据，并输入基于图像的神经网络中，预测了驾驶员是否在进行正常驾驶、攻击性驾驶、分心驾驶以及醉酒驾驶等行为，实现了对驾驶员状态更加广泛的分类。在变道任务中，设计实时连续的变道检测系统也是交通安全研究的重要组成部分，Jeong等<sup>[70]</sup>借助神经网络的优势，将车速、加速度、转向角和路况等多种信息融合在一起，使模型能够在短暂的时间内预测出车辆的未来位置，这样的预测不仅有助于自动驾驶系统作出更智能的决策，而且为道路安全提供了可靠的参考。除了前文提到的速度、转向角等常用的车辆动力学信息外，其他车辆1D状态信息（如油门、制动、控制指令等）对一些特定任务也发挥了重要作用，如安全驾驶、驾驶策略等。Yang等<sup>[71]</sup>提出了基于注意力机制的双向自动驾驶网络，并将图像信息、车辆行驶速度以及高级导航命令作为模型的输入，使模型可以更好地预测车辆行驶的路径点，实现了在复杂城市环境下车辆的自动驾驶。Kim等<sup>[72]</sup>在研究自动驾驶控制问题时，不同于传统的仅利用视觉信息，而是将视觉信息与底层控制命令（如制动和转向角）两种跨模态的信息相互映射，最终对车辆未来时刻的转向角、油门值和制动值作出更准确的预测。为了更好地实现用户与智能驾驶系统之间的交互，提高驾驶体验与可操作性，Katuwadeniya等<sup>[73]</sup>在设计共享导航框架的预测和控制循环中引入用户的输入，该方法在Udacity数据集上验证了其有效性，并且能够使用少量的训练数据将学习转移到不同的环境和移动平台中。

综上所述，相较于单一模态信息，多模态信息由于可以提供更为丰富、鲁棒性的多源信息而逐渐成为智能驾驶系统中的热点研究对象。研究人员通过设计准确高效的融合模块为下游任务提供高质量的判别性特征信息。另外，不同于传统的多个阶段任务方法（如感知、定位、路径规划和控制等）每个阶段独立设计模块和算法，端到端方法可以直接学习从传感器数据到车辆控制输出的映射关系，避免了中间阶段的手动设计。然而，仅使用多个模态的信息也存在一定的局限性，因为现有的多模态信息均来源于车辆的自身特征，缺少与周围环境以及其他车辆的交互，从而难以实现高效率的车路协同技术。为解决上述模态单一、缺乏考虑车路协同等问题，本文对基于多模态融合的端到端车路协同系统展开研究，以提高交通效率、优化交通管理并减

少交通事故的发生，从而推动未来智能交通系统的发展。

### 3.3 具身智能驾驶学习模式

传统智能驾驶系统具备的能力在模型训练完成后便已经确定，泛化能力差且后续只能通过对模型迭代更换，实现能力提升。而具身智能驾驶系统可以通过人-机-环境交互以及长期自主学习<sup>[37-43]</sup>的方式逐步提升模型的泛化能力，这也是具身智能驾驶系统同传统智能驾驶系统最大的区别。然而，由于数据的多样性匮乏、长尾场景的缺失与极端场景的采集困难，如何保证模型初始性能的优越性，并为系统的增量式更新与进化提供有价值的数据，仍然是一大挑战。解决方法之一便是利用Li等<sup>[74-75]</sup>提出的平行学习思想，通过人工系统构建与虚实互动数据智能保证模型的初始性能，并利用计算实验与平行执行实现系统的闭环优化与持续进化<sup>[76-77]</sup>。

与传统的机器学习理论不同，平行学习将智能代理和环境视为一个统一的系统，而不是两个对立的系统。通过构建一个与真实系统平行的人工系统，设计了包括描述学习、预测学习和指示学习在内的3个部分的学习模式。其中描述学习是指将知识系统和人工系统之一的数据作为输入，另一个系统的参数作为输出，将它们连接并聚合在一起，将数据呈现为从动作到状态的映射记录，并根据一定的目的，将真实系统的规律提炼为经验，然后用于修正人工系统。预测学习是指从平行系统中编码的先验知识开始，学习系统在简化的理想环境中的探索状态和行动空间，然后基于这一过程产生的理解，预测在真实世界中可能观察到的行为，并执行若干操作。从这些操作记录中，人工系统可以更好地理解真实系统，并根据这些新经验更新自身。指示学习是指通过现实世界中的决策与控制，将想象映射到现实。在平行系统中，经历自我提升过程后，学习系统将掌握采取某些行动后状态如何转变的信息，从而制定出最优策略。指示学习主要关注人工系统学习到的策略是否能在现实世界中被采用。

尽管通过持续学习和平行学习等学习模式，具身智能驾驶系统可以实现快速进化，但仍有一些问题亟待研究。首先是长期持续学习中不可避免的灾难性遗忘问题，现有的算法虽然可以在一定程度上降低模型的遗忘程度，但随着模型的不断成长与增量更新，其所面临的遗忘问题也会逐渐加重，将对模型性能产生较为严重的影响，因此亟须新的持续



学习方法能够改善甚至解决模型的灾难性遗忘问题。其次是有限计算资源与无限开放场景之间的矛盾问题。随着模型与环境及其他交通参与者的不断交互,记忆的交通场景逐渐增多,模型对计算资源的需求越来越大,并最终会超过具身智能车辆所能承载的极限。虽然可以通过云端存储的方式缓解一定的压力,但受现有通信能力的限制,采用云端存储的方式会不可避免地造成系统安全隐患。因此,仍需寻求有效的方式,以减少具身智能驾驶系统对计算资源的需求。

## 4 思考与展望

### 4.1 虚实互动数据智能

虚实互动数据智能利用诸如平行传感<sup>[78]</sup>、平行视觉<sup>[79-80]</sup>和平行雷达<sup>[81]</sup>等各种技术,将数据智能、虚拟现实交互和智能决策技术相结合,实现虚拟世界与现实世界之间的智能连接和交互。这项技术不仅涉及数字信息与物理环境之间的交互,还通过智能分析和决策应用为智能城市和智能制造等领域提供支持和解决方案。

基于此,具身智能驾驶可以实时感知周围环境,包括道路状况、车辆、行人和障碍物,并利用数据智能分析技术处理和理解这些数据,作出智能驾驶决策。在未来发展中,具身智能驾驶的感知能力将得到增强,使系统能够更准确地识别和理解复杂的驾驶环境,从而提高驾驶安全性和可靠性。此外,智能决策和路线规划将进一步优化,使系统能够根据实时交通状况和乘客需求,智能选择最佳驾驶路线,并避免交通拥堵和事故风险。随着机器学习和深度学习技术的发展,具身智能驾驶还将具备自学习和自适应能力,不断提高驾驶技能,并适应不同的驾驶场景和环境变化。最终,这项技术将为城市交通管理、移动服务和汽车行业带来革命性变革,为人们提供更安全、更便捷、更智能的出行体验。

### 4.2 基础模型与基础智能

具有大规模参数和复杂计算结构的机器学习模型被称为基础模型,最近,这些模型的研究引起了广泛关注<sup>[82-83]</sup>。另外,基础智能是指将更深层次的人工智能能力融入大型模型,通过优化模型架构、有效调整参数、丰富数据集等,进一步提升模型的智能水平。这使其能够更好地理解复杂问题,在自主决策、推理和创新方面展示出更高水平的智能。

在未来的研究中,基于基础模型和基础模型智

能的具身智能驾驶可以利用大规模神经网络等技术,通过深度学习算法对大规模数据集进行训练和学习。这使系统能够识别复杂的驾驶场景,理解交通规则和实时环境,并作出智能决策,实现安全高效的自动驾驶。基于基础智能的具身智能驾驶系统具有先进的感知能力、智能决策能力和稳健的适应能力,使其能够在复杂的路况下自动驾驶。

### 4.3 持续学习与平行智能

在工业5.0时代,平行智能领域进一步强调了可持续性与生态智能的相关思想<sup>[84-87]</sup>。类似地,具身智能的目标不是用模仿人类的机器人取代人类,而是用物理世界中的机器人和人工世界中的数字人来增强虚拟世界中的生物人类,形成新时代的平行智能<sup>[87]</sup>,通过3个世界的相互作用建立平行系统。具身智能中的持续学习是使智能体或组织自主获取新知识、适应新情况,并在不断变化的环境和需求中,持续提高其能力和性能。此外,利用去中心化自治组织(decentralized autonomous organization, DAO)的概念和技术,基于DAO的持续学习和平行智能建立能够持续学习和自主进化的智能体或组织。这些智能体或组织可以自主收集、分析和利用数据,通过智能合约等技术作出决策和执行任务,从而不断优化其功能和性能<sup>[84-85]</sup>。在这种模式下,DAO可以通过智能合约自动执行任务、分配资源和管理组织运营,无须人为干预。

基于持续学习和平行智能,具身智能驾驶系统能够不断学习和优化其能力,使其能够适应不同的路况、交通规则和车辆运行状态,从而实现更安全、更高效的自动驾驶。能够利用传感器、摄像头、雷达等设备感知周围环境,并通过实时数据分析和深度学习技术作出驾驶决策。在未来,具身智能驾驶将实现完全自主的车辆运行,进而消除对人类驾驶员的需求,并通过先进的自动驾驶技术,显著提高道路安全性,减少交通事故,提高交通效率。

## 5 结论

本文围绕面向自动驾驶的具身智能技术展开了深入探讨,从具身智能和自动驾驶的发展历程出发,提出了具身智能驾驶的概念,并构建了具身智能驾驶系统的核心框架,包括具身感知、具身执行和具身进化3个部分。系统将通过物理智能体、虚拟智能体、真实交通场景之间的持续交互,实现智能系统在复杂交通场景中的精准感知、高效执行、

自主进化。基于具身智能驾驶框架, 论文进一步剖析了相关技术的发展现状与存在问题, 同时探索了虚实互动数据智能、基础模型与基础智能、持续学习与平行智能在其中的重要作用和应用潜力。未来将基于具身智能驾驶理念开展深入的技术创新研究, 解决当前自动驾驶技术在开放场景中面临的挑战和问题, 并为其他移动机器人系统的发展提供一定的借鉴和思路。

## 参考文献:

- [1] WANG Y T, WANG X, WANG X X, et al. The ChatGPT after: building knowledge factories for knowledge workers with knowledge automation[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(11): 2041-2044.
- [2] TIAN Y L, LI X, ZHANG H, et al. VistaGPT: generative parallel transformers for vehicles with intelligent systems for transport automation[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(9): 4198-4207.
- [3] WANG F Y, MIAO Q H, LI X, et al. What does ChatGPT say: the DAO from algorithmic intelligence to linguistic intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2023, 10(3): 575-579.
- [4] MEMARIAN B, DOLECK T. Embodied AI in education: a review on the body, environment, and mind[J]. Education and Information Technologies, 2024, 29(1): 895-916.
- [5] 刘华平, 郭迪, 孙富春, 等. 基于形态的具身智能研究: 历史回顾与前沿进展[J]. 自动化学报, 2023, 49(6): 1131-1154.
- [6] LIU H P, GUO D, SUN F C, et al. Morphology-based embodied intelligence: historical retrospect and research progress[J]. Acta Automatica Sinica, 2023, 49(6): 1131-1154.
- [7] GUPTA A, SAVARESE S, GANGULI S, et al. Embodied intelligence via learning and evolution[J]. Nature Communications, 2021, 12(1): 5721.
- [8] HOWARD D, EIBEN A E, KENNEDY D F, et al. Evolving embodied intelligence from materials to machines[J]. Nature Machine Intelligence, 2019, 1(1): 12-19.
- [9] JIN D D, ZHANG L. Embodied intelligence weaves a better future[J]. Nature Machine Intelligence, 2020, 2: 663-664.
- [10] AI Y F, LIU Y H, GAO Y, et al. PMWorld: a parallel testing platform for autonomous driving in mines[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(1): 1402-1411.
- [11] YANG J J, HUANG Q K, GE S R, et al. On intelligent mining with parallel intelligence[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(10): 4296-4300.
- [12] CHEN L, XIE J K, ZHANG X T, et al. Mining 5.0: concept and framework for intelligent mining systems in CPSS[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(6): 3533-3536.
- [13] GU Y X, WANG Q L, QIN X L. Real-time streaming perception system for autonomous driving[C]//Proceedings of the 2021 China Automation Congress (CAC). Piscataway: IEEE Press, 2021: 5239-5244.
- [14] DENG Z Y, SHI Y J, SHEN W M. V2X-lead: LiDAR-based end-to-end autonomous driving with vehicle-to-everything communication integration[C]//Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway: IEEE Press, 2023: 7471-7478.
- [15] THRUN S. Toward robotic cars[J]. Communications of the ACM, 2010, 53(4): 99-106.
- [16] SCHWARTZ R. Safe integration of autonomous vehicles in urban traffic[J]. Nature Machine Intelligence, 2020.
- [17] LIU Y F, LIU Y Z, SHEN C. Combining minds and machines: investigating the fusion of cognitive architectures and generative models for general embodied intelligence[J]. Proceedings of the AAAI Symposium Series, 2024, 2(1): 307-314.
- [18] TURING A M. Computing machinery and intelligence[M]//Parsing the Turing Test. Dordrecht: Springer Netherlands, 2007: 23-65.
- [19] On-Road Automated Driving (ORAD) Committee. Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles[M]. [S.l.]: SAE International, 2021.
- [20] 沈宇, 王晓, 韩双双, 等. 代理技术 Agent 在智能车辆与驾驶中的应用现状[J]. 指挥与控制学报, 2019, 5(2): 87-98.
- [21] SHEN Y, WANG X, HAN S S, et al. Agent-based technology in intelligent vehicles and driving: state-of-the-art and prospect[J]. Journal of Command and Control, 2019, 5(2): 87-98.
- [22] WANG J G, WANG X, SHEN T Y, et al. Parallel vision for long-tail regularization: initial results from IVFC autonomous driving testing[J]. IEEE Transactions on Intelligent Vehicles, 2022, 7(2): 286-299.
- [23] BAI X Y, HU Z Y, ZHU X G, et al. TransFusion: robust LiDAR-camera fusion for 3D object detection with transformers[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 1080-1089.
- [24] RONG Y, WEI X Y, LIN T W, et al. DynStatF: an efficient feature fusion strategy for LiDAR 3D object detection[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2023: 3238-3247.
- [25] TIAN Y L, ZHANG X J, WANG X, et al. ACF-net: asymmetric cascade fusion for 3D detection with LiDAR point clouds and images[J]. IEEE Transactions on Intelligent Vehicles, 2023: 1-12.
- [26] SINDAGI V A, ZHOU Y, TUZEL O. MVX-net: multimodal VoxelNet for 3D object detection[C]//Proceedings of the 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 7276-7282.
- [27] 孙水发, 汤永恒, 王奔, 等. 动态场景的三维重建研究综述[J]. 计算机科学与探索, 2024, 18(4): 831-860.
- [28] SUN S F, TANG Y H, WANG B, et al. Review of research on 3D reconstruction of dynamic scenes[J]. Journal of Frontiers of Computer Science and Technology, 2024, 18(4): 831-860.
- [29] GRETTON A, SMOLA A, HUANG J Y, et al. Covariate shift and local learning by distribution matching[M]. Cambridge: MIT Press, 2009: 131-160.
- [30] GANIN Y, LEMPITSKY V. Unsupervised domain adaptation by back-propagation[EB]. arXiv preprint, 2014, arXiv: 1409.7495.
- [31] BERTHELOT D, ROELOFS R, SOHN K, et al. AdaMatch: a unified approach to semi-supervised learning and domain adaptation[EB]. arXiv preprint, 2021, arXiv: 2106.04732.
- [32] MENG D J, XIAO W, ZHANG L J, et al. Vehicle trajectory prediction based predictive collision risk assessment for autonomous driving in highway scenarios[J]. arXiv preprint, 2023, arXiv: 2304.05610.
- [33] WANG X W, ALONSO-MORA J, WANG M. Probabilistic risk metric for highway driving leveraging multi-modal trajectory predictions[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 19399-19412.
- [34] FANG L J, JIANG Q H, SHI J P, et al. TPNet: trajectory proposal network for motion prediction[C]//Proceedings of the 2020 IEEE/CVF

- Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 6796-6805.
- [32] 景荣荣, 吴兰, 张坤鹏. 基于 Transformer 的自动驾驶交互感知轨迹预测[J]. 科学技术与工程, 2023, 23(26): 11414-11423.
- JING R R, WU L, ZHANG K P. Transformer-based interaction-aware trajectory prediction for autonomous driving[J]. Science Technology and Engineering, 2023, 23(26): 11414-11423.
- [33] 李升波, 占国建, 蒋宇轩, 等. 类脑学习型自动驾驶决策系统的关键技术[J]. 汽车工程, 2023, 45(9): 1499-1515.
- LI S B, ZHAN G J, JIANG Y X, et al. Key technologies of brain-inspired decision and control intelligence for autonomous driving systems[J]. Automotive Engineering, 2023, 45(9): 1499-1515.
- [34] PENG X Y, CHANG C, WANG F Y, et al. Robust multitask learning with sample gradient similarity[J]. IEEE Transactions on Systems, Man, and Cybernetics: Systems, 2024, 54(1): 497-506.
- [35] 曹佳钰, 冷魁鹏, 张科. 面向自动驾驶应用的车联多智能体信息融合协同决策机制研究[J]. 物联网学报, 2020, 4(3): 69-77.
- CAO J Y, LENG S P, ZHANG K. Multi-agent driven collaborative decision mechanism of information fusion for autonomous driving vehicles[J]. Chinese Journal on Internet of Things, 2020, 4(3): 69-77.
- [36] WANG L Y, ZHANG X X, SU H, et al. A comprehensive survey of continual learning: theory, method and application[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2024, accepted.
- [37] RANNEN A, ALJUNDI R, BLASCHKO M B, et al. Encoder based lifelong learning[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2017: 1329-1337.
- [38] LIU X L, WU C S, MENTA M, et al. Generative feature replay for class-incremental learning[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2020: 915-924.
- [39] JOSEPH K J, BALASUBRAMANIAN V N. Meta-consolidation for continual learning[EB]. arXiv preprint, 2020, arXiv: 2010.00352.
- [40] WANG R Q, BAO Y X, ZHANG B C, et al. Anti-retroactive interference for lifelong learning[M]//Lecture Notes in Computer Science. Cham: Springer Nature, 2022: 163-178.
- [41] SHI Y J, ZHOU K Q, LIANG J, et al. Mimicking the oracle: an initial phase decorrelation approach for class incremental learning[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 16701-16710.
- [42] ZHANG G W, WANG L Y, KANG G L, et al. SLCA: slow learner with classifier alignment for continual learning on a pre-trained model [C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 19091-19101.
- [43] KIM G, XIAO C N, KONISHI T, et al. A theoretical study on solving continual learning[EB]. arXiv preprint, 2022, arXiv: 2211.02633.
- [44] MCCLOSKEY M, COHEN N J. Catastrophic interference in connectionist networks: the sequential learning problem[M]//Psychology of Learning and Motivation. Amsterdam: Elsevier, 1989: 109-165.
- [45] MASANA M, LIU X L, TWARDOWSKI B, et al. Class-incremental learning: survey and performance evaluation on image classification[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(5): 5513-5533.
- [46] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite[C]//Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2012: 3354-3361.
- [47] CAESAR H, BANKITI V, LANG A H, et al. NuScenes: a multimodal dataset for autonomous driving[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 11618-11628.
- [48] SUN P, KRETZSCHMAR H, DOTIWALLA X, et al. Scalability in perception for autonomous driving: waymo open dataset[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 2443-2451.
- [49] CHANG M F, LAMBERT J, SANGKLOY P, et al. Argoverse: 3D tracking and forecasting with rich maps[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 8740-8749.
- [50] HOUSTON J, ZUIDHOF G, BERGAMINI L, et al. One thousand and one hours: self-driving motion prediction dataset[EB]. arXiv preprint, 2020, arXiv: 2006.14480.
- [51] FENG D P, QI Y H, ZHONG S P, et al. S3E: a large-scale multimodal dataset for collaborative SLAM[EB]. arXiv preprint, 2022, arXiv: 2210.13723.
- [52] ALIBEIGI M, LJUNGBERGH W, TONDERSKI A, et al. Zenseact open dataset: a large-scale and diverse multimodal dataset for autonomous driving[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 20121-20131.
- [53] DOSOVITSKIY A, ROS G, CODEVILLA F, et al. CARLA: an open urban driving simulator[EB]. arXiv preprint, 2017, arXiv: 1711.03938.
- [54] HUANG X Y, CHENG X J, GENG Q C, et al. The ApolloScape dataset for autonomous driving[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway: IEEE Press, 2018: 1067-10676.
- [55] GEYER J, KASSAHUN Y, MAHMUDI M, et al. A2D2: audi autonomous driving dataset[EB]. arXiv preprint, 2020, arXiv: 2004.06320.
- [56] CHEN M, LIU P F, ZHAO H C. LiDAR-camera fusion: dual transformer enhancement for 3D object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 120: 105815.
- [57] AN P, LIANG J X, HONG X, et al. Leveraging self-paced semi-supervised learning with prior knowledge for 3D object detection on a LiDAR-camera system[J]. Remote Sensing, 2023, 15(3): 627.
- [58] ZHANG L, LI X, TANG K C, et al. FS-net: LiDAR-camera fusion with matched scale for 3D object detection in autonomous driving[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(11): 12154-12165.
- [59] LI Y W, YU A W, MENG T J, et al. DeepFusion: lidar-camera deep fusion for multi-modal 3D object detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 17161-17170.
- [60] KOH J, KIM J, YOO J H, et al. Joint 3D object detection and tracking using spatio-temporal representation of camera image and LiDAR point clouds[J]. Proceedings of the AAAI Conference on Artificial Intelligence, 2022, 36(1): 1210-1218.
- [61] HUANG K M, HAO Q. Joint multi-object detection and tracking with camera-LiDAR fusion for autonomous driving[C]//Proceedings of the 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). New York: ACM, 2021: 6983-6989.
- [62] ZHANG K P, LIU Y H, MEI F, et al. Boost correlation features with 3D-MIoU-based camera-LiDAR fusion for MODT in autonomous driving[J]. Remote Sensing, 2023, 15(4): 874.
- [63] WANG X Y, FU C Y, LI Z K, et al. DeepFusionMOT: a 3D multi-object tracking framework based on camera-LiDAR fusion with deep association[J]. IEEE Robotics and Automation Letters, 2022, 7(3):



- 8260-8267.
- [64] ZHU Z M, NIE J H, WU H, et al. MSA-MOT: multi-stage association for 3D multimodality multi-object tracking[J]. *Sensors*, 2022, 22(22): 8650.
- [65] WANG X Y, FU C Y, HE J W, et al. StrongFusionMOT: a multi-object tracking method based on LiDAR-camera fusion[J]. *IEEE Sensors Journal*, 2023, 23(11): 11241-11252.
- [66] BÜCHNER M, VALADA A. 3D multi-object tracking using graph neural networks with cross-edge modality attention[J]. *IEEE Robotics and Automation Letters*, 2022, 7(4): 9707-9714.
- [67] XIE B Q, YANG Z M, YANG L, et al. AMMF: attention-based multi-phase multi-task fusion for small contour object 3D detection[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2023, 24(2): 1692-1701.
- [68] CHIU H K, LI J, AMBRUS R, et al. Probabilistic 3D multi-modal, multi-object tracking for autonomous driving[C]//*Proceedings of the 2021 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway: IEEE Press, 2021: 14227-14233.
- [69] SHAHVERDY M, FATHY M, BERANGI R, et al. Driver behavior detection and classification using deep convolutional neural networks[J]. *Expert Systems with Applications*, 2020, 149: 113240.
- [70] JEONG D, BAEK M, LEE S S. Long-term prediction of vehicle trajectory based on a deep neural network[C]//*Proceedings of the 2017 International Conference on Information and Communication Technology Convergence (ICTC)*. Piscataway: IEEE Press, 2017: 725-727.
- [71] YANG L, LEI W M, ZHANG W, et al. Dual-flow network with attention for autonomous driving[J]. *Frontiers in Neurorobotics*, 2023, 16: 978225.
- [72] KIM C J, LEE M J, HWANG K H, et al. End-to-end deep learning-based autonomous driving control for high-speed environment[J]. *The Journal of Supercomputing*, 2021: 1961-1982.
- [73] KATUWANDENIYA K, MIRO J V, DANTANARAYANA L. End-to-end joint intention estimation for shared control personal mobility navigation[C]//*Proceedings of the 2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. Piscataway: IEEE Press, 2020: 1-6.
- [74] LI X, WANG Y T, WANG K F, et al. The ParallelEye-CS dataset: constructing artificial scenes for evaluating the visual intelligence of intelligent vehicles[C]//*Proceedings of the 2018 IEEE Intelligent Vehicles Symposium (IV)*. Piscataway: IEEE Press, 2018: 37-42.
- [75] LI X, WANG K F, GU X F, et al. ParallelEye pipeline: an effective method to synthesize images for improving the visual intelligence of intelligent vehicles[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(9): 5545-5556.
- [76] MIAO Q H, LV Y S, HUANG M, et al. Parallel learning: overview and perspective for computational learning across Syn2Real and Sim2Real[J]. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(3): 603-631.
- [77] LI L, LIN Y L, ZHENG N N, et al. Parallel learning: a perspective and a framework[J]. *IEEE/CAA Journal of Automatica Sinica*, 2017, 4(3): 389-395.
- [78] SHEN Y, LIU Y H, TIAN Y L, et al. Parallel sensing in metaverses: virtual-real interactive smart systems for “6S” sensing[J]. *IEEE/CAA Journal of Automatica Sinica*, 2022, 9(12): 2047-2054.
- [79] 王坤峰, 苟超, 王飞跃. 平行视觉: 基于 ACP 的智能视觉计算方法[J]. *自动化学报*, 2016, 42(10): 1490-1500.
- WANG K F, GOU C, WANG F Y. Parallel vision: an ACP-based approach to intelligent vision computing[J]. *Acta Automatica Sinica*, 2016, 42(10): 1490-1500.
- [80] 李轩, 王飞跃. 面向智能驾驶的平行视觉感知: 基本概念、框架与应用[J]. *中国图象图形学报*, 2021, 26(1): 67-81.
- LI X, WANG F Y. Parallel visual perception for intelligent driving: basic concept, framework and application[J]. *Journal of Image and Graphics*, 2021, 26(1): 67-81.
- [81] LIU Y H, SHEN L, FAN L L, et al. Parallel radars: from digital twins to digital intelligence for smart radar systems[J]. *Sensors*, 2022, 22(24): 9930.
- [82] LI X, TIAN Y L, YE P J, et al. A novel scenarios engineering methodology for foundation models in metaverse[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2023, 53(4): 2148-2159.
- [83] WANG F Y. New control paradigm for industry 5.0: from big models to foundation control and management[J]. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(8): 1643-1646.
- [84] LI J J, LIANG X L, QIN R, et al. From DAO to TAO: finding the essence of decentralization[C]//*Proceedings of the 2023 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. Piscataway: IEEE Press, 2023: 4283-4288.
- [85] LI J J, WANG F Y. The TAO of blockchain intelligence for intelligent web 3.0[J]. *IEEE/CAA Journal of Automatica Sinica*, 2023, 10(12): 2183-2186.
- [86] WANG Z R, LV C, WANG F Y. A new era of intelligent vehicles and intelligent transportation systems: digital twins and parallel intelligence[J]. *IEEE Transactions on Intelligent Vehicles*, 2023, 8(4): 2619-2627.
- [87] 王飞跃, 王雨桐. 数字科学家与平行科学: AI4S 和 S4AI 的本源与目标[J]. *中国科学院院刊*, 2024, 39(1): 27-33.
- WANG F Y, WANG Y T. Digital scientists and parallel sciences: the origin and goal of AI for Science and Science for AI[J]. *Bulletin of Chinese Academy of Sciences*, 2024, 39(1): 27-33.

#### [作者简介]



沈甜雨 (1996-), 女, 博士, 北京化工大学信息科学与技术学院副教授, 主要研究方向为机器视觉、智能感知与智能机器人系统。



李志伟 (1990-), 男, 博士, 北京化工大学信息科学与技术学院硕士生导师, 主要研究方向为自动驾驶融合感知、视觉语言大模型、智能机器人系统。



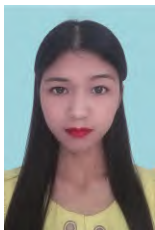
范丽丽 (1991-), 女, 博士, 北京理工大学信息与电子学院博士后, 主要研究方向为计算机视觉、跨模态感知与理解、类脑认知与决策。



张庭祯（2000- ），男，北京化工大学信息科学与技术学院硕士生，主要研究方向为计算机视觉、三维目标检测、平行视觉等。



刘华平（1976- ），男，博士，清华大学计算机科学与技术系教授，国家杰出青年科学基金获得者，主要研究方向为具身感知与学习。



唐丹丹（1997- ），女，燕山大学硕士生，主要研究方向为计算机视觉、三维目标检测。



王坤峰（1982- ），男，博士，北京化工大学信息科学与技术学院教授，主要研究方向为计算机视觉、多模态感知、智能无人系统。



周美华（2000- ），女，北京同仁医院眼科研究所科研助理，主要研究方向为机器学习、计算机视觉、自动驾驶。