

面向自动驾驶的感知决策一体化综述

刘泽禹, 张慧, 李浥东

(北京交通大学计算机科学与技术学院, 北京 100044)

摘要: 近年来, 感知决策一体化的端到端方案取得了突破性进展, 为提升自动驾驶的安全性和可靠性提供了新的思路。现有的综述多聚焦于单独车辆对外部环境的感知, 忽略了在多智能体协同下的复杂交互关系。基于当前的研究现状, 从应用方法的角度出发, 探讨了单智能体感知决策一体化方法, 并分析了在多智能体协同条件下的感知决策一体化方法。首先, 对感知决策一体化自动驾驶的基础技术进行了总结; 其次, 重点介绍了感知决策一体化自动驾驶技术的最新研究进展; 再次, 阐述了自动驾驶领域常用的大规模公共数据集; 接着, 介绍了目前常用的感知决策一体化自动驾驶评估方法; 最后, 对感知决策一体化自动驾驶进行了总结与展望。

关键词: 智慧交通; 自动驾驶; 感知决策一体化

中图分类号: TP39

文献标志码: A

doi: 10.11959/j.issn.2096-6652.202502

A review of integrated perception and decision-making for autonomous driving

LIU Zeyu, ZHANG Hui, LI Yidong

School of Computer and Information Technology, Beijing Jiaotong University, Beijing 100044, China

Abstract: In recent years, end-to-end solutions integrating perception and decision-making have made groundbreaking progress, providing new approaches to enhancing the safety and reliability of autonomous driving. Existing reviews often focus on the perception of the external environment by individual vehicles, overlooking the complex interactions in multi-agent collaboration. Based on the current state of research, integrated perception and decision-making methods for single agents were explored from an application perspective, and methods under multi-agent collaborative conditions were analyzed. Firstly, foundational technologies for integrated perception and decision-making in autonomous driving were summarized. Secondly, the latest research advancements in integrated perception and decision-making technologies for autonomous driving were highlighted. Thirdly, large-scale public datasets commonly used in the field of autonomous driving were elaborated upon. Then, the evaluation methods currently employed for integrated perception and decision-making in autonomous driving were introduced. Finally, a summary and outlook on integrated perception and decision-making in autonomous driving were presented.

Key words: intelligent transportation, autonomous driving, integrated perception and decision-making

0 引言

自动驾驶技术是人工智能领域中一个非常重要

的研究方向。随着信息技术的迅猛发展和交通需求的不断增长, 自动驾驶系统已成为现代交通领域的关键技术, 持续推动智慧交通系统的创新及升级。

收稿日期: 2024-10-16; 修回日期: 2025-01-10

通信作者: 张慧, huizhang1@bjtu.edu.cn

基金项目: 中央高校基本科研业务费专项资金资助 (No.2024JBMC045)

Foundation Item: The Fundamental Research Funds for the Central Universities (No.2024JBMC045)

该技术不仅着眼于实现简单的道路导航和车辆控制，而且致力于使车辆能够自主处理复杂环境下的感知与决策任务，从而确保自动驾驶技术在高级驾驶辅助系统、无人运输车等一系列实际应用中的安全性和可靠性。在传统“模块式”自动驾驶系统中，感知模块与决策模块之间的信息传递采用级联式架构，这一架构引起的累积误差会导致路径规划不准确、动态障碍物识别延迟等问题。因此，感知决策一体化的端到端方法成为研究者关注的焦点。端到端方法中的“端”指输入端和输出端，“端到端”即原始输入信息直接传输至输出端进行决策。整个过程通过一个或多个模型实现，直接将输入数据映射到输出控制决策，而不需要依赖中间的手工定义规则或经过多个独立的功能模块。与传统的自动驾驶技术相比，端到端方法在工程实现方面所需的工作量较少，但对数据量的需求则相对较高。传统自动驾驶技术与感知决策一体化驾驶技术的对比如图1所示。

端到端自动驾驶技术的发展经历了感知决策一体化程度不断提升的历史阶段，其源头可追溯至1988年神经网络中的自动驾驶陆地载具（autonomous land vehicle in a neural network, ALVINN）系统^[1]，这一系统采用相机和激光测距仪采集的数据通过简单神经网络生成车辆转向指令。随着计算机

视觉和深度学习技术的进步，2016年，Bojarski等^[2]提出了基于卷积神经网络（convolutional neural network, CNN）的端到端模型PilotNet，通过捕获多个摄像头的的数据，训练CNN以模拟人类驾驶员的行为，实现了车辆的自动导航，但PilotNet在处理复杂环境时的能力不足，这限制了其在真实世界中应用的适应能力，因此，人们开始考虑整合多模态数据来训练端到端模型。2021年，基于Transformer架构^[3]的TransFuser模型^[4]应运而生，该模型利用自注意力机制集成图像与光探测和测距（light detection and ranging, LiDAR）数据，为自动驾驶系统带来了更深入的环境理解能力。近年来，该领域涌现许多新兴算法和框架，显著提升了模型应对复杂驾驶场景的能力。这些进展从根本上推动了自动驾驶技术的发展，加快了其在实际应用中的落地。然而，感知决策一体化自动驾驶技术仍面临以下问题和难点。

（1）模型学习的自适应性

端到端模型是数据驱动的，其性能依赖于大规模且多样性的标记数据集。针对特定的任务，增加训练数据的规模和多样性可以提升端到端模型的泛化能力，增强其对环境的适应能力。当前的数据集主要分为两类：仿真数据集和真实数据集。仿真数据集在仿真环境中生成，能够模拟不同的场景和条

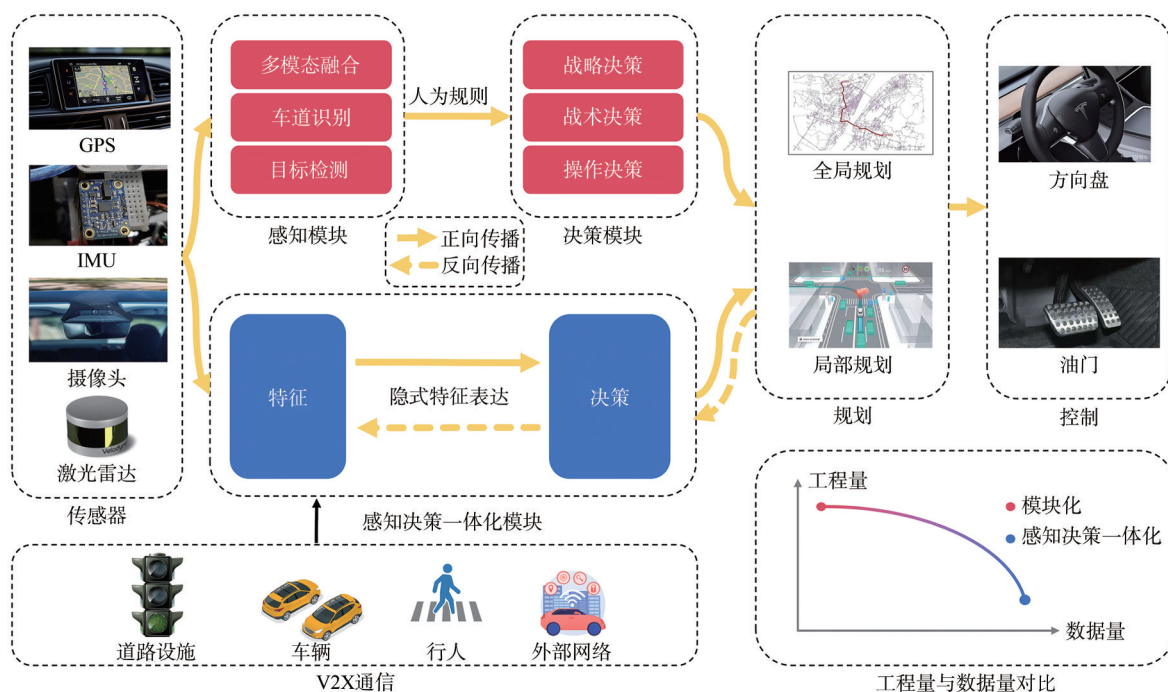


图1 传统自动驾驶技术与感知决策一体化驾驶技术的对比

件。然而,由于与真实环境之间存在分布差异,模型在实际应用中性能表现不佳。真实数据集则受限于是光照、天气、复杂背景和物体遮挡等因素的影响,难以全面覆盖所有的驾驶场景,导致模型在处理特定复杂情形时性能下降。因此,需要增强模型的自适应能力,使其能够有效应对不同环境中的变化和不可预见的情况,从而提升其在实际应用中的可靠性和鲁棒性。

(2) 端到端模型的“黑盒”特性

对于自动驾驶系统而言,整合感知与决策模块可以消除模块之间信息传递的累积误差,但是这种将原始输入数据直接映射到输出的结构,使得处理步骤和细节不再可见,使模型具有“黑盒”特性,这不仅增加了理解系统行为的难度,也给安全性和系统优化带来了额外的挑战。

(3) 计算资源的高需求性

自动驾驶系统需要处理来自多种传感器的多模态数据,这一过程要求系统具备强大的并行计算能力。然而,由于汽车环境的特殊性,自动驾驶汽车往往无法获取足够的计算资源来支持复杂的数据处理需求。因此,在高度动态的环境中,利用有限的计算资源对环境进行实时处理,是感知决策一体化自动驾驶技术发展的难点。

为了克服上述挑战,研究者已经提出了许多感知决策一体化自动驾驶方法。这些方法在数据预处理、特征提取和决策制定等各环节采用了不同的处理策略,成为当前的研究热点。本文从应用方法的角度出发,不仅对单智能体感知决策一体化方法进行了探讨,还分析了在多智能体协同条件下的感知决策一体化的方法。

1 感知决策一体化自动驾驶的基础技术

自动驾驶系统的核心在于有效获取和处理环境信息,以实现安全和高效行驶。其基础技术包括传感器技术、面向多智能体协同的车联网通信技术以及感知与决策技术。传感器技术通过多种传感器为系统提供实时的环境感知能力,以便准确识别周围的交通情况和障碍物;面向多智能体协同的车联网通信技术促进了不同车辆、行人及基础设施之间的信息共享与协同,增强了道路安全性和交通效率;感知与决策技术为自动驾驶系统的智能决策提供支撑,使其能够在复杂的环境下做出合理的判断和反应。本节详细介绍感知决策一体化自动驾驶的基础

技术,包括摄像头、雷达、全球导航卫星系统(global navigation satellite system, GNSS)模块和惯性测量单元(inertial measurement unit, IMU)等传感器技术,车对车(vehicle to vehicle, V2V)、车对人(vehicle to pedestrian, V2P)、车对基础设施(vehicle to infrastructure, V2I)和车对网络(vehicle to network, V2N)等通信技术,以及环境信息感知技术和决策规划技术。

1.1 传感器技术

(1) 摄像头

图像作为端到端模型的输入,在自动驾驶技术中发挥着重要作用。通过捕捉车辆周围的图像信息,系统能够实时分析和理解环境。摄像头主要分为三原色(red green blue, RGB)摄像头、深度摄像头和热成像摄像头。RGB摄像头是最常见的彩色摄像头,广泛应用于环境感知和物体识别,能够提供丰富的视觉信息;深度摄像头主要通过立体视觉(通过两个或多个摄像头的视差计算)或结构光技术(通过投射特定图案并观察其变形)来获取深度信息,帮助系统理解物体的距离和形状,提升环境感知的准确性;热成像摄像头在夜间或低能见度条件下尤为重要,能够检测温度差异,有效识别行人和动物,增强安全性。然而,摄像头的性能可能受到天气条件、物体遮挡和光照变化的影响,导致图像质量下降,进而影响系统的整体感知能力。因此,摄像头常常需要与其他传感器协同进行感知。

(2) 雷达

雷达通过发射声波或光信号,并接收反射信号来测量物体的距离和速度,用于获取实时的速度和距离信息,增强自动驾驶系统的安全性和可靠性。目前,雷达主要分为两类:超声波雷达和激光雷达。超声波雷达通过发射高频声波并测量反射时间来探测近距离物体,主要应用于停车辅助和低速行驶中的障碍物识别,它具有成本低和精度高的优点,但有效探测范围有限;激光雷达则通过发射激光束并测量反射时间,生成高精度的三维环境模型,广泛用于环境建模和物体检测,能够提供详细的深度信息,并适应复杂环境。

(3) GNSS模块

GNSS模块能够提供高精度的定位和导航信息,通过接收来自多个卫星的信号,实时计算车辆的地理位置、速度和行驶方向,以协助自动驾驶系统进行地图构建、环境感知和路径规划。然而,在

城市、峡谷或隧道等环境中,GNSS模块可能受到信号干扰,常常需要结合IMU以提高低信号环境下定位的准确性。

(4) IMU

IMU在自动驾驶中常用于辅助GNSS。IMU由加速度计、陀螺仪和磁力计这3部分组成。加速度计测量线性加速度,用于检测汽车的运动状态;陀螺仪测量角速度,以监测汽车的旋转和姿态变化;磁力计用于测量地磁场,以确定车辆在地磁坐标系中的方向。通过这3部分的组合,IMU能够实时测量和计算车辆的加速度、角速度和方向,为自动驾驶系统提供关键的运动数据。

1.2 车联网通信技术

(1) V2V

V2V通信技术通过无线通信技术实现车辆之间的实时信息交换。自动驾驶车辆能利用V2V通信系统共享位置、速度、加速度、转向意图等关键信息,有效预测潜在的风险并避免碰撞。在多车协同的感知决策一体化自动驾驶技术中,V2V通信能够显著提升车辆的感知能力,帮助自动驾驶系统在动态环境中做出更智能的决策。

(2) V2P

V2P通信技术关注于提高行人与车辆之间的交互安全。通过智能设备或路边传感器,车辆可以实时检测行人的位置和移动轨迹,并能够在行人即将进入行驶路径时发出预警并采取避让措施^[5]。这种安全信息的共享与交互能力,使自动驾驶系统能够更全面地感知周围环境,确保行人与车辆之间的顺畅互动。

(3) V2I

V2I通信技术实现了车辆与基础设施之间的智能互联。通过路侧单元(road side unit, RSU)、交通信号灯、摄像头等基础设施设备,车辆能够实时获取路况信息、交通管制指令、道路施工警告等关键数据^[6],协助自动驾驶车辆提前规划行驶路线,避免进入拥堵和危险区域,从而提高行驶效率。

(4) V2N

V2N通信技术将车辆与互联网深度融合,将智能车辆、路侧设备、交通地图、交通服务商等连接起来,通过局域优化方式规划智能汽车的行车路线^[7]。随着5G等新一代通信技术的不断发展,V2N通信技术将实现更高速度、更低时延的数据传输,为车联网和自动驾驶技术的发展注入新的动力。

1.3 感知与决策技术

(1) 环境信息感知技术

环境信息感知技术通过各种传感器和算法,收集、处理和分析周围环境的信息,主要分为鸟瞰图(bird's eye view, BEV)感知和三维占用(3D occupancy)感知。BEV感知将三维场景中的点或特征投影到一个水平面上,将3D坐标转换为2D坐标,以生成与车辆位置相对应的俯视图,为特征提取和运动规划提供了统一的表示空间^[8],有效解决了传统二维视角中的遮挡和比例失调问题。例如,Wu等^[9]提出了基于鸟瞰视图映射的感知和轨迹预测方法MotionNet,通过对BEV进行二维卷积来提取目标特征并实现路径规划。BEV感知因忽略了Z轴坐标的垂直结构信息,给环境感知带来了一定的局限性。此外,当场景较为复杂或物体之间的间隙较小时,BEV感知无法清晰区分物体的具体位置或相对运动。为了解决上述问题,一些研究者提出通过三维占用网络来对环境进行感知,该网络能够捕捉Z轴坐标的垂直结构特征,提供物体的深度、尺寸以及空间位置等信息,从而提升了场景理解的效果。例如,Peng等^[10]提出了结合单幅RGB图像和稀疏LiDAR点云的3D目标检测方法OccupancyM3D,将单个RGB图像提升至视锥体空间,生成视锥体特征以学习占用情况,通过网格采样,视锥体特征转换为体素化的3D特征,进而在3D空间中分析占用情况。

(2) 决策规划技术

决策与规划技术要求在复杂环境中做出合理的驾驶判断,传统的决策与规划策略在流程上大致分为3个步骤:行为预测、行为决策和运动规划^[11]。

行为预测阶段,系统通过分析交通参与者的历史表现和当前状态来推测他们的未来行为。行为预测方法可分为基于学习预测的方法和基于概率预测的方法。基于学习预测的方法利用长短时记忆网络等方法提取时空特征,再对车辆未来的行为进行预测^[12-14];基于概率预测的方法通过估算未来行为的概率,选出最有可能的结果^[15-17]。

行为决策阶段,系统会根据当前环境感知信息和预测结果选择合适的驾驶策略。国际自动机工程师学会(Society of Automotive Engineers International, SAE International)与国际标准化组织的SAE-J3016标准将行为决策分为战略层、战术层和操作层^[18]。战略层涉及车辆的出发时间、最佳路

线等全局策略选择。战术层定义了车辆在行驶时的机动操作,如是否以及何时超车或变道、选择合适的速度等,目前常用的方法有基于规则的决策和基于学习的决策。基于规则的决策常用状态机决策模型和决策树模型,两种模型都依赖于提前制定好的状态以及状态之间迁移的规则来选择满足条件的决策路径^[19-21],此类方法较为稳定,但人工制定的状态和规则无法涵盖现实世界中的所有情况,泛化能力不强。因此,有人提出了基于学习的决策,通过大量的数据,学习人类驾驶员在复杂场景中的行为来提升决策能力,如基于强化学习的自动驾驶模型,通过让智能体在模拟环境中不断试错后给予奖励或惩罚,形成反馈来优化智能体的驾驶行为^[22-25]。操作层定义了转向、制动和加速等微观修正,以维持交通中的车道位置或避免道路中的突发障碍及危险事件。

运动规划阶段,系统通过考虑车辆的动态特性,结合行为决策部分选择的驾驶策略计算出最佳的运动轨迹。运动规划分为全局规划和局部规划^[26]。全局规划的主要任务是在全局道路中选择最优路径,通过有向图表现道路网络,利用最短路径算法选择节点与节点之间成本最低的路径。局部规划关注车辆在选定路径上的具体行驶行为,包括避障、车道保持、车辆跟随等。全局规划为局部规划提供了大致的方向,局部规划则确保了车辆能够安全、高效地在全局规划的路径上行驶。

2 感知决策一体化自动驾驶技术的发展

感知决策一体化自动驾驶技术的根本问题是将环境感知和决策制定结合。按照数据来源的不同,感知决策一体化自动驾驶技术可以分为以视觉为主的感知决策一体化方法、融合多传感器信息的感知决策一体化方法和基于多智能体协同的感知决策一体化方法。以视觉为主的感知决策一体化方法的数据主要来自摄像头;融合多传感器信息的感知决策一体化方法的数据来自多种传感器;基于多智能体协同的感知决策一体化方法的数据来自车与多种交通参与者。本节从以视觉为主的感知决策一体化方法、融合多传感器信息的感知决策一体化方法和基于多智能体协同的感知决策一体化方法这3个方面分类介绍目前最新的感知决策一体化自动驾驶方案。感知决策一体化自动驾驶技术的分类和发展路线如图2所示。

2.1 以视觉为主的感知决策一体化方法

以视觉为主的感知决策一体化方法理论基础源于人类驾驶员对视觉信息的获取与处理过程,即通过视觉系统获取周围环境数据,并运用自然神经网络进行感知与决策。

鉴于此,许多研究借鉴了人类视觉处理和决策机制的策略,运用深度学习和深度强化学习技术实现端到端自动化驾驶。在车道保持场景中,Chen等^[27]提出了一种基于CNN的端到端学习方

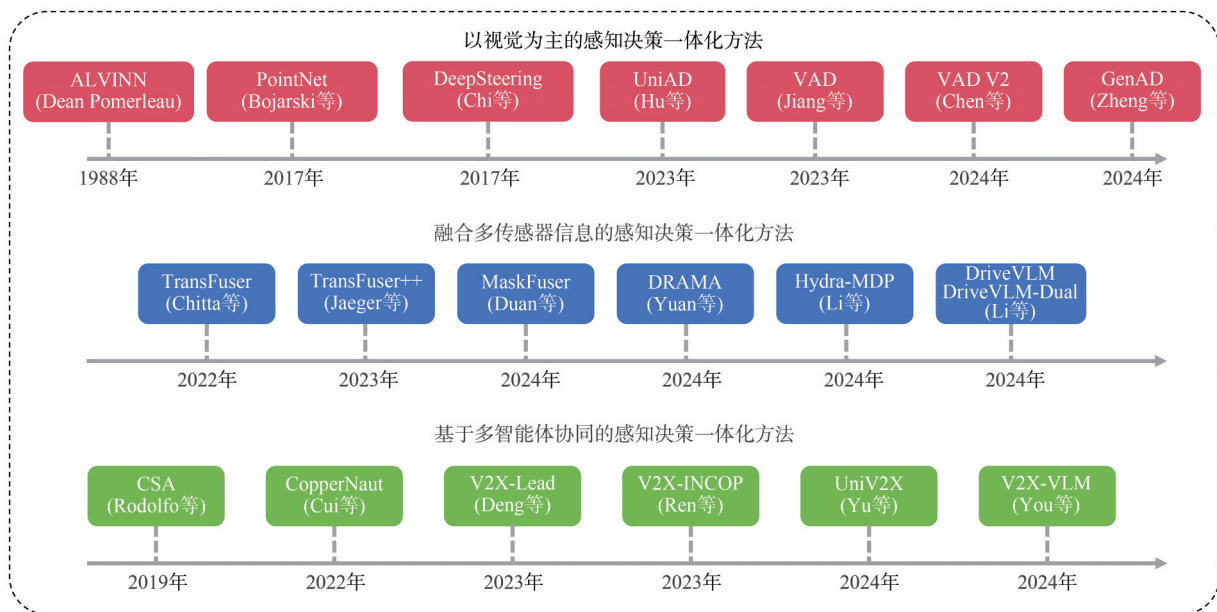


图2 感知决策一体化自动驾驶技术的分类和发展路线

法, 通过将图像帧输入CNN中, 直接输出相应的转向角度, 实现自动驾驶汽车的车道保持。尽管上述方法在转向任务中表现良好, 但单一任务的学习方式未能充分考虑速度、方向协同控制的重要性。针对此, Yang等^[28]提出了一种基于多任务学习的速度和方向盘角度估计方法, 通过CNN提取驾驶场景中的视觉特征, 同步预测车辆速度和方向盘角度。CNN有效提取了图像特征, 但无法利用一段时间内图像间的关联信息。为了解决这一问题, Chi等^[29]提出了一种基于改进的卷积长短期记忆网络结构, 通过结合空间和时间信息, 优化了模型对时间信息的利用。与上述方法不同, An等^[30]提出了一种基于机器学习的不确定性建模和运行时验证方法, 通过机器学习将人类驾驶风格分类, 将分类结果作为参数输入参数化随机混合状态图模型中, 再映射到网络概率定时自动机, 结合统计模型检查, 实现了车道变换场景中的决策。尽管多任务学习的方法在一定程度上提高了端到端模型的泛化能力, 但会出现共享信息对其他任务造成干扰, 使模型性能下降的情况。针对此, Hu等^[31]提出的统一自动驾驶(unified autonomous driving, UniAD)方法将感知、决策、规划三大类任务通过Transformer整合到统一的端到端网络框架下, 构建了全栈端到端自动驾驶模型。与UniAD不同, Jiang等^[32]采用稀疏的向量化场景表示方法, 提出了矢量化自动驾驶(vectorized autonomous driving, VAD)范式, 将多个视角的图像输入BEV解码器中, 转换为特征图, 利用矢量化运动变换器同时实现对动态目标的感知和轨迹预测, 通过矢量化地图变换器提取的静态目标元素, 将动态信息和静态信息一起输入规划变换器中, 实现对规划轨迹的矢量化约束。然而, BEV场景表征无法充分捕捉环境中的动态变化, 且需要较高的计算代价。针对这些问题, Chen等^[33]提出了基于概率规划的端到端向量化自动驾驶框架VADv2, 摒弃了以往的基于BEV的场景表征方法, 采用Token化的方式将源自多个视角的图像数据作为输入, 直接馈送至规划变换器中, 输出一系列动作的概率分布, 随后从中采样出一个具体的动作进行控制。此外, Zheng等^[34]提出了生成式端到端自动驾驶(generative end-to-end autonomous driving, GenAD)方法, 该方法通过将自动驾驶任务转化为轨迹生成问题, 构建了一个能够模拟真实交通场景的系统。GenAD方法利用ResNet50^[35]提取输入

图像的多尺度特征, 生成BEV表征, 结合高级特征和低级特征以增强环境理解, 采用变分自编码器进行轨迹先验建模, 利用时间模型捕捉自车与其他交通参与者的运动动态, 生成未来轨迹。上述这些方法在自动驾驶中往往依赖于密集或稀疏的场景表示, 导致在全面性和效率之间存在权衡。为了解决这一问题, Zheng等^[36]提出了以高斯为中心的端到端自动驾驶GaussianAD, 通过引入3D语义高斯作为核心表示, 结合动态和静态元素的特征, 更加全面地描述复杂环境, 使用稀疏卷积网络来进行三维感知和决策, 同时通过辅助的3D占用标签来自动化留存与训练, 减少了标注成本。顺序处理无法同时进行多个任务, 限制了系统的处理能力, 使其在面对复杂场景时难以迅速做出反应。为了提高端到端自动驾驶的并行处理能力, Weng等^[37]提出了并行端到端方法PARA-Drive, 通过将端到端自动驾驶系统划分为多个独立但相互关联的模块, 包括在线映射模块、运动预测模块、占用预测模块和运动规划模块, 采用共同训练的方法, 允许不同模块共享特征信息, 确保每个任务捕获到相应的上下文信息, 提高决策的准确性。Doll等^[38]提出了双流方法DualAD, 通过分别解耦动态代理和静态场景元素的表示动态对象使用基于对象的查询进行建模, 同时静态元素通过鸟瞰图网格进行表示, 可以更有效地补偿自车和其他物体之间的运动。DualVAD在DualAD的基础上增强了动态物体的运动感知和预测能力, 针对动态物体的处理进行了优化, 表现出更好的性能和适应性。

目前, 以视觉为主的感知决策一体化方法已在一些实际的自动驾驶系统中应用, 如特斯拉自动驾驶(full self-driving, FSD)技术。从FSD V12版本开始引入“端到端神经网络”技术, 通过对大量行车数据进行深度学习, 使用感知决策一体化大模型来直接操控车辆, 实现实际环境中从输入到输出的直接映射。与特斯拉一体化大模型不同, 极越汽车搭载了百度Apollo自动驾驶基础模型(autonomous driving foundation model, ADFM)方案, 采用了两段式端到端技术, 即感知的端到端和决策规划的端到端, 这种方式更利于收集长尾数据, 同时保证了智能驾驶的可控性。类似于Apollo ADFM, 卓驭科技提出了感知三网合一、预测-决策一体的端到端模型, 也采用了两段式端到端方案, 该方案利用动态BEV对移动障碍物进行检测

和追踪,静态BEV则处理道路基础设施、交通标志等静态环境元素。此外,双目占用(occupancy, OCC)网络利用立体视觉技术,通过双目摄像头获取的图像来估计物体的深度信息,再将感知信息输入预测-决策一体的端到端模型中,实现了实际环境下的端到端自动驾驶。与这些方案的目标不同,美国自动驾驶初创公司Comma.ai提出了开源项目Openpilot,旨在基于单一设备实现L2级别的辅助驾驶功能,Openpilot通过端到端Supercombo模型处理摄像头输入,以预测车道、障碍物位置及自行车轨迹。

2.2 融合多传感器信息的感知决策一体化方法

为了弥补单模态方法获取信息能力的不足,融合多传感器信息的感知决策一体化方法将多种传感器模态的数据整合后输入端到端模型中,实现对车辆的控制。

Prakash等^[4]将Transformer架构应用于多传感器模态融合的端到端技术中,提出了TransFuser方法,该方法将来自激光雷达的BEV场景表征和摄像头视觉数据输入多个编码器中,进行多尺度特征提取,从高分辨率的图像和低时延的LiDAR信息中获取丰富的环境上下文,提高了对复杂情况的适应能力。然而,TransFuser方法使用的目标点表示方式会导致预测结果不确定,并且对目标点的强依赖性使模型无法有效应对未见过的情况。针对此,Jaeger等^[39]提出了改进预测机制的TransFuser++,通过引入更精确的目标点条件化表示和两阶段训练策略,提升了对复杂交通场景的适应能力。然而,TransFuser依赖于直接特征拼接与变换,缺乏有效的跨模态特征对齐。为了解决上述问题,Duan等^[40]提出了带有交叉模态掩码自动编码器的MaskFuser方法,通过独立的CNN提取多模态传感器特征,并应用交叉模态掩码自动编码器进行掩蔽预训练,以增强对输入缺失的鲁棒性,利用混合的早期和晚期融合策略进行特征的统一语义标记,基于融合特征预测路径点转化为具体的车辆运动指令。端到端运动规划模型通常体积庞大,难以在资源受限的系统中部署,导致实际应用受限。知识蒸馏作为一种有效的模型压缩技术,通过将大型教师模型的知识传递给较小的学生模型,解决了传统端到端运动规划模型在资源受限系统中部署遇到的问题。Li等^[41]提出了采用多目标Hydra蒸馏的端到端多模态通用规划框架Hydra-MDP,通过Trans-

Fuser融合图像和LiDAR点云特征,采用多教师知识蒸馏的方式,结合人类教师和规则基础教师的知识,训练多头解码器模型以逆向学习环境对规划的影响,实现了端到端的可微分优化流程。Feng等^[42]提出了PlanKD方法,通过将大型教师模型中的关键知识转移至小型学生模型,利用信息瓶颈策略和安全感知轨迹点注意机制,实现了对端到端运动规划模型的有效压缩与性能提升。除了依赖传感器对车辆外部环境进行感知,人类驾驶员的注意力也能为端到端模型提供信息支持。Xu等^[43]提出了一种整合驾驶员的注意力感知决策一体化自动驾驶策略M2DA,利用LVAFusion模块整合图像和LiDAR点云数据,引入驾驶员注意力预测以模拟人类的视觉焦点,实现了对关键区域的有效识别。在决策方面,利用自回归路径点预测网络生成未来路径点,从而为控制模块提供决策信号。近年来,许多研究者针对Transformer架构复杂度、存储需求大等缺点做出了改进。Gu等^[44]提出了具有线性计算复杂度的状态空间模型Mamba框架,Mamba集成了选择性结构化状态空间模型(state space model, SSM),并采用了简化的神经网络架构,性能优于同尺寸的Transformer模型。Yuan等^[45]将Mamba引入自动驾驶领域,提出了基于Mamba的端到端运动规划器DRAMA,通过CNN和Mamba融合来自相机和激光雷达的多尺度特征,使用Mamba-Transformer解码器在保持低计算复杂度的同时,处理长序列输入,从而生成最终轨迹。

视觉语言模型(visual language model, VLM)已成为感知决策一体化自动驾驶系统的研究热点之一,通过结合视觉信息与自然语言描述,它能够更全面地解析复杂的交通环境。Wen等^[46]评估了GPT-4V在自动驾驶中感知决策的潜力,评估结果表明,GPT-4V在一些复杂的环境中展现出了良好的性能,但由于无法针对驾驶场景微调,GPT-4V只能进行上下文学习,因此在驾驶任务中表现不佳。针对VLM在自动驾驶中的感知与决策能力结合的问题,Tian等^[47]提出了结合VLM的感知决策一体化方法DriveVLM,该方法主要包括3个模块:场景描述、场景分析和分层规划。场景描述模块通过自然语言生成技术将驾驶环境中的关键元素用文本进行描述;场景分析模块对这些文本描述进行分析,评估环境元素属性,为决策提供充足的信息支

持；分层规划模块通过元动作、决策描述和轨迹点3种规划目标，将评估信息转化为具体的驾驶策略，系统根据规划的策略执行驾驶操作。为了测试场景理解与规划能力，Tian等^[47]提出了SUP-AD数据集及新的评估指标，SUP-AD数据集中的驾驶场景描述测试实验表明，DriveVLM场景描述达到了0.71，元动作得分达到了0.37，相较于GPT-4V在同样两个项目中分别获得的得分0.38和0.19，有显著的提升。尽管感知能力较强，VLM在空间推理速度和计算效率方面仍有不足。针对该问题，Tian等^[47]提出了DriveVLM-Dual，通过整合基于规则和模型驱动的传统自动驾驶策略，提升了系统对环境翻译的实时性。例如，通过与VAD的融合，DriveVLM能够通过分析来自不同驾驶场景的感知数据进一步优化其模型。由于视觉语言模型并不适合进行精确的数学计算，因此，使用它们来预测轨迹点可能会导致性能欠佳。针对轨迹点预测不准的问题，Jiang等^[48]提出了Senna方法，通过接收多个视角的图像和与驾驶任务相关的文本信息，对图像进行特征提取和文本编码，通过大语言模型（large language model, LLM，简称大模型）生成高层次的元动作决策，随后，Senna-E2E模块根据这些元动作进行低层轨迹规划，生成详细的运动轨迹，确保在多变的驾驶环境中能够有效执行。

融合多传感器信息的感知决策一体化方法也在实际驾驶场景中得到了应用，如理想汽车提出的端到端双系统大模型，通过VLM来模拟人类的思考过程，包含快速的直觉性思考和慢速的逻辑性思考。大模型中，系统1通过一体化大模型直接处理

来自摄像头和LiDAR的多模态数据，主要负责快速响应和处理简单的驾驶任务，系统2通过VLM处理更复杂的逻辑思考和决策过程。类似于理想汽车的端到端大模型，零一汽车端到端大模型通过多模态大语言模型的解码摄像头和导航信息，产生规划信号和逻辑推理信息，实现了对整车的控制。与生成式端到端方案不同，小米汽车提出的筛选式轨迹决策方案通过融合视觉与LiDAR信息，实时生成多条路径，最终选取一条最安全且最高效的路径。华为提出的ADS3.0通过融合视觉、毫米波雷达和LiDAR信息，利用通用障碍物检测（general obstacle detection, GOD）大网整合了障碍物检测和决策与规划，实现了对真实驾驶环境中复杂路况的处理。在自动驾驶场景中应用端到端方案的相关厂商如图3所示，部分厂商可能包含多种类型的方案，本文仅针对其中一种有代表性的方案进行讨论。

2.3 基于多智能体协同的感知决策一体化方法

基于单车多传感器模态的感知决策一体化方法不能利用周围其他智能体的信息，因此一些研究者开展了基于多智能体协同的感知决策一体化方法研究，主要采用车联网（vehicle to everything, V2X）技术，通过增强单个车辆的局限视野，提高自动驾驶系统对环境的适应能力。

Valiente等^[49]提出了一种通过CNN和长短期记忆网络直接输出转向角度的方法CSA，该方法通过车辆间的V2V信息共享，同时综合前方车辆和本车的图像数据，捕捉图像序列中的时序依赖性，输出本车的转向角度。Cui等^[50]提出了CooperNaut

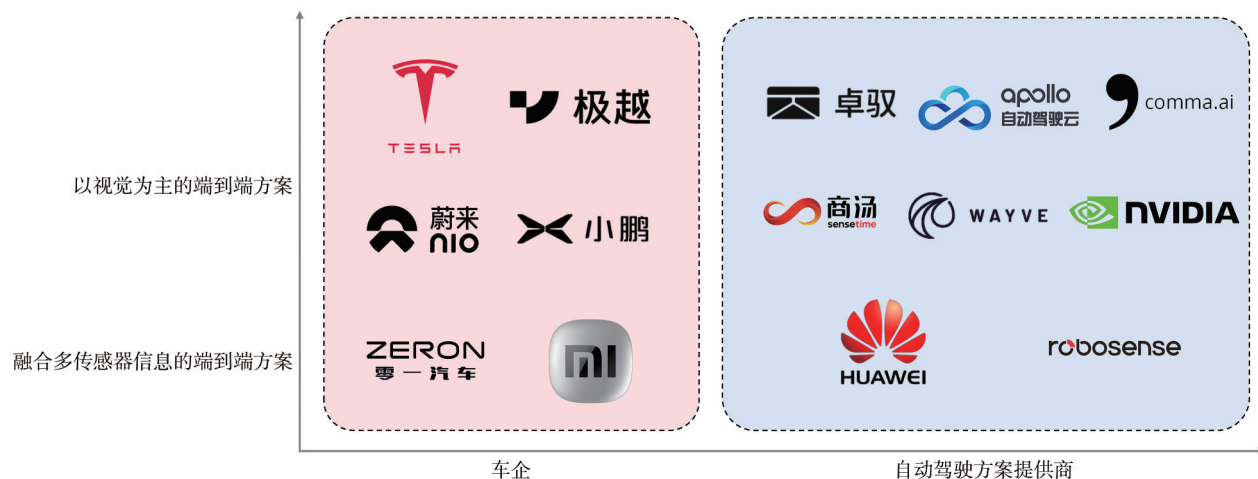


图3 在自动驾驶场景中应用端到端方案的相关厂商

方法,通过点编码器模块将LiDAR的原始点云数据处理为紧凑的特征表示,以减轻数据传输的带宽压力,利用空间变换和自注意力机制对来自其他车辆的信息进行聚合,增强对驾驶环境的全面理解,最后控制模块直接生成油门、刹车等控制指令。与CooperNaut利用LiDAR数据的方式不同,Deng等^[51]提出了V2X-Lead方法,采用无模型与离策略的深度强化学习算法,通过设定控制目标和设计多任务奖励函数,训练自动驾驶代理以提高在不同交通场景下的安全性和效率。现有的方法是在理想通信情况下提出的,缺乏在通信故障情况下对协作感知系统的改进方法。针对此,Ren等^[52]提出了V2X沟通中断感知的多智能体协同感知方法V2X-INCOP,通过设计通信自适应历史信息生成机制及多尺度时空预测模型,在通信中断的情况下恢复缺失的信息,利用知识蒸馏技术提升模型性能,采用课程学习策略来增强训练过程的稳定性和有效性。以上这些方法依赖于一种基础方法,使用简单的网络来优化规划和控制输出,缺乏对可解释性和泛化性的具体解决方法。Yu等^[53]提出了基于V2X的端到端自动驾驶框架UniV2X,将关键任务整合到单一模型中,优化最终的路径规划及各个环节的协同表现,该框架设计了一种稀疏-密集混合数据传输方法,以平衡车辆与基础设施之间的数据传输效率与信息量,在减少数据传输量的同时,也使模型在整体上具备了可解释性。此外,VLM在多智能体协同的感知决策一体化方案中提升了对环境信息的捕捉和解析能力,You等^[54]提出了结合VLM的V2X-VLM方法,通过车辆摄像头和基础设施传感器同时收集环境数据,构建了交通标志信息、路况说明等描述性文本的上下文,提取出更丰富的场景表示,实现了基于提取特征的决策制定和轨迹规划。

平行驾驶理论在多智能体协同的端到端自动驾驶中也有很大的潜力。平行驾驶的理论基础是平行系统理论^[55-56],该理论的架构由一个自然存在的现实系统以及与之相对应的一个或多个虚拟或理想的人工系统共同整合而成,二者通过平行执行协同运作形成有机统一体。平行系统理论在多智能体协同驾驶中展现了显著的发展潜力^[57],系统通过构建与实际物理世界对应的虚拟世界,对实际车辆和虚拟车辆进行同步运行和交互,实现对车辆的高效、安全控制。在平行驾驶系统中,实际车辆在物理世界中行驶,而虚拟车辆在虚拟世界中运行,两者相

互映射、相互影响。刘腾等^[58]提出了基于数字四胞胎的平行驾驶系统,将平行驾驶系统分为4个组成部分,其中包括一辆真实车辆和3个虚拟“守护天使”,这3个虚拟车辆分别是描述车辆、预测车辆和引导车辆,分别负责构建真实车辆及其环境的模型、进行决策及规划分析,以及指导真实车辆的行动。平行驾驶系统在矿山、物流等场景中已经被成功应用。在矿山场景中,平行驾驶系统通过虚拟平行矿山系统与真实的矿山设备相结合,实现了实时数据采集与分析,提升了安全作业效率。在物流场景中,无人驾驶物流车能够借助远程接管平台进行实时监控并实现主动避障,中心管控平台会构建虚拟交通场景,进行大量计算试验,以此来帮助优化车辆在实际驾驶环境中的表现,确保驾驶决策的安全性。在虚拟环境中,车辆与真实车辆的互动大幅提升了安全性,使无人驾驶物流车能够应对各种复杂驾驶场景。

上述3种自动驾驶方法对比如图4所示,其中(a)指以视觉为主的感知决策一体化方法感知部分,(b)指融合多传感器信息的感知决策一体化方法感知部分,(c)指基于多智能体协同的感知决策一体化方法感知部分,(d)指决策部分。

前文提到的部分方法的测试结果汇总见表1,NAVSIM测试结果汇总见表2,DAIR-V2X测试结果汇总见表3。

3 相关数据集

为了促进自动驾驶研究的进展,建设大规模的公共数据集已成为必然要求。当前,自动驾驶研究中常用的公共数据集有nuScenes数据集、KITTI数据集和DAIR-V2X数据集等。下面将从这些数据集包含的数据量大小、传感器、涵盖场景等方面对它们进行介绍。

3.1 单车多传感器数据集

(1) nuScenes数据集^[59]

nuScenes数据集专注于复杂城市环境中的感知与决策任务。它包含1 000个驾驶场景,覆盖波士顿和新加坡市两个城市,提供了丰富的传感器数据,包括6个摄像头、5个超声波雷达和1个激光雷达,以及全球定位系统(global positioning system, GPS)和IMU数据。该数据集标注了超过23个对象类别,并提供了140万个目标的3D边界框和对象属性信息。

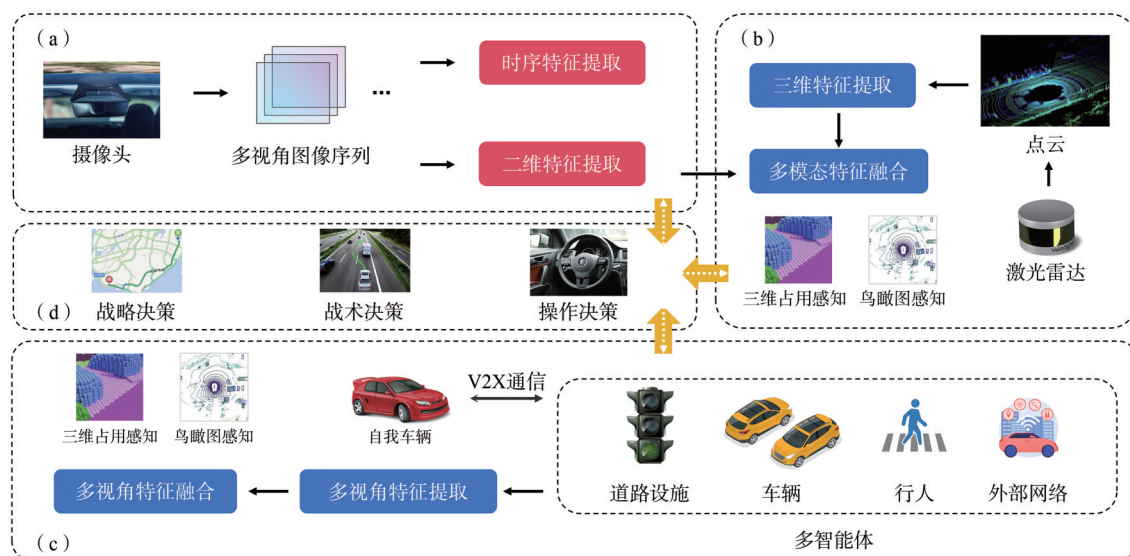


图4 3种自动驾驶方法对比

表1 部分方法的测试结果汇总

方法	模态	nuScenes		CARLA Town 05 Long			CARLA Town 05 Short		CARLA Longest6		
		L2 误差/m ↓	碰撞概率 ↓	驾驶评分 ↑	路线完成率 ↑	违规评分 ↑	驾驶评分 ↑	路线完成率 ↑	驾驶评分 ↑	路线完成率 ↑	违规评分 ↑
UniAD ^[31]	视觉	1.03	0.31%	—	—	—	—	—	—	—	—
VAD ^[32]	视觉	0.72	0.22%	30.31	75.20%	—	64.29	87.26%	—	—	—
GenAD ^[34]	视觉	0.91	0.43%	—	—	—	—	—	—	—	—
VAD V2 ^[33]	视觉	—	—	85.10	98.40%	0.87	89.70	93.0%	—	—	—
DualAD ^[38]	视觉	1.03	0.17%	—	—	—	—	—	—	—	—
DualVAD ^[38]	视觉	0.55	0.22%	—	—	—	—	—	—	—	—
Transfer ^[4]	多模态	—	—	31.0	47.5%	0.77	54.50	78.4%	46.95	89.64%	0.52
M2DA ^[43]	多模态	—	—	72.6	89.7%	0.80	—	—	—	—	—
MaskFuser ^[21]	多模态	—	—	—	—	—	—	—	49.05	92.85%	0.56
DriveVLM ^[47]	多模态	0.40	0.27%	—	—	—	—	—	—	—	—
DriveVLM-Dual ^{*[47]}	多模态	0.31	0.10%	—	—	—	—	—	—	—	—
PlanKD ^[42] (InterFuser)	多模态	—	—	—	—	—	93.69	96.43%	—	—	—
PlanKD ^[42] (TCP)	多模态	—	—	—	—	—	95.49	99.49%	—	—	—

注：*为与VAD协作的测试结果。

表2 NAVSIM测试结果汇总

方法	模态	NAVSIM					
		无碰撞指标 ↑	动作命中率 ↑	车辆行驶进度 ↑	预计碰撞时间 ↑	分类准确率 ↑	得分 ↑
HydraMDP ^{*[41]}	多模态	98.3%	96.0%	78.7%	94.6	100%	86.5
DRAMA ^[45]	多模态	98.0%	93.1%	80.1%	94.8	100%	85.5

注：*为Hydra-MDP-V8192-W-EP的测试结果。

(2) GTAV数据集^[60]

GTAV数据集是基于游戏《侠盗猎车手5》(Grand Theft Auto V, GTA5) 创建的一个合成三维数据集, 旨在为自动驾驶和计算机视觉研究提供丰富的资源。该数据集利用GTAV的游戏引擎生成

了仿真驾驶环境, 包含1 680×1 050 高分辨率的图像与深度图, 支持2D和3D车辆边界框标注。

(3) Synscapes数据集^[61]

Synscapes是一个专注于街道场景的合成数据集, 包含25 000张街景图像, 运用基于物理的渲染

表3 DAIR-V2X测试结果汇总

方法	数据传输技术	DAIR-V2X		
		L2 误差/m ↓	碰撞率 ↓	传输成本 ↓
V2X-VLM ^[54]	V2X	1.22	0.01%	1.24×10^7
V2X-VLM-Fast ^[54]	V2X	3.15	0.02%	1.24×10^7
UniV2X ^[53]	V2X	3.43	0.49%	8.09×10^5
CooperNaut ^[50]	V2V	5.26	1.38%	8.19×10^7

技术确保了高真实性和细节还原，其标注信息涵盖语义分割和深度图等。

(4) UrbanSyn 数据集^[62]

UrbanSyn 是一个合成的大规模城市自动驾驶数据集，模拟了城市驾驶场景，包含 75 000 张图像、语义分割注释、深度图以及 3D 目标框注释，支持目标检测、语义分割和深度估计等任务。

(5) ONCE 数据集^[63]

ONCE 数据集是一个自动驾驶场景 3D 物体检测数据集，包含 100 万个激光雷达场景和 700 万个对应的相机图像，记录了 144 h 的驾驶数据，覆盖 200 km² 的区域，并提供了 15 000 个完全注释的场景，包含多种环境条件，如昼夜、晴雨及城市郊区。

(6) KITTI 数据集^[64]

KITTI 数据集提供了市区、乡村和高速公路等多种场景下的真实图像数据，涵盖 389 对立体图像、39.2 km 的视觉测距序列，以及超过 20 万个 3D 标注物体的数据。KITTI 数据集的数据采集平台使用了多种传感器，包含灰度摄像机、3D 激光雷达和 GPS 导航，确保了数据的高质量和多模态特性。

(7) Cityscapes 数据集^[65]

Cityscapes 数据集是一个关注城市街道场景语义理解的大规模数据集，涵盖来自 50 个不同城市的多种立体视频序列，共计 5 000 帧高质量像素级注释和 20 000 个弱注释帧。该数据集提供了两套评测标准，其中 fine 标准包含 5 000 张精细标注的图像，coarse 标准则包含 5 000 张精细标注图像及 20 000 张粗糙标注图像。Cityscapes 数据集有 19 个类别的密集像素标注，覆盖率达到 97%，其中 8 个类别还具备实例级分割。

3.2 多智能体协同数据集

(1) OPV2V 数据集^[66]

OPV2V 数据集利用 OpenCDA^[67] 和 CARLA^[68] 仿真框架收集了 73 个场景，涵盖 6 种道路类型及 9 个城市，包含 1.2 万帧 LiDAR 点云和 4.8 万张 RGB 图像，标注了 23 万个 3D 检测框。OPV2V 数据集提

供了包括 4 种不同的点云检测 Backbone（如 PointPillar^[69]、VoxelNet^[70] 等）方法和多种信息融合策略（如单车无融合、点云融合和中间层融合），总计涵盖 16 个模型。此外，OPV2V 数据集开源了首个协同感知代码框架 OpenCOOD，提供 API 与多个模型代码，支持常见的多智能体融合策略。

(2) V2V4Real 数据集^[71]

V2V4Real 数据集涵盖 410 km 的驾驶区域，记录了两辆 L3 级别采集车在多种道路环境下的真实数据，包括激光雷达帧、RGB 图像和高清地图信息。数据集中标注了 80 万帧中的 2 万帧，提供了高精度的 3D 边界框标注，涵盖汽车、面包车、皮卡、半挂车和公共汽车等多种物体类别。

(3) DAIR-V2X 数据集^[72]

DAIR-V2X 数据集是一个基于真实场景的车路协同自动驾驶数据集，涵盖北京市高级别自动驾驶示范区的真实城市道路和高速公路场景。数据集包含 71 254 帧图像和点云数据，结合了车端和路端相机及激光雷达的信息，提供了多模态和多视角的数据。

(4) V2X-Sim 数据集^[73]

V2X-Sim 数据集利用 CARLA-SUMO 协同仿真环境，提供了高度真实的仿真数据。每辆车均配备有 6 个 RGB 摄像头、6 个深度摄像头、6 个语义分割摄像头、1 个 BEV 摄像头和 1 个 LiDAR 传感器。此外，路边设施上还配备了多个传感器，包括 4 个 RGB 摄像头、1 个 BEV 语义分割摄像头、1 个 LiDAR 传感器和 1 个语义 LiDAR 传感器，支持检测、跟踪和分割等多种感知任务。

(5) V2X-Seq 数据集^[74]

V2X-Seq 数据集提供了超过 1.5 万对时序图像和点云数据，以及完备的 3D 目标标注，支持车路协同检测和跟踪任务。V2X-Seq 还包含了覆盖 28 个真实路口的轨迹预测数据集，该数据集共包含超过 20 万个片段，为常见障碍物目标提供了详细的轨迹信息。

上述数据集详细对比见表 4。

4 性能评估方法

对自动驾驶系统的评估方法主要分为 3 种：现实世界评估、模拟器评估和离线评估。现实世界评估是指在实际道路环境中对自动驾驶车辆进行性能评估，依据指标评估模型的表现；模拟器评估是指

表 4 数据集详细对比

数据集	Agent	场景类型	数据量	传感器
nuScenes ^[59]	单车	城市环境	1 000 个场景	6 个摄像头、5 个雷达、1 个激光雷达、GPS、IMU
GTAV ^[60]	单车	合成城市环境	24 966 张图像	RGB 摄像头
Synscapes ^[61]	单车	街道场景	25 000 张图像	RGB 摄像头
UrbanSyn ^[62]	单车	城市环境	75 000 张图像	RGB 摄像头
ONCE ^[63]	单车	城市和乡村场景	100 万个 LiDAR 场景、700 万张图像	1 个激光雷达、多个 RGB 相机
KITTI ^[64]	单车	城市、乡村和高速公路	14 999 个图像、点云数据	灰度摄像机、激光雷达、GPS
Cityscapes ^[65]	单车	街道场景	25 000 帧图像	RGB 摄像头
OPV2V ^[66]	车与车	城市和郊区场景	73 个场景	LiDAR、RGB 相机
V2V4Real ^[71]	车与车	城市和郊区场景	410 km 驾驶区域	LiDAR、RGB 相机、高清地图
DAIR-V2X ^[72]	多智能体	城市道路、高速公路	72 890 帧图像数据、72 890 帧点云	车端和路端相机、激光雷达
V2X-Sim ^[73]	多智能体	合成场景	多样化场景	车端和路端激光雷达、RGB 摄像头
V2X-Seq ^[74]	多智能体	真实道路场景	95 个场景	RGB 摄像头、激光雷达

利用模拟器生成虚拟环境来测试自动驾驶系统的性能，同样依据指标评估模型的表现；离线评估通过对比人类驾驶数据和模型驾驶数据，根据误差大小来评估模型的表现。3 种评估方法的基本结构如图 5 所示。其中，(a) 指现实世界评估方法，(b) 指模拟器评估方法，(c) 指离线评估方法。

4.1 现实世界评估

现实世界评估方法通过在真实车辆上加装自动驾驶系统，并在公共道路上进行驾驶测试，直接对自动驾驶系统的各项数据指标进行评估。在国内举办的全国智能驾驶测试赛中，参赛车辆需要完成包括低速跟车、红绿灯识别、行人避让等多项测试项目，旨在模拟真实的驾驶场景。除竞赛外，一些专

业组织也为自动驾驶系统在真实环境中的评估制定了相应的测试指标，为自动驾驶车辆在实际应用中的研究提供了参考。《北京市自动驾驶车辆道路测试报告（2019 年）》^[75]中详述了多样化且系统化的自动驾驶系统评估方法，测试过程涉及设置标准化的测试条件和明确的测试指标，并利用真实路测数据进行动态分析，以实时监测车辆的感知、决策和控制能力。此外，美国国家公路交通安全管理局（National Highway Traffic Safety Administration, NHTSA）^[76]也提出了一个系统化的测试框架，该框架强调建立标准化的测试条件、确立明确的测试指标，并强调利用真实路测数据进行动态分析，以全面评估自动驾驶系统在复杂多变交通环境中的安

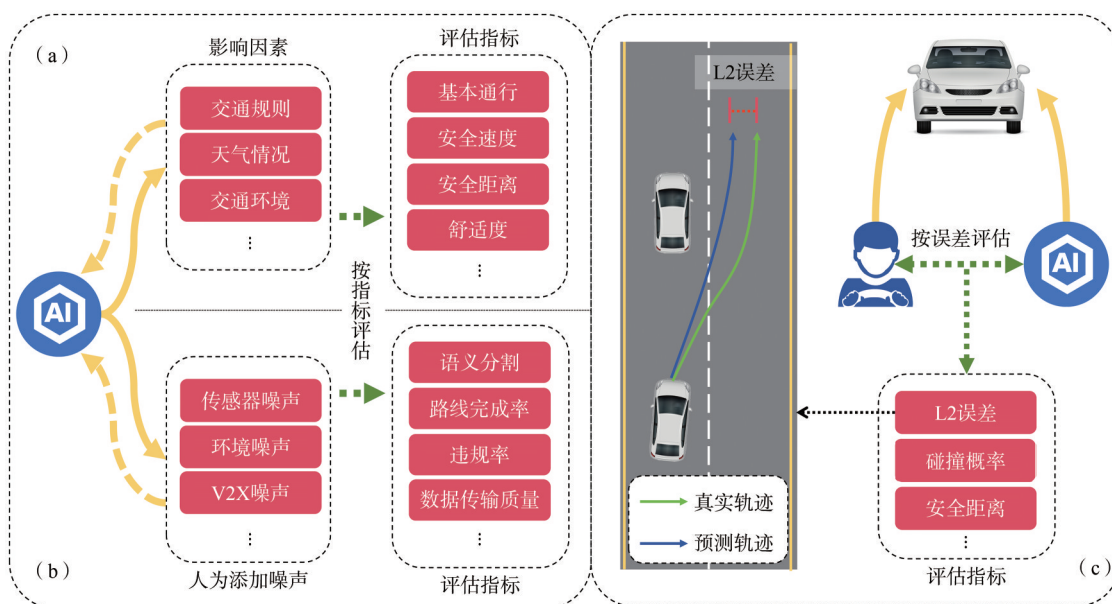


图 5 3 种评估方法的基本结构

全性和可靠性。

4.2 模拟器评估

模拟器评估方法通过在仿真环境中模拟现实世界的各种环境,协助研究者对自动驾驶算法进行研究与验证。相较于现实世界评估方法,模拟器评估方法能够显著降低评估成本,同时减少了对物理车辆和测试场地的依赖。模拟器评估过程大致可分为两个步骤:环境仿真和车辆动力学仿真。首先,模拟器利用从现实世界采集的数据或由程序生成的数据,对自动驾驶车辆的环境进行仿真。然后,模拟器针对车辆的动力学特性进行仿真,以模拟车辆在不同环境条件下的实际表现。这一基本流程已经被众多研究工作所采用,如自动驾驶仿真器CARLA^[68],它利用虚幻图形引擎和物理引擎技术,创建了一个高度真实的仿真环境,结合环境模型与车辆动力学模型,使研究人员能够全面测试自动驾驶算法在复杂场景下的性能和反应。在CARLA V2版本中,系统提供了多项评估指标,主要包括路线完成率、违规处罚,以及基于这两者的乘积得出的驾驶评分。为了评估多智能体协同自动驾驶方案,V2X-Sim^[73]基于CARLA提供了一套高度真实的仿真数据,添加了针对检测、跟踪和分割这3项任务的基准。此外,一些研究者还直接利用电子游戏提供的3D环境进行自动驾驶仿真,如以美国洛杉矶为原型构建虚拟城市洛圣都的游戏GTA5,该游戏为自动驾驶研究提供了丰富的城市场景^[60]。开源项目DeepGTAV则为GTA5游戏提供了相关插件,进一步便利了在模拟环境中对自动驾驶系统的研究。

4.3 离线评估

离线评估方法中,自动驾驶算法并不直接参与对车辆的控制,而是将传感器数据和目标位置作为输入,端到端自动驾驶系统将其预测的未来行驶轨迹与专家驾驶日志中的实际轨迹进行比较,以评估其性能。开环评估的一个显著优点是能够直接利用真实的交通和传感器数据进行测试,避免了对模拟器的依赖,使评估过程更为高效和真实。例如,NAVSIM测试平台、nuScenes数据集等就提供了此类评估方法。NAVSIM测试平台利用简化的BEV来模拟短时间的驾驶场景,评估指标包括动作命中率、车辆行驶进度得分、预计碰撞时间、分类准确率和综合得分。其中,动作命中率评估了代理在执行预定驾驶动作时的成功概率;车辆行驶进度得分

量化了车辆完成路径的能力及所需时间;预计碰撞时间用于评估系统预测潜在碰撞事件发生的时间;分类准确率反映了代理在识别和分类场景中对象的准确程度;综合得分是对以上所有指标的加权汇总。nuScenes数据集则提供了检测、追踪、预测、LiDAR分割、规划等多种评估任务。其中,规划任务有两种常用的指标:L2误差和碰撞概率。L2误差通过计算预测轨迹和真实轨迹之间的L2误差来评估预测轨迹的质量;碰撞概率通过计算预测轨迹与其他物体发生碰撞的概率来衡量自动驾驶系统的安全性。

5 展望

随着自动驾驶技术的快速演进,感知决策一体化的自动驾驶技术逐渐成为领域内的重要研究方向。该技术通过整合感知与决策过程,有效应对复杂多变的驾驶环境与动态需求,提高了系统的响应速度、准确性和鲁棒性。然而,当前的感知决策一体化自动驾驶方法仍存在若干亟待解决的问题。尽管多传感器的集成可显著提升感知性能,但其高昂的成本在一定程度上限制了大规模应用。部分方法虽然通过引入语言模型来增强泛化能力,并提升了端到端模型的可解释性,但随之增加的计算负担削弱了自动驾驶系统的实时性。因此,在具体应用中,研究者需要综合考量成本、计算性能、可解释性、安全性和泛化能力之间的平衡,以推动感知决策一体化自动驾驶技术的落地。未来,感知决策一体化自动驾驶技术将呈现如下发展趋势。

(1) 端到端与大语言模型结合

LLM在自动驾驶中的应用增强了模型对场景的理解能力。目前,这一应用主要分为两种方法:一种是通过问答任务进行驾驶场景理解,另一种则是在基于LLM的场景理解基础上融入规划功能。这两种方法的结合充分发挥了LLM在语义理解、信息生成和决策支持方面的优势,使自动驾驶系统不仅仅是在执行硬编码的动作,而是可以理解复杂的语境,并在此基础上进行自主决策。随着技术的进一步发展和优化,这些模型的应用将更加广泛,为自动驾驶车辆提供更为安全、灵活和高效的操作方式。

(2) 轻量化架构的引入

由于自动驾驶系统对实时性和计算性能的高要求,如何在有限的硬件资源上提高模型的计算效率,将成为重要的研究方向。未来,通过优化模型

架构以及引入轻量化算法,将能够显著提升算力的利用率,从而满足复杂环境下的实时需求。例如,RWKV通过融合循环神经网络的优势,能够高效地处理多模态传感器数据,支持快速的实时推理,适用于动态驾驶环境。Mamba则以线性时间复杂度著称,能够出色地应对长序列数据,确保信息在决策层的快速传递。KAN (Kolmogorov-Arnold network)通过引入可学习的激活函数,增强了对复杂场景的识别能力,适合在复杂交通状况下做出快速反应。未来,感知决策一体化自动驾驶技术将与这些架构进行融合,推动自动驾驶技术在性能与效率上实现进一步提升。

(3) 超视距导航能力

视距导航指的是车辆在当前感知范围之外,基于环境模型和历史数据来预测未来数秒甚至数十秒内的驾驶环境变化。随着V2X技术的发展,端到端自动驾驶系统能够实现低时延的数据交换,从而使超视距导航信息得以实时更新。这将使系统在高速行驶时也能快速响应外部变化,增强了驾驶的安全性与稳定性。

(4) 具身智能驱动

区别于传统自动驾驶系统通常依赖的静态场景模型,具身智能驱动下的自动驾驶技术将在传感器技术和多模态数据融合方面实现更高的成熟度。随着对动态场景理解的深入,未来的智能驾驶系统将集成领域自适应技术和模型无关学习,使系统在面对数据稀缺和高噪声环境时,仍能保持较高的泛化性能。

(5) 平行驾驶的进一步扩展

V2X技术的发展为平行驾驶提供了数据支持。在平行驾驶中,虚拟车可以对V2X获取的大量数据进行分析 and 处理,挖掘出更有价值的信息。同时,智能代理可以根据V2X提供的信息,及时调整虚拟车的行为和策略,使其更好地适应实际交通环境的变化,进而为真实车提供更精准的控制建议和服务,提升V2X在实际驾驶中的应用价值。

6 结束语

综上所述,以视觉为主的感知决策一体化方法基于人类视觉处理机制,通过直接提取图像来实现驾驶决策,提升了自动驾驶的规划性能。融合多传感器信息的感知决策一体化方法则通过整合多种传感器模态的数据,表现出了良好的效果。基于多智

能体协同的感知决策一体化方法借助V2X技术整合周围交通参与者的数据,打破了单个车辆视野的局限,进一步提升了自动驾驶系统对复杂交通环境的适应性。本文梳理了感知决策一体化技术在自动驾驶领域的最新研究进展和方法,旨在为后续研究提供参考和启示,助力推动该领域的持续创新与发展。

参考文献:

- [1] POMERLEAU D A. ALVINN an autonomous land vehicle in a neural network[C]//Advances in Neural Information Processing Systems 1. Massachusetts: MIT Press, 1989: 305-313.
- [2] BOJARSKI M, YERES P, CHOROMANSKA A, et al. Explaining how a deep neural network trained with end-to-end learning steers a car[EB]. arXiv preprint, 2017, arXiv: 1704.07911.
- [3] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. Advances in Neural Information Processing Systems, 2017, 30: 5998-6008.
- [4] PRAKASH A, CHITTA K, GEIGER A, et al. Multi-modal fusion transformer for end-to-end autonomous driving[EB]. arXiv preprint, 2021, arXiv: 2104.09224.
- [5] RAHIMIAN P, O'NEAL E E, ZHOU S W, et al. Harnessing vehicle-to-pedestrian (V2P) communication technology: sending traffic warnings to texting pedestrians[J]. Human Factors, 2018, 60(6): 833-843.
- [6] 廖勇, 尹子松, 田肖懿. 车联网V2I场景下基于GNN的SC-FDMA智能信道估计[J]. 电子学报, 2024, 52(3): 772-782.
- [7] LIAO Y, YIN Z S, TIAN X Y. Intelligent channel estimation of SC-FDMA based on GNN for V2I scenarios in Internet of vehicles[J]. Acta Electronica Sinica, 2024, 52(3): 772-782.
- [7] 李原, 潘定海, 王志刚, 等. 基于V2I/V2N的感知融合系统技术及应用研究[J]. 工业控制计算机, 2022, 35(2): 5-7, 10.
- [8] LI Y, PAN D H, WANG Z G, et al. Perception fusion system based on V2I/V2N[J]. Industrial Control Computer, 2022, 35(2): 5-7, 10.
- [8] XU H Y, CHEN J L, MENG S Y, et al. A survey on occupancy perception for autonomous driving: the information fusion perspective[J]. Information Fusion, 2025, 114: 102671.
- [9] WU P X, CHEN S H, METAXAS D N. MotionNet: joint perception and motion prediction for autonomous driving based on bird's eye view maps[C]//Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2020: 11234-11243.
- [10] PENG L, XU J K, CHENG H R, et al. Learning occupancy for monocular 3D object detection[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 10281-10292.
- [11] 朱冰, 贾士政, 赵健, 等. 自动驾驶车辆决策与规划研究综述[J]. 中国公路学报, 2024, 37(1): 215-240.
- [11] ZHU B, JIA S Z, ZHAO J, et al. Review of research on decision-making and planning for automated vehicles[J]. China Journal of Highway and Transport, 2024, 37(1): 215-240.
- [12] ZYNER A, WORRALL S, WARD J, et al. Long short term memory for driver intent prediction[C]//Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2017: 1484-1489.

- [13] DING W C, SHEN S J. Online vehicle trajectory prediction using policy anticipation network and optimization-based context reasoning[C]//Proceedings of the 2019 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2019: 2435-2441.
- [14] SUN H B, CHEN R F, LIU T Y, et al. LG-LSTM: modeling LSTM-based interactions for multi-agent trajectory prediction[C]//Proceedings of the 2022 IEEE International Conference on Multimedia and Expo (ICME). Piscataway: IEEE Press, 2022: 1-6.
- [15] POLYCHRONOPOULOS A, TSOGAS M, AMDITIS A J, et al. Sensor fusion for predicting vehicles' path for collision avoidance systems[J]. IEEE Transactions on Intelligent Transportation Systems, 2007, 8(3): 549-562.
- [16] BATZ T, WATSON K, BEYERER J. Recognition of dangerous situations within a cooperative group of vehicles[C]//Proceedings of the 2009 IEEE Intelligent Vehicles Symposium. Piscataway: IEEE Press, 2009: 907-912.
- [17] LIU W, KIM S W, PENDLETON S, et al. Situation-aware decision making for autonomous driving on urban road using online POMDP[C]//Proceedings of the 2015 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2015: 1126-1133.
- [18] SHADRIN S S, IVANOVA A A. Analytical review of standard SAE J3016 taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles with latest updates[J]. Avtomobil Doroga Infrastruktura, 2019, 3(21): 10.
- [19] 付新科, 蔡英凤, 陈龙, 等. 不确定性环境下的自动驾驶汽车行为决策方法[J]. 汽车工程, 2024, 46(2): 211-221, 259.
FU X K, CAI Y F, CHEN L, et al. Decision-making for autonomous driving in uncertain environment[J]. Automotive Engineering, 2024, 46(2): 211-221, 259.
- [20] ZHAO L H, ICHISE R, SASAKI Y, et al. Fast decision making using ontology-based knowledge base[C]//Proceedings of the 2016 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2016: 173-178.
- [21] OLSSON M. Behavior trees for decision-making in autonomous driving[R]. 2016.
- [22] MNIH V, KAVUKCUOGLU K, SILVER D, et al. Human-level control through deep reinforcement learning[J]. Nature, 2015, 518(7540): 529-533.
- [23] GLÄSCHER J, DAW N, DAYAN P, et al. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning[J]. Neuron, 2010, 66(4): 585-595.
- [24] BOJARSKI M, DEL TESTA D, DWORAKOWSKI D, et al. End to end learning for self-driving cars[EB]. arXiv preprint, 2016, arXiv: 1604.07316.
- [25] LILLICRAP T P, HUNT J J, PRITZEL A, et al. Continuous control with deep reinforcement learning[EB]. arXiv preprint, 2015, arXiv: 1509.02971.
- [26] TENG S Y, HU X M, DENG P, et al. Motion planning for autonomous driving: the state of the art and future perspectives[J]. IEEE Transactions on Intelligent Vehicles, 2023, 8(6): 3692-3711.
- [27] CHEN Z L, HUANG X M. End-to-end learning for lane keeping of self-driving cars[C]//Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2017: 1856-1860.
- [28] YANG Z Y, ZHANG Y X, YU J, et al. End-to-end multi-modal multi-task vehicle control for self-driving cars with visual perceptions[C]//Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR). Piscataway: IEEE Press, 2018: 2289-2294.
- [29] CHI L, MU Y D. Deep steering: learning end-to-end driving model from spatial and temporal visual cues[EB]. arXiv preprint, 2017, arXiv: 1708.03798.
- [30] AN D D, LIU J, ZHANG M, et al. Uncertainty modeling and runtime verification for autonomous vehicles driving control: a machine learning-based approach[J]. Journal of Systems and Software, 2020, 167: 110617.
- [31] HU Y H, YANG J Z, CHEN L, et al. Planning-oriented autonomous driving[EB]. arXiv preprint, 2022, arXiv: 2212.10156.
- [32] JIANG B, CHEN S Y, XU Q, et al. VAD: vectorized scene representation for efficient autonomous driving[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 8306-8316.
- [33] CHEN S Y, JIANG B, GAO H, et al. VADv2: end-to-end vectorized autonomous driving via probabilistic planning[EB]. arXiv preprint, 2024, arXiv: 2402.13243.
- [34] ZHENG W Z, SONG R Q, GUO X D, et al. GenAD: generative end-to-end autonomous driving[EB]. arXiv preprint, 2024, arXiv: 2402.11502.
- [35] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 770-778.
- [36] ZHENG W Z, WU J J, ZHENG Y, et al. GaussianAD: Gaussian-centric end-to-end autonomous driving[EB]. arXiv preprint, 2024, arXiv: 2412.10371.
- [37] WENG X S, IVANOVIC B, WANG Y, et al. PARA-drive: parallelized architecture for real-time autonomous driving[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 15449-15458.
- [38] DOLL S, HANSELMANN N, SCHNEIDER L, et al. DualAD: disentangling the dynamic and static world for end-to-end driving[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 14728-14737.
- [39] JAEGER B, CHITTA K, GEIGER A. Hidden biases of end-to-end driving models[C]//Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway: IEEE Press, 2023: 8206-8215.
- [40] DUAN Y Q, GUO X D, ZHU Z, et al. MaskFuser: masked fusion of joint multi-modal tokenization for end-to-end autonomous driving[EB]. arXiv preprint, 2024, arXiv: 2405.07573.
- [41] LI Z X, LI K L, WANG S H, et al. Hydra-MDP: end-to-end multi-modal planning with multi-target hydra-distillation[EB]. arXiv preprint, 2024, arXiv: 2406.06978.
- [42] FENG K T, LI C S, REN D C, et al. On the road to portability: compressing end-to-end motion planner for autonomous driving[C]//Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2024: 15099-15108.
- [43] XU D Y, LI H K, WANG Q F, et al. M2DA: multi-modal fusion transformer incorporating driver attention for autonomous driving[EB]. arXiv preprint, 2024, arXiv: 2403.12552.
- [44] GU A, DAO T. Mamba: linear-time sequence modeling with selective state spaces[EB]. arXiv preprint, 2023, arXiv: 2312.00752.
- [45] YUAN C R, ZHANG Z Q, SUN J W, et al. DRAMA: an efficient end-to-end motion planner for autonomous driving with mamba[EB]. arXiv preprint, 2024, arXiv: 2408.03601.
- [46] WEN L C, YANG X M, FU D C, et al. On the road with GPT-4V (ision): early explorations of visual-language model on autonomous driving[EB]. arXiv preprint, 2023, arXiv: 2311.05332.
- [47] TIAN X Y, GU J R, LI B L, et al. DriveVLM: the convergence of autonomous driving and large vision-language models[EB]. arXiv preprint, 2024, arXiv: 2402.12289.

- [48] JIANG B, CHEN S Y, LIAO B C, et al. Senna: bridging large vision-language models and end-to-end autonomous driving[EB]. arXiv preprint, 2024. arXiv preprint arXiv:2410.22313.
- [49] VALIENTE R, ZAMAN M, OZER S, et al. Controlling steering angle for cooperative self-driving vehicles utilizing CNN and LSTM-based deep networks[C]//Proceedings of the 2019 IEEE Intelligent Vehicles Symposium (IV). Piscataway: IEEE Press, 2019: 2423-2428.
- [50] CUI J X, QIU H, CHEN D, et al. Coopernaut: end-to-end driving with cooperative perception for networked vehicles[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 17231-17241.
- [51] 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.
- [52] REN S L, LEI Z X, WANG Z, et al. Interruption-aware cooperative perception for V2X communication-aided autonomous driving[J]. IEEE Transactions on Intelligent Vehicles, 2024, 9(4): 4698-4714.
- [53] YU H B, YANG W X, ZHONG J R, et al. End-to-end autonomous driving through V2X cooperation[EB]. arXiv preprint, 2024, arXiv: 2404.00717.
- [54] YOU J W, SHI H T, JIANG Z Y, et al. V2X-VLM: end-to-end V2X cooperative autonomous driving through large vision-language models[EB]. arXiv preprint, 2024, arXiv: 2408.09251.
- [55] WANG F Y, WANG C H. Agent-based control systems for operation and management of intelligent network-enabled devices[C]//Proceedings of 2003 IEEE International Conference on Systems, Man and Cybernetics. Conference Theme - System Security and Assurance. Piscataway: IEEE Press, 2003: 5028-5033.
- [56] 王飞跃. 平行系统方法与复杂系统的管理和控制[J]. 控制与决策, 2004, 19(5): 485-489, 514.
WANG F Y. Parallel system methods for management and control of complex systems[J]. Control and Decision, 2004, 19(5): 485-489, 514.
- [57] WANG F Y, ZHENG N N, CAO D P, et al. Parallel driving in CPSS: a unified approach for transport automation and vehicle intelligence[J]. IEEE/CAA Journal of Automatica Sinica, 2017, 4(4): 577-587.
- [58] 刘腾, 王晓, 邢阳, 等. 基于数字四胞胎的平行驾驶系统及应用[J]. 智能科学与技术学报, 2019, 1(1): 40-51.
LIU T, WANG X, XING Y, et al. Research on digital quadruplets in cyber-physical-social space-based parallel driving[J]. Chinese Journal of Intelligent Science and Technology, 2019, 1(1): 40-51.
- [59] 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: 11621-11631.
- [60] RICHTER S R, VINEET V, ROTH S, et al. Playing for data: ground truth from computer games[C]//Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 102-118.
- [61] WRENNINGE M, UNGER J. Synscapes: a photorealistic synthetic dataset for street scene parsing[EB]. arXiv preprint, 2018, arXiv: 1810.08705.
- [62] GÓMEZ J L, SILVA M, SEOANE A, et al. All for one, and one for all: UrbanSyn dataset, the third musketeer of synthetic driving scenes[EB]. arXiv preprint, 2023, arXiv: 2312.12176.
- [63] MAO J G, NIU M Z, JIANG C H, et al. One million scenes for autonomous driving: once dataset[EB]. arXiv preprint, 2021, arXiv: 2106.11037.
- [64] GEIGER A, LENZ P, STILLER C, et al. Vision meets robotics: the KITTI dataset[J]. International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [65] CORDTS M, OMRAN M, RAMOS S, et al. The cityscapes dataset for semantic urban scene understanding[EB]. arXiv preprint, 2016, arXiv: 1604.01685.
- [66] XU R S, XIANG H, XIA X, et al. OPV. V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication[C]//Proceedings of the 2022 International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2022: 2583-2589.
- [67] XU R S, GUO Y, HAN X, et al. OpenCDA: an open cooperative driving automation framework integrated with co-simulation[C]//Proceedings of the 2021 IEEE International Intelligent Transportation Systems Conference (ITSC). Piscataway: IEEE, 2021: 1155-1162.
- [68] DOSOVITSKIY A, ROS G, CODEVILLA F, et al. CARLA: an open urban driving simulator[EB]. arXiv preprint, 2017, arXiv: 1711.03938.
- [69] LANG A H, VORA S, CAESAR H, et al. PointPillars: fast encoders for object detection from point clouds[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 12689-12697.
- [70] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE Press, 2018: 4490-4499.
- [71] XU R S, XIA X, LI J L, et al. V. V. Real: a real-world large-scale dataset for vehicle-to-vehicle cooperative perception[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 13712-13722.
- [72] YU H B, LUO Y Z, SHU M, et al. DAIR-V2X: a large-scale dataset for vehicle-infrastructure cooperative 3D object detection[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 21329-21338.
- [73] LI Y M, MA D K, AN Z Y, et al. V2X-Sim: multi-agent collaborative perception dataset and benchmark for autonomous driving[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 10914-10921.
- [74] YU H B, YANG W X, RUAN H Z, et al. V2X-seq: a large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting[C]//Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2023: 5486-5495.
- [75] 吴琼. 北京自动驾驶车辆道路测试报告(2019)[J]. 智能网联汽车, 2020(2): 46-55.
WU Q. Road test report of self-driving vehicles in Beijing (2019)[J]. Intelligent Connected Vehicles, 2020(2): 46-55.
- [76] THORN E, KIMMEL S C, CHAKA M, et al. A framework for automated driving system testable cases and scenarios[R]. 2018.

[作者简介]



刘泽禹(2000—),男,北京交通大学计算机科学与技术学院硕士生,主要研究方向为多模态融合感知。



张慧 (1993-), 女, 北京交通大学计算机科学与技术学院副教授, 主要研究方向为复杂环境下的车辆感知、多传感器融合的协同检测、多模态目标检测、群体智能、平行视觉等。



李滢东 (1982-), 男, 北京交通大学计算机科学与技术学院教授、院长、博士生导师, 交通大数据与人工智能教育部重点实验室主任, 主要研究方向为大数据智能、数据隐私保护、先进计算、智能交通等。