

面向自动驾驶协同感知的高效传输方法综述

杜少毅¹, 孙远¹, 刘宇颖¹, 曹俱睿¹, 卓帅臣¹, 田智强¹, 郑心湖²

(1. 西安交通大学人工智能学院, 710049, 西安;

2. 香港科技大学(广州)系统枢纽智能交通学域, 511453, 广州)

摘要: 协同感知技术依托智能网联汽车与车路云协同网络,通过多智能体间的信息交互,有效弥补了单车智能在遮挡、远距离探测等长尾场景下的感知盲区。然而,受限的无线通信带宽与多源异构数据已成为系统规模化部署的核心瓶颈,因此围绕这一问题系统阐述了面向自动驾驶协同感知的通信高效传输方法。首先,从空间区域级和实例任务级两个层次,归纳了抑制冗余背景传输的特征筛选机制。其次,围绕交互拓扑构建、时空位姿对齐与跨模态语义对齐等多模态信息融合关键问题,总结了应对节点异构、时间异步与空间错位的代表性方法。然后,从特征压缩稀疏表示、生成式补充与大模型语义协同三条路径,梳理了带宽约束下兼顾通信效率与感知精度的实现方式。最后,系统整理了该领域的大规模开源数据集,并深入分析了当前在场景认知、动态网络安全以及生成式重构可信性等方面面临的瓶颈与未来发展趋势。该文旨在系统梳理相关研究进展,为构建更加鲁棒、安全的智能网联车路协同网络提供理论参考。

关键词: 自动驾驶;协同感知;智能网联汽车;多模态信息融合;通信效率

中图分类号: TP183 **文献标志码:** A

DOI: 10.7652/xjtuxb202610001 **文章编号:** 0253-987X(2026)10-0001-21

A Review of Efficient Transmission Methods for Cooperative Perception in Autonomous Driving

DU Shaoyi¹, SUN Yuan¹, LIU Yuying¹, CAO Jurui¹, ZHUO Shuaichen¹,
TIAN Zhiqiang¹, ZHENG Xinhui²

(1. College of Artificial Intelligence, Xi'an Jiaotong University, Xi'an 710049, China; 2. Intelligent Transportation Thrust of the Systems Hub, Hong Kong University of Science and Technology (Guangzhou), Guangzhou 511453, China)

Abstract: Based on intelligent connected vehicles and vehicle-road-cloud cooperative networks, the perception blind spots of single-vehicle intelligence in long-tail scenarios, such as occlusion and long-range detection, are effectively compensated by cooperative perception technology through information interaction among multiple agents. However, the large-scale deployment of such systems is critically bottlenecked by limited wireless communication bandwidth and multi-source heterogeneous data. To address this issue, communication-efficient transmission methods for cooperative perception in autonomous driving are systematically reviewed. First, feature filtering mechanisms for suppressing redundant background transmission are summarized at both the spatial region and instance task

收稿日期: 2026-03-30。 作者简介: 杜少毅(1980—),男,教授,博士生导师;郑心湖(通信作者),男,助理教授,博士生导师。 基金项目: 国家重点研发计划资助项目(2025ZD0124204);国家自然科学基金资助项目(U24A20252);中央高校基本科研业务费资助项目(xzy012026035)。

网络出版时间: 2026-06-24

网络出版地址: <https://link.cnki.net/urlid/61.1069.T.20260624.0956.003>

levels. Second, representative methods for addressing node heterogeneity, temporal asynchrony, and spatial misalignment are summarized, focusing on key issues in multimodal information fusion, including interaction topology construction, spatiotemporal pose alignment, and cross-modal semantic alignment. Then, implementation approaches that balance communication efficiency and perception accuracy under bandwidth constraints are reviewed along three pathways: sparse representation for feature compression, generative completion, and semantic collaboration with large models. Finally, large-scale open-source datasets in this field are systematically compiled, and current bottlenecks and future development trends in scene cognition, dynamic network security, and the credibility of generative reconstruction are deeply analyzed. Through this paper, relevant research progress is systematically reviewed, and a theoretical reference is provided for the construction of more robust and secure intelligent connected vehicle-road cooperative networks.

Keywords: autonomous driving; cooperative perception; intelligent connected vehicles; multimodal information fusion; communication efficiency

随着人工智能、先进传感器以及计算平台的快速发展,智能车辆已由辅助驾驶逐步迈向高阶自动驾驶^[1],其中单车感知是自动驾驶系统的重要基础。单车感知主要依赖摄像头、激光雷达等传感器,并结合深度学习算法^[2-4]实现对局部交通环境的多模态特征提取与三维重建。复杂动态交通场景下,单车系统仍受限于视距不足、遮挡严重和远距离目标稀疏等物理约束,容易在鬼探头等长尾场景下出现感知失效。单纯增加传感器数量或提升单车算力,不仅成本较高,也难以突破单点观测的物理上限。因此,如何突破单车感知局限、提升复杂开放场景下的全局感知能力,已成为自动驾驶发展的关键问题。

依托 5G 网络^[5]与车联网(IoV)通信技术的协同感知逐渐成为解决该问题的关键路径^[6-9]。该技术通过车与车、车与基础设施等多节点链路进行信息交互^[10],实现了跨视角的超视距感知与信息共享^[11]。典型的车路云协同感知系统架构如图 1 所示,该系统主要涵盖智能网联汽车、路侧边缘感知与

计算单元,以及云端协同调度中心等核心模块^[7]。但这种多节点、跨视角的持续信息共享,会不可避免地产生极高的通信负载。当前车联网分配的频谱资源难以支撑未来自动驾驶系统对高带宽传输的需求,而且无线信道的高动态性、多普勒效应及非视距遮挡等问题进一步加剧了通信的不确定性,导致系统在严格的时延约束下难以维系稳定、可靠的数据传输^[12]。

在此背景下,Yu 等^[10]在 UniV2X 框架中揭示了协同感知在实际部署中,感知需求与通信能力之间的核心矛盾。尽管路侧基础设施的广域感知优势能够有效弥补智能单车在视野盲区和远距离探测上的物理局限,但动态变化的无线信道和严苛的实时性要求,迫使系统必须在极低延迟与有限带宽的约束下完成高效的数据交互。这一矛盾凸显了通信带宽瓶颈对协同感知系统设计的根本性制约,即原始传感器数据虽信息丰富但传输成本高昂,而压缩后的特征或结果数据虽能显著降低通信负载,却可能损失关键语义信息,最终削弱多智能体融合的精度与决策的可靠性。

除严苛的通信带宽制约外,协同感知系统在实际部署中还面临着显著的多源异构挑战。参与协同的智能体可能配备不同类型和分辨率的传感器,例如摄像头、激光雷达和毫米波雷达等,也可能采用异构的感知模型与算法架构,导致数据格式、特征表征和语义空间存在本质差异。这些异构性不仅增加了跨域数据对齐与融合的技术难度^[11],还要求通信协议具备灵活的适配能力以支持多样化的数据交换需求。因此,如何在有限且动态变化的通信带宽约束下,实现异构智能体间高效、鲁棒的信息交互与语义对齐,成为协同感知走向实际应用的关键问题。

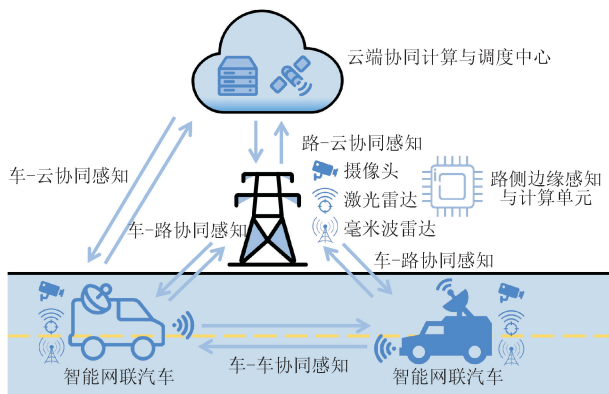


图 1 车路云协同感知系统架构

Fig. 1 Vehicle-road-cloud collaborative perception system architecture

为应对上述带宽与异构的双重挑战,现有协同感知方法主要形成了早期、晚期和中期三类融合策略。早期融合直接共享原始传感器数据^[13-14],信息保留最完整,但通信与计算负载最高^[15-16];晚期融合仅交互检测结果^[17],通信开销较低,但受限于单车感知上限,难以实现深层特征互补^[18];中期融合通过共享中间特征,在感知精度与传输效率之间取得了较好的平衡,因此成为当前研究的主流方向^[19]。该策略通过在节点间交互提取后的中间时空特征,既能实现空间或体素级语义的深度聚合^[19-20],又可通过轻量级架构^[21]或网络消息传递^[22],有效削减原始数据冗余。与此同时,针对多车协作中,固有的通信时延与传感器异步难题,中期融合还可通过对特征流进行显式的时间基准对齐与补偿^[23],进一步保障复杂交通流下的协同可靠性。

现有协同感知综述已从感知流程、融合策略、系

统架构、车路云协同、数据集与应用场景等方面形成了较为系统的总结,但对感知精度-通信负载权衡关系及其对应的高效传输机制关注仍相对不足。为明确本文与已有综述的差异,表 1 从综述主线、覆盖范围和关键问题等维度对代表性工作进行了比较。从表 1 可以看出,现有综述^[1,7-9,24-29]多侧重于系统架构、融合流程或应用挑战,而对受限通信条件下的信息筛选、异构对齐、压缩重构及其内在关联缺乏系统梳理。基于此,本文以通信高效传输为主线,重点围绕特征筛选、协同建模与对齐、信息编码与压缩 3 个层面构建分析框架,并进一步关注动态拓扑、异构多模态协同、生成式重构以及大模型语义协同等新近发展方向,具体技术分类如表 2 所示^[30-45]。本文将结合典型数据集对不同方法的适用评测场景进行对应分析,从而更系统地揭示协同感知中,感知性能、通信开销与系统鲁棒性之间的内在关联,整体框架如图 2 所示。

表 1 代表性协同感知综述工作对比

Table 1 Comparative summary of representative surveys on cooperative perception

文献	年份	主题	系统架构	通信效率	异构对齐	动态拓扑	压缩恢复	数据集
[1]	2022	协同感知流程、通信支撑、定位、检测/跟踪与地图应用	●	—	—	—	—	○
[7]	2022	IoV 环境下的感知、通信、融合及相关支撑技术	●	●	—	—	—	—
[24]	2023	方法分类、公开数据集、研究挑战与发展方向	○	○	—	—	—	●
[25]	2023	V2X 协同感知体系、方法脉络、数据集与挑战	●	●	○	●	—	○
[26]	2024	异构节点、层级协同、统一协同框架	●	—	●	—	—	—
[27]	2024	车路云协同体系、关键模块与应用支撑	●	○	—	—	—	—
[28]	2025	V2X 协同感知体系、关键技术、挑战与未来方向	●	●	○	●	○	●
[29]	2025	协同感知中的关键挑战、潜在解决思路与发展机遇	●	●	○	●	●	○
[8]	2025	车路群智协同感知方法、通信与信息融合问题	●	○	—	—	—	●
[9]	2025	车路协同感知策略、关键技术、相关数据与发展趋势	●	○	○	—	○	●
本文	2026	信息筛选、协作对象建模、时空/异构对齐、编码压缩、传输恢复	—	●	●	●	●	○

注:●表示重点覆盖;○表示较多涉及;—表示少量涉及。

表 2 协同感知高效传输方法的技术分类

Table 2 Technical taxonomy of communication-efficient transmission methods for cooperative perception

技术类别	核心问题	主要传输对象	典型机制	代表性方法	主要优势	主要局限
空间区域级 信息筛选	减少冗余背景 区域传输	鸟瞰视图(BEV) 区域、网格特征	置信度图、信息熵、 供需建模、时空 先验	CoSDH ^[30] 、 CoST ^[31] 、 UMC ^[32]	可显著降低通信负 载,易与 BEV 特征 融合框架结合	可能误删低置信度 但高价值的远距离 或遮挡目标
实例任务级 信息筛选	以目标或任务 为粒度选择 信息	目标查询、实例 特征	Query 筛选、Top-K 选择、困难目标优 先、稀疏查询	QUEST ^[33] 、 FocalComm ^[34]	传输语义更集中, 带宽需求更低	依赖前端查询质 量,冗余或错误查 询仍可能影响性能

续表

技术类别	核心问题	主要传输对象	典型机制	代表性方法	主要优势	主要局限
动态拓扑与节点选择	决定与谁通信、何时通信	低维请求向量、特征摘要、节点状态	握手通信、动态分量、优先级调度、协作建模	Who2com ^[35] 、When2com ^[36]	减少无效广播,提升多智能体协同效率	对节点关系、场景变化和链路状态较敏感
时空位姿对齐	缓解异步、定位漂移和空间错位	中间特征、感知图、位姿相关信息	隐式补偿、显式几何校准、多跳对齐、联合时延估计	V2X-ViT ^[37] 、GraphPS ^[38]	提升多车特征融合的稳定性 and 一致性	对定位精度、时间同步和共视条件仍较敏感
异构语义对齐	缓解传感器、模型和特征空间差异	多模态特征、公共表示、协议特征	域适应、参数高效微调、公共协议空间、图协同融合	HM-ViT ^[39] 、STAMP ^[40] 、HEAL ^[41]	支持跨模态、跨模型和跨平台协同	对齐成本较高,统一表示过程中可能产生信息损失
特征压缩与量化编码	进一步降低传输比特数	压缩特征、码字、索引	自编码器、特征解耦、码本量化、残差量化、信息瓶颈	What2comm ^[42] 、CodeFilling ^[43]	可显著降低通信代价,适合带宽受限场景	高压缩率下易损失细节信息,影响小目标和远距离目标检测
生成式特征重构	在少量传输条件下恢复缺失信息	残缺特征、提示词元、地图先验	生成式补全、先验引导、扩散模型、生成式通信	CORE ^[44] 、CoGMP ^[45]	可缓解信息缺失,提高低带宽条件下的感知完整性	重构结果的真实性、稳定性和安全性仍需验证

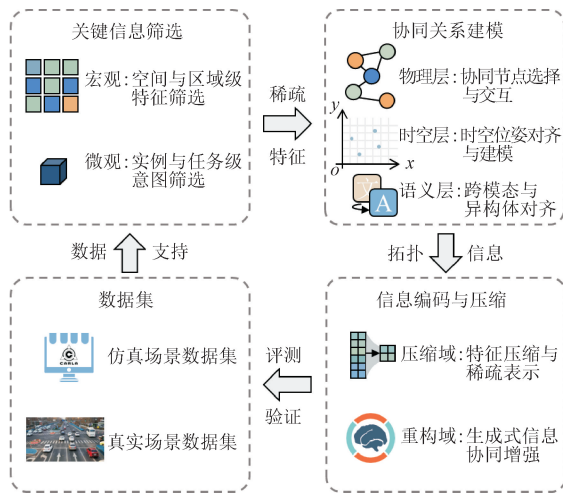


图 2 面向协同感知的高效传输方法综述框架

Fig. 2 Review framework of efficient transmission methods for cooperative perception

1 多层次特征筛选

中期特征融合有效降低了原始数据的传输量,但直接共享未经处理的稠密特征张量依然会带来不可忽视的通信负担。为平衡感知性能与受限通信带宽,研究者提出了多种关键信息筛选机制,以避免冗余特征的无效传输。本节主要讨论传什么的问题,即如何在发送端选择更有价值的

协同信息。由图 3 可知,现有特征筛选方法可按照信息粒度划分为空间与区域级筛选,以及实例与任务级筛选两个层面。

1.1 空间与区域级筛选

为应对通信带宽受限问题,中期融合方法逐步从全量特征共享转向基于任务相关性的选择性传输机制。早期方法多采用全量广播式策略,在智能体之间直接共享完整且未压缩的鸟瞰图(BEV)中间特征,但该方式由于缺乏有效筛选,不可避免地引入大量静态背景与低价值冗余信息,从而导致通信资源利用效率较低。为降低通信负载,Where2comm 方法提出基于空间置信度图的筛选机制,通过评估各空间网格的感知置信度并设定阈值,剔除低相关背景区域,使通信由全量传输转向稀疏传输,在降低通信负担的同时尽量保持感知性能^[46]。然而,单纯依赖置信度进行筛选仍存在局限。对于自车而言,低置信度区域既可能对应纯背景,也可能对应被严重遮挡或距离较远的目标,仅依靠单一置信度指标,容易误删潜在的高价值信息。为此,CoSDH 方法从供需感知角度重新建模筛选过程,通过分别刻画自车的信息需求与协作者的信息供给,将有限带宽更有针对性地分配给真正有助于补盲的前景区域^[30]。在此基础上,后续研究进一步将信息量评估和外部先

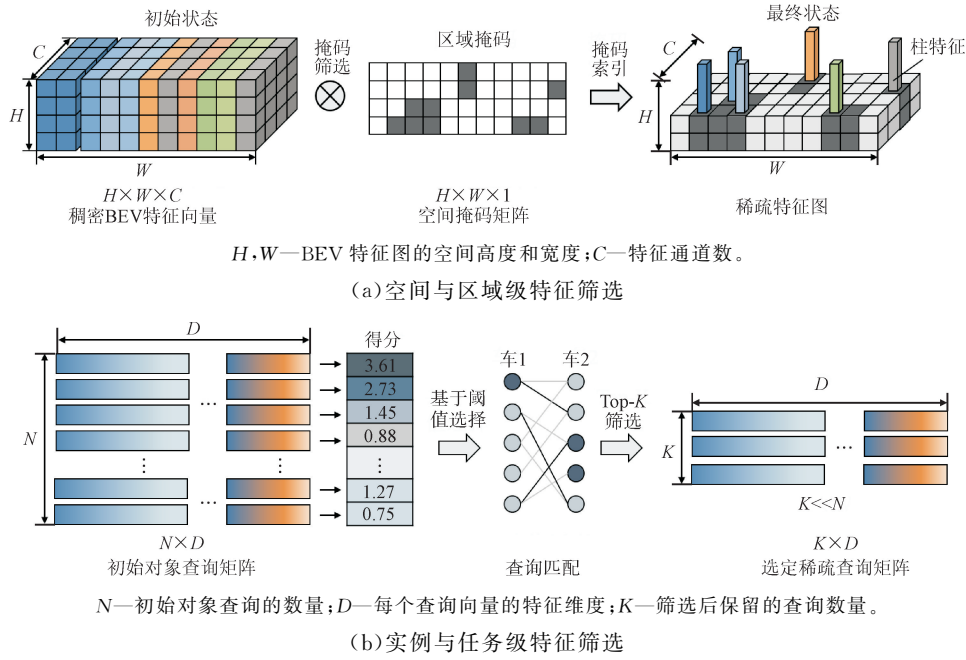


图 3 多层次特征筛选

Fig. 3 Multi-level feature filtering

验引入筛选过程,以提升通信资源分配的精细化水平。UMC方法将信息熵建模引入二维通信选择过程,通过量化多分辨率特征图的局部信息熵,自适应滤除低质量、低信息量的网格区域,从而兼顾全局结构与局部细节^[32]。除此之外,空间先验和时序先验也被用于引导通信方向与资源分配。Directed-CP方法结合自车预期行驶轨迹生成方向掩码,并借助主动注意力机制将通信资源集中于对行驶安全更关键的局部区域^[47]。CoST方法则利用时序差分机制,仅传输发生动态变化的对象特征,从而在维持场景语义完整性的前提下进一步降低带宽需求^[31]。

空间与区域级筛选方法主要以稠密 BEV 特征或局部网格为操作对象,通过置信度评估、信息量建模以及时空先验等机制抑制背景冗余,在宏观层面实现通信资源的有效分配,如图 3(a)所示。这类方法易于与现有 BEV 融合框架结合,但在感知范围持续扩大的同时,通信开销则可能增长,也较难显式表达实例级语义。

1.2 实例与任务级筛选

相较于区域级筛选,实例与任务级筛选进一步将传输单元由局部特征块细化为目标实例或稀疏查询。区域级筛选虽能有效降低通信量,但其依赖的稠密 BEV 特征仍包含不可忽略的背景冗余,且缺乏明确的实例语义。为此,协同感知的研究重心逐步由宏观区域裁剪转向以实例查询为最小单元的稀疏通信,其核心思路是利用 Transformer 中的目标查

询机制替代传统特征图表示,以更低带宽占用传递更集中的任务相关语义。

早期工作验证了基于查询的实例级传输在通信效率上的优势,QUEST 在智能体之间传递查询流,通过实例级特征交互实现共视目标增强与非共视目标互补^[33]。TransIFF 进一步表明,直接传输实例级特征而非全局特征图,能够显著降低通信带宽需求^[48]。针对多视角协同中,同一目标产生冗余查询的问题,CoopDETR 构建对象查询图并引入空间匹配机制,在提升检测精度的同时减少冗余交互带来的信息浪费^[49]。

在此基础上,研究进一步关注查询质量控制,以降低无效或噪声信息对协同过程的干扰。CoCMT 结合基于置信度的 Top-K 筛选策略与多重注意力掩码机制,将跨智能体交互限制在与目标高度相关的时空范围内,从而提升融合效率^[50]。ActFormer 则通过主动选择机制为每个查询分配兴趣权重,使模型能够自适应选择来自其他智能体的高相关信息^[51]。

近期工作又将查询筛选与任务特性及传感器物理模型结合,进一步提升复杂场景下的信息利用效率,如图 3(b)所示。针对严重遮挡或远距离条件下的困难目标,FocalComm 基于检测不确定性对查询进行动态排序,使通信资源优先分配给高价值的困难实例,从而改善长尾目标的检测性能^[34]。另一方面,SlimComm 利用 4D 毫米波雷达的多普勒信息构建稀疏查询,在前端过滤缺乏运动特征的冗余信

息,使通信过程更加聚焦于动态目标,在降低带宽占用的同时提升复杂场景下的感知可靠性^[52]。

总之,空间与区域级筛选和实例与任务级筛选都以降低协同通信负载为目标,但作用的信息粒度存在明显差异。前者主要面向稠密 BEV 特征或局部网格区域,更强调空间覆盖下的冗余抑制,具有适用场景广、易于与现有框架结合等优势;后者则以目标实例或稀疏查询为传输单元,更强调任务驱动下的语义聚焦,在语义可解释性和通信效率方面更具优势。然而,实例级筛选也更加依赖前端检测与查询建模质量,在严重遮挡或远距离小目标场景下仍可能遗漏关键信息。

本节从空间区域到实例查询两个层级梳理了特征筛选方法,反映出协同感知中的通信优化正由粗粒度裁剪逐步走向细粒度、任务驱动的信息选择。如何进一步融合空间先验与实例语义,在保证召回

能力的同时提高带宽利用率,仍是值得深入研究的重要方向。

2 协同建模与特征对齐

完成高价值信息筛选之后,协同感知系统尚需解决协同对象选择与特征对齐两大问题。面向多智能体网络的协同感知,不仅需要在有限带宽约束下构建高效的交互关系,还需要缓解由节点异构、时间异步和空间位姿误差带来的融合偏差。传统全向广播机制在复杂交通流中容易引发通信拥塞与接收端过载,而不同智能体在传感器配置、模型架构及时空状态上的差异,又进一步加剧了跨域信息融合的难度。因此,本节将从交互拓扑构建、时空位姿对齐以及跨模态语义映射这 3 个维度,系统梳理多智能体协同关系建模与特征对齐的关键方法,其原理图如图 4 所示。

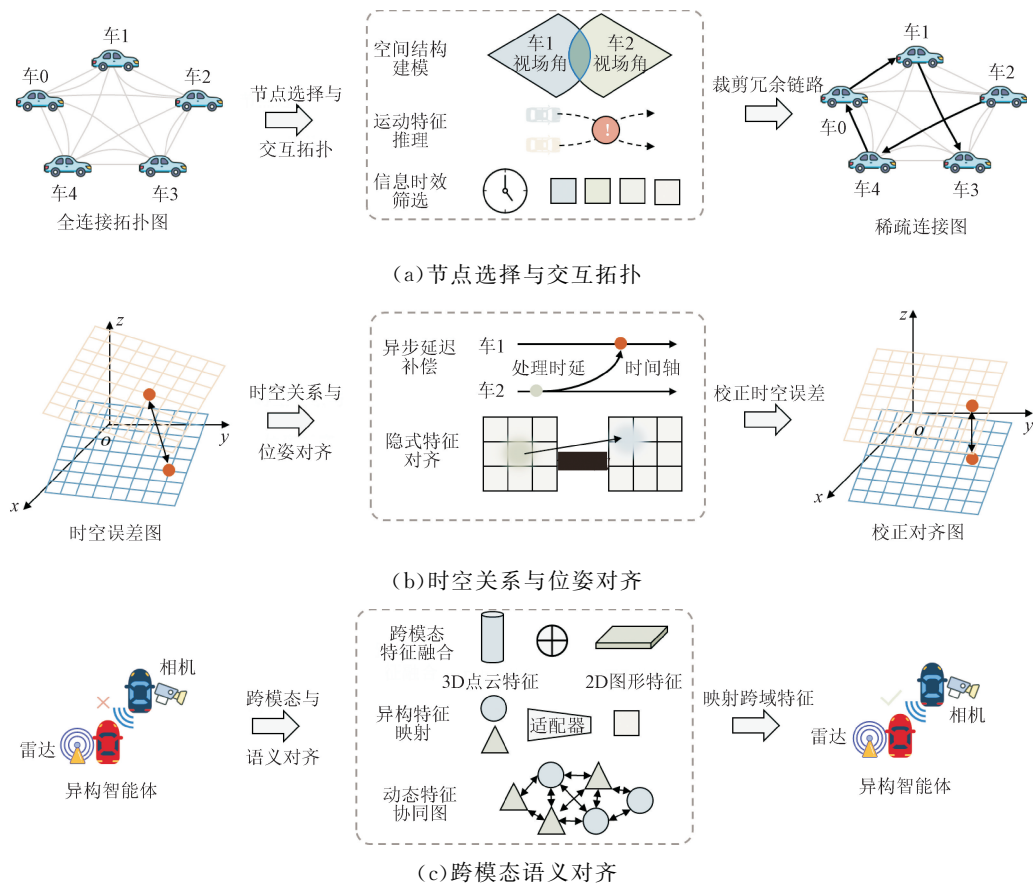


图 4 协同建模与特征对齐

Fig. 4 Cooperative modeling and feature alignment

2.1 节点选择与交互拓扑

在多智能体协同网络中,全连接静态通信拓扑会导致带宽消耗随节点数量的增加而快速增长,同时无差别的特征聚合容易引入来自视野边缘或无关

车道的干扰。因此,如图 4(a)所示,协同节点选择的关键在于根据当前场景从多个候选智能体中筛选出真正有价值的通信对象,进而构建稀疏、高效且低延迟的交互拓扑。

为降低全局广播带来的通信开销,早期研究主要从特征相关性出发构建通信群组。这类方法主要依赖全局特征的静态相似性匹配或注意力机制来决定交互对象,例如,Who2com 通过可学习握手协议,在正式建立通信前进行轻量级匹配,使因视野受限而感知退化的智能体,能够优先选择对自身最具补充价值的协作者^[35]。随后,When2com 将这一思路进一步扩展为动态群组划分,使智能体不仅能够决定与谁通信,还能够根据场景复杂度动态控制通信时机,从而提高带宽利用率^[36]。在此基础上,Li 等^[53]进一步从全局协作关系建模角度优化通信拓扑,通过知识蒸馏与协作图学习突出信息密集区域,并抑制来自无关节点的干扰。

尽管基于特征相似度或协作图学习的拓扑方法能够提升通信效率,但其物理可解释性仍然有限,且难以显式反映不同车辆之间由于相对运动和行驶意图差异所带来的潜在安全价值。因此,后续研究开始将运动意图和任务相关性引入协同节点选择。Chiu 等^[54]与 Wang 等^[55]将多智能体路径规划思想引入协同感知,通过评估邻居车辆对自车盲区的信息增益,实现更具针对性的节点选择与定向交互。进一步地,Fang 等^[56]与 Liu 等^[57]从优先级调度角度出发,结合空间相关性与时间时效性,为更新及时且高相关的节点分配更高的传输权重,从而使协同过程更加聚焦于当前任务最需要的信息来源。

然而,在真实车联网环境中,通信链路还会受到时延波动、消息丢失、特征缺失以及节点动态出入等持续扰动,单纯依赖静态或准静态拓扑往往难以维持稳定的协同收益。近年来,部分研究开始从非理想通信条件出发,探索动态网络下的鲁棒交互机制。针对消息中断与特征缺失,Ren 等^[58]通过显式建模中断状态提出中断感知协同框架,使系统在无法稳定接收邻居消息时仍能维持基本感知能力。Shi 等^[59]则设计损失感知的特征融合方法,以提高特征缺失条件下的检测鲁棒性。对于异步到达与时间失配问题,CoBEVFlow 方法通过鸟瞰图流场对异步特征进行运动补偿,将不同时刻的协同消息对齐到统一时间基准^[23];Dao 等^[60]则从工程部署角度构建了低带宽代价下的异步多智能体检测框架。当网络拓扑和信道状态持续变化时,协同系统则需具备动态自适应能力,如 ACC-DA 可根据实时信道状态调整通信图结构,并通过域对齐减轻异构数据带来的性能退化^[61]。此外,开放异构协同方法则进一步强调低成本接入、局部适配与渐进式特征对齐,以降低

节点的频繁加入与退出带来的重训练开销^[41]。

节点选择与交互拓扑方法已经由基于特征相关性的静态连接,逐步发展为兼顾任务需求、运动意图和网络扰动的动态协同机制。前一类方法结构简单、实现直接,但对场景变化和链路状态适应能力有限;后一类方法在真实部署条件下具有更强鲁棒性,但通常也带来更高的建模复杂度。如何在动态网络环境中同时兼顾通信效率、拓扑稳定性和协同收益,仍是交互建模需要持续解决的重要问题。

2.2 时空关系与位姿对齐

完成节点选择后,将多源感知数据统一到一致的时空坐标系,是协同感知走向实际部署的基础。在真实开放交通环境中,受卫星定位多径效应、异构传感器采样频率差异以及底层网络传输延迟等因素的影响,智能体之间普遍存在显著的空间位姿偏差与时间异步现象。若直接融合未经严格对齐的特征,容易引发特征重影和语义错位,进而导致整体感知性能的明显下降。因此,现有研究主要形成了隐式特征补偿与显式几何校准两类时空对齐策略,并逐步向弱定位依赖或无额外硬件辅助条件下的鲁棒对齐方案发展,如图 4(b)所示。

在隐式特征补偿方面,研究者主要依靠深层网络对高维特征交互过程中的时空误差进行自适应吸收和平滑。Xu 等^[37]提出融合多尺度窗口注意力与异构多智能体注意力的统一框架,并引入延迟感知的位置编码,以动态调节注意力权重,从而在存在定位漂移和通信延迟时实现异步特征的隐式对齐。类似地,V2VFormer 结合空间注意力与通道交互机制,通过对邻接节点特征权重的自适应调节,在特征层面缓解由位置坐标和航向角误差引起的语义偏移^[62]。隐式补偿方法在小幅度时空误差下具有较好的平滑能力,且避免了额外几何求解过程,但在大幅位姿偏差或严重异步条件下,其校正能力往往受到限制。

为提高大误差场景下的对齐稳定性,另一类研究不再完全依赖外部定位信息,而是利用感知结果本身的几何关系进行显式校准。Song 等^[63]利用各智能体独立检测到的目标边界框构建局部上下文图,并通过图匹配在不同视角之间寻找结构一致的目标关系,从而估计节点之间的相对旋转和平移,实现较为精确的空间对齐。然而,在复杂城市路口或长距离协作场景下,智能体之间往往由于建筑遮挡或视场受限而缺乏共视目标,此时若强制执行点对点图匹配,容易引发错误关联并导致位姿误差级联放大。为缓解这一问题,GraphPS 引入多跳协同思

想,通过计算各智能体感知图之间的相似度并构建最小生成树,在全局协作路径上逐步传递对齐信息^[38]。这样,即使部分节点之间缺乏直接共视目标,也能够借助中间节点完成间接对齐,从而在定位噪声较大或通信距离较长的条件下保持较稳定的联合位姿估计。

在此基础上,部分最新研究进一步尝试在同一框架下联合估计空间位姿与通信延迟,以减少对外部同步设备的依赖。Zhou 等^[64]提出的 V2XPnP 框架在多智能体之间构建显著目标关联图,一方面利用几何关系估计节点间相对位姿,另一方面结合历史特征序列推断通信过程中产生的时间延迟,从而将空间配准与时间同步统一到同一优化过程中。这表明,融合隐式补偿与显式校准优势的混合式时空对齐方法,正在成为一个更具潜力的发展方向。

总之,隐式补偿方法更适合吸收小幅、连续的时空扰动,显式校准方法则在大误差条件下通常具有更强的鲁棒性,但也更依赖共视目标和结构的一致性。未来的关键在于,如何在复杂动态交通环境下,将特征级鲁棒补偿与几何级显式校准更紧密地结合起来,以同时提升对齐精度、部署稳定性与泛化能力。

2.3 跨模态语义对齐

除节点关系和时空位置对齐外,协同感知系统在开放环境中还面临显著的跨模态与跨模型异构问题。不同厂商或不同部署平台的智能体往往搭载差异化的传感器组合,并采用不同的感知网络架构。由于硬件模态与软件结构的双重异构性,即使处于同一时空视角下,各节点提取的中间特征在维度、分辨率以及语义分布上仍可能存在显著的表征鸿沟。若直接对这些未经严格对齐的特征进行融合,容易引发语义混淆并削弱整体感知效果,如图 4(c) 所示。因此,如何在保障各节点模型独立性的前提下实现深度语义对齐,已成为开放式协同感知中的核心挑战。

面向传感器模态异构的协同场景,早期研究通常依赖专用跨模态融合结构缓解不同模态之间的表征差异。这类方法通常以 Transformer 或注意力机制为核心,将不同模态映射到统一的鸟瞰图空间后再进行交互融合。V2VFormer⁺⁺通过在 BEV 空间内引入动态通道融合与分层注意力机制,实现视觉图像与激光雷达点云之间的高精度聚合,并在部分节点仅具备单一模态的情况下保持相对稳定的感知输出^[65]。这类方法针对性较强,但当异构性由硬件模态进一步扩展到软件模型架构层面时,单纯依

赖专用融合模块的可扩展性会受到限制。

为支持更灵活的异构接入,后续研究开始转向统一协议空间或公共表示层的构建。通过域适应、特征尺度对齐和参数高效微调等机制,不同架构提取的特征可以被映射到较为一致的分布空间,从而减弱由模型差异带来的特征干扰^[66-68]。HEAL 方法提出统一的特征协议空间,使新接入节点仅需在本地微调前端编码器即可适配公共表示^[41],STAMP 方法则进一步通过适配器-还原器模块,将发送端私有特征转换为公共协议域,并在接收端恢复为本地可用表示,从而提高开放异构场景下的接入灵活性^[40]。然而,统一协议空间方法通常需要将异构特征映射到预先设定的公共域,当领域差异较大时,仍可能出现对齐困难、表达受限甚至信息损失。

针对固定公共空间的局限,近期研究进一步提出了动态协商与图协同对齐思路,根据当前参与协同的模态特征动态生成对齐策略,而非预先设定唯一的协议空间。Shao 等^[69]通过动态生成并优化公共表示,同时结合分布、结构和语义等多维对齐损失,以减小异构特征与协商空间之间的差异。Hetecooper 方法则通过构建特征协同图,在图空间内完成异构特征之间的信息传递与融合,避免传统插值或强制降维操作带来的信息压缩^[70]。相比静态公共协议空间,这类方法在保留模态原生结构信息和适应开放异构接入方面更具优势,但计算复杂度与训练稳定性仍有待进一步优化。

随着多模态协同的进一步发展,研究也开始由两两模态专用融合拓展到更复杂的具体传感器组合。对于相机与激光雷达,二者分别提供稠密纹理语义与稀疏三维几何信息,如 HM-ViT 方法^[39]和 HeCoFuse 方法^[71]通常先将图像与点云特征映射到统一的 BEV 空间,再利用交叉注意力建立像素语义与体素几何之间的对应关系。对于进一步引入 4D 毫米波雷达的场景,其多普勒速度信息与全天候感知能力能够为动态目标协同提供重要补充,但由于雷达回波本身稀疏且噪声较强,对齐问题也由静态几何匹配进一步拓展为几何-运动联合建模。V2X-R 以激光雷达点云为几何锚点,通过去噪扩散过程修正雷达特征偏移,并结合速度信息实现跨模态融合^[72]; V2X-Radar 方法^[73]与 SlimComm 方法^[52]也进一步表明,4D 毫米波雷达提供的速度先验有助于将跨模态交互集中于高价值动态区域。

现有跨模态语义对齐方法大致呈现出三类方法,第一类是面向特定模态组合的专用对齐方法,

例如相机与激光雷达之间基于 BEV 空间交叉注意力的融合机制^[65],其针对性较强,但扩展到更多异构模态时灵活性有限。第二类是以 HEAL^[41]、STAMP^[40]为代表的统一协议空间方法,通过构建公共表示层支持更灵活的异构模态接入,但在映射过程中仍可能受到领域差异和信息损失的影响。第三类则是动态协商与图协同方法,如 NegoCollab 方法^[69]和 Hetecooper 方法^[70],不再预先设定固定的公共空间,而是根据参与协同的模态特征动态生成对齐策略,从而在保留各模态原生结构信息方面更具优势。对于 4D 毫米波雷达等稀疏且噪声较强的异质模态,后两类方法通常表现出更好的适应性,但其计算复杂度与训练稳定性仍有待进一步优化。总之,跨模态语义对齐正在由静态映射走向动态协商,并逐步由相机-激光雷达等经典组合扩展到包含 4D 毫米波雷达在内的更复杂多模态统一表示框架。

3 信息编码与压缩

在车路协同感知系统中,受动态波动的无线通信信道与有限带宽约束的共同影响,直接传输高维连续浮点特征张量容易造成链路拥塞并引入额外时延。为在严格通信预算下保障信息传递的有效性,本节从前端特征压缩与稀疏表示、接收端生成式补全与高层语义增强两个层面,系统梳理协同信息的紧凑表征与恢复机制,如图 5 所示。

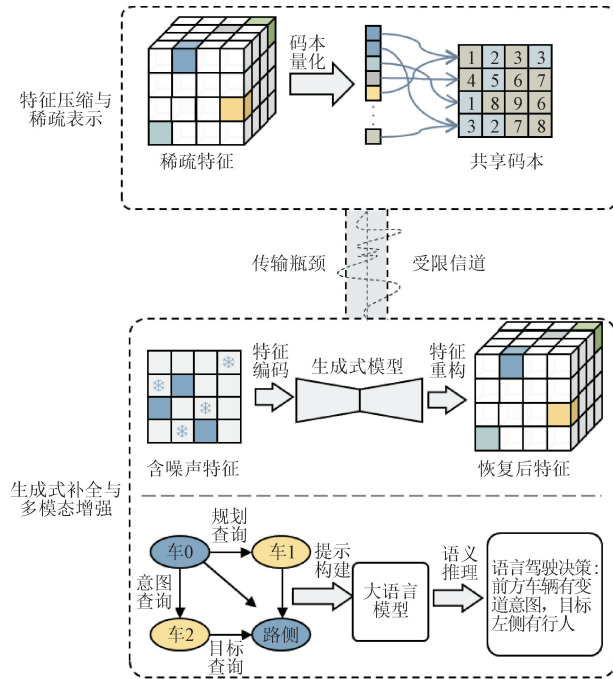


图 5 信息编码与压缩

Fig. 5 Information encoding and compression

3.1 特征压缩与稀疏表示

早期特征级融合方法通常将通信压缩建模为维度映射问题,借助轻量级自编码器沿通道维度对稠密特征图进行统一压缩^[22,28],这与视频编码领域中利用轻量级卷积网络对色度跨分量进行预测压缩、降低传输比特数的思路具有相似性^[74]。这类全局降维策略能够有效降低待传输特征的维度,但对不同空间位置采用相同处理方式,难以反映场景内部语义分布与几何重要性的差异。当压缩率进一步提高时,目标边缘和细粒度语义往往最先受损,从而影响下游检测与定位精度。

为缓解统一降维带来的语义损失,后续研究开始引入特征解耦与任务相关保留机制。What2comm 通过多头注意力与中心矩差异约束,将待传输信息划分为跨智能体共享的公共语义和视角相关增量两部分,从而优先保留任务关键内容^[42];CoMamba 则以状态空间模型替代传统自注意力,在较低参数量下实现高效特征表示与压缩,在通信负载与语义保真之间取得了更好的平衡^[75]。然而,这类方法虽然缓解了通道维度上的语义丢失,但其操作基础仍主要依赖稠密二维 BEV 网格,在大面积空旷区域和静态背景较多的自动驾驶场景中,稠密网格本身仍会带来显著冗余。

为进一步减少背景冗余,研究开始由稠密特征压缩转向结构稀疏化表示,CPPC 以三维稀疏实体替代规则网格,采用局部点簇作为通信单元,仅传输聚类中心及其聚合特征,从而在过滤背景噪声的同时保留较完整的三维几何细节^[76]。随后, SparseAlign 将稀疏表示扩展至特征提取、跨节点交互与空间对齐的全流程,仅维护少量前景查询向量,实现从感知到融合的端到端稀疏化,从而通过减少传输单元数量,从表示结构层面进一步压缩了通信负担^[77]。

当压缩机制仍以连续浮点特征为基本载体时,通信开销最终仍受数值精度限制。因此,需进一步研究引入离散量化与信息论约束来降低传输比特数。CodeFilling 在各节点预置共享码本,通过离散向量量化将连续特征映射为低位宽索引,使接收端仅通过查询本地码本即可恢复协同特征^[43]。为降低高压缩带来的量化噪声,ReVQom 采用残差向量量化,以多级码本渐进补偿量化残差,在相同带宽下提升重构质量^[78]。进一步地,InfoCom 将信息瓶颈理论引入协同感知压缩框架,通过变分优化在保留任务相关信息的同时抑制冗余噪声,并结合直通估计器与掩码引导的多尺度解码,实现通信负载与

感知精度之间的更优平衡^[79]。How2compress 则进一步将信息论约束与动态信道状态耦合,设计自适应量化深度与资源分配策略,使特征编码由静态压缩走向动态调节^[80]。

总之,特征压缩与稀疏表示方法大致沿着统一降维、结构稀疏化、离散量化与动态压缩的路径演进。前两类方法主要从表示维度和结构冗余角度降低通信负载,后两类方法则进一步从比特级传输代价出发优化通信效率。这4类方法通过在发送端减少特征维度、传输单元或编码比特数,有效降低通信开销,是高效传输研究中最直接、也最具工程可实施性的技术路线,但在高压压缩比条件下仍可能面临细节丢失、小目标退化和场景复杂性增强带来的性能下降。

3.2 生成式补全与多模态增强

当底层特征降维与离散量化已将通信负载压缩到较低水平后,接收端在极端受限带宽条件下仍不可避免地面临局部几何细节缺失和语义信息不完整等问题,单纯依赖发送端压缩往往已难以满足高保真环境感知需求^[81]。因此,研究进一步将生成模型和多模态语言模型引入协同感知,希望借助先验知识在有限通信条件下恢复更完整的场景表示。

在低层物理空间特征恢复方面,主要从生成式补全与先验引导重建展开研究。面对强压缩通信带来的特征空洞与结构畸变,Wang等^[44]构建基于局部稀疏观测的生成式补全框架,将协同节点传输的残缺特征作为条件输入,在接收端推断并恢复视野盲区内的隐藏信息。考虑到仅依赖局部稀疏观测往往难以准确刻画全局交通拓扑与场景约束,Fu等^[45]进一步将预训练模型中的道路几何先验引入重构过程,通过将连续感知特征映射为离散地图元素索引进行选择交互,并依托接收端的生成式地图先验对掩码区域进行定向恢复,从而提升复杂路网场景中的语义连续性与空间一致性,这种基于掩码区域恢复关键语义的思路在遮挡人脸识别等其他视觉任务中也得到了广泛应用^[82]。

为提升协同感知系统在恶劣天气、通信延迟及跨模态噪声等复合干扰条件下的恢复能力,研究又进一步引入扩散模型与生成式通信机制。Huang等^[72]与Zou等^[83]将去噪扩散模型应用于协同感知,将异构传感器获取的协同特征建模为受噪声干扰的潜在变量,并通过条件扩散过程逐步消除时空漂移和环境噪声,实现更稳健的多模态细节修复。随着生成先验能力的增强,通信内容本身也被继续

压缩。Zhou等^[84]提出生成式通信机制,使发送端仅交互轻量级空间引导提示,而由接收端依赖生成模型先验恢复与发送方语义一致的感知特征,从而进一步降低高维特征直接传输带来的负担。进一步地,Xu等^[85]将该思路扩展到三维表示空间,通过多视角观测的生成式聚合,实现动态交通流的高保真三维重建与自由视角场景表达。

除低层物理特征恢复外,相关研究还尝试将通信内容由特征级进一步提升为结构化语义信息,从而增强协同系统在高层场景的理解能力与决策层面的表达能力。在开放异构环境中,Gao等^[86]提出基于自然语言描述的协同机制,使智能体之间无需共享点云或图像矩阵,仅通过交互高度凝练的交通态势与驾驶意图文本即可实现极限压缩。自然语言虽具备较强语义概括能力,却难以稳定保留精确的几何位置关系。为减少语言语义与三维几何表征之间的差距,You等^[87]与Ma等^[88]提出视觉-语言协同对齐框架,通过对比学习将多视角视觉特征与文本描述映射到统一潜空间,并进一步输出带有坐标约束的结构化协同规划信息,从而增强高层语义交互对具体驾驶任务的支撑能力。

当协同任务进一步扩展到多车动态博弈与连续决策场景时,单轮语义交互已难以覆盖复杂交通环境中的长期协商需求。为此,Fang等^[89]与Hu等^[90]将检索增强生成、历史交互记忆以及闭环逻辑推理架构引入大模型框架,使智能体能够从历史经验中检索相似场景,并在动态博弈中生成更一致的通行时序与协作方案。与此同时,Chiu等^[91]提出面向车车协同自动驾驶的多模态大模型框架,通过V2V问答形式统一处理目标定位、显著目标识别与规划等任务。在此基础上,DriveLM^[92]与V2V-GoT^[93]进一步引入Graph VQA或Graph-of-Thoughts机制,将复杂驾驶任务分解为感知、预测与规划等多步推理过程,从而增强协同决策的可解释性与安全性。总体来看,大(视觉)语言模型正推动协同自动驾驶由低层特征交互进一步向高层语义理解与决策协同拓展^[94]。

为进一步客观评估协同感知方法在真实通信环境中的交互效率,近年研究在沿用常规检测精度指标的同时^[95],还引入了通信数据量^[46]和带宽提升得分^[96]等评价指标。表3汇总了代表性单帧协同感知方法在OPV2V^[97]和DAIR-V2X数据集^[98]上的检测精度与通信开销。其中,AP@0.5和AP@0.7分别表示交并比(IoU)阈值为0.5和0.7时的平均精度(AP),即预测框与真实框重叠区域占比达

到相应阈值时所对应的检测精度;通信开销采用 lb(Byte)作为统一度量单位,表示单帧协同传输的数据量,以统一不同方法的通信代价度量。该指标数值越小,说明所需传输带宽越低,通信效率越高。由表 3 可知,协同感知高效传输方法总体呈现出由区域级筛选向特征压缩与稀疏表征逐步演进的趋势。早期方法如 Where2comm 在 OPV2V 和 DAIR-V2X 上已能在有限通信开销下保持较好的检测性能,而 V2VNet 和 V2X-ViT 等则在相近通信量下进一步提升了检测精度。此外,CodeFilling、InfoCom 和 CPPC 等方法通过码本压缩、信息瓶颈或稀疏点簇表示,在更低通信开销下仍取得了较高的 AP 指标,体现出更优的通信—性能权衡关系。

表 3 典型高效协同感知方法在 OPV2V 和 DAIR-V2X 数据集上的通信开销与检测精度对比

Table 3 Comparison of communication cost and detection accuracy of typical efficient cooperative perception methods on OPV2V and DAIR-V2X

技术类别	方法	年份	OPV2V		DAIR-V2X		单帧通信 开销/lb(Byte)
			AP@0.5	AP@0.7	AP@0.5	AP@0.7	
空间级筛选	Where2comm ^[46]	2022	86.77	76.34	63.16	51.04	24.00
	EIMC ^[99]	2026	94.71	89.16	73.01	58.37	20.16
	DiscoNet ^[53]	2021	86.20	73.30	69.28	58.56	20.16
特征压缩与 编码传输	V2VNet ^[22]	2020	93.13	89.00	72.22	52.95	25.43
	V2X-ViT ^[37]	2022	93.66	86.06	71.87	55.46	24.00
	CodeFilling ^[43]	2024	86.84	75.84	76.90	61.06	15.00
	InfoCom ^[79]	2026	96.50	92.02	77.89	63.85	12.98
稀疏表征	CPPC ^[76]	2024	94.60	93.10	76.89	69.39	16.00
	CoLC ^[100]	2026	96.88	92.93	76.71	62.17	25.46

注:表中结果和方法据原文或其公开对比结果整理,受实验设置差异影响,仅用于代表性比较。

4 常用数据集

为系统评估各类协同感知算法的性能,本节对主流协同感知数据集进行梳理。受限于早期真实数据采集与高精度标定成本较高,相关研究主要依赖仿真环境开展方法验证。随着软硬件平台和数据采集技术的不断成熟,大规模真实世界数据集已逐渐成为协同感知研究的重要支撑。表 4 对主要数据集的采集场景、传感器配置和数据规模进行了整理。

4.1 仿真数据集

仿真数据集能够在可控环境下大规模生成多模态协同感知数据,并灵活构建复杂交通流、极端天气以及通信扰动等测试条件。通过在虚拟环境中精确控制车辆行为、交通参与者分布和传感器配置,研究

综上所述,特征压缩与生成式重构代表了应对极端带宽约束的两条互补路径。前者主要通过降维、量化、稀疏化或信息瓶颈等方式,在发送端尽可能保留任务相关的核心语义并显式压缩传输比特数,因而更适用于需要稳定控制通信开销的带宽受限场景。后者则进一步借助扩散模型、生成模型或大模型先验,在接收端主动恢复未被直接传输的场景信息,能够在极低通信预算下维持一定的感知与理解能力,但其重构结果的真实性、长尾场景下的泛化能力以及安全可控性仍面临挑战,促使研究者探索轻量压缩编码与生成式先验引导的结合,通过发送端显式压缩与接收端隐式补偿的联合优化,以更好地平衡通信效率与感知信息保真度。

者不仅能够系统评估算法在不同场景条件下的性能,还能够方便地引入通信延迟、数据丢包和定位误差等非理想因素。因此,仿真数据集特别适合用于算法验证、消融分析以及不同方法之间的可重复比较。

OPV2V 数据集^[97]依托 OpenCDA^[101]和 CARLA^[102]联合仿真平台构建,不仅包含 CARLA 默认虚拟城镇数据,还进一步引入了 Culver City 数字孪生场景,以评估模型在未见真实拓扑结构下的泛化能力。该数据集共包含 73 个复杂驾驶场景和 11 464 帧多模态协同感知数据,传感器配置较为完整,因此被广泛用于区域级特征筛选、特征压缩以及通信效率评估等任务。

V2XSet 在数据构建过程中显式模拟了现实协

同感知中常见的通信延迟和定位误差,旨在研究非理想通信条件下的多端混合同步[44]。该数据集包含 55 个高密度交通场景和 11 447 帧多模态数据,覆盖直道、弯道、街区中段、入口匝道和十字路口等典型道路类型。由于其明确纳入了时空噪声因素,因此常被用作评估协同 3D 检测与鸟瞰图语义分割模型抗噪能力和鲁棒性的重要基准。

V2X-R 是较早地在车路协同中系统引入 4D 毫米波雷达模态的大规模仿真数据集,包含 12 079 个场景,并覆盖晴天、雨天、雾天和雪天等多种天气条件及不同光照环境[72]。得益于 4D 雷达在恶劣天气下的穿透性和多普勒测速能力,该数据集为异构模态协同去噪、跨模态语义对齐和全天候鲁棒感知研究提供了重要支撑。

4.2 真实数据集

与仿真环境相比,真实世界数据集通常通过在实际城市道路或交叉路口部署测试车辆与路侧单元采集数据,能够更真实地反映车路协同系统在复杂交通环境中的运行状态。这类数据集不仅保留了真

实物理传感器噪声、复杂交通流动态以及不可避免的时空异步问题,还包含链路波动、遮挡变化和主体交互等工程部署中难以回避的现实因素。因此,真实数据集是评估协同感知系统可靠性、泛化能力与实际部署可行性的重要基础。

DAIR-V2X 是业界首个面向车路协同自动驾驶构建的大规模多模态真实世界数据集,在北京市高级别自动驾驶示范区的真实交通路口采集,由测试车辆与智能路侧基础设施同步收集激光雷达和图像数据,并提供高精度三维边界框标注[98]。该数据集不仅建立了车路协同三维目标检测的标准基准,也为研究真实物理场景下时序异步和传输成本约束提供了关键数据支撑。

V2X-Seq 在 DAIR-V2X 基础上进一步扩展为大规模真实世界车路协同序列数据集[103]。该数据集引入连续多模态观测、高精度矢量地图和实时交通信号灯信息,使研究视角由单帧感知拓展到时序感知、多目标跟踪与协同轨迹预测,从而更适合支撑时序对齐、运动补偿和动态场景理解相关研究。

表 4 自动驾驶协同感知数据集

Table 4 Datasets for cooperative perception in autonomous driving

来源	数据集	年份	发表来源	传感器类型	数据集大小/10 ³		
					图像	点云	包围框
仿真数据集	CODD ^[104]	2021	RA-L	L		13.5	204.0
	OPV2V ^[97]	2022	ICRA	C, L, R, M	44	11.4	232.9
	V2XSet ^[44]	2022	ECCV	L, C, M	44	11.4	233.0
	V2X-R ^[72]	2025	CVPR	L, C, R	151	37.7	170.8
真实数据集	DAIR-V2X ^[98]	2022	CVPR	L, C	71	71	1 196
	V2X-Seq ^[103]	2023	CVPR	L, C, M	71	15	464
	V2V4Real ^[105]	2023	CVPR	L, C, M	40	20	239
	Rcooper ^[106]	2024	CVPR	L, C	50	30	
	V2X-Radar ^[73]	2025	NeurIPS	L, C, R	40	20	350
	V2V-GoT-QA ^[91]	2026	ICRA	L, C		18	

注:C、L、R、M 分别表示相机、激光雷达、毫米波雷达和地图。

除上述通用真实数据集外,近年来也出现了若干针对特定问题构建的新型基准。V2X-Radar 通过引入 4D 毫米波雷达、激光雷达和多视角相机,在复杂城市路口环境中采集多天气、多光照条件下的协同数据,为研究复杂环境下的多模态融合与鲁棒协同感知提供了重要基础[73]。V2V-GoT-QA 则面向多模态大语言模型在协同自动驾驶中的应用而构建,在真实协同感知数据基础上设计了空间定位、遮

挡目标检索、显著风险识别以及未来轨迹规划等问答任务,用于评估多智能体融合多端信息后在高层场景理解和语义决策方面的能力[91]。

OPV2V 与 V2XSet 作为仿真基准,广泛应用于区域级特征筛选、特征压缩与稀疏表征等方法的性能对比;DAIR-V2X 与 V2V4Real 因其真实采集的时空异步与传感器异质特性,常被用于验证时空姿态对齐及车路异构融合的鲁棒性;V2X-R 与

V2X-Radar通过引入 4D 毫米波雷达模态,专门支撑异构语义对齐与多模态生成式重构研究;V2X-Seq 提供连续时序标注,适用于时序对齐与运动补偿评估;V2V-GoT-QA 则面向大模型语义协同,评估生成式通信与高层场景理解。

综上所述,仿真数据集与真实数据集在协同感知研究中承担着不同但互补的评测角色。前者强调环境可控、对比公平和消融便利,适合验证通信压缩、特征筛选和生成式恢复等方法的相对效果;后者则更强调真实噪声、异步扰动和工程复杂性,是检验时空对齐鲁棒性、异构融合能力和部署可行性的重要基准。未来的数据集构建仍有必要进一步加强对通信链路扰动、动态拓扑演化以及时空连续真值的标准化描述,从而更全面地支撑高效协同感知方法的比较、验证与落地。

5 挑战与展望

尽管协同感知高效传输研究已在信息筛选、协同建模、特征压缩与生成式恢复等方面取得显著进展,但在真实交通场景中,仍面临场景认知不足、动态网络扰动、开放拓扑安全风险以及重构可信性不足等深层挑战。为此,本节从场景认知瓶颈、动态拓扑与安全防御、物理约束与重构校验三个方面展开讨论。

5.1 场景认知瓶颈

在车路协同感知系统中,严格的无线通信带宽约束要求模型必须精准筛选并传输高价值的感知特征。现有方法为降低通信负载,通常采用空间置信度图、检测分数或响应强度等启发式指标进行特征筛选,并主要依赖局部检测响应进行决策。然而,这种由置信度得分驱动的筛选方式,缺乏对场景上下文、遮挡关系、道路结构以及潜在风险的联合建模能力,并忽略了多智能体协同中信息是否被自车需要与信息是否能被协作者有效提供之间的本质差异,导致通信交互在需求端与供给端均产生显著的语义错位。

在协同需求端,低置信度区域并不必然意味着存在值得请求的协同信息,其既可能对应于不存在目标的真实空旷背景,也可能对应于由障碍物遮挡、距离衰减或视角受限所形成的高风险感知盲区。若系统缺乏对场景几何结构和上下文语义的进一步辨别能力,仅依据低置信度触发协同请求,就可能将低价值背景区域与真正需要补盲的危险区域混为一谈,从而引入大量无效请求并造成通信带宽的浪费。

在协同供给端,仅依据置信度阈值进行传输裁

剪也可能造成关键信息误删。大量远距离、严重遮挡或处于长尾分布中的危险目标,虽然在单节点视角下只表现为微弱的响应和极低的置信度,却往往正是自车最需要的关键互补信息。若系统缺乏更深层的场景认知能力,仅凭单一置信度指标主导筛选,便可能将这些高价值困难样本误当作背景噪声直接剔除,从而削弱协同网络对真实感知盲区的弥补能力,并限制了整体协同收益的上限。

当前,高效传输方法在如何压缩与如何筛选方面已取得明显进展,但对何种信息在具体场景中真正值得传输的认知仍然不足。未来协同感知系统需要由基于局部响应的启发式筛选,进一步走向基于场景理解、风险语义和任务需求的认知驱动型通信决策。

5.2 动态拓扑与安全防御

真实交通网络具有显著的时空动态特征,车辆高速运动、路况持续变化以及通信条件波动共同导致节点间的连接关系不断重构。在车路协同场景中,车辆位置、相对距离以及可通信关系会随时间快速变化,使协同网络中的交互拓扑难以长期保持稳定。节点的频繁接入、退出与重连不仅增加了链路管理与协作组织的复杂度,也使多智能体之间难以形成持续稳定的协同关系。此外,不同节点在传感器采样频率、计算资源、通信时延以及消息转发路径上的差异,还会进一步造成协同信息在时间戳和空间坐标上的异步错位。若缺乏稳定有效的时空对齐机制,即便各节点传输的信息本身是准确的,也可能因拓扑变化与时序失配而削弱整体协同效果。

除动态时空扰动外,开放式车联网拓扑还引入了更高层面的安全风险。车联网本质上属于缺乏绝对中心控制的开放分布式网络,其去中心化特征虽然赋予系统较高灵活性,却也使其在面临恶意节点接入、特征污染与对抗攻击时暴露出显著的信任漏洞。任一被劫持节点都可能借助正常通信链路进入局部协同拓扑,并向邻近节点注入伪造感知结果、污染特征或对抗性扰动。由于协同感知具有显著的信息传播与聚合特性,局部异常一旦进入融合链路,便可能沿着拓扑持续扩散,对全局特征表达和下游决策造成连锁影响。

当前,大量协同感知框架依赖图神经网络或注意力机制进行多节点特征交互,这类机制天然具备信息传播与聚合能力,但通常缺乏对输入特征合法性的深度甄别能力。局部错误响应或恶意表征可能在多轮聚合后迅速放大,最终导致虚假目标生成、真实危险障碍物抹除,甚至引发系统性感知失效与决

策崩溃。因此,如何在克服高动态时空校准难题、保障极低交互时延的同时,在多节点开放拓扑中建立基于共识验证、交叉检验与异常隔离的零信任防御机制,已成为构建鲁棒车路协同网络的关键挑战。

5.3 物理约束与重构校验

在极端带宽受限条件下,基于统计先验的特征补全与场景重构为协同感知提供了一条新的技术路径^[107]。这类方法对缺失信息的恢复并不直接来源于新的传感器观测,而主要依赖训练数据分布与已有残缺输入的联合推断。因此,系统生成的重构内容未必能够准确对应真实环境中的物理实体。当依赖此类结果弥补通信压缩和视角遮挡带来的信息缺口时,统计推断结果与物理空间高保真建模要求之间的不一致性便逐渐显现,其潜在风险主要体现在长尾目标结构失真与多节点感知一致性下降两个方面。

在真实交通环境中,真正对行车安全构成威胁的往往是散落货物、侧翻车辆或形态异常障碍物等长尾风险目标。当这些目标因严重遮挡或通信受限仅保留有限特征线索时,基于概率推断的补全机制容易对异常结构进行平滑处理,并将其误判为常规目标或背景区域。由于补全过程并未引入新的物理观测依据,而只是基于有限输入对缺失区域进行统计意义上的填充,因此系统对高不确定区域的识别能力可能被表面上的完整重构结果所掩盖,从而延后风险告警、冗余确认与安全降级机制的触发时机,增加高风险目标漏检的概率^[108]。

在多节点协同场景中,这一问题还会进一步演化为跨节点一致性风险。协同决策的可靠性依赖不同通信节点围绕同一物理场景形成稳定一致的感知结果^[109-110]。然而,当多个节点在各自残缺观测基础上独立进行特征补全时,由于初始观测视角不同、推断路径存在随机性,不同节点在同一物理坐标域内生成的语义目标往往存在显著差异。这种跨节点语义分歧会降低多源冗余特征之间的可比性,使原本用于相互校验的协作数据转化为不一致的噪声输入,从而削弱全局协作决策的稳定性。因此,未来生成式协同方法不仅要关注能否补全,还必须进一步回答补全是否真实、是否可信、是否一致这3个更关键的问题。如何在利用统计先验恢复缺失特征的同时,引入物理几何约束、不确定性评估以及跨节点一致性校验机制,是生成式协同走向安全应用的关键前提。

5.4 发展趋势

(1) 基于场景认知的协同信息筛选。针对置信

度筛选导致的供需错位,未来研究需要构建融合场景上下文、遮挡关系与风险语义的深层认知机制,以更准确地区分低价值背景区域与高危感知盲区,实现通信资源与协同需求之间的高效匹配。通过增强系统对复杂环境的语义分辨能力,不仅能够抑制冗余信息传输,还能确保长尾困难目标在协同网络中得到可靠保留与有效交互^[111]。

(2) 面向动态拓扑的安全协同机制。针对高动态开放网络中的时空扰动与信任风险,未来亟需探索融合轻量级时空同步、异常检测与共识验证的安全协同架构。一方面,应提高跨节点时空对齐精度,缓解高速运动条件下的坐标漂移与时序失配;另一方面,应面向恶意节点的接入和异常特征的传播建立动态识别与隔离机制,从而有效阻断错误信息在开放拓扑中的级联扩散^[112]。

(3) 融合物理约束的特征补全与重构校验。针对统计重构引发的长尾目标失真与跨节点感知分歧,未来应在特征补全过程中引入三维几何、道路拓扑和先验地图等物理约束,并同步建立跨节点一致性校验和不确定性评估机制。这样既能够抑制异常结构的错误平滑,也能够帮助融合中心区分真实观测与统计推断结果,从而在观测证据不足时及时触发安全降级与风险控制。

6 结 论

车路云协同感知是突破单车智能物理瓶颈、迈向高阶自动驾驶的重要技术路径,有限的无线通信带宽与多源异构数据交互构成了其规模化部署的核心约束。本文围绕自动驾驶协同感知中的高效传输问题,系统梳理了近年来在特征筛选、协同建模与对齐、信息编码与压缩等方面的代表性方法,并总结了相关数据集及未来发展趋势。

在技术层面,现有研究一方面通过空间区域级与实例任务级的信息筛选机制抑制了冗余背景传输,提升了有限带宽下的通信利用效率。另一方面,通过动态拓扑构建、时空位姿对齐与跨模态语义映射,增强多智能体之间的信息协同能力。在此基础上,矢量量化、稀疏表示、生成式补全以及多模态大模型等技术的引入,进一步拓展了协同感知在极端带宽条件下的紧凑表达与高保真恢复能力。

协同感知高效传输研究正由单纯降低通信负载,逐步转向兼顾感知性能、系统鲁棒性与语义理解能力的综合优化范式。然而,在真实复杂交通场景中,系统仍面临场景认知不足、动态网络扰动、开放

拓扑安全风险以及生成式重构可信性不足等关键挑战。未来,协同感知网络需进一步融合场景深度认知、物理几何约束、跨节点一致性校验与安全协同机制,为构建全天候、高鲁棒、可落地的智能交通系统提供更坚实的技术支撑。

参考文献:

- [1] Caillot A, Ouerghi S, Vasseur P, et al. Survey on cooperative perception in an automotive context [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(9): 14204-14223.
- [2] Lang A H, Vora S, Caesar H, et al. PointPillars: fast encoders for object detection from point clouds [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2019: 12689-12697.
- [3] Yan Yan, Mao Yuxing, Li Bo. SECOND: sparsely embedded convolutional detection [J]. *Sensors*, 2018, 18(10): 3337.
- [4] 胡鹏, 夏晓华, 钟预全, 等. 采用多尺度特征增强的路面病害检测模型 [J]. *西安交通大学学报*, 2025, 59(2): 156-169.
Hu Peng, Xia Xiaohua, Zhong Yuquan, et al. Pavement disease detection model based on multi-scale feature reinforcement [J]. *Journal of Xi'an Jiaotong University*, 2025, 59(2): 156-169.
- [5] 孙士清, 彭建华, 游伟, 等. 5G 网络下资源感知的服务功能链协同构建和映射算法 [J]. *西安交通大学学报*, 2020, 54(8): 140-148.
SunShiqing, Peng Jianhua, You Wei, et al. A coordinating composition and mapping algorithm for a service function chain with resource-aware [J]. *Journal of Xi'an Jiaotong University*, 2020, 54(8): 140-148.
- [6] Gao Bolin, Liu Jiayi, Zou Hengduo, et al. Vehicle-road-cloud collaborative perception framework and key technologies: a review [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(12): 19295-19318.
- [7] Cui Guangzhen, Zhang Weili, Xiao Yanqiu, et al. Cooperative perception technology of autonomous driving in the internet of vehicles environment: a review [J]. *Sensors*, 2022, 22(15): 5535.
- [8] 洪东洋, 张慧, 张译匀, 等. 车路群智协同感知方法综述 [J]. *智能科学与技术学报*, 2025, 7(2): 143-164.
Hong Dongyang, Zhang Hui, Zhang Yiyun, et al. A review of collaborative intelligent perception methods for vehicle-road systems [J]. *Chinese Journal of Intelligent Science and Technology*, 2025, 7(2): 143-164.
- [9] 张新钰, 卢毅果, 高鑫, 等. 向智能网联汽车的车路协同感知技术及发展趋势 [J]. *自动化学报*, 2025, 51(2): 233-248.
- [10] ZhangXinyu, Lu Yiguo, Gao Xin, et al. Vehicle-road collaborative perception technology and development trend for intelligent connected vehicles [J]. *Acta Automatica Sinica*, 2025, 51(2): 233-248.
- [11] Yu Haibao, Yang Wenxian, Zhong Jiaru, et al. End-to-end autonomous driving through V2X cooperation [C]//Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2025: 9598-9606.
- [12] Kong Xianghao, Jiang Wentao, Jia Jinrang, et al. DU-SA: decoupled unsupervised Sim2real adaptation for vehicle-to-everything collaborative perception [C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 2023: 1943-1954.
- [13] Zhou Haibo, Xu Wenchao, Chen Jiacheng, et al. Evolutionary V2X technologies toward the internet of vehicles: challenges and opportunities [J]. *Proceedings of the IEEE*, 2020, 108(2): 308-323.
- [14] Chen Qi, Tang Sihai, Yang Qing, et al. Cooper: cooperative perception for connected autonomous vehicles based on 3D point clouds [C]//2019 IEEE 39th International Conference on Distributed Computing Systems (ICDCS). Piscataway, NJ, USA: IEEE, 2019: 514-524.
- [15] Arnold E, Dianati M, de Temple R, et al. Cooperative perception for 3D object detection in driving scenarios using infrastructure sensors [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2022, 23(3): 1852-1864.
- [16] Ye G, Spiegel P, Althoff M. Cooperative raw sensor data fusion for ground truth generation in autonomous driving [C]//2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC). Piscataway, NJ, USA: IEEE, 2020: 1-7.
- [17] Luo Guiyang, Shao Chongzhang, Cheng Nan, et al. EdgeCooper: network-aware cooperative LiDAR perception for enhanced vehicular awareness [J]. *IEEE Journal on Selected Areas in Communications*, 2024, 42(1): 207-222.
- [18] Zheng Zhaoliang, Xia Xin, Gao Letian, et al. CooperFuse: a real-time cooperative perception fusion framework [C]//2024 IEEE Intelligent Vehicles Symposium (IV). Piscataway, NJ, USA: IEEE, 2024: 533-538.

- [18] Barbieri L, Tedeschini B C, Brambilla M, et al. Deep learning-based cooperative LiDAR sensing for improved vehicle positioning [J]. *IEEE Transactions on Signal Processing*, 2024, 72: 1666-1682.
- [19] Chen Qi, Ma Xu, Tang Sihai, et al. F-cooper: feature based cooperative perception for autonomous vehicle edge computing system using 3D point clouds [C]// *Proceedings of the 4th ACM/IEEE Symposium on Edge Computing*. New York, USA: Association for Computing Machinery, 2019: 88-100.
- [20] Guo Jingda, Carrillo D, Tang Sihai, et al. CoFF: cooperative spatial feature fusion for 3-D object detection on autonomous vehicles [J]. *IEEE Internet of Things Journal*, 2021, 8(14): 11078-11087.
- [21] Guo Jingda, Carrillo D, Chen Qi, et al. Slim-FCP: lightweight-feature-based cooperative perception for connected automated vehicles [J]. *IEEE Internet of Things Journal*, 2022, 9(17): 15630-15638.
- [22] Wang T H, Manivasagam S, Liang Ming, et al. V2VNet: vehicle-to-vehicle communication for joint perception and prediction [C]// *ECCV 2020*. Cham, Switzerland: Springer International Publishing, 2020: 605-621.
- [23] Wei Sizhe, Wei Yuxi, Hu Yue, et al. Asynchrony-robust collaborative perception via bird's eye view flow [C]// *Proceedings of the 37th International Conference on Neural Information Processing Systems*. Red Hook, NY, USA: Curran Associates Inc., 2023: 28462-28477.
- [24] Han Yushan, Zhang Hui, Li Huifang, et al. Collaborative perception in autonomous driving: methods, datasets, and challenges [J]. *IEEE Intelligent Transportation Systems Magazine*, 2023, 15(6): 131-151.
- [25] Liu Si, Gao Chen, Chen Yuan, et al. Towards vehicle-to-everything autonomous driving: a survey on collaborative perception [PP/OL]. *arXiv* (2023-08-31) [2026-03-13]. <https://arxiv.org/abs/2308.16714>.
- [26] Bai Zhengwei, Wu Guoyuan, Barth M J, et al. A survey and framework of cooperative perception: from heterogeneous singleton to hierarchical cooperation [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(11): 15191-15209.
- [27] Gao Xin, Zhang Xinyu, Lu Yiguo, et al. A survey of collaborative perception in intelligent vehicles at intersections [J/OL]. *IEEE Transactions on Intelligent Vehicles*. <https://ieeexplore.ieee.org/document/10517450/>. DOI: 10.1109/TIV.2024.3395783.
- [28] Huang Tao, Liu Jianan, Zhou Xi, et al. Vehicle-to-everything cooperative perception for autonomous driving [J]. *Proceedings of the IEEE*, 2025, 113(5): 443-477.
- [29] Hu Senkang, Fang Zhengru, Deng Yiqin, et al. Collaborative perception for connected and autonomous driving: challenges, possible solutions and opportunities [J]. *IEEE Wireless Communications*, 2025, 32(5): 228-234.
- [30] Xu Junhao, Zhang Yanan, Cai Zhi, et al. CoSDH: Communication-efficient collaborative perception via supply-demand awareness and intermediate-late hybridization [C]// *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ, USA: IEEE, 2025: 6834-6843.
- [31] Tang Zongheng, Liu Yi, Sun Yifan, et al. CoST: efficient collaborative perception from unified spatiotemporal perspective [C]// *2025 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway, NJ, USA: IEEE, 2025: 1120-1129.
- [32] Wang Tianhang, Chen Guang, Chen Kai, et al. UMC: a unified bandwidth-efficient and multi-resolution based collaborative perception framework [C]// *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Piscataway, NJ, USA: IEEE, 2023: 8153-8162.
- [33] Fan Siqi, Yu Haibao, Yang Wenxian, et al. QUEST: query stream for practical cooperative perception [C]// *2024 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ, USA: IEEE, 2024: 18436-18442.
- [34] Shenkut D, Bhagavatula V. FocalComm: hard instance-aware multi-agent perception [C]// *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. Piscataway, NJ, USA: IEEE, 2026: 6277-6286.
- [35] Liu Y C, Tian Junjiao, Ma C Y, et al. Who2com: collaborative perception via learnable handshake communication [C]// *2020 IEEE International Conference on Robotics and Automation (ICRA)*. Piscataway, NJ, USA: IEEE, 2020: 6876-6883.
- [36] Liu Y C, Tian Junjiao, Glaser N, et al. When2com: multi-agent perception via communication graph grouping [C]// *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ, USA: IEEE, 2020: 4105-4114.
- [37] Xu Runsheng, Xiang Hao, Tu Zhengzhong, et al. V2X-ViT: vehicle-to-everything cooperative perception with vision transformer [C]// *ECCV 2022*. Cham, Switzerland: Springer Nature Switzerland, 2022: 107-124.
- [38] Li Chenyang, Xu Langkun, Jin Chenye, et al. GraphPS:

- graph pair sequences-based noisy-robust multi-hop collaborative perception [J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(2): 3895-3905.
- [39] Xiang Hao, Xu Runsheng, Ma Jiaqi. HM-ViT: hetero-modal vehicle-to-vehicle cooperative perception with vision transformer [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 284-295.
- [40] Gao Xiangbo, Xu Runsheng, Li Jiachen, et al. STAMP: scalable task-and model-agnostic collaborative perception [C]//Proceedings of the 2025 International Conference on Learning Representations. New York, USA: ICLR, 2025: 54656-54676.
- [41] Lu Yifan, Hu Yue, Zhong Yiqi, et al. An extensible framework for open heterogeneous collaborative perception [C]//Proceedings of the 2024 International Conference on Learning Representations. New York, USA: ICLR, 2024: 44576-44593.
- [42] Yang Kun, Yang Dingkan, Zhang Jingyu, et al. What2comm: towards communication-efficient collaborative perception via feature decoupling [C]//Proceedings of the 31st ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 2023: 7686-7695.
- [43] Hu Yue, Peng Juntong, Liu Sifei, et al. Communication-efficient collaborative perception via information filling with codebook [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 15481-15490.
- [44] Wang Binglu, Zhang Lei, Wang Zhaozhong, et al. Core: cooperative reconstruction for multi-agent perception [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 8676-8686.
- [45] Fu Jiahui, Gong Yue, Wang Luting, et al. Generative map priors for collaborative BEV semantic segmentation [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2025: 11919-11928.
- [46] Hu Yue, Fang Shaoheng, Lei Zixing, et al. Where2comm: communication-efficient collaborative perception via spatial confidence maps [C]//Proceedings of the 36th International Conference on Neural Information Processing Systems. Red Hook, NY, USA: Curran Associates Inc., 2022: 4874-4886.
- [47] Tao Yihang, Hu Senkang, Fang Zhengru, et al. Directed-CP: directed collaborative perception for connected and autonomous vehicles via proactive attention [C]//2025 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2025: 7004-7010.
- [48] Chen Ziming, Shi Yifeng, Jia Jinrang. TransIFF: an instance-level feature fusion framework for vehicle-infrastructure cooperative 3D detection with transformers [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2023: 18159-18168.
- [49] Wang Zhe, Xu Shaocong, Zhuang Xucai, et al. Coop-DETR: a unified cooperative perception framework for 3D detection via object query [C]//2025 International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2025: 2732-2739.
- [50] Wang Rujia, Xiang Hao, Li Jinlong, et al. CoCMT: communication-efficient cross-modal transformer for collaborative perception [C]//2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2025: 2471-2478.
- [51] Huang Suozhi, Zhang Juexiao, Li Yiming, et al. ActFormer: scalable collaborative perception via active queries [C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2024: 14716-14723.
- [52] Yazgan M, Wu Qiyuan, Hamdard I, et al. SlimComm: doppler-guided sparse queries for bandwidth-efficient cooperative 3-D perception [C]//2025 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Piscataway, NJ, USA: IEEE, 2025: 1803-1812.
- [53] Li Yiming, Ren Shunli, Wu Pengxiang, et al. Learning distilled collaboration graph for multi-agent perception [C]//Advances in Neural Information Processing Systems. [S.l.]: [s.n.], 2021: 29541-29552.
- [54] Chiu H K, Smith S F. Selective communication for cooperative perception in end-to-end autonomous driving [PP/OL]. arXiv (2023-05-26) [2026-02-26]. <https://arxiv.org/abs/2305.17181>.
- [55] Wang Wentao, Xu Haoran, Tan Guang. InterCoop: spatio-temporal interaction aware cooperative perception for networked vehicles [C]//2024 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2024: 14443-14449.
- [56] Fang Zhengru, Hu Senkang, An Haonan, et al. PACP: priority-aware collaborative perception for connected and autonomous vehicles [J]. *IEEE Transactions on Mobile Computing*, 2024, 23(12): 15003-15018.

- [57] Liu Yuntao, Huang Qian, Li Rongpeng, et al. Select2col: leveraging spatial-temporal importance of semantic information for efficient collaborative perception [J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(9): 12556-12569.
- [58] Ren Shunli, Lei Zixing, Wang Zi, et al. Interruption-aware cooperative perception for V2X communication-aided autonomous driving [J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(4): 4698-4714.
- [59] Shi Jingyue, Zhao Junhui, Zhuo Li, et al. V2V cooperative perception with adaptive communication loss for autonomous driving [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(10): 14866-14878.
- [60] Dao M Q, Berrio J S, Frémont V, et al. Practical collaborative perception: a framework for asynchronous and multi-agent 3D object detection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(9): 12163-12175.
- [61] Hu Senkang, Fang Zhengru, An Haonan, et al. Adaptive communications in collaborative perception with domain alignment for autonomous driving [C]//GLOBECOM 2024-2024 IEEE Global Communications Conference. Piscataway, NJ, USA: IEEE, 2024: 746-751.
- [62] Lin Chunmian, Tian Daxin, Duan Xuting, et al. V2VFormer: vehicle-to-vehicle cooperative perception with spatial-channel transformer [J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(2): 3384-3395.
- [63] Song Zhiying, Xie Tenghui, Zhang Hailiang, et al. A spatial calibration method for robust cooperative perception [J]. *IEEE Robotics and Automation Letters*, 2024, 9(5): 4011-4018.
- [64] Zhou Zewei, Xiang Hao, Zheng Zhaoliang, et al. V2XPnP: vehicle-to-everything spatio-temporal fusion for multi-agent perception and prediction [C]//2025 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ, USA: IEEE, 2025: 25399-25409.
- [65] Yin Hongbo, Tian Daxin, Lin Chunmian, et al. V2VFormer++: multi-modal vehicle-to-vehicle cooperative perception via Global-Local transformer [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2024, 25(2): 2153-2166.
- [66] Ma Yunsheng, Lu Juanwu, Cui Can, et al. MACP: efficient model adaptation for cooperative perception [C]//2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). Piscataway, NJ, USA: IEEE, 2024: 3361-3370.
- [67] Wei Quanmin, Dai Penglin, Li Wei, et al. CoPEFT: fast adaptation framework for multi-agent collaborative perception with parameter-efficient fine-tuning [C]//Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2025: 23351-23359.
- [68] Xu Runsheng, Li Jinlong, Dong Xiaoyu, et al. Bridging the domain gap for multi-agent perception [C]//2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2023: 6035-6042.
- [69] Shao Congzhang, Yuan Quan, Luo Guiyang, et al. NegoCollab: a common representation negotiation approach for heterogeneous collaborative perception [C]//39th Conference on Neural Information Processing Systems (NeurIPS 2025). Red Hook, NY, USA: Curran Associates Inc., 2025: 1-22.
- [70] Shao Congzhang, Luo Guiyang, Yuan Quan, et al. Hetecooper: feature collaboration graph for heterogeneous collaborative perception [C]//ECCV 2024. Cham, Switzerland: Springer Nature Switzerland, 2025: 162-178.
- [71] Wei Chuheng, Qin Ziye, Zimmer W, et al. HeCoFuse: cross-modal complementary V2X cooperative perception with heterogeneous sensors [C]//2025 IEEE 28th International Conference on Intelligent Transportation Systems (ITSC). Piscataway, NJ, USA: IEEE, 2025: 1214-1221.
- [72] Huang Xun, Wang Jinlong, Xia Qiming, et al. V2X-R: cooperative LiDAR-4D radar fusion with denoising diffusion for 3D object detection [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2025: 27390-27400.
- [73] Yang Lei, Zhang Xinyu, Li Jun, et al. V2X-radar: a multi-modal dataset with 4D radar for cooperative perception [C]//39th Conference on Neural Information Processing Systems (NeurIPS 2025) Track on Datasets and Benchmarks. Red Hook, NY, USA: Curran Associates Inc., 2025: 1-21.
- [74] 邹承益, 万帅, 朱志伟, 等. 采用轻量级卷积神经网络的 H.266/通用视频编码跨分量预测 [J]. *西安交通大学学报*, 2025, 59(2): 180-188.
- Zou Chengyi, Wan Shuai, Zhu Zhiwei, et al. Cross-component prediction for H.266/versatile video coding

- based on lightweight convolutional neural network [J]. Journal of Xi'an Jiaotong University, 2025, 59(2): 180-188.
- [75] Li Jinlong, Liu Xinyu, Li Baolu, et al. CoMamba: real-time cooperative perception unlocked with state-space models [C]//2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2025: 16993-17000.
- [76] Ding Zihan, Fu Jiahui, Liu Si, et al. Point cluster: a compact message unit for communication-efficient collaborative perception [C]//Proceedings of the 2025 International Conference on Learning Representations. Red Hook, NY, USA: Curran Associates Inc., 2025: 1-17.
- [77] Yuan Yunshuang, Xia Yan, Cremers D, et al. Sparse Align: a fully sparse framework for cooperative object detection [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2025: 22296-22305.
- [78] Shenkut D, Kumar B V K V. Residual vector quantization for communication-efficient multi-agent perception [C]//ICASSP 2026-2026 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2026: 19807-19811.
- [79] Wei Quanmin, Dai Penglin, Li Wei, et al. InfoCom: kilobyte-scale communication-efficient collaborative perception with information bottleneck [C]//Proceedings of the Fortieth AAAI Conference on Artificial Intelligence and Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence and Sixteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2026: 29731-29739.
- [80] Wu Yuheng, Nguyen T T, Liebe L, et al. How2compress: scalable and efficient edge video analytics via adaptive granular video compression [C]//Proceedings of the 33rd ACM International Conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 2025: 12006-12015.
- [81] 方芯, 王伟, 贺丽君, 等. 结构-语义-残差协同的图像压缩框架: 面向隐私保护与高保真重建 [J]. 西安交通大学学报, 2026, 60(7): 185-195.
- Fang Xin, Wang Wei, He Lijun, et al. A structure-semantics-residual collaborative image compression framework: toward privacy protection and high-fidelity reconstruction [J]. Journal of Xi'an Jiaotong University, 2026, 60(7): 185-195.
- [82] 王富平, 王定莎, 李藕, 等. 细粒度深度特征掩码估计的遮挡人脸识别算法 [J]. 西安交通大学学报, 2025, 59(2): 170-179.
- Wang Fuping, Wang Dingsha, Li Ou, et al. Occluded face recognition algorithm based on fine-grained deep feature mask estimation [J]. Journal of Xi'an Jiaotong University, 2025, 59(2): 170-179.
- [83] Zou Jiayu, Tian Kun, Zhu Zheng, et al. DiffBEV: conditional diffusion model for bird's eye view perception [C]//Proceedings of the Thirty-Eighth AAAI Conference on Artificial Intelligence and Thirty-Sixth Conference on Innovative Applications of Artificial Intelligence and Fourteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2024: 7846-7854.
- [84] Zhou Junfei, Dai Penglin, Wei Quanmin, et al. Pragmatic heterogeneous collaborative perception via generative communication mechanism [C]//39th Conference on Neural Information Processing Systems (NeurIPS 2025). Red Hook, NY, USA: Curran Associates Inc., 2025: 1-26.
- [85] Xu Haoran, Zhang Saining, Li Peishuo, et al. CRUISE: cooperative reconstruction and editing in V2X scenarios using Gaussian splatting [C]//2025 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2025: 12518-12525.
- [86] Gao Xiangbo, Wu Yuheng, Wang Rujia, et al. LangCoop: collaborative driving with language [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Piscataway, NJ, USA: IEEE, 2025: 4226-4237.
- [87] You Junwei, Jiang Zhuoyu, Huang Zilin, et al. V2X-VLM: end-to-end V2X cooperative autonomous driving through large vision-language models [J]. Transportation Research: Part C Emerging Technologies, 2026, 183: 105457.
- [88] Ma Yingzi, Cao Yulong, Sun Jiachen, et al. Dolphins: multimodal language model for driving [C]//ECCV 2024. Cham, Switzerland: Springer Nature Switzerland, 2025: 403-420.
- [89] Fang Shiyu, Liu Jiaqi, Cui Yiming, et al. Towards interactive and learnable cooperative driving automation: a large language Model-Driven Decision-Making framework [J]. IEEE Transactions on Vehicular Technology, 2025, 74(8): 11894-11905.
- [90] Hu Senkang, Fang Zhengru, Fang Zihan, et al. AgentsCoDriver: large language model empowered col-

- laborative driving with lifelong learning [PP/OL]. arXiv (2024-04-21) [2026-03-15]. <https://arxiv.org/abs/2404.06345>.
- [91] Chiu H K, Hachiuma R, Wang C Y, et al. V2V-LLM: vehicle-to-vehicle cooperative autonomous driving with multimodal large language models [PP/OL]. arXiv (2026-02-16) [2026-02-23]. <https://arxiv.org/abs/2502.09980>.
- [92] Sima Chonghao, Renz K, Chitta K, et al. DriveLM: driving with graph visual question answering [C]//ECCV 2024. Cham, Switzerland: Springer Nature Switzerland, 2025: 256-274.
- [93] Chiu H K, Hachiuma R, Wang C Y, et al. V2V-GoT: vehicle-to-vehicle cooperative autonomous driving with multimodal large language models and graph-of-thoughts [PP/OL]. arXiv (2026-02-16) [2026-03-11]. <https://arxiv.org/abs/2509.18053>.
- [94] Tian Hanlin, Reddy K, Feng Yuxiang, et al. Large (vision) language models for autonomous vehicles: current trends and future directions [J]. IEEE Transactions on Intelligent Transportation Systems, 2026, 27(1): 187-210.
- [95] Padilla R, Netto S L, da Silva E A B. A survey on performance metrics for object-detection algorithms [C]//2020 International Conference on Systems, Signals and Image Processing (IWSSIP). Piscataway, NJ, USA: IEEE, 2020: 237-242.
- [96] Li Yiming, Ma Dekun, An Ziyang, 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.
- [97] Xu Runsheng, Xiang Hao, Xia Xin, et al. OPV2V: an open benchmark dataset and fusion pipeline for perception with vehicle-to-vehicle communication [C]//2022 International Conference on Robotics and Automation (ICRA). Piscataway, NJ, USA: IEEE, 2022: 2583-2589.
- [98] Yu Haibao, Luo Yizhen, Shu Mao, et al. DAIR-V2X: a large-scale dataset for vehicle-infrastructure cooperative 3d object detection [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2022: 21329-21338.
- [99] Yang Kang, Wang Peng, Li Lantao, et al. EIMC: efficient instance-aware multi-modal collaborative perception [PP/OL]. arXiv (2026-03-03) [2026-03-16]. <https://arxiv.org/abs/2603.02532>.
- [100] Han Yushan, Zhang Hui, Xia Qiming, et al. CoLC: communication-efficient collaborative perception with LiDAR completion [PP/OL]. arXiv (2026-02-28) [2026-03-08]. <https://arxiv.org/abs/2603.00682>.
- [101] Xu Runsheng, Guo Yi, Han Xu, et al. OpenCDA: an open cooperative driving automation framework integrated with Co-Simulation [C]//2021 IEEE International Intelligent Transportation Systems Conference (ITSC). Piscataway, NJ, USA: IEEE, 2021: 1155-1162.
- [102] Dosovitskiy A, Ros G, Codevilla F, et al. CARLA: an open urban driving simulator [C]//Proceedings of the 1st Annual Conference on Robot Learning. Chia Laguna Resort, Sardinia, Italy: PMLR, 2017: 1-16.
- [103] Yu Haibao, Yang Wenxian, Ruan Hongzhi, et al. V2X-seq: a large-scale sequential dataset for vehicle-infrastructure cooperative perception and forecasting [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 5486-5495.
- [104] Arnold E, Mozaffari S, Dianati M. Fast and robust registration of partially overlapping point clouds [J]. IEEE Robotics and Automation Letters, 2022, 7(2): 1502-1509.
- [105] Xu Runsheng, Xia Xin, Li Jinlong, et al. V2V4Real: a real-world large-scale dataset for vehicle-to-vehicle cooperative perception [C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2023: 13712-13722.
- [106] HaoRuiyang, Fan Siqi, Dai Yingru, et al. RCooper: a real-world large-scale dataset for roadside cooperative perception [C]//2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2024: 22347-22357.
- [107] Ni Chaojun, Zhao Guosheng, Wang Xiaofeng, et al. ReconDreamer: crafting world models for driving scene reconstruction via online restoration [C]//2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2025: 1560-1569.
- [108] Ni Peizhou, Zhu Wenkai, Jiang Feng, et al. Perception balance with uncertainty-guided fusion and proposal-wise mixture-of-experts for robust multi-agent 3D object detection [J]. Expert Systems with Applications, 2026, 310: 131311.
- [109] Zhao Yangheng, Xiang Zhen, Yin Sheng, et al. MADE: malicious agent detection for robust multi-agent collaborative perception [C]//2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Piscataway, NJ, USA: IEEE, 2025: 13817-13823.

- [110] Hu Senkang, Tao Yihang, Xu Guowen, et al. CP-guard: malicious agent detection and defense in collaborative bird's eye view perception [C]//Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2025: 23203-23211.
- [111] Guan Runwei, Hu Rongsheng, Chen Shangshu, et al. RoadSceneVQA: benchmarking visual question answering in roadside perception systems for intelligent transportation system [C]//Proceedings of the Fortieth AAAI Conference on Artificial Intelligence and Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence and Sixteenth Symposium on Educational Advances in Artificial Intelligence. Palo Alto, CA, USA: AAAI Press, 2026: 4366-4375.
- [112] Wang Yulong, Sun Tong, Li Shenghong, et al. Adversarial attacks and defenses in machine learning-empowered communication systems and networks: a contemporary survey [J]. IEEE Communications Surveys & Tutorials, 2023, 25(4): 2245-2298.

(编辑 刘懿莹)