

多智能体协作感知的现状与展望



Multi-Agent Cooperative Perception: Current Status and Prospects

刘根嘉/LIU Genjia, 陈思衡/CHEN Siheng,
张文军/ZHANG Wenjun

(上海交通大学, 中国 上海 200240)
(Shanghai Jiao Tong University, Shanghai 200240, China)

DOI: 10.12142/ZTETJ.202504006

网络出版地址: <http://kns.cnki.net/kcms/detail/34.1228.TN.20250711.2051.008.html>

网络出版日期: 2025-07-14

收稿日期: 2025-05-08

摘要: 多智能体协作感知作为人工智能与分布式系统交叉的重要方向, 近年来在自动驾驶、无人系统编队、智能监控等领域受到广泛关注。梳理了该领域的研究背景与发展脉络, 重点讨论了通信机制优化、位姿噪声下的鲁棒性、异构协同、安全与隐私保护及仿真平台等研究热点。在此基础上, 探讨了实现高效通信的便捷方案, 并展望了未来发展趋势。可为相关研究提供参考与启示, 推动多智能体协作感知技术的持续进步。

关键词: 协作感知; 多智能体系统; 通信编码; 自动驾驶

Abstract: Multi-agent cooperative perception, as a key interdisciplinary direction between artificial intelligence and distributed systems, has attracted widespread attention in recent years in areas such as autonomous driving, unmanned system formation, and intelligent surveillance. This paper reviews the research background and development of this field, focusing on key topics including communication mechanism optimization, robustness under pose noise, heterogeneous cooperation, security and privacy protection, and simulation platforms. On this basis, this paper further explores a lightweight solution for efficient communication and discusses future research trends. This work aims to provide insights and references for researchers, promoting the continuous advancement of multi-agent cooperative perception technology.

Keywords: cooperative perception; multi-agent system; communication encoding; autonomous driving

引用格式: 刘根嘉, 陈思衡, 张文军. 多智能体协作感知的现状与展望 [J]. 中兴通讯技术, 2025, 31(4): 34-40. DOI: 10.12142/ZTETJ.202504006

Citation: LIU G J, CHEN S H, ZHANG W J. Multi-agent cooperative perception: current status and prospects [J]. ZTE technology journal, 2025, 31(4): 34-40. DOI: 10.12142/ZTETJ.202504006

1 多智能体协作感知背景

感知是智能系统面向物理世界的窗口, 是其执行任务的先决条件。2023年5月更新的《美国国家人工智能研发战略计划》^[1]指出, 增强智能系统的感知能力是一项需要长期投入的基础性研究工作。智能系统需整合各种来源(包括边缘设备)的数据, 以实现态势感知。然而传统单体感知方案无法克服因视线遮挡或远距离导致的感知信息缺失问题。这种物理限制和信息孤岛现象是单一智能体的固有瓶颈, 直接导致自动驾驶等智能系统难以在实际场景中部署落地。

针对上述问题, 借助现代通信技术, 多智能体协作感知技术应运而生。该技术在交通、生产、军事和科学探索等领域展现出显著的应用价值与发展潜力^[2]。传统感知方法通常基于单智能体架构进行环境信息采集。然而, 单一智能体受限于硬件和视野范围, 难以克服物体遮挡、视距约束及设备故障等固有问题。在此背景下, 多智能体协作感知技术通过

信息共享突破单智能体的感知局限, 增强系统容错能力, 成为诸多关键场景的重要解决方案^[3]。多智能体协作感知技术的发展可追溯至20世纪80年代的分布式人工智能研究, 1997年美国国家航空航天局(NASA)的“火星探路者”成功实现了人类首次地外多智能体协作探测, 为后续技术发展奠定了重要基础。进入21世纪后, 随着深度学习技术的突破性进展、5G低延迟通信和协作同步定位与地图构建(SLAM)算法等关键技术的日趋成熟, 多智能体协作感知技术正加速从实验室研究向实际应用跨越式发展。展望未来, 协作感知技术将向在智慧交通、远程医疗与国防等更广阔领域拓展。然而, 多智能体协作感知仍面临诸多挑战, 需提升六大关键能力: 协作通信效率能力、信息融合能力、异构协作能力、协作定位能力、安全隐私防护能力、仿真验证能力。

协作通信效率与信息融合能力是多智能体系统的神经中枢。协作感知面临的关键挑战是如何平衡感知性能与通信带宽。现实中的通信资源有限, 难以支持实时传输完整观测数据。为此, 系统需要筛选并传输紧凑有效的感知信息, 在降

基金项目: 国家自然科学基金项目(62450162)

低通信开销的同时保持良好的感知效果。通常自感知智能体首先根据与协作智能体的距离、自身的感知死角等信息动态连接协作智能体，随后协作智能体会对协作信息进行筛选和压缩，在进行低带宽传输的同时保证较低的感知失真。接收协作信息后，智能体需要对协作信息进行解码，并且通过数据融合模型，将自感知信息与协作感知信息融合，最终达到增强感知性能的目的。

异构协作与协作定位能力构成了系统的骨架支撑。随着协作规模扩大和下游任务复杂度提升，系统亟需具备动态成员管理能力，包括智能体的发现、身份认证与协作性验证等。特别是在多智能体之间存在传感器配置、任务目标和性能差异的异构场景下，如何实现跨异构智能体的有效协作，成为系统可扩展性与稳定性面临的重要挑战^[4-5]。精确的相对位姿估计是协作感知的基础，在全球定位系统（GPS）拒止环境中，需要基于视觉、激光雷达等传感器实现分布式协作定位，同时解决定位误差累积问题^[6]。

安全隐私与仿真验证能力是系统可靠应用的保障。安全方面需建立多层次防护体系：通信链路加密防止窃听，数据完整性校验抵御篡改^[7]，差分隐私技术保护敏感信息，如行人面部等。同时，系统应具备异常检测和自恢复能力，当恶意攻击者通过伪造或篡改传输数据（例如发送虚假障碍物信息）破坏系统可靠性时，系统需识别恶意数据并恢复原始信息。此外，仿真验证平台在降低开发成本和规避风险方面具有重要作用。理想的平台应支持复杂环境建模、智能体行为模拟以及感知算法的测试与评估。目前，英伟达的Isaac Sim、微软的AirSim等平台已初步具备多智能体协作感知的仿真能力，为相关研究与应用提供了有力支撑。

2 多智能体协作感知研究现状

2.1 通信效率

通信效率是协作感知系统实时部署和规模扩展所面临的关键瓶颈，这是因为在协作感知系统中，随着节点数量的增加和任务复杂性的提升，大规模协作不可避免。为保障系统的整体性能和通信资源利用效率，需要设计高效且紧凑的通信策略，以支撑大规模群体间的高效协作。信息完整度最高的通信策略是前期协作^[8]，如图1（a）所示，即在模型推理开始前进行协作，直接将原始观测数据作为协作信息进行传递。这种方法虽然保留了最完整的信息，但带宽开销巨大，难以

直接应用于高频实时的协作通信。为了降低通信量以适应有限的带宽，最直接的方式是进行后期协作，如图1（c）所示，即在模型推理完成后进行协作，仅传输感知结果，此时带宽利用效率较高。但这种方法对协作信息的融合总结能力有限，无法充分发挥协作优势。因此，中期协作开始受到研究者的关注，如图1（b）所示。这类方法传输压缩后的网络中间层特征，在通信开销与感知性能之间取得平衡。例如，V2VNet^[9]和DiscoNet^[10]分别引入熵编码技术与自动编码器，以便更高效地表示特征图；When2com^[11]则通过握手机制排除收益较低的协作对象，将全连接通信图优化为稀疏图结构，从而减少通信冗余。然而，现有方法大多只能采用固定压缩率，无法适应实际应用中的动态带宽情况。为此，一种可行的思路是，通过空间维度的特征筛选降低通信量，根据带宽限制自适应调整需要传输的和任务相关的特征区域。

2.2 位姿扰动下的鲁棒协作

在多智能体协作感知系统中，位姿对齐是实现多个智能体间数据同步和协作任务的重要基础。每个智能体通过其独立的定位系统来估算位姿。这些位姿数据需要在统一的坐标系中进行对齐，才能保证多个智能体的协作效率。然而，在实际应用中，定位系统并非总能提供精准的位姿信息，尤其是在复杂环境中，位姿估计常常受到传感器精度限制或环境干扰的影响，导致定位误差。这些位姿误差会影响智能体间的相对位姿估计，进而削弱协作感知的效果。为了应对这一问题，近年来研究者提出了多种方法来校正位姿扰动。普遍的思路是，根据 N 个视角对 M 个物体级或像素级的锚点

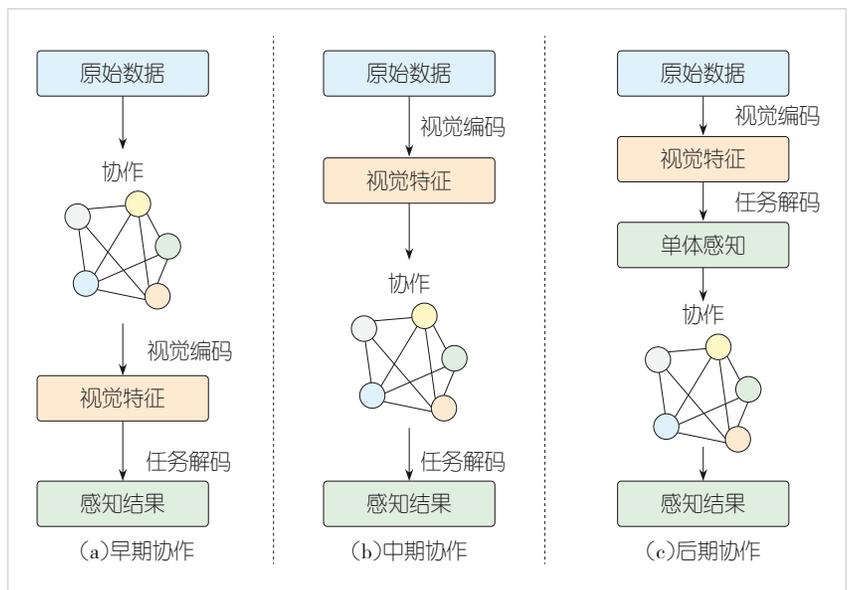


图1 前、中、后期协作示意图

$\left\{ \left\{ a_{ij} \right\}_{j=1}^M \right\}_{i=1}^N$ 的相对观测值, 对齐 N 个不同视角的位姿 $\{ \xi_i \}_{i=1}^N$,

可建模为如下优化问题, 其中 d 是位姿匹配误差度量函数:

$$\min_{\{ \xi_i \}_{i=1}^N} d \left(\left\{ \left\{ a_{ij} \right\}_{j=1}^M \right\}_{i=1}^N, \{ \xi_i \}_{i=1}^N \right) \quad (1).$$

例如, MASH^[12] 采用像素级锚点计算相对位姿关系, 通过构建相似度体积, 显式地学习像素级的对应关系, 从而避免使用噪声位姿进行推理。FPV-RCNN^[13] 通过推断关键点的语义标签, 并采用最大共识算法来建立智能体之间的对应关系, 校正位姿误差。CoAlign^[6] 通过智能体-物体图优化, 进一步提高了位姿一致性, 从而减少由定位误差引起的影响。RoCo^[14] 提出了一种创新的无监督框架, 通过迭代的物体匹配和智能体位姿调整, 解决协作感知中由于位姿误差导致的特征对齐问题, 并通过图优化过程最小化关联物体的对齐误差, 从而提升协作物体检测的性能和鲁棒性。尽管上述方法在一定程度上提升了位姿鲁棒性, 但仍存在对初始估计依赖较强、难以应对大尺度场景中高频动态变化的问题。未来研究可探索通过时序一致性约束实现更稳定的跨帧位姿优化。

2.3 异构场景下的协作

异构性是协作感知系统面临的关键挑战之一, 体现在多个层面: 不同传感器模态、不同模型架构和不同参数配置等, 每种异构性都会增加协作信息的融合难度。利用不同尺寸、模态、数据分布的视觉特征进行融合协作, 通常有两种方式: 1) 针对新的异构种类增添新的特征适配器或单独的特征融合模块; 2) 通过数据驱动的方式学习统一的特征空间。HMViT^[15] 提出了一种异构图注意力网络, 对不同种类的信息融合的学习进行区分, 从而降低异构协作的学习难度。MPDA^[16] 提出多智能体感知领域适配方法, 利用对抗领域对齐的学习框架, 促进异构智能体对之间的特征分布对齐。然而, 随着智能体数量的增加, 上述方法的扩展性受到限制。对此, HEAL^[4] 提出了反向对齐训练策略, 通过固定基础网络的解码器并仅训练编码器来创建异构模型, 实现现有智能体的协作。PolyInter^[17] 通过引入多态特征解释器, 在异构环境中无须训练新解释器, 利用可扩展的学习参数集成新智能体, 显著提高了协作感知的精度并保持了较低参数调整需求。STAMP^[18] 提出了一种可扩展的任务和模型无关的协作感知框架, 通过使用轻量级适配器-还原器对在共享协议域和特定智能体域之间转换鸟瞰图特征, 促进高效的特征共享与融合, 同时降低计算开销。已有方法在特定类型的异构协作中取得了效果, 但在面对传感器类型过多或新设备频繁接入时, 系

统适配可能仍面临困难。后续研究可以考虑构建泛化性更强的特征适配模块, 减少对每种设备单独设计的依赖。

2.4 安全与隐私

协作感知依赖智能体间的信息共享, 虽显著提升了整体感知能力, 但也带来了更多攻击风险与隐私泄露隐患。恶意智能体可能通过伪造、篡改或注入错误感知数据, 干扰系统判断, 甚至引发误识别、漏检等安全事故。因此, 构建协作环境下的安全防护机制已成为研究焦点。目前的主流方法聚焦于异常攻击检测, 通过对多智能体感知结果的一致性分析, 实现对潜在恶意行为的识别与隔离。例如, ROBOSAC^[19] 采用迭代方式并结合匈牙利匹配算法, 在协作过程中实现感知一致性, 以识别并剔除可能的恶意智能体。MADE^[7] 在此基础上引入多测试框架, 通过匹配损失与重建损失的联合判别策略, 实现更稳健的感知一致性验证。尽管上述基于异常检测的策略在识别明显攻击方面表现良好, 但面对更加隐蔽、精准的攻击时, 仍存在较大漏洞。为此, 最新研究开始集中于中间特征级别的一致性检测方向探索, 以提升检测的灵敏度和效率。CP-Guard+^[20] 通过引入特征级的恶意智能体检测机制以及双中心对比损失, 在无需验证最终感知结果的前提下有效区分正常与恶意特征, 大幅降低计算开销并增强系统鲁棒性。GCP^[21] 则结合空间一致性与时间行为分析, 利用贝叶斯-霍克测试与运动流重建, 实现对细粒度持续性攻击的高精度检测。目前的方法大多依赖已知攻击模式或规则进行检测, 对于未知类型或动态演化的攻击应对能力仍不足。未来可以考虑引入在线学习或自适应更新机制, 使系统在运行过程中持续优化其安全策略, 提升应对未知威胁的能力。

2.5 环境仿真

在协作感知研究中, 仿真环境已成为连接算法原型与实际道路的重要桥梁。相比高成本、难复现的实车实验, 仿真平台提供了“零风险、可控、可复现”的试验场。研究人员可灵活调节交通密度、天气等变量, 定向构建极端或长尾场景, 并进行多轮回放与消融实验, 系统分析性能瓶颈。

当前仿真方案可大体分为3类技术路线。第1类是端到端数字孪生型平台, 例如以CARLA^[22] 为代表的城市级仿真。作为协作感知场景最常用的仿真平台, 此类平台在视觉与物理真实感上保持高一致性, 支持多车、多传感器和交通行为建模, 可直接用于闭环自动驾驶实验。然而, 其高保真亦意味着对图形处理器 (GPU) 渲染与存储性能的极大消耗, 同时在快速导入真实道路素材、扩展路侧和空中视角方面仍需二次开发。第2类是图像重渲型方法, 以 RoCo-Sim^[23] 等工作

为典型，将高精度三维车辆资产嵌入真实路侧摄像机图像中，实现多视角几何一致渲染。此类技术继承了真实背景的照明与材质信息，生成速度可较端到端平台提高一个量级，尤其适合固定相机与车-路协作的感知-通信研究；但现阶段主要针对静态背景，对大范围动态场景的支持有限。第3类是显式体渲染或 3D Gaussian Splatting^[24] 方案，通过稀疏高斯体或稀疏体素完成连续视图重绘，可较传统基于三角网格的渲染速度提高数十到百倍，并天然支持可微分重定位与外观迁移，已成为快速合成大规模训练数据的新热点。

3 多智能体协作的通信便捷方案

传统基于熵编码和自编码器的压缩技术能在一定程度上减小协作通信量，但在实际感知任务中仍存在明显局限：大量空间区域并不包含关键信息，传输完整的特征图会造成显著带宽浪费；同时，高维特征通道也带来额外的通信开销。本文介绍一种多智能体协作的通信便捷方案，从空间和特征通道这两个维度筛选信息并压缩协作通信量，如图2所示。在空间维度上，采用一种基于置信度补全的信息选择机制：每个智能体根据自身感知结果生成信息评分图，用以评估不同空间区域的重要性，并根据多个协作者的信息评分图筛选关键区域的稀疏但有价值的特征。在特征通道维度上，引入可学习码本进行向量量化，用离散的码本索引代替高维连续特征，从而降低单位地块通信量。

3.1 基于置信度补全的信息选择

为了高效选择对其他智能体有支持作用的紧凑协作消

息，每个智能体应有选择性地分享与任务相关的关键信息，从而使多个智能体所提供的非冗余信息实现互补，满足彼此的信息需求。为实现这一目标，本信息选择策略包含信息评分生成和信息选择优化两个核心步骤。

首先，对于智能体 i ，每个智能体使用信息评分生成器 $\Phi_{\text{generator}}$ 基于其自身的特征图 \mathcal{F}_i 生成信息评分图 C_i ：

$$C_i = \Phi_{\text{generator}}(\mathcal{F}_i) \in [0, 1]^{H \times W} \quad (2)$$

图 C_i 反映其在不同空间区域上的可用信息量。之后该评分图被广播给所有其他智能体，从而促进智能体之间对彼此可支持信息的全面认知。其中 H 、 W 代表特征图的分辨率， $\Phi_{\text{generator}}$ 由感知解码器实现，通过赋予包含目标区域更高的信息评分，帮助补全漏检内容。

基于所有 N 个智能体共享的评分图 $\{C_i\}_{i=1}^N$ ，每个智能体在本地优化信息选择矩阵，选择最适合支持其他智能体的消息内容。由于每个智能体仅在某些空间区域需要特定信息来实现精准感知，超出部分的信息并不会带来明显收益，因此每个支持者需优先选择那些评分较高且与其他信息来源不冗余的空间区域来提供支持。接收者在信息需求被满足后将停止进一步的消息选择，以避免通信浪费。

具体而言，该问题被建模为一个约束优化问题，其目标是为每对支持者-接收者智能体计算出一个二值选择矩阵，以确定是否在特定空间位置传输信息。记 $M_{i \rightarrow j} \in \{0, 1\}^{H \times W}$ 为在鸟瞰图上定义的二值选择矩阵，其中矩阵中的每个元素表示：智能体 i 是否应在该空间位置向智能体 j 发送信息（1 表示发送，0 表示不发送），求解目标如下所示：

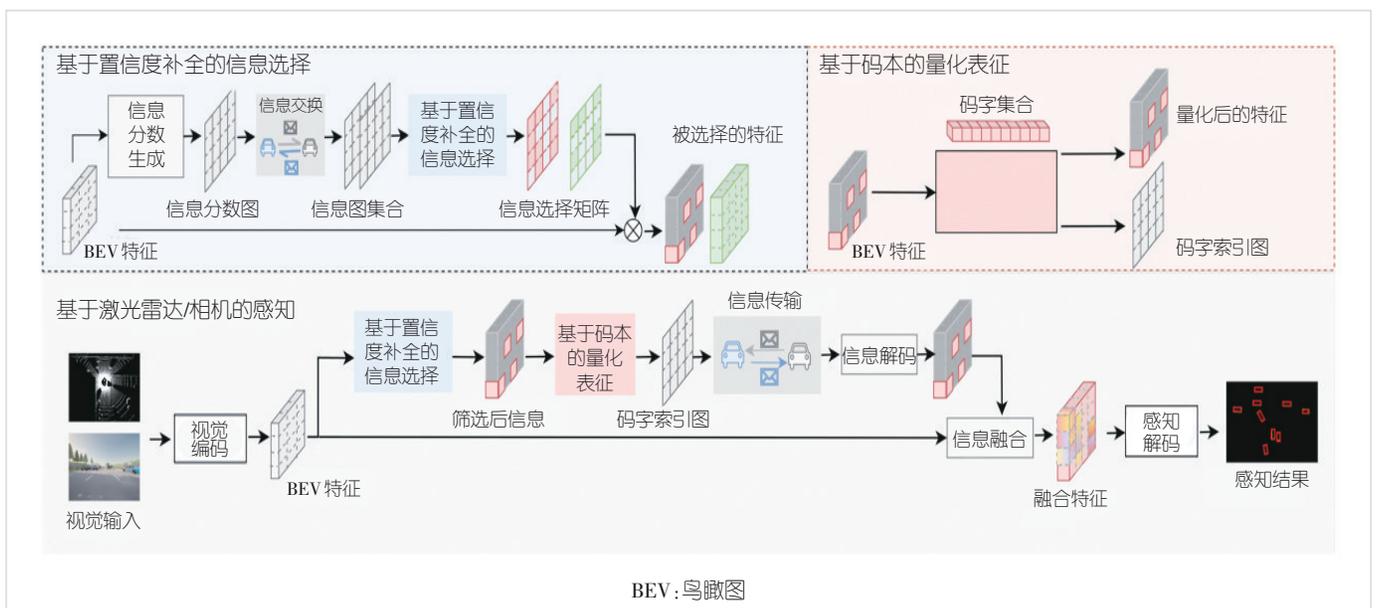


图2 多智能体协作的通信便捷方案

$$\{\mathbf{M}_{i \rightarrow j}^*\}_{i,j} = \operatorname{argmax}_M \sum_{j=1}^N f_{\min} \left(\mathbf{C}_j + \sum_{i=1, i \neq j}^N \mathbf{M}_{i \rightarrow j} \odot \mathbf{C}_i, \mathbf{u} \right) \quad (3),$$

其中, $\sum_{i,j=1, j \neq i}^N \mathbf{M}_{i \rightarrow j} \leq b, \mathbf{M}_{i \rightarrow j} \in \{0,1\}^{H \times W}$, \odot 代表点乘, u 是一个代表信息需求阈值的超参数, b 代表通信量预算, f_{\min} 计算选择矩阵与阈值之间的较小值。基于选择矩阵 $\{\mathbf{M}_{i \rightarrow j}^*\}_{j=1}^N$, 智能体 i 为 j 发送的信息最终为 $\mathcal{Z}_{i \rightarrow j} = \mathbf{M}_{i \rightarrow j}^* \odot \mathcal{F}_i$ 。

3.2 基于码本的表征量化压缩

为高效表示高维特征图 $\mathcal{Z}_{i \rightarrow j}$, 智能体采用一种基于码本的消息表示方法, 以降低在通道维度上的通信开销。其核心思想是: 利用一个任务驱动的码本, 将高维特征向量近似为码本中最接近的一个码字, 从而只需传输整数形式的码字索引, 无须传输完整的浮点特征向量, 大幅减少通信数据量。

具体而言, 令 $F = \{\mathcal{F}^{(i,s)}\}_{i=1, s=1}^{N,S}$ 表示由所有 N 个智能体在 S 个训练场景中通过观测编码器提取的鸟瞰图 (BEV) 特征集合。令 $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_{n_L}] \in \mathbb{R}^{C \times n_L}$ 表示码本, $\mathbf{D}_{[l]} = \mathbf{d}_l \in \mathbb{R}^C$ 表示第 l 个码字, n_L 代表码字总数, C 代表原始特征维度。码本学习优化目标可表示为:

$$\mathbf{D}^* = \operatorname{arg min}_{\mathbf{D}} \sum_{\mathcal{F} \in F} \sum_{h,w} \min_l \left(\Psi(\mathbf{D}_{[l]}) + (\mathcal{F}_{[h,w]} - \mathbf{D}_{[l]})_2^2 \right) \quad (4),$$

其中, $\Psi(\cdot)$ 表示将原始特征用码字近似表征后的目标检测表现。公式 (4) 中第一项对应下游检测任务的需求, 第二项则反映了原始特征向量与所选码字之间的重构误差。尽管该近似在重构上是有损的, 但对于感知任务而言是无损的, 从而在不影响感知能力的前提下有效降低通信开销。在测试阶段, 基于共享的码本 \mathbf{D} , 每个智能体可以将所选的稀疏特征图 $\mathcal{Z}_{i \rightarrow j}$ 替换为一系列码字索引 $\mathcal{I}_{i \rightarrow j}$ 。对于每一个 BEV 空间位置 (h, w) , 其对应的码字索引通过以下方式获得:

$$\left(\mathcal{I}_{i \rightarrow j} \right)_{[h,w]} = \operatorname{arg min}_l \left(\left(\mathcal{Z}_{i \rightarrow j} \right)_{[h,w]} - \mathbf{D}_{[l]} \right)_2^2 \quad (5)。$$

基于码本的向量量化压缩在降低通信开销方面具有显著优势, 但需要注意多个关键难点。一是, 码本容量的选择需根据具体任务复杂度进行动态调整: 容量过小可能导致信息表达能力不足, 影响感知精度; 容量过大则会增加存储与通信负担, 削弱压缩效果。二是, 码本训练需在特征重建率与任务适配性之间实现权衡。为此, 码本训练同时采用了重建损失和感知损失。这种双重优化机制要求量化后的向量尽可能保留原有信息, 能对关键任务保持高敏感性。

这种基于码本表示方式具有三大优势: 1) 高效性: 通过传输轻量级的码字索引, 大幅减少通信负载; 2) 适应性: 可通过调整码本配置灵活适应不同的通信资源条件, 较

小的码本提升通信效率, 较大的码本增强感知性能; 3) 可扩展性: 提供统一的共享表示格式, 使得新加入的异构智能体能够通过将其有效的感知特征基加入码本, 轻松融入协作系统。信息接收端将根据索引和共享的码本还原协作信息, 并与本地信息融合用于最终的感知推理。

3.3 实验分析

图 3 展示了置信度筛选和码本压缩在 DairV2X^[25] 真实协作数据集上不同通信带宽下的感知精度。实验设置包括: 检测范围为 204.8 m × 102.4 m, 感知分辨率为 0.4 m × 0.4 m, 最大通信半径为 100 m, 码本容量为 256, 评估指标为在重合度 (IoU) 阈值为 0.3、0.5、0.7 时的平均精度 (AP)。从结果中可以观察到, 该方法展现出以下显著优势: 1) 在所有通信带宽条件下, 置信度筛选+码本压缩方法均优于传统的前期协作与后期协作策略。尤其值得注意的是, 该方法仅使用约 1/256 的通信量, 便可达到前期协作的感知精度。2) 基于置信度筛选的机制使系统具备了自适应通信能力, 即使在 60 倍压缩率下仍保留 95% 的完整通信精度。相比之下, 其他方法往往只能在固定带宽下发挥作用, 缺乏灵活性。3) 基于码本学习的压缩方案显著提升了特征通道维度上的压缩效率, 实现了高达 100 倍的通信压缩比, 充分展现了码本在协作感知中的有效性与可扩展性。

4 未来研究展望

4.1 空天地异构

空天地一体化系统正成为多智能体协作感知的重要发展方向。在灾害应急场景中, 卫星可快速获取全局图像, 无人机提供中尺度巡查, 地面传感器则用于局部高精度监测, 三类智能体形成互补。在智慧交通中, 卫星导航、空中平台监控与地面路侧单元协同, 可实现复杂路况下的实时感知与调度。此外, 该体系在低空物流、森林防火、农业遥感等新型场景中同样具备应用潜力, 已被《数字交通“十四五”规划》与多地“空天信息产业行动计划”纳入重点建设内容。但由于空、天、地智能体之间存在显著异构性, 协作面临多方面技术挑战: 感知数据在分辨率、时间同步性和观测范围上存在差异, 难以直接融合; 通信协议与链路多样, 需应对跨平台互通、链路切换与频谱冲突问题; 计算资源分布不均, 星载/空载平台计算能力受限, 需动态任务分配; 智能体运动特性不同, 轨迹调度复杂。此外, 多域协作中还涉及数据安全、身份认证等系统性难题。

针对这些问题, 未来研究可从以下几个方向推进: 构建

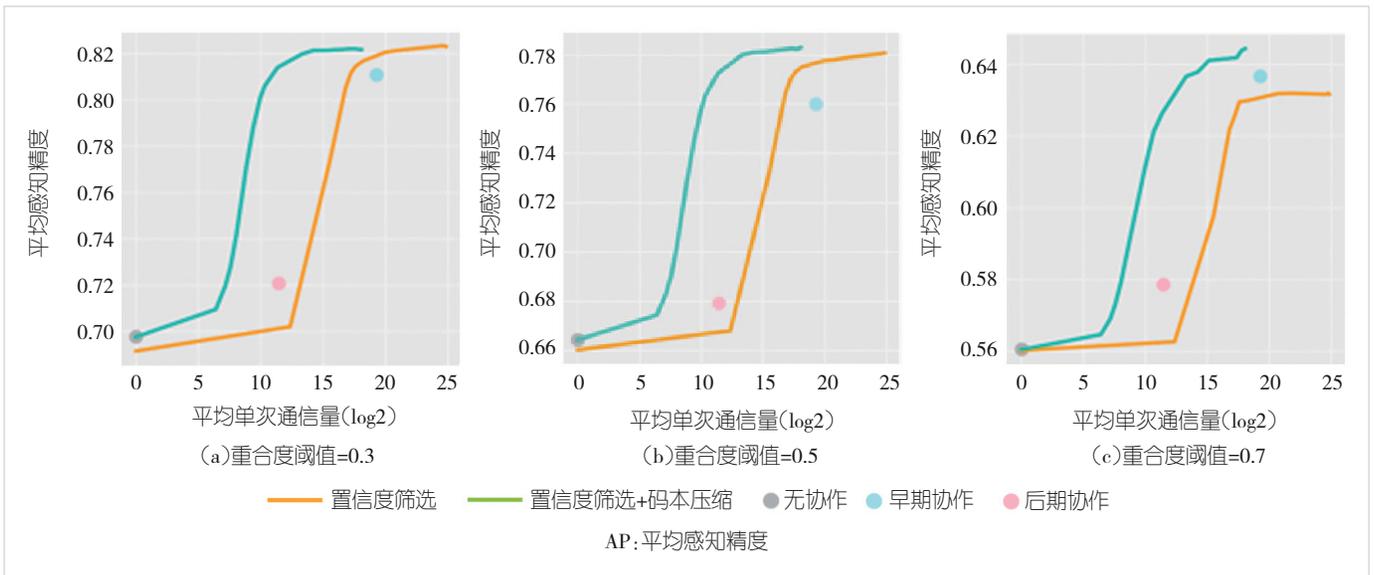


图3 不同通信量下的协作感知精度

统一的多源异构数据融合框架，实现时空对齐与语义整合；设计适用于星-空-地的通信协议栈与弹性链路调度机制；设计感知任务驱动的计算调度算法，实现跨平台的推理负载均衡；开发面向异构运动模型的多智能体协作控制策略，提升任务执行一致性；同时，完善分层安全机制，确保跨域协同过程中的数据隐私与系统安全。

4.2 多具身智能协作

具身智能正从单一智能体系统迈向“多形态、多功能、跨时空”的协作体。在工业、物流、交通等领域，地面车辆、机械臂、空中无人机等多种智能体协同完成复杂任务，展现出显著的效率与灵活性。例如，配送车与空中无人机的协同投递、道路清障机械臂与巡检车的协同运维，以及车与路侧单元组成的混合网络^[26]，均已进入研究视野。这一趋势不仅为传统移动机器人注入新动能，也对感知、任务分配与实时调度提出了全新要求。

与此同时，大规模协作也伴随协作安全、通信拥塞与伦理治理等挑战。集中式规划在大规模系统中难以适用，去中心化的图神经网络与事件驱动通信机制成为研究热点^[27]。目前，缺乏覆盖多种智能体综合任务的统一评测体系，这限制了算法的横向比较与产业落地。未来研究需在安全性、实时性与可评测性之间寻求平衡，并引入更加开放的跨域验证基准，促使学术界与产业界达成共识。

5 结束语

本文梳理了多智能体协作感知的研究脉络与技术发展，围绕通信效率、位姿鲁棒性、异构协作、安全隐私和仿真平

台五大关键问题分析了研究现状，提出了一种面向动态带宽的通信便捷方案，通过空间维度的置信度筛选与特征通道维度的码本量化压缩，实现了通信负载与感知精度之间的高效权衡。实验结果表明，该方法在多种带宽条件和协作场景下均显著优于现有主流方案，具备良好的带宽自适应性和较高感知精度。此外，本文还对空天地一体化协作与多具身系统的发展方向进行了展望，指出协作感知将在异构智能体融合、跨域通信调度与安全协同等方面持续面临挑战。总体而言，本文梳理了协作感知研究中的关键问题，并针对通信效率这一问题提供了一种可行的解决方案。

致谢

感谢上海交通大学在读研究生胡安宁、刘畅行，以及南京理工大学在读研究生臧子晨对本研究工作的支持！

参考文献

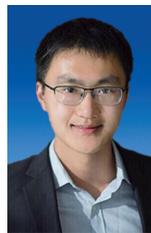
- [1] National Artificial Intelligence Agency. Research and development blueprint plan: 2023 update [EB/OL]. (2023-05-23) [2025-04-05]. <https://www.nitrd.gov/pubs/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf>
- [2] ALOTAIBI E T, ALQEFARI S S, KOUBAA A. LSAR: multi-UAV collaboration for search and rescue missions [J]. IEEE access, 2019, 7: 55817-55832
- [3] CHEN Q, MA X, TANG S H, 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. ACM, 2019: 88-100. DOI: 10.1145/3318216.3363300
- [4] LU Y F, HU Y, ZHONG Y Q, et al. An extensible framework for open heterogeneous collaborative perception [EB/OL]. (2024-01-25) [2025-04-05]. <https://arxiv.org/abs/2401.13964>
- [5] MENG F, HUANG Y M, LU Z H, et al. Multi-user mmWave beam

- tracking via multi-agent deep Q-learning [J]. ZTE communications, 2023, 21(2): 53–60. DOI: 10.12142/ZTECOM.202302008.
- [6] LU Y F, LI Q H, LIU B A, et al. Robust collaborative 3D object detection in presence of pose errors [C]//Proceedings of IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023: 4812–4818. DOI: 10.1109/ICRA48891.2023.10160546
- [7] ZHAO Y H, XIANG Z, YIN S, et al. MADE: malicious agent detection for robust multi-agent collaborative perception [C]//Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2024: 13817–13823. DOI: 10.1109/IROS58592.2024.10801337
- [8] CHEN Q, TANG S H, YANG Q, et al. Cooper: cooperative perception for connected autonomous vehicles based on 3D point clouds [C]//Proceedings of IEEE 39th International Conference on Distributed Computing Systems (ICDCS). IEEE, 2019: 514–524. DOI: 10.1109/icdcs.2019.00058
- [9] WANG T H, MANIVASAGAM S, LIANG M, et al. V2VNet: vehicle-to-vehicle communication for joint perception and prediction [M]//Computer Vision – ECCV 2020. Cham: Springer International Publishing, 2020: 605–621. DOI: 10.1007/978-3-030-58536-5_36
- [10] LI Y, REN S, WU P, et al. Learning distilled collaboration graph for multi-agent perception [EB/OL]. (2021-11-21)[2025-04-05]. <https://arxiv.org/abs/2111.00643>
- [11] LIU Y C, TIAN J J, GLASER N, et al. When2com: multi-agent perception via communication graph grouping [C]//Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020: 4105–4114. DOI: 10.1109/CVPR42600.2020.00416
- [12] GLASER N, LIU Y, TIAN J, et al. Overcoming obstructions via bandwidth-limited multi-agent spatial handshaking [EB/OL]. (2021-07-01)[2025-04-05]. <https://arxiv.org/abs/2107.00771>
- [13] YUAN Y S, CHENG H, SESTER M. Keypoints-based deep feature fusion for cooperative vehicle detection of autonomous driving [J]. IEEE robotics and automation letters, 2022, 7(2): 3054–3061. DOI: 10.1109/LRA.2022.3143299
- [14] HUANG Z, WANG S, WANG Y, et al. RoCo: robust cooperative perception by iterative object matching and pose adjustment [C]//Proceedings of the 32nd ACM International Conference on Multimedia. ACM, 2024: 7833–7842
- [15] XIANG H, XU R S, MA J Q. HM-ViT: hetero-modal vehicle-to-vehicle cooperative perception with vision transformer [C]//Proceedings of the IEEE/CVF international conference on computer vision. IEEE, 2023: 284–295
- [16] XU R S, LI J L, DONG X Y, et al. Bridging the domain gap for multi-agent perception [C]//Proceedings of IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023: 6035–6042. DOI: 10.1109/ICRA48891.2023.10160871
- [17] XIA Y C, YUAN Q, LUO G Y, et al. One is plenty: a polymorphic feature interpreter for immutable heterogeneous collaborative perception [C]//Proceedings of the Computer Vision and Pattern Recognition Conference. IEEE, 2025: 1592–1601
- [18] GAO X, XU R, LI J, et al. STAMP: scalable task and model-agnostic collaborative perception [EB/OL]. (2025-01-24)[2025-04-05]. <https://arxiv.org/abs/2501.18616>
- [19] LI Y M, FANG Q, BAI J M, et al. Among us: adversarially robust collaborative perception by consensus [C]//Proceedings of IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2023: 186–195. DOI: 10.1109/ICCV51070.2023.00024
- [20] HU S, TAO Y, FANG Z, et al. CP-guard+: a new paradigm for malicious agent detection and defense in collaborative perception [EB/OL]. (2025-02-07) [2025-04-05]. <https://arxiv.org/abs/2502.07807>
- [21] TAO Y, HU S, HU Y, et al. GCP: guarded collaborative perception with spatial-temporal aware malicious agent detection [EB/OL]. (2025-01-05)[2025-04-05]. <https://arxiv.org/abs/2501.02450>
- [22] DOSOVITSKIY A, ROS G, CODEVILLA F, et al. CARLA: an open urban driving simulator [EB/OL]. (2017-11-10)[2025-04-05]. <https://arxiv.org/abs/1711.03938>
- [23] DU Y W, HU A N, CHAO Z C, et al. RoCo-sim: enhancing roadside collaborative perception through foreground simulation [EB/OL]. (2025-03-13)[2025-04-05]. <https://arxiv.org/abs/2503.10410>
- [24] KERBL B, KOPANAS G, LEIMKUEHLER T, et al. 3D gaussian splatting for real-time radiance field rendering [EB/OL]. (2023-08-08)[2025-04-05]. <https://arxiv.org/abs/2308.04079>
- [25] 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 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2022: 21329–21338. DOI: 10.1109/CVPR52688.2022.02067
- [26] LIU G J, HU Y, XU C X, et al. Towards collaborative autonomous driving: simulation platform and end-to-end system [J]. IEEE transactions on pattern analysis and machine intelligence, 2025: 1–18. DOI: 10.1109/TPAMI.2025.3560327
- [27] LIU C X, LIU G J, WANG Z J, et al. CoLMDriver: LLM-based negotiation benefits cooperative autonomous driving [EB/OL]. (2025-03-11)[2025-04-05]. <https://arxiv.org/abs/2503.08683>

作者简介



刘根嘉, 上海交通大学在读博士研究生; 主要研究领域为端到端自动驾驶、多智能体协作等。



陈思衡, 上海交通大学人工智能学院副教授; 主要研究领域为多智能体系统、大模型智能体等; 入选国家级人才计划青年项目, 承担了国家自然科学基金原创探索计划项目及面上项目、科技部人工智能2030重大项目、上海市科委人工智能专项等科研项目; 曾获得IEEE信号处理协会最佳青年作者论文奖、ASME结构检测协会最佳论文Runner-Up奖、2018 GlobalSIP会议最佳论文奖、三菱电机实验室总统奖等; 发表论文100余篇, 谷歌引用10 000余次。



张文军, 上海交通大学讲席教授、国家自然科学基金委数字媒体通信创新群体带头人、杰出青年基金获得者、教育部首批“长江学者”特聘教授、国家百千万工程领军人才、国务院特殊津贴专家、中国电子学会会士、IEEE Fellow; 主要研究领域为生成式视频通信与媒体网络融合; 以第一完成人获得国家科技进步奖二等奖4项, 并获中国专利金奖、何梁何利科技进步奖和广电行业CCBN学术贡献奖等荣誉; 发表论文140余篇, 获授权发明专利236项(含35项国际发明专利), 主编工业和信息化部“十四五”规划教材《视频通信》。