文章编号:1671-4598(2026)06-0229-07 DOI:10.16526/j.cnki.11-4762/tp.2026.06.029 中图分类号:TP391 文献标识码:A

基于边云协同的车辆场景多模态语义补全与语义压缩方法

叶仕通
(广州华商学院 人工智能学院,广州 511300)

摘要:面向城市道路的遮挡补全与通信受限问题,提出一种车、路、云协同的多模态语义补全与语义压缩方法;该方法在车端融合相机、雷达与惯导,完成首次补全并仅上行紧凑潜表示,在路侧聚合多车潜码进行时空对齐,生成区域引导向量与占据/遮挡提示下发至相关车辆,在云端按区域热度迭代代码本与参数;通过重要度驱动的区域化码率分配、关键帧与差分一体化消息形态降低上传开销,并以双头解码与小步细化提升被遮挡区域的形状与语义质量;实验测试表明,mIoU_occ 相对强基线平均提升约 3%~4%,复杂场景分位段的平均上行开销整体更低,常用带宽条件下端到端 P95 时延约为 100 ms,满足车路协作部署的实时性与带宽约束要求。

关键词:跨语义占据补全;车路云协同;语义压缩;带宽自适应;端到端时延

Multi-Modal Semantic Completion and Compression Method for Vehicle Scenarios Based on Edge-Cloud Collaboration

YE Shitong
(School of Artificial Intelligence, Guangzhou Huashang College, Guangzhou 511300, China)

Abstract: To address occlusion completion and limited communication on urban roads, a multi-modal semantic completion and compression method for vehicle-road-cloud collaboration is proposed, which integrates cameras, radars, and an inertial navigation system (INS) to achieve the first completion and only move up the compact latent representation, aggregate the latent code of multiple vehicles on the roadside for the spatio-temporal alignment, generate and transmit the region guidance vectors and occupancy/occlusion cues to relevant vehicles, and iterate the codebook and parameters based on the cloud regional heatmap, which reduces upload overhead through the importance-driven regional code rate allocation, key frames and differential messages, enhancing the shape and semantic quality of occluded regions via dual-head decoding and small-step refinement. Experimental results demonstrate that compared to a strong baseline, the mIoU_occ index achieves an average improvement of approximately 3%~4%, exhibiting a slower average uplink overhead in complex scene segments, with an end-to-end P95 latency of about 100 ms under typical bandwidth conditions, thereby meeting the real-time and bandwidth constraints for vehicle-road collaboration deployment.

Keywords: semantic occupancy completion; vehicle-road-cloud collaboration; semantic compression; bandwidth adaptation; end-to-end latency

0 引言

面向自动驾驶的场景理解正在从以目标框为核心的检测识别,转向以三维语义占据为载体的整体建模与补全。相比只描述离散目标,占据预测能够同时表达任意形状的实体、遮挡关系与自由空间结构,使规划与决策可以直接在三维语义栅格与可通行空间上进行推理,从而更适应复杂交通参与者、非规则形状目标以及多遮挡交互等城市道路典型情形[1-3]。在真实道路环境中,大型车辆遮挡、路侧设施遮挡、临停目标遮挡与植被遮挡

收稿日期:2025-11-04; **修回日期**:2025-12-26。
基金项目:2024 年广东省普通高校自然科学类平台和项目(2024ZDZX3035)。
作者简介:叶仕通(1981-),男,博士,教授。
引用格式:叶仕通.基于边云协同的车辆场景多模态语义补全与语义压缩方法[J].计算机测量与控制,2026,34(6):229-235,243.

会持续造成观测缺失，远距区域还叠加视差减小与纹理稀疏，使得遮挡边界附近的形状外延与类别判断更易出现偏移，误差随时间传播后会进一步影响风险评估与行为决策的可信度。车路协同逐步走向规模化应用，多主体在空间上形成互补视域，在时间上积累可复用线索，为遮挡区域提供跨主体补充证据，使遮挡推断不再完全依赖单车单帧的局部观测[4]。但协同感知的落地同时受到通信带宽、端到端时延、丢包抖动与软硬件异构等现实约束的牵制，若仍以大尺寸连续特征或密集栅格进行共享，不仅会抬升链路开销与排队时延，也会增加路侧计算与缓存压力，使协同收益在中低带宽预算与多车并发条件下被系统负担抵消[5-6]。因此，在不依赖原始图像大规模上行的前提下，如何组织可传输的信息形态，如何把协同增益更集中地作用于遮挡区域，如何在补全质量与通信开销之间形成可部署的折中，成为面向真实道路部署亟需解决的问题。

国内外研究围绕占据建模与协同感知两条主线已形成代表性路线。占据方向上，SurroundOcc 将多相机特征映射到体素空间并结合 3D 解码与多尺度监督提升整体占据质量，OpenOccupancy 与 Occ3D 从标注密度提升、可见性建模与评测口径统一等角度完善了训练与评价体系[7-9]。协同方向上，V2X ViT 采用跨主体注意力增强多主体融合表达，How2comm 围绕传什么与何时传开展通信侧优化，相关综述进一步系统梳理了异构条件下的协同问题、带宽与时延约束下的通信设计以及评测范式[10-12]。现有工作或侧重占据预测的单体建模，或侧重协同机制与传输策略，在重遮挡与远距信息稀缺条件下的 amodal 补全质量提升，以及在通信受限条件下的信息形态组织与分层分工方面仍有提升空间。针对上述不足，本文提出一种车、路、云协同的多模态语义补全与语义压缩方法，车端完成首次补全并仅上行紧凑潜表示，路侧聚合多车潜码进行时空对齐与一致融合后生成区域引导向量与占据遮挡提示按需回传，通信侧以区域重要度驱动的码率分配与关键帧、差分一体的消息形态组织传输，从而在中低带宽预算下兼顾遮挡区域的 amodal 补全质量与上行开销控制。

1 整体方法框架

本文面向车路云协同的城市道路场景，围绕遮挡条件下的三维语义占据补全与通信受限下的语义级传输，构建一套从车端感知到路侧协作再到云端更新的整体方法，如图 1 所示。整体流程以车端为最小闭合计算单元，在每帧对相机、毫米波雷达与惯导数据进行特征提取与融合，先完成一次本地语义补全，得到可见层结果与 amodal 结果，并同步输出用于通信的低体量语义级表示。随后，车端将关键帧潜表示、差分残差与事件提示作为上行消息发送到路侧单元，路侧在同域范围内聚合多车信息，通过时空对齐与一致融合形成更可靠的占据与遮挡估计，再将压缩后的区域引导向量与遮挡提示按需回传到相关车辆，用于条件化车端解码与细化，从而在不上传原始图像的前提下提高遮挡区域的补全质量与边界表现。

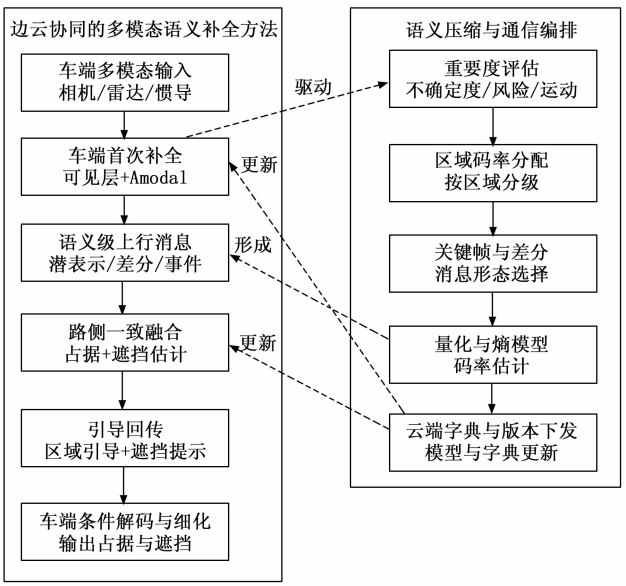

图 1 方法示意图

在通信与调度层面，本文将补全质量需求与链路预算统一到语义压缩与编排策略中，使上行负担随场景复杂度自适应变化。车端先依据不确定度、风险与运动等信息生成重要度图，对候选区域进行分级，并据此分配区域化码率，决定哪些区域需要更精细的表达。时间维度采用关键帧与差分一体的消息形态，当场景变化较大或事件触发时优先发送关键帧潜表示以刷新参考，当变化较小则仅发送稀疏残差以减少开销，同时在分组层面引入与丢包和速率波动相关的冗余配置，缓解弱网络条件下的信息缺失。云端以更长时间尺度汇聚匿名化索引与统计量，周期性更新模型与压缩字典并下发到路侧与车端，路侧负责转播与轻量微调，从而形成分工明确的部署路径，使系统在 50～200 kB 每帧的预算区间内兼顾占据补全效果与平均上行成本。

2 边云协同的多模态语义补全方法

本节给出车端与路侧协同的补全流程。车端负责提取新近细节并完成首次补全，路侧负责融合多车信息并返回紧凑的引导向量。

为统一后续表述，先在车端构建多模态特征入口：

$$F_t = \Gamma\left[f_{\mathrm{cam}}(X_t^{\mathrm{cam}}), f_{\mathrm{rad}}(X_t^{\mathrm{rad}}), f_{\mathrm{imu}}(X^{\mathrm{imu}})_t\right] \quad (1)$$

式中，X_t^{cam} 表示相机帧，X_t^{rad} 表示雷达回波或点迹张量，X_t^{imu} 表示惯导量测，f_{cam}，f_{rad}，f_{imu} 为各模态的特征提

取子网，Γ 为融合算子，可取拼接后再投影或门控加权，输出 F_t 作为统一表征。

为了减少相邻帧差异对补全的影响，对特征进行轻量对齐：$\widetilde{F}_t = W(F_t, H_t, U_t)$，$W$ 为基于单应矩阵与稀疏光流的变换算子，H_t 由路面近似平面假设与相机外参估计得到，U_t 为稀疏光流场。

路侧对多车上传的低比特潜码进行时空融合，先为每辆车分配可信度权重：

$$\alpha_i = softmax_i(\kappa_1 r_i + \kappa_2 IoU_i + \kappa_3 e^{-d_i}) \tag{2}$$

式中，α_i 为车辆 i 的融合权重，r_i 表示其历史不确定度的等效可靠度，IoU_i 为与参考视角的重叠度指标，d_i 为与目标区域的空间距离，κ_1，κ_2，κ_3 为标量系数并在训练中学习。

基于权重，路侧形成占据与遮挡的一致估计，并作为后续引导的基础：

$$O^s = \sum\nolimits_i \alpha_i O_i, M^s = \sum\nolimits_i \alpha_i M_i \tag{3}$$

其中：O_i 为车辆 i 上传的占据概率图或体素概率，M_i 为遮挡置信图，O^s，M^s 分别为路侧融合后的占据与遮挡结果。

路侧将拓扑与运动线索转化为方向提示图，用于指导车端的轮廓延展：

$$D = norm[K * (T, V)] \tag{4}$$

式中，T 为车道中心线与停止线等拓扑栅格，V 为由多车轨迹估计的局部速度场，K 为少量卷积核，$*$ 表示卷积运算，$norm$ 将向量场归一化得到单位方向提示图 D。为了压缩下行开销，路侧仅在候选区域上生成小集合的引导向量：

$$q_k = AttnPool_{R_k}[(O^s, M^s, D)], k = 1, \cdots, K \tag{5}$$

式中，R_k 为第 k 个候选区域，$AttnPool$ 表示注意力加权的区域池化，输出 q_k 为用于条件化解码的小向量集合，K 是本帧的候选数。

车端以对齐特征与引导向量为输入，采用双头解码得到可见层与 amodal 层：

$$[\hat{V}_t, \hat{A}_t] = \mathrm{Dec}(\widetilde{F}_t, \{q_k\}_{k=1}^K) \tag{6}$$

其中：Dec 为共享主干的双头解码器，$\hat{V}$ 为可见语义与边界，$\hat{A}_t$ 为 amodal 估计，二者在细粒度掩膜与边界分支上互不共享以减少信息泄漏。

图 2 给出了车端的计算路径与输出形态，展示了从对齐特征与路侧引导的条件融合进入共享主干，再分出可见与 amodal 两个不共享分支的解码结构。

针对远距与重遮挡区域，引入小步迭代的细化更新，使边界与轮廓逐步变得更贴合：

$$z_t^{s+1} = z_t^s + \eta R[z_t^s, \widetilde{F}_t, (q_k)], z_t^0 = \hat{A}_t \tag{7}$$

式中，z_t^s 表示第 s 次细化的中间结果，η 为步长，R 为残差预测算子，输入包含当前结果与条件引导。实际实现时 s 取极小常数即可带来可见改善。为了在视频中保持前后一致，引入基于光流的时序约束[13]：

$$\mathcal{L}_{\mathrm{temp}} = mean_u \left\| \hat{A}_t(u) - warp(\hat{A}_{t-1}, U_t(u)) \right\|_1 \tag{8}$$

其中：$warp$ 为用光流 U_t 将上一帧结果映射到当前帧的算子，u 为像素或体素索引，mea 表示在前景与重要区域上的平均。路侧与车端之间存在重叠视域时，再增加跨车一致项以减少多视角下的差别：

$$\mathcal{L}_{\mathrm{cons}} = mean_{i,j} \left\| \Pi_{i \to j}(\hat{A}_t^i) - \hat{A}_t^j \right\|_1 \tag{9}$$

式中，$\hat{A}_t^i$ 为车辆 i 的 amodal 结果，$\Pi_{i \to j}$ 表示用外参与深度将 i 的结果投影到 j 的视域中，再在重叠区域计算差异。

训练阶段的总体目标整合监督与自监督两类项：

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{vis}} + \lambda_2 \mathcal{L}_{\mathrm{amo}} + \lambda_3 \mathcal{L}_{\mathrm{bnd}} + \lambda_4 \mathcal{L}_{\mathrm{temp}} + \lambda_5 \mathcal{L}_{\mathrm{cons}} + \lambda_6 \mathcal{L}_{\mathrm{unc}} \tag{10}$$

其中：$\mathcal{L}_{\mathrm{vis}}$ 与 $\mathcal{L}_{\mathrm{amo}}$ 分别用于可见与 amodal 掩膜的监督，$\mathcal{L}_{\mathrm{bnd}}$ 为边界对齐项，$\mathcal{L}_{\mathrm{temp}}$ 与 $\mathcal{L}_{\mathrm{cons}}$ 如式（10）所示，$\mathcal{L}_{\mathrm{unc}}$ 为不确定度校准项，λ 为相应权重。

为提高对难样本区域的表达，采用异方差建模并最小化像素级负对数似然[14]：

$$\mathcal{L}_{\mathrm{unc}} = mean_u \left[\frac{1}{2} \log \sigma_u^2 + \frac{1}{2} \frac{(y_u - \mu_u)^2}{\sigma_u^2} \right] \tag{11}$$

式中，μ_u，σ_u^2 分别为位置 μ 的预测均值与方差，y_u 为监督信号。该项促使模型在信心不足处自动降低过度确定的输出。

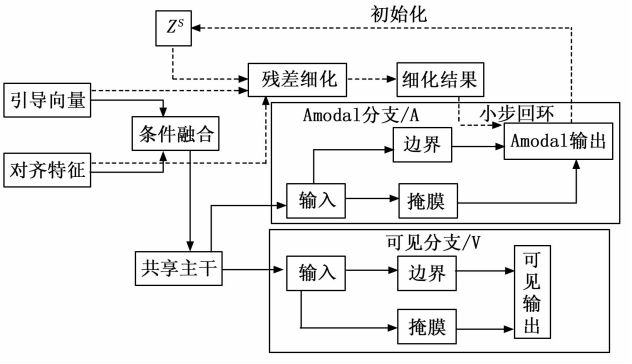

图 2　车端解码与细化回路

3　语义压缩与通信编排

本节围绕车端编码、路侧转发、云端更新三方协作给出压缩与调度的计算链。目标是在有限比特与时延约束下保持补全结果的质量与连贯性。

本文设计每个时刻的综合目标函数，用于平衡任务误差、比特开销与传输时延：

$$J_t = D_t + \beta R_t + \gamma \tau_t \tag{12}$$

式中，D_t 为任务相关的失真项，包含可见层与 amodal 的误差组合。R_t 为本时刻预计发送的比特量，τ_t 为从编码到可用的实际延迟，β，γ 为权重系数，随网络状

态与任务优先级自适应设定。任务失真由 3 个子项构成，以保证区域语义与边界细节同时得到关注：

$$D_t = \omega_1 \mathcal{L}_{\text{vis},t} + \omega_2 \mathcal{L}_{\text{amo},t} + \omega_3 \mathcal{L}_{\text{bnd},t} \tag{13}$$

式中，$\mathcal{L}_{\text{vis},t}$与$\mathcal{L}_{\text{amo},t}$分别度量可见掩膜与 amodal 掩膜的误差$\mathcal{L}_{\text{bnd},t}$为边界对齐误差 ω_1，ω_2，ω_3 为非负权重，用验证集标定。为了指导区域化码率分配，在车端生成重要度图，结合不确定度、风险与运动状态：

$$S_t(u) = \sigma(a_1 \sigma_u^{-2} + a_2 \rho_u + a_3 m_u) \tag{14}$$

式中，u 为像素或体素索引 σ_u^2 为该位置的方差预测，其倒数反映信心 ρ_u 为风险评分，如易发生交互或靠近道路边界。m_u 为运动强度或视差幅值，a_1，a_2，a_3 为可学习系数 $\sigma(\cdot)$ 为逻辑函数，将结果映射到 [0，1] 范围。

基于重要度图，本文在每个候选区域为潜表示分配目标码率，形成空间自适应比特预算：

$$r_k = r_{\min} + (r_{\max} - r_{\min}) \text{mean}_{u \in R_k} S_t(u) \tag{15}$$

式中，R_k 为第 k 个候选区域，$r_{\min}$分别为该区域允许的最小与最大码率，右端的均值表示区域平均重要度。为直观展示重要度如何驱动区域码率分配，本文在图 3 中以热力形式给出 S_t 并标注候选区域。

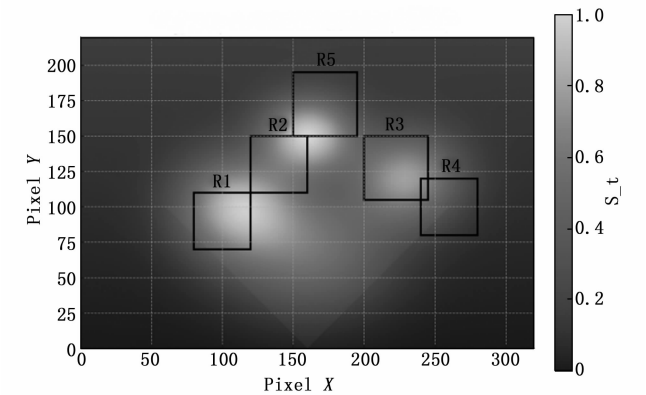

图 3 热力图

潜表示采用通道自适应的量化步长来匹配不同语义成分的敏感度：

$$\hat{z}_c = \text{round}\left(\frac{z_c}{\Delta_c}\right)\Delta_c \tag{16}$$

式中，z_c 为第 c 个通道的连续潜变量，Δ_c 为该通道的量化步长$\hat{z}$ 为量化后的结果，round（·）表示就近整数取整。码率由熵模型进行估计[15-16]，作为训练与调度时的比特成本：

$$\hat{R}_t = -\sum_c \sum_u \log_2 p_c(\hat{z}_{c,u} \mid \psi) \tag{17}$$

式中，$p_c(\cdot \mid \psi)$ 为带超先验参数 ψ 的条件概率模型，$\hat{z}_{c,u}$为通道 c 在位置 u 的量化值，求和覆盖所有通道与位置。时间维度上采用关键帧与差分帧策略，通过性价比指标决定是否发送关键帧：

$$I_t = 1\left[\frac{D_t^{\text{delta}} - D_t^{\text{full}}}{R_t^{\text{full}} - R_t^{\text{delta}}} > \theta\right] \tag{18}$$

式中，I_t 为时刻 t 的关键帧指示，D_t^{full}，R_t^{full} 分别为全量潜码时的失真与码率。D_t^{delta}，R_t^{delta} 为差分编码时的对应量，θ 阈值，由带宽与时延目标给出。差分帧只发送参考与当前之间的显著残差，在重要区域内更精细地保留细节[17-19]：

$$\in_t^{\text{sel}}(u) = 1[S_t(u) > \tau] \cdot [\phi(\widetilde{F}_t)(u) - \phi(\widetilde{F}_{t^*})(u)] \tag{19}$$

式中，$\phi(\cdot)$ 为从对齐特征提取的发送向量，t^* 为最近的关键帧时刻，τ 为重要度阈值 $\in_t^{\text{sel}}$ 为参加编码的稀疏残差。突发事件到来时按严重程度提升帧级预算，以保证高价值信息优先传输：

$$R_t^{\text{budget}} = \min[R_{\max}, \bar{R} + (1 + \sum_{e \in \varepsilon_t} \beta_e s_e)\delta R] \tag{20}$$

式中，$R_{\max}$为链路上限 $\bar{R}$ 为平滑后的基准速率，ε_t 为当前检测到的事件集合，s_e 为事件严重度，β_e 为事件类型权重，δR 为可增配的速率幅度。

为降低丢包引起的空洞，在分组层面按链路状态自适应地配置冗余比率：

$$P_t = \min[P_{\max}, \eta_1 \hat{p}_{\text{loss}} + \eta_2 \text{var}(R_{t-w:t})] \tag{21}$$

式中，P_t 为本时刻的校验比特比例，$\hat{p}_{\text{loss}}$为近期丢包率估计，var（$R_{t-w:t}$）为窗口长度 w 内的速率波动。η_1，η_2 为非负系数，$P_{\max}$为冗余上限。整体调度在帧级与区域级同时进行来满足预算与时延约束：

$$\min_{\{r_k, I_t, P_t\}} \sum_t \mathcal{J}_t$$
$$\text{s. t.} \sum_k r_k + P_t \leqslant R_t^{\text{budget}}, \tau_t \leqslant \tau_{\max} \tag{22}$$

式中，决策变量为区域码率 r_k，关键帧指示 I_t，冗余比例 P_t，约束包括帧级预算与最大容许延迟 $\tau_{\max}$。

本文设计路侧与云端的轻量更新规则，使用小规模参数并不增加传输负担：

$$\psi \leftarrow \psi + \eta_\psi \nabla_\psi [-\sum_{c,u} \log p_c(\hat{z}_{c,u} \mid \psi)] \tag{23}$$

$$\Delta_c \leftarrow \Delta_c + \eta_\Delta \nabla_{\Delta_c} \mathcal{J} J \tag{24}$$

式中，ψ 为熵模型超先验参数，Δ_c 为通道量化步长，二者以小步更新，由云端离线统计与路侧在线微调共同驱动，其中 η_ψ，η_Δ 为学习率。

通过以上各项，车端先以 S_t 完成区域化码率分配，再以通道自适应量化和熵模型估计 R_t，时间维度由 I_t 与$\in_t^{\text{sel}}$控制发送粒度，事件调度由 R_t^{budget} 动态扩展空间，分组层面的 P_t 提供对弱网络条件的缓冲，最终在预算与时延约束内优化$\mathcal{J}_t$。

本文提出的多模态语义补全与语义压缩方法的整体算法伪代码如下所示：

算法 1：边—云协同的多模态语义补全与语义压缩

输入：相机 X_t^{cam}，雷达 X_t^{rad}，惯导 X_t^{imu}，道路拓扑 T，交通轨迹 V

输出：可见语义 $\hat{V}_t$，Amodal 占据 $\hat{A}_t$，上行负载 $\hat{z}_t$

初始化：云端先验(ψ,Δ)，带宽预算 R^{budget}，时延阈值 $\tau_{\max}$

for 每一帧 t do

边缘特征融合：

$$F_t = \Gamma\left[f_{\text{cam}}(X_t^{\text{cam}}), f_{\text{rad}}(X_t^{\text{rad}}), f_{\text{imu}}(X_t^{\text{imu}})\right]$$

轻量时序对齐：$\widetilde{F}_t = W(F_t, H_t, U_t)$

重要度估计：对每个单元 u 计算 $S_t(u) = \sigma(a_1\sigma_u^{-2} + a_2\rho_u + a_3 m_u)$

关键帧判定：$I_t = 1\left[\dfrac{D_t^{\text{delta}} - D_t^{\text{full}}}{R_t^{\text{full}} - R_t^{\text{delta}}} > \theta\right]$

if $I_t = 1$ then

$z_t \leftarrow E_K F(\widetilde{F}_t)$

else

$\in_t^{\text{sel}} \leftarrow 1[S_t > \tau] \odot [\varphi(\widetilde{F}_t) - \varphi(\widetilde{F}_{t^-})]$

$z_t \leftarrow E_\Delta(\in_t^{\text{sel}})$

end if

量化与码率估计：$\hat{z}_c = \text{round}\left(\dfrac{z_c}{\Delta_c}\right)\Delta_c$

$\hat{R}_t = -\sum_c \sum_u \log_2\ p_c(\hat{z}_{c,u} \mid \psi)$

冗余与发送：$P_t \leftarrow \min[P_{\max}, \eta_1 \cdot \hat{p}_{\text{loss}} + \eta_2 \cdot \text{var}(R_{t-w:t})]$，上行发送 $\hat{z}_t$

路侧融合：$\alpha_i \propto \exp[\kappa_1 \cdot r_i + \kappa_2 \cdot IoU_i + \kappa 3 \cdot e^{-d_i}]$

$O^r s \leftarrow \sum_i \alpha_i \cdot O_i, M^r s \leftarrow \sum_i \alpha_i \cdot M_i$

生成引导：$D \leftarrow \text{norm}[K * (T, V)]$

$q_k \leftarrow AttnPool_{R_k}[(O^s, M^s, D)]$，下行发送$\{q_k\}$

车端解码：$[\hat{V}_t, \hat{A}_t] \leftarrow Dec[\widetilde{F}_t, (q_k)]$

小步细化：令 $z^0 \leftarrow \hat{A}_t$

for s=0…S−1 do

$z^{s+1} \leftarrow z^s + \eta \cdot R[z^s, \widetilde{F}_t, (q_k)]$

end for

if 训练阶段 then

最小化：$\mathcal{L} = \lambda_1\ \mathcal{L}_{\text{vis}} + \lambda_2\ \mathcal{L}_{\text{amo}} + \lambda_3\ \mathcal{L}_{\text{bnd}} + \lambda_4 \mathcal{L}_{\text{temp}} + \lambda_5\ \mathcal{L}_{\text{cons}} + \lambda_6\ \mathcal{L}_{\text{unc}}$

end if

周期性云端更新：

$\psi \leftarrow \psi + \eta_\psi\ \nabla_\psi[-\sum_{c,u} \log p_c(\hat{z}_{c,u} \mid \psi)]$

$\Delta_c \leftarrow \Delta_c + \eta_\Delta \nabla_{\Delta_c}\ \mathcal{L}\ J$

end for

4　实验仿真

4.1　实验环境及设置

实验操作系统为 Ubuntu22.04，深度学习框架为 PyTorch2.3，CUDA12.1 与 cuDNN，Python3.10。硬件配置为四块 RTX 4090 24 GB，CPU 为 64 线程级别，内存不低于 256 GB。车端推理与通信开销在 Jetson AGX Orin 64 GB 上实现，路侧服务以容器方式部署在一台带 24 核 CPU 与 128 GB 内存的边缘服务器上，云端训练与蒸馏在独立 GPU 集群上运行。网络条件通过 Linux tc netem 仿真，上下行带宽按 50、100、200 kB 每帧三档设置，单向时延在 20～100 ms 范围内采样，丢包率在 0～10%范围内分段取值。

数据选择覆盖单车语义占据补全与车路协同两条主线，语义占据部分使用 nuScenes 的语义占据扩展版本与 SemanticKITTI 的体素标注版本。训练与验证按官方划分，输入为六路相机图像，图像分辨率下采样到 640×352。协同实验选用 DAIR-V2X-C 与 OPV2V 两个公开数据集，分别覆盖真实道路的车路协同与仿真条件下的多车协同，采用标准训练与验证划分，时序窗口为 3 帧，路侧视角与车端视角通过提供的外参进行统一标定。nuScenes 以城市道路日常交通为主，包含多种光照与天气条件的样本，但对强雨雾等极端气象，以及道路施工、交通事故、临时封闭等特殊交通事件的覆盖并不集中，相关情形在数据中的占比相对有限。SemanticKITTI 更侧重车载序列与体素标注的连续性验证，对典型道路结构与常见交通参与者有较好支持，但对车路协同要素与复杂事件的覆盖相对不足。DAIR V2X C 能反映真实道路车端与路侧的协同观测特征，但受采集条件限制，其环境类型与事件多样性仍有限，OPV2V 便于在可控条件下评估协同机制，但与真实世界在传感器噪声与突发事件分布上存在差异。因此，本文实验结论主要适用于常见城市道路与常见光照天气条件下的遮挡补全与协同传输场景，但该组合能够同时覆盖单车占据基准能力、真实车路协同特征与可控的多车协同评测，有利于在一致设置下从补全质量与通信代价两方面对方法进行更全面的验证。

对比方法包含两类代表工作，单车语义占据补全采用 SurroundOcc[7] 作为强基线，通信侧协同采用 CodeFilling[20] 式的码本填充与区域选择方法作为高效协同基线。

测评指标包括：

1）语义占据 $mIoU_{\text{occ}}$：在固定体素网格与评测范围下计算全类平均交并比，用来衡量补全后的三维语义准确度。

2）平均上行开销（kB/帧）：对整段测试数据统计上传比特总量并除以帧数，直接反映通信成本，便于与通信友好的基线对比。

3）端到端时延 P95（ms）：从车端采集得到可用补全结果的时间，取 95 分位数，体现在真实链路下的响应速度。

4.2　实验分析

在 nuScenes-Occupancy 验证集上，以 0～100 m、每 5 m 为一档共 20 段评测 $mIoU_{\text{occ}}$。实验过程中的输入为多相机视角，各方法采用统一的体素分辨率与评测范围。训练与预处理一致，仅网络结构按各自论文实现。$mIoU_{\text{occ}}$ 指标的实验结果如图 4 所示，图中越靠上表示补全越好。

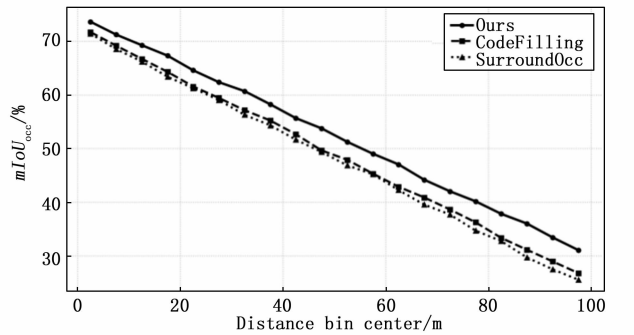

图 4 $mIoU_{occ}$ 指标实验结果

本文方法在全距离段均略高于两条基线，20 段相对于 SurroundOcc 平均提升约 4.28%，相对于 CodeFilling 约提升 3.46%。在 60～100 m 的远距段，曲线间距扩大，说明在遮挡与视差增大的路段仍能保持更好的占据一致性。需要指出的是，图中 $mIoU_{occ}$ 在不同距离段存在一定波动，其来源主要包括三类因素：1）远距段本身可用观测更稀疏，体素占据对少量误差更敏感，导致不同方法在 5 m 分箱统计上更容易出现起伏；2）nuScenes 的占据标注在远距与遮挡区域的可见证据较弱，边界体素与薄结构更容易出现标注噪声或不确定性，从而放大分箱后的统计波动；3）多相机融合在远距段受视差与遮挡变化影响更明显，个别样本的遮挡格局会使某些分箱段的难度显著升高，进一步带来曲线的局部起伏。相对而言，本文方法在远距段仍保持更平滑的趋势，表明路侧融合与引导回传对稀疏证据下的占据推断具有一定缓冲作用。近距 0～20 m 三者差距较小，主要受传感器分辨率与标注饱和影响。总体看，本文方法在不改变采集配置与评测口径的前提下带来稳定的小幅增益，更适合作为边路协同下的基础占据补全方案。

接着本文在 nuScenes-Occupancy 验证集上，按场景复杂度分位点从 P5 到 P100 划分 20 个档位，统计平均上行开销（kB/帧）。输入与训练设置统一，仅通信侧与编码方式不同。复杂度越高，目标与遮挡越多，理论上需要更高的上传量，实验结果如图 5 所示。

结果显示本文方法在全档位都低于 CodeFilling，并显著低于 SurroundOcc，相比 CodeFilling 平均减少约 12.5 kB/帧，相比 SurroundOcc 平均减少约 110.3 kB/帧。在高复杂度区间 P80～P100，差距进一步扩大且本文方法仍位于 100～120 kB/帧的预算线附近，说明区域选择与差分策略在拥挤场景依然有效。对于不同复杂度档位下开销存在的差异与阴影带宽度变化，其原因同样主要包含下列因素：复杂度分位点是按样本统计划分，高复杂度档位内部仍包含遮挡格局差异较大的片段，导致同一档位中需要上传的有效区域数与残差幅值

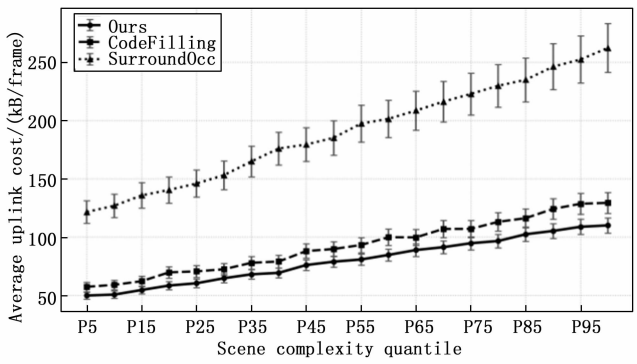

图 5 平均上行链路成本与场景复杂度对比

不完全一致。通信仿真中带宽采样与丢包注入具有随机性，即使对同一段数据，重传与冗余触发次数也会在不同随机种子下产生差别，从而反映为均值附近的波动。模型在不同场景中对不确定度与重要度的响应存在差异，例如在密集交互与强遮挡片段中重要区域更集中，区域化码率分配会触发更明显的预算倾斜，使开销随复杂度上升呈现非线性增长。综合来看，这些误差主要来源于场景内部差异、标注不确定性以及通信仿真的随机扰动，本文在 3 种随机种子下统计均值与标准差，能够较客观地反映方法在不同难度档位下的开销水平与波动范围。

本文以端到端 P95（ms）衡量从车端采集到获得可用补全结果的时延上界，固定基础网络条件为基线往返时延 50 ms、丢包 5%、视频帧周期约 33 ms，在 50～220 kB/帧的带宽预算范围内评测 3 种方法。P95 由 4 部分构成：基线网络排队与调度、车端计算开销、序列化与上传时间以及由丢包诱发的抖动项。3 种方法的上传量分别取中等复杂度场景的代表值，用统一公式换算为对应预算下的上传时间，再与其计算开销叠加得到 P95。实验结果如图 6 所示，为反映稳定性，图中给出了按 6%比例估计的一倍标准差阴影带，并用虚线标注 100、120、150 ms 共 3 条常见服务阈值。

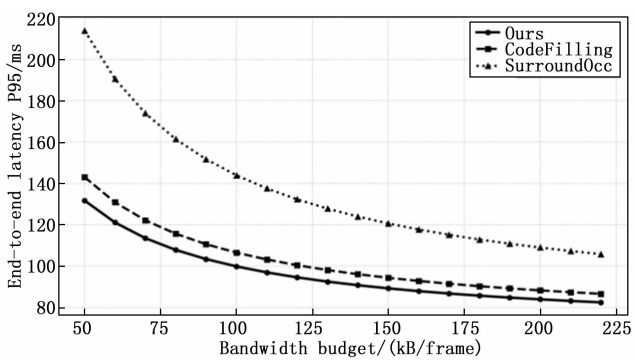

图 6 P95 与带宽预算的关系
（丢包率＝5%，基准往返时间＝50 ms）

从图 6 中的结果可以看出，P95 随带宽预算提升而单调下降。全区间平均差距显示，本文方法相对 CodeFilling 约低 6.2 ms，相对 SurroundOcc 约低 40.0 ms，优势主要来自更小的每帧上传量与较低的序列化占用。在 100～150 kB/帧的常用预算段，本文方法基本位于 100 ms 附近并随预算平滑下降，能够满足多数车路协作场景对响应速度的要求。而在 50～80 kB/帧的低预算段，三者差距进一步拉开，说明消息粒度与差分策略对时延上界影响显著。阴影带宽度随预算收敛，表明当预算不再成为瓶颈时，链路抖动与计算开销的波动主导 P95 的分散度。综上，在统一条件下，本文方法以更小的上传量换取更短的高分位时延，并在低预算与中高复杂度场景中保持更稳的响应表现。

本文固定车端上行消息形态与网络仿真条件不变，仅改变同域协同车辆数 N，取 $N\in\{1, 2, 3, 4, 6, 8, 10, 12\}$。路侧端部署与前述一致，采用边缘服务器容器化服务，统计路侧在线处理阶段的时延 P95，在线处理包括消息接入与缓存，时空对齐，多车一致融合，引导生成与下发队列。3 种方法在相同输入与评测流程下运行，结果如图 7 所示。

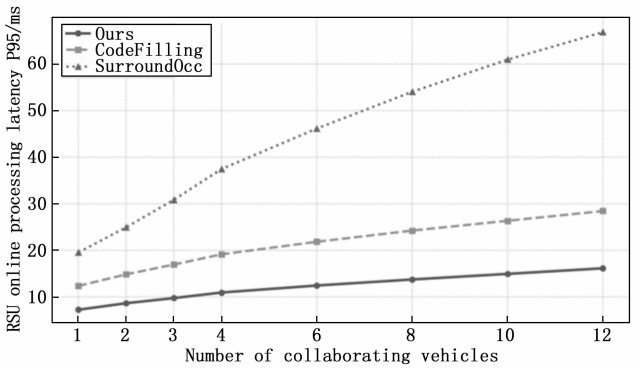

图 7　路侧在线处理时延 P95 随协同车辆数变化

从图 7 可以看出，3 种方法的路侧在线时延均随协同车辆数增加而上升，但本文方法增长更缓，在 12 车协同时 P95 为 16.1 ms，低于 CodeFilling 的 28.4 ms 与 SurroundOcc 的 66.8 ms。该结果说明本文在路侧以语义级潜表示进行融合并生成紧凑引导回传，避免了对大体量中间特征或重型解码过程的依赖，使多车规模增大时仍能保持较低的在线计算开销，更有利于在边缘服务器算力受限条件下扩展协同车辆数。

5　结束语

本文提出一种面向车、路、云协同的车辆场景多模态语义补全与语义压缩方法。车端以轻量主干融合相机、雷达与惯导信息，在对齐后的特征上采用可见与 amodal 的双头解码并配合小步迭代细化，路侧基于多车潜表示生成占据与遮挡提示以及区域引导向量，下行以少量条件向量提升被遮挡区域的外形与语义质量。通信层面，结合重要度与不确定度进行区域化码率分配，配合关键帧与差分策略以及事件优先调度，在不改变感知配置的前提下，显著降低平均上行开销并缩短高分位时延。实验结果说明，在相同的感知与训练条件下，本文方法能以更小的上传量带来更完整的补全输出，同时保持更快的响应速度与更好的序列连贯性。

未来工作将面向跨城迁移与道路形态快速变化探索更高效的在线自适应更新策略，并加强对极端天气与夜间弱纹理场景的适配能力。在更严格的带宽与高时延网络下，研究多粒度索引化表示与更精细的事件触发机制，扩展到更低预算的量产级通信环境，并进一步验证方法的可扩展性与工程可用价值。

参考文献：

[1] 晁　琪，赵燕东，刘圣波. 多模态融合的三维语义分割算法研究 [J]. 红外与激光工程，2024，53 (5)：253－267.

[2] 赵树恩，袁　亮，赵东宇. 基于要素信息补全的自动驾驶复杂场景语义理解 [J]. 仪器仪表学报，2025，46 (4)：295－305.

[3] 仇志江，张　林，姚　垚，等. 三维点云场景语义分割研究进展 [J]. 中国图象图形学报，2025，30 (7)：2325－2342.

[4] HU Y，FANG S，LEI Z，et al. Where2comm：communication-efficient collaborative perception via spatial confidence maps [J]. Advances in Neural Information Processing Systems，2022，35：4874－4886.

[5] 邢焕来，况田泽宇，徐乐西，等. 基于对比学习和扩散模型的图像语义通信方法 [J]. 计算机学报，2025，48 (5)：1151－1167.

[6] 申　滨，李　旋，赖雪冰，等. 基于 Swin Transformer 的宽带无线图传语义联合编解码方法 [J]. 电子与信息学报，2025，47 (8)：2665－2674.

[7] WEI Y，ZHAO L，ZHENG W，et al. Surroundocc：multi-camera 3d occupancy prediction for autonomous driving [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision，2023：21729－21740.

[8] WANG X，ZHU Z，XU W，et al. Openoccupancy：A large scale benchmark for surrounding semantic occupancy perception [C] //Proceedings of the IEEE/CVF International Conference on Computer Vision，2023：17850－17859.

[9] TIAN X，JIANG T，YUN L，et al. Occ3d：A large-scale 3d occupancy prediction benchmark for autonomous driving [J]. Advances in Neural Information Processing Systems，2023，36：64318－64330.

（下转第 243 页）