语义赋能的空间感知编解码技术研究

陈鸣锴，刘沁妍
（南京邮电大学通信与信息工程学院，江苏 南京 210003）

摘　要：针对特种机器人在大规模场景实时重建时所面临的计算复杂性、网络传输效率和高保真度语义重建等问题，提出了一种语义赋能的空间感知编解码技术。首先，采用BEV特征映射将点云转换为结构化张量；其次，基于语义信息生成二值语义掩码实现关键区域定位和数据稀疏化；再次，构建层次化熵编码框架，通过可学习量化和超先验概率模型实现高效压缩；最后，解码采用具身智能场景语义对齐重建确保几何与语义高保真恢复。实验结果表明，所提方法在保持良好重建质量的同时实现了高效的数据传输，充分验证了该方法的有效性和鲁棒性。

关键词：具身智能；空间感知；语义赋能；熵编码；场景重建
中图分类号：TN92
文献标志码：A
DOI：10.11959/j.issn.1000–436x.2026069

Research on semantic-enhanced spatial perception codec technology

Chen Mingkai, Liu Qinyan
School of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: To address the challenges of computational complexity, network transmission efficiency, and high-fidelity semantic reconstruction faced by specialized robots in large-scale scene real-time reconstruction, a semantic-enhanced spatial perception codec technology was proposed. Firstly, BEV feature mapping was adopted to convert point clouds into structured tensors. Secondly, binary semantic masks were generated based on semantic information to achieve key region localization and data sparsification. Thirdly, a hierarchical entropy coding framework was constructed to realize efficient compression through learnable quantization and hyper-prior probability models. Finally, decoding adopts embodied intelligent scene semantic alignment reconstruction to ensure high-fidelity restoration of geometry and semantics. Experimental results demonstrate that the proposed method achieves efficient data transmission while maintaining good reconstruction quality, which fully validates the effectiveness and robustness of the method.

Keywords: embodied intelligence, spatial perception, semantic-enhanced, entropy coding, scene reconstruction

0　引言

随着人工智能、机器学习和传感器技术的飞速发展，搜救机器人、管道检测机器人、核环境机器人等特种机器人在危险环境勘察、应急救援和未知区域探索中应用日益广泛[1]。特种机器人通常依赖激光雷达、RGB-D相机等传感器获取环境的三维

收稿日期：2025–11–26；**修回日期：**2026–03–14
通信作者：陈鸣锴，mkchen@njupt.edu.cn
基金项目：国家自然科学基金资助项目（No.62001246）；江苏省重点研发计划基金资助项目（No.BE2023035）；江苏省通信与网络技术工程研究中心开放课题基金资助项目（No.JERCCN202301）
Foundation Items: The National Natural Science Foundation of China (No.62001246), The Key Research and Development Program of Jiangsu Province (No.BE2023035), Open Research Fund of Jiangsu Engineering Research Center of Communication and Network Technology (No.JERCCN202301)

空间信息，构建精确的场景模型[2]。传统空间感知方法主要基于几何特征提取和点云处理技术[3]，在静态环境中具有较高的重建精度，但在复杂动态环境下往往忽略语义信息，难以支撑任务导向分析。同时，感知算法[4]常采用统一处理策略，难以随任务自适应调整，在资源受限条件下难以兼顾精度与效率。

在灾区、地下管廊等极端环境中，通信带宽受限、时延增大使传统“先感知、后传输、再处理”的串行架构面临显著瓶颈[5]，大量点云与图像数据难以及时回传，制约了远端态势感知与决策效率。尽管现有压缩传输方法可在一定程度上缓解带宽压力，但多采用通用压缩策略，未显式利用语义与任务相关性，易在高压缩率下造成关键语义与细节信息丢失[6]。

具身智能作为特种机器人智能化的重要方向，强调机器人基于环境感知结果进行自主决策与行为规划，这不仅要求系统感知几何结构，还需理解环境中物体的语义属性与空间关系[7]。由于不同任务对环境信息的关注重点存在显著差异[8]，现有以几何为主的空间感知与数据处理方法缺乏任务自适应能力，难以实现面向具体任务的高效信息筛选与传输。

因此，在实际应用中，各类特种机器人通常面临 3 个关键挑战。

1) 如何应对极端或通信带宽受限的环境。在救灾、地下管廊、核环境等场景中，通信带宽极其有限且时延高，需要一种机制来确保任务关键语义信息能被传递，以免因数据不全或时延导致指挥误判。

2) 如何减轻前端计算与传输压力。在嵌入式或能源受限平台上，算力与存储资源都有限，因此必须减少点云原始数据的传输与处理量，否则长期运行难以维系。

3) 如何提高终端决策效率与可靠性。在远端决策节点，若收到的是冗余原始数据，则需要额外的解析和筛选，既耗时又可能出错，因此传送符合任务要求的子集便成了保障终端决策效率与可靠性的关键。

针对上述挑战，本文提出了一种以感知–语义理解–编码–重建为一体的空间信息编解码方案，旨在通过具身智能提供高效的空间感知与传输支持。本文主要贡献如下。

1) 提出了语义驱动的主动编解码方法，将任务相关性引入量化、掩码与熵建模中，不仅突破了传统“被动压缩”的局限，还能在资源约束下优先保证任务关键信息的保真度。

2) 构建了端到端的语义一致性编解码框架，在解码重建过程中引入“优先级权重→点云语义映射”机制。该机制利用解码端为适应后端任务分配的优先级权重，将任务优先级直接注入重建点云的语义标注中，从而做到“编码端按任务优先级分配资源→解码端按优先级直接标注语义”，保证任务相关语义在低码率下得到优先恢复。

3) 提出了面向具身智能的动态资源调度机制，将网络状态、任务优先级与实时性约束纳入编码决策，使系统能够在复杂异构环境中保持高效、鲁棒的空间感知与重建能力。

1　相关工作

空间场景感知与理解、空间数据压缩与传输以及具身智能决策的相关研究各有侧重，共同构成了空间环境感知与交互系统的核心支撑。这三大方向的深度汇聚正驱动远程场景感知与交互技术的快速演进。本节将系统回顾其研究现状、探讨应用前景，并指出各自面临的关键挑战，为后续提出语义赋能的空间感知编解码技术奠定坚实基础。

1.1　空间场景感知与理解

近年来，随着深度学习、计算机视觉、传感器技术与图形渲染等多学科的融合，空间场景感知与理解取得了飞速发展。Azuma 等[9]提出的 ScanQA 框架首次在 ScanNet 彩色深度点云与文本问答对上联合训练，通过结合 3D 对象提议与问句嵌入，开启了空间数据与语言推理的探索。Leutenegger[10] 进一步引入闭环检测与惯性测量单元辅助，实现场景图的持续校正，展现实时流式感知趋势。Bultmann 等[11]设计的分布式边缘传感器网络将实时语义标注与几何结构深度融合，支持对大规模工业厂区和室外公共空间的低时延监控与交互。

随着大型预训练模型与多模态学习的普及，空间场景感知与理解正进一步向“语言–视觉–几何”协同的方向演进。Chen 等[12]利用大规模语言模型增强室内物体与功能区分类，将常识推理融入场景建模。Jia 等[13]和 Li 等[14]在 OccupancyDETR 和 OccScene 中统一物体检测、占用预测与三维生成，推动感知与生成的双向理解。最新如 Qi 等[15]的

GPT4Scene 与 Jiang 等[16]的社交视觉–语言–动作联合模型，将视觉语言模型扩展至视频、语音与三维交互，显著提升了人机协作与空间指令执行能力。未来，多粒度语义表示、时空一致性与大模型推理的深度融合，必将推动空间场景感知与理解迈向更高智能化阶段。

1.2 空间数据压缩与传输

近年来，随着遥感、激光雷达、RGB-D 相机等多源空间数据量爆发式增长，高效存储与传输在时效与质量间的平衡成为研究热点。Fu 等[17]将残差网络嵌入传统变换流程，显著提升了高频纹理保真。同时，针对图像自相似与细节保持，分形编码与小波变换仍具优势。He 等[18]结合分形编码与小波多尺度分解，在保持细节的同时加快编码。为兼顾视觉感知，Yuan 等[19]引入人类视觉系统模型，对低比特率水平下的图像压缩进行感知优化，确保带宽受限下的视觉质量。

在传输端，边缘计算与物联网催生轻量、低时延的一体化压缩方案。Yang 等[20]在电力相机数据中采用预测量化与比特打包三阶段算法，大幅降低传输负载。Lu 等[21]提出了基于 Transformer 的空频混合解码，适用于高帧率、低时延视频压缩。面向高光谱和天文数据的无损压缩，Khoshkhahtinat 等[22]分别利用递归神经网络与时空冗余学习定制专业场景。未来，端到端学习、动态自适应与跨模态协同将进一步推动海量空间数据的智能传输与实时应用。

1.3 具身智能决策

近年来，具身智能决策在虚拟与物理环境中均取得显著进展。在沉浸式场景下，Yasuda 等[23]提出的多模态智能调解框架通过融合真实与合成运动传感器数据，将事件触发时的视觉信息分发至 YOLO 集群，实现了去中心化的实时决策。针对动态突发环境，Zhou 等[24]设计的“HAZARD”基准测试模拟多变干扰，评估智能体在不确定条件下的应急响应能力。在工业 4.0 和制造领域，具身智能决策同样大放异彩。Gao 等[25]基于网络物理生产系统与非线性核映射，构建了自适应健康诊断与维护决策方法，实现了机器人伺服系统的在线故障预测。Coito 等[26]通过融合传感器与业务数据，搭建了实时预测性维护与调度系统。

深度强化学习与多模态大模型的融入，为具身智能决策带来新机遇。Ahn 等[27]提出的 SayCan 框架展示了大语言模型与领域知识融合的具身智能决策模式。展望未来，随着时空记忆、语义推理和大模型推理能力的深度融合，具身智能决策将进一步提升自主性、适应性与可解释性，助力智慧城市、数字孪生和机器人导航等领域的智能化革新。

1.4 出发点与挑战

本文面向具身智能的高效空间感知与传输设计专属的编解码技术，但在设计过程中面临多重挑战。一是大规模空间环境重建不仅需要高精度几何保持，而且还要精细建模语义层级。在计算与通信受限条件下，如何实现“语义优先”的重建是首要难点。二是在灾害救援、地下管廊等极端环境中，网络极不稳定，带宽受限且时延高。如何在动态条件下实时分配资源与调度编码，确保关键语义可靠传递是核心问题。三是现有点云压缩多采用统一策略，忽视任务对区域和语义的差异化需求。如何在编码端保留任务优先级，并在解码端实现语义对齐，是由“压缩优化”迈向“任务驱动”的关键。概括而言，本文需应对三大挑战：计算与通信约束下的语义优先级重建；极端环境中的网络瓶颈与实时传输难题；任务自适应的语义一致性与资源调度机制缺乏。

2 语义赋能编解码技术总体框架

面对大规模空间场景的实时感知与远程传输需求，传统的点云压缩方法往往忽略了数据中蕴含的语义信息，导致在带宽受限环境下关键信息丢失。

本文提出的具身智能赋能的空间感知编解码技术首先对空间点云根据任务相关性进行语义驱动的特征筛选，通过生成语义掩码与优先级权重，实现对关键区域的显式提取与稀疏化表示。接着，结合层次化熵编码框架与可学习量化机制，在考虑语义权重与网络状态的条件下动态调整压缩策略，生成兼顾效率与语义保真的紧凑比特流。在解码端，系统引入语义一致性重建机制，将任务优先级直接映射至点云的语义标注，实现几何结构与语义信息的协同恢复，并支持具身智能场景下的任务自适应解析。整个技术流程实现了从原始点云数据的语义筛选到高效编码传输，再到目标场景重建的完整闭环处理，总体框架结构如图 1 所示，主要包括三大模块：空间场景感知与语义理解、任务语义权重动态

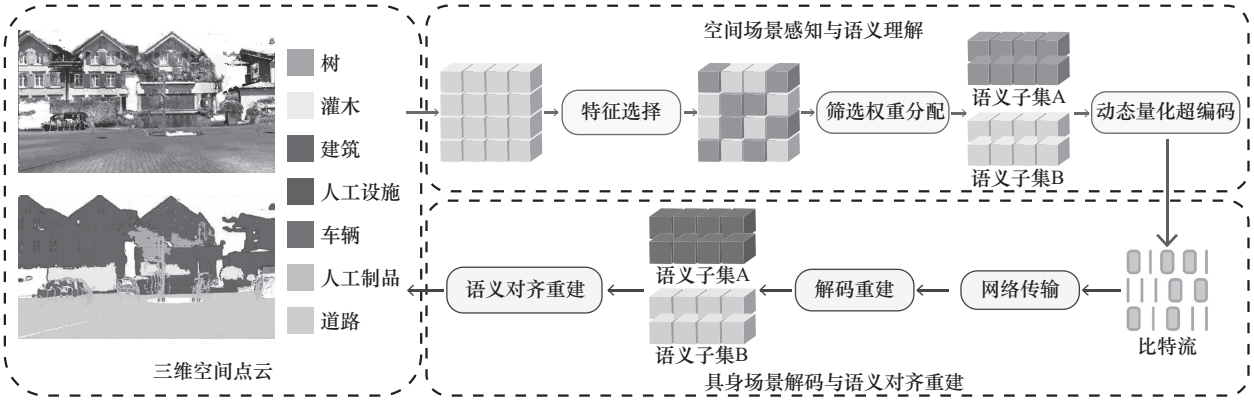

图1　总体框架结构

量化超编码和具身场景解码与语义对齐重建。后文将对这3个模块进行详细介绍。

2.1　空间场景感知与语义理解

在空间场景感知与理解的过程中，原始点云数据通常呈现出结构复杂、数据量大等特点，需要通过有效的特征提取和表示方法来捕获其中的几何结构和语义信息。为了实现这一目标，本文采用特征映射方法，其核心在于将原始稠密的三维空间点云转换为结构化张量，既保留了点云中几何与语义信息，又为后续的网络处理和压缩编码提供了有效的表示。输入是一组高精度重建的三维空间点云，点的数量在数十万级别，并且每个点不仅包含简单的(x,y,z)坐标，还携带协方差描述、颜色分量和语义标签。具体而言，将点云集合记为

$$P=\{\boldsymbol{p}_i|i=1,2,\cdots,N\} \tag{1}$$

其中，第i个点$\boldsymbol{p}_i$包含空间坐标(x_i,y_i,z_i)、协方差矩阵$\boldsymbol{\Sigma}_i$、颜色分量(r_i,g_i,b_i)和语义标签向量$\boldsymbol{s}_i$，其中协方差矩阵可简化为局部尺度$(\sigma_{x,i}^2,\sigma_{y,i}^2,\sigma_{z,i}^2)$或椭球参数。多维属性使每个点既具有高精度空间信息，也可反映局部不确定性、外观特征与语义归属。

为了将稠密的点集合映射为规则张量，首先对三维点云进行空间划分，大致流程如图2所示。基于PointPillars[28]的算法，首先在水平方向上进行二维网格划分，获得$H\times W$个垂直柱状单元。记网格步长为Δx和Δy，平面覆盖范围为$[X_{\min},X_{\max})\times[Y_{\min},Y_{\max})$。任意点$\boldsymbol{p}_k=(x_k,y_k,z_k)$被映射到垂直柱状单元$(h,w)$。

$$h=\left\lfloor\frac{x_k-X_{\min}}{\Delta x}\right\rfloor, w=\left\lfloor\frac{y_k-Y_{\min}}{\Delta y}\right\rfloor \tag{2}$$

其中，$\lfloor\cdot\rfloor$表示向下取整，$h\in\{0,\cdots,H-1\}$，$w\in\{0,\cdots,W-1\}$，$H=\left\lceil\frac{X_{\max}-X_{\min}}{\Delta x}\right\rceil$，$W=\left\lceil\frac{Y_{\max}-Y_{\min}}{\Delta y}\right\rceil$。这样整个场景在水平方向上被划分为$H$行$W$列的柱状网格，共$H\times W$个垂直柱状单元，而每个单元中汇聚了$z$方向上落在对应$(x,y)$范围内的所有点。

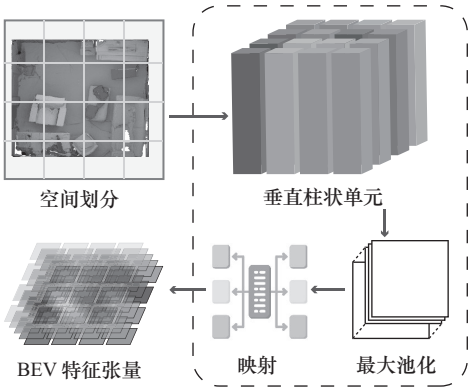

图2　特征映射流程

对于每个垂直柱状单元内的点集，本文在垂直方向上进一步做离散划分以保留竖直信息。设柱高为$[Z_{\min},Z_{\max})$，网格高为Δz，则任一点$\boldsymbol{p}_k$所属的竖直格索引为

$$c=\left\lfloor\frac{z_k-Z_{\min}}{\Delta z}\right\rfloor, c\in\{1,\cdots,C\} \tag{3}$$

其中，$C=\left\lceil\frac{Z_{\max}-Z_{\min}}{\Delta z}\right\rceil$。设柱状单元$(h,w)$内的点集为$P_{h,w,c}=\{\boldsymbol{p}_k\}$，对于每个点$\boldsymbol{p}_k$，首先构建原始输入向量，表示为

$$\begin{aligned}\boldsymbol{u}_k=[&x_k,y_k,z_k,r_k,g_k,b_k,\boldsymbol{s}_k,\\&x_k-x_{h,w,c},y_k-y_{h,w,c},z_k-z_{h,w,c}]\end{aligned} \tag{4}$$

其中，$(x_{h,w,c},y_{h,w,c},z_{h,w,c})$为所在柱的中心，随后利用

多层感知器将 u_k 映射到 C 维的点级特征向量 $\boldsymbol{\Phi}(\boldsymbol{p}_k)\in\mathbb{R}^C$，该向量包含绝对位置、相对于柱中心的偏移、颜色和语义信息等。然后对柱内所有 $\boldsymbol{\Phi}(\boldsymbol{p}_k)$做最大池化，得到定长柱特征向量。

$$\boldsymbol{F}_{h,w,c}=\operatorname{maxpool}\left\{\boldsymbol{\Phi}(\boldsymbol{p}_k)\,\middle|\,\boldsymbol{p}_k\in P_{h,w,c}\right\}\in\mathbb{R}^C \tag{5}$$

至此，便将不定数目的点集压缩为一个固定维度的向量，既保留了该柱内所有点的几何与语义聚合信息，也为后续卷积操作提供了统一格式。

随后，将所有柱状特征按照其在水平平面上的网格位置，摆放到一个形状为 $H\times W\times C$ 的张量中，成为鸟瞰图（bird's-eye-view, BEV）特征图。

$$\begin{aligned}&\boldsymbol{F}=[\boldsymbol{F}_{h,w,c}],\\&0\leqslant h<H-1,0\leqslant w<W-1,1\leqslant c\leqslant C\end{aligned} \tag{6}$$

其中，$\boldsymbol{F}_{h,w,c}$ 是前面得到的固定维数为 C 的柱级向量。BEV 特征图将三维点云“投影”到二维平面，但保留了每个网格柱状单元在 z 方向上的信息，同时可直接使用二维卷积网络进行后续处理。

至此，特征映射完成了从大规模点云到 $H\times W\times C$ 张量的转化，既确保了结构化与计算效率，也为下游的语义掩码生成、空间感知特征筛选与压缩编码奠定了坚实基础。

在完成三维点云的 BEV 特征图构建后，不同于传统方法的“先压缩再分析”，本文采用“先理解再筛选”的策略。根据预定义的任务需求，动态生成二值语义掩码，以便将任务关注的区域从无序点云中“提取”出来，供后续的空间感知特征筛选与压缩编码使用。

语义掩码生成的关键在于，利用与特征映射步骤相同的柱状单元分区方式，将点云上的语义标签投影到 BEV 特征图上，再以二值化形式表示每个柱状单元内部是否存在“任务关注”的语义类别。具体而言，对于每个柱内的所有点集合 $\{\boldsymbol{p}_k\}$，提取其语义标签向量 $\boldsymbol{s}_k$。定义二值化语义掩码图为 $\boldsymbol{S}\in\{0,1\}^{H\times W\times C}$，其中 $S_{h,w,c}\in\{0,1\}$ 为标量。语义掩码生成规则为：若点集 $P_{h,w,c}$ 中至少存在一个点 $\boldsymbol{p}_k$ 的语义标签 $\operatorname{sem}(\boldsymbol{p}_k)$ 属于预先通过人工定义得到的任务关注语义集合 $\mathcal{C}_{\text{task}}$，则令 $S_{h,w,c}=1$，否则令 $S_{h,w,c}=0$，数学表述为

$$S_{h,w,c}=\begin{cases}1, & \exists\boldsymbol{p}_k\in P_{h,w,c}\ \text{s.t.}\operatorname{sem}(\boldsymbol{p}_k)\in\mathcal{C}_{\text{task}}\\0, & \text{其他}\end{cases} \tag{7}$$

其基于集合包含关系的数学表述具有天然的通用性，不受特定语义类别限制。这种处理确保了三维体素化索引与后续处理步骤保持一致，并避免了再次执行三维体素化带来的计算负担。

执行完上述聚合之后得到的二值化语义掩码图 $\boldsymbol{S}\in\{0,1\}^{H\times W\times C}$ 就是一个仅包含 0 或 1 的多通道二维特征图，其中 H 和 W 分别对应 BEV 特征图的行和列维度，C 为该特征图的通道数。若 $S_{h,w,c}=1$，则说明该网格单元内至少存在一个语义属于集合 $\mathcal{C}_{\text{task}}$ 的点；若 $S_{h,w,c}=0$，则表示该网格单元内所有点不在任务关注类别中。最终，这张二值化语义掩码图在空间上的分布形态与特征映射得到的 BEV 特征图一一对应，在图像上可直观地看到目标物体或关注区域在特种机器人视角下的俯视投影。

为了只保留与下游任务密切相关的空间区域特征，不包含目标语义的区域将会被全部置零，进而实现对特征图的稀疏化处理，从而在保证感知精度的同时显著减少待传输和解码的特征数据量，如图 3 所示。

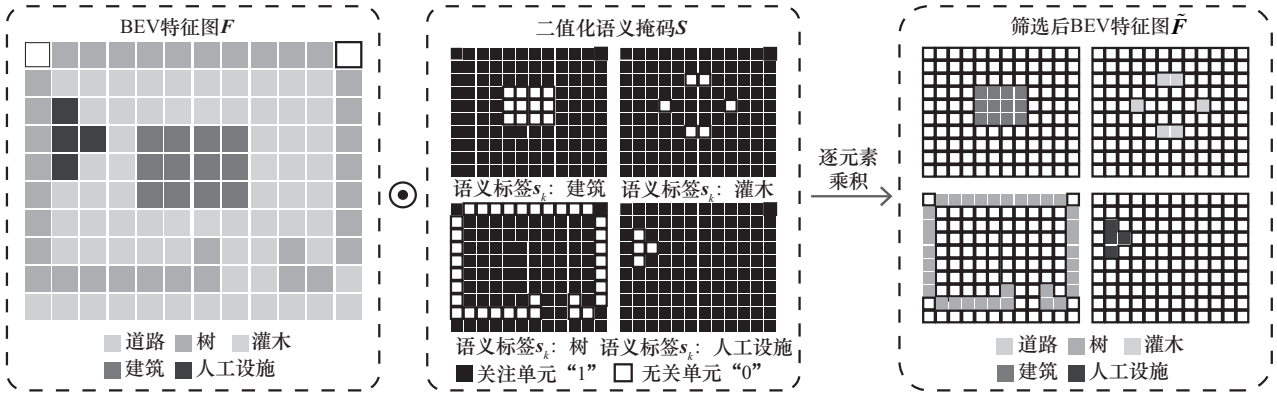

图 3 空间感知特征选择结果对比

在生成任务相关的语义掩码后，为了更精细地控制编码资源，需根据多维度因素动态计算每个任务语义区域的编码权重。首先在$S_{h,w,c}=1$的位置计算各点的语义权重。

$$m_{h,w,c}=\alpha I_{h,w,c}+\beta N(t)+\gamma T_{h,w,c}+\delta C(t) \tag{8}$$

其中，$I_{h,w,c}$为语义重要性因子，可根据任务相关类别的优先级计算，定义为

$$I_{h,w,c}=\frac{\sum_{\mathrm{sem}(\boldsymbol{p}_k)\in\mathcal{C}_{\mathrm{task}}}\mathbb{I}\left(\boldsymbol{s}_k\in P_{h,w,c}\right)\mathrm{priority}(\boldsymbol{s}_k)}{\left|\mathcal{C}_{\mathrm{task}}\right|} \tag{9}$$

其中，$\mathrm{priority}(\boldsymbol{s}_k)$为该语义的任务优先级。网络状态因子反映了可用与需求带宽的比值，即

$$N(t)=\mathrm{clip}\left(\frac{B_{\mathrm{avail}}(t)}{B_{\mathrm{req}}(t)},0,1\right)\cdot \exp\left(-\eta\max\left(0,\frac{D(t)-D_0}{D_0}\right)\right) \tag{10}$$

其中，$B_{\mathrm{avail}}(t)$为时刻t网络可用带宽，$B_{\mathrm{req}}(t)$为当前语义帧在单位时间内的需求带宽，其由编码端生成的语义比特数与传输时间窗口共同决定；$D(t)$为端到端传输时延，D_0为任务可接受的时延门限，η为时延惩罚系数，用于控制超过时延门限的衰减强度。$T_{h,w,c}$为任务优先级因子，可根据任务类型调整语义权重的分布，如搜救任务强化人员与通道、导航任务强化障碍与路径。时间约束因子$C(t)=\exp\left(-\lambda(t-t_0)\right)$用于刻画语义信息随任务推进而逐渐失效的时效性特征，其中t_0为当前任务的起始时刻，λ为时间衰减率。此外，为避免量纲与数值范围上的不同，各时间约束因子先归一化为$[0,1]$，并通过加权线性组合的方式共同决定最终语义权重。

随后将每个权重值映射到其对应的网格位置，得到一个与多通道二值特征图空间位置严格对应的连续权重矩阵$\boldsymbol{M}$。该方法在保持语义筛选优势的同时，引入任务、网络与时间的多维约束，实现了更精细化和自适应的资源调度，为后续熵编码与解码提供了更合理的输入分布。

空间感知特征选择的具体实现非常直观，即

$$\tilde{\boldsymbol{F}}=(\boldsymbol{S}\odot\boldsymbol{F})\odot\boldsymbol{M} \tag{11}$$

其中，$\odot$表示在通道维度广播后做逐元素乘积，$\tilde{\boldsymbol{F}}$表示稀疏化后的BEV特征图，表示为

$$\tilde{\boldsymbol{F}}_{h,w,c}=\left(\boldsymbol{S}_{h,w,c}\times\boldsymbol{F}_{h,w,c}\right)\times\boldsymbol{M}_{h,w,c},\quad 0\leqslant i<H-1,0\leqslant j<W-1,1\leqslant c\leqslant C \tag{12}$$

通过上述的逐元素乘积操作将对应的特征全部置零，并对置一部分的所有特征进行权重分配，这一操作不仅减少了后续需要进行编码和传输的数据量，也使后续压缩模块能更高效地利用稀疏结构进行零值压缩，从而显著降低比特率。这样做又可保证后续能够获取丰富的几何特征、颜色特征和深度语义特征，满足下游任务（如目标检测、分类）对精度和鲁棒性的要求。

2.2　任务语义权重动态量化超编码

在完成空间感知特征筛选后，得到已被稀疏化的BEV特征图$\tilde{\boldsymbol{F}}\in\mathbb{R}^{H\times W\times C}$，其绝大多数位置已被置零，仅保留与任务相关的网格区域。尽管此时数据量已大幅降低，但考虑到网络带宽限制问题，即使是筛选后的数据，在极端环境下仍然面临传输瓶颈，因此仍需对$\tilde{\boldsymbol{F}}$进行高效压缩。运用层次化熵模型[29]，在保证重建保真度的前提下实现接近最小平均比特率的量化与编码，从而实现高效的空间数据压缩。

由于BEV特征图$\tilde{\boldsymbol{F}}$中保存的是浮点数特征，后续的熵编码器无法直接处理浮点数，因此必须执行离散化量化操作。令$\mathcal{Q}(\cdot)$表示一个均匀量化函数，将每个通道值近似映射到一个离散整数集合中，具体过程可表示为

$$\hat{\boldsymbol{F}}=\mathcal{Q}(\tilde{\boldsymbol{F}}) \tag{13}$$

$$\hat{\boldsymbol{F}}_{h,w,c}=\mathrm{round}\left(\frac{\tilde{\boldsymbol{F}}_{h,w,c}}{q}\right)q \tag{14}$$

其中，q是根据语义权重得到的动态量化步长，表示为

$$q=\Delta_q\left(2-\overline{m}_{h,w,c}\right) \tag{15}$$

其中，Δ_q表示均匀量化步长，$\overline{m}_{h,w,c}$由语义权重$m_{h,w,c}$做归一化处理后得到，语义权重越高的区域，量化步长越小，编码精度越高，反之则采用更大步长以提升压缩率，$\mathrm{round}(\cdot)$表示四舍五入操作。量化的作用在于将连续特征映射到有限集合，从而可以用离散概率模型表示每个量化值出现的概率分布，由于传输的点云已经是按语义筛选而得到的子集点云，语义标签$\boldsymbol{s}_i$被视为已知先验条件，因此不需要把$\boldsymbol{s}_i$作为量化对象单独传输，而是在任务端直接赋予重建的子集点云，从而在保证语义一致性的

同时降低传输开销。

在获得离散化量化后的特征图$\hat{\boldsymbol{F}}$后，由于直接对$\hat{\boldsymbol{F}}$建立全局的联合概率模型计算复杂度往往极高，且难以准确建模所有通道之间以及相邻单元之间的联合分布，为此可以利用层次化熵模型先对$\hat{\boldsymbol{F}}$本身构造一层辅助的隐藏特征，这些隐藏特征能捕捉$\hat{\boldsymbol{F}}$在空间、通道和结构上的统计依赖关系，进而以条件概率的方式更准确地建模$\hat{\boldsymbol{F}}$分布。

首先，将$\hat{\boldsymbol{F}}$输入一个超网络中，记为

$$\boldsymbol{z} = \mathcal{H}_{\text{enc}}(\hat{\boldsymbol{F}}) \tag{16}$$

其中，$\mathcal{H}_{\text{enc}}(\cdot)$表示超编码器，该编码器由3层下采样卷积构成，其卷积核大小为3，下采样因子为2；$\boldsymbol{z} \in \mathbb{R}^{H' \times W' \times D}$表示超隐藏特征张量，其相比于$\hat{\boldsymbol{F}}$形状更小，但蕴含了$\hat{\boldsymbol{F}}$的整体分布信息，$H' \times W'$通常与$H \times W$相同或等比下采样，$D$表示超特征的通道数。超网络由若干卷积和下采样层构成，通过对$\hat{\boldsymbol{F}}$进行数次卷积与调参，提取出该层张量在空间位置、通道相关性和稀疏结构方面的显著统计信息。

接下来，需要对$\boldsymbol{z}$进行离散化量化操作，得到可供熵编码器压缩的整数张量$\hat{\boldsymbol{z}}$，同样采用均匀量化策略。

$$\hat{\boldsymbol{z}} = \mathcal{Q}(\boldsymbol{z}) \tag{17}$$

$$\hat{\boldsymbol{z}}_{h',w',d} = \text{round}\left(\frac{\boldsymbol{z}_{h',w',d}}{\Delta_{q'}}\right)\Delta_{q'} \tag{18}$$

其中，$\Delta_{q'}$是对超隐藏张量使用的量化步长，该步长以模型内部隐藏层特征的数值单位为量纲。为了使整个网络可微，在训练阶段通常会采用噪声替代量化的方法，即在正向传播的时候，用均匀的噪声近似量化扰动；在反向传播的时候，则保留近似梯度，使训练能够进行，因此选取默认值1作为对超隐藏张量使用的量化步长值。

随后，为了使熵编码的码率更接近数学上的最小熵，需先利用$\hat{\boldsymbol{z}}$预测出$\hat{\boldsymbol{F}}$的概率，再做精细化编码，记概率模型为$q(\hat{\boldsymbol{z}};\phi)$，其中$\phi$是在训练时学习得到的模型参数。编码端根据概率模型对$\hat{\boldsymbol{z}}$进行算术编码，生成比特流$\mathcal{B}_z$。

当对$\hat{\boldsymbol{F}}$进行熵编码时，先在概率上假设每个离散的$\hat{\boldsymbol{F}}_{h,w,c}$满足

$$q(\hat{\boldsymbol{F}}_{h,w,c}|\hat{\boldsymbol{z}}) = \left(\mathcal{N}(\mu_{h,w,c},\sigma^2_{h,w,c}) \mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)\right)(k) \tag{19}$$

其中，$\mathcal{N}(\mu,\sigma^2)\mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)$表示先假设浮点特征$\tilde{\boldsymbol{F}}_{h,w,c}$在量化前后与离散化后值之差满足标准均匀分布$\mathcal{U}\left(-\frac{1}{2},\frac{1}{2}\right)$，再将离散化后的值建模为高斯分布$\mathcal{N}(\mu,\sigma^2)$，该高斯分布由超解码器输出的$(\mu_{h,w,c},\sigma^2_{h,w,c})$来控制，$k$表示$\hat{\boldsymbol{F}}_{h,w,c}$对应的离散整数索引。借助这一混合模型，可以近似地计算每个$\hat{\boldsymbol{F}}_{h,w,c}$在给定条件$\hat{\boldsymbol{z}}$下的概率，如式(20)所示。

$$P(\hat{\boldsymbol{F}}_{h,w,c}|\hat{\boldsymbol{z}}) = \int_{\hat{\boldsymbol{F}}_{h,w,c}-\frac{1}{2}}^{\hat{\boldsymbol{F}}_{h,w,c}+\frac{1}{2}} \mathcal{N}(x;\mu,\sigma^2)\text{d}x \tag{20}$$

由于积分区间为1，使每个离散值都能得到准确的概率估计，因此将所有概率$P(\hat{\boldsymbol{F}}_{h,w,c}|\hat{\boldsymbol{z}})$连乘，便可得到整个$\hat{\boldsymbol{F}}$的条件概率为

$$q(\hat{\boldsymbol{F}}|\hat{\boldsymbol{z}}) = \prod_{h=0}^{H-1}\prod_{w=0}^{W-1}\prod_{c=1}^{C}\mathcal{P}(\hat{\boldsymbol{F}}_{h,w,c}|\hat{\boldsymbol{z}}) \tag{21}$$

在编码端，根据所得概率$q(\hat{\boldsymbol{F}}|\hat{\boldsymbol{z}})$对$\hat{\boldsymbol{F}}$进行算数编码，生成比特流$\mathcal{B}_F$。此时的熵编码器充分利用$\hat{\boldsymbol{z}}$所提供的先验信息，在对每个$\hat{\boldsymbol{F}}_{h,w,c}$进行熵编码时动态调整码字长度，即当$\hat{\boldsymbol{F}}_{h,w,c}$的先验概率高时，则分配更短的码字；当$\hat{\boldsymbol{F}}_{h,w,c}$的先验概率低时，则分配更长的码字。此时，经过量化与熵编码后得到的压缩结果形式主要为二进制比特流文件，其大小远小于原始浮点量化特征需要的存储空间。在网络传输阶段，该比特流不仅能够快速通过带宽受限的链路，还可以降低端到端时延，从而实现高效的空间数据传输。

2.3 具身场景解码与语义对齐重建

为了保证接收端场景重建的几何与语义精度，解码与重建的重点在于保证语义与几何的一致性，并尽可能地降低解码和计算开销，实现通信传输的高效性与感知精度的平衡。解码端恢复稀疏BEV特征图后，结合传输的稀疏位置索引，逐一找到非零网格单元对应的三维柱状区域，直接恢复原始点云中落入该区域的点，即恢复原始点索引子集，从而保证重建点云的几何一致性。

首先，接收端从通信信道中接收到由前面层次化熵编码模型得到的两端比特流$\mathcal{B}_z$和$\mathcal{B}_F$。当$\mathcal{B}_z$传输到接收端后，接收端会使用与编码端相同的概率模型$q(\hat{\boldsymbol{z}};\phi)$来执行解算术编码，恢复出$\hat{\boldsymbol{z}}$的确切离散值，记为

$$\hat{\boldsymbol{z}} = \mathcal{E}_{\text{dec}}(\mathcal{B}_z) \tag{22}$$

然后，接收端使用超解码器$\mathcal{H}_{\text{dec}}(\cdot)$估计原始特征的条件分布参数，即

$$(\mu,\sigma^2) = \mathcal{H}_{\text{dec}}(\hat{\boldsymbol{z}}) \tag{23}$$

其中，$\mathcal{H}_{\text{dec}}(\cdot)$的结构与编码端的$\mathcal{H}_{\text{enc}}(\cdot)$是相对应的，即$\mathcal{H}_{\text{dec}}(\cdot)$使用对称的反卷积结构逐步上采样，其卷积核大小和上采样因子的数值与超编码器相同。输出的$\mu,\sigma^2 \in \mathbb{R}^{H\times W\times C}$分别表示在每个BEV网格和通道下，对离散特征$\hat{\boldsymbol{F}}_{h,w,c}$的高斯潜在分布的均值与方差的估计值。接收端在还原出$\hat{\boldsymbol{z}}$后，又通过超解码器$\mathcal{H}_{\text{dec}}(\cdot)$计算出精确的$\mu,\sigma^2$，根据在编码阶段所假设的$\hat{\boldsymbol{F}}_{h,w,c}$服从$q(\hat{\boldsymbol{F}}_{h,w,c}|\hat{\boldsymbol{z}})$便可对比特流$\mathcal{B}_F$进行最后的熵解码，最终还原出原始特征向量$\hat{\boldsymbol{F}}$。

随后，需要对$\hat{\boldsymbol{F}}$进行反量化操作，以近似恢复成连续值张量$\tilde{\boldsymbol{F}}$。编码阶段所用的量化步长为q，则反量化过程可表示为

$$\begin{aligned}&\tilde{\boldsymbol{F}}_{h,w,c} = \hat{\boldsymbol{F}}_{h,w,c} \times q,\\ &0 \leqslant h < H, 0 \leqslant w < W, 1 \leqslant c \leqslant C\end{aligned} \tag{24}$$

在完成熵解码与反量化操作之后，$\tilde{\boldsymbol{F}} \in \mathbb{R}^{H\times W\times C}$即接收端恢复的稀疏化BEV特征图，其仅在少数$\boldsymbol{S}_{h,w} = 1$的柱状单元网格位置含有非零数值，其他位置均为零。为保证重建的点云与语义能保持一致，先将连续权重矩阵$\boldsymbol{M}$量化为K个离散优先级别，如一、二、三级。

$$L = \mathcal{Q}(\boldsymbol{M}) \in \{1,\cdots,K\} \tag{25}$$

并定义从优先级到任务相关语义集合$\mathcal{C}_{\text{task}}$的映射，即

$$\boldsymbol{N}:\{1,\cdots,K\} \to \mathcal{C}_{\text{task}} \tag{26}$$

例如，当$\{1,2,3\}$对应的语义优先级集合为{车辆、树木、草丛}时，若$L = 1$，则$N(1) =$车辆；对于任意L，$\boldsymbol{N}(L)$为单一确定语义标签，且不同优先级不共享同一语义标签，从而保证了后续语义赋值在解码端无歧义、不需要额外标识即可恢复。随后，结合先前保存的柱状单元(h,w)将$\tilde{\boldsymbol{F}}$映射回三维点云子集$\hat{\boldsymbol{P}}_{\text{subset}}$，并对由BEV特征图$\tilde{\boldsymbol{F}}$重建的所有点采用该特征图的任务语义优先级直接赋予语义标签，即对任意的$\boldsymbol{p}_k \in \hat{\boldsymbol{P}}_{\text{subset}}$有

$$\boldsymbol{s}_k' = \boldsymbol{N}(L) \tag{27}$$

该硬映射保证了后端任务优先级直接、统一地注入重建点云上，且解码端仅凭L即可完成细粒度语义恢复，因此不需要额外传输类别标识。值得指出的是，上述优先级到语义的一对一硬映射是针对特种机器人任务场景的确定性设计选择。在目标应用中，任务关注的语义类别数量有限，且可在系统部署前预先配置，因此一对一映射足以覆盖实际需求并保证解码端语义恢复的确定性与低开销。至此，子集点云中的每一个点都与原始点云中的某一对应点完全一致，确保了最终输出的三维点云具备“几何一致性”与“语义相关性”两大特点。遍历所有任务相关语义及点云后，即完成了场景解码与重建。

3　实验与分析

本文基于S3DIS[30]和Semantic3D[31]数据集进行相关实验验证，两个数据集都提供了丰富的语义分割标注。实验采用80%和20%的数据集划分策略，随机将原始数据集分为训练集和测试集。需要说明的是，本文实验中的语义标签来自数据集提供的逐点GT语义标注，不需要额外的预测网络。为了训练模型，使用配备了4块NVIDIA GeForce RTX 3080 GPU并行计算的工作站，同时配置了128 GB内存以处理大规模数据集的加载和模型参数存储。资源开销统计基于PyTorch框架，在推理阶段通过内置接口记录模型参数量、计算复杂度（单位为MAC）与GPU峰值显存，占用统计过程不影响原有实验流程与性能评估结果。在具体实现中，BEV特征图映射网格尺寸设置为$(\Delta x,\Delta y,\Delta z) = (0.1,0.1,0.2)$ m，点云坐标范围根据每个场景实际边界自适应计算，均匀量化步长$\Delta_q = 0.000\,5$，网络状态因子计算中默认$B_{\text{avail}}(t)$=10 Mbit/s，D_0=200 ms，$\eta = 1$，时间约束因子采用半衰期参数化，实验中默认采用均匀权重配置，相关权重系数之和满足归一化约束。在码率统计方面，本文的压缩增益与数据缩减比均已将超先验辅助比特流和稀疏位置索引的开销纳入总码率计算。BEV特征图网格超参数作为双端预先约定的配置信息，不计入传输开销。任务优先级到语义类别映射采用一对一策略，即每个离散优先级仅绑定一个语义标签，用于保证解码端语义恢复的确定性与低开销。以下将展示评估语义赋能的空间感知编解码技术性能的实验结果。

本文采用编解码时间、编码压缩增益、点数保持率[32]、Chamfer距离[33]、语义重心偏移[34]、峰值信噪比（peak signal-to-noise ratio, PSNR）[35]这6项评价指标。编解码时间包括压缩和解压所需的平均总时间开销，直接影响算法的实时性能；编码压缩增益衡量数据压缩效果，定义为压缩前后数据大小的比值，在编码性能统计过程中，压缩后数据大小统一按编码比特数进行统计，并可换算为标准码率指标bit/point，以保证不同方法之间比较的公平性；点数保持率量化重建后保留的点云比例，数值越接近1意味着保留了更多的原始信息点；Chamfer距离通过双向最近邻距离评估两个点云的几何相似性，Chamfer距离越接近0表示系统输出与原始点云的差异越小；语义重心偏移测量语义类别重心的位移程度，越小表示输出语义与原文语义的偏移越小；PSNR作为全局误差度量指标，计算对象为重建与原始BEV特征图之间的差异，计算式为 $\mathrm{PSNR}=10\lg\left(\frac{\mathrm{MAX}^2}{\mathrm{MSE}}\right)$，其中，MAX为BEV特征图的最大幅值，MSE为重建与原始特征图的逐元素均方误差（mean square error, MSE），PSNR越高意味着压缩算法在噪声条件下“恢复”得越好，该模型的抗干扰能力越强。

图4展示了掩码生成方法在3类干扰条件下的鲁棒性。在物体重叠与点云稠密度干扰条件下，F1分数分别为0.954和0.941；在语义边界模糊干扰条件下，F1分数下降至0.888。召回率在所有干扰条件下均高于0.995，精确率在最严苛干扰条件下仍维持在0.799以上，验证了该方法在复杂环境下的稳定性。

图5对比了不同编解码方法在处理时间与压缩效率方面的性能。为保证压缩性能比较的公平性，不同编解码方法均在语义筛选后的点云子集上进行。在传统方法中，Draco编码压缩增益最高约23.6，但处理时间达1 933.8 ms；MPEG-GPCC编码压缩增益约3.9，耗时2 180.5 ms；Open3D处理时间最快，但编码压缩增益较低；3DAC与Learned-PCGC分别在538.8 ms和464.6 ms内实现了10.94与21.87的编码压缩增益。本文方法兼顾语义筛选与局部编码压缩，在853.9 ms内实现了约5.54的编码压缩增益，表明任务驱动的语义筛选可在系统层面显著缩减传输数据规模。

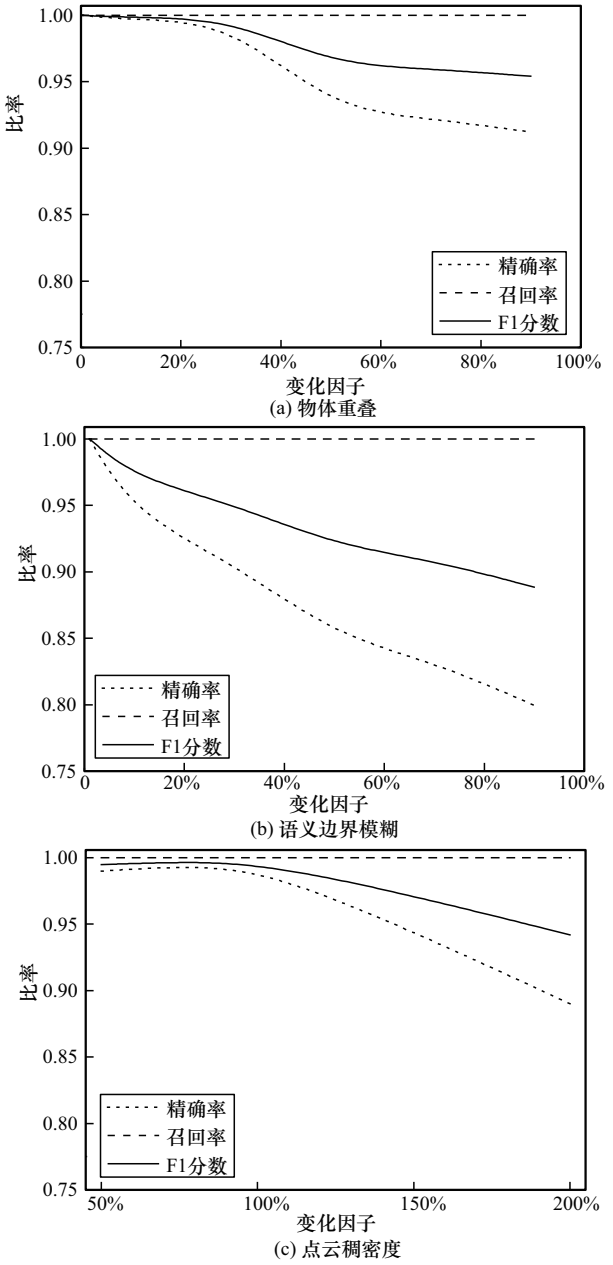

图4　不同干扰条件下掩码生成方法的鲁棒性比较

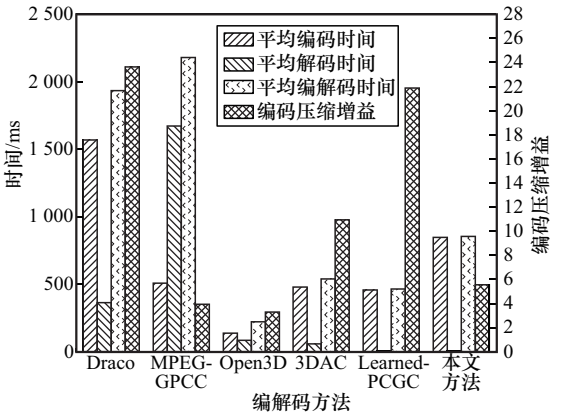

图5　编解码方法比较

为评估本文方法在资源受限平台上的可部署性，本文进一步分析了推理阶段的计算开销。语义筛选模块仅包含约 0.36×10^6 可训练参数，计算复杂度约为 0.131 GMAC（约 0.26 GFLOP），单场景单帧推理所需 GPU 峰值显存约为 327 MB。结果表明，该方法在模型规模、算力需求与存储开销方面均保持较低水平，具备良好的工程可实现性。

表 1 比较了各方法在任务关注区域的重建质量性能。Draco 点数保持率为 1，但统一量化导致 Chamfer 距离偏大；MPEG-GPCC 几何细节损失明显；3DAC 与 Learned-PCGC 受生成模型精度限制存在一定几何偏移。需指出，点数保持率与 Chamfer 距离衡量不同维度，前者反映数量保留，后者更准确地刻画空间几何差异。本文方法在点数保持率接近 1 的同时，实现了更低的 Chamfer 距离，并将语义重心偏移控制在 0.012 28 以内，在几何精度与语义保持之间取得良好平衡。

表 1　　重建质量比较

编解码方法	点数保持率	Chamfer 距离	语义重心偏移
Draco	1	0.008 53	0.001 45
MPEG-GPCC	0.561 68	2.577 03	0.068 03
Open3D	0.303 2	1.552 9	0.013 58
3DAC	1	0.026 3	0.013 89
Learned-PCGC	1	0.028 7	0.014 58
本文方法	0.949 99	0.001 5	0.012 28

图 6 展示了不同编解码方法在高斯噪声、突发噪声、比特翻转噪声和量化噪声条件下的 PSNR，其中本文采用的比特翻转噪声并非直接作用于算术编码输出的压缩比特流，而是作用于量化后和熵编码前的离散特征值，以此模拟量化误差或信道扰动对特征表示的影响，同时确保解码流程正常运行。本文方法的重建 PSNR 均保持在 28~31 dB，整体性能表现显著优于对比方法，验证了其在多种信道扰动条件下的稳定性与抗噪能力。

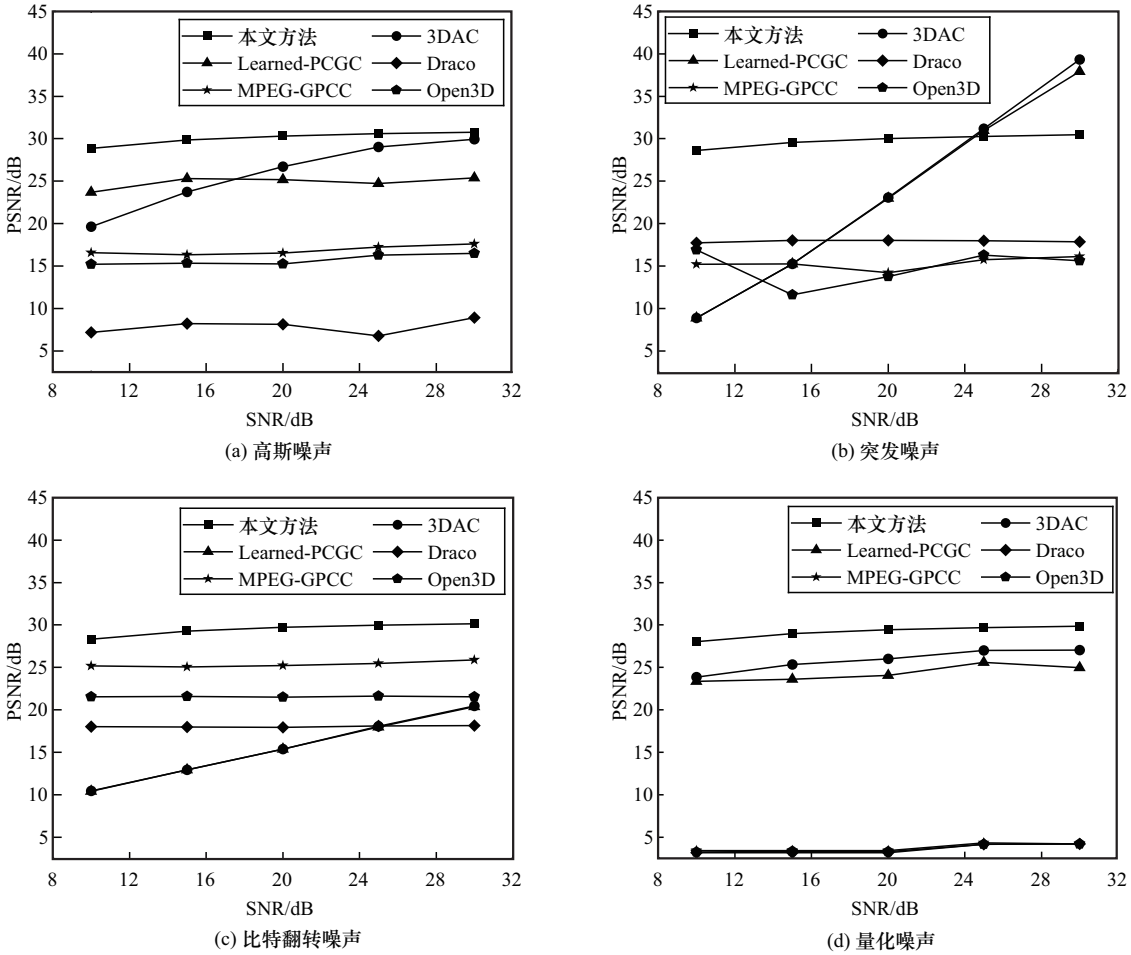

图 6　不同编解码方法在不同类型噪声下的 PSNR 值比较

实验结果表明，本文方法在各种噪声环境下均表现出优异的抗噪性能和稳定性，相比传统方法具有显著的鲁棒性优势，验证了该方法在复杂噪声环境下的实用性和可靠性。

4 结束语

本文强调了语义赋能的空间感知编解码技术在推动特种机器人环境感知与具身智能决策方面的重要意义，巧妙地解决了大规模场景实时重建的计算复杂性、受限网络环境下的数据传输效率以及语义信息高保真度重建等关键挑战。通过严格的实验验证，证明了本文方法在不同信噪比条件下均能保持优异的重建质量和语义一致性。实验结果表明，本文方法在各种噪声环境下均表现出卓越的抗干扰性能和稳定性，相比传统编解码方法具有显著优势。在编解码时间、压缩率、几何精度保持以及语义完整性等关键指标上，本文方法实现了良好的综合平衡。特别是在恶劣网络条件下，本文方法仍能保持较高的重建质量，验证了其在实际应用中的实用性和可靠性。未来工作将进一步通过系统性的消融实验，对掩码机制、动态量化、超先验建模以及语义对齐模块在压缩效率与任务性能中的贡献进行更加深入的分析。

本文计划扩展该框架的适用性，并在更广泛的场景和环境中严格测试其鲁棒性。未来的工作将重点关注以下几个方向：1) 构建任务与语义映射的数学模型，引入更一般化的映射机制，从信息论与决策论角度刻画“任务约束下的最优编码目标”，为动态量化与权重分配提供理论支撑；2) 设计更优的自适应权重分配算法，使系统在多任务并发与资源竞争场景下保持全局性能最优；3) 开发语义优先级决策树或策略库，提升系统对烟雾、遮挡、强反光等极端干扰条件的鲁棒性；4) 扩展多模态与协同感知能力，探索在多机器人协同作业场景中的应用潜力。相信随着技术的不断进步和应用场景的拓展，语义赋能的空间感知编解码技术将在智慧城市、应急救援、工业检测等领域发挥更加重要的作用，为构建更加智能化的机器人生态系统贡献力量。

参考文献：

[1] Bogue R. The role of robots in environmental monitoring[J]. Industrial Robot: The International Journal of Robotics Research and Application, 2023, 50(3): 369-375.

[2] Lee D, Jung M, Yang W, et al. LiDAR odometry survey: recent advancements and remaining challenges[J]. Intelligent Service Robotics, 2024, 17(2): 95-118.

[3] Guo Y L, Wang H Y, Hu Q Y, et al. Deep learning for 3D point clouds: a survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021, 43(12): 4338-4364.

[4] Hughes N, Chang Y, Hu S Y, et al. Foundations of spatial perception for robotics: hierarchical representations and real-time systems[J]. The International Journal of Robotics Research, 2024, 43(10): 1457-1505.

[5] Dawarka V, Bekaroo G. Building and evaluating cloud robotic systems: a systematic review[J]. Robotics and Computer-Integrated Manufacturing, 2022, 73: 102240.

[6] Fang G C, Hu Q Y, Wang H Y, et al. 3DAC: learning attribute compression for point clouds[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 14799-14808.

[7] Song C H, Blukis V, Tremblay J, et al. RoboSpatial: teaching spatial understanding to 2D and 3D vision-language models for robotics[C]//Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2025: 15768-15780.

[8] Shridhar M, Manuelli L, Fox D. CLIPort: what and where pathways for robotic manipulation[C]//Proceedings of the Conference on Robot Learning. New York: PMLR, 2022: 894-906.

[9] Azuma D, Miyanishi T, Kurita S, et al. ScanQA: 3D question answering for spatial scene understanding[C]//Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2022: 19107-19117.

[10] Leutenegger S. OKVIS2: realtime scalable visual-inertial SLAM with loop closure[PP]. V2. (2022-08-12) [2025-11-26]. arXiv: arXiv.2202.09199.

[11] Bultmann S, Behnke S. 3D semantic scene perception using distributed smart edge sensors[C]//International Conference on Intelligent Autonomous Systems. Berlin: Springer, 2023: 313-329.

[12] Chen B Y, Xia F, Ichter B, et al. Open-vocabulary queryable scene representations for real world planning[C]//Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway: IEEE Press, 2023: 11509-11522.

[13] Jia Y P, He J, Chen R Z, et al. OccupancyDETR: making semantic scene completion as straightforward as object detection[PP]. V3. (2024-05-18) [2025-11-26]. arXiv: arXiv.2309.08504.

[14] Li B H, Jin X, Wang J N, et al. OccScene: semantic occupancy-based cross-task mutual learning for 3D scene generation[PP]. V2. (2025-08-22) [2025-11-26]. arXiv: arXiv.2412.11183.

[15] Qi Z Y, Zhang Z X, Fang Y, et al. GPT4Scene: understand 3D scenes from videos with vision-language models[PP]. V4. (2025-03-11) [2025-11-26]. arXiv: arXiv.2501.01428.

[16] Jiang J P, Xiao W Y, Lin Z Y, et al. SOLAMI: social vision-language-action modeling for immersive interaction with 3D autonomous characters[C]//Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2025: 26887-26898.

[17] Fu H S, Liang F, Lin J P, et al. Learned image compression with Gaussian-Laplacian-logistic mixture model and concatenated residual modules[J]. IEEE Transactions on Image Processing, 2023, 32: 2063-2076.

[18] He J, Gong H X, Lu H Y. Design of fractal image coding compression and transmission model based on wavelet transform[C]//International Conference on Advanced Hybrid Information Processing. Berlin: Springer, 2022: 15-25.

[19] Yuan F, Zhan L H, Pan P W, et al. Low bit-rate compression of underwater image based on human visual system[J]. Signal Processing: Image Communication, 2021, 91: 116082.

[20] Yang J H, Yu H, Li P, et al. Real-time D-PMU data compression for edge computing devices in digital distribution networks[J]. IEEE Transactions on Power Systems, 2024, 39(4): 5712-5725.

[21] Lu M, Guo P Y, Shi H Q, et al. Transformer-based image compression[PP]. V1. (2021-11-12) [2025-11-26]. arXiv: arXiv.2111.06707.

[22] Khoshkhahtinat A, Zafari A, Mehta P M, et al. Neural-based video compression on solar dynamics observatory images[J]. IEEE Transactions on Aerospace and Electronic Systems, 2024, 60(5): 6685-6701.

[23] Yasuda M, Ohishi Y, Saito S, et al. Multi-view and multi-modal event detection utilizing transformer-based multi-sensor fusion[C]//Proceedings of the ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway: IEEE Press, 2022: 4638-4642.

[24] Zhou Q H, Chen S L, Wang Y S, et al. HAZARD challenge: embodied decision making in dynamically changing environments[PP]. V1. (2024-01-23) [2025-11-26]. arXiv: arXiv.2401.12975.

[25] Gao G B, Zhou D M, Tang H, et al. An intelligent health diagnosis and maintenance decision-making approach in smart manufacturing[J]. Reliability Engineering & System Safety, 2021, 216: 107965.

[26] Coito T, Firme B, Martins M S E, et al. Intelligent sensors for real-time decision-making[J]. Automation, 2021, 2(2): 62-82.

[27] Ahn M, Brohan A, Brown N, et al. Do as I can, not as I say: grounding language in robotic affordances[PP]. V2. (2022-08-16) [2025-11-26]. arXiv: arXiv.2204.01691.

[28] Lang A H, Vora S, Caesar H, et al. PointPillars: fast encoders for object detection from point clouds[C]//Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2019: 12689-12697.

[29] Ballé J, Minnen D, Singh S, et al. Variational image compression with a scale hyperprior[PP]. V2. (2018-05-01) [2025-11-26]. arXiv: arXiv.1802.01436.

[30] Armeni I, Sener O, Zamir A R, et al. 3D semantic parsing of large-scale indoor spaces[C]//Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE Press, 2016: 1534-1543.

[31] Hackel T, Savinov N, Ladicky L, et al. Semantic3D.net: a new large-scale point cloud classification benchmark[PP]. V1. (2017-04-12) [2025-11-26]. arXiv: arXiv.1704.03847.

[32] Wang J Q, Zhu H, Ma Z, et al. Learned point cloud geometry compression[PP]. V1. (2019-09-26) [2025-11-26]. arXiv: arXiv.1909.12037.

[33] Wu T, Pan L, Zhang J Z, et al. Density-aware chamfer distance as a comprehensive metric for point cloud completion[PP]. V1. (2021-11-24) [2025-11-26]. arXiv: arXiv.2111.12702.

[34] Mari D, Camuffo E, Milani S. CACTUS: content-aware compression and transmission using semantics for automotive LiDAR data[J]. Sensors, 2023, 23(12): 5611.

[35] Xu Y W, Chen D F, Fang Y, et al. Efficient vibrotactile codec based on nbeats network[J]. IEEE Signal Processing Letters, 2024, 31: 2845-2849.

[作者简介]

陈鸣锴（1989– ），男，福建宁德人，博士，南京邮电大学副教授、硕士生导师，主要研究方向为多媒体通信与计算、无线网络中的资源分配与信号处理等。

刘沁妍（2001– ），女，江苏南京人，南京邮电大学硕士生，主要研究方向为多媒体通信、图像处理与触觉信号压缩等。