

结合有序建模和空间感知的点云语义分割算法

王一凡¹, 刘 骊^{1,2}, 付晓东^{1,2}, 刘利军^{1,2}, 彭 玮^{1,2}

¹(昆明理工大学 信息工程与自动化学院, 昆明 650500)

²(昆明理工大学 信息工程与自动化学院 云南省计算机技术应用重点实验室, 昆明 650500)

E-mail: ieall@kust.edu.cn

摘要: 针对三维点云语义分割中的无序性、几何信息缺失及边界模糊等挑战, 提出一种结合有序建模和空间感知的点云语义分割算法。首先通过融合 Z 阶与希尔伯特空间填充曲线及其镜像变体进行多种空间填充曲线的互补有序建模, 构建空间结构保持性互补的点云编码序列; 然后设计动态空间编码机制, 根据不同阶段任务需求自适应调整编码维度, 提升空间编码效率; 最后构建空间感知聚合模块, 融合空间引导的局部特征传播与多层感知机全局上下文感知, 实现特征高效混合学习, 增强特征的空间稳定性与几何一致性。在 S3DIS、ScanNet v2 和 ScanObjectNN 3 个公开点云数据集上的实验表明, 所提算法实现了高精度语义分割, 有效提升了对复杂边界区域和小物体的空间感知能力和语义理解精度。

关键词: 三维点云语义分割; 空间填充曲线; 互补有序建模; 动态空间编码; 空间感知聚合

中图分类号: TP391

文献标识码: A

文章编号: 1000-1220(2026)05-1198-07

Ordered Modeling and Spatial Perception for Point Cloud Semantic Segmentation Algorithm

WANG Yifan¹, LIU Li^{1,2}, FU Xiaodong^{1,2}, LIU Lijun^{1,2}, PENG Wei^{1,2}

¹(Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

²(Computer Technology Application Key Lab of Yunnan Province, Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650500, China)

Abstract: To address the challenges of disorder, missing geometric information, and blurred boundaries in 3D point cloud semantic segmentation, this paper proposes an algorithm that combines complementary ordered modeling with spatial perception. A complementary ordered modeling module is first constructed by integrating Z-order curves, Hilbert curves, and their mirrored variants, forming spatially consistent and structurally complementary point cloud encoding sequences. Then, a dynamic spatial encoding mechanism is introduced to adaptively adjust the encoding dimensions according to the requirements of different stages, thereby improving the efficiency of spatial representation. Finally, a spatial-aware aggregation module is designed to fuse spatially guided local feature propagation with multilayer perceptron-based global context modeling, achieving efficient hybrid feature learning while enhancing spatial stability and geometric consistency. Experiments conducted on three public datasets—S3DIS, ScanNet v2, and ScanObjectNN—demonstrate that the proposed algorithm delivers high-accuracy semantic segmentation and significantly enhances spatial perception and semantic understanding in complex boundary regions and small object scenarios.

Keywords: 3D point cloud semantic segmentation; space-filling curves; complementary ordered modeling; dynamic spatial encoding; spatial-aware aggregation

0 引言

三维点云语义分割^[1]是将点云中的每个点赋予特定的语义标签, 实现场景的精确理解与解析, 可广泛应用于智能建筑中的室内结构识别、自动驾驶中的道路场景分析、增强现实中的环境理解与交互等。传统的点云分割主要使用机器学习方法, 首先从点云数据中提取手工设计的几何或统计特征, 然后使用支持向量机(Support Vector Machine, SVM)等分类器训练特征。如 Yu 等^[2]采用欧氏聚类算法将点云分割为独立的目标工作区域, 从分割后的点云中提取特征, 然后利用随机

森林的集成投票机制对特征进行学习, 但其对数据质量与手工经验要求较高, 且特征提取与学习过程分离, 难以适应大规模、非结构化的真实场景点云分割任务。

区别于以上方法, 基于深度学习的点云分割能够自动从数据中学习多尺度、多通道的几何和语义特征, 针对点云无序、非结构化特性, Qi 等^[3]提出 PointNet 方法, 使用全局最大池化作为对称函数对点云特征进行聚合, 实现点集的排列不变性, 但缺乏对局部结构的建模, 难以捕获细粒度的局部特征。Zhao 等^[4]利用 Transformer 架构的排列不变性, 通过注意力机制建模点之间的关系, 并通过学习得分对点进行排序, 实

收稿日期: 2025-05-30 收修改稿日期: 2025-07-01 基金项目: 国家自然科学基金项目(62262036, 62362043)资助; 云南省兴滇英才支持计划项目(KKXY202203008)资助。 作者简介: 王一凡, 男, 2001年生, 硕士研究生, CCF 学生会员, 研究方向为计算机图形学与计算机视觉; 刘 骊(通信作者), 女, 1979年生, 博士, 教授, 博士生导师, CCF 高级会员, 研究方向为计算机图形学与计算机视觉、图像处理; 付晓东, 男, 1975年生, 博士, 教授, 博士生导师, CCF 高级会员, 研究方向为服务计算、决策理论与方法; 刘利军, 男, 1978年生, 博士, 副教授, CCF 会员, 研究方向为图像处理、云计算; 彭 玮, 女, 1980年生, 博士, 教授, 博士生导师, CCF 会员, 研究方向为机器学习、数据挖掘。

现输入的置换不变性,但参数数量会随通道数量的增加而增加,限制了模型的深度与精度.为了提升效率,许多研究引入空间填充曲线^[5]对点云进行有序建模,Chen 等^[6]利用 Z 阶曲线对原始点云进行编码,在采样时优先保留具有几何相关性的点,以增强后续特征提取对形状结构的敏感度,但仅使用单一空间填充曲线,使模型难以适应非均匀点云的局部几何变换.为此,Qi 等^[7]提出相对空间编码,在特征提取过程中引入点与点之间的相对位置信息,增强了模型对空间结构的理解能力;Qian 等^[8]去除原始的显式相对空间编码,通过局部点集特征学习来隐式建模空间关系,然而上述方法均采用固定维度的静态空间编码机制,限制了其在处理复杂几何形状场景时的精度和适应性.

综上所述,三维点云语义分割仍存在以下难点:1)由于点云天然无拓扑,现有基于空间填充曲线的序列化方法主要依赖单一曲线,导致点云信息与序列化结构缺失,难以适应非均匀点云的局部几何变化;2)现有空间编码多采用固定维度的静态编码机制,面对不同尺度和几何结构的点云数据时,难以适应不同阶段的任务需求,实现细粒度分割.

针对上述难点,本文提出一种结合有序建模和空间感知的三维点云语义分割算法.该算法构建互补有序建模模块,采用多种空间填充曲线对点云数据进行互补有序建模,以适应非均匀点云的结构变化;同时结合动态空间编码与空间感知聚合模块,根据不同训练阶段与特征尺度动态调整空间编码维度与特征聚合深度,协同建模局部特征和全局特征,以提高点云语义分割的精度和鲁棒性.

1 相关工作

基于深度学习的三维点云语义分割根据实现方法的不同大致可分为三种:基于投影的方法、基于体素的方法和基于点的方法.

基于投影的方法将三维点云投影到二维图像平面上,利用二维卷积神经网络(Convolutional Neural Network, CNN)进行特征提取.Chen 等^[9]利用多视图方法,将激光雷达点云投影到二维视图鸟瞰图与前视图中,并使用 CNN 从二维视图中提取特征.为解决视图对齐和视角遮挡问题,Xu 等^[10]提出端到端的 SqucczcScgV3 网络,对球面投影后生成的二维图像应用空间自适应卷积,自适应调整卷积操作.虽然上述方法有效解决了多视图的部分缺陷,但仍无法充分利用点云的几何和结构信息,影响分割边界的准确性.

基于体素的方法将连续的点云数据划分为固定大小的三维体素网格,再利用三维 CNN 进行特征提取和语义标签预测.Wu 等^[11]将三维点云表示为三维体素网格上的二值变量概率分布,然后利用卷积深度置信网络学习复杂形状分布,实现高效点云学习.Chen 等^[12]提出 VoxelNeXt 的全稀疏体素网络,直接从稀疏体素特征中检测与分割三维目标,并结合稀疏高度压缩和体素修剪提升效率.然而,以上方法仍存在内存消耗大和细节模糊等问题,难以有效应用于大规模点云场景的语义分割.

基于点的方法直接处理非结构化的点云数据,PointNet 方法^[3]是直接处理点云数据的代表方法.为解决点云密度不

均匀特性,Lu 等^[13]结合图神经网络与边缘损失约束,使用动态图卷积网络进行特征学习的同时增加边缘损失,提高分割边界的准确性.Lai 等^[14]提出 Stratifiedformer 方法,对每个查询点的近邻点进行密集采样,远距离点进行分层稀疏采样,并融合局部和远程特征提高分割性能.为进一步提升效率,Wu 等^[15]采用空间填充曲线代替传统的 k 近邻搜索,对点云进行序列化,并基于序列化局部注意力机制进行条件位置编码.然而,上述方法的计算复杂度会随着输入长度的增加而显著增长,为此,Liang 等^[16]将状态空间模型(State Space Model, SSM)应用到点云分割中,提出 PointMamba 方法,有效提高模型的表达能力.Han 等^[17]则引入双向 SSM 模型,以增强局部几何特征与全局上下文的融合能力.

针对点云无序性,Wang 等^[18]利用八叉树结构,将点云递归地划分为 8 个子区域,在每个节点上提取局部特征,有效解决点云稀疏性问题,但树深度增加会显著影响分割效率.Zhang 等^[19]通过重新组合点云数据的 x、y、z 坐标顺序,生成 6 种序列化变体,并引入提示机制使模型感知当前点序列的排列规则,提高了模型的鲁棒性,但简单的坐标轴变换会破坏点云的局部性.Li 等^[20]提出多路径序列化策略,融合 Z 阶曲线和希尔伯特曲线对点云进行序列化,以增强空间信息保留能力.

由于上述方法仍未能同时解决点云无序性与空间编码机制的扩展性限制问题,为克服现有方法的局限性,本文通过互补有序建模和动态空间编码机制,有效解决三维点云语义分割中的无序性、几何信息缺失及边界模糊等问题.

2 本文算法

本文通过互补有序建模对点云进行高效建模,并采用动态空间编码与空间感知聚合机制对点云进行高效处理,其流程图如图 1 所示.首先构建互补有序建模模块,使用多种空间填充曲线对点云进行多视角序列化建模,在不破坏点云物理结构的前提下,构建 4 种空间结构保持性互补的编码序列;其次,针对不同阶段需求建立动态空间编码机制,以实现自适应空间感知的编码;最后,提出空间感知聚合,融合局部特征传播(Local Feature Propagation, LFP)与多层感知机(Multilayer Perception, MLP),构建高效特征混合机制,并引入空间感知池化替代传统 k 近邻边缘最大池化^[21],同时通过相邻坐标预测作为正则化项,强化特征稳定性与空间一致性.

2.1 互补有序建模

为解决三维点云的无序性与非结构化问题,本文设计了互补有序建模模块,其流程图如图 2 所示,借助空间填充曲线的空间遍历能力,在保留几何接近性的前提下,将原始点云序列化为结构化数据.空间填充曲线是将多维空间映射至一维空间的数学工具,代表性的曲线为阶曲线和希尔伯特曲线,其具有良好的空间连续性保留特性,在数学上可定义为双射函数 $\varphi = \mathbb{Z}^n \mapsto \mathbb{Z}$,其中 n 是空间维数,在三维点云中为 3.在点云语义分割中,不同曲线对几何邻近性的敏感性不同.Z 阶曲线具有计算简单、易于实现的优势,能够快速将点云映射到一维空间,但其局部邻近性保留能力较弱,尤其在点云分布不均匀时表现不佳.相比之下,希尔伯特曲线具有更强的局部邻近性保留能力,

能够更好地捕捉点云的几何结构,但其计算复杂度较高.现有方法多采用单一空间填充曲线进行点云序列化,无法有效结合

两种曲线的优势.本文构建互补有序建模模块,将两种曲线有机结合,在充分发挥各自优势的同时弥补各自的缺陷.

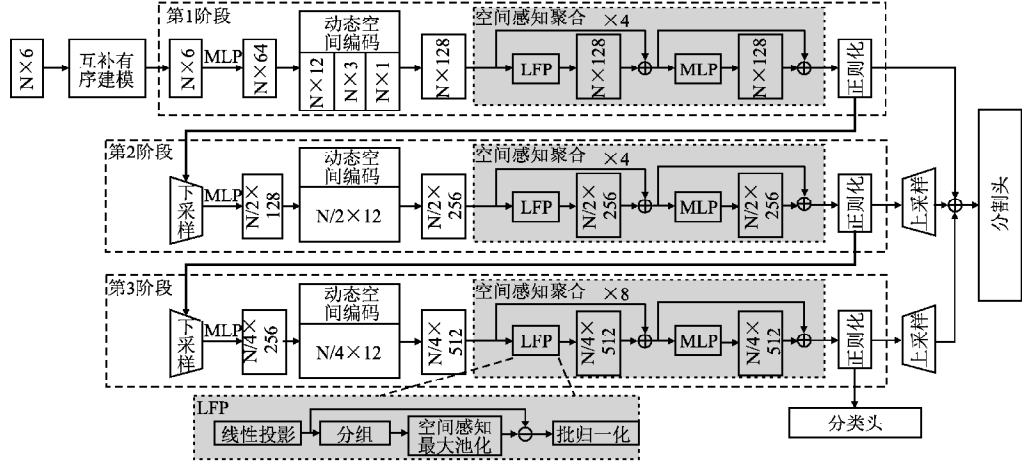


图1 所提方法框架

Fig. 1 Framework of the proposed method

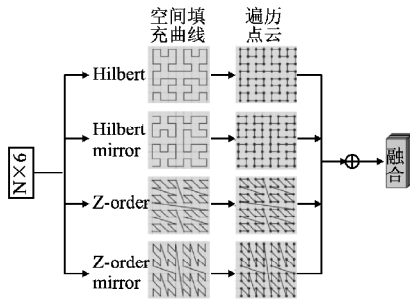


图2 互补有序建模模块

Fig. 2 Complementary ordering module

对于给定点云集合 $P = \{p_i \in \mathbb{R}^3\}_{i=1}^N$, 首先采用固定网格尺度 δ 构建三维整数网格坐标 g_i , 该操作可视为一种空间离散化, 将连续点云嵌入到统一的整数体素空间中. 每个点的栅格坐标通过以下方式计算:

$$g_i = (x_i, y_i, z_i) = \left\lfloor \frac{p_i - \min(P)}{\delta} \right\rfloor, g_i \in \mathbb{Z}^3 \quad (1)$$

其中 $g_i \in \mathbb{Z}^3$ 表示第 i 个点的三维栅格索引, $\min(P)$ 表示点云最小边界, $\lfloor \cdot \rfloor$ 表示逐元素向下取整操作.

然后采用空间填充曲线进行编码操作. Z 阶曲线编码通过按位交错排列的每个位构造一维序列, 公式如下:

$$Z(g_i) = \sum_{i=0}^{d-1} ((x_i \ll (3i+2)) | (y_i \ll (3i+1)) | (z_i \ll (3i))) \quad (2)$$

其中 x_i, y_i, z_i 为坐标的第 i 个二进制位, d 为编码深度, \ll 表示左移, $|$ 表示按位或.

Z 阶曲线通过编码为每个点赋予顺序信息. 为保证与原始点云的空间一致性, 本文不对点云进行重新物理排序, 仅记录序列化过程生成的映射, 因此需要对其进行解码, 将一维 Z 阶编码重新转回三维坐标. 解码过程使用逆向提取交错位, 最终恢复栅格坐标, 再逆离散化还原点坐标 $z(\hat{p}_i)$, 过程如下:

$$\begin{cases} x_i = \sum_{i=0}^{d-1} ((key \& (1 \ll (3i+2))) \gg (2i+2)) \\ y_i = \sum_{i=0}^{d-1} ((key \& (1 \ll (3i+1))) \gg (2i+1)) \\ z_i = \sum_{i=0}^{d-1} ((key \& (1 \ll (3i))) \gg (2i)) \end{cases} \quad (3)$$

$$z(\hat{p}_i) = \delta \cdot g_i + \min(P) \quad (4)$$

相比于 Z 阶编码, 希尔伯特编码通过递归构造的空间曲线具有更高的空间连续性和聚类性. 定义 d 为希尔伯特曲线的递归深度, 令 g_i 被映射为希尔伯特路径上的整数序列 $H(g_i) \in \mathbb{N}$, 其构造过程为:

$$H(g_i) = \sum_{i=0}^{d-1} h_k^{(i)} \cdot 2^{3k} \quad (5)$$

其中 $h_k^{(i)}$ 表示第层的希尔伯特子块索引, 每一层根据上层状态进行旋转和反射后确定子块索引. 希尔伯特曲线的解码需要恢复灰码序列, 再依次逆旋转变换每层子块, 最后通过逆离散化 $g_i = H^{-1}(H(g_i)) = (x_i, y_i, z_i)$ 还原点坐标 $h(\hat{p}_i)$. 通过结合两种曲线, 能够适应不同场景下点云的分布特性. 以上两种曲线通过分别沿着 x、y 和 z 轴进行顺序遍历处理三维空间. 通过改变遍历的顺序, 将 y 轴优先于 x 轴, 可以得到标准空间填充曲线的镜像变体, 生成镜像 Z 阶与镜像希尔伯特, 表示为:

$$\begin{cases} z'(\hat{p}_i) = z(\delta[:, [1, 0, 2]]) \\ h'(\hat{p}_i) = h(\delta[:, [1, 0, 2]]) \end{cases} \quad (6)$$

最后将不同曲线遍历生成的 4 种路径 $z(\hat{p}_i), h(\hat{p}_i), z'(\hat{p}_i), h'(\hat{p}_i)$ 进行融合. 虽然 4 条曲线之间存在信息冗余, 但它们在不同区域会提供互补信息, 因此本文引入全部 4 种曲线数据, 以 1:1:1:1 的权重进行融合. 通过多序列融合, 得到包含多样性点云结构特征 F 的新的点云数据 P , 为空间关系提供不同视角, 捕捉标准曲线可能忽略的特殊局部关系, 以提高语义分割的边界准确性.

2.2 动态空间编码

传统的相对空间编码采用固定维度的编码策略, 面对不同尺度和几何结构的点云数据时, 需要进行逐点卷积操作, 导致编码计算效率低且复杂, 为此, 本文引入动态空间编码机制, 根据不同阶段的任务需求动态调整编码维度. 对 3 个学习阶段 (第 1 阶段、第 2 阶段和第 3 阶段) 分别设置索引, 通过索引识别目前所处阶段, 并对应设置空间编码维度.

对于每一阶段, 输入点集为 $P = \{p_i \in \mathbb{R}^3\}_{i=1}^N$, 其特征为 $F = \{f_i \in \mathbb{R}^C\}$. 为提升计算效率并控制空间感受野, 在第 2、3

阶段首先对点集进行体素下采样,下采样主要由一层局部特征传播层和一层线性层构成.具体地,使用 KNN 构建每个点的 K 近邻索引 $N_{(i)}$,对原始特征执行局部特征传播操作进行特征增强,以提升对边界和密度变化的敏感性,再使用线性层将增强特征统一升维,使用体素网格尺度 δ 构建下采样点集 P^d 与特征 F^d .

对下采样后的点云执行动态空间编码,其核心为空间编码矩阵.先通过斐波那契球体采样在单位球面上生成 D 个方向向量 v_d ,形成球面均匀分布的方向初始集合.再为每个方向添加随机旋转矩阵 R_d ,每个方向向量配合可学习缩放参数 s_d 旋转轴 w_d 和旋转角 $\theta_d \in [0, \pi]$,构建变换矩阵 T_d ,并组合成最终的编码矩阵 M_d ,形式如下:

$$\begin{cases} T_d = (\sin\theta_d \cdot A + (1 - \cos\theta_d) \cdot A^2 + I) \cdot \text{diag}(s_d) \\ M_d = [T_1 \parallel v_1 \oplus \dots \oplus [T_d \parallel v_d]] \end{cases} \quad (7)$$

其中 A 为旋转轴的反对称向量, I 为单位阵, $\text{diag}(\cdot)$ 表示对角矩阵.动态空间编码根据当前所处阶段设置空间编码维度.第一阶段,由于输入较为复杂,包含基本的 12 维空间编码矩阵 M_d ,3 维颜色编码 RGB_i 与 1 维高度编码 h_i ,因此空间编码维度设置为 16.颜色编码与高度编码作为可学习参数,与空间编码矩阵一起,为点云数据提供更丰富的特征表示.第 2、3 阶段主要处理空间关系,不再需要编码原始特征和引入额外的编码信息,直接使用生成的空间编码矩阵,空间编码维度设置为 12,以节约计算开销.最终构造的编码输入特征形式为:

$$f_i^d = \begin{cases} [M_d, RGB_i, h_i] \in \mathbb{R}^{16}, d=0 \\ M_d \in \mathbb{R}^{12}, d \geq 1 \end{cases} \quad (8)$$

最后,采用 PointNet++ 的相对空间编码方法^[7]进行空间编码操作,得到最终的编码特征:

$$F = \text{encode}\{p_j - p_i, f_i^d\} \quad (9)$$

其中 p_i 为中心点坐标, p_j 为中心点的最近邻点坐标, $p_j - p_i$ 表示局部相对坐标.

2.3 空间感知聚合

受到 MetaFormer^[22] 提出的密集混合策略启发,本文提出空间感知聚合模块,高效学习空间编码后的特征.该模块中, LFP 作为轻量级令牌混合器捕捉局部几何结构, MLP 负责通道维度的信息融合,提供全局上下文感知.

LFP 中,首先通过线性投影层对输入特征进行特征变换,然后将特征分为 4 组,通过坐标和缩放参数处理每组特征,结合空间编码和 K 近邻信息进行空间感知池化,最后利用批归一化稳定特征分布.

由于传统 K 近邻边缘最大池化^[21] 操作仅计算邻域点与中心点的特征差值,这种池化操作忽略了空间线索,导致局部几何信息丢失,模型难以有效区分几何结构相似的区域(如平坦墙面与倾斜桌面).本文提出空间感知池化机制,根据点与中心点的空间距离调整池化权重,表示为:

$$\begin{cases} F_i^{\text{out}} = \text{BN}(\sum_{j \in N(i)} W_c(\Delta p_{ij}) \odot f_j) \\ W_c = f \times e^{-r^2} \end{cases} \quad (10)$$

其中 W_c 为空间自适应权重, r 为点与中心点的欧氏距离,靠近几何中心的点权重更高.通过空间感知池化机制,算法能捕捉更多点云的局部几何结构,提高语义分割的精度.

在空间感知聚合中, LFP 与 MLP 的比例为 2:1. 首先经

过两层 LFP,并引入残差连接,将原始特征与局部特征增强特征融合,避免过拟合现象,再使用一层 MLP 进行全局上下文感知.为弥补下采样造成的感受野损失,第 1、第 2 阶段的特征聚合深度为 4,第 3 阶段特征聚合深度为 8.为减少聚合过程中潜在的模糊性,在空间感知聚合后通过正则化项预测点与邻近点的相对坐标.具体地,在每个阶段结束时,为每个点随机选取一个邻近点 j ,使用相减操作构建特征差 Δf_{ij} 与真实坐标差 Δp_{ij} .再将特征差输送到由多层感知机构成的分类器中得到预测坐标偏移 $\Delta \hat{p}_{ij}$,定义正则项 \mathcal{L}_{reg} 为其与真实值的均方误差.最后对 3 个阶段取平均并与调节权重相乘,得到最终的特征差损失 $\mathcal{L}_{\text{total}}$:

$$\begin{cases} \Delta \hat{p}_{ij} = \text{MLP}(\Delta f_{ij}) \\ \mathcal{L}_{\text{reg}} = \frac{1}{N} \sum_{i=1}^N \|\Delta \hat{p}_{ij} - \Delta p_{ij}\|_2^2 \\ \mathcal{L}_{\text{total}} = \lambda \cdot \frac{1}{D} \sum_{d=1}^D \mathcal{L}_{\text{reg}} \end{cases} \quad (11)$$

3 实验及结果分析

3.1 实验环境及设置

本文实验环境基于 Ubuntu22.04 操作系统,计算机硬件为 NVIDIA GeForce RTX 4090 GPU,24 GB 内存,深度学习框架采用 PyTorch 2.0.1.为确保训练样本在不同尺度上的结构一致性与特征可比较性,本文在训练前对点云数据进行预处理,将点云平移至坐标原点并缩放 40 倍至单位球体内,同时将 RGB 颜色值归一化至 $[0, 1]$ 区间,以匹配网络输入要求并增强训练稳定性.优化器选择 Adam W,损失函数使用带标签平滑的交叉熵(平滑率 0.2),权重衰减为 0.05.学习率策略采用余弦退火衰减(衰减参数 0.05).归一化和激活函数分别设置为 Batch Norm 和 GELU.算法关键超参数设置如下:正则化强度 λ 为 0.003;3 个阶段的空间感知聚合模块深度分别为(4,4,8);3 个阶段的空间编码特征维度分别为(16,12,12).

3.2 数据集及评估指标

实验选用 3 个代表性数据集验证算法性能. S3DIS 数据集用于室内场景语义分割评估,包含 6 个大型室内区域(Areal-Area6),共 13 个语义类别,来源于地面激光扫描,点密度高,每个点包含 XYZ 坐标、RGB 颜色及所属房间的归一化坐标.本文设置最大点数为 30000,3 个阶段的体素网格尺度分别为(0.04,0.08,0.16),坐标标准差分别为(1.6,3.8,7.6),批次大小 8,初始学习率 0.006. ScanNet v2 是大规模室内场景数据集,包含 1513 个场景、21 个语义类别,由 RGB-D 视频采集.点云密度较低,本文采用更细的网格划分,设置最大点数为 80000,3 个阶段的体素网格尺度分别为(0.02,0.04,0.08),坐标标准差分别为(1.6,3.0,5.7),批次大小 8,初始学习率 0.006. ScanObjectNN 数据集用于点云对象分类任务,包含 2902 个独立对象实例,总样本 15000 左右,分为 15 个常见物体类别.提供多个难度级别的变体,用于评估模型在遮挡、旋转扰动和背景噪声等真实挑战下的表现,3 个阶段的点数划分为(1024,512,256),3 个阶段的坐标标准差分别为(2.2,4.4,8.8),批次大小 32,初始学习率 0.003.

本文采用点云语义分割与分类领域的标准评估指标进行定量评估:总体准确率(Overall Accuracy, OA)、平均准确率

(Mean Class Accuracy, mAcc)和平均交并比(Mean Intersection over Union, mIoU).这3个指标分别从整体样本准确性、各类别平均准确性以及分割区域重叠度等角度全面评估模型性能.

3.3 定量结果分析

3.3.1 基于 S3DIS 数据集的定量评估

基于 S3DIS 数据集的语义分割定量对比结果如表 1 所示,所对比方法的定量结果均源于原文献及其开源代码复现.实验选取 Areal、2、3、4、6 区域为训练集,训练 100 轮,Area5

作为测试集,以 mIoU 及 9 个主要类别的 IoU 作为主要评估指标,并展示 OA 与 mAcc 指标的对比结果.

本文方法以 75.6% 的 mIoU 领先于所有对比方法,较次优的 PCM 方法^[19]提升 1.5%.在 9 个主要类别中有 7 个类别取得最优结果,其中在“椅子”(+3.2%)、“沙发”(+1.4%)等小物体类别上提升尤为明显,说明模型具有较好泛化能力.

与结构建模能力强的 Point Transformer^[4]相比,本文方法在 mIoU 上提升 5.2%,主要归功于互补多曲线引导的有序序

表 1 S3DIS 数据集语义分割结果(%)

Table 1 Semantic segmentation results of S3DIS dataset(%)

方法	OA	mAcc	mIoU	柱子	窗户	门	桌子	椅子	沙发	书柜	黑板	杂物
PointTransformer ^[4]	90.8	76.5	70.4	38.0	63.4	74.3	89.1	82.4	74.3	80.2	76.0	59.3
PointNeXt ^[8]	91.0	77.2	71.1	37.7	59.3	74.0	83.1	91.6	77.4	77.2	78.8	60.6
StratifiedFormer ^[14]	91.5	78.1	72.0	46.1	60.0	76.8	92.6	84.5	77.8	75.2	78.1	64.0
PTv2 ^[23]	91.6	78.0	72.6	34.4	64.7	77.9	93.1	84.4	77.3	86.3	84.5	62.2
DeLA ^[24]	91.4	79.1	73.6	52.3	58.8	74.9	82.6	93.1	83.5	79.1	79.9	66.7
PTv3 ^[15]	91.7	79.0	73.6	55.8	63.7	77.1	83.8	93.3	79.1	79.4	85.4	61.7
PCM ^[19]	92.9	81.6	74.1	55.2	61.5	76.0	84.3	93.3	85.3	79.4	81.2	65.2
本文	91.8	79.3	75.6	56.5	62.8	80.8	83.9	93.4	86.7	82.4	87.4	66.9

注:粗体表示各列最优值,下划线表示次优值.

列化和动态空间编码机制,保持了空间邻域的结构稳定性,同时避免了在密集场景中注意力矩阵过度依赖带来的局部信息缺失问题.与 StratifiedFormer^[14]、PointNeXt^[8]和 DeLA^[24]等分层聚合方法相比,本文在多数类别上取得更优交并比,尤其在柱子、书柜和黑板等需要建模平面结构与上下文语义连续性的类别上优势明显,表明本文所引入的空间感知聚合模块在全局一致性建模方面具有较强适应性.

3.3.2 基于 ScanNet v2 数据集的定量评估

基于 ScanNet v2 数据集的语义分割定量对比结果如表 2 所示.实验选取包含 1201 个场景的 Scannet v2 train 作为训练集,训练 100 轮,包含 312 个场景的 Scannet v2 val 作为验证集.以 mIoU 和 mAcc 为评估指标.

表 2 ScanNet v2 数据集语义分割结果(%)

Table 2 Semantic segmentation results of ScanNet v2 dataset(%)

方法	Params	mIoU	mAcc
PTv2 ^[23]	12.8M	75.5	85.0
OctFormer ^[18]	44.0M	75.7	83.1
DeLA ^[24]	39.2M	76.2	83.4
DeepLA-Net ^[25]	61.2M	77.6	-
PTv3 ^[15]	46.2M	77.6	85.0
Pamba ^[20]	82.2M	77.6	-
CDSegNet ^[1]	101.4M	77.9	85.2
本文	66.6M	78.1	85.3

注:粗体表示各列最优值,下划线表示次优值.

表 2 中分别展示了各方法在验证集(Val)上的参数量、mIoU 和 mAcc 指标.本文方法在验证集上取得 78.1% 的 mIoU 和 85.3% 的 mAcc,在所有方法中取得最优结果.与采用传统分层全局上下文建模的 PTv2^[23]、OctFormer^[18]和 DeLA^[24]方法相比,本文的语义分割性能取得了极大的提升,说明所设计的动态空间编码和空间感知聚合机制能够更有效捕捉点云的结构关系.参数量对比中,本文的参数量适中.在与更复杂模型的对比中,本文方法在参数规模小于 CDSeg-

Net^[1](101.4M)和 Pamba^[20](82.2M)的情况下,依然在 mIoU 上分别领先 0.2% 和 0.5%,展现出良好的效率-精度平衡特性.虽然 DeepLA-Net^[25](61.2M)和 PTv3^[15](46.2M)在参数量方面更为轻量,但性能相比本文方法较差,并且由于所用硬件算力的提升,参数量的增加不会对本文模型效率产生过大的影响.上述对比验证了本文的互补有序建模在保持空间连续性,以及空间感知聚合对全局语义与邻域几何的协同表达的优势.

3.3.3 基于 ScanObjectNN 数据集的定量评估

为验证本文方法的通用性,在 ScanObjectNN 点云分类数据集上进行了评估,以 OA 为评估指标,结果如表 3 所示.实验选取数据集的 3 个标准测试变体:OBJ-ONLY 是基础版本,

表 3 ScanObjectNN 数据集分类结果(%)

Table 3 Classification results of ScanObjectNN dataset(%)

方法	OBJ-BG	OBJ-ONLY	PB_T50_RS
PointNet + + ^[7]	82.3	84.3	77.9
PointNeXt ^[8]	92.5	91.7	87.7
PCM ^[19]	-	-	88.4
Point Mamba ^[16]	94.3	92.6	89.3
DeLA ^[24]	92.8	92.0	90.4
DeepLA-Net ^[25]	-	-	90.6
MAMBA3D ^[17]	92.9	92.1	91.8
本文	94.4	92.4	92.1

注:粗体表示各列最优值,下划线表示次优值.

仅包含原始目标点云,无背景点和旋转扰动;OBJ-BG 是在 OBJ-ONLY 基础上加入真实场景的背景点,引入非目标干扰点;PB_T50_RS 则是最复杂的变体,融合旋转扰动、点云遮挡和背景点干扰,最贴近实际应用场景.本文在 PB_T50_RS 进行训练,训练 400 轮,在 3 个测试变体分别进行测试.

本文方法在 OBJ-BG 和 PB_T50_RS 中分别取得 94.4% 和 92.1% 的最优性能,在 OBJ-ONLY 变体中取得 92.4%,略低于 Point Mamba^[16]的 92.6%,但在无额外预训练、无外部监督的条件下依然取得了高精度,说明本文方法在真实点云

中的可迁移性与鲁棒性.尤其在 PB_T50_RS 变体上,本文方法优于所有对比方法,验证了其在稀疏、扰动、多姿态环境下的点云识别能力.

3.4 定性结果分析

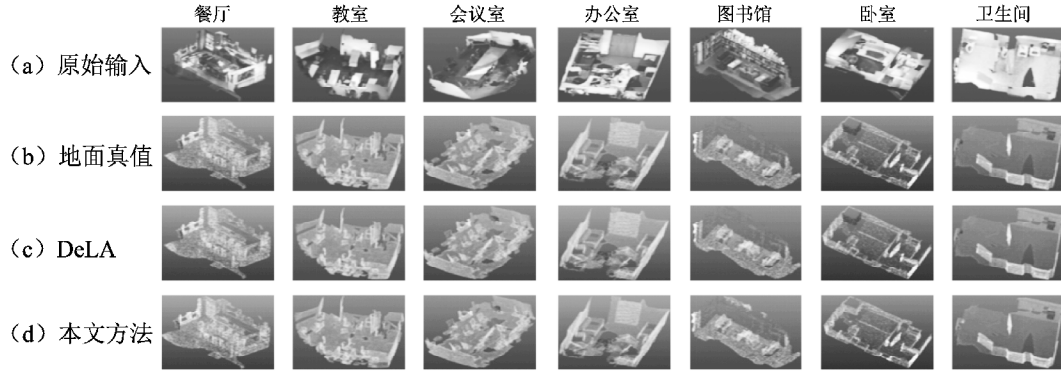


图3 ScanNet v2 数据集语义分割结果可视化

Fig. 3 Visualization of semantic segmentation results from ScanNet v2 dataset

卫生间)进行语义分割结果的可视化对比,每个可视化结果包含4行:原始输入RGB图像、地面真值、DeLA^[24]方法以及本文方法预测结果.

与DeLA^[24]方法相比,本文方法在多个场景中都表现出更为清晰且连续的语义边界,尤其在“门-墙”、“书柜-墙”和“椅子-地板”等类别邻近的区域,本文展现出更清晰的边界判断能力.例如,在会议室和教室等场景中,DeLA方法^[24]对桌子与地板边缘存在明显模糊,而本文方法能准确定位边界,保持语义一致性.

此外,本文方法对“垃圾桶”、“台灯”、“书架角”和“洗手池”等小型物体类别具有更高的识别精度和完整性,体现出对低密度或遮挡物体的良好感知能力.这得益于动态空间编码在不同阶段对局部空间结构的适配性增强,以及空间感知聚合模块对上下文的多尺度融合,使得模型不仅在墙体、地板等大尺度结构中准确划分,也在细节区域保持更高的语义分辨率.

3.5 消融实验

为验证本文方法各个模块的有效性,对互补有序建模模块、空间编码模块、特征聚合模块以及正则化模块进行系统分析,所有消融实验均在S3DIS数据集上进行,以mIoU为主要评估指标.

3.5.1 互补有序建模有效性分析

表4列出了互补有序建模模块中不同序列化组合的语义分割结果.将序列化组合划分为5种:Z阶、希尔伯特、Z阶+

表4 互补有序建模模块消融结果

Table 4 Complementary ordered modeling module ablation results

序列化组合	mIoU
Z阶	74.6
希尔伯特	75.0
Z阶+镜像Z阶	74.8
希尔伯特+镜像希尔伯特	75.1
Z阶+镜像Z阶+希尔伯特+镜像希尔伯特	75.6

镜像Z阶、希尔伯特+镜像希尔伯特、Z阶+镜像Z阶+希尔伯特+镜像希尔伯特.结果表明,单独采用Z阶或希尔伯特的序列化构造已能带来一定的性能提升,并且基于希尔伯特

为验证本文算法在结构建模与语义识别方面的有效性,本文在ScanNet v2数据集上进行定性结果分析.图3展示了ScanNet v2数据集典型场景语义分割可视化结果.本文选取7个典型室内场景(餐厅、教室、会议室、办公室、图书馆、卧室、卫生间)

的模式优于基于Z阶的模式,这验证了希尔伯特曲线具有更好的空间连续性.在引入镜像曲线后,模型性能进一步提升,这证明不同的序列化模式为点云中的空间关系提供了不同的视角,互补有序建模模块为后续空间建模奠定了更稳定的结构基础.

3.5.2 空间编码有效性分析

表5列出不同空间编码模式下的语义分割结果.实验结果显示,动态空间编码结合空间感知池化能达到最优性能(75.6%),较传统相对空间编码与k近邻边缘最大池化方法^[21]提升1.3%.通过分析可见,从相对空间编码到动态空间编码提升了0.7%,而引入空间感知池化替代k近邻边缘

表5 空间编码模块消融结果

Table 5 Spatial encoding module ablation results

空间编码模式	mIoU
相对空间编码+K近邻边缘最大池化 ^[21]	74.3
相对空间编码+空间感知池化	74.4
动态空间编码+K近邻边缘最大池化 ^[21]	75.0
动态空间编码+空间感知池化	75.6

最大池化^[21]则贡献了0.6%的增益.这一结果充分证明了本文提出的动态空间编码模块能够根据不同阶段需求自适应调整编码维度,而空间感知池化则通过考量点的空间距离关系,有效增强了对局部几何结构的建模能力.

3.5.3 不同深度下特征聚合有效性分析

表6给出了特征聚合模块在不同深度设置下的语义分割

表6 特征聚合深度消融结果

Table 6 Feature aggregation depths ablation results

深度设置	总层数	mIoU	Params
[1,1,2]	4	72.6	27.3M
[2,2,4]	8	74.1	40.7M
[4,4,8]	16	75.6	66.6M
[8,8,16]	32	75.1	90.7M

性能及参数量.实验结果表明,当深度设置为[4,4,8]时,效果达到最优;当深度设置为[1,1,2]和[2,2,4]时,虽然参数量显著减少,但由于感受野不足,无法有效识别复杂的空间结构.而过深的网络结构[8,8,16]不仅参数量大幅增加,还出

现梯度消失和过拟合问题,反而导致性能下降.这表明合理的层级设计对于平衡模型表达能力和训练难度至关重要,因此本文最终采用[4,4,8]的深度配置,既保证了充分的特征提取能力,又避免了过度复杂化.

3.5.4 正则化强度有效性分析

表7列出了不同正则化强度对语义分割性能的影响.无正则项时,模型 mIoU 仅为 72.8%.随着正则化强度 λ 的增强,性能逐步提升,在 $\lambda = 0.001 \sim 0.003$ 区间达到峰值(75.6%),相比无正则化基线提升 2.8%.这表明适当强度的结构约束有助于模型学习更加平滑且具空间一致性的特征表示.当 λ 过小时,正则化效果不足;而当 λ 过大(如 0.01 和 1)时,则会抑制特征表达能力,造成性能下降.因此,本文最终选择 $\lambda = 0.003$ 作为正则化强度.

表7 正则化强度消融结果

Table 7 Regularization strength ablation results

正则化	0	1e-4	3e-4	1e-3	3e-3	1e-2	1
mIoU	72.8	73.5	75.0	75.6	75.6	74.4	72.8

上述一系列消融实验全面验证了本文所提出的各个关键模块的有效性,证明了结合有序建模和空间感知的点云语义分割算法在多个维度上的优化设计是合理且必要的.

4 结束语

本文提出了一种结合有序建模和空间感知的点云语义分割算法.该算法通过互补有序建模模块对点云进行高效序列化建模,采用动态空间编码机制根据不同阶段需求自适应调整特征表示,并引入空间感知聚合模块协同建模局部几何与全局语义信息.在多个公开基准数据集上的实验结果表明,本文方法在保持点云空间拓扑稳定性的同时,有效提升了语义分割的精度和鲁棒性,特别是在复杂边界区域和小物体分割方面表现出显著优势.

但本文方法仅在室内静态扫描点云数据集上进行训练与验证,在户外、移动平台等动态或低密度场景中的适应性有待进一步验证;其次,正则项仅采用基于单邻域点的偏移监督,空间关系建模仍较为浅层.在后续研究中,将考虑扩展算法至更多样化的数据场景,探索更有效的结构化正则化方法,优化算法计算效率,提升其通用性和适应性.

References:

[1] Qu W, Wang J, Gong Y, et al. An end-to-end robust point cloud semantic segmentation network with single-step conditional diffusion models[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025:27325-27335.

[2] Zhang J, Lin X, et al. SVM-based classification of segmented airborne LiDAR point clouds in urban areas[J]. Remote Sensing, 2013, 5(8):3749-3775.

[3] Qi C R, Su H, Mo K, et al. PointNet: deep learning on point sets for 3d classification and segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017:652-660.

[4] Zhao H, Jiang L, Jia J, et al. Point transformer[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021:16259-16268.

[5] Peano G. Sur une courbe, qui remplit toute une aireplane[M]. Vi-

enna: Springer Vienna, 1990:71-75.

- [6] Chen X, Wu Y, Xu W, et al. Pointscnet: point cloud structure and correlation learning based on space filling curve-guided sampling[J]. Symmetry, 2021, 14(1):8, doi:10.3390/sym14010008.
- [7] Qi C R, Yi L, Su H, et al. Pointnet++: deep hierarchical feature learning on point sets in a metric space[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems, 2017:5105-5114.
- [8] Qian G, Li Y, Peng H, et al. Pointnext: revisiting pointnet++ with improved training and scaling strategies[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022:23192-23204.
- [9] Chen X, Ma H, Wan J, et al. Multi-view 3D object detection network for autonomous driving[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017:1907-1915.
- [10] Xu C, Wu B, Wang Z, et al. Squeezesegv3: spatially-adaptive convolution for efficient point-cloud segmentation[C]//Proceedings of the European Conference on Computer Vision, 2020:1-19.
- [11] Wu Z, Song S, Khosla A, et al. 3D shapenets: a deep representation for volumetric shapes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2015:1912-1920.
- [12] Chen Y, Liu J, Zhang X, et al. Voxelnext: fully sparse voxelnet for 3d object detection and tracking[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023:21674-21683.
- [13] LU C J, PAN X. Multiscale dynamic graph CNN with edge supervision loss for point cloud segmentation[J]. Journal of Chinese Computer Systems, 2024, 45(9):2215-2220.
- [14] Lai X, Liu J, Jiang L, et al. Stratified transformer for 3D point cloud segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:8500-8509.
- [15] Wu X, Jiang L, Wang P S, et al. Point transformer v3: simpler faster stronger[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024:4840-4851.
- [16] Liang D, Zhou X, Xu W, et al. Pointmamba: a simple state space model for point cloud analysis[C]//Proceedings of the 38th International Conference on Neural Information Processing Systems, 2024:32653-32677.
- [17] Han X, Tang Y, Wang Z, et al. Mamba3d: enhancing local features for 3D point cloud analysis via state space model[C]//Proceedings of the 32nd ACM International Conference on Multimedia, 2024:4995-5004.
- [18] Wang P S. OctFormer: octree-based transformers for 3D point clouds[J]. ACM Transactions on Graphics, 2023, 42(4):1-11.
- [19] Zhang T, Yuan H, Qi L, et al. Point cloud mamba: point cloud learning via state space model[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025:10121-10130.
- [20] Li Z, Ai Y, Lu J, et al. Pamba: enhancing global interaction in point clouds via state space model[C]//Proceedings of the AAAI Conference on Artificial Intelligence, 2025:5092-5100.
- [21] Li G, Müller M, Thabet A, et al. DeepGCNs: can GCNs go as deep as CNNs? [C]//Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019:9267-9276.
- [22] Yu W, Luo M, Zhou P, et al. Metaformer is actually what you need for vision[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:10819-10829.
- [23] Wu X, Lao Y, Jiang L, et al. Point transformer v2: grouped vector attention and partition-based pooling[C]//Proceedings of the 36th International Conference on Neural Information Processing Systems, 2022:33330-33342.
- [24] Chen B, Xia Y, Zang Y, et al. Decoupled local aggregation for point cloud learning[EB/OL]. <https://arxiv.org/pdf/2308.16532>, 2023.
- [25] Zeng Z, Dong M, Zhou J, et al. Deepla-Net: very deep local aggregation networks for point cloud analysis[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2025:1330-1341.

附中文参考文献:

- [13] 卢超杰, 潘翔. 边缘损失约束下的多尺度动态图卷积点云分割网络[J]. 小型微型计算机系统, 2024, 45(9):2215-2220.