LAE-Transformer：融合局部感知增强的机载LiDAR点云分割网络

库闵凡[1]，张良[1,2,3]，邓继伟[2,3]，王广帅[2,3]，王新文[1]，刘恒志[1]

1. 湖北大学 资源环境学院, 武汉 430062;
2. 天津市轨道交通导航定位及时空大数据技术重点实验室, 天津 300251;
3. 中国铁路设计集团有限公司, 天津 300251

摘 要：针对现有机载LiDAR点云深度学习分割算法难以充分利用地物局部拓扑信息，不同尺度的特征表达能力较弱，小尺度标分割效果不够理想的问题，本研究提出了一种融合局部感知增强的机载LiDAR点云分割网络LAE-Transformer（Local-Aware Enhanced Transformer）。首先，通过提取浅层拓扑特征构建点云局部几何结构图，增强模型对地物细节的捕捉能力；随后，串连下采样与区域点联合Transformer模块提取点云深层特征，增强模型在多尺度下的特征感知能力；最后，在上采样过程引入动态残差连接，自适应地融合不同感受野下的关键信息；此外，构建了基于注意力池化与最大池化的混合池化层，弥补信息损失的问题。通过DALES点云数据集和LASDU点云数据集的测试本文提出结果表明，本文网络的总体精度（OA）与平均交并比（mIoU）分别达到97.8%与80.8%以及87.2%与68.5%，其中DALES数据集中的卡车、电线杆和围栏等小尺度目标的交并比（IoU）则分别达到了42.1%、75.4%和63.8%，优于其他大部分主流网络，验证了本文网络在机载LiDAR点云分割中的可靠性。

关键词：机载LiDAR点云，深度学习，语义分割，自注意力机制，局部感知增强，复杂场景，小尺度目标，Transformer

中图分类号：P237/P2

引用格式：**库闵凡，张良，邓继伟，王广帅，王新文，刘恒志.2025.LAE-Transformer：融合局部感知增强的机载LiDAR点云分割网络.遥感学报，29(11)：3363-3377**

Ku M F，Zhang L，Deng J W，Wang G S，Wang X W and Liu H Z. 2025. LAE-Transformer：An airborne LiDAR point cloud segmentation network incorporating local-aware enhancement. National Remote Sensing Bulletin，29(11)：3363-3377[DOI：10.11834/jrs.20254478]

1 引 言

机载激光雷达（Airborne LiDAR）整合了激光测距、全球定位系统GNSS（Global Navigation Satellite System）和惯性导航系统INS（Inertial Navigation System）技术，能够快速采集室外场景点云数据（Stilla和Jutzi，2008）。机载LiDAR点云密度大、精度高、覆盖区域广，目前已广泛应用于自动驾驶（Cai等，2023）、智慧城市（Xiong等，2024）及林业（Zhang等，2024b）等众多领域。对点云进行语义分割，赋予数据不同的类别标签，实现更精确的环境感知，是开展进一步研究与应用的关键前提（杨必胜 等，2021）。然而，由于LiDAR点云的数据量大且分布不均、地理环境复杂等原因，实现精确的语义分割具有一定挑战性。

早期的点云分割（卢维欣 等，2015）大部分通过手动提取特征完成，依赖人工先验知识（Weinmann等，2015），难以表征点云的深层次语义信息（Gevaert等，2016），且在复杂场景下的泛化能力不足，具有较大的局限性。随着深度学习

收稿日期：2024-10-25；**预印本**：2025-04-17

基金项目：天津市轨道交通导航定位及时空大数据技术重点实验室开放课题基金（编号：TKL2023A13，TKL2023B15）；中国国家铁路集团有限公司科技研究开发计划项目（编号：L2023G016）；国家自然科学基金（编号：41601504）

第一作者简介：库闵凡，研究方向为机载LiDAR点云数据处理和点云分割。E-mail：kuminfan@qq.com

通信作者简介：张良，研究方向为机器学习、遥感影像智能分类和三维点云处理。E-mail：zhangliang@hubu.edu.cn

在计算机视觉领域的不断深化，学者们开始将深度学习方法应用于三维点云处理中。由于点云数据具有非结构化和离散性等特点，经过特定处理才能输入到深度神经网络中，根据处理方式的不同，点云深度学习方法可分为基于体素的方法、基于多视图的方法和基于点的方法（Zhang等，2023a）。

基于体素的方法通过将三维空间划分为均匀体素，实现点云与体素的映射，从而将不规则点云结构化，便于深度学习网络处理。Maturana和Scherer（2015）提出的VoxNet结合体积占有率网格和3D卷积神经网络（CNN）实现点云分割，克服了大规模点云处理的难题；Zhou和Tuzel（2018）的VoxelNet构建了以体素特征为单元的编码层，聚合局部结构与全局形状特征；PointGrid网络（Le和Duan，2018）混合了体素与点的三维表达形式，以更好地表达体素尺度变化与局部细节。基于多视图的方法借鉴深度学习图像处理技术，通过将三维点云投影到不同视角的二维平面，整合多视角特征以捕捉点云全局特征。Su等（2015）提出的MVCNN利用CNN从多视角提取特征，通过最大池化聚合各视角特征，完成点云的分类与分割；Hamdi等（2021）提出的MVTN通过自适应学习相关投影视角信息，以获取更准确的几何特征。尽管基于体素和基于多视图的方法解决了点云数据难以直接应用于深度学习模型的问题，但体素化和视图转换的过程可能会破坏点云的原始结构，导致几何信息丢失，限制模型效果。

基于点的方法直接输入原始点云数据，通过多层训练实现语义分割，能够充分利用点云的原始特征，避免数据细节损失。2017年，Charles等（2017）提出的PointNet模型通过多层感知机MLP（Multilayer Perceptron）提取高维特征，并利用最大池化聚合全局特征，广泛应用于点云分类与分割。然而，PointNet仅处理单个点云，未考虑局部邻域特征，限制了其复杂场景下的分割效果。为此，Qi等（2017）提出了PointNet++，通过局部区域划分逐一提取特征，提升了分割效果；此后，Ye等（2018）提出3P-RNN，通过金字塔池化模型和双向循环神经网络捕捉点间全局关系；Wang等（2019）提出了动态图卷积神经网络模型DGCNN，将点与邻域点之间的依赖关系视为图像结构，提取更深层次的拓扑信息；Landrieu和Simonovsky（2018）通过聚类点云局部特征划分超点，减小了模型训练内存开销；Thomas等（2019）提出的KPConv设计灵活可变形卷积核，以调整点云特征图形状，提取局部特征；Hu等（2020）提出的RandLA-Net利用随机点采样和局部特征聚合模块有效处理大规模点云数据的复杂性；Ma等（2022）通过堆叠残差MLP层学习点云特征，提出轻量级网络架构PointMLP；一些研究通过引入多尺度架构（Jiang等，2018），设置不同的邻域大小（赵传 等，2020），以学习更丰富的几何特征（Li等，2022）。上述模型大多直接采用多层感知机和卷积神经网络等方式捕捉特征，在点云局部特征提取方面展现出了一定的效果。然而，受点云数据稀疏性等因素影响，最大池化操作和卷积层在特征升降维过程中可能会遗漏重要的局部几何细节，进而限制复杂场景分割结果的准确性。

近年来，Transformer在自然语言处理（NLP）领域取得显著成果（Vaswani等，2017）。考虑到点云数据的序列化和非结构化等特性，这与Transformer擅长处理序列建模任务的特点高度契合，因此一些研究开始将Transformer及其自注意力机制扩展到点云深度学习领域。例如，Point Cloud Transformer（Guo等，2021）采用隐式拉普拉斯算子和改进的自注意力机制，以适应无序点云数据；Point Transformer（Zhao等，2021）将自注意力机制应用于每个采样点，从而映射到整块点云中；Wang（2023）提出了一种基于八叉树的Transformer结构，实现更高效的编码速率；Wei等（2023）提出的MGT将数据划分为大小不一的切片，提高了捕获多尺度特征的能力；Robert等（2023）结合自适应超点聚类与Transformer架构，提出了一种高效的点云处理方法SuperPoint Transformer；Zhang等（2024a）提出的TCFAP-Net引入跨特征融合与自适应感知的方法，增强了模型对复杂点云特征的理解与感知。然而，在复杂机载点云场景中，由于地物之间尺度不一，Point Transformer（Zhao等，2021）等主流模型的特征提取方法难以有效区分地物边缘特征，对不同尺度地物之间的差异性特征表达能力较弱，限制了其在机载点云语义分割中的准确性，具体表现为对如道路、建筑物等局部同质性较强的大尺度目标分割效果较好，但对如卡车、围栏等尺度较小、特征表达较弱的小尺度目标分割效果有限。

综上，本文提出了一种融合局部感知增强的机载LiDAR点云分割网络LAE-Transformer（Local-Aware Enhanced Transformer），旨在解决机载点云场景中小尺度目标地物难以有效区分的问题，增强模型的分割性能。首先，为了增强模型对地物细节与拓扑信息的捕捉能力，设计一个局部拓扑信息增强模块LTIE（Local Topological Information Enhancement），充分利用点的拓扑特征 和局部上下文信息，构建局部几何结构图，获取细粒度更为丰富的特征编码；随后，为了增强对不同尺度目标的特征感知能力，通过串联多个下采样层与区域点联合Transformer模块RPT（Regional Point Transformer）逐层提取点云的深层次特征并扩充网络模型的感受野；最后，在上采样过程中引入动态残差连接模块DRC（Dynamic Residual Connection），自适应地融合不同感受野下的关键信息，增强不同阶段语义特征间的关联性；此外，为了解决最大池化导致局部细节缺失的问题，构建基于注意力池化与最大池化的混合池化方法，该方法用于网络处理过程中全局与局部信息的聚合，聚焦关键信息的同时减少其他有效信息的损失。通过DALES机载点云数据集（Varney等，2020）和LASDU机载点云数据集（Ye等，2020）开展实验与分析，以期提升复杂场景的总体分割与小尺度目标的提取精度。

2 研究方法

2.1 网络总体架构

如图1所示，本文网络整体上采用了编码器—解码器与跳跃连接的U型结构。首先，将尺寸为（N，d）的原始点云集合输入网络（其中N为点个数，d为点的特征维度），通过LTIE提取点的拓扑特征和局部上下文信息，并将其作为编码器的输入数据。编码器—解码器由MLP、上采样与下采样、RPT和DRC组成。在编码器阶段，输入的点云浅层特征经由MLP映射得到高维特征后，再输入RPT中学习以建立局部区域内与区域之间的自注意力关系；随后通过串连下采样与RPT的方式稀疏点云数量，扩充模型感受野，提取不同分辨率下的特征。下采样过程中，每层的点数量按照1/4的比例逐层递减，特征通道数则以2倍的比例逐层递增，最终经过4层下采样操作，得到维度为$\left(\frac{N}{256},512\right)$的特征张量，作为解码器的输入数据。在解码器阶段，通过跳跃连接与DRC模块动态融合编码层与解码层之间的信息，获取不同尺度下更丰富的上下文特征；同时，串连上采样与RPT，在逐层恢复点云特征维度的过程中进一步提取采样点的深层次特征。上采样过程中，点数量按照4倍的比例逐层递增，特征通道数则以1/2的比例逐层递减，最后，通过4层上采样操作与全连接FC（Fully Connected）层输出得到维度为（N，C）的特征张量（C为场景点云中的类别个数），即网络模型对点云语义分割的预测结果。

2.2 局部拓扑信息增强模块

机载LiDAR点云蕴含丰富的拓扑信息，提取点云的局部几何结构特征并体现场景中不同地物的显著差异，是实现点云分割的关键前提。现有方法大多直接使用点的三维坐标、RGB和强度等信息作为特征输入，忽略了点云局部区域的拓扑特征，限制了模型对点云几何边缘信息的判别能力，进而影响小尺度目标的分割性能。本文提出的LTIE通过计算点云的局部拓扑特征，并整合局部邻域内不同点间的特征差异，以获取细粒度更为丰富的特征编码，提高对点云区域间几何结构关系的判别能力。

如图2所示，该模块首先进行浅层拓扑特征的提取，利用K-最近邻算法（KNN）搜索采样点的邻域，确定该区域内的K个最邻近点，并根据这些最邻近点的三维坐标集合构建协方差矩阵，利用主成分分析（PCA）计算得到归一化特征值λ_1、λ_2、λ_3（其中$\lambda_1 \geqslant \lambda_2 \geqslant \lambda_3$，$\lambda_1+\lambda_2+\lambda_3=1$）。随后基于表1所列公式，进一步计算并筛选得到线性度（L_λ）、平面度（P_λ）、散射度（S_λ）、全方差（O_λ）、各向异性（A_λ）、曲率变化度（C_λ）和特征熵（E_λ）等浅层拓扑特征。

上述所提基于局部区域的浅层拓扑特征是对点云原始输入特征的有效补充，然而，由于部分地物类别之间的相似度较高，简单的浅层拓扑特征难以帮助模型实现不同地物类别间的准确区分。因此，为了更充分的探索局部区域的几何结构信息，文中基于输入点$\boldsymbol{P}=\{p_i\}_{i=1}^{N}$与对应的特征张量集合$\boldsymbol{F}=\{f_i\}_{i=1}^{N}$（其中$N$为采样点个数），设计了一种基于上下文的特征增强方法。首先，利用

KNN对每个采样点进行分组聚合，构建以采样点p_i为中心的局部邻域点集合$\boldsymbol{P}_i'=\{p_{i,j}\}_{j=1}^{K}$与特征集合$\boldsymbol{F}_i'=\{f_{i,j}\}_{j=1}^{K}$。在此基础上，通过提取采样区域的边缘特征$\boldsymbol{E}_i'=\{e_{i,j}\}_{j=1}^{K}$（其中$e_{i,j}=f_{i,j}-f_i$），并结合邻域内的采样点特征构建局部邻接图$\boldsymbol{G}_i$，捕获得到细粒度更为丰富的特征编码。局部邻接图$\boldsymbol{G}_i$可以表示为

$$\boldsymbol{G}_i=\mathrm{concat}\left(f_i\times K,\boldsymbol{E}_i'\right) \tag{1}$$

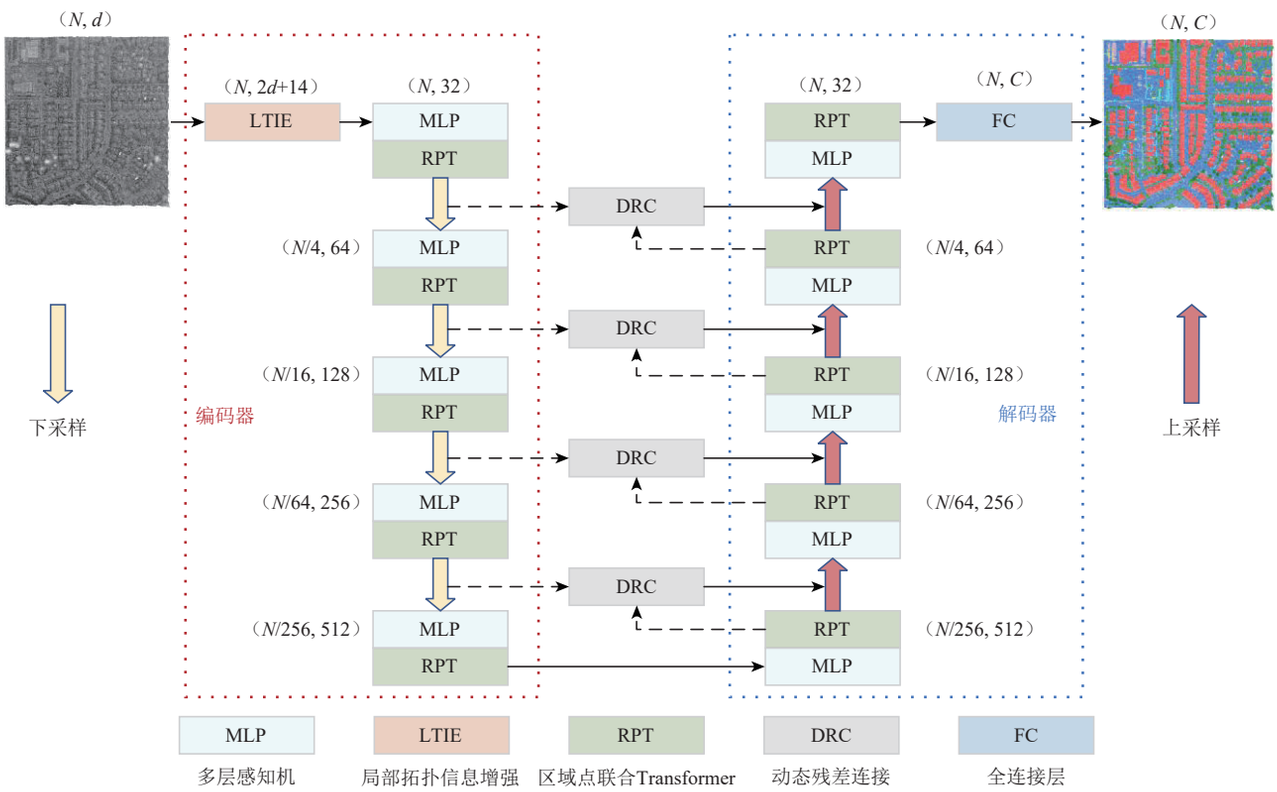


图1 LAE-Transformer总体架构

Fig. 1 The overall framework of LAE-Transformer

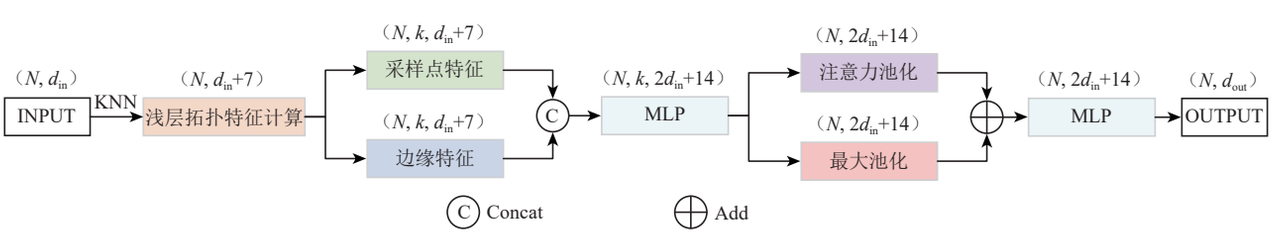


图2 局部拓扑信息增强模块

Fig. 2 Module of local topological information enhancement

表1 浅层拓扑特征计算公式

Table 1 Calculation formula for surface topological features

	拓扑特征						
	线性度	平面度	散射度	全方差	各向异性	曲率变化度	特征熵
计算公式	$\boldsymbol{L}_\lambda=\frac{\lambda_1-\lambda_2}{\lambda_1}$	$\boldsymbol{P}_\lambda=\frac{\lambda_2-\lambda_3}{\lambda_1}$	$\boldsymbol{S}_\lambda=\frac{\lambda_3}{\lambda_1}$	$\boldsymbol{O}_\lambda=\sqrt[3]{\lambda_1\lambda_2\lambda_3}$	$\boldsymbol{A}_\lambda=\frac{\lambda_1-\lambda_3}{\lambda_1}$	$\boldsymbol{C}_\lambda=\frac{\lambda_3}{\lambda_1+\lambda_2+\lambda_3}$	$\boldsymbol{E}_\lambda=-\sum_{i=1}^{3}\lambda_i\ln(\lambda_i)$

随后，引入由最大池化和注意力池化构成的混合池化层，对局部几何结构特征进行进一步聚合。最大池化保留显著特征，抑制不重要的信息和噪声；注意力池化利用注意力机制动态调整特

征聚合权重，增强关键特征的敏感性，同时补偿最大池化可能造成的空间信息损失。其具体形式如下：

$$m_i = \mathrm{Maxpooling}\left(\boldsymbol{G}_i\right) \tag{2}$$

$$t_i = \sum_{i=1}^{K}\left(\boldsymbol{G}_i \odot g\left(\boldsymbol{G}_i, W\right)\right) \tag{3}$$

式中，m_i为采用最大池化操作得到的特征；t_i为采用注意力池化操作得到的特征；$\odot$为点积相乘；W为局部邻接图结构中每个特征张量的权重系数；g则为K个近邻点学习的注意力得分。

最终，通过聚合压缩最大池化特征和注意力池化特征，生成综合局部拓扑结构的新的增强特征，并作为模块的输出。

2.3 区域点联合Transformer模块

Point Transformer （Zhao等，2021）将注意力机制层应用于局部区域内的每个点，虽然有效解决了Transformer难以直接应用到大规模点云数据的问题，但其关注重心主要集中在局部区域的单点特征，忽略了局部区域内的上下文关系，从而限制了模型对不同尺度地物特征的感知能力。为此，本文基于局部区域编码单元与向量自注意力单元构建了区域点联合Transformer模块RPT，如图3所示。局部区域编码单元能够帮助模型捕捉局部区域间的上下文关系；而向量自注意力单元则帮助模型建立点间的特征关联。通过堆叠这两个基本单元，并结合用于恢复输入特征维度的上采样插值操作，形成了一个由点自注意力层与局部区域自注意力层组成的并联结构。这一结构不仅能够提取局部区域的关键信息，还能有效构建局部区域间的自注意力关系，从而增强模型对不同尺度地物点的语义判别能力。

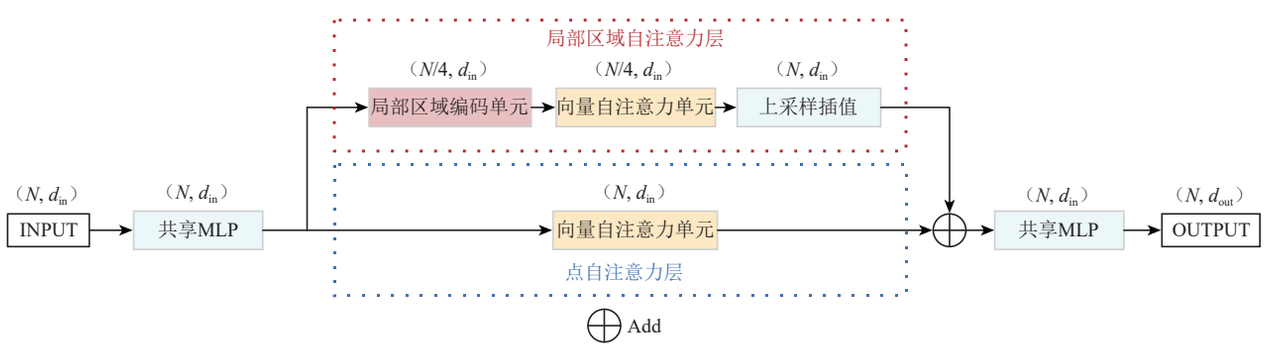


图3 区域点联合Transformer模块

Fig. 3 Module of regional point Transformer

2.3.1 局部区域编码单元

为了有效组织起局部区域之间的关联，本文通过局部区域编码单元对点云进行局部区域结构的划分。本文假设每个局部区域包含了一个采样点及其K个邻近点，并以聚合采样点与这K个邻近点特征的方式来代表该局部区域。如图4所示，首先，对于输入点$P = \left\{p_i\right\}_{i=1}^{N}$，使用最远点采样法（FPS）进行采样点的选取，得到采样点的点集合$\tilde{\boldsymbol{P}} = \left\{\tilde{p}_i\right\}_{i=1}^{N/4}$与特征集合$\tilde{\boldsymbol{F}} = \left\{\tilde{f}_i\right\}_{i=1}^{N/4}$，并利用KNN得到以采样点$\tilde{p}_i$为中心的局部邻域点集合$\tilde{\boldsymbol{P}}_i' = \left\{\tilde{p}_{i,j}\right\}_{j=1}^{K}$与特征集合$\tilde{\boldsymbol{F}}_i' = \left\{\tilde{f}_{i,j}\right\}_{j=1}^{K}$；同时，考虑到位置信息对于局部区域关联的重要性，使用一个相对位置编码层来增强点云空间位置信息的表达，具体公式如下：

$$\tilde{l}_{i,j} = \mathrm{concat}\left(\tilde{f}_{i,j}, \mathrm{mlp}\left(\tilde{p}_i - \tilde{p}_{i,j}\right)\right) \tag{4}$$

式中，$(\tilde{p}_i - \tilde{p}_{i,j})$代表采样点与其局部邻域点的相对坐标，依此类推得到聚合了相对位置编码的特征集合$\tilde{\boldsymbol{L}}_i' = \left\{\tilde{l}_{i,j}\right\}_{j=1}^{K}$，随后，通过一个混合池化层聚合局部特征，输出一个新的特征张量，以表示每个采样点对应的局部区域；最后，利用一个MLP层对齐特征维度，输入到后续的向量自注意力层中。

2.3.2 向量自注意力单元

针对点云的注意力机制研究（Feng等，2020）大多采用基于标量点积注意力（scalar dot-product attention）的方法。标量点积注意力使用单一的标量权重值来衡量2个向量之间的相似度或关联性，

限制了其捕捉点云间复杂依赖关系的能力，不利于应对复杂机载点云场景；相较之下，向量注意力机制（vector attention）的计算方式更为灵活，通过学习向量之间的上下文关系并计算不同维度的权重进行特征聚合，能够有效捕捉点云之间的复杂模式和长程依赖关系，在复杂机载点云场景中能够表现出更好的性能。本文利用KNN确定采样点的K个邻近点，将向量自注意力单元应用在K个邻近点组成的局部区域，并通过减法来表示采样点与其邻近特征之间的关系，获取更为丰富的点云自相关特征，具体计算公式可以描述为

$$\bar{f}_i = \sum_{j=0}^{K} \rho\left(\gamma\left(q\left(f_i\right) - k\left(f_{ij}\right) + \delta\left(p_i - p_{ij}\right)\right)\right) \odot \left(v\left(f_{ij}\right) + \delta\left(p_i - p_{ij}\right)\right) \tag{5}$$

式中，f_i代表采样点p_i的输入特征张量，$f_{i,j}$代表采样点p_i第j个邻近点的特征张量；q、k、v由线性层（Linear）与MLP层组成，对应了标量注意力机制中的查询（Query）、键（Key）和值（Value）。随后，$q\left(f_i\right) - k\left(f_{i,j}\right)$通过减法描述了采样点与邻近点之间的特征关系，$\delta\left(p_i - p_{i,j}\right)$则通过一个Linear层与RELU函数对采样点与邻近点之间的位置状态进行特征编码；γ表示为MLP组成的映射函数，对点特征进行线性变换；ρ表示为Softmax函数，对特征进行归一化操作，并计算特征张量之间的相似度；$\odot$则为点积相乘，建立起q、k、v之间的关联；最后通过求和运算聚合上述操作生成的特征张量值，得到向量自注意力单元的输出特征$\bar{f}_i$。

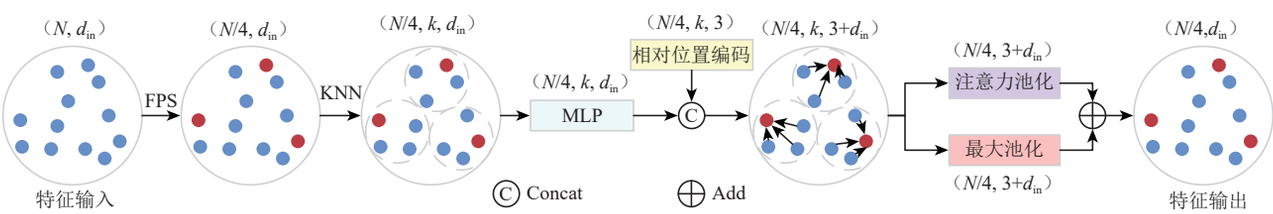


图4 局部区域编码单元

Fig. 4 Local regional encoding unit

2.4 动态残差连接模块

编码器经历了多次下采样，导致部分特征信息丢失，为了有效保留编码器的特征并获取更多细节语义信息，本文在模型的上采样过程中引入了动态残差连接模块。该模块根据特征间的相关性，动态融合编码器与解码器之间的语义特征，捕捉和聚合不同阶段的关键信息，从而增强模型对不同地物点间特征差异性的感知能力。具体来说，如图5所示，该模块的输入解码器的高级语义特征（INPUT）和编码器跳跃连接传递的低级语义特征（SKIP）。首先，这两个输入特征相加并使用ReLU函数进行激活，通过线性层（Linear）和批量归一化层BN（BatchNorm）对相加特征进行降维和归一化处理；随后，使用Sigmoid激活函数计算各特征维度的权重系数，学习特征间的相关性，并将这些权重系数与SKIP特征相乘，获取低级语义特征中与任务高度相关的关键特征；最后，对编码器低级语义特征与解码器高级语义特征分别进行残差连接，以防止模型过拟合，并减轻特征计算过程中的信息损失对模型性能的影响。输出特征（OUTPUT）具体表示为

$$y = x + \left(\text{skip} + \text{skip} \odot \left(\sigma\left(\text{ReLU}\left(x + \text{skip}\right)\right)\right)\right) \tag{6}$$

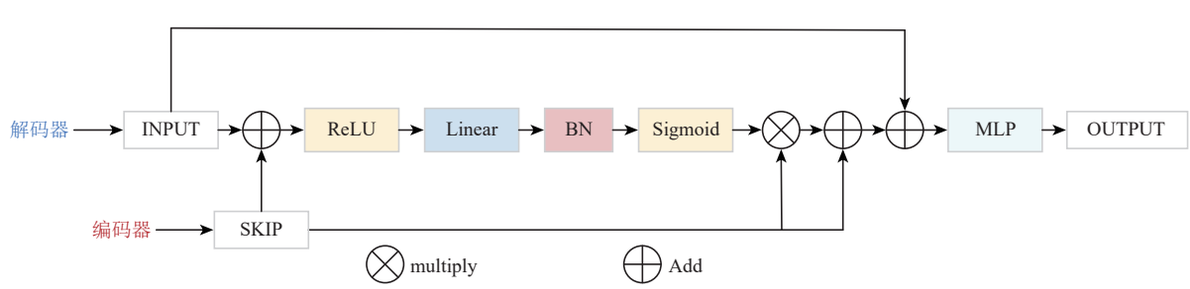


图5 动态残差连接模块

Fig. 5 Module of dynamic residual connection

式中，y为输出特征，x，skip分别为解码器输入特征与编码器输入特征，⊙为点积相乘，σ表示为由linear层、BatchNorm层和Sigmoid层的组合。

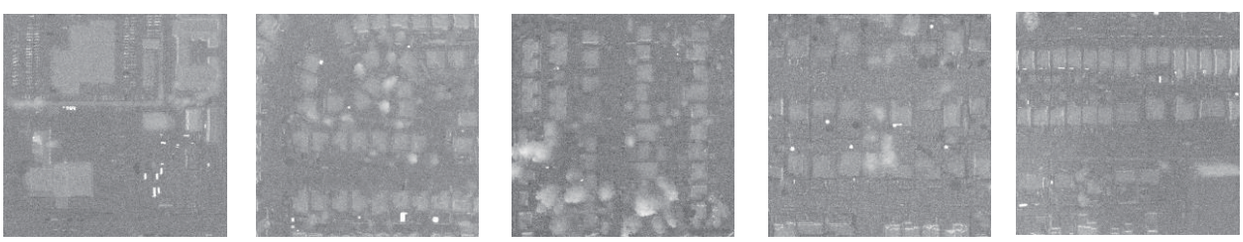
3 实验验证与结果分析

3.1 实验数据及评价指标

3.1.1 实验数据集

本文选用的实验数据集一是由戴顿大学（University of Dayton）推出的Dayton Annotated Laser Earth Scan（DALES）数据集（Varney等，2020）。DALES数据集是一个大型机载LiDAR点云数据集，涵盖了10 km^{2}范围的29个训练集和11个测试集场景，拥有包含地面、建筑物、汽车、卡车、电线杆、电力线、

围栏、和植被等8个类别的近5亿个点。图6（a）—（e）展示了其中5个具有代表性的不同场景中的原始数据。

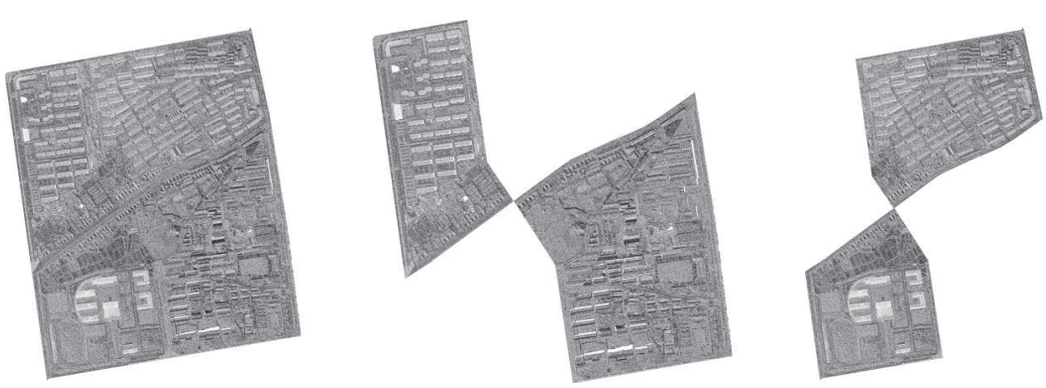


（a）场景一 （a）Scene 1　（b）场景二 （b）Scene 2　（c）场景三 （c）Scene 3　（d）场景四 （d）Scene 4　（e）场景五 （e）Scene 5

图6 DALES数据集部分场景原始数据

Fig. 6 Original data from various scenes in DALES dataset

本文选用的数据集二是由同济大学与慕尼黑工业大学推出的LASDU数据集（Ye等，2020）。LASDU数据集是HiWATER（Heihe Watershed Allied Telemetry Experimental Research）项目（Li等，2013）数据中的一部分，涵盖了1 km^{2}范围的黑河流域密集城市地区场景，拥有包含地面、建筑物、树木、低矮植被和人工制品等5个语义类别。图7（a）—（c）展示了该数据集的原始数据、训练数据与测试数据。



（a）原始数据 （a）Original data　（b）训练数据 （b）Training data　（c）测试数据 （c）Testing data

图7 LASDU数据集原始数据

Fig. 7 Original data in LASDU dataset

3.1.2 评价指标

为了有效评估网络在机载LiDAR点云语义分割中的性能，本文采用各类别交并比IoU（Intersection over Union）作为单个类别的评价指标；并使用总体准确度OA（Overall Accuracy）和平均交并比mIoU（mean Intersection over Union）作为整个测试集的

评价指标：

$$\mathrm{IoU} = \frac{T_\mathrm{P}}{T_\mathrm{P} + F_\mathrm{P} + F_\mathrm{N}} \tag{7}$$

$$\mathrm{OA} = \frac{T_\mathrm{P} + T_\mathrm{N}}{T_\mathrm{P} + T_\mathrm{N} + F_\mathrm{P} + F_\mathrm{N}} \tag{8}$$

$$\mathrm{mIoU} = \frac{\sum_{i=1}^{N} \frac{T_\mathrm{P}}{\left(T_\mathrm{P} + F_\mathrm{P} + F_\mathrm{N}\right)}}{N} \tag{9}$$

式中，T_P为某类样本正确分类的点数，F_P为某类样本错误分类的点数；T_N为其他样本正确分类的点数，F_N为其他样本错误分类的点数；N代表样本类别的数量。

3.2 实验环境设计

本文实验在NVIDIA RTX 4070Ti Super显卡、Intel i9-14900K的硬件环境与Ubuntu22.04、Cuda11.1、Python3.7、Pytorch1.9的软件环境下进行测试。模型在训练阶段中使用了SGD优化器，其中初始学习率设置为0.1，权重衰减率设置为0.001，动量系数为0.9，批处理大小为4，迭代次数为100次。此外，文中所用KNN算法的邻近点数量K均设置为16；损失函数选用加权交叉熵损失函数，权重值根据其对应类别所占比重动态决定，具体计算公式如下所示：

$$W_\mathrm{c} = \frac{1}{\ln\left(\alpha + \rho_\mathrm{c}\right)} \tag{10}$$

式中，ρ_c为该类别点数占总数量的比例，α为一个额外的超参数，设置为1.4，使各类别权重W_c的取值趋于［1，3］的范围中。

3.3 对比实验结果及分析

3.3.1 在DALES数据集上的实验结果及分析

针对DALES数据，选取PointNet++（Qi等，2017）、SPG（Landrieu和Simonovsky，2018）、KPConv（Thomas等，2019）、RandLA-Net（Hu等，2020）、Point Transformer（Zhao等，2021）、Superpoint Transformer（Robert等，2023）、EyeNet（Yoo等，2023）和TCFAP-Net（Zhang等，2024a）等点云深度学习网络进行定量对比，具体分割结果如表2所示。可见：本文提出的LAE-Transformer的OA与mIoU分别达到了97.8%和80.8%，优于其他大部分主流网络，与其中表现最好的KPConv（Thomas等，2019）处于同一水平；与同样基于自注意机制的Point Transformer（Zhao等，2021）相比，LAE-Transformer改进显著，其中OA提升了0.5%，mIoU提升了3.6%，说明了本文方法在机载点云场景下整体分割性能的优越性。此外，针对单个类别的分割，LAE-Transformer也取得了不错的效果，超过了表2中所列大部分的深度学习方法，在卡车、电线杆和围栏3个小尺度目标的IoU中实现了最佳，达到了42.1%、75.4%和63.8%，相较于Point Transformer（Zhao等，2021）的结果分别提高了12.7%、6.3%和5.6%，在电力线和汽车2个小尺度目标，IoU为94.5%和83.5%，略低于KPConv，在所有算法中排第二。上述结果验证了本文所提方法在小尺度目标提取中的有效性。

表2 DALES数据集中不同方法的语义分割结果定量对比

Table 2 Quantitative comparison of semantic segmentation results on different methods in DALES dataset

方法	总体准确度	交并比IoU/%								
	OA/%	平均交并比mIoU/%	地面	建筑物	汽车	卡车	电线杆	电力线	围栏	植被
PointNet++	95.7	68.3	94.1	89.1	75.4	30.3	40.0	79.9	46.2	91.2
SPG	95.5	60.6	94.7	93.4	62.9	18.7	28.5	65.2	33.6	87.9
EyeNet	97.2	79.6	**97.1**	**96.8**	83.5	41.3	72.5	94.6	57.3	93.4
KPConv	**97.8**	**81.1**	**97.1**	96.6	**85.3**	41.9	75.0	**95.5**	63.5	**94.1**
RandLA-Net	97.1	78.8	96.8	96.4	81.3	38.6	72.3	94.4	57.7	93.1
TCFAP-Net	97.1	78.3	96.5	96.5	82.3	33.2	69.9	94.0	60.0	93.7
Point Transformer	97.3	77.2	96.9	96.0	81.6	29.4	69.1	93.2	58.2	93.2
Superpoint Transformer	97.5	79.6	96.9	96.2	82.0	37.6	73.4	94.3	61.9	**94.1**
LAE-Transformer（本文方法）	**97.8**	80.8	97.0	**96.8**	83.5	**42.1**	**75.4**	94.5	**63.8**	93.6

注：数值粗体表示最优结果。

图8展示了标准分割结果和Point Transformer（Zhao等，2021）与LAE-Transformer在不同场景数据中的对比。对比图8（a）—（e）圆形虚线标注区域可以看出，Point Transformer（Zhao等，2021）

在植被、汽车、卡车、围栏和建筑物等地物之间存在明显的错分漏分现象，这主要是由于这些地物在局部区域内互相包含，且具有相近的几何特征，网络倾向于将其识别为相同类别，这也是导致小尺度目标地物分割困难的原因之一；相较之下，本文方法通过融合地物自身特征与其边缘特征，增强不同局部邻域间的关联，使得分割效果有了显著提升。图9展示了LAE-Transformer在局部放大区域的分割效果。可见虽然图中仍存在少量如汽车与其他类别之间的混淆，但整体上对地物结构的判别更准确，错分漏分的问题较少，特别是针对小尺度目标地物能够实现更加完整、准确的提取。

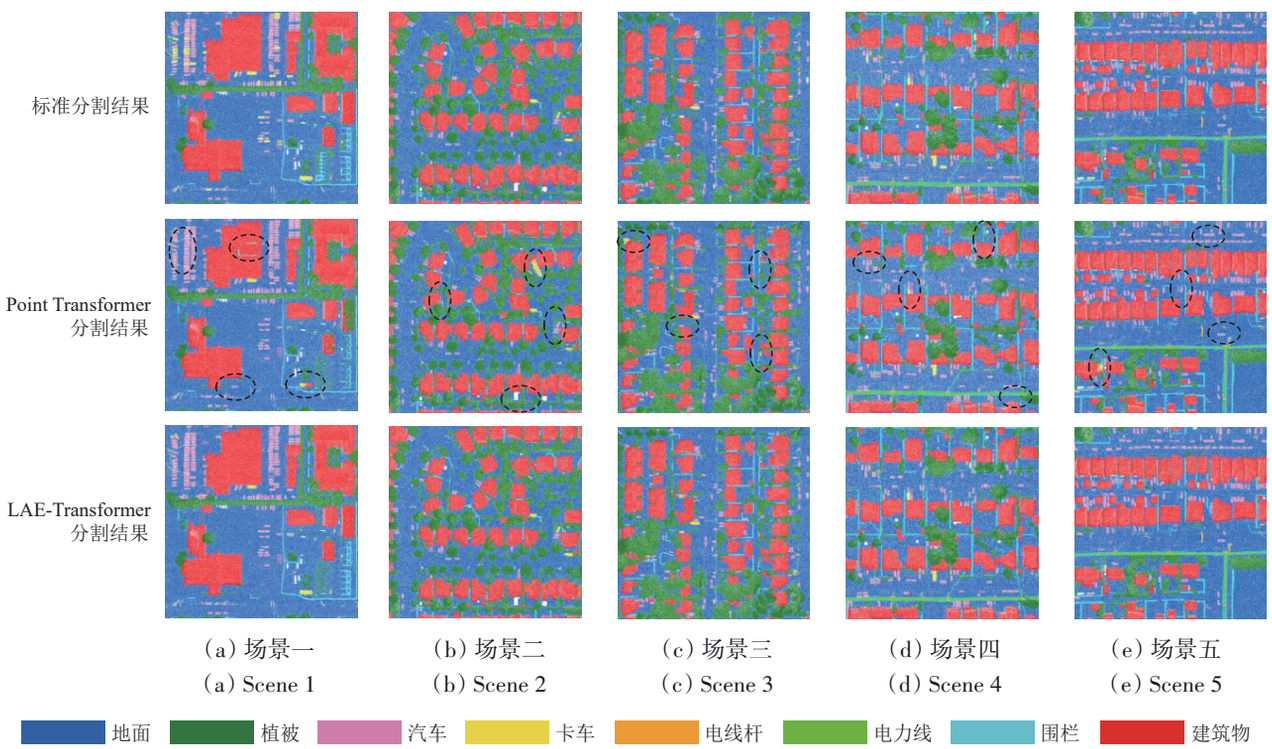


图8 DALES数据集部分场景语义分割结果

Fig. 8 Semantic segmentation results for various scenes in DALES dataset

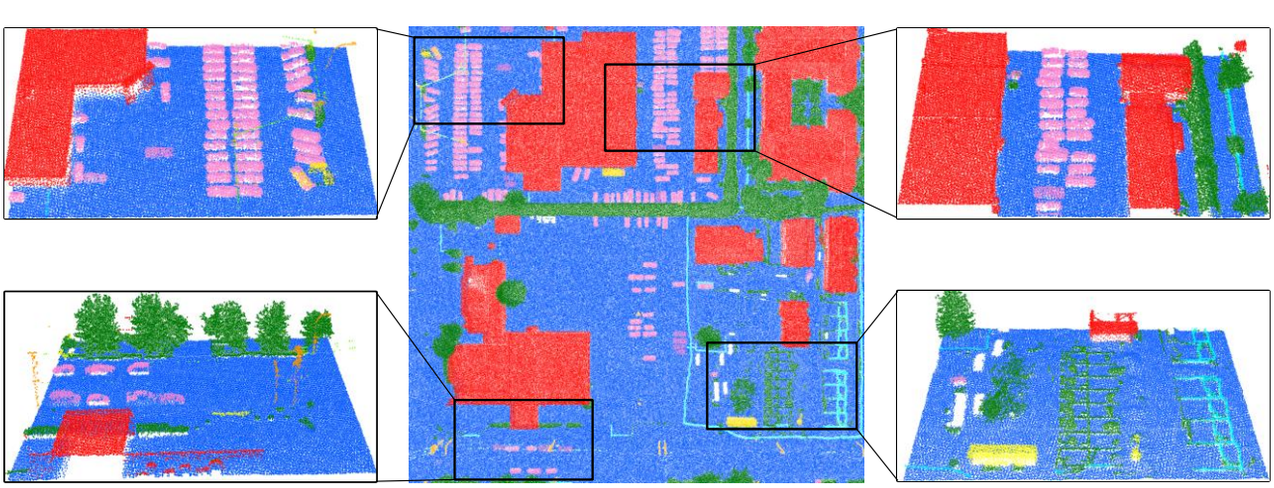


图9 LAE-Transformer局部区域分割效果

Fig. 9 LAE-Transformer local region segmentation performance

3.3.2 在LASDU数据集上的实验结果及分析

针对LASDU数据，选取PointNet++（Qi等，2017）、DGCNN（Wang等，2019）、KPConv（Thomas等，2019）、HDA-PointNet++（Huang等，2020）、

IPCONV（Zhang等，2023b）、DAKAG-Net（Zhao等，2024）和Point Transformer（Zhao等，2021）等点云深度学习网络进行定量对比，具体分割结果如表3所示。可见：本文提出的LAE-Transformer的OA与mIoU分别达到了87.2%和68.5%，与其中表现最好的DAKAG-Net（Zhao等，2024）处于同一水平，优于其他大部分网络，特别是在与Point Transformer（Zhao等，2021）相比时，OA提升了2.6%，mIoU提升了4.2%；在单个类别的分割效果上，LAE-Transformer同样取得了优秀的表现，在树木、低矮植被和人工制品3个类别的IoU中实现了最佳，达到了78.5%、57.7%和32.7%，相较于Point Transformer（Zhao等，2021）的结果分别提高了1.8%、8.1%和5.7%；在地面和建筑物类别的IoU则分别为81.8%和91.8%，接近于目前表现最好的网络DAKAG-Net（Zhao等，2024），这些结果进一步验证了LAE-Transformer在复杂机载点云场景下的实用性。

表3 LASDU数据集中不同方法的语义分割结果定量对比

Table 3 Quantitative comparison of semantic segmentation results on different methods in LASDU dataset

方法	总体准确度OA/%	交并比IoU/%					
		平均交并比mIoU/%	地面	建筑物	树木	低矮植被	人工制品
PointNet++	82.8	59.2	78.0	82.8	69.5	46.2	18.5
DGCNN	85.5	61.5	82.6	87.3	69.0	46.2	22.7
KPConv	83.7	60.1	80.3	87.6	71.2	42.6	19.1
HDA-PointNet++	84.4	61.5	79.7	87.2	69.8	48.4	22.6
IPCONV	86.7	64.6	82.6	92.8	75.1	42.5	30.2
DAKAG-Net	**87.4**	66.9	**83.2**	**93.3**	75.4	51.3	31.5
Point Transformer	84.6	64.3	78.4	89.6	76.7	49.6	27.0
LAE-Transformer(本文方法)	87.2	**68.5**	81.8	91.8	**78.5**	**57.7**	**32.7**

注：数值粗体表示最优结果。

图10展示了标准分割结果、Point Transformer（Zhao等，2021）和LAE-Transformer在测试集上的预测结果。对比方形框选区域可见，本文提出的LAE-Transformer在大多数场景下均表现出良好的性能，尽管仍存在少量的错分和漏分现象，但在人工制品等小尺度目标地物的分割效果上明显优于Point Transformer，有效避免了大范围地物点之间的分割混淆。

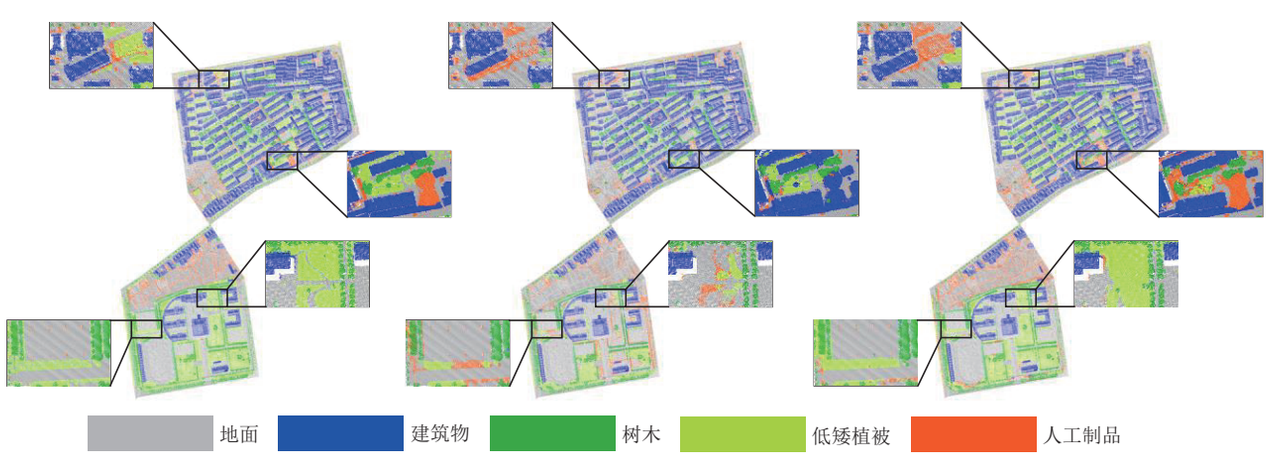


图10 LASDU数据集语义分割结果

Fig. 10 Semantic segmentation results in LASDU dataset

3.4 消融实验

为了进一步验证文中模型的有效性，本文基于DALES数据集设置了消融实验，用以评估不同的模块和参数对整体分割与小尺度目标类别分割效果的影响。

3.4.1 不同模块的消融实验

为了评估每个模块在本文方法中的作用与贡献，本文针对LTIE、RPT和DRC等模块进行消融

实验。设置了M1、M2、M3、M4、M5共5种不同模块的组合方式进行对比实验。实验方法的设定如表4所示。

表4 不同模块消融实验方法设定

Table 4 Setting of ablation experiment methods on different module

实验方法	模块		
	局部拓扑信息增强模块LTIE	区域点联合Transformer模块RPT	动态残差连接模块DRC
M1			
M2			√
M3		√	
M4	√		
M5(本文)	√	√	√

注：“√”表示选中该模块；空白表示未选中该模块。

不同模块消融实验结果见表5。可见：缺少全部模块的M1方法中各项精度都远低于增加了任意模块的M2、M3、M4等其他组合方法；同时，增加了全部模块的M5方法整体分割性能提升显著，相对于M1方法，OA提升了2.4%，mIoU提升了11.7%，针对小尺度目标的分割精度也有着不同程度的提升，验证了本文所提模块对小尺度目标地物提取与整体分割的有效性。

3.4.2 不同K近邻的消融实验

本文主要使用KNN进行邻近点的搜索与提取，为了评估不同K值对分割性能的影响，选取了K值4—64的共5组实验进行对比分析。实验结果如表6所示。可见：随着K值由4增长到16，OA与mIoU分别上涨了1.7%和6.5%，这是由于随着K邻近点的增长，采样点能够获取更为丰富的邻域信息，进而使得分割性能有所提升；但随着K值由16增长到64，OA与mIoU则分别下降了1.2%和4.3%，这是由于K邻近点的个数过多，造成了数据冗余，影响了关键点信息的提取。由此推断得出，本文应选取16作为最佳K近邻点数。

表5 不同模块消融实验结果

Table 5 Ablation experiment results on different Module

实验方法	总体准确度OA/%	平均交并比mIoU/%	部分小尺度目标地物交并比IoU/%				
			汽车	卡车	电线杆	电力线	围栏
M1	95.4	69.1	73.6	30.9	57.8	79.1	48.2
M2	96.1	74.9	77.5	33.4	63.2	86.4	51.6
M3	96.9	76.5	79.8	36.0	68.3	91.1	53.7
M4	96.3	76.9	78.4	37.2	69.5	90.3	53.1
M5(本文方法)	97.8	80.8	83.5	42.1	75.4	94.5	63.8

表6 不同K近邻消融实验结果

Table 6 Ablation experiment results on different numbers of KNN

K近邻点数	总体准确度OA/%	平均交并比mIoU/%
4	96.1	74.3
8	97.2	79.7
16(本文方法)	97.8	80.8
32	97.2	78.4
64	96.6	76.5

3.4.3 不同池化方法的消融实验

为了验证不同池化方法对分割性能的影响，选取了最大池化、平均池化和本文所提混合池化3种方法进行消融实验，实验结果如表7所示。可见，采用混合池化方法的模型分割性能最好，这是由于混合池化方法在聚焦关键信息的同时能够保留更多有效特征，减少了信息的损失。

表7 不同池化方法消融实验结果

Table 7 Ablation experiment results on different pooling methods

实验方法	总体准确度OA/%	平均交并比mIoU/%
最大池化	97.7	79.6
平均池化	97.6	79.2
混合池化(本文方法)	97.8	80.8

3.4.4 不同自注意力机制的消融实验

为了评估不同自注意力机制对分割性能的影响，选取了直接使用MLP、标量点积自注意力与文中所用向量自注意力进行分析对比，实验结果如表8所示。可见采用向量自注意力机制实验方法

的分割性能较好，这是由于向量自注意力对每个维度的权重进行单独计算，可以捕捉到更多的细粒度信息与点云之间的复杂依赖关系。

表8 不同池化方法消融实验结果

Table 8 Ablation experiment results on different Self-Attention mechanisms

实验方法	总体准确度OA/%	平均交并比mIoU/%
MLP	96.3	73.9
标量点积自注意力	97.3	79.0
向量自注意力(本文方法)	97.8	80.8

为了验证不同池化方法对分割性能的影响，选取了最大池化、平均池化和本文所提混合池化3种方法进行消融实验，实验结果如表7所示，可以看到，采用混合池化方法的模型分割性能最好，这是由于混合池化方法在聚焦关键信息的同时能够保留更多有效特征，减少了信息的损失。

3.4.5 不同浅层拓扑特征数量的消融实验

本文在LTIE中提取了7个点云浅层拓扑特征，作为模型输入。为了评估这些特征之间的相关性及其对分割性能的影响，采用ReliefF算法(Kononenko，1994)对特征进行了初步筛选和排序，该算法通过计算特征对样本点类别区分能力的权重，衡量各特征的重要性。

根据图11中的特征权重排序结果，本文进一步分析并对比了不同数量特征组合的表现，实验结果如表9所示。结果表明，随着特征数量的增加，模型的OA和mIoU逐渐提升，并在使用全部浅层拓扑特征时，达到了最佳的分割精度。

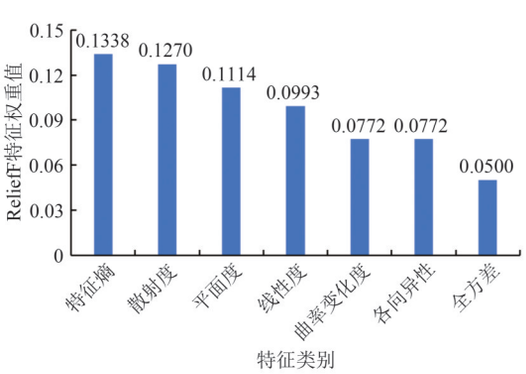


图11 ReliefF浅层拓扑特征权重排序

Fig. 11 Surface topological features weight ranking by ReliefF

表9 不同浅层拓扑特征数量消融实验结果

Table 9 Ablation experiment results on different numbers of surface topological features

浅层拓扑特征数量	总体准确度OA/%	平均交并比mIoU/%
0	96.6	76.4
3	97.1	77.7
5	97.5	79.3
7(本文方法)	97.8	80.8

4 结 论

为提升机载LiDAR点云语义分割能力，特别是小尺度地物的识别精度，本文以局部感知增强为核心，探究多模块协同的点云分割网络。主要结论如下：(1)所提出的融合局部拓扑增强、区域点联合Transformer和动态残差连接的神经网络结构，能有效提升点云语义分割的整体性能；(2)该网络在特征提取过程中显著增强了几何结构与边缘信息的表达能力，并建立起局部区域间的语义关联，强化了细节表征能力；(3)在DALES和LASDU数据集上的实验表明，本方法在整体精度和小尺度目标分割效果上优于多数对比模型，验证了其有效性与优越性。

然而，目前本文采用KNN选取局部采样点，其中K值的设定比较依赖人工经验，过大或过小的K值都容易造成数据缺失，影响最终的分割结果；同时，大规模点云浅层拓扑特征的计算也会带来额外的时间成本，并影响模型的运行效率。另一方面，不同采集设备、不同场景的点云数据差异很大，虽然本文在DALES和LASDU数据集上显示出较好的性能，但尚未在其他数据集上得到充分验证，可能存在数据集依赖性的问题。因此，将本文模型拓展至更多类型的点云数据集，并针对数据特点设计更为合适的局部采样方法，提高模型的泛化性和鲁棒性，将是后续研究的重点。

参考文献(References)

Cai H X, Zhang Z Y, Zhou Z Y, Li Z Y, Ding W B and Zhao J H. 2023. BEVFusion4D: learning LiDAR-Camera fusion under bird's-eye-view via cross-modality guidance and temporal aggregation. arXiv: 2303.17099 [DOI: 10.48550/arXiv.2303.17099]

Charles R Q, Su H, Kaichun M and Guibas L J. 2017. PointNet: deep learning on point sets for 3D classification and segmentation//

2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu: IEEE: 77-85 [DOI: 10.1109/CVPR.2017.16]

Feng M T, Zhang L, Lin X F, Gilani S Z and Mian A. 2020. Point attention network for semantic segmentation of 3D point clouds. Pattern Recognition, 107: 107446 [DOI: 10.1016/j.patcog.2020.107446]

Gevaert C M, Persello C, Sliuzas R and Vosselman G. 2016. Classification of informal settlements through the integration of 2D and 3D features extracted from UAV data. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, III-3: 317-324 [DOI: 10.5194/isprs-annals-III-3-317-2016]

Guo M H, Cai J X, Liu Z N, Mu T J, Martin R R and Hu S M. 2021. PCT: point cloud transformer. Computational Visual Media, 7(2): 187-199 [DOI: 10.1007/s41095-021-0229-5]

Hamdi A, Giancola S and Ghanem B. 2021. MVTN: multi-view transformation network for 3D shape recognition//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 1-11 [DOI: 10.1109/ICCV48922.2021.00007]

Hu Q Y, Yang B, Xie L H, Rosa S, Guo Y L, Wang Z H, Trigoni N and Markham A. 2020. RandLA-Net: efficient semantic segmentation of large-scale point clouds//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 11105-11114 [DOI: 10.1109/CVPR42600.2020.01112]

Huang R, Xu Y S, Hong D F, Yao W, Ghamisi P and Stilla U. 2020. Deep point embedding for urban classification using ALS point clouds: a new perspective from local to global. ISPRS Journal of Photogrammetry and Remote Sensing, 163: 62-81 [DOI: 10.1016/j.isprsjprs.2020.02.020]

Jiang M Y, Wu Y R, Zhao T Q, Zhao Z L and Lu C W. 2018. PointSIFT: a SIFT-like network module for 3D point cloud semantic segmentation. arXiv: 1807.00652 [DOI: 10.48550/arXiv.1807.00652]

Kononenko I. 1994. Estimating attributes: analysis and extensions of RELIEF//European Conference on Machine Learning. Catania: Springer: 171-182 [DOI: 10.1007/3-540-57868-4_57]

Landrieu L and Simonovsky M. 2018. Large-scale point cloud semantic segmentation with superpoint graphs//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4558-4567 [DOI: 10.1109/CVPR.2018.00479]

Le T and Duan Y. 2018. PointGrid: a deep network for 3D shape understanding//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 9204-9214 [DOI: 10.1109/CVPR.2018.00959]

Li K Q, Zhao M Y, Wu H Y, Yan D M, Shen Z, Wang F Y and Xiong G. 2022. GraphFit: learning multi-scale graph-convolutional representation for point cloud normal estimation//17th European Conference on Computer Vision. Tel Aviv: Springer: 651-667 [DOI: 10.1007/978-3-031-19824-3_38]

Li X, Cheng G D, Liu S M, Xiao Q, Ma M G, Jin R, Che T, Liu Q H, Wang W Z, Qi Y, Wen J G, Li H Y, Zhu G F, Guo J W, Ran Y H, Wang S G, Zhu Z L, Zhou J, Hu X L and Xu Z W. 2013. Heihe watershed allied telemetry experimental research (HiWATER): scientific objectives and experimental design. Bulletin of the American Meteorological Society, 94(8): 1145-1160 [DOI: 10.1175/BAMS-D-12-00154.1]

Lu W X, Wan Y C, He P P, Chen M L, Qin J X and Wang S Y. 2015. Extracting and plane segmenting buildings from large scene point cloud. Chinese Journal of Lasers, 42(9): 0914004 (卢维欣, 万幼川, 何培培, 陈茂霖, 秦家鑫, 王思颖 . 2015. 大场景内建筑物点云提取及平面分割算法 . 中国激光, 42(9): 0914004) [DOI: 10.3788/cjl201542.0914004]

Ma X, Qin C, You H X, Ran H X and Fu Y. 2022. Rethinking network design and local geometry in point cloud: a simple residual MLP framework. arXiv: 2202.07123 [DOI: 10.48550/arXiv.2202.07123]

Maturana D and Scherer S. 2015. VoxNet: a 3D Convolutional Neural Network for real-time object recognition//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Hamburg: IEEE: 922-928 [DOI: 10.1109/IROS.2015.7353481]

Qi C R, Yi L, Su H and Guibas L J. 2017. PointNet++: deep hierarchical feature learning on point sets in a metric space. arXiv: 1706.02413 [DOI: 10.48550/arXiv.1706.02413]

Robert D, Raguet H and Landrieu L. 2023. Efficient 3D semantic segmentation with superpoint transformer. arXiv: 2306.08045 [DOI: 10.48550/arXiv.2306.08045]

Stilla U and Jutzi B. 2008. Waveform analysis for small-footprint pulsed laser systems//Shan J and Toth C K, eds. Topographic Laser Ranging and Scanning: Principles and Processing. Boca Raton: CRC Press: 215-234

Su H, Maji S, Kalogerakis E and Learned-Miller E. 2015. Multi-view convolutional neural networks for 3D shape recognition//2015 IEEE International Conference on Computer Vision (ICCV). Santiago: IEEE: 945-953 [DOI: 10.1109/ICCV.2015.114]

Thomas H, Qi C R, Deschaud J E, Marcotegui B, Goulette F and Guibas L J. 2019. KPConv: flexible and deformable convolution for point clouds//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE: 6410-6419 [DOI: 10.1109/ICCV.2019.00651]

Varney N, Asari V K and Graehling Q. 2020. DALES: a large-scale aerial LiDAR data set for semantic segmentation//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle: IEEE: 717-726 [DOI: 10.1109/CVPRW50498.2020.00101]

Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser Ł and Polosukhin I. 2017. Attention is all you need//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc.: 6000-6010

Wang P S. 2023. OctFormer: octree-based transformers for 3D point clouds. ACM Transactions on Graphics, 42(4): 155 [DOI: 10.1145/3592131]

Wang Y, Sun Y B, Liu Z W, Sarma S E, Bronstein M M and Solomon J M. 2019. Dynamic graph CNN for learning on point clouds. ACM Transactions on Graphics, 38(5): 146 [DOI: 10.1145/3326362]

Wei X, Wang M Y, Lin S H J, Li Z Y, Yang J, Al-Jawari A and Tang X. 2023. Multi-scale geometry-aware transformer for 3D point cloud classification. arXiv: 2304.05694 [DOI: 10.48550/arXiv.2304.05694]

Weinmann M, Jutzi B, Hinz S and Mallet C. 2015. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing, 105: 286-304 [DOI: 10.1016/j.isprsjprs.2015.01.016]

Xiong K Z, Zheng M J, Xu Q S, Wen C L, Shen S Q and Wang C. 2024. SPEAL: skeletal prior embedded attention learning for cross-source point cloud registration. arXiv: 2312.08664 [DOI: 10.48550/arXiv.2312.08664]

Yang B S, Han X and Dong Z. 2021. A deep learning network for semantic labeling of large-scale urban point clouds. Acta Geodaetica et Cartographica Sinica, 50(8): 1059-1067 (杨必胜, 韩旭, 董震. 2021. 适用于城市场景大规模点云语义标识的深度学习网络. 测绘学报, 50(8): 1059-1067) [DOI: 10.11947/j.AGCS.2021.20210093]

Ye X Q, Li J M, Huang H X, Du L and Zhang X L. 2018. 3D recurrent neural networks with context fusion for point cloud semantic segmentation//15th European Conference on Computer Vision–ECCV 2018. Munich: Springer: 415-430 [DOI: 10.1007/978-3-030-01234-2_25]

Ye Z, Xu Y S, Huang R, Tong X H, Li X, Liu X F, Luan K F, Hoegner L and Stilla U. 2020. LASDU: a large-scale aerial LiDAR dataset for semantic labeling in dense urban areas. ISPRS International Journal of Geo-Information, 9(7): 450 [DOI: 10.3390/ijgi9070450]

Yoo S, Jeong Y, Jameela M and Sohn G. 2023. Human vision based 3D point cloud semantic segmentation of large-scale outdoor scenes//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Vancouver: IEEE: 6577-6586 [DOI: 10.1109/CVPRW59228.2023.00699]

Zhang H, Wang C S, Tian S W, Lu B L, Zhang L P, Ning X and Bai X. 2023a. Deep learning-based 3D point cloud classification: a systematic survey and outlook. Displays, 79: 102456 [DOI: 10.1016/j.displa.2023.102456]

Zhang J J, Jiang Z P, Qiu Q J and Liu Z. 2024a. TCFAP-Net: transformer-based cross-feature fusion and adaptive perception network for large-scale point cloud semantic segmentation. Pattern Recognition, 154: 110630 [DOI: 10.1016/j.patcog.2024.110630]

Zhang R X, Chen S Y, Wang X Y and Zhang Y S. 2023b. IPCONV: convolution with multiple different kernels for point cloud semantic segmentation. Remote Sensing, 15(21): 5136 [DOI: 10.3390/rs15215136]

Zhang S, Chen Y P, Wang B, Pan D, Zhang W M and Li A G. 2024b. SPTNet: sparse convolution and transformer network for woody and foliage components separation from point clouds. IEEE Transactions on Geoscience and Remote Sensing, 62: 5702718 [DOI: 10.1109/TGRS.2024.3376454]

Zhao C, Guo H T, Lu J, Yu D X and Zhang B M. 2020. Airborne LiDAR point cloud classification based on deep residual network. Acta Geodaetica et Cartographica Sinica, 49(2): 202-213 (赵传, 郭海涛, 卢俊, 余东行, 张保明. 2020. 基于深度残差网络的机载LiDAR点云分类. 测绘学报, 49(2): 202-213) [DOI: 10.11947/j.AGCS.2020.20190004]

Zhao H S, Jiang L, Jia J Y, Torr P and Koltun V. 2021. Point transformer//2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal: IEEE: 16239-16248 [DOI: 10.1109/ICCV48922.2021.01595]

Zhao J B, Zhou H Y and Pan F F. 2024. A dual attention KPConv Network combined with attention gates for semantic segmentation of ALS point clouds. IEEE Transactions on Geoscience and Remote Sensing, 62: 5107914 [DOI: 10.1109/TGRS.2024.3422829]

Zhou Y and Tuzel O. 2018. VoxelNet: end-to-end learning for point cloud based 3D object detection//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE: 4490-4499 [DOI: 10.1109/CVPR.2018.00472]

LAE-Transformer: An airborne LiDAR point cloud segmentation network incorporating local-aware enhancement

KU Minfan[1], ZHANG Liang[1,2,3], DENG Jiwei[2,3], WANG Guangshuai[2,3], WANG Xinwen[1], LIU Hengzhi[1]

1. Faculty of Resources and Environmental Science, Hubei University, Wuhan 430062, China;
2. Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology, Tianjin 300251, China;
3. China Railway Design Group Limited, Tianjin 300251, China

Abstract: Existing semantic segmentation algorithms that are based on deep learning for airborne LiDAR point clouds encounter specific challenges when they deal with complex scenes. These challenges include insufficient utilization of local topological information, limited capability in multiscale feature representation, and low segmentation accuracy for small-scale objects, such as trucks, poles, and fences. These issues directly hinder the deployment of airborne LiDAR systems in practical applications. An airborne LiDAR point cloud segmentation network incorporating local-aware enhancement, namely, LAE-Transformer, is proposed in this study to address the aforementioned challenges. It aims to improve the accuracy and reliability of semantic segmentation tasks for airborne LiDAR point clouds by effectively utilizing local topological information and multiscale features.

The proposed LAE-Transformer adopts an encoder–decoder structure with skip connections and incorporates four major innovations. First, the local topological information enhancement module is introduced to extract shallow topological features by using a K-nearest neighbor (KNN) approach. This module constructs a local geometric structure graph that allows the model to capture the fine-grained spatial

variations and geometric edges of objects. Second, during the encoding phase, a combination of downsampling layers and Regional Point Transformer (RPT) modules is employed to extract deep semantic features while expanding the receptive field. The RPT modules utilize self-attention mechanisms to model intra- and inter-region dependencies, substantially improving the model's ability to represent features at multiple scales. Third, during the decoding phase, a dynamic residual connection module is incorporated to adaptively fuse features from different semantic layers, thus ensuring the continuity and integrity of spatial information in the upsampling process. Last, a mixed pooling layer that integrates max pooling and attention pooling is designed to aggregate local salient features and global contextual information, effectively reducing information loss and maintaining rich feature diversity during processing.

Extensive experiments are conducted on two publicly available benchmark datasets DALES and LASDU. The proposed LAE-Transformer achieves remarkable improvements in segmentation performance compared with several state-of-the-art models. Specifically, it attains an overall accuracy of 97.8% and 87.2% and mean Intersection over Union (IoU) scores of 80.8% and 68.5% on DALES and LASDU datasets, respectively. In addition, the model demonstrates substantial advantages in segmenting small-scale objects. On the DALES dataset, it achieves IoU scores of 42.1% for trucks, 75.4% for poles, and 63.8% for fences, outperforming widely used methods, such as point Transformer. These results strongly validate the model's superior accuracy, robustness, and generalization capabilities in handling complex airborne LiDAR scenes.

The proposed LAE-Transformer effectively overcomes the limitations of current semantic segmentation models by integrating local topological feature enhancement, multiscale Transformer-based feature extraction, dynamic residual connections, and a mixed pooling mechanism. These innovations allow the network to efficiently extract and fuse local and global features, thereby improving the segmentation accuracy for small-scale objects and the overall reliability of semantic segmentation in airborne LiDAR point cloud. The method's practical value and potential for deployment in airborne LiDAR applications across complex environments are also validated.

Key words：airborne LiDAR point cloud, deep learning, semantic segmentation, self-attention mechanism, local-aware enhancement, complex scenes, small-scale objects, Transformer

Supported by Tianjin Key Laboratory of Rail Transit Navigation Positioning and Spatio-temporal Big Data Technology (No. TKL2023A13, TKL2023B15); Science and Technology Research and Development Program of China State Railway Group Co. Ltd. (No. L2023G016); National Natural Science Foundation of China (No. 41601504)