自适应显隐结构融合的铁道点云分割网络

吴非凡, 巨志勇, 夏涵锐

(上海理工大学光电信息与计算机工程学院,上海 200093)

摘　要：针对现有铁路场景点云语义分割方法难以兼顾跨场景的鲁棒性与物体分割精度。选用 RandLA-Net 作为基础框架，开发了专门针对实际铁道工程场景的显隐式三维结构自适应融合网络（RandLA-Net with Adaptive Fusion of Explicit Structure，RandLA-AFES），在编码器阶段并行引入了显式结构特征提取模块，利用局部空间分区策略捕获结构核，生成显式几何先验，有效补偿了下采样过程中的细粒度结构损失；利用邻域特征采样模块聚合深层局部上下文信息，以保证点云高维隐式语义表征的完整性；设计了动态自适应融合机制，通过引入可学习权重参数，动态校准显式几何特征与隐式语义特征的融合比例，实现了双模态特征的最优耦合，有效抑制了冗余噪声。在公开铁路场景数据集 WHU-Railway3D 上的大量实验表明，提出的方法对比基线模型，仅在增加极少量参数和推理时间的基础上，在平均交并比、总体精度等指标上均优于现有主流方法，更在不同复杂道路场景中展现出卓越的语义分割能力，验证了其有效性与工程应用潜力。

关键词：激光点云；语义分割；深度学习；铁路场景

中图分类号:TP391.41　　**文献标志码**：A

引　　言

三维点云语义分割在铁路基础设施的数字化巡检与智能维保中发挥着至关重要的作用。然而，真实铁路大场景具有极端的多尺度与空间异质性，其不仅包含大面积低频的平坦区域（如道床、地面），又包含大量高频稀疏且极细长的拓扑结构（如接触网、支柱、电杆）。这种复杂的场景特性对分割算法的跨场景泛化能力与细粒度几何保持能力提出了严峻挑战。

近年来，深度学习在点云处理领域取得了显著进展，目前关于点云语义分割的方法大致可分为 4 类：基于投影的方法、基于体素化的方法、基于点的方法，基于多模态的方法。基于投影的方法[1-3]需将三维点云映射至二维平面或多视角图像，导致几何信息损失、遮挡及尺度畸变，且性能严重依赖视角选择；基于体素的方法[4-6]受限于立方体划分，面临分辨率以及内存权衡，高分辨率时显存需求过大，低分辨率时边界与细小结构被过度平滑；基于多模态的方法[7-9]虽然利用图像、深度等多源信息，但依赖跨模态精确配准与同步，训练流程复杂，且在光照不足、纹理缺失场景下性能骤降。基于点的方法直接以原始无序点云为输入，避免了投影、体素化或多模态融合带来的中间表示，因而在精度、内存及灵活性等方面具有独特优势。而以 PointNet++[10]为代表的基于点的方法，由于依赖复杂的图构建或最远点采，难以应对铁路级百万规模点云的内存与算力消耗。为此，RandLA-Net[11]等网络通过“随机采样+局部特征聚合”的范式，在大规模场景分割中实现了效率与精度的扩展平衡，成为目前主流的轻量级基线模型。然而其随机采样机制在细粒度细节保持及密度鲁棒性方面仍存固有局限。且对于点云几何结构特征的利用有限，导致局部与全局信息利用不足。

尤其对于铁道场景，现有 RandLA-Net 等方法在处理铁路接触网、支柱等细长结构时，极易破坏对象的几何拓扑完整性。为此，本文以 RandLA-Net 网络为基础框架，提出显隐式三维结构自适应融合网络 RandLA-AFES。本文定义隐式特征为网络深层提取的语义表征，定义显式特征为基于局部邻域协方差计算的几何属性。引入显式特征的目的是在下采样过程中显式地补偿丢失的几何细节。具体地，在编码器的每一级局部特征聚合模块之后，并行插入显式结构特征提取模块，利用原始点云的数据构造轻量级几何先验，显式补偿随机采样造成的结

基金项目：国家自然科学基金资助项目(81101116)。
收稿日期：2026-01-25　**修回日期：**2026-07-05

构信息损失，有效解决了基线模型在稀疏细长结构上分割精度不足的问题。引入自适应融合模块，通过学习从而动态调整网络中隐式高维语义特征与输出的显式几何特征的融合比重。

本文的主要贡献总结如下：

（1）提出了一种显隐式三维结构自适应融合网络。针对大规模点云处理中随机采样策略导致跨场景泛化能力弱的问题，本文在轻量级基线网络的基础上引入自适应显隐式结构融合模块，在保持高效推理的同时，有效弥补了网络对局部与全局几何信息利用不足的缺陷。

（2）设计了显式结构特征提取模块。针对铁路场景中支柱、接触网等细长物体在下采样时极易丢失几何拓扑的问题，本模块引入局部空间分区策略提取显式结构核，并将其作为几何先验注入特征流。这一机制有效补偿了细粒度的细节损失，显著增强了网络对局部几何结构的显式刻画能力。

（3）构建了基于可学习权重的动态自适应特征融合机制。针对铁路场景显著的空间异质性，该机制能够动态校准深层隐式语义特征与显式几何先验的融合比重。通过实现语义与几何表征的最优耦合，有效抑制了冗余噪声，增强了特征表达的稳定性与鲁棒性。

1 RandLA-Net 网络框架

RandLA-Net 是一种精简高效的神经网络架构，专为从大规模点云数据中直接推断语义信息而设计。该方法的核心在于将点云随机采样(Random Sample,RS)与局部特征聚合(Local Feature Aggregation,LFA)策略相结合，LFA 模块通过对邻域点进行 K 近邻查找，利用 MLP(Multi-Layer Perceptron, MLP)和池化操作提取局部上下文特征，MLP 作为一种前馈人工神经网络，在此处利用其共享权重的特性，对局部邻域内的点云特征进行高维非线性映射与空间变换，从而有效地提取并聚合深层次的局部上下文语义信息。从而实现高效且精准的点云语义分割。整体采用编码器-解码器结构，编码阶段通过随机采样逐步降采样并用局部特征聚合模块补偿随机采样丢失的信息，解码阶段上采样恢复分辨率，最终高效输出逐点语义标签，可直接处理百万级点云，兼顾精度与速度，适用于大规模实时三维场景理解，结构如图 1 所示。

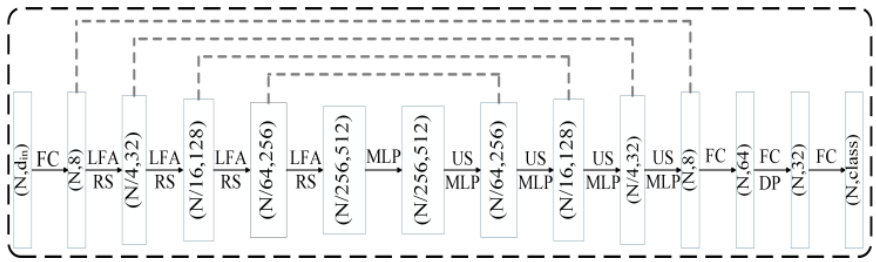

图 1　RandLA-Net 整体框架

Fig.1　Overall Framework of RandLA-Net

2 改进的 RandLA-AFES 网络

RandLA-Net 在原始实现中仅依赖 K 近邻采样和欧氏距离进行特征提取，尽管这种方法简单直观，大幅提升了计算效率能在大规模场景语义分割中展现出显著优势。但由于采用随机采样策略带来的跨域泛化能力弱，且对点云特征的利用不足导致网络局部-全局信息特征提取不完全。为此，本文提出了一种更为高效的网络架构 RandLA-AFES。网络在随机下采样阶段，引入了自适应显隐式结构融合模块(Adaptive Fusion of Explicit-Implicit Structure Module,AFEM)，该模块主要由显式结构特征提取模块(Explicit Structural Feature Extractor,ESFE)，邻域特征采样模块(Neighborhood Feature Sampling Module,NFSM)以及自适应融合模块(Adaptive Fusion Module,AFM)组成，如图 2 所示该模块通过显式结构特征提取模块来构建显式特征先验来补偿损失的特征，并与提取深度语义信息的邻域特征采样模块通过自适应融合模块融合点云语义与几何特征各自权重，增强点云特征表达的稳定性，解决下采样时损失的结构信息损失。计算公式如下：

$$\hat{F}_i = \alpha \hat{F}_i^{(1)} + \beta \hat{F}_i^{(2)} \tag{1}$$

其中 α 与 β 为经过自适应融合模块学习后每个点的语义和几何特征的融合权重。$\hat{F}_i^{(1)}$ 为经过邻域特征采样模块提取到的隐式特征，$\hat{F}_i^{(2)}$ 为经过显式结构特征提取模块得到的显式特征。

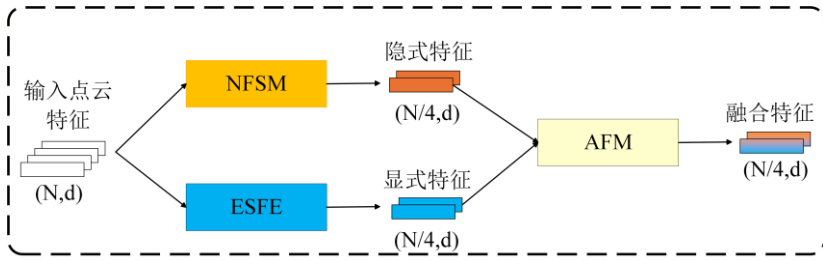

图 2　自适应显隐式结构融合模块

Fig.2 Adaptive Fusion of Explicit-Implicit Structure Module

2.1 隐式结构特征提取模块

隐式特征是指通过深层神经网络自动学习到的特征表示。这种特征虽然缺乏直观的物理几何解释，但其每个通道都编码了局部邻域内复杂的点分布模式、相对位置关系以及潜在的语义上下文。这种高维抽象能力使得网络能够理解场景的整体语义，而不仅仅局限于局部的几何形状。为了高效地捕获点云的上下文深层隐式语义，我们保留了 RandLA-Net 中提出的局部特征聚合结构来尽可能完整提取点云中的高维信息，而后通过邻域特征采样模块避免下采样时，重要特征信息的丢失。

2.1.1 局部特征聚合

对于一个点云 P，为了能让网络学习到其中复杂的局部结构，局部特征聚合结构首先对输入的每个中心点 p_i 使用 K 近邻算法搜索其空间邻域点集 $\{p_i^1,...,p_i^k\}$，而后将点集引入相对位置编码，并通过卷积将其映射到高维特征空间，得到位置编码特征 r_i^k。经过聚合而后与该点的原始特征 p_i^k 跳跃连接，最终生成中心点 p_i 的隐式特征向量 f_i。

$$f_i = \text{Aggregation}(\text{Conv}(r_i^k + f_i^k)) \tag{2}$$

2.1.2 邻域特征采样模块

对于点云下采样中丢失的高维嵌入特征，采用邻域特征采样模块通过聚合每个点的邻域特征来加强点特征的表达。如图 3 所示该模块的主要功能是对输入的点云隐式特征进行下采样，并通过最大值池化的方式聚合每个点的 K 近邻点的特征。

$$\hat{F}_i^{(1)} = \text{MaxPooling}(f_i^k) \tag{3}$$

方法可以有效地减少数据量并保留重要的特征信息，使用最大值池化是确保每个隐式特征向量 f_i 中包含了其 k 个近邻点中最具代表性的特征，从而提升了特征的鲁棒性和表达能力。

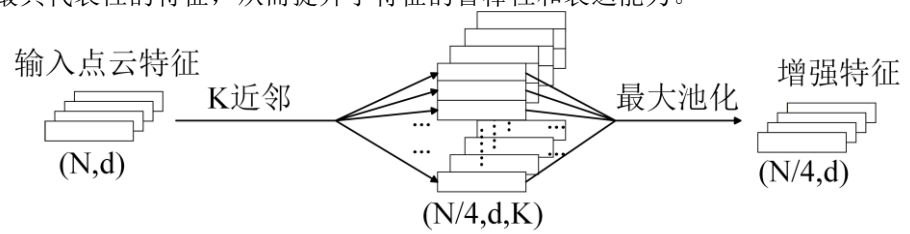

图 3　邻域特征采样模块

Fig.3　Neighborhood Feature Sampling Module

2.2　显式结构特征提取模块

根据奈奎斯特空间采样定理，随机采样作为一种各向同性的均匀下采样，当其局部采样密度低于结构特征的空间变化率时，必然导致高频几何结构发生混叠与丢失。基于文献[12]，本文提出了显式结构特征提取模块。其作用等效于在随机采样前引入了一个空间高通滤波器。它通过提前计算并保留局部邻域的空间分布统计量，在欧氏空间中强制保留了形状先验，从而在理论层面上补偿了高维特征映射过程中的局部高频几何信息损失。

该模块首先通过 PointHop[13]方法在输入三维空间中捕获显式的局部结构信息，并利用这些信息生成具有共享权重的动态核函数，将其作为显式结构核。在后续的特征提取网络中，通过提供几何特征的初步嵌入来计算

每个点与局部结构信息之间的相关性，从而提取点云中的显式结构局部特征，并通过动态位置增强嵌入特征的表达能力，而后结合静态位置信息和动态特征信息，并行通过全局和局部特征池化来增强下采样后点云的特征整体表达，生成最终的特征输出。

2.2.1 PointHop 方法

PointHop 方法的主要功能如图 4 所示，是对点云数据进行特征提取，提取的特征包括点云的标准差、中心点坐标以及基于邻域点不同分区的平均特征。通过这种方式，模块能够捕捉到点云数据的局部几何信息，并可以直接生成高维的点云特征。相比于传统基于协方差或法向量的几何提取，PointHop 通过基于空间象限的均值聚合，天然克服了点云无序性，并在保持极低计算复杂度的同时，能显式刻画细长结构的垂直度与线性度，极大地弥补了随机采样对铁路细微物体带来的拓扑截断问题。

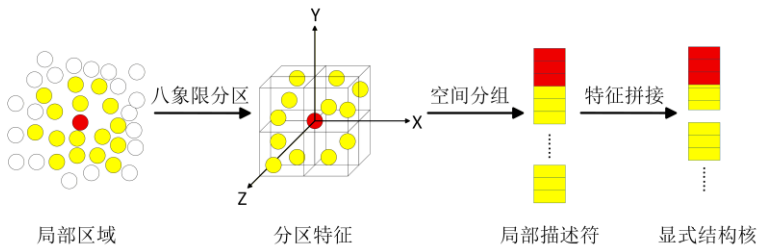

图 4　PointHop 实现流程图

Fig.4　PointHop Implementation Flowchart

PointHop 的具体处理流程为从输入的三维点云数据中每个点及邻居点按照八个空间进行划分以中心点为局部坐标系原点，依据三维坐标将邻域划分为八个象限，在每个象限内对邻点属性取均值以形成方向性的低阶统计量，如公式(4)所示。

$$\bar{a}_i^p = \frac{1}{n_i^p}\sum_{j=1}^{k} r_{i,j} t_{i,j}^p \, p = 1,\cdots 8 \tag{4}$$

其中 $\bar{a}_i^p$ 表示第 i 个点的第 p 个分区的质心坐标，n_i^p 代表该分区的邻域点数量。$p_{i,j}$ 第 i 个中心点的第 j 个邻域点，$t_{i,j}^p$ 表示 $r_{i,j}$ 是否属于第 p 个分区，其表达式可写为：

$$t_{i,j}^p = \begin{cases} 0, & t_{i,j}^p \notin \xi_p \\ 1, & t_{i,j}^p \in \xi_p \end{cases} \tag{5}$$

其中 ξ_p 表示第 p 个分区。

而后将八个进行过空间分组的描述向量拼接成固定维度的空间向量作为后续的动态结构核。

$$ESK_i = \mathrm{Concat}\left\{\bar{a}_i^p\right\}_{j=1}^{8} \tag{6}$$

该方法以象限编码天然消除了点云无序性，以均值聚合提升对扫描扰动的鲁棒性，同时保持对局部几何结构的显式刻画。

2.2.2 显式结构特征提取网络

为避免高维嵌入逐渐稀释原始几何细节，本文设计了一个轻量级并行分支网络。

该分支利用 PointHop 在输入空间捕获局部结构核，随后将几何先验注入下采样后的特征流，并通过全局-局部联合池化强化表征。

该网络如图 5 所示，分别对显式特征和隐式特征进行下采样以及 K 近邻方法，得到了下采样后带有邻域的特征 $r_{i,j}$ 和 $f_{i,j}$。随后对于显式特征 $r_{i,j}$，将先前计算得到的显式结构核 ESK_i 作为提供几何特征的初步嵌入，得到点云结构位置信息的特征 $\hat{p}_{i,j}$。

$$\hat{p}_{i,j} = r_{i,j} \oplus \mathrm{MLP}\left(ESK_i\right) \tag{7}$$

而 $f_{i,j}$ 则与映射至同一维度的显式特征相融合，通过动态位置嵌入得到 $\hat{f}_{i,j}$。

$$\hat{f}_{i,j} = f_{i,j} \oplus \mathrm{MLP}(p_{i,j}) \tag{8}$$

而后将 $\hat{p}_{i,j}$ 与 $\hat{f}_{i,j}$ 相加，分别在全局池化空间和局部邻域空间，进行最大池化，通过 MLP 得到强化后的特

征 $\hat{F}_i^{(1)}$。

$$\hat{F}_{i,j} = \hat{f}_{i,j} + \hat{p}_{i,j} \tag{9}$$

$$\hat{F}_i^{(2)} = \mathrm{MLP}(\mathrm{Concat}[\mathrm{GlobalPooling}(\hat{F}_{i,j}), \mathrm{LocalPooling}(\hat{F}_{i,j})]) \tag{10}$$

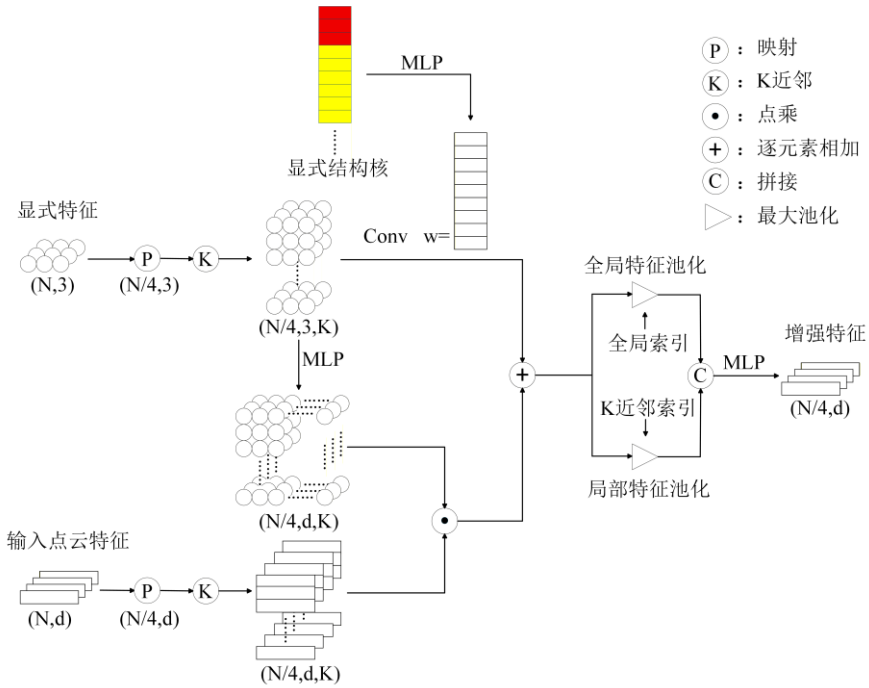

图 5　显式结构特征提取网络图

Fig.5　Explicit Structural Feature Extraction Network Diagram

2.3　自适应融合模块

考虑到铁路场景单帧点云规模庞大，若采用一般的通道注意力机制，将急剧增加计算开销与推理延迟。以此本文设计自适应标量权重，其核心目的在于以极低参数量在空间维度实现显隐式特征的有效加权，从而在保证大场景推理速度的同时，实现模型精度与轻量化的最优平衡。

如图 6 所示自适应融合模块通过引入可学习的参数来自适应二者权重，实现语义和几何最优耦合，避免引入无关噪声。权重计算方式如下：

$$\alpha, \beta = \mathrm{SoftMax}(\mathrm{Conv}(\mathrm{Concat}(\hat{F}_i^{(1)}, \hat{F}_i^{(2)}))) \tag{11}$$

其中 α 与 β 为经过学习后每个点的语义和几何特征的融合权重，被设定为全局可学习参数，在网络训练阶段通过反向传播自动优化更新，以实现该权重经过充分学习后全局通用，能够自适应不同场景下的多模态特征分布。

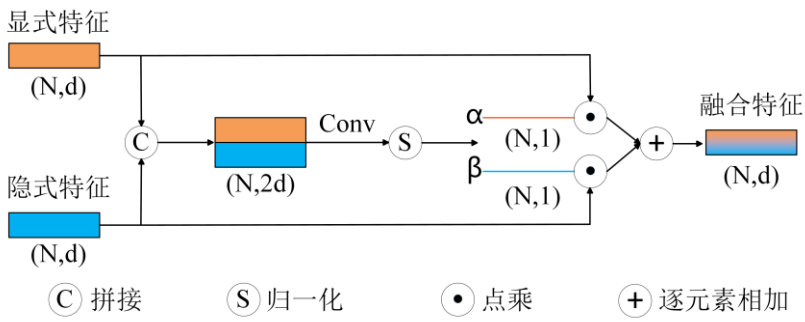

图 6　自适应融合模块结构图

Fig.6　Adaptive Fusion Module Structure Diagram

2.4　RandLA-AFES 语义分割网络

如图 7 所示，网络采用 5 层对称的编码-解码架构。为兼顾效率与精度，仅在编码器第 4、5 层引入自适应显隐式融合模块（AFEM），前 3 层则保留邻域特征采样模块（NFSM）。网络以原始点云数据(N，d)作为输入，其中 N 表示点云中的点数，d 为输入点云特征的维度(坐标值和强度)。输入点云经全连接映射后进入编码器，在渐进式降采样的同时，特征维度从 8 维逐级扩展至 1024 维。解码器通过最近邻插值上采样，并利用级联融合细化编码端特征。最后，融合特征经由 MLP、全连接层及 Dropout 层实现语义标签预测。该架构相较于基线模型增加了一层特征提取层，旨在增强对铁路复杂场景的层次化特征捕获能力。

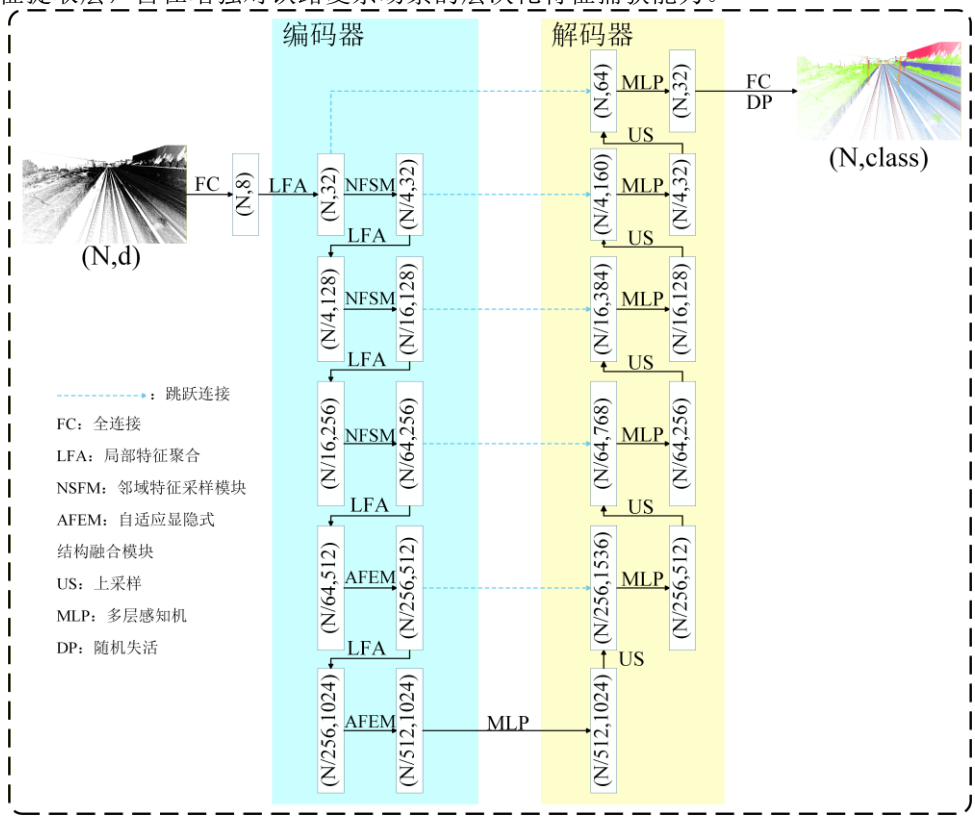

图 7　RandLA-AFES 点云分割网络结构图

Fig.7　RandLA-AFES Point Cloud Segmentation Network Architecture Diagram

3　实验

3.1　数据集设置与初始参数

为了评估所构建的显隐式三维结构自适应融合的三维点云语义分割模型的有效性，选用大规模铁路场景数据集 WHU-Railway3D[14]进行实验。

本次使用的 WHU-Railway3D 是一个专门为铁路场景设计的多样化点云语义分割数据集，基于场景复杂度和语义类分布将铁路分为城市铁路、农村铁路和高原铁路。该数据集跨度约 30 公里，包含 46 亿个点，分为 11 类，如钢轨、道床和建筑等。除了三维坐标，WHU-Railway3D 还提供了丰富的属性信息，如反射强度、扫描角度和返回次数，数据集按照场景类型被划分为 Rural、Plateau 和 Urban 三个子集。统计信息如表 1 所示。

表 1 WHU-Railway3D 数据集

Table 1　WHU-Railway3D Dataset

Dataset	Urban	Rural	Plateau
Length	10.7km	10.6km	10.4km

Points	695M	3475M	444M
Class	11	11	8

其中 Rural 子集共包含 10.6km 的铁路走廊，点云总量 3.5×10^{10} 个，涵盖轨道、地面、支柱、植被等 11 类地物，类别最全、点数最多，故将其用作对比。Plateau 与 Urban 子集分别对应高原、城市两种差异显著的场景，用于跨场景泛化实验，以测试模型在不同环境下的鲁棒性。本次实验系统为 Ubuntu 18.04，硬件为 Intel(R) Xeon(R) Platinum 8358P CPU @2.60GHz 和 NVIDIA RTX3090 GPU(24 GB)，深度学习框架为 PyTorch1.11.0。实验初始学习率设置为 0.001，轮次设置为 100，采用 Adam 优化器，动量因子 β_1，β_2 分别设置为 0.9，0.999，损失函数采用交叉熵损失。

3.2 评价性能指标

为了验证模型的性能，本文采用平均交并比(mean intersection over union,mIoU)和总体准确率(overall accuracy,OA)和 FPS(frames per second)来评价模型性能。

$$OA=\frac{TP+TN}{TP+TN+FP+FN} \tag{12}$$

$$mIoU=\frac{1}{n}\sum_{i=1}^{n}\frac{TP}{TP+FP+FN} \tag{13}$$

其中，TP 表示被预测为正类的正样本，TN 表示被预测为负类的负样本，FP 表示被预测为正类的负样本，FN 表示被预测为负类的正样本，n 为数据集中的类别数。

FPS 是用于衡量模型检测点云速度的常用指标，FPS 数值越高，表示模型的推理速度越快，其计算式如下所示。

$$FPS=\frac{N}{t} \tag{14}$$

其中 N 为点云图像数量，t 为模型检测点云图像所需时间。

3.3 铁路数据集语义分割

为了验证所提算法的有效性，本次实验将本文方法与近年基于点的点云语义分割方法进行比较，初始点云均设置为 100k 个点以应对大规模场景点云输入。

WHU-Railway3D 数据集分割算法分割结果如表 2 所示，本次实验均使用同种训练策略进行了统一训练，在相同实验环境中，本文算法不仅获得了出色的平均交并比，同时保持了领先的推理帧率，综合性能优于现有对比方法。

从表中分析可知，道床、支柱及电杆等几何显著但尺寸有限的类别受益最为明显。显式结构核以原始几何先验补偿下采样造成的细节损失，显著增强了对垂直细长结构的特征提取。自适应融合模块通过动态校准语义与几何特征的权重，抑制冗余并强化判别性表征。相比之下钢轨、地面等大尺度、低曲率区域因显式特征提取模块的过度抑制而出现轻微分割精度回退，但仍处于可接受范围。

整体而言，本方法在基线 RandLA-Net 的基础上将 mIoU 提升 1.2%，且细长结构识别精度优于现主流模型。验证了显隐式特征融合策略在大规模铁路场景点云语义分割中的有效性。

表 2 WHU-Railway3D 数据集语义分割结果

Method	mIoU	OA	钢轨	道床	支柱	支撑	触线	栅栏	电杆	植被	建筑	地面	其他	FPS
PointNet++[17]	28.7	68.4	11.4	20.6	25.2	24.1	51.2	17.7	25.5	65.6	37.2	19.8	17.4	25
SPG[15]	32.5	70.5	17.6	22.8	47.2	18.7	28.1	24.7	27.1	73.6	35.3	38.4	24.0	4
PointCNN[16]	34.4	72.4	16.6	35.9	31.8	40.9	34.7	21.5	42.5	72.3	13.8	41.4	27.0	12
BAAF-Net[17]	38.6	74.7	21.9	40.7	43.1	37.2	45.1	29.8	26.1	75.4	30.8	45.7	28.8	20
SCF-Net[18]	42.6	78.9	30.9	23.2	35.8	33.1	65.6	42.4	30.0	77.8	43.2	58.8	27.8	27
GACNet[19]	56.8	85.4	64.2	63.3	56.5	52.3	55.2	47.1	60.9	80.6	52.0	65.1	25.6	12
PAConv[20]	68.4	92.2	**86.0**	87.7	54.9	63.5	93.1	46.7	62.1	84.5	66.8	81.9	25.2	15
DGCNN[21]	70.4	93.8	79.2	91.7	51.2	68.7	91.3	51.8	68.6	87.1	63.4	**93.4**	28.0	8

KPconv[22]	71.8	94.0	76.9	92.3	49.2	65.4	**93.6**	54.9	81.4	**88.9**	70.3	90.0	26.9	20
RandLA-Net[13]	72.2	94.1	75.6	92.3	60.1	67.9	93.1	41.8	86.4	84.6	71.6	87.3	27.5	**95**
PointMetaBase[23]	72.6	94.1	79.1	90.7	72.0	68.3	83.7	48.0	84.3	86.6	79.2	87.8	30.9	83
PointTransformer[24]	72.9	94.2	77.3	87.7	69.0	77.3	77.1	49.5	88.0	81.4	79.6	81.6	33.4	17
PointVector[25]	73.2	94.3	71.9	86.1	74.8	71.5	75.4	**56.0**	84.4	84.9	77.3	84.4	28.5	42
Ours	73.4	94.4	76.5	**92.8**	**75.5**	65.7	92.8	44.6	**88.3**	87.6	74.3	81.6	27.7	86
PTV3[26]	**73.7**	**94.6**	79.2	89.7	69.5	**76.1**	80.4	50.3	87.2	82.6	**80.1**	81.9	**34.1**	52

Table 2　WHU-Railway3D Semantic Segmentation Results

表 3 跨场景语义分割结果

Table 3　Cross-scenario Semantic Segmentation Results

表 4 Semantic3D 数据集语义分割结果

Table 4　Semantic3D Semantic Segmentation Results

Method	mIoU	OA	man-made.	natural.	high veg.	low veg.	buildings	hard scape	scanning art.	cars
KPconv[11]	74.6	92.9	90.9	82.2	84.2	47.9	94.9	40.0	77.3	79.7
RandLA-Net[13]	77.4	94.8	95.6	91.4	86.6	51.5	95.7	51.5	69.8	76.8
SCF-Net[21]	77.6	94.7	97.1	91.8	86.3	51.2	95.3	50.5	67.9	80.7
RFCR[28]	77.8	94.3	94.2	89.1	85.7	54.4	95.0	43.8	76.2	83.7
Ours	78.8	95.2	96.2	91.2	87.5	55.2	97.2	54.6	72.2	77.6

而后验证模型的跨场景泛化性，与基线模型分别在 Urban 和 Plateau 子数据集上重新进行对比实验，实验结果如表 3 所示。本文方法均取得更高的分割性能，比基线模型分别高 1.7%和 2.8%的分割精度。为了进一步验证本模型的有效性，本文选择在 Semantic3D[27]数据集上进行实验验证，Semantic3D 是一个大规模户外三维点云语义分割数据集，包含了超过 40 亿个经过人工精确标注的点云，其对比方法结果均引用原论文，实验结果如表 4 所示。本文方法在 mIoU 与 OA 指标上，均优于对比模型，取得了较好的分割效果。

3.4　可视化结果分析

在可视化环节，本文从 WHU-Railway3D 中选取了同时包含钢轨、道床、支柱、接触线、栅栏、电杆及植被等典型地物的区段作为代表性场景。

		mIoU	钢轨	道床	支柱	支撑	触线	栅栏	电杆	植被	建筑	地面	其他
Rural	Baseline	72.2	75.6	92.3	60.1	67.9	93.1	41.8	86.4	84.6	71.6	87.3	27.5
	Ours	73.4	76.5	92.8	75.5	65.7	92.8	44.6	88.3	87.6	74.3	81.6	27.7
Urban	Baseline	72.4	66.6	89.5	75.3	46.1	91.9	62.1	43.2	86.0	61.8	89.6	84.3
	Ours	74.1	67.2	89.9	82.3	46.9	91.8	60.5	54.0	84.3	64.8	87.4	86.0
Plateau	Baseline	84.0	85.4	96.5	-	-	89.8	81.6	58.1	-	82.3	99.3	78.7
	Ours	86.8	86.0	96.6	-	-	87.3	81.8	72.4	-	87.2	99.2	83.6

表 2 的分割结果已表明，本方法在道床、支柱和电杆三类上的 IoU 较基线模型提升最为显著。据此，图 8 进一步针对支柱与电杆的预测结果开展可视化对比：将两类误分割区域以红色矩形框标注，并对局部细节进行放大呈现。可以观察到，基线网络倾向于将支柱、电杆与邻近结构混淆，导致几何边界模糊。而本文方法通过显式几何先验与自适应特征耦合，显著增强了对细长垂直结构的判别能力，从而在视觉上实现了更为精准的边缘保持与类别分离。

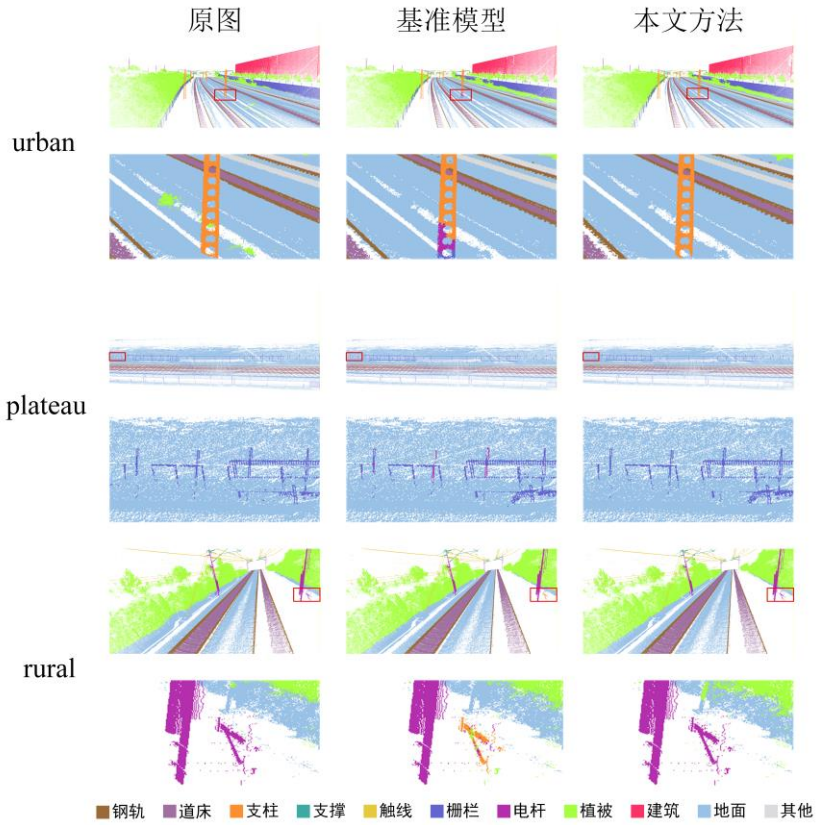

图 8 WHU-Railway3D 语义分割可视化

Fig.8 WHU-Railway3D Semantic Segmentation Visualization

3.5 层级模块划分

考虑到显式结构特征提取模块会进行特定的聚集操作(如求和、平均等)。这些操作需要遍历和操作数据，当数量过大时，会消耗较多的计算资源和时间。从而会影响模型推理速度，因此本文为了在保证整体模型轻量化的前提下最大化改进模块所带来的精度增益，本文对在编码器中各层级是否使用自适应显隐式结构融合模块进行实验。经过多次实验，结果表明本文模型在各项评价指标上的波动极小，模型整体性能稳定，具备良好的工程鲁棒性。实验结果如表 5 所示。

表 5 跨场景语义分割结果不同层级 AFEM 模块表现

Table 5 Performance of AFEM Modules at Different Levels

编码器 前 5 层	参数量 (Paras)	推理速度 (ms/batch)	mIoU
[× , × , × , × , ×]	1.27 M	121	72.2
[√ , × , × , × , ×]	1.32 M	1342	72.4
[× , × , × , × , √]	1.32 M	133	72.6
[√ , √ , × , × , ×]	1.37 M	1576	72.8
[× , × , × , √ , √]	**1.37 M**	**148**	**73.4**
[√ , √ , √ , × , ×]	1.42 M	1833	73.2
[× , × , √ , √ , √]	1.42 M	257	73.3
[√ , √ , √ , √ , ×]	1.47 M	1905	73.5
[× , √ , √ , √ , √]	1.47 M	571	73.6
[√ , √ , √ , √ , √]	1.52 M	1996	73.8

其中×表示当前层仅使用邻域特征采样模块，√表示当前层使用自适应显隐式结构融合模块。推理速度表示推理一个批次所需的时间，时间越小表示推理速度越快。如表 5 实验所示，当在浅层引入自适应显隐式结构融合模块时，由于点云数量庞大，带来了巨大的推理延迟，且浅层特征多为低级特征，mIoU 提升有限。而随着网络加深，点云数量经过下采样大幅减少，此时在深层引入自适应显隐式结构融合模块，能以极小的计算代价显著捕获高维度的几何与语义补充信息。我们可以从表中发现当编码器后两层使用自适应显隐式结构融合模块时，对比精度提升 1.4%，参数量仅增加 0.1M，推理速度相较于原始网络仅增加 27ms，实现了最佳的性能比。而后推理速度呈指数上升，而精度最多仅增加 0.4%。故本文选择编码器后两层使用自适应显隐式结构融合模块。

3.6 消融实验

为系统评估 RandLA-AFES 网络中各模块有效性，本文在 WHU-Railway3D 数据集上以 RandLA-Net 为基线设计了三组递进式消融实验。

表 6 自适应显隐式结构融合模块有效性验证

Table 6 Validation of the Adaptive Implicit-Explicit Structure Merging Module

邻域特征采样模块	显式结构特征提取模块	自适应融合模块	mIoU
√	×	×	72.2
×	√	×	72.6
√	√	×	72.9
√	√	√	73.4

如表 6 所示，首先仅引入显式结构特征提取模块(ESFE)，mIoU 由 72.2%提升至 72.6%。这一提升主要归因于 ESFE 对随机采样缺陷的修正。基线模型使用的随机采样策略极易导致支柱、接触线等稀疏细长物体的几何拓扑结构断裂，而 ESFE 通过并行计算局部协方差等显式特征，强行保留了这些微小目标的几何线性度与垂直度信息，从而在细粒度类别上显著改善了分割边界。继而将 ESFE 与邻域特征采样模块(NFSM)以固定权重进行融合，mIoU 进一步增至 72.9%。这验证了高维语义特征与低维几何特征具有高度互补性，语义特征提供了上下文场景理解，而几何特征提供了精确的形状约束，两者相互融合可提升模型性能。最终嵌入自适应融合模块(AFM)实现动态加权，模型实现了最佳性能，mIoU 达 73.4%，相对基线提升 1.2%。该结果表明了自适应机制在解决铁路场景空间异质性上的关键作用。铁路场景包含大面积平坦区域和细长复杂结构。固定权重融合容易在平坦区域引入高频几何噪声。而 AFM 通过学习特征通道的响应，能够在平坦区域自动降低几何特征权重来抑制噪声，同时在支柱和轨道边缘通过提高几何特征权重来锐化边界。这种自适应机制解决了特征冗余问题，实现了语义与几何表征的最优一致性，从而充分证明了所提模块对大规模铁路场景点云语义分割性能的显著提升作用。

3.7 自适应融合权重统计分析

为了深入理解自适应融合模块在语义特征与几何特征耦合过程中的调节机制，验证其自适应能力的有效性，本文对模型训练全过程中的语义权重参数 α 和几何权重参数 β 进行了统计分析。

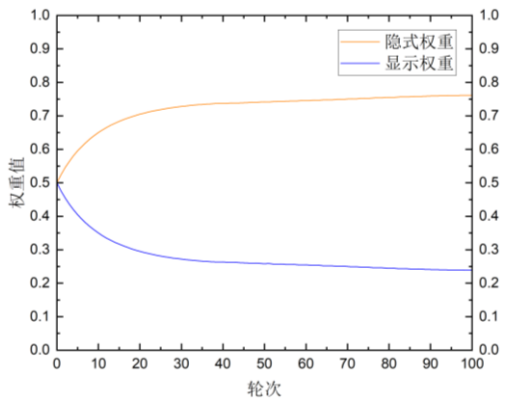

图 9 融合权重训练演变曲线

Fig.9 Evolution of Fusion Weights

如图 9 所示，在训练初始阶段，将α和β的初始值均设为 0.5。随着训练迭代次数的增加，网络通过反向传播算法根据损失函数动态调整权重。在 30 个 Epoch 后，权重逐渐趋于稳定。最终，语义特征权重收敛于 0.76 左右，而几何特征权重收敛于 0.24 左右。这一统计结果说明深层语义特征提供了主要的上下文判别信息，是分类的基础。证明了显式几何特征并非冗余，其作为必要的补充，有效地修正了随机采样带来的结构丢失问题。

为了验证自适应显隐融合机制对铁道场景中细长结构的增益作用，本文选取了道床、支柱、电杆三个典型细长结构类别，对比了不同融合策略与本文自适应权重策略的性能差异。

如图 10 所示，相较于人工设定的固定权重，本文提出的自适应策略在在这些细长结构类别上的增益较为显著，比如在支柱类别上，mIoU 相比于固定权重(0.5:0.5)提升了 6.8%;而电杆类别也提升了 2.8%。

这一现象证明了自适应融合模块的作用机理，对于地面、建筑等大尺度平面物体，单一的语义特征已经足够。但对于几何特征显著的细长物体，模型通过自适应学习，自动赋予了显式几何特征更恰当的关注度，这表明本文提出的自适应机制成功实现了精准地利用显式几何信息来修复细长结构的拓扑完整性。

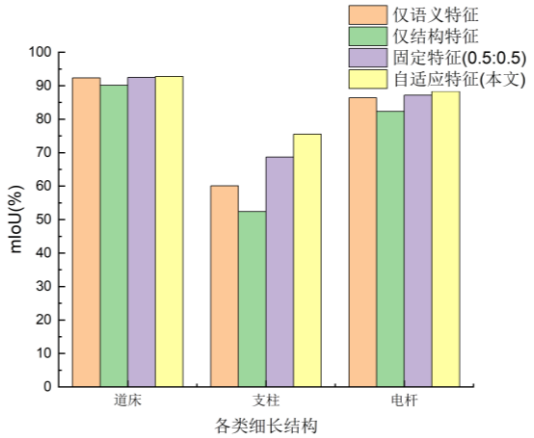

图 10 不同权重策略在细长结构类别上的性能对比

Fig.10 Performance Comparison of Different Weighting Strategies on Slender Structures

4 结束语

针对铁路复杂场景跨域分割精度低的问题，本文以 RandLA-Net 为基线，提出显隐式三维结构自适应融合网络 RandLA-AFES。该网络通过并行引入显式特征提取分支，以极小算力代价校准了降采样过程中的几何拓扑偏差。同时利用自适应融合模块动态优化语义与几何特征权重，实现了多模态特征的最优耦合。实验证明，该方法在 WHU-Railway3D 及其跨场景子集中有效权衡了推理速度与分割精度，并在 Rural 场景取得最优表现。而对于不同铁道场景子数据集 Urban 和 Plateau 相较于基线网络在损失少量推理速度的基础上，在分割精度方面均有一定的提升。后续可以优化采样算法，原始模型由于随机采样而丢弃许多细节，因此可以通过优化采样方法得到对物体更加准确的特征描述，从而提升检测性能。

参考文献：

[1] WU B C, WAN A, YUE X Y, et al. SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud[C]// IEEE International Conference on Robotics and Automation. Piscataway, NJ: IEEE,2018: 1887-1893.

[2] MILIOTO A, VIZZO I, CHLEY J, et al. RangeNet++: Fast and Accurate LiDAR Semantic Segmentation[C]//IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE,2019: 4213-4220.

[3] LAI X, CHEN Y K, LU F B, et al. Spherical Transformer for LiDAR-based 3D Recognition[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE,2023: 17545-1755.

[4] LIU Z J, YANG X Y, TANG H T, et al. FlatFormer: Flattened Window Attention for Efficient Point Cloud Transformer[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE,2023: 1200-1211.

[5] MATURANA D, SCHERER S, IEEE. VoxNet: A 3D Convolutional Neural Network for Real-Time Object Recognition[C]// IEEE/RSJ International Conference on Intelligent Robots and Systems. Piscataway, NJ: IEEE, 2015: 922-928.

[6] ZHU X G, ZHOU H, WANG T, et al. Cylindrical and Asymmetrical 3D Convolution Networks for LiDAR Segmentation[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE,2021: 9934-9943.

[7] LI J L, DAI H, HAN H, et al. MSeg3D: Multi-modal 3D Semantic Segmentation for Autonomous Driving[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 21694-21704.

[8] LI S X, ZHAO Y, LI C J, et al. Fusion-3D: Integrated Acceleration for Instant 3D Reconstruction and Real-Time Rendering [C]// 57th International Symposium on Microarchitecture. Piscataway, NJ: IEEE, 2024: 78-91.

[9] 田明昊, 李攀, 阎肃, 等. 基于二维激光雷达和多视角相机数据级融合的 3D-RGB 点云成像[J].数据采集与处理, 2025,40(6):1505-1517.
TIAN Minghao, LI Pan, YAN Su, et al. 3D-RGB Point Cloud Imaging Based on Data-Level Fusion of 2D LiDAR and Multi‐view Camera[J]. Journal of Data Acquisition and Processing, 2025,40(6):1505-1517.

[10] QI C R, YI L, SU H, et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[C]// The 31st Annual Conference on Neural Information Processing Systems. Piscataway, NJ: IEEE, 2017, 30.

[11] HU Q Y, YANG B, XIE L H, et al. RandLA-Net: Efficient Semantic Segmentation of Large-Scale Point Clouds[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 11105-11114.

[12] SUN S F, RAO Y M, LU J W, et al. X-3D: Explicit 3D Structure Modeling for Point Cloud Recognition[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2024: 5074-5083.

[13] ZHANG M, YOU H X, KADAM P, et al. PointHop: An Explainable Machine Learning Method for Point Cloud Classification[J]. IEEE Trans Multimedia, 2020, 22(7): 1744-1755.

[14] QIU B, ZHOU Y Z, DAI L, et al. WHU-Railway3D: A Diverse Dataset and Benchmark for Railway Point Cloud Semantic Segmentation[J]. IEEE Trans Intell Transp Syst, 2024, 25(12): 20900-20916.

[15] LANDRIEU L, SIMONOVSKY M, IEEE. Large-scale Point Cloud Semantic Segmentation with Superpoint Graphs[C]// 31st IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 4558-4567.

[16] LI Y Y, BU R, SUN M C, et al. PointCNN: Convolution On X-Transformed Points[C]// 32nd Conference on Neural Information Processing Systems. Piscataway, NJ: IEEE, 2018, 31.

[17] QIU S, ANWAR S, BARNES N, et al. Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 1757-1767.

[18] FAN S Q, DONG Q L, ZHU F H, et al. SCF-Net: Learning Spatial Contextual Features for Large-Scale Point Cloud Segmentation [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 14499-14508.

[19] JIANG L, ZHAO H S, LIU S, et al. Hierarchical Point-Edge Interaction Network for Point Cloud Semantic Segmentation [C]// IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 10432-10440.

[20] XU M T, DING R Y, ZHAO H S, et al. PAConv: Position Adaptive Convolution with Dynamic Kernel Assembling on Point Clouds [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 3172-3181.

[21] WANG Y, SUN Y B, LIU Z W, et al. Dynamic Graph CNN for Learning on Point Clouds [J]. ACM Trans Graph, 2019, 38(5): 12.

[22] THOMAS H, QI C R, DESCHAUD J E, et al. KPConv: Flexible and Deformable Convolution for Point Clouds [C]// IEEE/CVF International Conference on Computer Vision. Piscataway, NJ: IEEE, 2019: 6420-6429.

[23] LIN H J, ZHENG X W, LI L J, et al. Meta Architecture for Point Cloud Analysis [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 17682-17691.

[24] ENGEL N, BELAGIANNIS V, DIETMAYER K. Point Transformer [J]. IEEE Access, 2021, 9: 134826-134840.

[25] DENG X, ZHANG W Y, DING Q, et al. PointVector: A Vector Representation In Point Cloud Analysis [C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2023: 9455-9465.

[26] WU X Y, JIANG L, WANG P S, et al. Point Transformer V3: Simpler, Faster, Stronger [C]// 2024 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. Piscataway, NJ: IEEE, 2024: 4840-4851.

[27] HACKEL T, WEGNER J D, SAVINOV N, et al. Large-Scale Supervised Learning For 3D Point Cloud Labeling: Semantic3d.Net [J]. PHOTOGRAMMETRIC ENGINEERING AND REMOTE SENSING, 2018, 84(5): 297-308.

[28] GONG J Y, XU J C, TAN X, et al. Omni-supervised Point Cloud Segmentation via Gradual Receptive Field Component Reasoning

[C]// 2021 IEEE/CVF CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION. Piscataway, NJ: IEEE, 2021:11668-11677.

作者简介：

吴非凡（2000-），男，硕士研究生，研究方向：计算机视觉，点云分割，E-mail: feifan_123@foxmail.com。

巨志勇(1975-)，**通信作者**，男，博士，讲师，研究方向：模式识别，机器视觉，E-mail: juzy@usst.edu.cn。

夏涵锐（2002-）男，硕士研究生，研究方向：计算机视觉，三维重建, E-mail:2422605050@st.usst.edu.cn。

Adaptive Fusion of Explicit-Implicit Structures for Railway Point Cloud Segmentation Network

Wu Feifan, Ju Zhiyong*, Xia Hanrui

（School of Optoelectronic Information & Computer Engineering, University of Shanghai for Science and Technology, Shanghai 200093, China）

Abstract：Existing point cloud semantic segmentation methods for railway scenes struggle to simultaneously balance cross-scenario robustness and object segmentation accuracy. Using RandLA-Net as the basic framework, this paper develops an implicit-explicit 3D structure adaptive fusion network (RandLA-Net with Adaptive Fusion of Explicit Structure, RandLA-AFES) specifically tailored for actual railway engineering scenarios. In the encoder stage, an explicit structural feature extraction module is introduced in parallel, which utilizes a local spatial partitioning strategy to capture structural kernels and generate explicit geometric priors, effectively compensating for the fine-grained structural loss during the downsampling process. Meanwhile, a neighborhood feature sampling module is utilized to aggregate deep local contextual information, ensuring the integrity of the high-dimensional implicit semantic representation of the point cloud. Furthermore, a dynamic adaptive fusion mechanism is designed. By introducing learnable weight parameters, it dynamically calibrates the fusion ratio of explicit geometric features

and implicit semantic features, achieving the optimal coupling of dual-modal features and effectively suppressing redundant noise. Extensive experiments on the public railway scene dataset WHU-Railway3D demonstrate that, compared with the baseline model, the proposed method outperforms existing mainstream methods in metrics such as mean intersection over union (mIoU) and overall accuracy (OA) with only a minimal increase in parameters and inference time. Moreover, it exhibits excellent semantic segmentation capabilities across various complex scenarios, verifying its effectiveness and potential for engineering applications.

Highlights:

1.Proposed RandLA-AFES, a novel network to preserve the topological integrity of thin structures in large-scale railway point clouds.

2.Introduced an Explicit Structure module using spatial moments (PointHop) to efficiently capture linear priors without complex eigen-decomposition.

3.Designed a lightweight Adaptive Fusion mechanism with globally learnable weights to dynamically balance implicit semantics and explicit geometry.

4.Achieved significant IoU gains for challenging categories (e.g., poles, pillars) with minimal computational overhead and robust engineering stability.

Key words: laser point cloud; semantic segmentation; deep learning; railway scene

Foundation item：National Natural Science Foundation of China(No.81101116)。
Received：2026-01-25 **Revised**：2026-07-05
***Corresponding author**, **E-mail:** juzy@usst.edu.cn