



引文格式: 马天恩, 刘涛, 杜萍, 等. 一种聚合全局上下文信息的三维点云语义分割方法[J]. 武汉大学学报(信息科学版), 2025, 50(12): 2548-2559. DOI: 10.13203/j.whugis.20230143

Citation: MA Tianen, LIU Tao, DU Ping, et al. A 3D Point Cloud Semantic Segmentation Method for Aggregating Global Context Information[J]. Geomatics and Information Science of Wuhan University, 2025, 50(12): 2548-2559. DOI: 10.13203/j.whugis.20230143

# 一种聚合全局上下文信息的三维点云语义分割方法

马天恩<sup>1,2,3</sup> 刘涛<sup>1,2,3</sup> 杜萍<sup>1,2,3</sup> 陈朴一<sup>1,2,3</sup> 令振飞<sup>1,2,3</sup>

1 兰州交通大学测绘与地理信息学院, 甘肃 兰州, 730070

2 地理国情监测技术应用国家地方联合工程研究中心, 甘肃 兰州, 730070

3 甘肃省地理国情监测工程实验室, 甘肃 兰州, 730070

**摘要:** 现有基于深度学习的三维点云语义分割方法在很大程度上忽略了全局上下文信息, 未能充分利用点云的局部几何结构、颜色信息和高层语义特征之间的互补性。针对该问题, 提出一种融合局部特征编码和密集连接的点云语义分割模型。首先, 设计一个局部特征提取模块, 使模型能够同时捕获空间几何结构、颜色信息和语义特征; 然后, 结合局部特征聚合模块保留原始点云数据中丰富的几何信息, 减少特征提取过程中几何信息的损失; 最后, 利用密集连接模块聚合全局上下文信息, 实现低层特征和高层语义信息的互补。采用 S3DIS 和 Semantic3D 两个大型基准数据集进行实验验证, 发现所提模型在两个数据集上的平均交并比分别达到 71.8% 和 77.8%。实验结果表明, 所提模型在三维点云语义分割方面具有较好的性能。

**关键词:** 点云语义分割; 深度学习; 局部特征编码; 密集连接; 全局上下文信息

中图分类号: P237

文献标识码: A

收稿日期: 2024-09-08

DOI: 10.13203/j.whugis.20230143

文章编号: 1671-8860(2025)12-2548-12

## A 3D Point Cloud Semantic Segmentation Method for Aggregating Global Context Information

MA Tianen<sup>1,2,3</sup> LIU Tao<sup>1,2,3</sup> DU Ping<sup>1,2,3</sup> CHEN Puyi<sup>1,2,3</sup> LING Zhenfei<sup>1,2,3</sup>

1 School of Geomatics and Geographic Information, Lanzhou Jiaotong University, Lanzhou 730070, China

2 National-Local Joint Engineering Research Center of Technologie and Applications for National Geographic State Monitoring, Lanzhou 730070, China

3 Gansu Province Engineering Laboratory for National Geographic State Monitoring, Lanzhou 730070, China

**Abstract: Objectives:** The deep learning-based 3D point cloud semantic segmentation methods often overlook global contextual information and do not fully leverage the synergy between the local geometric structure, color information, and high-level semantic features of point cloud. It is essential to effectively capture the geometric structure, color variations, and semantic features of point clouds while retaining global context information. **Methods:** This paper proposes a point cloud semantic segmentation model that integrates local feature encoding and dense connectivity. First, a local feature extraction module is employed to enable the model to concurrently capture spatial geometric structure, color information, and semantic features. Second, a local feature aggregation module is incorporated to preserve the rich geometric data within the original point cloud, minimizing information loss during feature extraction. Finally, we utilize a dense con-

**基金项目:** 国家自然科学基金(42261076, 42061060); 甘肃省科技重大专项(22ZD6GA010); 地理信息工程国家重点实验室/自然资源部测绘科学与地球空间信息技术重点实验室联合开放基金(2022-03-03); 兰州交通大学优秀平台支持项目(201806); 兰州交通大学天佑创新团队项目(TY202001)。

**第一作者:** 马天恩, 博士生, 研究方向为三维点云语义识别和场景理解。mte\_shuimu@163.com

**通信作者:** 刘涛, 博士, 教授。liutao@lzjtu.edu.cn

nectivity module to aggregate contextual semantic information, and promote synergy between low-level features and high-level semantic data. **Results:** The proposed model is benchmarked on two large datasets, S3DIS and Semantic3D. The results show that the proposed model achieves an overall accuracy (OA) of 88.3% and mean intersection over union (mIoU) of 71.8% on S3DIS dataset, improving the baseline set of RandLA-Net by 0.3% and 1.8%, respectively. On Semantic3D dataset, we register an OA of 94.9% and an mIoU of 77.8%, marking respective improvements of 0.1% and 0.4% over RandLA-Net. **Conclusions:** The proposed model effectively preserves local geometric and color information through local feature encoding. The local feature aggregation module refines point proximity along boundaries to align with similar feature domains, and dense connections successfully integrate global context and key geometric features. Overall, the proposed model delivers more accurate semantic labels and a superior geometric feature representation, enhancing the precision of local segmentations.

**Key words:** point cloud semantic segmentation; deep learning; local feature encoding; dense connection; global context information

点云已成为继矢量地图和遥感影像之后的第三类重要的时空数据源<sup>[1]</sup>,它能够较好地表达复杂场景和物体的几何形状,在物体空间关系的表达方面也具有独特的优势。点云语义分割是三维场景理解和分析的基础,旨在识别点云场景中每个点的语义标签,在基础测绘<sup>[2]</sup>、智慧城市建设<sup>[3]</sup>、高精度语义提取<sup>[4]</sup>、工程施工<sup>[5]</sup>和自动驾驶<sup>[6]</sup>等多个领域有着广泛的应用。

传统的点云语义分割方法主要包括基于属性聚类<sup>[7]</sup>、模型拟合<sup>[8]</sup>以及区域增长<sup>[9]</sup>等,这些方法通常需要手工设定特征,极大地限制了点云的广泛应用。近年来,由于深度学习在图像语义分割方面取得了重大进展,相关研究也逐步扩展到点云数据处理领域。与二维图像等结构化数据不同,三维点云具有无序性、不均匀性、非结构化等特点,致使传统的卷积神经网络(convolutional neural network, CNN)无法直接处理三维点云数据。因此,基于深度学习的三维点云语义分割已成为当前的研究重点。

目前,基于深度学习的三维点云语义分割方法主要分为三类:基于投影的方法、基于体素的方法和基于点的方法。(1)基于投影的方法是将三维点云投影到二维图像中<sup>[10]</sup>,然后利用成熟的2D深度学习算法进行分割。文献[11]将点云投影到不同视点的图像中,并使用2D卷积神经网络处理图像,然后将不同图像的特征融合成一个简单的形状描述算子用于点云的分割。文献[12]首先将点云转换为RGB和几何合成图像,然后使用改进的SegNet<sup>[13]</sup>和U-Net<sup>[14]</sup>对图像进行语义分割,最后根据点云与图像的对应关系得到点云的分割结果。文献[15]通过球面投影将点云转换为图像,并设计了一个独立学习和融合架

构来解决模态间隙问题。然而,由于数据转换,这些方法通常会丢失一些几何结构信息,从而影响分割性能。(2)基于体素的方法是将不规则的点云转换为规则的体素网格,然后使用3D CNN进行分析和处理。文献[16]将点云转换为 $20^3$ 大小的体素网格,通过两层三维卷积和最大池化操作获得体素特征,然后采用全连接网络和softmax函数得到分割结果。文献[17]提出一种超体素随机森林与长短期记忆神经网络(long short-term memory, LSTM)联合优化的室内点云高精度分类方法,解决三维点云数据分割时分类目标内部不一致的问题。基于体素的方法在体素分辨率较低的情况下容易丢失三维场景的细节信息,不利于小尺度目标的识别。(3)基于点的方法是直接作用于不规则点云实现分割。文献[18]提出了PointNet网络,将提取的点特征和全局特征相结合,输入分类器中得到语义分割结果,但它难以获取点云局部特征以及上下文信息。文献[19]采用一种层次化的编码器-解码器框架改进PointNet网络,进而提出PointNet++网络,并基于 $K$ -最近邻算法提取邻域信息,来增强局部特征的提取。文献[20]提出PointSIFT算法,通过三层卷积运算对不同尺度的多方位空间信息进行叠加编码,然后将多尺度特征进行级联,实现方位编码和尺度感知。文献[21]提出PointWeb使邻域球中的每一个点都能聚集其他点的特征,从而能够提取丰富的几何信息。文献[22]扩展2D密集连接网络到三维点云来学习上下文信息。为了增加点特征之间的区别,文献[23]提出了一种特征增强结构,根据中心点的类别是否与邻域点的类别相同来选择特征聚合路径。一些特征提取模块<sup>[24-26]</sup>针对点云的空间坐标和语义特征进

行编码,将邻域点的信息和质心结合起来提高模型的感知能力,实现局部特征的提取。还有一些深度学习网络<sup>[27-29]</sup>通过聚集局部区域的信息来丰富点的特征,如RandLA-Net<sup>[30]</sup>对空间坐标进行编码来表示局部几何结构,BAAF-Net<sup>[31]</sup>编码语义特征和空间坐标来增强局部上下文信息。

点云的颜色信息以及点间的语义信息对于物体的分类以及分割都是至关重要的。上述方法大多只关注点云数据全局或者局部特征的提取,忽略了全局上下文信息,未能充分利用点云的局部几何结构、颜色信息和高层语义特征之间的互补性。同时,缺乏对颜色信息的感知能力,对点云的颜色信息特征提取能力不足。因此,本文提出一种融合局部特征编码和密集连接模块(local feature extraction and aggregation module, LFEAM-DCM)的网络结构。首先对点的空间坐标、颜色信息和语义特征进行显式编码;然后利用图结构和局部卷积算子学习局部特征,减少特征提取过程中的信息损失;最后利用密集连接模块来聚合局部上下文信息,通过低层和高层特征的融合来捕获全局上下文信息,实现三维点云语义分割。

## 1 研究方法

### 1.1 LFEAM-DCM网络体系结构

本文提出的LFEAM-DCM网络体系结构如图1所示。该网络包含3个关键模块:局部特征提取模块(local feature extraction module, LFEM)、局部特征聚合模块(local feature aggregation module, LFAM)和密集连接模块(dense

connection module, DCM)。由于U-Net<sup>[14]</sup>体系结构简单、分割高效,故LFEAM-DCM采用该架构,也分为编码层和解码层。编码阶段设置5个编码层,前一层输出的点特征作为下一层的输入;每个编码层包括DCM和下采样操作,每个DCM由LFEM和LFAM组成。本文使用随机采样对点云进行下采样,这是因为随机采样的效率高,适合大规模点云场景。3D点云的处理方式与传统CNN处理2D图像的方式类似,按照 $N \rightarrow \frac{N}{4} \rightarrow \frac{N}{16} \rightarrow \frac{N}{64} \rightarrow \frac{N}{256}$ 的顺序逐渐稀疏点云,其中 $N$ 为点的个数。同时,每个点的特征维度按 $16 \rightarrow 32 \rightarrow 128 \rightarrow 256 \rightarrow 512$ 的顺序增加以保留更多信息。解码阶段设置5个解码层,每个解码层包含上采样操作和多层感知机(multi-layer perception, MLP),特征在编码层和解码层之间基于跳跃连接进行传输。最后,使用3个全连接层和一个Dropout层来预测每个点的标签分类结果。

### 1.2 点云局部特征提取

在编码器的逐级下采样过程中,不可避免地会遗漏一些关键点,这可能会导致局部信息的丢失,因此本文采用LFEM来提取局部上下文特征。LFEM由本质信息编码和多维特征编码两个神经单元组成,如图2所示,其中 $xyz$ 表示空间坐标, $rgb$ 表示颜色, $f$ 表示多维语义特征, $P_i^k$ 、 $C_i^k$ 、 $F_i^k$ 分别为空间信息、颜色信息、语义信息, $B_j$ 是本质信息编码单元, $M_j$ 是多维特征编码单元。 $B_j$ 主要是对点云的空间坐标和颜色信息进行编码, $M_j$ 是对点云语义特征进行显式编码,通过两个编码单元实现点云局部特征的提取。

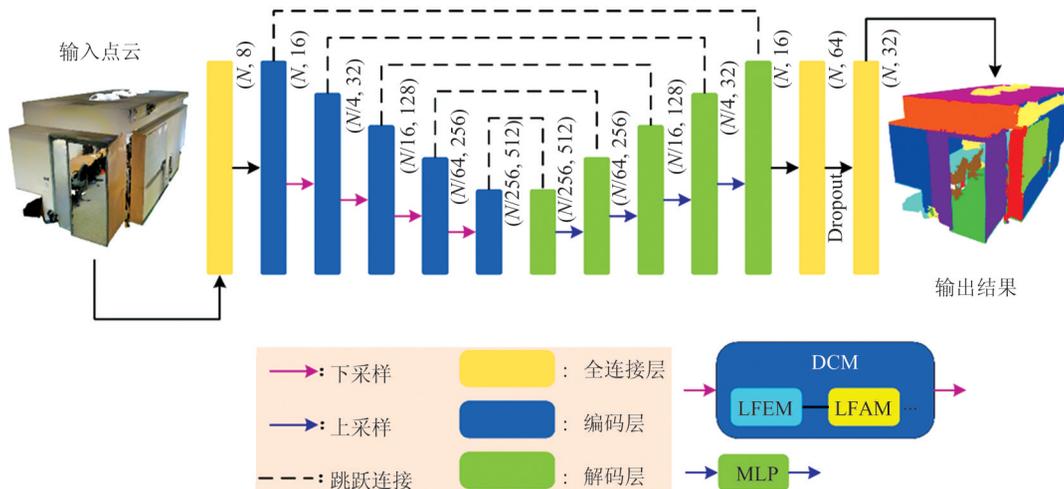


图1 LFEAM-DCM网络体系结构

Fig. 1 Architecture of LFEAM-DCM Network

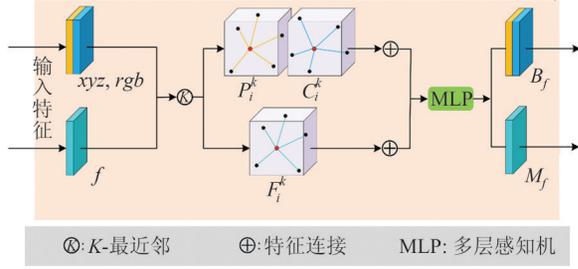


图2 局部特征提取模块

Fig. 2 Local Feature Extraction Module

### 1.2.1 本质信息编码

点云数据通常包含空间坐标和颜色信息,这是点云数据的基础特征。现有的局部特征编码方法通常只关注点云的空间坐标,而颜色信息仅作为初始特征输入到网络中,并未加以充分利用。本质信息编码是对点云的空间坐标和颜色信息进行编码,以学习点云的几何结构和颜色差异,如图3所示。

对于点云  $P$  中的每一个点  $p_i$  和其对应的颜色信息  $c_i$ ,最近  $K$  个点的空间坐标和颜色信息被融合到中心点。这些点是由三维空间中欧氏距离的  $K$ -最近邻算法确定的,确保了上下文信息的提取。为了合理编码空间坐标,同时保留点云的局部几何结构,本文采用 RandLA-Net<sup>[30]</sup>的方法,将中心点及其邻近点的绝对位置  $p_i, p_i^k$ ,以及二者的相对距离  $p_i - p_i^k$  和绝对距离  $\|p_i - p_i^k\|$  作为空间

信息,具体计算式为:

$$P_i^k = (p_i; p_i^k; p_i - p_i^k; \|p_i - p_i^k\|) \quad (1)$$

式中,  $\| \quad \|$  表示计算三维空间中的欧氏距离。受文献[32]的启发,与空间坐标的编码类似,本文将中心点及其邻近点对应的颜色信息  $c_i, c_i^k$ ,以及二者的相对色差  $c_i - c_i^k$  和绝对色差  $\Delta c_i$  结合起来,其中  $\Delta c_i = \text{sum}(r_i + g_i + b_i) - \text{sum}(r_i^k + g_i^k + b_i^k)$ 。颜色信息可表示为:

$$C_i^k = (c_i; c_i^k; c_i - c_i^k; \Delta c_i) \quad (2)$$

将点云的空间坐标和颜色信息作为一组 6D 向量输入到网络中,意味着空间坐标和颜色信息在实际计算过程中作为一个变量整体,可称为点云的本质信息。为了解决整体形态信息利用不足的问题,本文提出本质信息编码包括空间坐标编码和颜色信息编码,将中心点及其邻近点的本质信息  $I_i = (p_i; c_i), I_i^k = (p_i^k; c_i^k)$ ,以及二者的相对差  $I_i - I_i^k$  和绝对差  $\Delta I_i = (\|p_i - p_i^k\|; \Delta c_i)$  连接起来。本质信息编码可表示为:

$$B_f = (I_i \oplus I_i^k \oplus (I_i - I_i^k) \oplus \Delta I_i) \quad (3)$$

式中,  $\oplus$  表示特征连接操作。本质信息编码包含空间坐标编码和颜色信息编码,能够同时保留局部几何结构和局部颜色信息,使网络更加有效地提取可区分的局部特征,显著缓解随机下采样操作造成的局部信息损失。

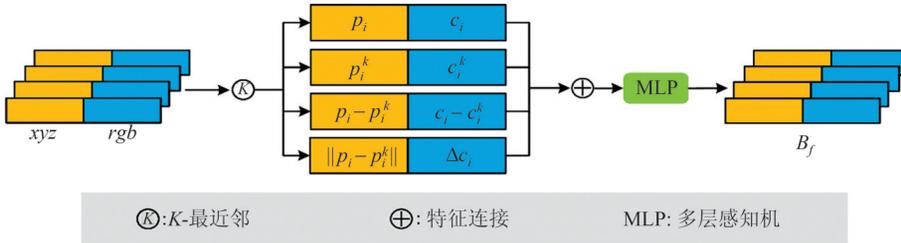


图3 本质信息编码

Fig. 3 Essential Information Encoding

### 1.2.2 多维特征编码

多维特征编码对点云语义特征  $f$  进行编码,使网络能够充分利用局部语义特征。多维特征编码对每个点  $p_i$  对应的语义特征  $f_i$  和邻近点  $p_i^k$  对应的语义特征  $f_i^k$  进行编码,如图4所示。将中心点及其邻近点对应的语义特征  $f_i, f_i^k$ ,以及相对特征差  $f_i - f_i^k$  结合,得到的语义信息可表示为:

$$F_i^k = (f_i; f_i^k; (f_i - f_i^k)) \quad (4)$$

形成的多维特征编码可表示为:

$$M_f = (f_i \oplus f_i^k \oplus (f_i - f_i^k)) \quad (5)$$

与本质信息编码不同,多维特征编码不使用绝对差值,避免由于特征维数过高导致的计算资源消耗。

### 1.3 点云局部特征聚合

LFEM 的输出包含两组编码,分别是本质信息编码  $B_f$  和多维特征编码  $M_f$ ,LFAM 能够有效聚合两种编码,充分利用点云的局部信息,如图5所示。给定点云  $P = \{p_1, p_2, \dots, p_N\} \in \mathbb{R}^{N \times K \times 6}$ ,其中  $N$  为输入点的数量,  $K$  为相邻点的个数。首先将点云的空间坐标和颜色信息作为输入,进行

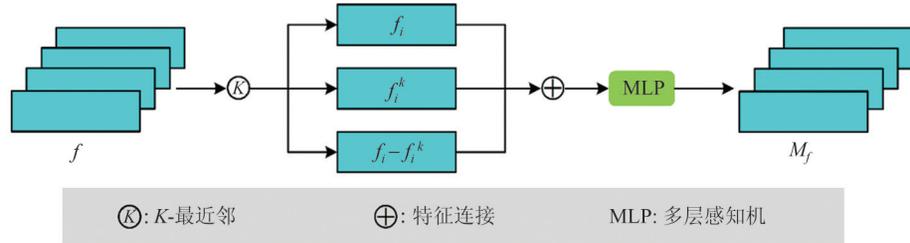


图4 多维特征编码

Fig. 4 Multi-dimensional Feature Encoding

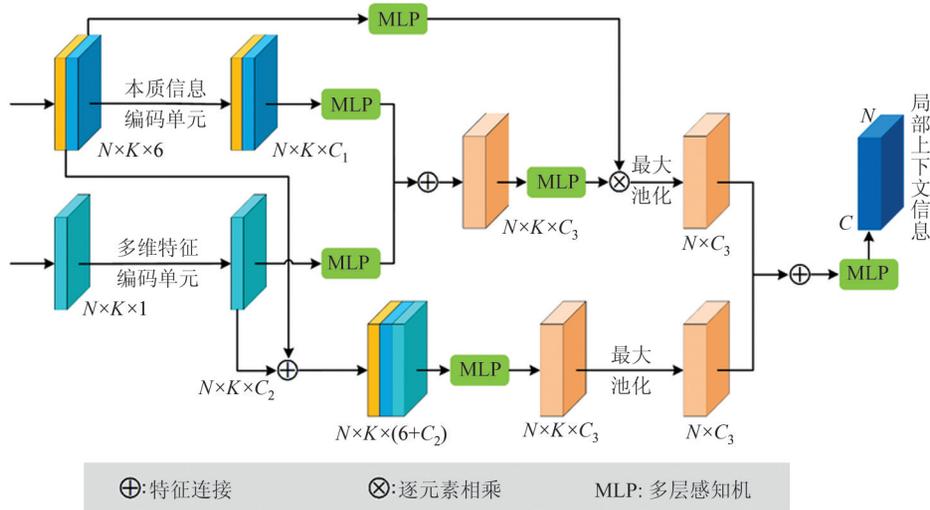


图5 局部特征聚合模块

Fig. 5 Local Feature Aggregation Module

编码得到特征图  $F_1 = \{f_1, f_2, \dots, f_N\} \in \mathbb{R}^{N \times K \times C_1}$ ; 然后对点云的语义特征进行编码, 得到特征图  $F_2 = \{f'_1, f'_2, \dots, f'_N\} \in \mathbb{R}^{N \times K \times C_2}$ ; 最后将 MLP 分别应用于特征图  $F_1$  和特征图  $F_2$ , 通过特征连接得到特征图  $F_3 = \{f''_1, f''_2, \dots, f''_N\} \in \mathbb{R}^{N \times K \times C_3}$ 。由此得到新的本质信息编码和多维特征编码分别是:

$$\hat{B}_f = m(I \oplus I_i^k \oplus (I_i - I_i^k) \oplus \Delta I_i) \quad (6)$$

$$\hat{M}_f = m(f_i \oplus f_i^k \oplus (f_i - f_i^k)) \quad (7)$$

式中,  $m(\cdot)$  表示 MLP, 用于所有点之间共享。

上述操作主要用于聚集每个点的局部特征, 但是这种方式忽略了邻域内各点之间的关系。为此, 本文基于 PointConv<sup>[27]</sup> 设计了局部邻域的卷积算子, 采用邻域内的坐标点作为 MLP 的输入权重函数, 与特征图  $F_3$  进行逐元素相乘, 采用最大池化操作得到特征图  $F'_3 \in \mathbb{R}^{N \times C_3}$ , 其计算过程如下:

$$\check{M}_f = \text{MaxPool}(m(p_i^k) \cdot [\hat{B}_f, \hat{M}_f]) \quad (8)$$

式中,  $\text{MaxPool}(\cdot)$  表示最大池化操作, 它用于保留特征图像中最突出的特征;  $[\cdot]$  表示特征图的

串联。同时使用邻域内的坐标点作为 MLP 的权重函数, 与特征图  $F_2$  进行特征连接, 得到  $F'_2 \in \mathbb{R}^{N \times K \times (6 + C_2)}$ ; 采用 MLP 来学习映射  $\mathbb{R}^{N \times K \times (6 + C_2)} \rightarrow \mathbb{R}^{N \times K \times C_3}$ , 并采用最大池化操作得到  $F''_2 \in \mathbb{R}^{N \times C_3}$ ; 将  $F''_2 \in \mathbb{R}^{N \times C_3}$  和  $F'_3 \in \mathbb{R}^{N \times C_3}$  进行特征连接, 然后通过 MLP 学习得到特征图  $F_{\text{out}} \in \mathbb{R}^{N \times C}$ , 聚合过程如下:

$$\check{M}'_f = \text{MaxPool}(m([\hat{B}_f, \hat{M}_f])) \quad (9)$$

$$\check{M}_f = m([\check{M}_f, \check{M}'_f]) \quad (10)$$

LFAM 利用图结构和局部卷积算子来学习局部邻域的空间坐标、颜色信息和语义特征, 通过对局部特征加权, 进一步聚合局部上下文信息, 以增强特征对局部邻域的表达力。

#### 1.4 全局上下文信息聚合

DCM 的主要作用是增大特征提取的接受域。已有研究<sup>[33-34]</sup>大多参考 ResNet<sup>[26]</sup> 结构, 通过不断叠加网络层来增加特征提取的接受域, 但易导致原有几何信息丢失, 并产生巨大的网络参数。因此, 本文使用 DCM 来反映目标对象与环境之间的依赖性, 聚合上下文语义信息。聚合的信息可表示为:

$$f^l = A\left(\Psi\left([f_{i1}^0, \dots, f_{i1}^{l-1}]\right), \Psi\left([f_{i2}^0, \dots, f_{i2}^{l-1}]\right), \dots, \Psi\left([f_{ij}^0, \dots, f_{ij}^{l-1}]\right), \dots\right) \quad (11)$$

式中,  $f^l$  表示  $l$  层的特征;  $f_{ij}^{l-1}$  是  $f^l$  的上层特征;  $\Psi(\cdot)$  表示映射函数;  $A(\cdot)$  表示聚合函数。  $f^l$  可以利用 DCM 来提取多层次特征。

为了进一步说明 DCM 的优点, 比较了传统的核点卷积神经网络(kernel point convolutional neural network, KPCNN)与 DCM 之间的区别, 如图 6 所示, 其中  $N$  表示输入点的数量,  $C$  表示输入特征维度, 即原始特征通道数,  $T$  表示中间层的特征维度, 即卷积操作输出后的特征通道数。

KPCNN 通过逐层线性连接从低层到高层学习特征, 把上一层输出的特征作为下一层的输入, 最后获得高层语义信息。虽然该方法也可以获得更大的接受域, 但在逐层特征提取过程中可能会丢失几何细节。而 DCM 通过在不同层次之间建立连接, 不仅可以增加特征提取的接受域, 还可以获得不同层次的原始几何信息。随着网络层的逐渐加深, 深层上下文信息可以提高高级语义的表达。

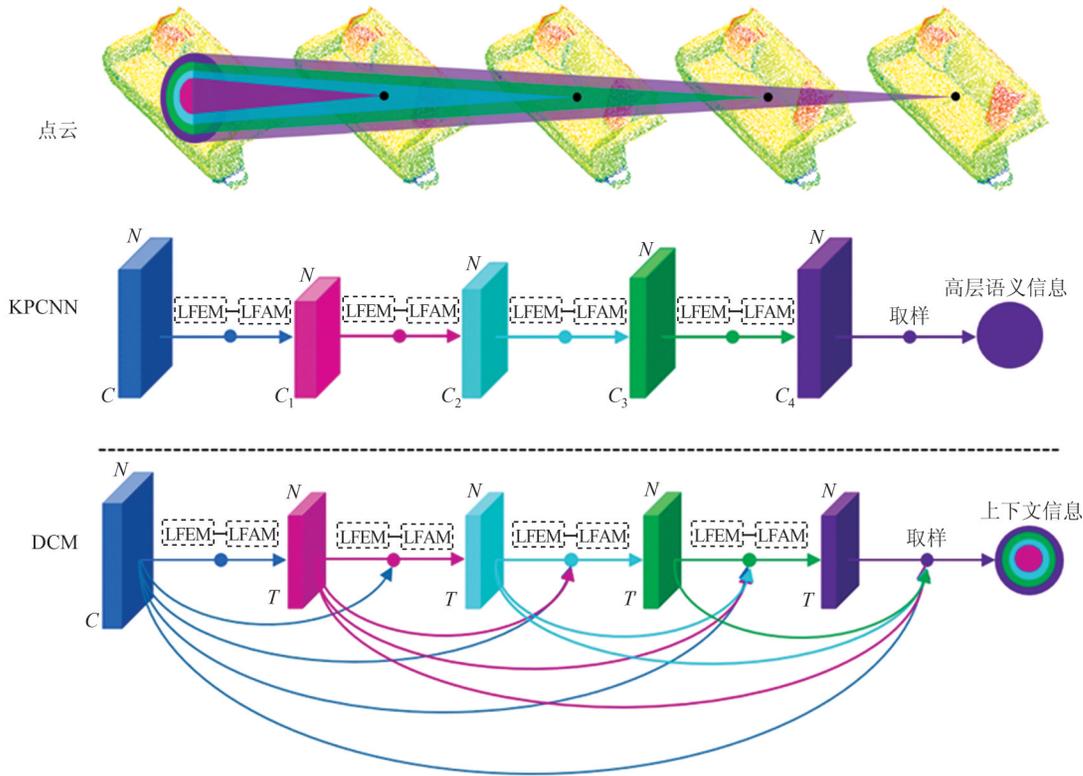


图 6 核点卷积神经网络与密集连接模块对比

Fig. 6 Comparison of Kernel Point Convolutional Neural Network and Dense Connection Module

本文利用 DCM 对多尺度特征进行深度提取, 获得丰富的语义信息。其中低层次特征通常表达的是局部的结构细节, 而高层次特征则包含抽象的语义信息。随着网络层的逐渐加深, 深层的上下文信息反过来可以改进高层语义信息的表达。此外, 网络的每一层只学习具有较少特征维度  $T$  的特征, 从而减少了参数的数量。通过 DCM 在不同编码层之间建立连接, 将不同层输出的局部上下文信息进行聚合, 最后捕获到全局上下文信息。

## 2 实验与分析

### 2.1 实验数据

本文选用室内数据集 S3DIS<sup>[35]</sup> 和室外数据

集 Semantic3D<sup>[36]</sup> 进行实验。S3DIS 数据集由 6 个室内区域组成, 共有 272 个房间, 包含 13 个语义标签, 其中每个点都表示为 6 维向量, 即空间坐标和颜色信息。Semantic3D 数据集是一个大型室外 3D 点云数据集, 包含 8 个语义标签, 分为 15 个训练场景和 15 个在线测试场景, 总点数超过  $40 \times 10^9$ 。

### 2.2 评估指标及参数设置

为了充分评估 LFEAM-DCM 网络在点云语义分割中的有效性, 分别在 S3DIS 和 Semantic3D 数据集进行定量和定性分析。定量分析使用总体精度 (overall accuracy, OA) 和平均交并比 (mean intersection over union, mIoU) 作为评估指标。OA 表示预测正确的类别在整体输入类别中

所占的比例,用于评估整体分割的准确性;交并比(intersection over union, IoU)是每个类别真实值与预测值的交集/并集,用于评估所有对象类别分割的准确性,mIoU表示所有类别的平均交并比。计算式分别为:

$$E_{OA} = \frac{\sum_{i=0}^k c_{ii}}{\sum_{i=0}^k \sum_{j=0}^k c_{ij}} \times 100\% \quad (12)$$

$$E_{IoU} = \frac{c_{ii}}{\sum_{j=0}^k c_{ij} + \sum_{j=0}^k (c_{ji} - c_{ii})} \times 100\% \quad (13)$$

$$E_{mIoU} = \frac{\sum_{i=0}^k E_{IoU_i}}{k+1} \times 100\% \quad (14)$$

式中, $c_{ii}$ 表示真实类别*i*预测为类别*i*的点数; $c_{ij}$ 表示真实类别*i*预测为类别*j*的点数; $c_{ji}$ 表示真实类别*j*预测为类别*i*的点数; $k$ 是数据集中语义类别的数量。

所有实验均在装有Ubuntu20.04操作系统的工作站上进行,该工作站配备Inter(R) Core(TM) i9-10900K @3.70 GHz处理器、64 GB的内存、1 TB固态硬盘以及NVIDIA GeForce RTX 3090 24 GB显卡用以加快模型训练,所有网络模型均在Python 3.6+TensorFlow\_gpu 2.6.0+CUDA 11.4的深度学习环境下实现。使用Adam优化器和交叉熵损失函数对网络进行训练,训练期间批大小设置为4,历元设置为100,初始学习

率设置为0.01, $K$ -最近邻算法的搜索范围设置为16。

### 2.3 室内场景语义分割

为了验证LFEAM-DCM网络的有效性,本文选取PointNet<sup>[18]</sup>、SPG<sup>[37]</sup>、PointCNN<sup>[38]</sup>、PointWeb<sup>[21]</sup>、ShellNet<sup>[39]</sup>、KPCConv<sup>[33]</sup>、BAF-LAC<sup>[25]</sup>以及RandLA-Net<sup>[30]</sup>等8种经典网络作为对比方法,在室内S3DIS点云数据集Area5上进行测试,定量分析结果如表1所示。由表1可知,与RandLA-Net<sup>[30]</sup>和其他深度学习方法相比,LFEAM-DCM在墙壁、梁和板类别上实现了最佳性能,同时获得了88.3%的OA和71.8%的mIoU,相较于RandLA-Net分别提高了0.3%和1.8%。此外,在全部13个类别中,有7个类别(墙壁、梁、柱、门、沙发、书柜和板)的分类结果优于基线方法RandLA-Net。

LFEAM-DCM和RandLA-Net在S3DIS数据集上的定性结果比较如图7所示。其中,黑色圆圈显示出RandLA-Net中分割错误或边界分割不显著的区域。从结果来看,本文方法在分割书架、板和墙壁类别上表现出较好的性能,相较于RandLA-Net能更平滑和准确地分割出对象的目标边界。这主要是因为LFEM通过局部特征编码来保留局部几何结构和颜色信息,而不同物体之间的颜色差异明显,LFEM通过编码局部颜色信息可以捕获这种颜色差异。同时,LFAM进一步加深了局部特征的学习,使得边界周围的点更接近具有相似特征的邻域点。

表1 S3DIS数据集Area5的实验结果/%

Table 1 Experimental Results on Area5 of S3DIS Dataset/%

方法	OA	mIoU	IoU												
			屋顶	地板	墙壁	梁	柱	窗户	门	桌子	椅子	沙发	书柜	板	其他
PointNet <sup>[18]</sup>	78.6	47.6	88.0	88.7	69.3	42.4	23.1	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
SPG <sup>[37]</sup>	86.4	62.1	89.9	95.1	76.4	62.8	47.1	55.3	68.4	73.5	69.2	63.2	45.9	8.7	52.9
PointCNN <sup>[38]</sup>	88.1	65.4	94.8	97.3	75.8	63.3	51.7	58.4	57.2	71.6	69.1	39.1	61.2	52.2	58.6
PointWeb <sup>[21]</sup>	87.3	66.7	93.5	94.2	80.8	52.4	41.3	64.9	68.1	71.4	67.1	50.3	62.7	62.2	58.5
ShellNet <sup>[39]</sup>	87.1	66.8	90.2	93.6	79.9	60.4	44.1	64.9	52.9	71.6	84.7	53.8	64.6	48.6	59.4
KPCConv <sup>[33]</sup>	—	70.6	93.6	92.4	83.1	63.9	54.3	66.1	76.6	57.8	64.0	69.3	74.9	61.3	60.3
BAF-LAC <sup>[25]</sup>	88.2	71.7	92.5	95.9	81.3	63.2	57.8	63.0	79.9	70.3	74.6	60.6	67.2	65.3	60.4
RandLA-Net <sup>[30]</sup>	88.0	70.0	93.1	96.1	80.6	62.4	48.0	64.4	69.4	69.4	76.4	60.0	64.2	65.9	60.1
LFEAM-DCM	88.3	71.8	93.6	95.6	83.4	64.5	51.4	64.2	75.2	68.3	73.5	65.8	70.1	67.4	60.3

### 2.4 室外场景语义分割

在室外Semantic3D点云数据集上,LFEAM-DCM与多种经典算法进行场景语义分割,其定量结果如表2所示。由表2可知,与RandLA-Net<sup>[30]</sup>及其他深度学习方法相比,LFEAM-DCM

在建筑(IoU为96.4%)和汽车(IoU为81.3%)类别上实现了最佳性能。同时,获得了94.9%的OA和77.8%的mIoU,相较于RandLA-Net分别提高了0.1%和0.4%,同时在地面、草坪、树木和扫描伪影类别上性能提升显著。

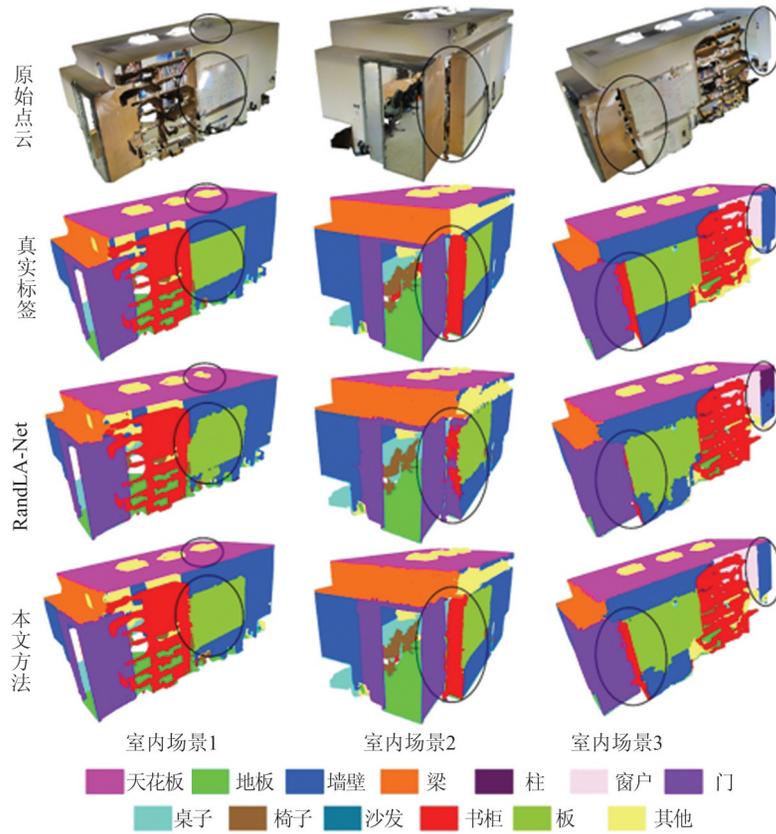


图 7 S3DIS 数据集上的语义分割结果

Fig. 7 Semantic Segmentation Results on S3DIS Dataset

表 2 Semantic3D 数据集 reduced-8 测试数据上的实验结果/%

Table 2 Experimental Results on reduced-8 Test Data of S3DIS Dataset/%

方法	OA	mIoU	IoU							
			地面	草坪	树木	灌木	建筑	花坛	扫描伪影	汽车
SnapNet <sup>[12]</sup>	88.6	59.1	82.0	77.3	79.7	22.9	91.1	18.4	37.3	64.4
SEGCloud <sup>[40]</sup>	88.1	61.3	83.9	66.0	86.0	40.5	91.1	30.9	27.5	64.3
ShellNet <sup>[39]</sup>	93.2	69.3	96.3	90.4	83.9	41.0	94.2	34.7	43.9	70.2
GACNet <sup>[41]</sup>	91.9	70.8	86.4	77.7	88.5	60.6	94.2	37.3	43.5	77.8
SPG <sup>[37]</sup>	94.0	73.2	97.4	92.6	87.9	44.0	83.2	31.0	63.5	76.2
KPConv <sup>[33]</sup>	92.9	74.6	90.9	82.2	84.2	47.9	94.9	40.0	77.3	79.7
RandLA-Net <sup>[30]</sup>	94.8	77.4	95.6	91.4	86.6	51.5	95.7	51.5	69.8	76.8
LFEAM-DCM	94.9	77.8	97.2	92.3	88.2	51.3	96.4	43.6	71.9	81.3

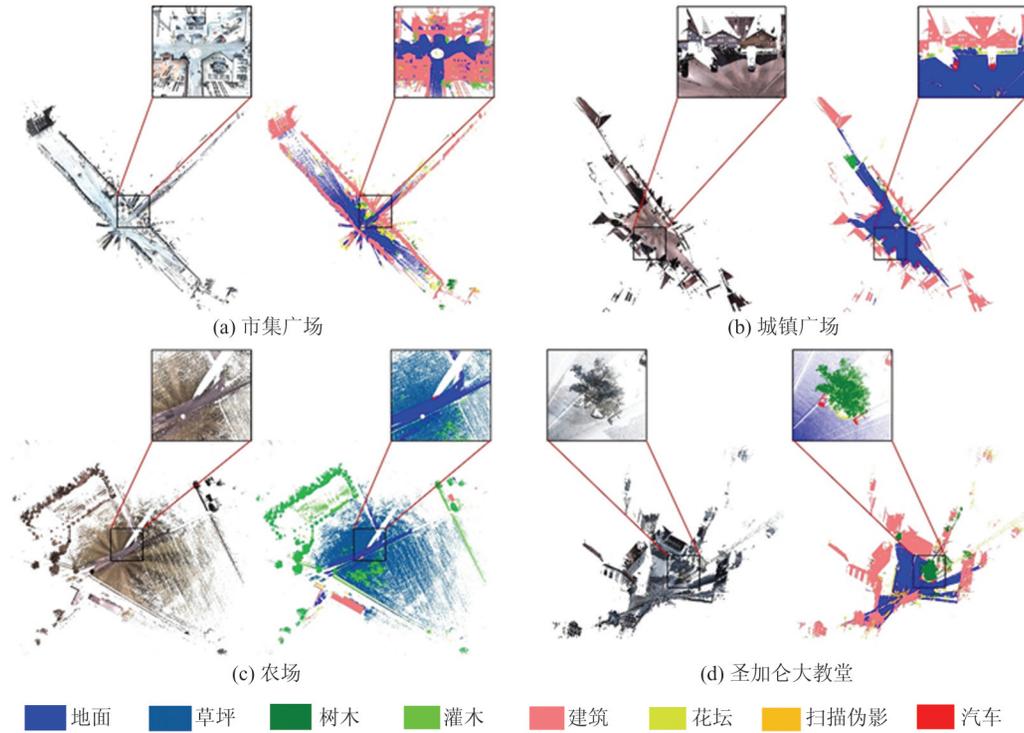
LFEAM-DCM 在 Semantic3D 的 reduced-8 测试数据上的定性结果如图 8 所示,其中每个子图的左、右分别对应真实场景 RGB 点云和语义分割点云,黑色框显示局部详细结果。从图 8 可以看出,LFEAM-DCM 可以将大多数类别准确分割,尤其是建筑物和汽车。之所以能准确地分割,得益于本文提出的 DCM 对多尺度特征进行深度提取,获得丰富的上下文语义信息。此外,有些类别之间没有明显的几何差异,如灌木和树木。因为它们具有相似的结构,所以很难清晰、准确地分割这些类别的边界,但它们之间存在颜

色差异,通过对颜色信息进行编码来学习质心和邻域之间的颜色差异。因此利用颜色信息编码可以使网络清晰、准确地分割出这些类别的边界,提取出可区分的特征,显著提高分割性能。

### 3 消融实验

#### 3.1 LFEAM-DCM 网络模型的消融

为了进一步说明 LFEAM-DCM 网络中各模块的有效性,在 S3DIS 数据集的 Area5 测试集上进行消融实验,结果如表 3 所示。3 项消融实验分



注:每个子图的左、右图分别对应真实场景RGB点云和语义分割点云。

图8 Semantic3D基准数据集 reduced-8测试数据上的语义分割结果

Fig. 8 Semantic Segmentation Results on reduced-8 Test Data of Semantic3D Dataset

别为:(1)移除LFEM;(2)移除DCM,即只需将前一层的输出作为下一层的输入;(3)移除LFAM。

表3 LFEAM-DCM网络的消融实验

Table 3 Ablation Experiment of LFEAM-DCM

模型编号	模块			mIoU/%
	LFEM	LFAM	DCM	
1		✓	✓	57.4
2	✓	✓		61.3
3	✓		✓	62.1
4	✓	✓	✓	64.8

由表3可知,模型4的mIoU相较于模型1高出7.4%,说明移除LFEM对网络性能影响最大,原因在于该模块可以有效地从局部邻域中学习点的特征,从而使每个点的特征得到一定程度的增强。模型4的mIoU相较于模型3高出2.7%,说明LFAM采用邻域内的坐标点作为MLP的输入权重函数,通过最大池化操作能够进一步聚合局部特征,增强特征对局部邻域的描述能力。比较模型2和模型4可以发现,模型4的mIoU相较于模型2高出3.5%,说明DCM能够捕获到全局上下文信息和有价值的几何特征,进而提升了网络模型的语义分割性能。

### 3.2 LFEM的消融

为了验证LFEM的有效性,本文采用不同编码方法进行了实验,研究空间坐标编码、颜色信息编码和多维语义特征编码对LFEM的影响,实验结果如表4所示。

表4 局部特征提取模块的消融实验。

Table 4 Ablation Studies About Local Feature

Extraction Module		
模型	属性编码组合	mIoU/%
$M_0$	<i>rgb</i>	58.9
$M_1$	<i>xyz</i>	60.7
$M_2$	<i>rgb + f</i>	61.6
$M_3$	<i>xyz + f</i>	62.4
$M_4$	<i>xyz + rgb</i>	63.3
$M_5$	<i>xyz + rgb + f</i>	64.8

由表4可知,空间坐标编码对网络的分割性能影响最大,这是因为它能够有效保留点云的局部几何信息;颜色信息编码对网络的分割性能影响次之,虽然获取颜色信息时会产生一些噪声,但颜色信息的编码仍然可以学习局部色差;多维语义特征编码对网络的分割性能影响与颜色信息编码相当。因此,空间坐标编码、颜色信息编码和多维语义特征编码在点云语义分割中能够有效提高网络的分割性能。

## 4 结 语

针对大规模点云场景语义分割问题,本文提出一种新的端到端的深度神经网络 LFEAM-DCM。首先,采用 LFEM 对空间坐标、颜色信息和语义信息进行显式编码,利于网络同时捕获空间几何结构、颜色信息和语义特征;然后,采用 LFAM 来保留原始点云数据中丰富的几何细节,减少特征提取过程中的信息损失;最后,提出一个 DCM,通过关注低层(局部)和高层(全局)特征,将低层的局部特征和高层抽象的语义信息进行连接,聚合局部上下文信息,实现不同层级的特征互补,得到全局上下文信息。在 S3DIS 和 Semantic3D 两个不同基准数据集上对所提模型进行了实验,结果表明,LFEAM-DCM 在 S3DIS 数据集上的 OA 为 88.3%,mIoU 为 71.8%;在 Semantic3D 数据集上的 OA 为 94.9%,mIoU 为 77.8%。与几种深度学习方法的结果对比表明,LFEAM-DCM 具有更准确的语义标注和更强大的几何特征表达能力,对局部细节的分割更加精确。未来将进一步研究更有效的大规模点云场景语义分割方法,如通过添加相对角度特征并与其他初始特征相结合作为模型的输入,最大限度地利用点间的相对几何特征来增强局部上下文等。

## 参 考 文 献

- [1] 张云生,王旭颖,陈斯颺,等. 点云智能理解:从点云中心至语言中心的转变[J]. 武汉大学学报(信息科学版), 2025, DOI: 10.13203/j.whugis20240497.
- ZHANG Yunsheng, WANG Xuying, CHEN Siyang, et al. Point Cloud Intelligent Understanding: Transition from Point-Centric to Language-Centric Paradigm [J]. [J]. *Geomatics and Information Science of Wuhan University*, 2025, DOI: 10.13203/j.whugis20240497.
- [2] 张永军,邹思远,刘欣怡. 稀疏点云引导的航空影像数字表面模型生成方法[J]. 武汉大学学报(信息科学版), 2023, 48(11): 1854-1862.
- ZHANG Yongjun, ZOU Siyuan, LIU Xinyi. Sparse Point Cloud Guided Digital Surface Model Generation for Aerial Images[J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(11): 1854-1862.
- [3] PARK Y, GULDMANN J M. Creating 3D City Models with Building Footprints and LiDAR Point Cloud Classification: A Machine Learning Approach [J]. *Computers, Environment and Urban Systems*, 2019, 75: 76-89.
- [4] 李庞胤,米晓新,丁鹏辉,等. 融合车载影像与点云的道路边界提取与矢量化[J]. 武汉大学学报(信息科学版), 2024, 49(4): 631-639.
- LI Pangyin, MI Xiaoxin, DING Penghui, et al. Fusion of Vehicle-Mounted Imagery and Point Cloud for Road Boundary Extraction and Vectorization [J]. *Geomatics and Information Science of Wuhan University*, 2024, 49(4): 631-639.
- [5] 毛庆洲,徐浩轩,朱璟,等. 利用高密度点云现场检测在建隧道超欠挖[J]. 武汉大学学报(信息科学版), 2025, 50(3): 562-570.
- MAO Qingzhou, XU Haoxuan, ZHU Jing, et al. Immediate Detection of Over-Under Break of Tunnel Under Construction Using High-Density Point Cloud [J]. *Geomatics and Information Science of Wuhan University*, 2025, 50(3): 562-570.
- [6] MILIOTO A, VIZZO I, BEHLEY J, et al. RangeNet ++: Fast and Accurate LiDAR Semantic Segmentation [C]//IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Macau, China, 2020.
- [7] CASTILLO E, LIANG J, ZHAO H K. Point Cloud Segmentation and Denoising via Constrained Nonlinear Least Squares Normal Estimates [M]//Innovations for Shape Analysis: Models and Algorithms. Berlin, Heidelberg: Springer, 2013: 283-299.
- [8] 黄绪洲,赵建虎. 一种联合区域生长和布料模拟的点云滤波算法[J]. 测绘地理信息, 2024, 49(5): 24-27.
- HUANG Xuzhou, ZHAO Jianhu. A Point Cloud Filtering Algorithm for Combining Region Growth and Cloth Simulation [J]. *Journal of Geomatics*, 2024, 49(5): 24-27.
- [9] VO A V, TRUONG-HONG L, LAEFER D F, et al. Octree-Based Region Growing for Point Cloud Segmentation [J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2015, 104: 88-100.
- [10] 熊汉江,郑先伟,丁友丽,等. 基于2D-3D语义传递的室内三维点云模型语义分割[J]. 武汉大学学报(信息科学版), 2018, 43(12): 2303-2309.
- XIONG Hanjiang, ZHENG Xianwei, DING Youli, et al. Semantic Segmentation of Indoor 3D Point Cloud Model Based on 2D-3D Semantic Transfer [J]. *Geomatics and Information Science of Wuhan University*, 2018, 43(12): 2303-2309.
- [11] SU H, MAJI S, KALOGERAKIS E, et al. Multi-View Convolutional Neural Networks for 3D Shape

- Recognition[C]//IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 2016.
- [12] BOULCH A, GUERRY J, LE SAUX B, et al. SnapNet: 3D Point Cloud Semantic Labeling with 2D Deep Segmentation Networks[J]. *Computers & Graphics*, 2018, 71: 189-198.
- [13] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(12): 2481-2495.
- [14] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional Networks for Biomedical Image Segmentation [M]//Medical Image Computing and Computer-Assisted Intervention. Cham: Springer, 2015: 234-241.
- [15] XIAO A R, YANG X F, LU S J, et al. FPS-Net: A Convolutional Fusion Network for Large-Scale LiDAR Point Cloud Segmentation[J]. *ISPRS Journal of Photogrammetry and Remote Sensing*, 2021, 176: 237-249.
- [16] HUANG J, YOU S Y. Point Cloud Labeling Using 3D Convolutional Neural Network [C]//The 23rd International Conference on Pattern Recognition (ICPR), Cancun, USA, 2016.
- [17] 汤圣君, 张韵婕, 李晓明, 等. 超体素随机森林与 LSTM 神经网络联合优化的室内点云高精度分类方法[J]. *武汉大学学报(信息科学版)*, 2023, 48(4): 525-533.
- TANG Shengjun, ZHANG Yunjie, LI Xiaoming, et al. A High-Precision Indoor Point Cloud Classification Method Jointly Optimized by Super Voxel Random Forest and LSTM Neural Network [J]. *Geomatics and Information Science of Wuhan University*, 2023, 48(4): 525-533.
- [18] CHARLES R Q, HAO S, MO K C, et al. PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation [C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, USA, 2017.
- [19] QI C R, YI L, SU H, et al. PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space[EB/OL]. [2017-02-13]. <https://arxiv.org/abs/1706.02413>.
- [20] JIANG M Y, WU Y R, ZHAO T Q, et al. PointSIFT: A SIFT-Like Network Module for 3D Point Cloud Semantic Segmentation [EB/OL]. [2018-06-05]. <https://arxiv.org/abs/1807.00652>.
- [21] ZHAO H S, JIANG L, FU C W, et al. PointWeb: Enhancing Local Neighborhood Features for Point Cloud Processing [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2019.
- [22] LIU Y C, FAN B, MENG G F, et al. DensePoint: Learning Densely Contextual Representation for Efficient Point Cloud Processing [C]// IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 2019.
- [23] LU T, WANG L M, WU G S. CGA-Net: Category Guided Aggregation for Point Cloud Semantic Segmentation [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.
- [24] 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 (CVPR), Nashville, USA, 2021.
- [25] SHUAI H, XU X, LIU Q S. Backward Attentive Fusing Network with Local Aggregation Classifier for 3D Point Cloud Semantic Segmentation [J]. *IEEE Transactions on Image Processing*, 2021, 30: 4973-4984.
- [26] 程志新, 邹进贵, 贺亦峰, 等. 基于最优邻域特征提取和改进随机森林的建筑立面结构点云分类方法[J]. *测绘地理信息*, 2025, 50(5): 111-116.
- CHENG Zhixin, ZOU Jingui, HE Yifeng, et al. Point Clouds Classification for Building Facade Structures Based on Optimal Neighborhood Feature Extraction and Improved Random Forest [J]. *Journal of Geomatics*, 2025, 50(5): 111-116.
- [27] WU W X, QI Z A, LI F X. PointConv: Deep Convolutional Networks on 3D Point Clouds [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2020.
- [28] HUANG Q G, WANG W Y, NEUMANN U. Recurrent Slice Networks for 3D Segmentation of Point Clouds [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.
- [29] 杨森, 花向红, 李丞, 等. 融合特征选择的点云语义分割方法[J]. *测绘地理信息*, 2025, 50(2): 19-25.
- YANG Sen, HUA Xianghong, LI Cheng, et al. A Point Cloud Semantic Segmentation Method Based on Fusion Feature Selection [J]. *Journal of Geomatics*, 2025, 50(2): 19-25.
- [30] 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 (CVPR), Seattle, USA, 2020.
- [31] QIU S, ANWAR S, BARNES N. Semantic Segmentation for Real Point Cloud Scenes via Bilateral Augmentation and Adaptive Fusion [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, USA, 2021.
- [32] ZENG Z Y, XU Y Y, XIE Z, et al. LEARD-Net: Semantic Segmentation for Large-Scale Point Cloud Scene [J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112: 102953.
- [33] 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 (ICCV), Seoul, Korea, 2019.
- [34] MA L F, LI Y, LI J, et al. Multi-Scale Point-Wise Convolutional Neural Networks for 3D Object Segmentation from LiDAR Point Clouds in Large-Scale Environments[J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22 (2) : 821-836.
- [35] ARMENI I, SENER O, ZAMIR A R, et al. 3D Semantic Parsing of Large-Scale Indoor Spaces[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, USA, 2016.
- [36] HACKEL T, SAVINOV N, LADICKY L, et al. SEMANTIC3D. NET: A New Large-Scale Point Cloud Classification Benchmark[J]. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2017, IV-1/W1: 91-98.
- [37] LANDRIEU L, SIMONOVSKY M. Large-Scale Point Cloud Semantic Segmentation with Superpoint Graphs [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018.
- [38] LI Y Y, BU R, SUN M C, et al. PointCNN: Convolution on X-Transformed Points [C]//Neural Information Processing Systems, Montréal, Canada, 2018.
- [39] ZHANG Z Y, HUA B S, YEUNG S K. ShellNet: Efficient Point Cloud Convolutional Neural Networks Using Concentric Shells Statistics [C]//IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 2020.
- [40] TCHAPMI L, CHOY C, ARMENI I, et al. SEG-Cloud: Semantic Segmentation of 3D Point Clouds [C]//International Conference on 3D Vision, Qingdao, China, 2018.
- [41] WANG L, HUANG Y C, HOU Y L, et al. Graph Attention Convolution for Point Cloud Semantic Segmentation [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Long Beach, USA, 2020.