

DOI:10.19651/j.cnki.emt.2519645

多尺度局部特征融合的弱监督点云语义分割<sup>\*</sup>武佳颖<sup>1,2,3</sup> 杨晓文<sup>1,2,3</sup> 韩 燮<sup>1,2,3</sup> 韩慧妍<sup>1,2,3</sup> 张 元<sup>1,2,3</sup> 赵 融<sup>1,2,3</sup>(1. 中北大学计算机科学与技术学院 太原 030051; 2. 机器视觉与虚拟现实山西省重点实验室 太原 030051;  
3. 山西省视觉信息处理及智能机器人工程研究中心 太原 030051)

**摘要:** 针对现有基于弱监督的语义分割模型无法同时顾及点云局部细节特征的高相关性和泛化性,以及特征利用不充分的问题,以 RAC-Net 为基线模型,提出了多尺度局部特征融合的弱监督点云语义分割模型 WS-MLF。首先,原始点云数据作为输入,提出了多尺度球域采样的方法 MSSM 提取多层的不同半径特征;其次,设计了特征增强模块 MFA,充分利用点邻域内的几何特征;再次,提出了注意力模块 SCH-Att 以增强关键通道和关键点的特征捕捉;最后,运用解码器进行上采样,生成每个点不同的语义标签,完成语义分割任务。该模型在大规模室内场景数据集 S3DIS 和 ScanNet-v2 上进行了实验验证,结果表明在 S3DIS 数据集上,标签比率为 0.02% 和 0.06% 时,mIoU 分别较 RAC-Net 提升了 2.71% 和 0.54%,在 ScanNet-v2 数据集上,标签比率为 20 pt 时,mIoU 较 RAC-Net 提升了 1.55%。实验结果验证了该模型在弱监督场景中对点云关键特征的良好提取能力,提升了基于弱监督的点云语义分割精度。

**关键词:** 弱监督;点云语义分割;多尺度;注意力机制;特征聚合

**中图分类号:** TP391;TN958.98 **文献标识码:** A **国家标准学科分类代码:** 520.603

Weakly supervised point cloud semantic segmentation via  
multi-scale local feature fusionWu Jiaying<sup>1,2,3</sup> Yang Xiaowen<sup>1,2,3</sup> Han Xie<sup>1,2,3</sup> Han Huiyan<sup>1,2,3</sup> Zhang Yuan<sup>1,2,3</sup> Zhao Rong<sup>1,2,3</sup>

(1. School of Computer Science and Technology, North University of China, Taiyuan 030051, China;

2. Shanxi Key Laboratory of Machine Vision &amp; Virtual Reality, Taiyuan 030051, China;

3. Shanxi Vision Information Processing and Intelligent Robot Engineering Research Center, Taiyuan 030051, China)

**Abstract:** To address the limitations of existing weakly supervised semantic segmentation models for point clouds, which struggle to balance local feature correlation, generalization, and feature utilization. This paper proposes WS-MLF, a weakly supervised point cloud semantic segmentation model via multi-scale local feature fusion, based on the RAC-Net baseline. Firstly, the raw point cloud data is taken as input, and a multi-scale spherical sampling methods (MSSM) is employed to capture hierarchical features across varying spatial radii. Secondly, a multi-local feature aggregation enhancement module (MFA) is designed to refine geometric context within neighborhoods. Thirdly, a spatial-channel-fused hybrid attention module (SCH-Att) is proposed to prioritize discriminative channels and key points. Finally, a decoder is utilized for upsampling to generate point-level semantic labels, thereby completing the semantic segmentation task. The proposed model is evaluated on large-scale indoor scene datasets, S3DIS and ScanNet-v2. Experimental results demonstrate that on the S3DIS dataset, when the label ratios are 0.02% and 0.06%, the mIoU surpasses RAC-Net by 2.71% and 0.54%, respectively. On the ScanNet-v2 dataset, with a label ratio of 20 pt, the mIoU increases by 1.55% compared with RAC-Net. These results validate WS-MLF's effectiveness in extracting key features under weak supervision, enhancing segmentation accuracy.

**Keywords:** weakly supervised;point cloud semantic segmentation;multi-scale;attention mechanism;feature aggregation

## 0 引 言

点云语义分割<sup>[1]</sup>是三维计算机视觉的一项关键任务,

广泛应用于城市建模<sup>[2]</sup>、机器人的辅助应用<sup>[3]</sup>以及跨领域的创新应用<sup>[4]</sup>等,为现实场景理解提供技术支持。随着点云采集设备的持续发展,近年来已构建多个涵盖不同场景

收稿日期:2025-08-21

\* 基金项目:国家自然科学基金(62272426)、山西省重点研发计划项目(202402020101001)、山西省自然科学基金(202303021211153, 202403021212166)项目资助

与物体类型的大规模点云数据集<sup>[5-6]</sup>,这不仅为点云语义分割的研究提供了丰富的数据基础,也为数据标注带来了更高的挑战。深度学习的发展为点云语义分割注入了新的活力,并在分割精度方面取得了显著突破。

目前,基于深度学习的点云语义分割方法主要分为基于投影、体素和原始点的方法。前两类方法是指将点云转化为一组规则的二维图像或三维体素的格式,利用卷积进行特征提取,该点云转换过程会造成特征的损失,对语义分割的精度产生影响。相比之下,直接在原始点云上进行建模的方法能够更充分地保留点云的空间结构信息,并在复杂场景的处理上展现出明显优势,已成为近年来的研究热点之一。2017 年所提出的 PointNet<sup>[7]</sup>模型,是首个直接处理原始点云的深度学习模型,随后提出的 PointNet++<sup>[8]</sup>、Point Transformer<sup>[9]</sup>及 Swin3D<sup>[10]</sup>等模型,有效整合了局部和全局特征,提高了点云语义分割精度。基于原始点的点云语义分割方法有效避免了数据格式转换过程中信息丢失的问题,能够更完整地保留点云的几何结构特征,因此在应对复杂三维场景时展现出显著优势。

然而,点云数据本身具有无序排列、不规则结构以及密度分布不均等特性,同时大规模点云数据的存储与处理也带来了计算和资源方面的压力,使得传统的全监督方法在实际应用中面临标注成本高昂、人工消耗巨大、效率低下等问题<sup>[11]</sup>。在此背景下,如何在标注数据有限的条件下仍能实现高质量的语义分割成为当前研究的关键挑战之一。为此,研究者逐渐将目光转向弱监督学习方法,尝试借助有限的监督信息或辅助信息,实现对点云的高效语义理解,从而缓解全监督方法在大规模应用场景下的标注负担。现有的弱监督点云语义分割方法通过扩展弱标注信息、构建一致性约束、设计稳健的伪标签筛选机制等手段以减弱对精确标注的依赖。2022 年 Hu 等<sup>[12]</sup>提出了 SQN (weakly-supervised semantic segmentation of large-scale 3D point clouds),利用点相关性来查询标记点的邻近点,聚合点云更全面的局部特征,最大限度地利用弱标记信号,通过增强局部特征的聚合提升对未标记区域的分割精度;同年 Li 等<sup>[13]</sup>提出的 HybridCR (weakly-supervised 3D point cloud semantic segmentation via hybrid contrastive regularization)引入了动态点云增强器为训练数据生成不同的转换,且提出了一个混合框架整合一致性和对比约束,利用数据增强和对比学习有效增强模型对不同数据分布的适应能力,提高其鲁棒性和泛化能力;2024 年 Wu 等<sup>[14]</sup>提出的 RAC-Net (reliability-adaptive consistency regularization for weakly-supervised point cloud segmentation)将预测置信度和不确定性纳入伪标签筛选以提高伪标签的可靠性,增强模型训练;同年魏东等<sup>[15]</sup>提出的 CCFE (weakly-supervised point cloud semantic segmentation with consistency constraint and feature enhancement)是基于 RandLA-Net 模型<sup>[16]</sup>,引入投影一致性损失与边界对比损失,提升了模型对噪声和

复杂场景的泛化能力,增强了边界区域分割精度;2025 年 Su 等<sup>[17]</sup>基于 SQN<sup>[12]</sup>引入了 2D 视觉边界引导机制,将 2D 图像中的语义边界信息投影至点云空间,对伪标签进行边界感知修正与传播,进一步改进了伪标签的筛选机制,从而显著减少边界区域的误标现象。近期也有研究尝试从模型结构层面提升弱监督性能,2024 年 Niu 等<sup>[18]</sup>基于 RandLA-Net<sup>[16]</sup>提出了一种场景一致性建模方法,通过对完整场景与不完整场景的协同建模,增强监督信号的质量;2025 年 Wang 等<sup>[19]</sup>基于 PointNet++<sup>[8]</sup>提出了一种多引导特征细化方法,通过显式几何先验、邻域关系与全局原型的多重引导,实现局部平滑与全局一致的特征优化,在稀疏标注条件下显著提升分割精度和泛化能力。本研究延续这一思路,从结构层面对 RAC-Net 进行改进。

尽管弱监督方法在一定程度上缓解了点云数据标注成本高昂的压力,但仍存在无法同时顾及点云局部细节特征的高相关性和泛化性、局部与全局特征交互不足以及特征利用不充分的问题,因此提出了一种多尺度局部特征融合的弱监督点云语义分割模型 (weakly supervised point cloud semantic segmentation via multi-scale local feature fusion, WS-MLF)。具体而言:1) 提出了多尺度的球域采样方法 (multi-scale spherical sampling methods, MSSM),在不同尺度下系统地采集邻域信息以增强特征多样性;2) 设计了一种多尺度局部特征聚合增强模块 (multi-local feature aggregation enhancement module, MFA),将不同半径的特征进行自适应融合增强,提升特征表述的准确性;3) 提出了融合空间-通道的混合注意力机制模块 (spatial-channel-fused hybrid attention module, SCH-Att),利用金字塔式的空间与通道注意力分配不同尺度特征的权重,强化局部与全局特征的语义交互,提升模型的语义分割性能。

## 1 本文方法

### 1.1 总架构

本文提出了一种多尺度局部特征融合的弱监督点云语义分割模型 WS-MLF,如图 1 所示,其中图 1(a)表示基线模型 RAC-Net,图 1(b)表示 WS-MLF。

RAC-Net 首先将包含有空间坐标 XYZ 和颜色信息 RGB 等特征的原始点云数据作为输入,使用全连接层 (FC) 和数据增强层 (AUG) 对原始数据进行维度对齐与噪声抑制;其次,在编码器中采用 K 最近邻采样方法逐层下采样,构建高层语义特征;最后,解码器对其进行上采样,并通过逐点分类模块对每个点生成语义标签,完成点云的语义分割任务。

WS-MLF 沿用了 RAC-Net 提出的基于置信度和不确定性联合评估伪标签的机制,以确保伪标签数据的准确性和模型训练的稳定性。在编码器中,WS-MLF 采用最近点采样方法 (farthest point sampling, FPS) 选择中心点,设计基于多尺度球域采样方法 MSSM 的多尺度局部特征聚合

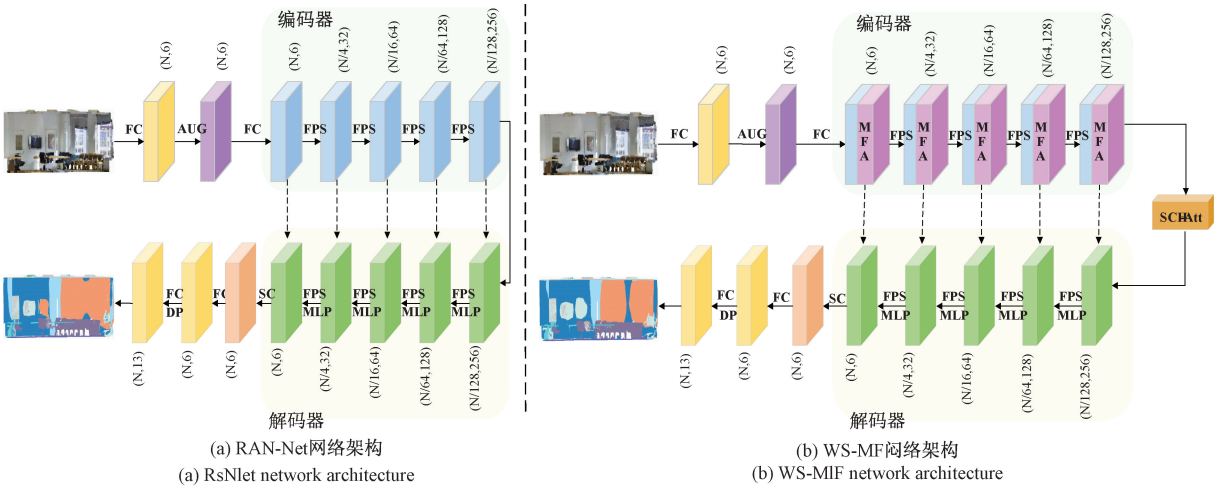


图1 网络总架构

Fig. 1 General network architecture

增强模块 MFA,经中心点构建多尺度局部邻域并对来自两种不同采样半径的特征进行聚合与增强;提出融合空间-通道的混合注意力机制模块 SCH-Att 对增强特征赋予不同的权重,引导网络关注更具语义判别力的区域,用于强化局部特征与全局语义之间的高效交互,提升了模型的分割精度。

### 1.2 多尺度的球域采样方法 MSSM

小尺度的单尺度球域采样方法聚合特征更具有关联性,但是缺少泛化性;而大尺度的聚合特征可能会引入较多不相关的特征,影响聚合特征的表达,因此,仅利用单一半径的方法不能同时顾及局部特征的相关性和泛化性。针对缩小局部特征聚合半径以增强相关性的同时,扩大聚合半径以增强泛化性的特征提取需求,本文提出了多尺度的球域采样方法,利用小尺度的聚合特征增强中心点周围的强相关点的细节特征表示,并引入大尺度局部聚合特征以增强泛化性。

首先,基于球域的多尺度采样方法 MSSM 针对点云数量庞大的问题,采用最远点采样方法获得球域采样的中心点;其次,构建不同半径的球形邻域,捕获不同半径邻域内的多尺度上下文信息,由于点云的密度不均匀性导致在同半径不同位置的中心点周围的邻域点数量不同,存在以下两种情况:

1)当球邻域中与中心点距离小于  $r$  的点超过所需的  $k$  个时,按照点的距离进行升序排序,并选取前  $k$  个距离最小的点以提取对应局部特征,如式(1)所示:

$$q(r, k) = \{ \|s - s_0\| \leq r, \text{count}(s) \leq k, \forall s \in N \} \quad (1)$$

其中,  $\text{count}(s) \leq k$  表示在球邻域中选取前  $k$  个距离小于  $r$  的点。

2)当球邻域中与中心点距离小于  $r$  的点不足所需的  $k$  个,仅  $a$  ( $a < k$ ) 个时,使用  $K$  最近邻采样方法对邻域内的

点与中心点的距离进行升序排序,选取距离中心点距离最近的  $(k - a)$  个点进行补齐  $k$  个,提取该半径内的局部特征,如式(2)所示:

$$q(r, k) = \{ \|s - s_0\| \leq r, \text{count}(D(s)) \leq k - a, \forall s \in N \} \quad (2)$$

其中,  $D(s)$  表示根据球邻域内的点与中心点的距离对邻域内的点升序排序。若  $(k - a) > a$ ,则重复执行式(2),直至邻域内点数达  $k$ 。

改进的基于多尺度的球域采样的特征提取方法能够减少引入无关点影响聚合特征的表达,有效覆盖不同局部细节的特征表达,同时顾及局部特征的强相关性和泛化性,从而增强特征表达的多样性和鲁棒性。

### 1.3 多尺度局部特征聚合增强模块 MFA

针对利用少量标注信息提升分割精度问题,没有标签的点需要利用邻域内点特征进行聚合、推测,获得可靠的语义标签信息。但是,大多数点云语义分割网络往往仅对各点原始的语义信息进行聚合、卷积等,忽略了带有丰富信息的点位置信息。因此本文提出了多尺度局部特征聚合增强模块 MFA,如图2所示,在每个单层部分引入位置信息融合模块(location information fusion module, LIF),对不同尺度邻域特征进行层次化融合实现特征表达优化。

MFA 包含3阶段(以半径  $r_1$  的邻域为例说明):

1)对原始点云数据每个采样点执行不同半径球查询操作,利用 MSSM 构建  $r_1$  尺度球型邻域,生成覆盖不同空间范围的邻域结构。

2)LIF 模块提取  $r_1$  邻域中点的几何特征信息,运用坐标差分运算计算点间相对位置向量、欧氏距离矩阵等信息,实现几何特征的显式表达。各半径邻域共享同一特征增强范式,如式(3)所示:

$$D_i^k = d_i \odot d_i^k \odot (d_i - d_i^k) \odot \|d_i - d_i^k\| \quad (3)$$

其中,  $D_i^k$  表示第  $i$  个中心点利用 MSSM 采样的  $k$  个

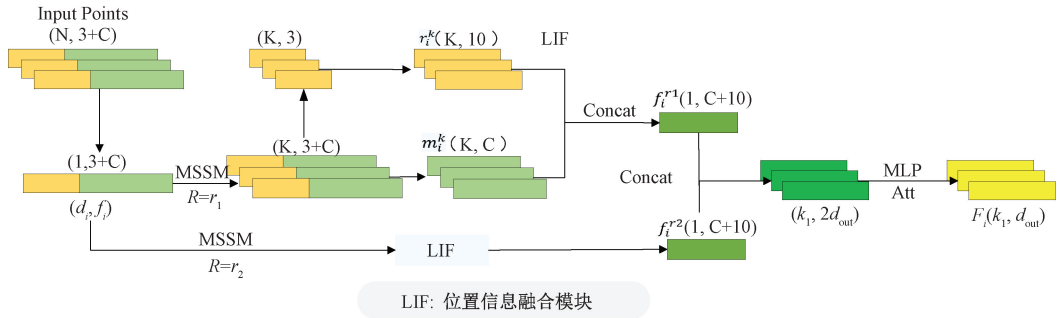


图 2 多尺度局部特征聚合增强模块(MFA)

Fig. 2 Multi-local feature aggregation enhancement module(MFA)

邻域点的距离特征,  $d_i$  表示第  $i$  个中心点的初始 XYZ 几何特征,  $d_i^k$  表示第  $i$  个中心点的  $k$  个邻域点的初始 XYZ 特征,  $(d_i - d_i^k)$  表示第  $i$  个中心点与  $k$  个邻域点的相对位置特征,  $\|d_i - d_i^k\|$  表示第  $i$  个中心点与  $k$  个邻域点的欧氏距离,  $\odot$  表示 Concat 连接操作。

3) 特征  $D_i^k$  经标准化处理形成增强几何特征描述符, 如式(4)所示:

$$r_i^k = \text{LeakyReLU}(\text{BN}(D_i^k)) \quad (4)$$

其中,  $r_i^k$  表示第  $i$  个中心点利用球域采样得到的  $k$  个邻域点聚合的几何特征, BN 表示 Batch Normalization 批量归一化层, LeakyReLU 表示激活函数层。

4) 增强的几何特征  $r_i^k$  与原起点云的语义特征  $m_i^k$  结合, 输入到注意力池, 如式(5)所示:

$$f_i^{r_1} = \text{Att}(r_i^k \odot m_i^k) \quad (5)$$

其中,  $f_i^{r_1}$  表示半径为  $r_1$  邻域内点的聚合特征,  $m_i^k$  表

示第  $i$  个中心点的  $k$  个邻域点的语义特征, Att 表示注意力池。

5) 将  $r_2$  邻域内的点特征  $f_i^{r_2}$  与  $r_1$  邻域内的点特征  $f_i^{r_1}$  连接, 输入到 MLP 和注意力池, 得到增强特征  $F_i$ , 如式(6)所示:

$$F_i = \text{Att}(\text{MLP}(f_i^{r_1} \odot f_i^{r_2})) \quad (6)$$

其中,  $f_i^{r_2}$  表示聚合两个不同半径尺度后的增强特征,  $f_i^{r_2}$  表示半径为  $r_2$  邻域内点的聚合特征, MLP 表示多层感知机。增强特征  $F_i$  同时蕴含不同半径内的局部几何细节与上下文信息。

#### 1.4 融合空间-通道的混合注意力机制模块 SCH-Att

本文设计了一种融合空间-通道的混合注意力机制 SCH-Att, 进一步增强多尺度特征的表达能。该注意力机制包括两个部分: 通道注意力(channel attention, CA)和空间注意力(spatial attention, SA), 如图 3 所示。

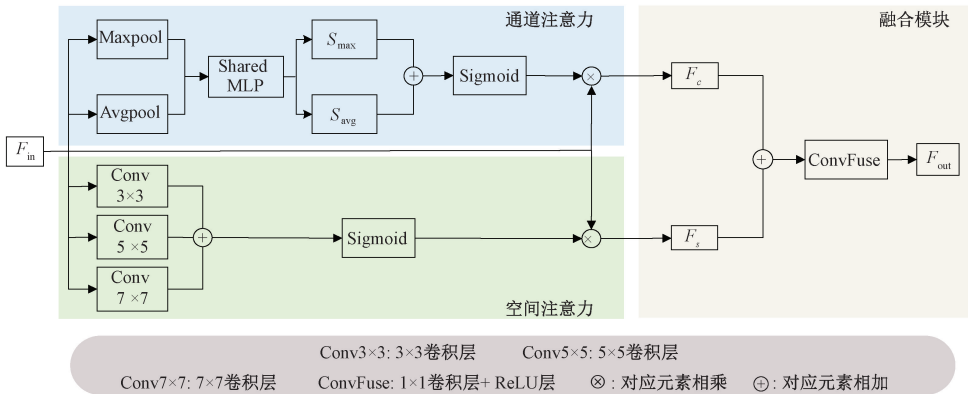


图 3 融合空间-通道的混合注意力机制模块(SCH-Att)

Fig. 3 Spatial-channel-fused hybrid attention mechanism module(SCH-Att)

首先, 通道注意力<sup>[20]</sup>自适应地调整不同特征间的重要性, 学习每一维特征加权系数以强调具有显著性的特征信息, 运用全局平均池化与最大池化提取整体上下文信息, 采用共享的多层感知机(shared MLP)建模通道间的相关性, 具体表示为式(7)和(8):

$$S_{\max} = \text{Share}(\text{Maxpool}(F_{\text{in}})) \quad (7)$$

$$S_{\text{avg}} = \text{Share}(\text{Avgpool}(F_{\text{in}})) \quad (8)$$

其中,  $S_{\max}$  和  $S_{\text{avg}}$  分别表示最大池化和平均池化的注意力分数,  $F_{\text{in}}$  表示经过 5 层下采样模块得到的特征, Share 是共享 MLP 层。  $S_{\max}$  和  $S_{\text{avg}}$  输入到 Sigmoid 函数生成各通道的权重分布, 以强调重要特征通道, 再将权重与原始输入特征图  $F_{\text{in}}$  对应通道进行逐元素乘法, 得到特征  $F_c$ 。具体可以表示为式(9)所示:

$$F_c = \text{Sigmoid}(S_{\max} + S_{\text{avg}}) \otimes F_{\text{in}} \quad (9)$$

其中,  $\otimes$  表示逐元素相乘。

其次,空间注意力利用点和点之间的相对位置关系动态聚合空间的特征,定位需要重点关注的空间位置。如图3中的空间注意力模块所示,先利用3个卷积核分别为  $3 \times 3$ 、 $5 \times 5$ 、 $7 \times 7$  的金字塔卷积操作获得不同感受野空间特征,再执行 Sigmoid 函数生成空间注意力映射,最后与原始特征图  $F_{in}$  逐元素相乘,有效提升模型在语义分割任务中的局部细节特征表达能力和泛化能力。具体表示如式(10)所示:

$$F_s = \text{Sigmoid} \left( \begin{array}{c} \text{Conv}^{3 \times 3}(F_{in}) + \text{Conv}^{5 \times 5}(F_{in}) \\ + \text{Conv}^{7 \times 7}(F_{in}) \end{array} \right) \otimes F_{in} \quad (10)$$

其中,  $\text{Conv}^{3 \times 3}$ 、 $\text{Conv}^{5 \times 5}$  和  $\text{Conv}^{7 \times 7}$  分别表示  $3 \times 3$ 、 $5 \times 5$  和  $7 \times 7$  的卷积网络。

最后,两种注意力机制特征  $F_c$  和  $F_s$  并联进行特征融合,赋予特征不同的注意力权重,得到兼顾不同特征表示和不同感受野的特征  $F_{out}$  并传递给解码器,增强局部上下文信息的表达,提升对不规则几何的优化特征提取<sup>[21]</sup>。具体可以表示为式(11):

$$F_{out} = \text{ConvFuse}(F_c + F_s) \quad (11)$$

其中,  $\text{ConvFuse}$  表示  $1 \times 1$  的卷积网络和 Relu 激活函数层。

## 2 实验结果与分析

### 2.1 数据集

本文应用两个大型室内点云数据集 S3DIS 和 ScanNet-v2 数据集进行了 WS-MLF 网络的训练和测试。

S3DIS 数据集是一个大规模室内 3D 点云数据集,包含丰富的几何信息和颜色信息。该数据集提供了丰富的室内环境数据,包括 6 个不同建筑的空间,如办公室、会议室和大厅等,涵盖了多个室内环境以及 13 个不同类别,包括墙壁、地板、桌子等常见的室内物体。由于区域 5 与其他区域的数据集不在同一个建筑中,场景中的物品有一定的差异性,所以本文选取区域 5 作为测试数据集,这样可以更准确地测试网络语义分割的鲁棒性和泛化能力。

ScanNet-v2 数据集提供了高质量的 3D 点云数据和对应的彩色图像。该数据集包含 1 513 个室内场景,涵盖了各种不同的室内环境,如客厅、厨房和浴室等,包括了 20 种物体类别,如墙壁、地板和沙发等。选取了其中的 312 个场景进行测试,其余场景用于训练。

### 2.2 环境配置以及评估指标

本文实验环境基于 Ubuntu 20.04,显卡是 NVIDIA GeForce RTX 3090,内存是 24 GB,搭载 CUDA11.3,使用 Torch1.11.0 和 Python3.8。批大小设置为 2,初始学习率是 0.01,训练轮次设置为 100,使用 SGD 优化器。

本文实验的评估指标是总体准确率(overall accuracy, OA)和平均交并比(mean intersection over union, mIoU)。OA 指正确预测的类别数占需预测类别总

数的百分比;交并比指预测区域和实际区域的交集与预测区域和实际区域的并集的比值,重复计算各个类别的 IoU,最后取其平均值即为 mIoU,如式(12)和(13)所示:

$$OA = \frac{1}{N} \sum_{i=1}^C TP_i \quad (12)$$

$$mIoU = \frac{1}{C} \sum_{i=1}^C \frac{TP_i}{TP_i + FP_i + FN_i} \quad (13)$$

其中,  $N$  表示样本的总数;  $C$  表示分割类别的数量;  $TP$  表示正确预测为正例样本的点数,  $TN$  表示正确预测为反例样本的点数,  $FP$  表示错误预测为正例样本的点数,  $FN$  表示错误预测为反例样本的点数。

### 2.3 对比试验

#### 1) S3DIS 数据集对比实验

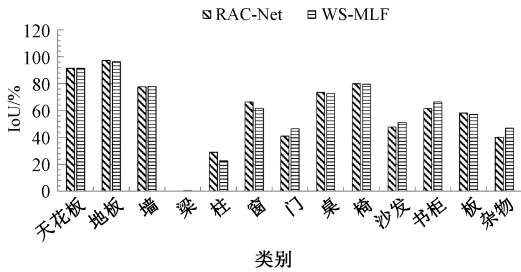
本文方法 WS-MLF 与现有的先进弱监督点云语义分割方法进行对比,表 1 展示了不同方法在 S3DIS 数据集区域 5 上得到的结果。当标签比率为 100% 和 0.1% 时,现有方法的 mIoU 分别可达到 63.73% 和 63.40%;而本文提出的 WS-MLF 模型在标签比率大幅降低至 0.06% 和 0.02% 时, mIoU 仍分别可达 59.25% 和 57.85%,相较于基线网络 RAC-Net 分别提升了 2.71% 和 0.54%,并且超过 Wang 等<sup>[19]</sup>所提出的网络 0.05%,充分体现了在极低标注比例下的显著竞争优势。图 4 中展示了 WS-MLF 和

表 1 不同方法在 S3DIS 数据集的区域 5 上的分割精度对比

Table 1 Comparison of segmentation accuracy of different methods on region 5 of the S3DIS dataset

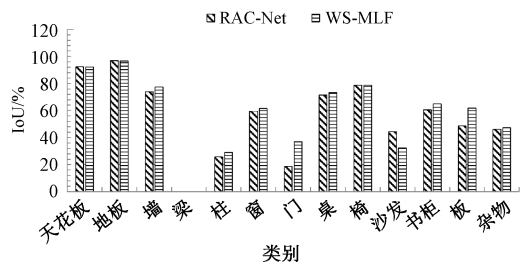
标签比例/%	方法	OA/%	mIoU/%
100	PointNet <sup>[7]</sup>	—	41.10
	RandLA-Net <sup>[16]</sup>	86.70	63.00
	SQN <sup>[12]</sup>	—	63.73
0.10	RandLA-Net <sup>[16]</sup>	—	52.90
	SQN <sup>[12]</sup>	—	61.41
0.06	PointMatch <sup>[22]</sup>	—	63.40
	OTOC <sup>[23]</sup>	—	55.30
	DAT <sup>[24]</sup>	—	58.50
	RAC-Net <sup>[14]</sup>	84.99	58.71
	WS-MLF(本文)	85.55	59.25
0.03	PA-Net <sup>[25]</sup>	80.90	51.60
	OTOC <sup>[23]</sup>	—	50.10
0.02	DAT <sup>[24]</sup>	—	56.50
	PointMatch <sup>[22]</sup>	—	55.30
	RAC-Net <sup>[14]</sup>	83.00	55.14
	Wang <sup>[19]</sup>	—	57.80
	WS-MLF(本文)	<b>85.26</b>	<b>57.85</b>

注:“—”表示原论文未公开该数据。加粗表示在同一标注比率时,性能最高。



(a) 标签率为0.06%的各类别分割精度

(a) Segmentation accuracy for each category at a labeling rate of 0.06%



(b) 标签率为0.02%的各类别分割精度

(b) Segmentation accuracy for each category at a labeling rate of 0.02%

图 4 S3DIS Area-5 数据集各类别的分割精度

Fig. 4 Segmentation accuracy for each category of S3DIS Area-5 dataset

RAC-Net 在 S3DIS 数据集区域 5 上各类别的分割精度,其图 4(a)表示的是标签比率为 0.06%的结果,图 4(b)表示的是标签比率为 0.02%的结果,可以看出在 0.06%的设置下,WS-MLF 在门、沙发和书柜等类别都得到了较高的分割精度;在 0.02%的设置下,WS-MLF 在墙、板和杂物等类别都得到了较高的分割精度。

从定性角度分析,WS-MLF 和 RAC-Net 网络在 0.02%标签率的弱监督下的分割结果可视化如图 5 所示,图 5(a)为输入点云,图 5(b)为参考标准,图 5(c)为 RAC-Net 的分割结果,

图 5(d)为 WS-MLF 的分割结果。由上至下是来自区域 5 的 5 个不同的房间,分别是房间 1~5。虚线框表示 WS-MLF 相较于 RAC-Net 表现更好的部分。对这 5 个房间的分割过程中,RAC-Net 错误地将房间 1 内的杂物划分为窗户、窗户间的墙壁连接处错分为窗户、房间 2 内的大面积墙壁错分为柱子以及房间 3 内的桌子上的小型杂物被错误划分为其他物体等。而 WS-MLF 在以上所述的位置都取得了较好的分割效果,对其错误分割都有一定的改进。相较于 RAC-Net,WS-MLF 在细节和大面积部分的分割性能都有所改进。

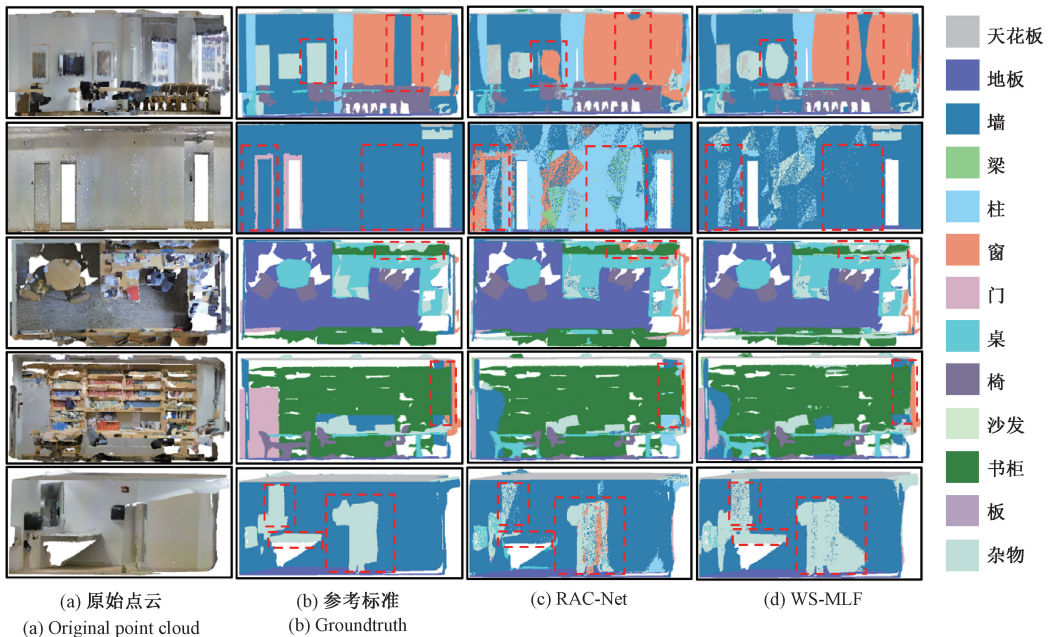


图 5 0.02%标签率下 S3DIS 数据集区域 5 上的分割结果可视化对比

Fig. 5 Comparison of visualization of segmentation results on region 5 of S3DIS dataset at 0.02% labeling rate

本文方法 WS-MLF 和基线网络 RAC-Net 在 0.06% 标签率下的分割结果可视化如图 6 所示,其中虚线框框出的部分表示 WS-MLF 相较于基线网络表现更好的部分。在对 5 个房间的分割过程中,RAC-Net 错将房间 1 内的椅子分割为桌子、房间 2 内的门错分为墙、房间 3 内的与墙连接部分的书柜错分为墙、房间 4 内的墙角处的杂物错分为墙以及房间 5 内的杂物错分为书柜等。而 WS-MLF 在上

述位置都展现了显著的优越性,有较好的分割表现。可见,相较于 RAC-Net,WS-MLF 在大面积事物以及细节部分的分割效果都有一定的改进。

2) ScanNet-v2 数据集对比实验

ScanNet-v2 测试集上的对比实验结果如表 2 所示。从表 2 可以看出,在 20 pt 的标签比率下,本文提出的 WS-MLF 模型相较于基线网络 RAC-Net 的 mIoU 提升了 1.55%,

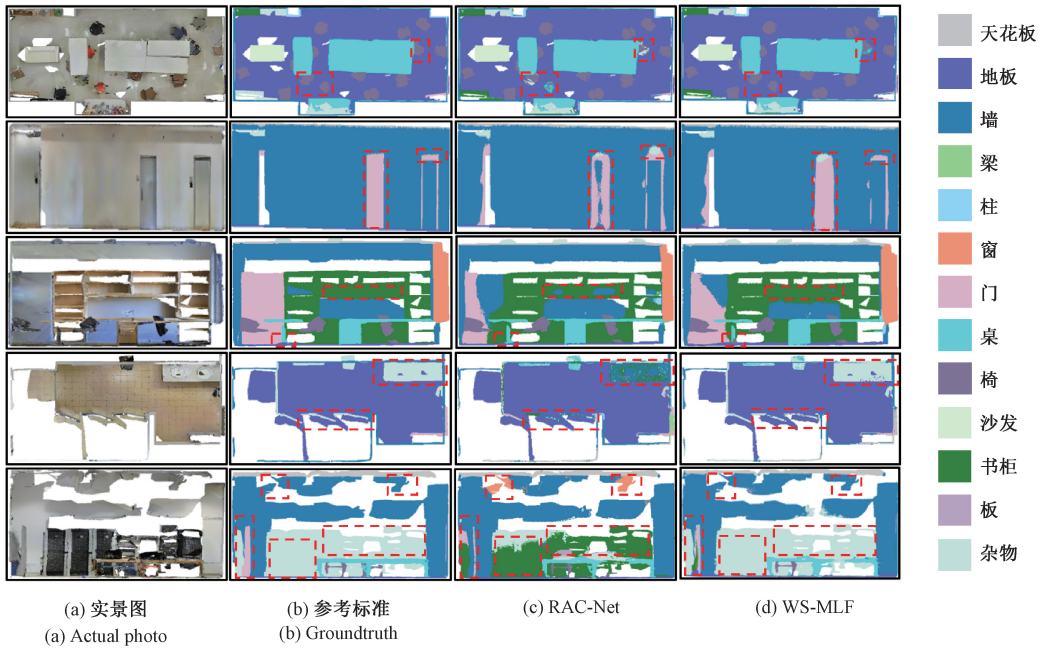


图6 0.06%标签率下 S3DIS 数据集区域 5 上的分割结果可视化对比

Fig. 6 Comparison of visualization of segmentation results on region 5 of S3DIS dataset at 0.06% labeling rate

表2 不同方法在 ScanNet-v2 数据集上的分割精度对比

Table 2 Comparison of segmentation accuracy of different methods on ScanNet-v2 dataset

标签比率	方法	OA/%	mIoU/%
100%	Pointnet++ <sup>[8]</sup>	—	33.90
	PointCNN <sup>[26]</sup>	—	45.80
	KPConv <sup>[27]</sup>	—	68.40
	MinkowskiNet <sup>[28]</sup>	—	73.60
1%	PSD <sup>[29]</sup>	—	54.70
	Niu <sup>[18]</sup>	—	51.42
0.10%	SQN <sup>[12]</sup>	—	56.90
	PointMatch <sup>[22]</sup>	—	68.80
	Niu <sup>[18]</sup>	—	43.46
20 pt	Wang <sup>[19]</sup>	—	57.40
	CSC_LA_SEM <sup>[30]</sup>	—	53.10
	Viewpoint_BN_LA_AIR <sup>[31]</sup>	—	54.80
	PointContrast_LA_SEM <sup>[32]</sup>	—	55.00
	DAT <sup>[24]</sup>	—	55.20
	MIL-derived transformer <sup>[33]</sup>	—	54.40
	RAC-Net <sup>[14]</sup>	80.56	54.09
	WS-MLF(本文)	<b>83.70</b>	<b>55.64</b>

注:“—”表示原论文未公开该数据。加粗表示在同一标签比率时,性能最高。“20 pt”表示使用官方给定的 20 个标注点作为稀疏标签来训练模型。

其分割精度已超过 1% 标签比率下的结果,展现出优异的弱监督性能。进一步对比可见,在 0.1% 标签比率下,Niu

等<sup>[18]</sup>和 Wang 等<sup>[19]</sup>提出的方法 mIoU 分别为 43.46% 和 57.40%,而 WS-MLF 在低于其标注比率的情况下仍取得较高精度,显示出在极低标注条件下的明显竞争优势。ScanNet-v2 测试集在 20 pt 标签比率时各类别的分割精度如图 7 所示,可以看出本文方法 WS-MLF 在门、柜台、浴帘和厕所等类别上展现出了较高的分割效果。

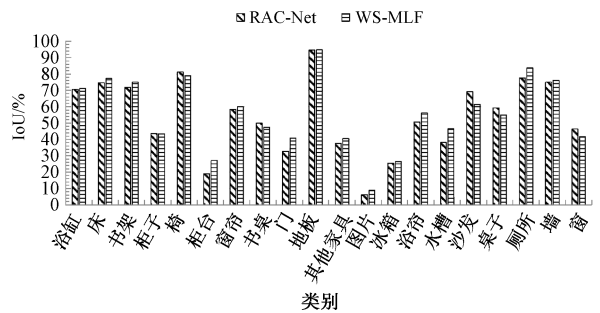


图7 ScanNet-v2 数据集各类别的分割精度

Fig. 7 Segmentation accuracy for each category of ScanNet-v2 dataset

将 RAC-Net 与本文方法 WS-MLF 在 ScanNet-v2 测试集上进行可视化的对比,如图 8 所示,虚线框表示 WS-MLF 比 RAC-Net 分割效果好的部分。可以看出,在对由上至下的 5 个不同房间分割过程中,RAC-Net 将房间 1 和 3 内的未分类的其他家具错分为椅子、墙和地板、房间 2 内的外形相似的柜子和柜台混淆、房间 4 内的门错分为墙、房间 5 内的桌子错分为书桌等。而 WS-MLF 对上述的部分都有一定改善,利用 MSSM 方法增强了对细节部分特征的提取,获得比 RAC-Net 更准确的分割结果。

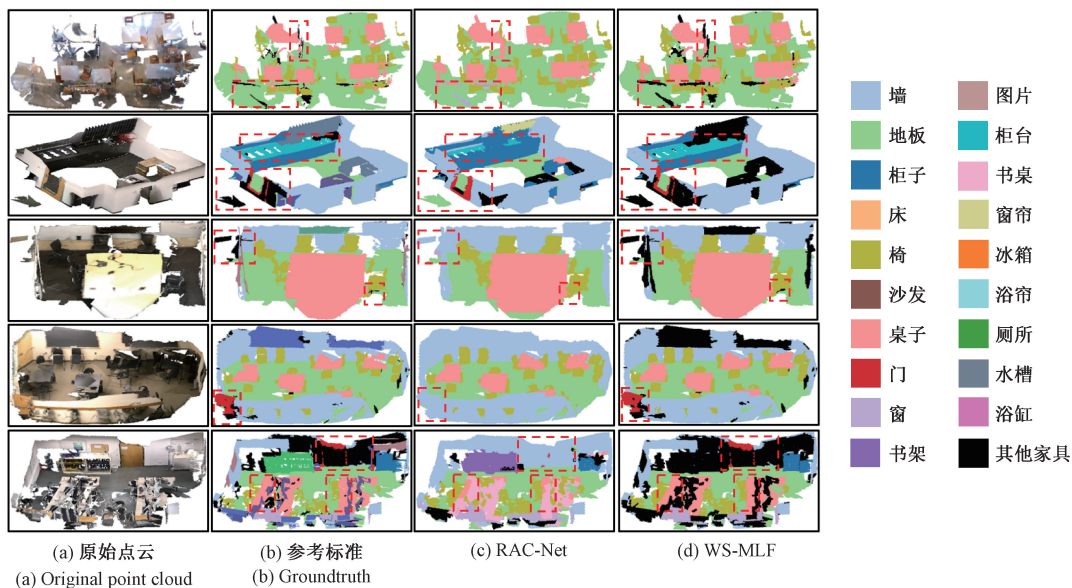


图 8 20 pt 标签率下 ScanNet-v2 数据集上的分割结果可视化对比

Fig. 8 Comparison of visualization of segmentation results on ScanNet-v2 dataset at 20 pt labeling rate

### 2.4 消融实验

为了定量评估本文提出模块的有效性,对 MSSM、MFA 和 SCH-Att 模块的不同组合进行了消融实验。将 S3DIS 数据集区域 5 划分为测试集,其他区域作为训练集,在标签比率为 0.02% 的条件下进行测试,结果如表 3 所示。1)方法 1 对应的是基线网络 RAC-Net,其利用 K 最近邻采样方法进行点云下采样,在区域 5 的 mIoU 达到 55.14%。K 最近邻采样忽略几何结构,在非均匀区域选择较远的邻域点,破坏了局部特征的一致性。2)方法 2 利用本文提出的 MSSM 替代基线中的 K 最近邻采样后,模型在区域 5 的 mIoU 提升至 57.26%。基于最远点采样的中心点选择,实现了对点云的更均匀覆盖;多尺度球域采样利用不同半径范围,提取与中心点紧密相关的局部特征,有效保持了物体的几何细节,增强了特征的表现力和泛化能力。3)方法 3 引入多尺度局部特征聚合增强模块 MFA,使得网络在区域 5 的 mIoU 提升至 57.60%,得益于 MFA 引入了各点的位置信息,获取的点云特征更加丰富,增加了网络训练的可用信息,缓解由于弱监督带来的有效信息不足的缺点。4)方法 4 引入融合空间-通道的混合注

意力机制模块 SCH-Att,使得网络在区域 5 的 mIoU 达到 57.85%,其中的通道注意力机制可以增强某些重要通道的特征表达,金字塔形的空间注意力机制可以利用不同大小的卷积核覆盖不同的感受野,提高细节特征的表达能力和泛化性,结合二者提升网络的分割性能。

### 3 结 论

弱监督点云语义分割利用少量的真实标签训练,对物体进行语义分割,该方法很大程度上降低了标注成本,提升了数据的利用率。针对现有方法无法同时顾及点云局部细节特征的高相关性和泛化性以及特征利用不充分的问题,本文基于 RAC-Net,提出了一种多尺度局部特征融合的弱监督点云语义分割方法 WS-MLF。首先,提出了多尺度球域采样方法 MSSM 构建层次化邻域结构并进行特征提取,增强了点特征间的强关联性和泛化性;其次,设计了局部特征聚合增强模块 MFA,深度挖掘邻域内的几何信息,缓解了聚合过程中特征利用不充分的问题;最后,提出了融合空间-通道的混合注意力机制模块 SCH-Att,利用动态权重分配增强关键信息的表达能力,提升局部上下文信息的表达,从而得到更准确的语义分割结果。基于大规模室内场景数据集的语义分割实验结果表明,相较于基线方法 RAC-Net,在 S3DIS 数据集上标签率为 0.02% 时提高了 2.71%,标签率为 0.06% 时提高了 0.54%,在 ScanNet-v2 数据集上标签率为 20 pt 时提高了 1.55%,以上结果验证了本文方法的有效性。针对边缘区分不清晰和算法效率不高的局限性,未来拟结合边缘损失函数改善边缘分割问题以及模型的轻量化,提升模型的训练效率和推理速度。

表 3 不同模块在 S3DIS 数据集区域 5 上的分割精度对比  
Table 3 Comparison of segmentation accuracies of different modules on region 5 of the S3DIS dataset

方法	MSSM	MFA	SCH-Att	mIoU/%
1	×	×	×	55.14
2	√	×	×	57.26
3	√	√	×	57.60
4	√	√	√	57.85

## 参考文献

- [1] 仇志江,张林,姚焱,等. 三维点云场景语义分割研究进展[J]. 中国图象图形学报, 2025, 30(7): 2325-2342.  
 QIU ZH J, ZHANG L, YAO Y, et al. Survey on semantic segmentation in 3D point cloud scenes[J]. Journal of Image and Graphics, 2025, 30(7): 2325-2342.
- [2] 刘洪凯,王少红,左云波,等. 室内障碍物点云分割的可变阈值联合聚类算法研究[J]. 电子测量技术, 2024, 47(9): 70-78.  
 LIU H K, WANG SH H, ZUO Y B, et al. Research on variable threshold joint clustering algorithm for indoor obstacle point cloud segmentation [J]. Electronic Measurement Technology, 2024, 47(9): 70-78.
- [3] 张静,黄国方,刘晓铭,等. 配网带电作业机器人精准作业定位方法[J]. 电网技术, 2022, 46(2): 812-819.  
 ZHANG J, HUANG G F, LIU X M, et al. Precise positioning for live working on distribution line[J]. Power System Technology, 2022, 46(2): 812-819.
- [4] 牛丽娟,李立功. 点云数据与 BIM 的古建筑三维模型构建方法研究[J]. 电子测量技术, 2021, 44(8): 115-119.  
 NIU L J, LI L G. Research on 3D modeling of ancient buildings based on point cloud data and BIM [J]. Electronic Measurement Tecgnology, 2021, 44(8): 115-119.
- [5] SILBERMAN N, HOIEM D, KOHLI P, et al. Indoor segmentation and support inference from RGBD images[C]. 12th European Conference on Computer Vision, 2012: 746-760.
- [6] JANOCH A, KARAYEV S, JIA Y Q, et al. A category-level 3-D object dataset: Putting the Kinect to work[C]. 2011 IEEE International Conference on Computer Vision Workshops, 2011: 1168-1174.
- [7] QI C R, SU H, KAICHUN M, et al. PointNet: Deep learning on point sets for 3D classification and segmentation[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017: 652-660.
- [8] QI C R, YI L, SU H, et al. PointNet++: Deep hierarchical feature learning on point sets in a metric space[J]. ArXiv preprint arXiv: 1706.02413, 2017.
- [9] ZHAO H SH, JIANG L, JIA J Y, et al. Point Transformer [C]. 18th IEEE/CVF International Conference on Computer Vision, 2021: 16259-16268.
- [10] YANG Y Q, GUO Y X, XIONG J Y, et al. Swin3D: A pretrained transformer backbone for 3D indoor scene understanding [J]. Computational Visual Media, 2025, 11(1): 83-101.
- [11] 刘玉珍,张冬霞,陶志勇. 基于自适应图卷积和注意力池化的点云分类与分割[J]. 计算机工程与科学, 2024, 46(5): 872-880.  
 LIU Y ZH, ZHANG D X, TAO ZH Y. Point cloud classification and segmentation based on adaptive graph convolution and attention pooling[J]. Computer Engineering & Science, 2024, 46(5): 872-880.
- [12] HU Q Y, YANG B, FANG G CH, et al. SQN: Weakly-supervised semantic segmentation of large-scale 3D point clouds[C]. 17th European Conference on Computer Vision, 2022: 600-619.
- [13] LI M T, XIE Y, SHEN Y H, et al. HybridCR: Weakly-supervised 3D point cloud semantic segmentation via hybrid contrastive regularization[C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 14930-14939.
- [14] WU ZH H, WU Y CH, LIN G SH, et al. Reliability-adaptive consistency regularization for weakly-supervised point cloud segmentation[J]. International Journal of Computer Vision, 2024, 132: 2276-2289.
- [15] 魏东,白宜凡,孙赫,等. 一致性约束与特征增强的弱监督点云语义分割[J]. 激光与光电子学进展, 2025, 62(8): 405-415.  
 WEI D, BAI Y F, SUN H, et al. Weakly-supervised point cloud semantic segmentation with consistency constraint and feature enhancement [J]. Laser & Optoelectronics Progress, 2025, 62(8): 405-415.
- [16] 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, 2020: 11108-11117.
- [17] SU ZH, ZHOU L, TAN Y D, et al. Visual boundary-guided pseudo-labeling for weakly supervised 3D point cloud segmentation in indoor environments [J]. IEEE Transactions on Visualization and Computer Graphics, 2025, 31(9): 5962-5974.
- [18] NIU Y CH, YIN J Q, QI CH, et al. Weakly supervised point cloud semantic segmentation based on scene consistency[J]. Applied Intelligence, 2024, 54: 12439-12452.
- [19] WANG Y F, ZHAO Q F, XIA Z Y. Multi-guided feature refinement for point cloud semantic segmentation with weakly supervision[J]. Knowledge-Based Systems, 2025, 311: 113050.
- [20] 郑美琳,高建瓴. 融合多注意力机制与 PointRCNN 的三维点云目标检测[J]. 电子测量技术, 2022, 45(9): 127-132.  
 ZHENG M L, GAO J L. 3D point cloud target detection based on attention mechanism and pointrcnn [J].

- Electronic Measurement Technology, 2022, 45(9): 127-132.
- [21] HAN G G, HUANG SH L, ZHAO F, et al. SIAM: A parameter-free, spatial intersection attention module[J]. Pattern Recognition, 2024, 153:110509.
- [22] WU Y SH, YAN Z ZH, CAI SH C, et al. PointMatch: A consistency training framework for weakly supervised semantic segmentation of 3D point clouds[J]. Computers & Graphics-Uk, 2023, 116: 427-436.
- [23] LIU ZH ZH, QI X J, FU C, et al. One thing one click: A self-training approach for weakly supervised 3D semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 1726-1736.
- [24] XU X W, WANG Y F, ZHENG Y, et al. Back to reality: Weakly-supervised 3D object detection with shape-guided label enhancement [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022:8438-8447.
- [25] NIU Y CH, YIN J Q. PA-Net: Trustworthy weakly supervised point cloud semantic segmentation with primary-auxiliary structure [J]. Computers and Electrical Engineering, 2024, 119: 109555.
- [26] LI Y Y, BU R, SUN M CH, et al. PointCNN: Convolution on X-transformed points [J]. ArXiv preprint arXiv: 1807.07791, 2018
- [27] 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, 2019:6411-6420.
- [28] CHOY C, GWAK J Y, SAVARESE S, et al. 4D spatio-temporal convNets: Minkowski convolutional neural networks[C]. 32nd IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 3075-3084.
- [29] ZHANG Y CH, QU Y Y, XIE Y, et al. Perturbed self-distillation: weakly supervised large-scale point cloud semantic segmentation [C]. 18th IEEE/CVF International Conference on Computer Vision, Piscataway, 2021:15520-15528.
- [30] HOU J, GRAHAM B, NIESSER M, et al. Exploring data-efficient 3D scene understanding with contrastive scene contexts [C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 15587-15597.
- [31] LUO L Y, TIAN B W, ZHAO H, et al. Pointly-supervised 3D scene parsing with viewpoint bottleneck[J]. ArXiv preprint arXiv:2109.08553, 2021.
- [32] XIE S N, GU J T, GUO D M, et al. PointContrast: Unsupervised pre-training for 3D point cloud understanding [C]. 16th European Conference on Computer Vision, 2020:574-591.
- [33] YANG CH K, WU J J, CHEN K S, et al. An ml-derived transformer for weakly supervised point cloud segmentation [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11830-11839.

### 作者简介

**武佳颖**, 硕士研究生, 主要研究方向为点云语义分割。

E-mail: wujiaying0316@163.com

**杨晓文**(通信作者), 副教授, 博士, 主要研究方向为计算机视觉、人工智能。

E-mail: wenyang1314@nuc.edu.cn

**韩燮**, 教授, 博士, 主要研究方向为计算机视觉、人工智能。

E-mail: hanxie@nuc.edu.cn

**韩慧妍**, 副教授, 博士, 主要研究方向为计算机视觉、人工智能。

E-mail: hhy980344@163.com

**张元**, 副教授, 硕士, 主要研究方向为计算机视觉、人工智能。

E-mail: zhangyuan@nuc.edu.cn

**赵融**, 讲师, 博士, 主要研究方向为计算机视觉、人工智能。

E-mail: 20240008@nuc.edu.cn