

文章编号:1001-5078(2024)02-0185-08

· 激光应用技术 ·

顾及多尺度监督的点云语义分割

文阳晖^{1,2,3}, 杨晓文^{1,2,3}, 张元^{1,2,3}, 韩燮^{1,2,3}, 况立群^{1,2,3}, 薛红新^{1,2,3}

(1. 中北大学, 计算机科学与技术学院, 山西太原 030051;

2. 山西省视觉信息处理及智能机器人工程研究中心, 山西太原 030051;

3. 机器视觉与虚拟现实山西省重点实验室, 山西太原 030051)

摘要:针对复杂场景点云分割精度不高、神经网络隐藏单元缺乏直接监督,难以提取语义明确的点云特征等问题,提出了一种将多尺度监督和 SCF-Net 相结合的点云语义分割网络。首先构建了一个类别信息生成模块,记录编码器中隐藏单元感受野内的类别,用于解码器中辅助分类器的监督学习。其次将解码阶段的点云类别预测任务分解成一系列点云感受野类别预测任务,通过对解码器中每一层添加辅助分类器,预测当前阶段点云感受野类别,编码阶段生成的类别信息作为标签监督网络学习。模型从粗到细地推理点云感受野类别,最终预测得到点云语义标签。实验结果表明,该方法能够有效提取点云关键信息,提高语义分割精度。

关键词:三维点云;语义分割;多尺度监督;深度学习;SCF-Net

中图分类号:TN958.98;TP391.41 文献标识码:A DOI:10.3969/j.issn.1001-5078.2024.02.004

Point cloud semantic segmentation considering multi-scale supervision

WEN Yang-Hui^{1,2,3}, YANG Xiao-wen^{1,2,3}, ZHANG Yuan^{1,2,3}, HAN Xie^{1,2,3},
KUANG Li-qun^{1,2,3}, XUE Hong-xin^{1,2,3}

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

2. Shanxi Province's Vision Information Processing and Intelligent Robot Engineering Research Center, Taiyuan 030051, China;

3. Shanxi Key Laboratory of Machine Vision and Virtual Reality, Taiyuan 030051, China)

Abstract: In this paper, a point cloud semantic segmentation network combining multi-scale supervision and SCF-Net is proposed to address the problems of low segmentation accuracy of point cloud in complex scene, the lack of direct supervision in neural network hidden units, and the difficulty in extracting specific point cloud features. A category information generation module is first constructed to record the receptive field categories of hidden unit in the encoder, which is used for the supervised learning of auxiliary classifiers in the decoder. Secondly, the point cloud category prediction task in the decoding stage is decomposed into a series of point cloud receptive field category prediction tasks. By adding auxiliary classifiers to each layer of the decoder, the point cloud receptive field category of the current stage is predicted and the category information generated in the coding stage is used as the label to supervise network learning. The model infers point cloud receptive field categories from coarse to fine, and finally predicts point cloud semantic labels. The experimental results show that the method can effectively extract key information of point cloud and

基金项目:国家自然科学基金项目(No. 62272426; No. 62106238);山西省回国留学人员科研资助项目(No. 2020-113);山西省科技成果转化引导专项项目(No. 202104021301055)资助。

作者简介:文阳晖(1999-),男,硕士研究生,CCF会员类型(N8029G),研究方向为三维图像分类分割技术。
E-mail:15536871044@sina.com

通讯作者:杨晓文(1980-),女,博士,副教授,硕士生导师,研究方向为计算机视觉、虚拟仿真与可视化技术。
E-mail:wenyang1314@nuc.edu.cn

收稿日期:2023-04-04

improve the accuracy of semantic segmentation.

Keywords: three-dimensional point cloud; semantic segmentation; multi-scale supervision; deep learning; SCF-Net

1 引言

近年来,随着无人驾驶、计算机视觉、人工智能等领域的快速发展,深度传感技术日益成熟,三维点云的语义分割在计算机视觉领域引起了越来越多的关注^[1]。伴随大型三维点云数据集的出现以及计算机硬件的更新换代,在点云语义分割领域中基于深度学习的算法逐步占据主导地位^[2]。

基于深度学习的点云分割模型,大多采用编码器-解码器架构,网络仅由最后一层中的点云语义标签监督,而其他层中的隐藏单元缺乏直接监督,无法提取具有明确语义信息表示的特征,引入多尺度监督对于点云分割任务是十分必要的。在二维视觉领域中, Kirillov 等^[3]提出 PointRend,对低分辨率图像进行分割,迭代地对输出结果采用上采样并微调 and 再预测,直至获取最终结果,实现对不同尺度的监督。Huang 等^[4]提出 Unet3+,在每个解码层后都连接了中继监督层,将特征上采样至原图大小,利用真实标签统一进行监督,增强网络捕获边界细粒度结构的能力。在三维视觉领域中,Xu 等^[5]提出了一种场景描述符,学习表示点云场景中存在的类别,监督网络最中间层,过滤不属于场景中的类别,指导点级语义分割。

本文提出一种以 SCF-Net 为基础融合多尺度监督的点云语义分割网络 (Multi-scale Supervision SCF-Net, MSS-SCF-Net),将点云类别预测任务分解成一系列点云感受野类别预测任务,在编码阶段利用类别信息生成模块 (Category Information Generation Module, CIG) 生成类别信息,在解码阶段通过类别信息预测模块 (Category Information Prediction Module, CIP) 结合浅层以及深层编码特征对各个尺度的点云进行预测,类别信息作为标签监督隐层特征学习,增强对点云特征的挖掘能力,提升网络整体分割精度。

2 相关研究

近些年来,国内外研究人员提出了众多基于深度学习的点云分割网络,主要包括三类方法:基于投影的网络^[6-7]、基于体素化的网络^[8-9]和基于点云的网络。基于投影和基于体素化的网络由于将点云转化为其他形式的数据,导致数据额外损失,分割效果不佳。基于点云的网络能够直接对点云进行语义分割,2017年,Qi 等人^[10]提出的 PointNet 网络是该

类方法的开拓者,能够直接在点云数据上进行特征学习,利用共享感知机捕获逐点特征,但忽略了局部特征的学习。为了加强局部特征提取能力,Qi 等人^[11]在 PointNet 的基础上提出了 PointNet++,采用分层思想,逐步扩大感受野进而聚合局部特征。2019年 Zhao 等人^[12]提出了 PointWeb,通过自适应特征调整模块在每个成对的 3D 点之间交换信息,增强局部邻域的代表能力。2021年,Fan 等人^[13]提出 SCF-Net,将点云以极坐标形式表示,利用双距离注意力池化模块,基于几何距离与特征距离自动学习有效的局部特征,在室内场景的语义分割中取得了很好的效果。2023年,敖等人^[14]提出 CsegNet,利用逐点卷积和边缘卷积提取局部特征,有效改善边缘不平滑现象。以上网络模型的提出,为后续基于深度学习的点云语义分割模型指引了方向,具有重要的理论研究价值。

3 本文方法

3.1 结合多尺度监督和 SCF-Net 的神经网络

本文方法以 SCF-Net 为主干网络,采用带有跳跃连接的编码器-解码器架构,为了避免随机采样丢弃关键信息,替换为最远点采样,能够覆盖整个点云,增强网络对局部特征的提取。

网络具体结构如图 1 所示。网络的输入是大小为 $N \times D$ 的点云,其中 N 为点云个数, D 为特征维数。首先通过全连接层将特征维数统一为 8,逐步利用五个编码层对特征进行编码,使用最远点采样算法对点云进行下采样,利用 SCF 模块来学习上下文特征。同时,点云的标签信息也被输入至网络中,大小为 $N \times C$, C 为类别数目。输入的标签信息依次经过 CIG 模块和最远点采样处理。CIG 模块用于获取点云感受野中的类别信息,类别信息作为标签监督指导解码阶段点云特征提取。类别信息生成与编码阶段特征提取共享下采样,生成的目标类别信息能精确地记录感受野的类别成分。接下来,使用五个解码层对特征进行解码。采用最近邻插值进行上采样,利用跳跃连接与编码层生成的中间特征进行拼接,将拼接后的特征输入到共享 MLP。编码特征通过五个解码层逐渐恢复至原始分辨率。此外解码器中每层点云特征将输入到类别信息预测模块 (CIP) 以获取预测类别信息,利用编码阶段生成的

目标类别信息通过分层监督来指导预测,实现点级语义标签从全局到局部、从粗到细的预测。最后使

用三个连续的全连接层预测最终标签,输出大小为 $N \times C$ 的分割预测结果。

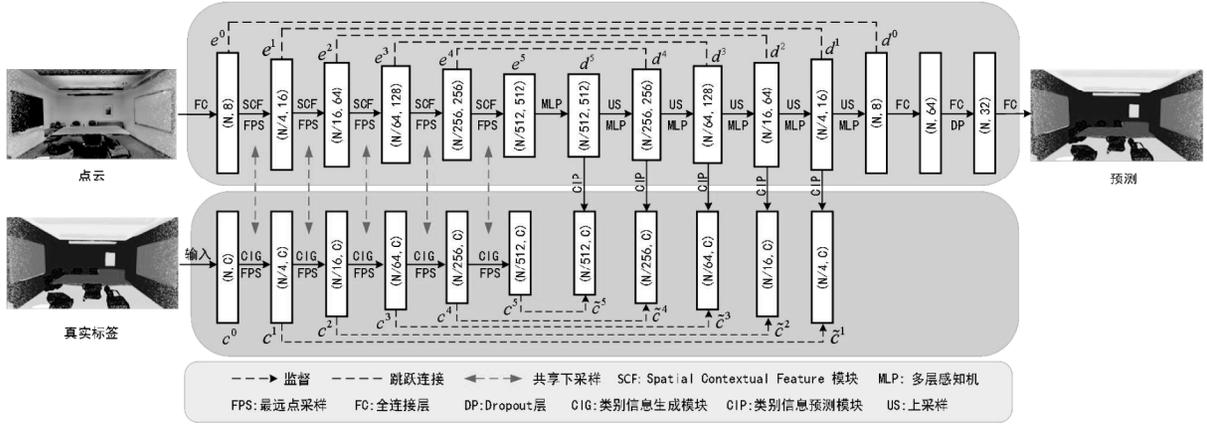


图1 整体网络结构

Fig. 1 Overall network structure

3.2 类别信息生成模块

在图像领域中,多尺度监督的应用^[4,15]通常是将解码器中每一层生成的特征上采样到原始图片的分辨率大小,添加辅助分类器,计算特征与真实标签产生的额外损失。然而,由于点云的无序性,无法像图像那样与下采样后的数据具有固定的映射关系,直接将隐层点云上采样至原始分辨率与真实标签计算损失,会对隐层特征引入错误的监督,导致网络性能的下降^[16]。

由于无法直接将解码器每层生成的点云特征上采样到原始分辨率,对下采样后的点云提供标签是非常重要的。当点云经过下采样后,点云的标签仅包含自身所对应的类别是不够的,还需要体现邻域内其他点的类别。因此,本文提出了一种类别信息生成模块,用于表示采样点云感受野内的所有类别。

本文以 p_i^k 表示编码器第 k 层中第 i 个点,将类别信息向量 c_i^k 定义为一个 $1 \times C$ 的二进制向量,表示点 p_i^k 的感受野内存在的对象类别。感受野中存在第 j 个类别, $c_i^k[j]$ 元素为 1,如果不存在,则为 0。特别的,将 c_i^0 定义为输入点云的类别信息向量,此时感受野中只含自身点,具体定义如式(1)所示:

$$c_i^0[j] = \begin{cases} 1, l_i \text{ 所对应的类别} \\ 0, \text{其他} \end{cases} \quad (1)$$

其中, l_i 是输入点云 p_i 的标签。编码器第 k 层的类别信息向量可以通过公式(2)从 $k-1$ 层推得而出。

$$c_i^k[j] = \bigvee_{n \in N(i)} \{c_n^{k-1}[j]\} \quad (2)$$

其中, n 是 p_i^k 在 $k-1$ 层感受野中的点云序号; $N(i)$

表示 p_i^k 在 $k-1$ 层感受野中的点云集合; \bigvee 表示逻辑或。将式(2)操作定义为类别信息聚合。类别信息生成模块如图 2 所示,首先将上一层的类别信息向量 (N, C) 作为输入,对于单个类别信息向量 $(1, C)$,通过 K 近邻算法获取邻域内的 K 个类别信息向量 (K, C) ,经过一次类别信息聚合操作得到能够表示其邻域的类别信息向量 $(1, C)$ 。由于基础网络的 SCF 模块在聚合局部点云特征时重复了两次 K 近邻操作,使得点云感受野扩张了两次。为了能够精准的记录感受野中的类别成分,在类别信息生成模块中将上述操作也重复两次。

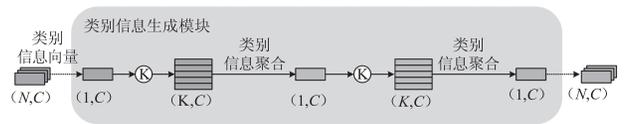


图2 类别信息生成模块

Fig. 2 Category information generation module

将前一层的类别信息向量输入到类别信息生成模块,得到后一层的类别信息向量,重复该过程,直至生成编码器最后一层点云的类别信息向量。生成的向量将用于解码阶段,作为语义标签监督指导解码器以提取更加具体、明确的语义信息,提升网络分割性能。

3.3 多尺度监督

在本文的方法中,将解码阶段的点云类别预测任务分解成一系列点云感受野类别预测任务,除解码器最后一层,剩余层都需要预测当前阶段点云的感受野类别。随着解码层数的增加,点云感受野逐渐减小,对于解码器的最后一层,感受野

只含有自身点,感受野类别预测转化为最终的语义类别预测。

为预测点云的感受野类别,本文设计了类别信息预测模块。如图3所示,解码特征 d^k 对应于编码特征 e^k, e^{k-1} 相对于 e^k 是邻近的浅层编码特征,拥有更加丰富的细粒度局部结构信息。 e^{k+1} 相对于 e^k 是邻近的高层编码特征,拥有更多的高层语义上下文信息。类别信息预测模块如图3所示,通过MLP和最远点采样将邻近浅层编码特征 e^{k-1} 与解码特征 d^k 对齐,利用MLP和最近邻插值将邻近高层编码 e^{k+1} 与 d^k 对齐。将对齐后的浅层和高层特征拼接后输

入MLP,得到融合了浅层几何信息和高层语义信息的增强特征 f^k ,定义如式(3)所示:

$$f^k = \text{MLP}(\text{FPS}(\text{MLP}(e^{k-1})) \oplus \text{NI}(\text{MLP}(e^{k+1}))) \quad (3)$$

式中,FPS表示最远点采样;NI表示最近邻插值; \oplus 表示拼接操作。最后将增强特征 f^k 与解码特征 d^k 拼接后输入MLP,通过Sigmoid函数预测类别信息向量 \tilde{c}^k ,定义如式(4)所示。

$$\tilde{c}^k = \text{Sigmoid}(\text{MLP}(d^k \oplus f^k)) \quad (4)$$

利用邻近浅层和高层编码特征能够进一步改善点云感受野类别的预测精度。

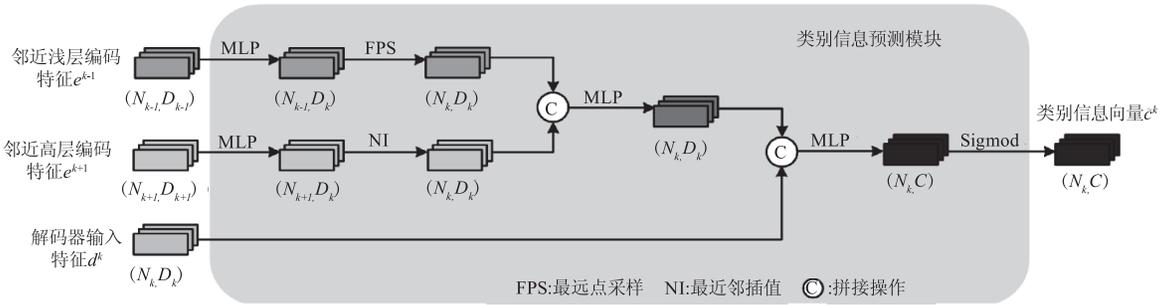


图3 类别信息预测模块

Fig. 3 Category information prediction module

编码阶段生成的目标类别信息向量 c_i^k 直接用于监督指导 \tilde{c}_i^k 的预测,损失函数的计算如式(5)所示:

$$L_c^k(i, j) = c_i^k[j] \log(\tilde{c}_i^k[j]) + (1 - c_i^k[j]) \log(1 - \tilde{c}_i^k[j]) \quad (5)$$

除了解码器的最后一层,剩余层都通过公式(6)计算本层类别信息向量预测所产生的损失函数:

$$L_c^k = -\frac{1}{CN} \sum_{i=1}^N \sum_{j=1}^C L_c^k(i, j) \quad (6)$$

式中, N 表示编码器第 k 层中的采样点云数; C 表示类别数。

在本文网络中 d^5 为解码器最中间层的特征,包含语义丰富的全局上下文信息。将 d^5 输入CIP模块中预测得到具有最大感受野的 \tilde{c}^5 ,编码阶段生成的 c^5 用于正则化 \tilde{c}^5 ,帮助 d^5 学习更加具体明确的语义表示。之后对于解码器的下一层,将 d^5 上采样并通过跳跃连接与编码器中的 e^4 特征连接后输入MLP中得到 d^4 。同样, d^4 将被输入CIP模块中用来预测得到更加局部、类别明确的 \tilde{c}^4 。整个过程持续到 $k=1$,用于正则化类别信息预测所产生的损失函数如式(7)所定义:

$$L_c = \frac{1}{5} \sum_{k=1}^5 L_c^k \quad (7)$$

对于解码器的最后一层,将输出得到的特征 d^0 输入到三个连续的全连接层,并使用softmax函数来预测最终语义类别,使用常见交叉熵损失函数来监督网络最终输出的语义标签。网络的整体损失函数如式(8)所定义:

$$L = L_n + L_c \quad (8)$$

整体损失函数由两部分构成, L_n 表示语义分割中常见交叉熵损失, L_c 表示类别信息预测损失。

4 实验结果与分析

4.1 网络参数设置

算法实验环境如下:硬件为RTX3090 24GB GPU,软件为UBUNTU16.04操作系统,CUDA11.1, TensorFlow 1.15.4。网络训练参数如下:使用初始学习率为0.01的Adam优化器训练网络,批处理大小设置为4,网络训练100个epoch,每个epoch后学习率下降5%,最近邻 K 大小为16。

4.2 实验数据集及评价指标

为了验证所提方法有效性,在两个大规模语义分割数据集S3DIS和SemanticKITTI上进行对比验证。

S3DIS 数据集是一个大型室内点云数据集,由斯坦福大学从室内工作环境中采集得到,包含 6 个区域,由 271 个房间组成。总计扫描约 2.73 亿个点,每个点包含 XYZ 坐标信息以及 RGB 颜色信息。

SemanticKITTI 是一个大规模室外场景数据集,通过机载激光扫描系统采集,拥有复杂的室外交通场景,包含 22 个点云序列,总计 45.49 亿个点,含有 19 个有效类别。

语义分割的性能评估指标采用总体精度(OA)、平均交并比(mIoU)、平均类精度(mAcc),可定义为:

$$OA = \frac{\sum_{i=0}^{N-1} P_{nn}}{\sum_{n=0}^{N-1} \sum_{m=0}^{N-1} P_{nm}} \quad (9)$$

$$mAcc = \frac{1}{N} \sum_{n=0}^{N-1} \frac{P_{nn}}{\sum_{m=0}^{N-1} P_{nm}} \quad (10)$$

表 1 不同方法在 S3DIS 数据集上的分割精度对比(单位:%)

Tab. 1 Comparison of segmentation accuracy of different methods on S3DIS dataset

Methods	mIoU	mAcc	OA	Ceil	Floor	Wall	Beam	Col
PointNet ^[10]	47.6	66.2	78.6	88.0	88.7	69.3	42.4	23.1
RSNet ^[17]	56.5	66.5	-	92.5	92.8	78.6	32.8	34.4
SPG ^[18]	62.1	73.0	86.4	89.9	95.1	76.4	62.8	47.1
PointCNN ^[19]	65.4	75.6	88.1	94.8	97.3	75.8	63.3	51.7
RandLA-Net ^[20]	70.0	82.0	88.0	93.1	96.1	80.6	62.4	48.0
KPConv ^[21]	70.6	79.1	-	93.6	92.4	83.1	63.9	54.3
SCF-Net ^[13]	71.6	82.7	88.4	93.3	96.4	80.9	64.9	47.4
BAF-LAC ^[22]	71.7	81.3	88.2	92.5	95.9	81.3	63.2	57.8
MSS-SCF-Net(Ours)	73.1	83.5	89.1	93.4	96.8	81.7	63.7	50.7
Methods	Wind	Door	Table	Chair	Sofa	Book	Board	Clut
PointNet ^[10]	47.5	51.6	54.1	42.0	9.6	38.2	29.4	35.2
RSNet ^[17]	51.6	68.1	59.7	60.1	16.4	50.2	44.9	52.0
SPG ^[18]	55.3	68.4	73.5	69.2	63.2	45.9	8.7	52.9
PointCNN ^[19]	58.4	57.2	71.6	69.1	39.1	61.2	52.2	58.6
RandLA-Net ^[20]	64.4	69.4	69.4	76.4	60.0	64.2	65.9	60.1
KPConv ^[21]	66.1	76.6	57.8	64.0	69.3	74.9	61.3	60.3
SCF-Net ^[13]	64.5	70.1	71.4	81.6	67.2	64.4	67.5	60.9
BAF-LAC ^[22]	63.0	79.9	70.3	74.6	60.6	67.2	65.3	60.4
MSS-SCF-Net(Ours)	66.2	72.5	73.2	85.4	70.5	65.7	68.0	62.5

图 4 为本文算法与 SCF-Net 网络的可视化分割结果图。其中图 4(a)为输入点云,图 4(b)为 SCF-Net 方法的分割结果,图 4(c)为本文方法的分割结果,图 4(d)为参考标准。从上至下依次为房间 1、房间 2 和房间 3。可以看出,相较于 SCF-Net,本文

$$mIoU = \frac{1}{N} \sum_{n=0}^{N-1} \frac{P_{nn}}{\sum_{m=0}^{N-1} P_{nm} + \sum_{m=0}^{N-1} P_{mn} - P_{nn}} \quad (11)$$

式(9)、(10)、(11)中, N 为类别数; n 为真实类别; m 为预测类别; P_{nn} 为类别预测正确点数; P_{nm} 表示假负例点数; P_{mn} 表示假正例点数。

4.3 实验结果及分析

4.3.1 对比实验

(1) S3DIS

在 S3DIS 数据集上采用 6 折交叉验证方法,总体精度、平均交并比和平均类精度作为度量指标。表 1 为本文方法与其他方法的度量指标对比。由表 1 可知,本文方法在总体精度、平均交并比和平均类精度优于其他方法,并且在 13 个类别中的 5 类别(窗户、椅子、沙发、木板和杂物)取得最佳性能。与 RandLA-Net 相比,本文方法在 mIoU 提高了 3.1%,与基础网络 SCF-Net 相比,mIoU 提高了 1.5%,且在 12 个类别中的 IoU 均有提升。

方法在物体的细节处取得了更好的分割效果。在房间 1 中,SCF-Net 将桌腿错误的划分为了杂物,圆柱与墙面连接处划分不够平滑,造成了圆柱的局部丢失;在房间 2 中,SCF-Net 对于门框与墙之间的边界划分的不够清晰,且书架中的局部位置出现分割错

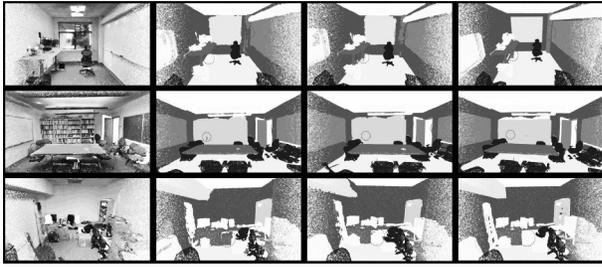


图4 S3DIS数据集分割结果可视化对比

Fig. 4 Visualization comparison of segmentation results of S3DIS dataset
 误;在房间3中,SCF-Net 错误地将颜色相近的杂物划分为了桌子。而本文方法在以上位置取得了相对较好的分割效果,说明多尺度监督的引入加强了网络局部信息挖掘能力,提升了各类目标的轮廓分割能力,对于物体连接处的分割更加平滑。

表2 不同方法在 SemanticKITTI 数据集上的分割精度对比(单位:%)

Tab. 2 Comparison of segmentation accuracy of different methods on SemanticKITTI dataset

Methods	mIoU	car	bicycle	motorcycle	truck	other-vehicle	person	bicyclist	motorcyclist	road
RangeNet53 ++ ^[23]	52.2	91.4	25.7	34.4	25.7	23.0	38.3	38.8	4.8	91.8
SCF-Net ^[13]	53.3	93.5	27.1	28.9	37.5	35.9	49.1	44.4	5.3	90.5
RandLA-Net ^[20]	53.9	94.2	26.0	25.8	40.1	38.9	49.2	48.2	7.2	90.7
PolarNet ^[24]	54.3	93.8	40.3	30.1	22.9	28.5	43.2	40.2	5.6	90.8
BAF-LAC ^[22]	54.9	94.0	27.6	30.4	39.9	40.1	46.3	49.1	10.5	90.9
MSS-SCF-Net(Ours)	55.2	94.6	28.9	28.6	41.7	38.7	46.5	47.1	10.5	91.2
Methods	parking	sidewalk	other-ground	building	fence	vegetation	trunk	terrain	pole	traffic-sign
RangeNet53 ++ ^[23]	65.0	75.2	27.8	87.4	58.6	80.5	55.1	64.6	47.9	55.9
SCF-Net ^[13]	58.2	73.3	19.7	88.3	57.7	81.1	59.5	66.1	50.6	45.5
RandLA-Net ^[20]	60.3	73.7	20.4	86.9	56.3	81.4	61.3	66.8	49.2	47.7
PolarNet ^[24]	61.7	74.4	21.7	90.0	61.3	84.0	65.5	67.8	51.8	57.5
BAF-LAC ^[22]	61.4	74.0	22.2	88.2	57.0	81.3	61.7	65.3	49.8	53.9
MSS-SCF-Net(Ours)	61.7	75.6	20.0	89.5	59.5	82.0	63.5	68.3	52.3	48.0

将 SCF-Net 与本文 MSS-SCF-Net 在 SemanticKITTI 的验证集上进行可视化结果对比,如图 5 所示。可以看出,SCF-Net 错误地将行人识别为骑脚踏车者,而将骑脚踏车者辨别为行人以及骑机车者,这三者在外形上极其相似且体型较小,导致未能正确识别。并且对于形状类似的植被和栅栏,SCF-Net 也容易将其混淆。而本文方法对于上述情况均有改善,能够正确识别行人与骑机车者这些小目标物体。这得益于多尺度监督的引入,加强网络对隐层特征的学习,提升对小目标物体及相似物体的识别能力。

4.3.2 消融实验

为了定量评估多尺度监督和最远点采样的有效性,对这两种方法不同组合进行了消融实验。选取

(2) SemanticKITTI

在 SemanticKITTI 测试集上的对比实验结果如表 2 所示。从表 2 可以看出,本文方法在点云的平均交并比上,相比基于投影的主流方法 RangeNet53 ++ 提升了 3.0 %,相较于基础网络 SCF-Net 提升了 1.9 %。SCF-Net 采用随机采样处理大规模点云数据,利用 SCF 模块学习空间上下文特征,逐步扩大点的感受野范围,防止丢失关键点信息。但对于稀疏性较大的 SemanticKITTI 数据集,随机采样无法避免丢失物体边缘信息,导致网络出现错误分割。本文方法采用最远点采样保留物体边缘的完整,通过对多个尺度点云特征进行监督学习,促使隐藏层学习到的特征易区分、更具鲁棒性,增强网络对物体边缘的分割能力。

S3DIS 数据集区域 5 为测试集,其他区域为训练集,结果如表 3 所示。(1)基线方法为 SCF-Net,其采用的点云采样算法为随机采样,在区域 5 的 mIoU 为 63.4 %。随机采样算法的使用可能会导致某些关键点特征丢失,使得网络对物体局部描述能力下降。(2)将基线方法的随机采样替换为最远点采样后,网络在区域 5 的 mIoU 提升为 64.1 %,这是因为最远点采样能够更好覆盖整个点云,不会导致物体某一部分经过下采样后丢失,有效保留目标几何结构,增强网络特征提取能力以及稳定性。(3)将多尺度监督引入基线方法,使得网络在区域 5 的 mIoU 提升为 65.2 %,原因在于多尺度监督引入了辅助分类器,能够判断隐藏层中点云特征图质量的好坏,促使隐藏层学习

过程更加直接透明,网络提取到语义明确的点云特征。(4)将多尺度监督与最远点采样算法组合使用,网络模型的分割性能得到了进一步改善,分割精度达到最好效果,在区域6的mIoU提升至65.8%。

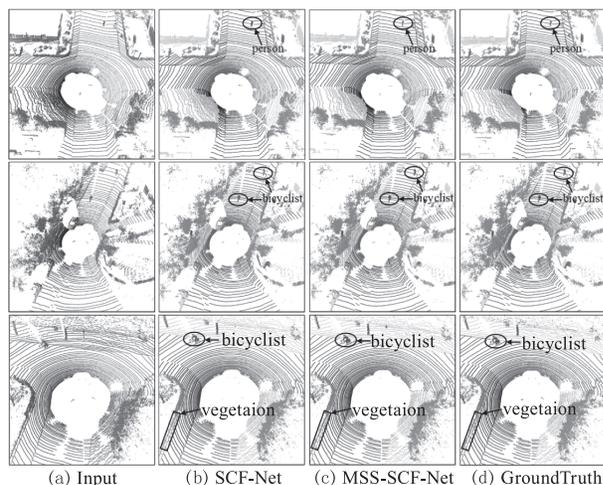


图5 SemanticKITTI数据集分割结果可视化对比

Fig.5 Visualization comparison of segmentation results of

SemanticKITTI dataset

表3 S3DIS区域5的消融实验结果

Tab.3 Results of ablation experiments in region 5 of S3DIS

随机采样	最远点采样	多尺度监督	mIoU/%
√	×	×	63.4
×	√	×	64.1
√	×	√	65.2
×	√	√	65.8

实验进一步探究了浅层和高层编码特征对类别信息预测模块的影响,如表4所示,当只利用解码特征去预测类别信息向量时,网络在区域5的mIoU为64.9%。同时考虑解码特征以及对应的浅层和高层编码特征,网络在区域5的mIoU提升为65.8%。浅层编码特征拥有更为具体的空间信息,高层编码特征拥有更为抽象的语义信息,二者的结合进一步增强类别信息的预测精度,从而提高整个网络的分割能力。

表4 类别信息预测模块消融实验结果

Tab.4 Ablation experimental results of category information prediction module

Method	mIoU/%
类别信息预测模块(仅考虑解码特征)	64.9
类别信息预测模块	65.8

5 结 语

针对室内以及室外大场景的语义分割,本文提出了一种结合多尺度监督和SCF-Net的神经网络,将解码阶段的点云类别预测任务分解成一系列点云感受野类别预测任务,通过对解码器中每一层进行额外的监督,以在分割任务中提取信息丰富语义明确的特征,进一步提升网络对各类物体的分割效果。针对SCF-Net采用随机采样算法可能导致关键信息丢失这一问题,本文将其替换为最远点采样算法,能够更好的覆盖整个点云,增强网络对局部特征的提取能力。实验结果表明,多尺度监督的引入能够有效提高网络整体分割精度,促使网络隐藏层学习到的特征易区分、更具鲁棒性,改善网络对物体边缘的分割效果。当然,本文提出的网络模型还有很多地方需要改善,如何进一步简化模型、降低模型复杂度是下一步的工作重点。

参考文献:

- [1] Liang Zhiqiang, Chen Chunmei, Chen Yanjie, et al. Point cloud classification and segmentation network based on double local features fusion of high-dimensional and low-dimensional[J]. Laser & Infrared, 2022, 52(10): 1557 - 1564. (in Chinese)
梁志强, 陈春梅, 陈妍洁, 等. 高低维双局部特征融合的点云分类分割网络[J]. 激光与红外, 2022, 52(10): 1557 - 1564.
- [2] Zhang Chunjiao, Xu Shenghua, Liu Zhengjun, et al. Semantic segmentation of point cloud based on normal vector and RandLA - Net[J]. Science of Surveying and Mapping, 2022, 47(7): 126 - 134. (in Chinese)
张春娇, 徐胜华, 刘正军, 等. 顾及法向量的点云语义分割方法[J]. 测绘科学, 2022, 47(7): 126 - 134.
- [3] Kirillov A, Wu Y, He K, et al. Pointrend: Image segmentation as rendering [C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020: 9799 - 9808.
- [4] Huang H, Lin L, Tong R, et al. Unet 3 + : a full-scale connected unet for medical image segmentation [C]//ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020: 1055 - 1059.
- [5] Xu J, Gong J, Zhou J, et al. Scene encoder: scene-aware semantic segmentation of point clouds with a learnable scene descriptor [C]//Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence, 2021: 601 - 607.

- [6] Ahn P, Yang J, Yi E, et al. Projection-based point convolution for efficient point cloud segmentation[J]. *IEEE Access*, 2022, 10: 15348 – 15358.
- [7] Kellner M, Stahl B, Reiterer A. Fused projection-based point cloud segmentation[J]. *Sensors*, 2022, 22(3): 1139.
- [8] Meng H Y, Gao L, Lai Y K, et al. Vv-net: voxel vae net with group convolutions for point cloud segmentation [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 8500 – 8508.
- [9] Zhao L, Xu S, Liu L, et al. SVASeg: sparse voxel-based attention for 3D LiDAR point cloud semantic segmentation [J]. *Remote Sensing*, 2022, 14(18): 4471.
- [10] Qi C R, Su H, Mo K, et al. Pointnet: deep learning on point sets for 3D classification and segmentation [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017: 652 – 660.
- [11] Qi C R, Yi L, Su H, et al. Pointnet ++: deep hierarchical feature learning on point sets in a metric space [J]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [12] Zhao H, Jiang L, Fu C W, et al. Point web: enhancing local neighborhood features for point cloud processing [C]//*2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019: 5560 – 5568.
- [13] Fan S, Dong Q, Zhu F, et al. SCF-net: learning spatial contextual features for large-scale point cloud segmentation [C]//*2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021: 14499 – 14508.
- [14] Ao Jian Feng, Pan Zhong Tai, Chen Xiao Long. Research on indoor semantic segmentation of 3D point cloud based on CSegNet [J]. *Laser & Infrared*, 2023, 53(2): 194 – 201. (in Chinese)
敖建锋, 潘仲泰, 程小龙. 基于 CSegNet 的三维点云室内语义分割研究 [J]. *激光与红外*, 2023, 53(2): 194 – 201.
- [15] Zhou Z, Rahman Siddiquee M M, Tajbakhsh N, et al. Unet ++: a nested u-net architecture for medical image segmentation [C]//*Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop*, 2018: 3 – 11.
- [16] Gong J, Xu J, 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 (CVPR)*, 2021: 11668 – 11677.
- [17] Huang Q, Wang W, Neumann U. Recurrent slice networks for 3D segmentation of point clouds [C]//*2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018: 2626 – 2635.
- [18] Landrieu L, Simonovsky M. Large-scale point cloud semantic segmentation with superpoint graphs [C]//*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018: 4558 – 4567.
- [19] Li Y Y, Bu R, Sun M C, et al. PointCNN: convolution on X-transformed points [C]//*32nd Conference on Neural Information Processing Systems (NIPS)*, 2018.
- [20] Hu Q, Yang B, Xie L, et al. RandLA-net: efficient semantic segmentation of large-scale point clouds [C]//*2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 11105 – 11114.
- [21] Thomas H, Qi C R, Deschaud J-E, et al. Kpconv: flexible and deformable convolution for point clouds [C]//*Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 6411 – 6420.
- [22] Shuai H, Xu X, Liu Q. Backward attentive fusing network with local aggregation classifier for 3D point cloud semantic segmentation [J]. *IEEE Trans Image Process*, 2021, 30: 4973 – 4984.
- [23] Milioto A, Vizzo I, Behley J, et al. Rangenet ++: fast and accurate lidar semantic segmentation [C]//*2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019: 4213 – 4220.
- [24] Zhang Y, Zhou Z, David P, et al. Polarnet: an improved grid representation for online lidar point clouds semantic segmentation [C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9601 – 9610.