改进 RangeNet++ 损失函数的车载点云小目标语义分割方法

张硕^{1,2)}, 叶勤^{1)*}, 史婧¹⁾, 刘行¹⁾

¹⁾ (同济大学测绘与地理信息学院 上海 200092)

²⁾ (自然资源部城市国土资源监测与仿真重点实验室 深圳 518034)
(yeqin@tongji.edu.cn)

摘 要: 针对道路场景点云全语义分割对行人等重要移动小目标实时分割效果差的问题, 提出一种基于 RangeNet++ 深度神经网络并对损失函数进行优化改进的小目标语义分割方法. 首先对原本使用交叉熵损失函数的 RangeNet++ 网络进行改进; 然后采用 Focal Loss 损失函数调节稀少但重要的移动小目标的权重, 能在卷积层数更少的 DarkNet21 特征提取网络下, 通过少次训练就能提高行人等重要移动小目标类别的检测和分割精度. 在 SemanticKITTI 数据集上的实验表明, 与原有的 RangeNet++ 相比, 该方法使用的 backbone 卷积层数和网络训练次数都更少, 但对移动小目标的语义分割达到了更高的准确率和精度.

关键词: 车载点云; 移动小目标; 损失函数; 实时语义分割

中图分类号: TP391.41 DOI: 10.3724/SP.J.1089.2021.18581

A Semantic Segmentation Method of In-Vehicle Small Targets Point Cloud Based on Improved RangeNet++ Loss Function

Zhang Shuo^{1,2)}, Ye Qin^{1)*}, Shi Jing¹⁾, and Liu Hang¹⁾

¹⁾ (College of Surveying and Geo-Information, Tongji University, Shanghai 200092)

²⁾ (Key Laboratory of Urban Land Resources Monitoring and Simulation, Ministry of Natural Resources, Shenzhen 518034)

Abstract: Pedestrians and other moving small-targets plays an important role in road scene point cloud. However, they perform poorly for real-time semantic segmentation. Therefore, a small-target semantic segmentation method based on RangeNet++ and optimization of loss function is proposed. Firstly, the Focal Loss function is used to improve the original RangeNet++ which uses cross entropy as loss function. Then, the weight of some rare, important moving small-targets in Focal Loss function is adjusted. The proposed method can improve the detection and segmentation accuracy of important moving small-targets categories such as pedestrians with fewer training sessions under the DarkNet21 which has fewer convolutional layers. Experiments on the SemanticKITTI dataset show that compared with the original RangeNet++, the improved method uses fewer convolutional layers backbone and fewer training sessions, but achieves higher accuracy and precision in semantic segmentation of moving small-targets.

Key words: in-vehicle point cloud; moving small-targets; loss function; real-time semantic segmentation

收稿日期: 2020-08-18; 修回日期: 2021-02-23. 基金项目: 国家自然科学基金(41771480); 自然资源部城市国土资源监测与仿真重点实验室开放基金资助课题(KF-2019-04-015). 张硕(1995—), 女, 硕士研究生, 主要研究方向为摄影测量与遥感、点云数据处理; 叶勤(1970—), 女, 博士, 副教授, 博士生导师, 论文通讯作者, 主要研究方向为数字摄影测量与遥感、计算机视觉; 史婧(1997—), 女, 学士; 刘行(1995—), 男, 硕士研究生, 主要研究方向为计算机视觉.

在自动驾驶领域,通过对车载 LiDAR 获取道路场景点云数据并进行实时语义分割,得到道路、车辆、行人和障碍物等的准确位置信息是研究的热点,而其中移动小目标(如行人等)的准确分割既是保证驾驶安全的重要内容也是其难点。传统点云分割方法一般可分为基于边缘、区域增长、属性聚类、模型拟合以及基于图^[1]的算法;但其存在以下问题:(1) 分割的规则和特征依赖于人为设计,泛化能力差^[2];(2) 多阶段的处理流程会产生复合累积错误,在缺乏目标周围的环境信息时,对单帧激光扫描的点云的分割计算时间和精度不稳定,不能满足应用要求。为解决上述问题,人们将深度神经网络引入点云语义分割。在图像分割中, Alex-Net^[3]在大规模图像数据集上分割精度和效果远优于传统的 hand-craft 特征方法;随后的 VGG^[4], GoogLeNet^[5], ReNet^[6]和 ResNet^[7]等经典模型对 2D-2.5D 图像分割取得了瞩目的优良效果^[8]。但对多为无序的 3D 点云数据,直接采用深度学习处理仍是难点。

点云的深度学习语义分割通常分为基于点云、体素、投影和多视图 4 种。PointNet^[9]及 PointNet++^[10]可直接基于原始不规则的无序点云进行处理,使用共享参数的多层感知机(multi layer perceptron, MLP)学习逐点特征,并使用对称池函数获取全局特征;PointNet++弥补了 PointNet 缺失局部特征的缺点,以层次结构学习点云特征的深度网络,分层渐进式地获取较大局部区域每个点附近的精确几何特征,实现局部特征与全局特征的融合。但其体系结构复杂,计算复杂度高且耗时多,多用于静态点云,在动态连续的自动驾驶场景中较少使用。SqueezeSeg^[11-12]对点云数据进行球面投影后得到深度图,将其输入到超轻量级全卷积神经网络(fully convolutional networks, FCN)SqueezeNet^[12]进行特征提取;经条件随机场(conditional random field, CRF)平滑校正后,逐像素地进行分类,其边界模糊;再使用 DBSCAN 算法^[13]进行聚类,最后将聚类分割后的深度图中的点离散化到 3D 空间中。其模型具有参数较少、运算速度快、内存消耗少以及分割效果佳等特点,更符合车载场景需求。在 SqueezeSeg 和 SqueezeSegV2^[14]的基础上提出的 RangeNet++^[15]模型,沿用了对点云的球形投影,但是使用 Darknet^[16]特征提取骨架,且用基于 GPU 的 KNN (*K*-nearest neighbors)搜索代替 CRF,实现自动驾驶场景下 LiDAR 点云准确、快速、实时地全语义分割。RangeNet++可在任何基于序列图像

的卷积神经网络(convolutional neural networks, CNN)主干上运行,推断出完整的原始点云的语义标签;同时仍以帧速率运行而不受限于 LiDAR 旋转扫描频率,适用于机器人和移动平台。

上述深度学习网络对大类点云分割效果较好,但对场景中小目标语义分割的精度和准确率较低。因小目标占比少、易与背景混淆、同时自身包含的信息有限,一直是目标检测和分割中未很好解决的问题,特别是对车载场景中的行人等小目标。本文针对自动驾驶获取的车载点云,基于 RangeNet++模型,就造成点云小目标分割效果不佳的因素展开分析,针对样本的类别不平衡和简单/困难样本不平衡这 2 个关键问题,对模型原有的 one-stage detector 交叉熵(cross entropy, CE)损失函数进行分析,改进损失函数为 two-stage 的 Focal Loss(FL)损失函数^[17],并权衡样本的不均衡性、调节不同类别的权重分配,重点提高道路车载点云场景中小目标语义分割的效果,在此基础上加快网络收敛速度。本文利用 KITTI 数据集进行实验,验证了损失函数改进并调整小目标权重的方法在车载点云小目标语义分割上具有良好效果;同时使用了卷积层更少的 Darknet21 骨架,达到了比原有 RangeNet++模型对行人等小目标更高的分割准确率和精度。

1 小目标与 RangeNet++

1.1 小目标的检测与分割

小目标中“小”的定义,是指待检测的目标自身真实的空间尺寸过小或距传感器较远导致其在整个场景中占比较小^[18]。小目标检测识别结合了定位与检测 2 项任务,要求在复杂场景中准确地找到每一个目标的精确边框,并判断其所属类别。其研究现存 2 个问题:(1) 依赖传统的目标检测中经典成熟的方法,但后续推理速度缓慢,不能满足实时性需求;(2) 现有的方法泛化性差,随小样本类别增加,检测分割效果明显下降^[19]。由于小目标整体数量占比较小,点云簇相对较小且不具有规则形状,样本数量更少,行人小目标大多处于缓慢运动状态,传统的语义分割方法很难将其有效地识别并分割出来;而现有的深度学习方法大都也因此出现训练过程中学习率不够而导致欠拟合,经常将其与周围构建筑物识别为一体、形成遮挡,或者根本无法识别分割较远处的动态小目标,整体检测分割精度较低。目前已经具有的较强学习能力的 CNN 在深度神经网络架构不变的基础上,

融入了具有更多先验信息的高阶 CRF, 引入多尺度特征的概念, 文献[20-22]分别以多尺度卷积核、多尺度输入图像和多尺度池化层的方式, 将不同大小的感受野融合到单个神经网络模型中^[23], 它们均通过增加网络结构以实现对小目标的有效分割, 但对硬件资源消耗较大. 总结复杂场景下小目标分割的挑战如下: (1) 较难分割小目标或目标中的条状部位; (2) 在模糊边缘处较难分割精确的轮廓; (3) 较难区分不同目标的相似部位^[23]. 在自动驾驶领域, 行人及其非机动车工具在道路场景中相对于周围固定的构建筑物属于小目标, 且在连续场景中会发生短距离的相对移动变化; 它们的准确识别对无人驾驶场景相当重要, 其错误分割会带来严重的安全隐患. 因此, 提高道路场景车载点云的小目标高效准确分割是一个亟需解决的问题.

1.2 RangeNet++原理

因 3D 场景中数据标注比在 2D 图像中更困难, 导致 3D 点云训练样本缺乏, 现阶段 2D 中的训练数据远比 3D 中丰富. RangeNet++模型^[15]将点云投影到 2D 图像中解决训练样本不足问题, 利用 2D CNN 对深度图分割后, 再使用 KNN 后置算法将其映射回 3D 空间, 得到点云分割结果. 其使用 YOLO V3^[23]中提出的特征提取骨干网络 Darknet53, 借鉴金字塔特征图思想并结合 ResNet 的特点, 既保证对特征进行超强表达, 又避免了网络过深带

来的梯度问题, 但其卷积层较多, 对机器性能要求较高.

1.3 RangeNet++模块构成及本文方法流程

针对车载点云中数量少但重要的移动小目标分割精度和准确度不佳的问题, 本文改进并优化了基于实时分割效率较高的 RangeNet++方法^[15]. 现有方法 CE 损失函数对样本只有 positive/negative 的区分, 无 easy/hard 的区分, 由此对损失函数重新考量, 并将其中的 Darknet53 换为卷积层数更少的 Darknet21(含残差块以提高特征提取能力)来降低计算要求, 提升对车载点云场景的小目标分割效果与效率, 具体流程如图 1 所示. RangeNet++全卷积语义分割结构^[15]如图 2 所示.

图 1 基于损失函数改进的处理流程

图 2 RangeNet++全卷积语义分割结构^[15]

图 1 中, RangeNet++将每一个去偏斜点云 $P_i = (x, y, z)$ 转换为距离表示, 通过公式^[15]

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{1}{2} [1 - \arctan(x, y) \pi^{-1}] \omega \\ [1 - (\arcsin(z, r^{-1}) + f_{up}) f^{-1}] h \end{pmatrix} \quad (1)$$

映射到球面坐标系, 得到图像坐标系的深度图. 其中, (u, v) 为图像坐标系; (h, w) 是期望深度图像表示的高度和宽度; $f = f_{up} + f_{down}$ 表示传感器的垂直俯视图; $r = \|P_i\|_2$ 表示每个点到传感器的距离.

用 Darknet21 对深度图进行特征提取和语义分割: 将转化后的数据输入带有显著下采样编码器的 2D 语义分割 CNN, 只在水平方向执行下采样. 距离图像中每个像素的概率分布为

$$\hat{y}_c = \frac{\exp(M_{\text{Logit}_c})}{\sum_{c=1}^C \exp(M_{\text{Logit}_c})} \quad (2)$$

其中, M_{Logit_c} 表示 Logit 模型.

用距离信息、像素坐标和传感器参数来实现 $\mathbb{R}^2 \rightarrow \mathbb{R}^3$ 的映射: 将距离图像与对应于每个点的图像坐标进行配对和索引, 快速无损地生成语义

标签,用KNN方法获取最终3D点云的语义分割结果(认为相邻点语义类别相同)。

本文改用FL损失函数^[17]解决对“行人”类的重要小目标的分割精度和准确率不甚理想的问题;采用卷积层数更少的Darknet21骨架代替结构复杂、对GPU要求高的Darknet53,提高分割效率,节约计算资源。

2 FL损失函数优化与网络训练

2.1 损失函数选择与改进

现有RangeNet++模型采用对每个类别增加权重的CE损失函数来缓解数据不平衡问题,即

$$CE_{loss} = -\sum_{i=1}^n y_i \cdot CE(p_{ii}) = -\sum_{i=1}^n y_i \cdot \text{lb}(p_{ii}) \quad (3)$$

其中, $y \in \{-1, +1\}$ 表示类别标签;

$$p_i = \begin{cases} p, & \text{if } y = 1 \\ 1-p, & \text{otherwise} \end{cases} \quad (4)$$

CE损失函数为只考虑样本中positive/negative区分的one-stage检测算法,单独评估每个像素矢量的类预测;然后对所有像素求平均值,即无差别平等学习图像中的像素。实际上,不同类别样本可根据交并比(intersection over union, IoU)分为hard positive(IoU>0.5), easy positive, hard negative(IoU<0.4)以及easy negative 4类,现实自动驾驶场景中常出现类别不均衡(class imbalance)的问题,由此导致训练过程被像素或点云较多的类主导,很难学习到较小物体的特征,即网络有效性较低。CE不能解决由于样本分类难度不平衡、样本数量稀少导致的小目标分割效果差的问题。为此,本文选用two-stage的FL损失函数^[17],加大hard negative类别的权重,优化训练,重点学习和提取困难样本特征。通常two-stage检测速度较one-stage慢,但准确性高,具体公式为

$$FL(p_i) = -\alpha_i (1-p_i)^\gamma \text{lb}(p_i) \quad (5)$$

其中, γ 表示注意力参数($\gamma \geq 0$),取 $\gamma=0$ 时即为CE,其值越大,调节因子 $(1-p_i)^\gamma$ 的影响越大,简单负样本对模型的贡献越低; $p \in [0,1]$ 是类别预测概率。

相较于CE而言,FL关注于困难样本稀疏子集,增加类别权重 α_i 对FL函数起到平衡作用,增加权重调节因子 $(1-p_i)^\gamma$,减少easy example样本的损失贡献,并拓宽了样本接收低损失的范围,提升易错分类样本的重要性和模型精确性,避免在

训练过程中大量的简单负样本淹没检测器; γ 增强了调节因子的作用,其值越大,权重调节因子的影响越大,简单负样本对模型的贡献越低。由于行人等重要但难以识别的动态小物体数量远小于其他类别,结合FL的特性均衡样本分布差异性,本文在模型训练时对其设置类别权重大于其他样本,同时结合Sigmoid激活函数提升数据稳定性,以提高网络模型的准确度。

2.2 网络模型优化及训练

2.2.1 数据集介绍

KITTI数据集是目前国际上最大的自动驾驶场景下的计算机视觉算法评测集,由389对立体图像和光流图、39.2 km点云序列,以及超过20万个3D标注物体的图像组成,以10 Hz频率采样并同步。原始数据集总体分为Road, City, Residential, Campus和Person;按各个类占总区域的比值,行人小目标大约只有1.23%。

本文使用数据集SemanticKITTI^[24]是对KITTI视觉里程计所有序列全类别标注(20类),提供了汽车激光雷达的完整360°视野的密集逐点注释。转换后的数据集共11个序列(11600帧,共124669个点云数据);其中,序列00~09为训练集,10为验证集,11~21作为测试集标签未公开;验证模型分割性能需交由SemanticKITTI网站进行测试。基于该数据集的基准任务^[24]: (1) 使用单个扫描对点云进行语义分割; (2) 使用之前的多次扫描进行语义分割; (3) 对未来的语义场景进行预测,完成语义场景理解。为进行后续点云分割的训练与评测,需将标注文件从相机坐标系转换到激光雷达点云坐标系,映射关系为

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} f_u & 0 & u_0 \\ 0 & f_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (6)$$

其中, (u,v) 表示相机数据坐标; (x,y,z) 表示LiDAR点云; f_u 表示相机水平方向焦距, f_v 表示相机垂直方向焦距值; u_0 表示像素坐标系的中心点横坐标, v_0 表示像素坐标系中心点纵坐标, R 表示旋转矩阵; t 表示平移分量。

2.2.2 模型优化及最佳参数选择

本文实验基于6GB的GeForceGTX 1060显卡

进行训练, 考虑模型优化速度以及超参数鲁棒性等, 采用动量梯度优化算法(Momentum)作为网络优化器, 以提高收敛速度, 防止将完整样本划分为若干小 batch 后下降过程中产生扰动, 在一定程度上具有增加稳定性、摆脱局部最优的能力. 训练时将参数初始化: 初始学习率为 0.01, batch_size=2, 以注意力参数 β (代表 v_i 的权重) 初始化为 0.9, 带动量 v_i 梯度下降来寻找模型参数的最优解并更新, 学习率衰减因子 lr_decay 为 0.995, 参数更新为

$$v_t = \gamma v_{t-1} + \eta \nabla_{\theta} J(\theta) \quad (7)$$

$$\theta = \theta - v_t \quad (8)$$

预训练迭代 10 次共 32h, 损失函数值下降趋势如图 3 所示, 易看出改进 FL 函数后较 CE 函数收敛速度明显加快.

图 3 损失函数值下降过程(每 50 batch 间隔取值)

实验数据集规模较大而计算资源有限, 每次基于当前结果微调网络超参数, 对比设置 FL 函数中小类别 $\gamma=2$, 对无权值的 FL 函数测试结果如图 4 所示. 可以看到, 整体 IoU 与同样使用 Darknet21 作为 backbone 的 CE 方法略有提高, 但 Person 类别识别效果相差不大, 即采用不依赖权重参数的 FL 函数也能泛化得到与依赖权值的 CE 函数达到同样的效果.

图 4 无权值的 FL 函数测试结果

后续重点针对与 Person 相关的 Motorcycle,

Bicycle, Bicyclist 和 Motorcyclist 类别调节权重, 每次权重设置增进 0.1 再进行网络训练; 其余非重要类取 0.2, 并对受影响的 Car 权重进行优化, 如图 5 所示.

图 5 Car 类 α_i 参数调整分析

经过多次实验最终调整 Person 权重为 3, Motorcycle/Bicycle 权重为 2, Bicyclist/Motorcyclist 权重为 1, 对预训练模型进行微调, 在此基础上设置 20 个 epoch 进行训练, 最终值极小, 趋于平缓, 如图 6 所示. 使用该最佳参数模型在验证集上得到了对 Person 较好的分割效果, 并生成序列 11~21 测试集语义标签并提交到 SemanticKITTI 网站验证分割效果.

图 6 训练损失函数随 epoch 增加而下降的趋势

3 小目标点云语义分割实验及评价

对 SemanticKITTI 中测试数据集中的连续动态车载点云数据, 利用第 2.2.2 节优化训练好的模型进行语义分割实验. 为比较本文方法对重要小目标语义分割性能, 实验了 3 种方法: 本文采用 Darknet21 骨架改进 FL 损失函数并调节权重参数 (方法 1), RangeNet++ 模型分别采用 Darknet21 骨架和 CE 损失函数 (方法 2), 以及 Darknet53 骨架和 CE 损失函数 (方法 3)^[15], 完成了对车辆行驶过程中动态场景的实时精准语义分割, 生成动态深度图和分割场景.

3.1 点云语义分割效果的定性评价

对行进过程中动态获得的点云进行球面投影后得到深度图,图7和图8所示为将训练好的模型对深度图进行语义分割的结果,标记 Car 类为蓝色,Person 类为红色,Road 类为紫色,未分类为黄色。可以看出,本文方法对 Bicycle 和 Motorcyclist 的实时语义分割识别更清晰。分析图7和图8中的这2帧场景点云发现,方法3未能分割出主干道上

的大量行人以及非机动车辆;而方法1对主干道不同景深的不同类型移动小目标分割正确且更完整,包括远处体量极小的行人和自行车(图7和图8中间的白线圈内)、部分身体被车辆遮挡的行人(图7右侧),以及正在过马路的行人及非机动车(图8右侧),且将非机动车与行人类别分离,语义信息更精细、准确;而方法3对图中这些都未能正确识别。

图7 场景1深度图对比

图8 场景2深度图对比

使用 KNN 方法将语义分割后的深度图映射回 3D 点云空间并可视化,与方法3对比明显看出本文方法的优越性。如图9所示,对白线圈中的行人,方法1对驾驶场景下行人类别分割效果良好且清晰完整,中间的人物头部和四肢清晰可见,同时不同类别的点云分割也十分清晰;而方法3对白线圈内左边人物的头部识别效果较差,中间人物的上半身都无法辨别,效果逊于本文方法。

如图10所示,方法1在点云稀疏且类别杂乱的区域也能将行人等移动小目标分割出来(图10中白线圈处),而方法3没有识别分割出栅栏和建筑物交界处的行人。结合实际出行情况和分割效果图像可知,Person类经常出现在道路两旁,与建筑物、树木和栅栏等比较接近,存在一定的遮挡情况,出现部分误分类情况。

图9 点云语义分割效果对比1

图10 点云语义分割效果对比2

3.2 评价指标

由于序列11~21标签未公开,因此将本文方法所的结果提交到 SemanticKITTI 网站进行评估,采用得到均交并比值(mean intersection-over-union, mIoU)作为评价指标展示语义分割效果,如表1所示。其中,TP代表预测为正且预测正确的数量,TN代表预测为负且预测正确的数量,FP代表预测为负且预测错误的数量,FN代表预测为负且预测错误的数量。mIoU表征预测结果集合和真实结果集合的交集与并集的比值,即检测的准确率,

$$mIoU = \frac{1}{k+1} \sum_{i=1}^n \frac{TP_i}{TP_i + FP_i + FN_i} \quad (9)$$

由表1中 mIoU 对比可知,方法1在验证序列10上对 Person 取得了显著优于方法2和方法3的

效果,在测试序列 11~21 上对小目标分割效果提升并不明显,整体分割 mIoU(关注预测为正样本的部分)相对下降,而 SemanticKITTI 测评准确率达到 0.8512(考虑全部样本).在测试集对 Bicycle 和 Motorcyclist 的分割效果略优于方法 2 和方法 3,但对 Person 的分割精度并未提升.

表 1 mIoU 值对比(未使用 KNN 后处理)

类别	方法 1		方法 2	方法 3
	验证序列 10	序列 11~21	序列 11~21	序列 11~21
Car	0.748	0.833	0.854	0.864
Bicycle	0.074	0.280	0.262	0.245
Motorcycle	0.167	0.228	0.265	0.327
Truck	0.028	0.138	0.186	0.255
Other-vehicle	0.003	0.179	0.156	0.226
Person	0.487	0.276	0.318	0.362
Bicyclist	0.000	0.299	0.336	0.336
Motorcyclist	0.000	0.090	0.040	0.047
Road	0.789	0.892	0.914	0.918
Parking	0.459	0.538	0.570	0.648
Sidewalk	0.658	0.709	0.740	0.746
Other-ground	0.000	0.126	0.264	0.279
Building	0.726	0.769	0.819	0.841
Fence	0.555	0.436	0.523	0.550
Vegetation	0.742	0.771	0.776	0.783
Trunk	0.397	0.463	0.484	0.501
Terrain	0.496	0.634	0.636	0.640
Pole	0.393	0.289	0.360	0.389
Traffic-sign	0.433	0.415	0.500	0.522
mIoU _{AVG}	0.398	0.436	0.474	0.499

注:加粗表示本文方法优于原方法的语义分割结果.

原 RangeNet++ 的分割效果已是当时最佳方法,而一方面本文训练次数 20 epoch,远远低于另外 2 种方法的 150 epoch;另一方面,由于小目标本身占比很小,测试集 11~20 和前面训练集数据在小目标上也有差异,这种差异会因为占比小而对训练结果产生更大的影响,由此也会造成本文方法在预测集的改进不明显.即使这样,在其他类别分割效果与方法 2、方法 3 基本接近而稍有逊色的情况下,方法 1 得到了对使用此 2 种方法分割效果较差的 Bicycle 和 Motorcyclist 目标类别的提高,同时 Other-vehicle 的分割效果也优于方法 2;结合图 6 中下降趋势,表明 FL 损失函数确实能有效地改善困难稀疏子集均衡性,值减小到 10^{-3} 量级,表明模

型拟合效果较好.

4 结 语

本文提出了一种道路场景车载点云小目标精准分割的方法,在 RangeNet++ 的基础上,改用对不平衡困难类别样本识别更为有效的 FL 损失函数,并优化权重系数,且采用卷积层数更少的 Darknet21 骨架,对其参数和超参数进行改进.在对道路场景点云全语义分割的基础上,提高了重要但体量较小且不规则的小目标的识别和分割的效果,特别在计算条件有限的情况下,可以改进提高点云小目标识别;也证明了 FL 损失函数能有效地改善困难稀疏子集均衡性.由于数据集中小目标的形态多但数量较少,本文研究是通过改变权重以及损失函数选择来解决此问题,后续研究还可以考虑:(1)采用人工生成模拟小目标学习样本的方法,增加数据集中小目标的学习样本数量,以进一步完善网络对小目标的学习,增强语义分割效果;(2)针对出现的其他类别分割精度下降问题,可对数据进行预处理,先过滤出点云的前景点并生成掩模,以使网络能更着重于对小目标的训练但不影响其他类别;(3)在网络训练阶段过采样包含小目标的点云、调节网络结构增大感受野等,以此来平衡小目标与其他类别物体的识别度.此外,考虑点云拥有的辐射信息,后续还可尝试将其引入深度学习,以提高对复杂重要小目标检测与分割效果.

参考文献(References):

- [1] Zhao Ruibin, Pang Mingyong, Zhang Yanling, *et al.* Progressively extracting accurate building roofs from airborne LiDAR data[J]. Journal of Computer-Aided Design & Computer Graphics, 2017, 29(4): 624-631(in Chinese)
(赵瑞斌, 庞明勇, 张燕玲, 等. 机载 LiDAR 点云中精细建筑物顶面的渐进提取[J]. 计算机辅助设计与图形学学报, 2017, 29(4): 624-631)
- [2] Dong Zihao, Shao Xiuli. A multi-category edge perception method for semantic segmentation[J]. Journal of Computer-Aided Design & Computer Graphics, 2019, 31(7): 1075-1085(in Chinese)
(董子昊, 邵秀丽. 多类别的边缘感知方法在图像分割中的应用[J]. 计算机辅助设计与图形学学报, 2019, 31(7): 1075-1085)
- [3] Krizhevsky A, Sutskever I, Hinton G E, *et al.* ImageNet classification with deep convolutional neural networks[C]

- //Proceedings of the 25th International Conference on Neural Information Processing Systems. New York: ACM Press, 2012: 1097-1105
- [4] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[OL]. [2020-08-18]. <https://arxiv.org/pdf/1409.1556v6.pdf>
- [5] Szegedy C, Liu W, Jia Y Q, *et al.* Going deeper with convolutions[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2015: 1-9
- [6] Visin F, Kastner K, Cho K, *et al.* ReNet: a recurrent neural network based alternative to convolutional networks[OL]. [2020-08-18]. <https://arxiv.org/abs/1505.00393.pdf>
- [7] He K M, Zhang X Y, Ren S Q, *et al.* Deep residual learning for image recognition[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 770-778
- [8] Garcia-Garcia A, Orts-Escolano S, Oprea S, *et al.* A review on deep learning techniques applied to semantic segmentation[OL]. [2020-08-18]. <https://arxiv.org/abs/1704.06857v1.pdf>
- [9] Charles R Q, Su H, Mo K C, *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. Los Alamitos: IEEE Computer Society Press, 2017: 77-85
- [10] Qi C R, Yi L, Su H, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space[OL]. [2020-08-18]. <https://arxiv.org/abs/1706.02413>
- [11] 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] //Proceedings of the IEEE International Conference on Robotics and Automation. Los Alamitos: IEEE Computer Society Press, 2018: 1887-1893
- [12] Iandola F N, Han S, Moskewicz M W, *et al.* SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5 MB model size[OL]. [2020-08-18]. <https://arxiv.org/abs/1602.07360.pdf>
- [13] Ester M, Kriegel H P, Sander J, *et al.* A density-based algorithm for discovering clusters in large spatial databases with noise[C] //Proceedings of the 2nd International Conference on Knowledge Discovery and Data Mining. Palo Alto: AAAI Press, 1996: 226-231
- [14] Wu B C, Zhou X Y, Zhao S C, *et al.* SqueezeSegV2: improved model structure and unsupervised domain adaptation for road-object segmentation from a LiDAR point cloud[OL]. [2020-08-18]. <https://arxiv.org/abs/1809.08495.pdf>
- [15] Milioto A, Vizzo I, Behley J, *et al.* RangeNet++: fast and accurate LiDAR semantic segmentation[C] //Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems. Los Alamitos: IEEE Computer Society Press, 2019: 4213-4220
- [16] Redmon J, Farhadi A. YOLOv3: an incremental improvement[OL]. [2020-08-18]. <https://arxiv.org/abs/1804.02767.pdf>
- [17] Lin T Y, Goyal P, Girshick R, *et al.* Focal loss for dense object detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(2): 318-327
- [18] Huang Jipeng. Research and application of small object detection based on deep learning[D]. Nanjing: Nanjing University, 2019(in Chinese)
(黄继鹏. 基于深度学习的小目标检测研究与应用[D]. 南京: 南京大学, 2019)
- [19] Pan Xingjia, Zhang Xulong, Dong Weiming, *et al.* A survey of few-shot object detection[J]. Journal of Nanjing University of Information Science & Technology: Natural Science Edition. 2019, 11(6): 698-705(in Chinese)
(潘兴甲, 张旭龙, 董未名, 等. 小样本目标检测的研究现状[J]. 南京信息工程大学学报: 自然科学版, 2019, 11(6): 698-705)
- [20] Chen L C, Papandreou G, Kokkinos I, *et al.* DeepLab: semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 40(4): 834-848
- [21] Chen L C, Yang Y, Wang J, *et al.* Attention to scale: scale-aware semantic image segmentation[C] //Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2016: 3640-3649
- [22] Zhao H S, Shi J P, Qi X J, *et al.* Pyramid scene parsing network[C] //Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition. Los Alamitos: IEEE Computer Society Press, 2017: 6230-6239
- [23] Hu Tai. Research on small object semantic segmentation algorithm based on deep neural networks[D]. Nanjing: Nanjing Normal University, 2018(in Chinese)
(胡太. 基于深度神经网络的小目标语义分割算法研究[D]. 南京: 南京师范大学, 2018)
- [24] Behley J, Garbade M, Milioto A, *et al.* SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences[C] //Proceedings of the IEEE/CVF International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2019: 9296-9306