高技术通讯 2024 年 第 34 卷 第 10 期:1091-1097

doi:10.3772/j.issn.1002-0470.2024.10.008

基于编解码结构的激光雷达动态物体点云分割网络①

赵宇轩② 贾克斌③ 李志坚

(北京工业大学信息学部 北京 100124)

(北京工业大学计算智能与智能系统北京市重点实验室 北京 100124)

(先进信息网络北京实验室 北京 100124)

摘 要 随着科技的不断发展,自动驾驶技术越来越多地进入到人们的生活中。动态物体点云分割是其中十分关键的一项任务,它可以为地图构建、路径规划等任务提供前置帮助。本文提出一种基于编码器-解码器结构的激光雷达动态物体点云分割网络,使用自校正卷积替换上下文特征提取模块中的普通卷积,提升神经网络的特征学习能力;并在网络解码阶段加入通道注意力机制,提升网络对重要特征通道的关注学习程度,从而达成更好的分割效果。本文在 SemanticKITTI MOS 数据集上进行实验,实验结果表明,本文所提出的动态物体点云分割网络相比原有方法取得更优表现,交并比(IoU)达到 72.1%。

关键词 激光雷达;深度学习;点云分割;自动驾驶

近些年来,激光雷达由于其测量精确度高、扫描范围广、受周围环境影响小等特点[1],被越来越多地应用到人们的日常生活当中,激光雷达在自动驾驶、机器人自主定位、无人机导航等领域承担着重要的环境感知任务[2-3]。

准确的环境感知能力是自动化设备能够自主工作的重要前提,理想的运行环境均基于静态假设,然而在实际数据采集感知过程中,周边环境中动态物体的出现是不可避免的。若在运行过程中不能准确识别出动态物体,则会导致同步定位、路径规划、静态地图建立等细分任务的效果不理想,最终影响到自动化设备的运行,甚至威胁到人们的安全。可见,激光雷达动态物体点云分割是自动化设备实现自主导航必不可少的能力。

激光雷达点云分割主要分为基于点的方法、基于体素的方法与基于投影的方法[4],通过不同的点云处理方式对点云进行分割。基于点的分割方法直接作用于不规则点云,可以有效提取每个点的特征,但由于其计算效率较低,难以在大规模点云下使用;基于体素的分割方法将点云以几何块划分、从而分块进行处理,此方法可以有效减少计算量,但会导致更多信息损失,影响精度;基于投影的方法将点云从三维空间投影至二维平面,可以一定程度上兼顾计算效率与信息保留,但在二维平面转换至三维空间的重投影阶段,可能会出现边缘错误分类的情况。

本文针对更具挑战性的激光雷达动态物体点云分割技术,不同于激光雷达点云语义分割,它只专注于将点云划分为动态与静态两类,而不是汽车、行人、道路等多种分类,已成为点云分割领域新的研究热点。考虑到目前动态对象分割精确度仍有提高的空间,本文在现在工作的基础上设计了一种新的激光雷达动态物体点云分割网络,进一步提高动态物体点云分割效果。综上所述,本文研究成果概括如下:(1)提出了一种基于编解码结构的激光雷达动态物体点云分割网络,使用自校正卷积设计了新的上下文模块;同时引入通道注意力机制,提升神经网

① 北京市自然科学基金(4212001)资助项目。

② 男,1999 年生,硕士生;研究方向:激光雷达点云处理;E-mail: zyx99@emails.bjut.edu.cn。

③ 通信作者,E-mail: kebinj@bjut.edu.cn。

(收稿日期:2024-03-12)

络的特征学习能力。(2)本文所提出的网络在 SemanticKITTI MOS 数据集上与原有方法相比,取得了更优表现,交并比(intersection over unin,IoU)达到 72.1%。

1 相关工作

1.1 激光雷达点云语义分割

对于点云语义分割,目前主要采用深度学习的方法来完成,可以分为 3 种方式:基于点的方法、基于体素的方法与基于投影的方法。

基于点的方法直接处理点云中的每一个点,将每个点依次进行预测,输出结果。2016 年,Qi 等人[5]提出了 PointNet,PointNet 使用共享多层感知机(shared multilayer perceptron,Shared MLP)来提取点云中每个点的特征,再经过池化得到全局特征,从而完成分割任务。PointNet 方法通过提取全局特征完成了分割,但对于点云局部信息提取不足。为解决此问题,Qi 等人[6]继续提出 PointNet ++,此方法通过将点云进行分层分组,分别进行特征提取,从而解决了局部信息提取不足的问题,达成更好的分割效果。此类基于点的分割方法能够在较小尺寸的点云上取得较好效果,然而当输入较大场景点云时,其运行效率则会有较大下降。

基于体素的方法将输入点云分成多个几何块,例如体素、圆柱体等,这样的方法能够一定程度上解决点云的无序性和非结构化问题。例如 SEGCloud[7],将点云划分为多个体素块后,进行特征提取。但基于体素的方法分割性能与所选体素大小直接相关,由于不同的体素分辨率设置大小,可能会出现丢失局部信息与计算资源浪费等情况。

基于投影的方法将三维点云投影到二维像素平面上,后对二维图像进行特征提取,将三维点云分割任务转换为图像分割任务,从而完成点云分割。2019 年,Milioto 等人[8]提出了 RangeNet ++,将三维点云转换为深度图后,对点云进行处理,后将深度图重投影至三维空间,完成点云分割。此方法相比于另外两种方法,能相对兼顾计算效率与分割精度。

1.2 激光雷达动态物体点云分割

激光雷达动态物体点云分割属于激光雷达语义分割的更细分领域,此任务并不将点云分割为多种物体,而是只区分点云动态与否。目前常用的分割方法分为传统方法与深度学习方法。

在非学习的传统方法中, Schauer 等人[9]提出 peopleremover,此方法基于体素射线法,根据栅格地图中栅格是否被击穿来判断动态点云,完成动态点云分割。Kim 等人[10]提出通过多分辨率深度图像,将雷达帧与地图逐帧对比,并逐步调节分辨率进行移除、恢复操作,从而达成分割动态点云的效果。文献[11]通过对比激光雷达扫描帧与相应地图,计算了扇区占用率,并根据阈值在疑似动态区拟合出地平面,从而分割出动态点云。传统方法耗费时间较多,且通常都需要预先构建地图。

随着人工智能技术的不断发展,深度学习方法已经广泛应用于各个领域,同样也应用于激光雷达动态物体点云分割。Ruchti 等人[12]提出,通过神经网络的学习,预测每个点处于动态的概率,将其从点云中移除,达成分割动态点的目的。Pfreundschuh 等人[13]提出,通过生成的动态点云标注,训练已有网络,分割其中动态点云。Chen 等人[14]基于已有的投影法激光雷达语义分割网络,开创性地提出结合激光雷达扫描帧与残差图像进行激光雷达动态物体点云分割,结合点云基本属性与时序信息,生成点云深度图与残差图,一同作为输入完成激光雷达动态点云分割,取得良好效果。

2 基于编解码结构的动态点云分割网络

本文提出的激光雷达动态点云分割网络结构如图 1 所示,下面将依次介绍点云到深度图像的投影转换、残差图像的生成、编解码结构网络、网络结构的改进以及从深度图像恢复到激光雷达点云的重投影过程与原理。

2.1 深度图生成

激光雷达点云是一种在三维空间的非结构化离散数据,直接对其进行处理需要消耗大量的时间,本文使用投影法,将每个三维空间坐标系中的激光雷达点 $p(x, y, z)$ 转换至二维图像坐标系中的点 $i(u, v)$,从而避免复杂计算。三维点云到深度图像的转

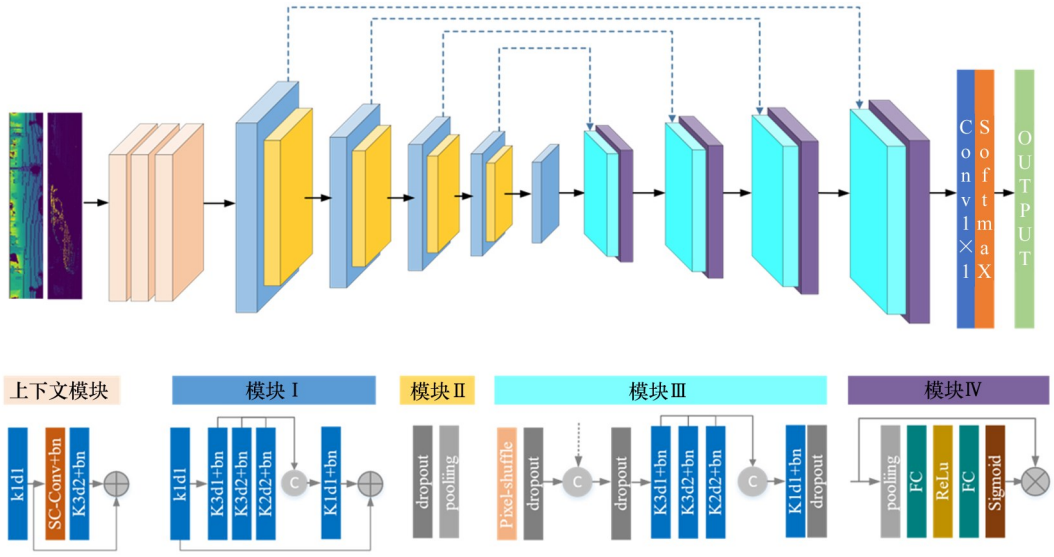


图 1　网络结构图

化如式(1)所示。

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

其中,$f = f_{up} + f_{down}$,f 是激光雷达传感器的视场大小,f_{up} 与 f_{down} 分别是激光雷达垂直视场角的上、下限值;(u,v) 是深度图像的二维坐标;$r = \|p\|^2$ 代表每个点的深度;h 与 w 代表深度图像的高度与宽度。在经过此公式转换后,可通过二维图像坐标索引获取点的三维坐标等信息。所生成的深度图像如图 2 所示,由此方法将三维点云转换至二维平面,通过深度区分出物体的几何轮廓,将三维分割任务转换为二维分割任务。

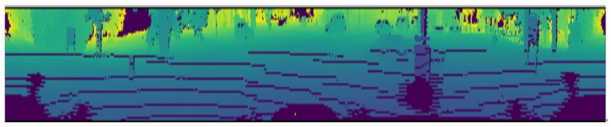


图 2　点云深度图

2.2　残差图生成

为了更加准确地分割出动态物体点云,本文生成残差深度图,从而结合时空信息进行特征学习。本文使用文献[14]的方法生成当前帧 l 与先前帧 k 间的残差图。将先前扫描帧转换至当前坐标下,通过式(1)将点云转为深度图像,后通过式(2)生成残差图像,将深度图与残差图一同作为网络输入。

$$d_{k,i}^{l} = \frac{|r_i - r_i^{k\to l}|}{r_i} \tag{2}$$

其中,r_i 是当前坐标 (u_i,v_i) 对应点的深度值,$r_i^{k\to l}$ 是转换后帧在此相同像素的深度值。残差图生成的结果如图 3 所示,生成的残差图像作为辅助信息,能够引导神经网络更好地区分动态与非动态物体点云。

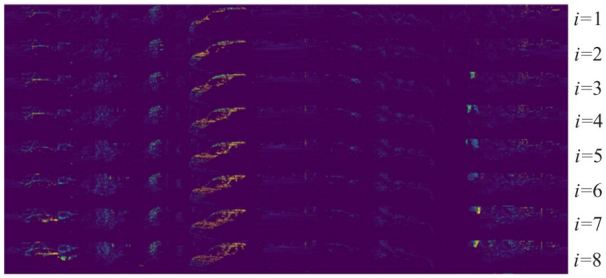


图 3　点云残差图

2.3　结合自校正卷积的上下文模块

上下文关系指目标像素与其周围邻域信息的关系,在语义分割任务中,上下文关系起着至关重要的作用,影响着语义分割任务表现的好与坏。受到文献[15]的启发,本文使用自校正卷积替换普通卷积,对网络中的上下文模块进行改进。自校正卷积在不改变主体网络结构的情况下,帮助神经网络生成更具区分性的特征,提升网络在分割任务中的表现。自校正卷积的结构如图 4 所示。

自校正卷积分为上下2个分支,首先输入经过

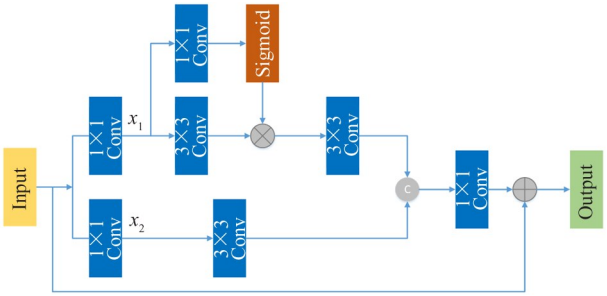


图 4　自校正卷积

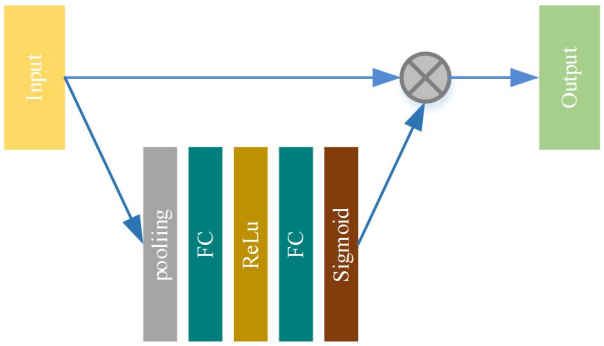


图 5　SE 模块结构

两个 1×1 卷积层进入上下两分支，得到 x_1 与 x_2，x_1 与 x_2 的特征通道数均为输入 *Input* 通道数的一半。上半分支包含 2 个 3×3 卷积层，其中在第 1 个卷积层配备了 1×1 卷积层与 Sigmoid 层，下半分支包含一个 3×3 卷积层进行特征提取，上下分支进行拼接后，再经过 1×1 卷积层并与跳跃连接相加，最终得到与输入特征维度相同的输出 *Output*。

2.4　编码器-解码器网络结构

本文使用编码器-解码器结构进行进一步特征提取。其结构由上图 1 中的模块Ⅰ至模块Ⅳ组成。其中，在编码阶段的第 1 层至第 4 层，网络通过编码器对输入特征图进行下采样并增加通道数，从而达成提升特征维度进行特征提取的目的；第 5 层编码器不进行下采样，保留原本维度。在解码阶段，4 层解码器通过对特征图进行上采样，逐步恢复图像的分辨率尺寸，完成输出。但在编码器的下采样过程中，由于池化操作将导致一些信息丢失，而丢失的信息并不能通过解码器的上采样过程恢复，因此在编码器进行下采样操作前，使用跳跃连接对较高分辨率特征图的信息进行保留，待解码器阶段与上采样后的特征进行融合，丰富特征细节信息。

2.5　结合通道注意力机制的解码器结构改进

在编解码器网络结构中有众多特征通道，普通神经网络往往不注重特征图不同通道权重的重要性，导致任务效果不佳。为了提升网络对重要特征通道的关注程度，本文在解码器部分添加了通道注意力机制 SE（squeeze and excitation）模块[16]，SE 模块的结构如图 5 所示。

SE 模块使得神经网络获取到特征图的每个通道的重要程度，从而让神经网络重点关注对当前任务更有效的特征通道，最终提升动态物体点云分割效果。首先将输入依次通过池化层（pooling）、全连接层（full connection，FC）、ReLu 层、全连接层与 Sigmoid 层，得到每个特征通道的权重值，最终与输入相乘，得到经过通道权重加权后的输出特征。

2.6　点云重投影与 KNN 后处理

基于投影的点云分割方法，需要将预测结果从二维平面重投影至三维空间。由于在点云和深度图像转换过程中，多个激光雷达点云可能会分配给同一相同像素，这会导致分割边缘的错误分类情况发生（在二维平面分割完成将其映射至三维空间时，可能会有边缘分割错误分类的情况发生）。本文使用文献[8]中提出的 K 近邻算法（k-nearest neighbor，KNN）聚类方法，划定搜索窗口并计算最中心点云周围标签的出现次数，随后选择出现次数最多的标签结果作为中心点云的预测标签结果，优化边缘错误分类情况。

3　实验结果与分析

3.1　数据集与评价标准

本文使用 SemanticKITTI MOS 数据集进行实验，SemanticKITTI MOS 是一个建立在 SemanticKITTI[17] 上，将点云分为动态点云与静态点云的三维激光雷达点云数据集。数据集包含 22 个激光雷达点云序列，其中，00～07 和 09～10 序列共 19 130 帧用于训练，序列 08 共 4 071 帧用于验证，序列 11～21 共 20 351 帧用于测试。KITTI-road 作为 SemanticKITTI MOS 数据集的补充，继续标记了序列 30～41 共 12 个序列，包含更多的动态物体点云。其中序列 30～34 和序列 40 共 2 905 帧用于训练，序列 35～39

和序列 41 共 2 889 帧用于验证。

为了量化网络性能表现,本文使用 IoU 进行网络性能度量,如式(3)所示。

$$IoU = \frac{TP}{TP + FP + FN} \tag{3}$$

其中 TP、FP 与 FN 分别代表动态点云的真阳性、假阳性和假阴性。

3.2 实验环境

本文在 Ubuntu 20.04 环境下,使用 NVIDIA RTX 3090 GPU 与 Pytorch 环境进行实验,本文按照文献[14]方法进行模型训练,设置初始学习率为 0.01,优化器动量为0.9,权重衰减系数设置为 0.000 1,训练模型 150 轮。

3.3 消融实验

为讨论本文所提出模块的改进效果,本文分别在输入残差图像数量为 1 和 8 的情况下,使用 SemanticKITTI MOS 数据集进行实验,分别对每种方法训练 150 轮,实验在验证集(序列 08)上进行,结果如表 1 与 2 所示。

表 1 输入残差图数量为 1 时结果

方法	*IoU*/%
LMNet	60.5
Our-a	61.8
Our-b	62.8
Our	63.7

表 2 输入残差图数量为 8 时结果

方法	*IoU*/%
LMNet	63.5
Our-a	65.1
Our-b	66.5
Our	67.7

其中,Our-a 为本文使用自校正卷积改进的网络,Our-b 为本文使用通道注意力改进的网络,Our 为同时使用自校正卷积和通道注意力机制的网络。由上表可见,在残差图像输入数量为 1 的情况下,将 4 种方法训练 150 轮,Our-a、Our-b、Our 方法相对于 LMNet(LiDAR-based multi-class abject detection network)分别提升了 1.3%、2.3%与 3.2%。在残差图像输入数量为 8 的情况下,同样将 4 种方法训练 150 轮,Our-a、Our-b、Our 方法相对于 LMNet[14] 分别提升了 1.6%、3.0%与 4.2%。可见,本文提出的方法在不同输入数量情况下对模型的学习能力均有提升。

由于 SemanticKITTI MOS 数据集中,动态点云比例较少,KITTI-road 选取一些包含动态点较多的序列对数据集进行补充,本文同时探讨网络在 KITTI-road 补充数据集上的性能表现,在输入残差图像数量为 8 的情况下进行实验,实验在验证集(序列 08、序列 35 ~ 39、序列 41)的结果如表 3 所示。

表 3 结合补充数据集时结果

方法	*IoU*/%
LMNet (结合 KITTI-road)	77.0
Our(结合 KITTI-road)	86.1

由表 3 可知,本文提出的方法在结合补充数据集时,相比 LMNet 效果同样提升。

3.4 动态物体点云分割性能对比

将本文输入残差数量为 8 实验模型的测试集(序列 11 ~ 21)推断结果上传至 SemanticKITTI MOS 数据集评估服务器,得到动态点云分割 IoU,并将结果与已有方法进行对比,对比结果如表 4。

表 4 分割性能对比结果

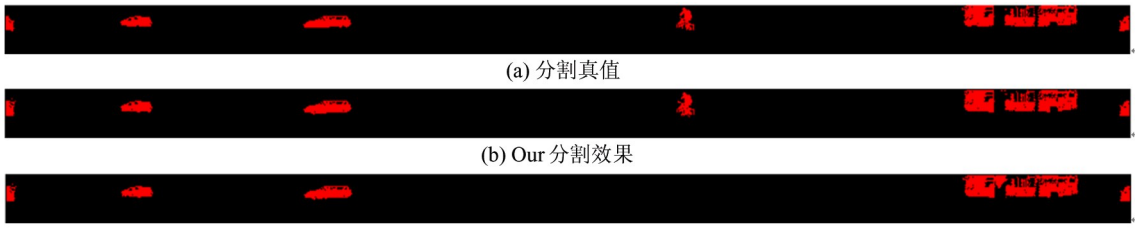
方法	*IoU*/%
Flownet3d	28.7
SqSequence	43.2
SalsaNext	46.6
KPConv	60.9
LMNet	57.7
Our	62.0
LMNet (结合 KITTI-road)	63.7
Our(结合 KITTI-road)	72.1

Flownet3d[18]、SqSequence[19]、SalsaNext[20] 在测试集的动态点云分割 IoU 分别为 28.7%、43.2%、46.6%。KPConv[21] 可达到 60.9%,但 KPConv 使用

基于点的分割方法,计算量大,推理时间无法满足实时需求。结合残差图输入的 LMNet 其 *IoU* 可达到 57.7%,而本文提出同样结合残差图输入的方法可达到62.0%,动态点云分割性能均优于上述网络。在使用补充数据集 KITTI-road 后,LMNet 达到 63.7%,本文提出的方法 *IoU* 最高可达 72.1%,分割性能最佳。

3.5　分割效果与推断时间对比

使用残差图输入的 LMNet 与本文提出方法达成了更佳性能,其分割效果如图 6(a)至图 6(c)所示,分别为分割真值、本文提出方法(Our)分割效果与 LMNet 分割效果。由图 6 可见,本文所提出方法与 LMNet 均可以在多数情况下取得较好分割效果,但对于图片中间部分,相对较小的骑行人员动态点云,本文提出的方法取得了更好的分割效果,优于 LMNet。



(a) 分割真值

(b) Our 分割效果

(c) LMNet 分割效果

图 6　分割效果对比

同时本文使用 NVIDIA RTX 3090 GPU 进行推断时间测试,实验结果如表 5 所示。

表 5　推断时间对比结果

方法	推断时间/ms	参数量/ $\times 10^6$
LMNet	13.88	6.71
Our	14.67	6.72

由表中数据可知,LMNet 平均推断时间为 13.88 ms,参数量为 6.71×10^6,本文提出的方法平均推断时间为 14.67 ms,参数量为 6.72×10^6。两种方法推断时间均可满足小于激光雷达单帧时间间隔的条件。不难看出,本文提出的方法在参数量与推断时间增长较少的情况下,取得更优的动态物体点云分割效果。

4　结　论

本文提出了一种基于深度图投影法的编码器解码器结构激光雷达动态物体点云分割网络,在上下文模块与解码器阶段对网络进行了改进,在点云重投影阶段结合 KNN 算法对点云进行后处理。通过对比实验证明,本文所提出的神经网络在相比原有方法增加少量参数与运算时间的情况下,有效提升了激光雷达动态物体点云分割任务的精度。目前方法仅通过激光雷达点云数据进行处理,后续计划继续结合图像信息,达成更好的动态物体点云分割效果。

参考文献

[1] 刘铭哲,徐光辉,唐堂,等. 激光雷达 SLAM 算法综述[J]. 计算机工程与应用,2024,60(1):1-14.

[2] 许智理,闫倬豪,李星星,等. 面向智能驾驶的高精度多源融合定位综述[J]. 导航定位与授时,2023,10(3):1-20.

[3] 王世强,孟召宗,高楠,等. 激光雷达与相机融合标定技术研究进展[J]. 红外与激光工程,2023,52(8):119-132.

[4] 杨雪磊. 三维点云语义分割研究综述[J]. 长江信息通信,2022,35(10):49-51.

[5] QI C, 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. Honolulu, USA:IEEE, 2017:77-85.

[6] QI C, YI L, SU H, et al. PointNet++: deep hierarchical feature learning on point sets in a metric space[C]//Advances in Neural Information Processing Systems. Long Beach, USA:MIT Press, 2017:5099-5108.

[7] TCHAPMI L, CHOY C, ARMENI I, et al. SEGCloud: semantic segmentation of 3D point clouds[C]//2017 International Conference on 3D Vision. Qingdao, China: IEEE, 2017:537-547.

[8] 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. Macao, China:IEEE, 2019:4213-4220.

[9] SCHAUER J, Nü CHTER A. The peopleremover——removing moving objects from 3D point cloud data by traversing a voxel occupancy grid[J]. IEEE Robotics and Automation Letters, 2018, 3(3): 1679-1686.

[10] KIM G, KIM A. Remove, then revert: static point cloud map construction using multiresolution range images[C]//2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. Las Vegas, USA: IEEE, 2020: 10758-10765.

[11] LIM H, HWANG S, MYUNG H. ERASOR: egocentric ratio of pseudo occupancy-based dynamic object removal for static 3D point cloud map building[J]. IEEE Robotics and Automation Letters, 2021,6(2):2272-2279.

[12] RUCHTI P, BURGARD W. Mapping with dynamic-object probabilities calculated from single 3D range scans[C]//2018 IEEE International Conference on Robotics and Automation. Brisbane, Australia:IEEE, 2018:6331-6336.

[13] PFREUNDSCHUH P, HENDRIKX H, REIJGWART V, et al. Dynamic object aware LiDAR SLAM based on automatic generation of training data[C]//2021 IEEE International Conference on Robotics and Automation. Xi'an, China: IEEE, 2021:11641-11647.

[14] CHEN X, LI S, MERSCH B, et al. Moving object segmentation in 3D LiDAR data: a learning-based approach exploiting sequential data[J]. IEEE Robotics and Automation Letters, 2021, 6(4): 6529-6536.

[15] ZHAO H, KONG X, HE J, et al. Efficient image super-resolution using pixel attention[C]//2020 European Conference on Computer Vision. Glasgow, UK: Springer, 2020:56-72.

[16] HU J, SHEN L, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020, 42(8):2011-2023.

[17] BEHLEY J, GARBADE M, MILIOTO A, et al. SemanticKITTI: a dataset for semantic scene understanding of LiDAR sequences[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea: IEEE, 2019:9296-9306.

[18] LIU X, QI C, GUIBAS L. FlowNet3D: learning scene flow in 3D point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Los Angeles, USA:IEEE, 2019:529-537.

[19] SHI H, LIN G, WANG H, et al. SpSequenceNet: semantic segmentation network on 4D point clouds[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, USA:IEEE, 2020:4573-4582.

[20] CORTINHAL T, TZELEPIS G, AKSOY E. SalsaNext: fast, uncertainty-aware semantic segmentation of LiDAR point clouds for autonomous driving[C]//International Symposium on Visual Computing. San Diego, USA: Springer, 2020:207-222.

[21] THOMAS H, QI C, DESCHAUD J, et al. KPConv: flexible and deformable convolution for point clouds[C]//2019 IEEE/CVF International Conference on Computer Vision. Seoul, Korea:IEEE, 2019:6410-6419.

Dynamic object point cloud segmentation network for LiDAR based on encoder-decoder structure

ZHAO Yuxuan, JIA Kebing, LI Zhijian

(Faculty of Information Technology, Beijing University of Technology, Beijing 100124)

(Beijing Key Laboratory of Computational Intelligence and Intelligent System, Beijing 100124)

(Beijing Laboratory of Advanced Information Networks, Beijing 100124)

Abstract

With the continuous development of science and technology, automatic driving technology has increasingly entered people's lives. Dynamic object point cloud segmentation is one of the key tasks, which can provide pre-help for map construction, path planning and other tasks. In this paper, a LiDAR dynamic object point cloud segmentation network based on encoder-decoder structure is proposed, which uses self-correcting convolution to replace common convolution in context feature extraction module to improve the feature learning ability of neural network. The channel attention mechanism is added in the decoding stage of the network to improve the learning degree of the network's attention to important feature channels, so as to achieve better segmentation effect. In this paper, experiments are carried out on the SemanticKITTI MOS dataset, and the experimental results show that the proposed dynamic object point cloud segmentation network achieves better performance than the original method, with an intersection over union (IoU) of 72.1%.

Key words: LiDAR, deep learning, point cloud segmentation, autonomous driving