

文章编号: 2096-1618(2023)05-0535-08

基于语义信息的三维点云全景分割方法研究

任不凡, 黄小燕, 吴思东, 蒋涛, 袁建英

(成都信息工程大学自动化学院, 四川 成都 610225)

摘要:针对端到端点云全景分割网络精度不足的问题,设计一种基于点云语义信息的全景分割算法。首先利用语义分割模型获取点云数据语义信息,然后结合点云语义和空间信息,对前景目标(车、人等)进行聚类。具体地,为避免同类相邻目标被聚类为一个目标,提出融合法向量夹角特征、空间位置、语义信息的聚类算法进行准确的前景实例分割。最后,提出一种新的类别划分方法,在不影响后续决策处理情况下,显著增加分割质量。SemanticKITTI数据集上的实验结果表明,提出的方法在全景质量、分割质量、识别质量、平均交并比4个指标上取得了较好的效果,分别达到56.6%、82.3%、68.2%、68.1%,并保持较快的速度(175 ms),充分证明其有效性和实用性。

关键词:全景分割;深度学习;三维点云;聚类算法

中图分类号:TP391

文献标志码:A

doi:10.16836/j.cnki.jcuit.2023.05.007

0 引言

在智能机器人^[1]、自动驾驶^[2]等领域,利用三维激光点云数据进行目标检测、语义分割等工作发展迅速^[3],许多应用场景都需要三维激光点云处理,并进行环境感知和准确解析。目标检测需要识别周围环境中相关的目标物体(如汽车、行人),并对检测到的物体实时快速反应;语义分割即对三维点云中的每个点赋予其实际意义的标签,是环境解析中一项非常重要的任务。但目前利用三维激光点云数据对场景内的环境解析还不够精细,如语义分割不能对同一类型不同的个体进行区分,而这些解析对于整个智能系统是否能做出稳健和安全决策至关重要。如何利用点云准确对场景进行全方位解析,获取精细化的信息,受到越来越多学者的关注。

目前,针对三维点云的大多研究仍集中在对分割的单个类别的推理或单个对象的识别,无法满足应用场景中对整体环境进行解析的实际需求。全景分割是图像领域^[4]引入的问题,它将场景分为不可数目标(stuff)和可数目标(things)两种类别,并生成全局的、统一的分割场景。其中,对如天空、道路、草地等目标分为背景stuff类别,进行语义分割,分配语义标签;对如汽车、自行车、行人等目标分为前景things类别,进行实例分割并分配实例标签和ID,以区分出他们相同

类别中的不同个体。全景分割任务以及全新评价指标的提出,使该项任务得到了广泛的应用与关注。因此,利用三维点云数据信息,进行更加精细的全景分割方法研究成为一个热门的场景解析问题。

随着深度学习技术的发展,涌现出许多用神经网络学习算法来解决全景分割问题。现有的全景分割方法主要分为将点云投影至二维图像的方法、三维目标检测与语义分割相结合的方法和一站式端到端的全景分割深度学习网络算法等。Zhou等^[5]提出的Panoptic-PolarNet网络通过将点云投影至二维图像,利用鸟瞰图的方法以PolarNet为骨干网络设计再通过投票机制来融合生成的语义信息和实例信息,得到最终的全景分割。Gasperini等^[6]提出的Panoster网络通过添加一个改进学习聚类方法的实例分支解析对象来得到实例信息,为语义分割网络带来全景能力。Hong等^[7]提出的DS-Net网络采用专门为三维点云设计的柱面卷积,通过可学习的聚类模块动态地处理非均匀的点云的特征,来融合语义和实例两个分支模块。此外,将语义分割与三维目标检测相结合进行全景分割方法也是解决方案的一个途径。如语义分割网络KPConv^[8]与三维目标检测网络PointPillars^[9]或者PV-RCNN^[10]相结合,语义分割网络RangeNet^[11]与三维目标检测网络PointPillars相结合,在三维点云全景分割中被证明是一种十分有效的解决方案。

以上方法大多在实现前景类分割时存在解析精度不够、解析运行所耗费的系统资源较高和运算时间较长等问题。且端到端的分割网络,若网络结构某一部分不能准确分割,不仅会导致整个网络的性能不良,而且会影响全景分割整体精度的平衡。为提升全景分割

收稿日期:2022-09-22

基金项目:国家自然科学基金资助项目(62103064);四川省科技厅资助项目(2021YFG0295, 2021YFG0133, 2021YFN0104, 2021YFH0069, 2022YFN0020, 2022YFS0565);四川省无人系统智能感知控制技术工程实验室开放课题资助项目(WRXT2020-001, WRXT2020-002, WRXT2020-005)

通信作者:吴思东. E-mail:wsd@cuit.edu.cn

的性能,提高推理速度,本文提出一种基于语义信息的全景分割方法,以更快速精准地进行场景解析与感知。该方法利用语义分割网络获取整个场景的语义信息,再对 things 类别语义信息进行整合,结合语义、空间及法向量信息,利用轻量快速的聚类算法对实例进行分割,最后结合 stuff 类别语义信息获取整个场景的全景分割结果。

本文的主要贡献:

(1)设计了一种基于语义信息的全景分割方法。利用语义分割网络得到语义信息,结合语义信息和空间信息对前景 things 类别进行实例分割,从而实现全景分割。该方法将全景分割进行解耦,减轻了网络模型压力,获得了很好的分割效果。

(2)设计了一种结合语义信息和空间信息的聚类分割方法。对相邻目标提出以法向量夹角的判断方法,准确分割出不同目标,进一步提升全景分割的效果。

(3)在大规模交通场景数据集 Semantic KITTI 上进行验证,与已有的全景分割网络 Panoptic RangeNet、Panoster、Panoptic-PolarNet、DS-Net 对比,实验结果表明本文的全景分割方法性能达到了优异的水平:PQ 为 56.5%、RQ 为 67.9%、SQ 为 82.3%、mIoU 为 68.2%。而且,对前景 things 类别的实例聚类分割部分只额外增加 24 ms 的处理时间。

1 方法

1.1 整体网络模型框架

全景分割网络框架由 3 个部分组成:利用点云语义分割模型对点云进行语义分割获取点云语义信息;根据语义信息划分前景类别和背景类别,对前景类别利用聚类算法进行实例分割;将背景类的语义信息与前景类的实例分割结果结合,实现整个应用场景的全景分割解析。完整的全景分割网络结构如图 1 所示。

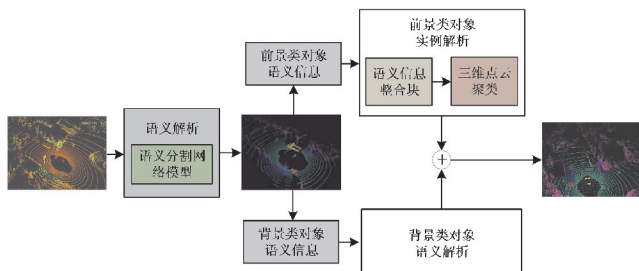


图1 全景分割整体网络结构框架

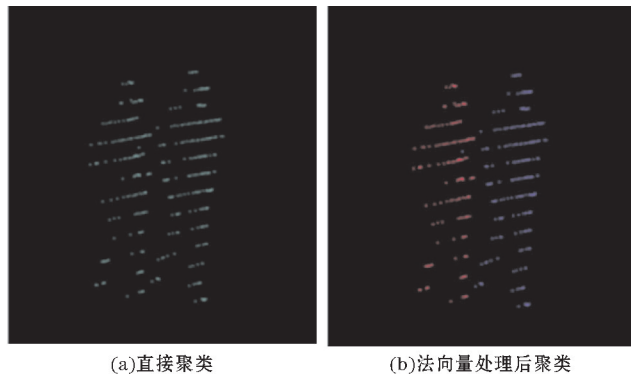
1.2 语义解析

直接利用现有的语义分割网络对点云进行处理获

取语义信息,换句话说,任意语义分割网络都能够适用于本文的全景框架。在此,选择目前最先进的语义分割网络 Cylinder3D^[12] 获取每个点的语义信息。并根据全景分割的定义,将语义信息按照前景 things 类别和背景 stuff 类别将点云划分。

1.3 前景类对象实例解析

针对前景类对象,全景分割需进一步对每个目标进行准确分割。因此,将语义分割模型结果与点云空间信息进行结合,利用聚类算法对同类别目标点进行聚合,从而实现单个个体的分割。直接利用语义信息可将不同类别信息进行初始划分,随后利用空间距离对点云聚类实现目标个体的分割。但如果存在多个个体距离很近的情况,基于空间距离的聚类方法可能会失效,如图 2(a)所示,当两个行人靠近时,基于空间距离的聚类方法不能准确分割单个个体。由于同一物体表面法向量变化缓慢,相邻个体通常存在明显间隔,边界区域法向量存在明显变化,因此本文提出一种基于法线夹角的分割策略。该策略利用法线夹角检测相邻目标边缘并保存,将边缘点剔除加大相邻目标间隔,从而分开相邻目标。



(a)直接聚类 (b)法向量处理后聚类

图2 多个行人距离近时分割效果

基于法线夹角的边缘点检测流程(图 3)如下:

步骤 1 采用体素下采样以减少后续检测流程的处理时间。

步骤 2 因为点云中单个点没有办法计算法向量,所以只能通过拟合平面估计局部邻域内的点云法向量,使用 k 近邻算法搜索距离目标点 s 最近的 k 个近邻点 p ,距离小于阈值即为近邻点,其搜索近邻点距离的核心公式:

$$d(s, p_i) = \sqrt{\sum_{t=1}^m (s_t - p_{it})^2}$$

式中, d 为近邻点 p_i 与目标点 s 之间的距离; s_t 为目标点 s 的第 t 个属性; p_{it} 为第 i 个近邻点的第 t 个属性; m 为属性总数,取值范围为 $[1, 3]$ 。

步骤3 利用目标点 s 连接近邻点 p_i, p_{i-1} 构建向量组成平面,求出最近的 k 个近邻点估计出的平均法向量 \mathbf{n} :

$$\mathbf{n} = \frac{1}{k} \sum_{i=1}^k |\overrightarrow{sp_{i-1}} \times \overrightarrow{sp_i}| (\overrightarrow{sp_0} = \overrightarrow{sp_k}) \quad (i=1, 2, \dots, k)$$

式中, \mathbf{n} 为平均法向量; $\overrightarrow{sp_{i-1}}$ 和 $\overrightarrow{sp_i}$ 为目标点与近邻点构建的向量。

步骤4 在这个局部邻域内,对需要判定的点,计算点法向量与平均法向量之间的法向量夹角 α ,判断 α 与夹角阈值 T 的大小关系,若 $\alpha < T$,即若小于一定阈值则判断为边缘点。

$$\alpha = \arccos \frac{\mathbf{n}_i \mathbf{n}}{|\mathbf{n}| |\mathbf{n}_i|}$$

式中, \mathbf{n}_i 为点法向量, \mathbf{n} 为平均法向量,它们之间的夹角为 α 。

步骤5 判断是否所有点均被检测。若未检测完,继续跳回步骤2执行;若检测完毕,则完成边缘点的判断,结束流程。

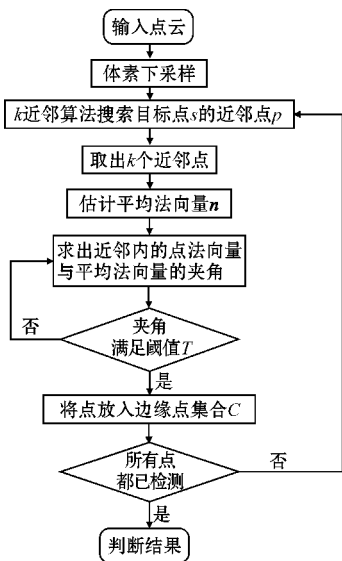


图3 基于法线夹角的边缘点检测流程

在获取边缘点之后,将边缘点从原数据中分离出来,并对提取后剩余的原始数据利用聚类算法进行点云聚类。随后,还需要将已分离出集合 C 里的边缘点还原,因此需将这些点根据 K 近邻来判断所靠近近邻点多数的标签作为该点的标签,从而完成实例分割。图2(b)为利用法线夹角分割策略后的聚类结果,可以看出相邻行人被完整地分割成两个目标。

聚类算法方面,选择利用 Scan-line Run (SLR) 聚类算法^[13]对前景类对象进行分割,获得实例分割结果。Scan-line Run (SLR) 聚类针对实际应用,以快速和低复杂性的方式来解决三维点云分割问题。该算法在提取地面后,通过分析三维点云的结构特点,利用点

云的结构对剩余的非地面点进行聚类。在激光雷达 360° 扫描的情况下,所扫描到的每一层被称为按层组织的一组点云。每个层中的点也以椭圆的方式组织,所有椭圆层的起点被认为共享相同的方向,并利用智能索引来执行有效的分割。激光点云数据的多线结构导致的分层,很类似于二维图像多行堆叠结构,激光发射器垂直方向排列,射线形式发出,使射线方向类似图像的列排列,不同之处在于每层元素分布不均匀以及激光点云数据独特的圆环结构。本方法将点视为像素,以二进制图像的形态学操作为基础,并采用二值图像的两次连通分量标记技术^[14]来完成实时点云聚类方案。SLR 算法具体流程(图4)如下:

步骤1 第一层点初始化。按顺序遍历点,点间距离小于一定值便判定为同一类,若存在距离或空元素突变便对接下来的点重新赋予初始类别,接收新标签(三角形中数值)。

步骤2 新标签的分配和两个标签的传播。8点最近的非地面相邻点是2,但它们的距离大于预先设置的距离阈值。在这种情况下,为第8点赋予新的簇类标签。接下来,最近的10的非地面邻居是3,其距离小于阈值,点10继承点3的簇类标签1。同理,点12和13都接近它们各自的相邻点5和6,获得与相邻点一样的簇类标签。

步骤3 中间点处理。其中点17和19有相邻点10和12,它们属于不同的簇并且都满足于继承标签的条件。本方案设定,此类情况下,按两个标签中的更小标签(即标签1)继承。

步骤4 簇间的合并。当存在步骤3中属于不同的簇且都满足继承标签的条件情况时,若两簇连通,则将两簇合并,赋予较小标签。

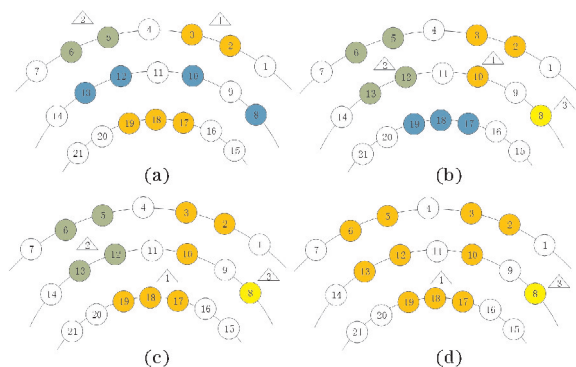


图4 SLR 聚类算法流程^[13]

如图4所示,白色点代表地面点(此时已完成地面点剔除,相当于空),而彩色点为留下来的非地面点。非地面点以不同颜色区分簇类,蓝色指还未进行分类划分的非地面点,三角形代表对当簇的标签划分。至此,完成分割任务。

此外,表 1 将 SLR 算法与欧几里得聚类法和深度聚类算法进行比较,以验证 SLR 聚类算法的性能。欧几里德聚类算法中,将距离阈值(Dth)设置为中等阈值0.5 m。因为大的距离阈值会将近距离的对象组合在一起,导致过分割;而小的距离阈值将只能检测空间距离间隔较大的目标,对于空间距离间隔较小的多个目标极易造成目标欠分割的现象,导致欠分割。对于

深度聚类算法,角度阈值 θ 的选择为 10° 。对于 SLR 聚类,两个距离阈值取值分别为0.5和1.0。在 3 种点云聚类方法中,SLR 聚类的分割效果最好。同时在耗时方面,表 1 评估了使用不同聚类方法在平均每一帧上的实例分割处理速度,处理速度最快的是欧几里得聚类算法,仅18.7 ms,而深度聚类算法和 SLR 聚类处理也不慢,分别为19.2 ms和29.4 ms。

表 1 点云聚类方法对比

聚类方法	参数设置	PQ	PQ+	RQ	SQ	mIoU	Time/ms
Euclidean Cluster	Dth=0.5 m	55.9	61.5	67.7	82.2	67.7	18.7
Depth Cluster	$\theta=10^{\circ}$	54.2	60.9	66.3	81.6	67.3	19.2
Scan-line Run	thrun, thmerge = 0.5, 1.0	56.6	62.9	68.2	82.3	68.1	29.4

1.4 语义信息整合模块

本文的全景分割是在语义分割结果的基础上进一步处理的结果,若语义分割结果出错,会导致聚类方法不可能获得准确的聚类效果。因此,为减少因语义分割结果错误导致的目标实例分割出错的问题,本文对 Cylinder3D 算法在 Semantic KITTI 数据集^[15]上的各类前景目标分类错误概率进行了统计,如表 2 所示。可以看到:骑摩托车的人[motorcyclist]错分到其他类别的概率达到99.61%。其中,进一步对骑摩托车的人、骑自行车的人、摩托车、自行车的具体误分的类别进行了统计,如表 3 所示。可以看出,骑摩托车的人[motorcyclist]误分到骑自行车的人[bicyclist]的比率占58.01%,骑自行车的人[bicyclist]误分到骑摩托车的人[motorcyclist]的比率占0.17%,而摩托车[motorcycle]和自行车[bicycle]相互间的误分率也很高。其原因正是在行驶过程中,由于激光雷达有着近密远疏的特性,在整个大场景下采集到的骑摩托车的人和骑自行车的人的点云数据太过稀疏,特征细节不够丰富。相比于相机,由于激光雷达采集到的点云数据稀疏、不具有颜色和纹理信息,并且骑自行车的人与骑摩托车的人在行为姿态特征上又太过相似,因此骑摩托车的人极其容易误分为骑自行车的人。

表 3 Semantic KITTI 数据集主要类别的误分率 单位:%

类别	bicyclist 误分率	motorcyclist 误分率	bicycle 误分率	motorcycle 误分率
car	0.1519	23.6652	0.3865	1.2123
bicycle	0.0089	0.0000	89.8123	0.7211
motorcycle	0.0000	0.0000	0.5784	90.5183
truck	0.0393	0.0000	0.0000	0.2106
other-vehicle	0.1486	5.3391	9.0242	4.6422
person	0.8306	12.4098	0.0428	0.0398
bicyclist	98.6466	58.0087	0.1468	0.0699
motorcyclist	0.1740	0.5772	0.0090	2.5857

有些类别间的差异很小,拥有相似的点云几何特征,但对于自动驾驶来说他们却是拥有相似决策逻辑的目标对象。如图 5 所示,在激光雷达扫描到的点云数据中自行车与摩托车拥有几乎相同的几何特征。

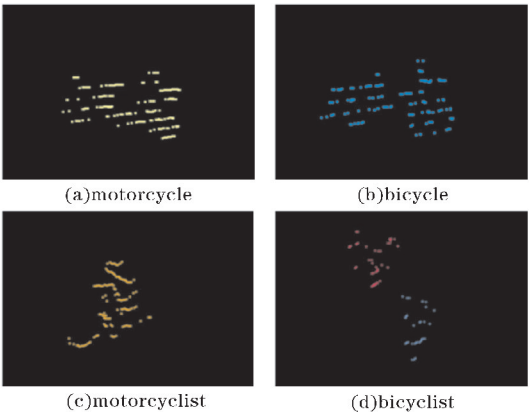


图 5 类别差异图

若是刻意去判断这些相似类别间的差异,将会在算力和性能上付出更多的成本代价。尽管目标对象有差异,但是在实际应用场景中^[16],都会在激光雷达中扫描到具体的点云信息。因为无论前方是一个自行车还是摩托车,自动驾驶任务会做出同样的决策结果。以实际情况中的这些问题上来说,对这两者的区分通过更加均衡的标签分布反而可以获得更加良好的分割

表 2 Semantic KITTI 数据集类别错分概率 单位:%

类别	错分概率
motorcyclist	99.607
bicycle	37.221
person	17.406
other-vehicle	15.485
motorcycle	15.285
truck	8.028
bicyclist	5.400
car	1.646

表现。若在这个阶段错分,则会导致实例分割精度低下。同样地,骑自行车的人和骑摩托车的人两个类别之间的相似特征差异很小,也会影响分割质量。

而对于自动驾驶汽车来说,骑自行车的人和骑摩托车的人、自行车和摩托车类别间的区分对后续自动驾驶汽车决策影响不大,但由于其语义分割误分率较大,使全景分割无法准确识别各个实例个体。因此,将数据集中的类别进行重划分操作,对标签类别微调整合,得到如表 4 所示的划分结果。

表 4 类别重划分

初始序号	初始类别	重划分后标签	重划分后类别
2	bicycle	2	bicycle
3	motorcycle	2	bicycle
7	bicyclist	7	bicyclist
8	motorcyclist	7	bicyclist

2 实验结果与分析

2.1 实验环境与测试集

本文实验的硬件平台为搭载了 Inter (R) Core (TM) i9-10850K CPU, GeForce RTX 3090 GPU 硬件配置的服务器,操作系统为 Ubuntu 18. 04 LTS, 软件架构具体的环境参数为 Python 3. 8, CUDA 11. 1. 0, Cudnn 8. 0. 4, PyTorch 1. 8. 0。

为体现本文方法的有效性,采用的是自动驾驶领域现有的大规模公开点云分割数据集 Semantic KITTI。数据集 Semantic KITTI 主要由郊区街道组成,交通密度相对较低,将激光雷达点云逐点分为了 28 个类别,忽略点数极少的小类别后保留 19 类用作点云分割。且用标签对其中每个点进行注释,并对 360°全场景进行数据采集和标注,显示了大量具有挑战性的交通情况和环境类型。

2.2 评价指标

本文以平均交并比 (mIoU)、全景分割质量 (PQ)、分割质量 (SQ) 和识别质量 (RQ) 作为实验的主要评价

指标,能更加综合地对全景分割进行性能评估。IoU 在数学中用于计算两个集合的交集在并集上的占比,在计算机视觉中用于评估分割精度,而 mIoU 是所有类别 i 的平均值,具体定义:

$$mIoU = \frac{1}{C} \sum_i \frac{TP_i}{TP_i + FP_i + FN_i}$$

其中 C 表示类别数, TP_i 、 FP_i 和 FN_i 分别表示对 i 类的预测中,预测值和真值一致的点集,真值不是 i 类却被分割为 i 类的点集,真值为 i 类却被预测为其他类的点集。

全景分割的评估指标是全景质量、分割质量和识别质量, PQ 可以解释为分割质量 (SQ) 和识别质量 (RQ) 的乘积,具体定义如下:

$$PQ = RQ \times SQ$$

$$RQ = \frac{|TP|}{|TP| + \frac{1}{2}|TN| + \frac{1}{2}|FP|}$$

$$SQ = \frac{\sum TP \cdot IoU}{|TP|}$$

2.3 对比实验及分析

将本文的全景分割方案与其他方案如语义与目标检测结合的方案、端到端的全景分割方案等进行比较,可以看出本文提出的方法在综合各类全景质量 (PQ)^[17]、分割质量 (SQ) 和识别质量 (RQ) 以及平均交并比 (mIoU) 评测指标下均有优异表现。可以看到,将语义分割与三维目标检测相结合进行全景分割方法也是解决方案的一个分支,但三维目标检测需要一个庞大且复杂的网络,并且有冗余的计算标签的成本。如表 5 所示,本文的全景分割方案性能优于语义分割与三维目标检测结合的解决方案,与 KPConv + PV-RCNN 等相比,同样分割精度明显提升。与基于点云投影至二维图像的方法 (如基于鸟瞰图的 Panoptic-PolarNet) 相比,本文方法在 mIoU 指标上有 8. 6% 的提升。与一站式端到端的全景分割网络模型 Panoster 和 DS-Net 相比,本文方法在 mIoU 指标上分别有 8. 2%、6. 5% 的提升,在各小类别上的对比在表 5 中有详细说明,大部分类别指标均有不同程度的提升。

表 5 全景分割方法对比

方法	类型	PQ	PQ+	RQ	SQ	PQth	RQth	SQth	PQst	RQst	SQst	mIoU
KPConv+PointPillars	sem+box	44. 5	52. 5	54. 4	80. 0	32. 7	38. 7	81. 5	53. 1	65. 9	79. 0	58. 8
RangeNet+PointPillars	sem+box	37. 1	45. 9	47. 0	75. 9	20. 2	25. 2	75. 2	49. 3	62. 8	76. 5	52. 4
KPConv+PV-RCNN	sem+box	50. 2	57. 5	61. 4	80. 0	43. 2	51. 4	80. 2	55. 9	68. 7	79. 9	62. 8
Panoptic RangeNet	panoptic	38. 0	47. 0	48. 2	76. 5	25. 6	31. 8	76. 8	47. 1	60. 1	76. 2	50. 9
Panoster	panoptic	52. 7	59. 9	64. 1	80. 7	49. 4	58. 5	83. 3	55. 1	68. 2	78. 8	59. 9
Panoptic-PolarNet	panoptic	54. 1	60. 7	66. 0	81. 4	53. 3	60. 6	87. 2	54. 8	68. 1	77. 2	59. 5
DS-Net	panoptic	55. 9	62. 5	66. 7	82. 3	55. 1	62. 8	87. 2	56. 5	69. 5	78. 7	61. 6
Ours	sem+cluster	56. 6	62. 9	68. 2	82. 3	51. 9	62. 0	84. 8	59. 1	71. 1	79. 6	68. 1

图6展示了本文所提出的方法在 SemanticKITTI 数据集上的分割结果,从左至右依次为原始点云、语义分割、实例分割的结果。可以看到,本文方法的语义分割部分将每一类目标分割开,对于重要交通参与者(行人、汽车、大型车辆等)即前景类对象,能够获得准确清晰目标边缘信息,并且对回归得到准确语义信息赋予不同的显示颜色对应不同的实例 ID 标签。图7展示了经过法向夹角边缘检测优化前后三维点云聚类的分割细节对比,图中靠得很近的行人能够准确地分割开来,并能够实现在道路交通场景下的逐点语义解析工作,对于可行驶区域也能获得可靠信息以供自动驾驶后续的规划使用,网络的综合性能良好。在实际场景里,真实道路场景解析的需求及环境相对简单,类别的划分不必如此精细。而类别划分需求的精简,相

似类的合并,可减少网络类别的识别,增大类间的差异,减轻整个分割框架的负担,提高网络对类别识别的综合性能,减少计算和内存压力。因此,本文也对经过标签类别重划分后的性能指标做了对比评估,如表6所示。最后,本文在表7还对比了与其他全景分割框架的耗时统计。本文的框架利用聚类方法,对硬件要求低,能够在耗费少量资源的条件下仅调用 CPU 即可快速准确地对语义信息进行处理。如表7所示,其他端到端的全景分割的方法,如 KPConv+PointPillars,处理每一帧点云的耗时高达500 ms,而本文方法在点云场景中能保持较快的推理速度,特别是在对前景类的实例分割速度仅耗时29.4 ms,却也可以达到较高精度,证明了本文方法计算的高效性。

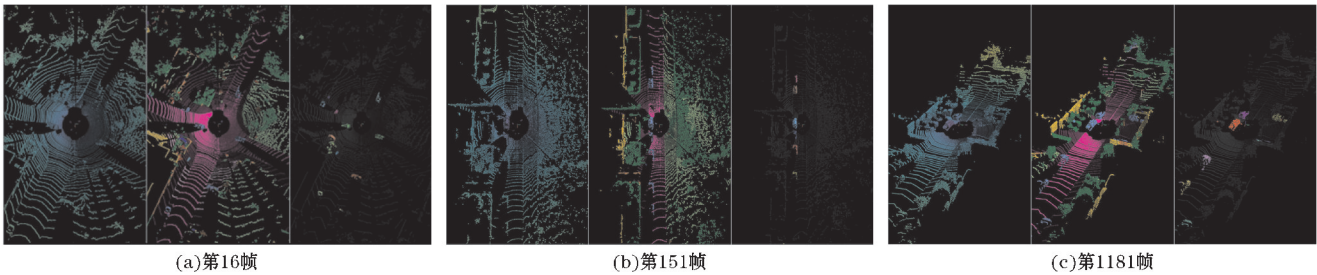


图6 数据集 Semantic KITTI 的分割效果图

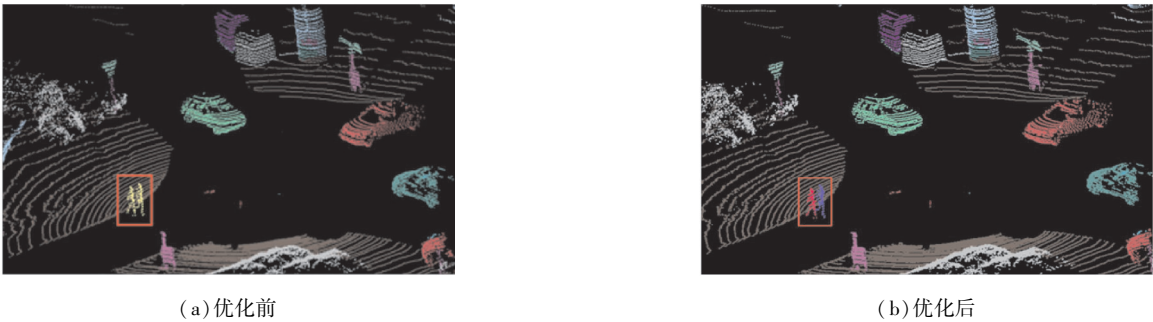


图7 点云聚类分割优化细节

表6 类别重划分后全景分割方法对比

方法	类型	PQ	PQ+	RQ	SQ	PQth	RQth	SQth	PQst	RQst	SQst	mIoU
KPConv+PointPillars	sem+box	44.5	52.5	54.4	80.0	32.7	38.7	81.5	53.1	65.9	79.0	58.8
DS-Net	panoptic	55.9	62.5	66.7	82.3	55.1	62.8	87.2	56.5	69.5	78.7	61.6
Ours	sem+cluster	56.6	62.9	68.2	82.3	51.9	62.0	84.8	59.1	71.1	79.6	68.1
Ours (类别重划分后)	sem+cluster	60.1	65.3	71.8	81.8	65.8	76.4	85.7	56.4	68.9	79.4	70.0

表7 全景分割耗时对比

方法	类型	PQ	PQ+	RQ	SQ	mIoU	Time/ms
KPConv+PointPillars	sem+box	44.5	52.5	54.4	80.0	58.8	537.3
DS-Net	panoptic	55.9	62.5	66.7	82.3	61.6	316.7
Ours(Cylinder3D+SLR)	sem+cluster	56.6	62.9	68.2	82.3	68.1	146+29.4

3 结束语

本研究有助于实现在人工智能领域中对复杂场景进行快速准确的理解。具体来说,提出了一个基于语义信息的三维激光点云全景分割方法。该方法在第一阶段利用现有的语义分割网络方法得到语义信息。在第二阶段使用轻量且快速的点云聚类算法将前面得到的前景类语义分割信息进行后续处理,并增强实例分割的推理速度。此外,这一方法在对类别重划分后大大减轻了种类分布不均造成的不利影响,且在实际道路场景中也能够高效准确地进行场景解析。这种方法也适用于其他激光雷达点云分割任务。

为了评估所提出的方法,本文选择了带有逐点语义和实例信息注释的 SemanticKITTI 数据集,基于不同的分割网络模型进行了算法实验。同时,也对比了几种点云聚类算法在全景分割中的性能表现。实验结果表明,本文提出的方法可以实现准确性和效率的有效平衡,其中 PQ 为56.5%、RQ 为67.9%、SQ 为82.3%、mIoU 为68.2%。在全景分割质量精度上,相较三维目标检测与语义分割相结合的方法,如 KPConv+PointPillars,提高了12.1%,相较全景分割网络模型(如 DS-Net)提高了0.7%。总而言之,本文研究是以激光雷达三维点云全景分割为目的对人工智能领域和自动驾驶场景感知与理解的积极探索。未来将专注于对复杂动态环境进行快速准确的解析。

致谢:感谢成都信息工程大学科研基金(KYTZ202109, KYTZ202142)对本文的资助

参考文献:

- [1] Zermas D, Izzat I, Papanikolopoulos N. Fast segmentation of 3d point clouds: A paradigm on lidar data for autonomous vehicle applications[C]. 2017 IEEE International Conference on Robotics and Automation(ICRA). IEEE, 2017: 5067–5073.
- [2] Shin M O, Oh G M, Kim S W, et al. Real-time and accurate segmentation of 3-D point clouds based on Gaussian process regression[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, 18(12): 3363–3377.
- [3] Yurtsever E, Lambert J, Carballo A, et al. A survey of autonomous driving: Common practices and emerging technologies[J]. IEEE access, 2020, 8: 58443–58469.
- [4] Kirillov A, Girshick R, He K, et al. Panoptic feature pyramid networks[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 6399–6408.
- [5] Zhou Z, Zhang Y, Foroosh H. Panoptic-polarnet: Proposal-free lidar point cloud panoptic segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13194–13203.
- [6] Gasperini S, Mahani M A N, Marcos-Ramiro A, et al. Panoster: End-to-end panoptic segmentation of lidar point clouds[J]. IEEE Robotics and Automation Letters, 2021, 6(2): 3216–3223.
- [7] Hong F, Zhou H, Zhu X, et al. Lidar-based panoptic segmentation via dynamic shifting network[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13090–13099.
- [8] 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.
- [9] Lang A H, Vora S, Caesar H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]. Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2019: 12697–12705.
- [10] Shi S, Guo C, Jiang L, et al. Pv-rcnn: Point-voxel feature set abstraction for 3d object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10529–10538.
- [11] Singh G, Gupta S, Lease M, et al. Range-Net: A High Precision Streaming SVD for Big Data Applications[J]. arXiv preprint arXiv:2010.14226, 2020.
- [12] Zhou H, Zhu X, Song X, et al. Cylinder3d: An effective 3d framework for driving-scene lidar semantic segmentation[J]. arXiv preprint arXiv:2008.01550, 2020.
- [13] Zermas D, Izzat I, Papanikolopoulos N. Fast segmentation of 3d point clouds: A paradigm on lidar data for autonomous vehicle applications[C]. 2017 IEEE International Conference on Robotics and

- Automation(ICRA). IEEE,2017:5067–5073.
- [14] He L,Chao Y,Suzuki K. A run-based two-scan labeling algorithm[J]. IEEE transactions on image processing,2008,17(5):749–756.
- [15] 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,2019:9297–9307.
- [16] Schumann O,Hahn M,Scheiner N,et al. Radar-Scenes:A real-world radar point cloud data set for automotive applications[C]. 2021 IEEE 24th International Conference on Information Fusion (FUSION). IEEE,2021:1–8.
- [17] Kirillov A,He K,Girshick R,et al. Panoptic segmentation[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition,2019:9404–9413.

Research on Panoptic Segmentation of 3D Point Clouds based on Semantic Information

REN Bufan, HUANG Xiaoyan, WU Sidong, JIANG Tao, YUAN Jianying
(College of Automation,Chengdu University of Information Technology,Chengdu 610225,China)

Abstract: The accuracy of end-to-end point cloud panoptic segmentation network is insufficient, this paper designs a panoptic segmentation algorithm based on point cloud semantic information to address this problem. This paper uses a semantic model to obtain the semantic information of point cloud, and then combined with spatial information to cluster the foreground objects (cars, people, etc.). Specifically, to avoid the adjacent objects of the same category from being clustered into one object, a clustering algorithm that integrates normal angle, spatial localization, and semantic information is proposed to accurately segment foreground instance. Finally, through further analysis and research on data categories, a new category classification method is proposed, which can significantly increase the segmentation quality without affecting the subsequent decision processing. The experimental results on the SemanticKITTI dataset show that the proposed method has achieved good results in four indicators: panoptic quality, segmentation quality, recognition quality, and average intersection ratio, respectively reaching 56.6%, 82.3%, 68.2%, 68.1%, and maintain a relatively fast speed (175 ms), which further indicates the effectiveness and practicability of this method.

Keywords: panoptic segmentation; deep learning; point cloud; clustering algorithm