

DOI:10.3969/j.issn.1003-5060.2026.07.001

基于改进的 PointPillars 小目标检测算法

张炳力^{1,2}, 杨程磊^{1,2}, 潘泽昊^{1,2}, 王恽昕^{1,2}, 王欣雨^{1,2}, 王焱辉^{1,2}

(1. 合肥工业大学 汽车与交通工程学院, 安徽 合肥 230009; 2. 安徽省智能汽车工程研究中心, 安徽 合肥 230009)

摘要:针对传统点云检测算法中小目标漏检、误检等问题,文章提出一种改进的 PointPillars 小目标检测算法。首先,在主干网络中引入特征金字塔模块,融合高层次语义信息与低层次定位信息来优化对小目标的检测;其次,加入残差模块和卷积块注意力模块(convolutional block attention module,CBAM),结合通道信息与空间信息来完成对深层特征的提取;最后,通过三检测头输出不同尺寸的特征图,以实现对不同类别目标的精确识别。在公开数据集 KITTI 上进行的可行性分析结果表明,相较于传统 PointPillars 算法,该文算法对小目标识别的平均精度均值提升了 1.74%。实车试验结果表明,该文算法对道路目标检测结果准确,检测频率达 65 Hz,满足实时性需求。

关键词:特征金字塔;残差模块;注意力机制;道路目标;智能驾驶

中图分类号:U471.15

文献标志码:A

文章编号:1003-5060(2026)07-0865-07

Small target detection algorithm based on improved PointPillars

ZHANG Bingli^{1,2}, YANG Chenglei^{1,2}, PAN Zehao^{1,2},

WANG Yixin^{1,2}, WANG Xinyu^{1,2}, WANG Yanhui^{1,2}

(1. School of Automobile and Traffic Engineering, Hefei University of Technology, Hefei 230009, China; 2. Anhui Provincial Research Center for Intelligent Automotive Engineering, Hefei 230009, China)

Abstract:In view of the missing and false detection of small and medium targets in the traditional point cloud detection algorithm, this paper proposes a small target detection algorithm based on improved PointPillars. Firstly, feature pyramid module is added into the backbone network to optimize the detection of small targets by combining high-level semantic information and low-level positioning information. Secondly, by adding the residual module and the convolutional block attention module (CBAM), the channel information and spatial information are combined to complete the extraction of deep features. Finally, three detection heads output feature maps of different sizes to achieve accurate recognition of different types of targets. The feasibility analysis of the proposed algorithm is carried out on KITTI data set. Compared with traditional PointPillars algorithm, the proposed algorithm improves the mean average precision(mAP) for small target recognition by 1.74%. The results of real vehicle test show that the algorithm is accurate for road target detection, and the detection frequency of 65 Hz meets the real-time requirement.

Key words:feature pyramid; residual module; attention mechanism; road target; intelligent driving

随着智能驾驶技术的日益发展,多模态感知应运而生。借助密集的激光点云信息,激光雷达

可实现更高的感知精度、更高的识别准确率以及更直接的数据获取。凭借强大的环境感知和处理

收稿日期:2024-03-15;修回日期:2024-04-15

基金项目:安徽省科技重大专项资助项目(202203a05020008);长三角科技创新共同体联合攻关专项资助项目(2022CSJGG1501)和济宁市产业创新重大技术“全球揭榜”资助项目(2022JBZP002)

作者简介:张炳力(1968—),男,安徽合肥人,博士,合肥工业大学教授,博士生导师,通信作者,E-mail:zhangbingli@hfut.edu.cn.

能力,激光雷达逐渐成为实现高阶辅助驾驶功能的重要硬件配置之一,因此国内外智能车企业纷纷启用激光雷达方案。

由于处理器算力有限,业界主要通过聚类算法对激光雷达点云信息进行处理。常见的聚类算法包括 K -means、DBSCAN 等。 K -means 算法易于实现,但容易将噪声点归入距离最近的簇中一同聚类。针对此类缺陷,文献[1]提出一种借助阈值的 K -means 聚类算法,根据阈值限定不同中心间的最小距离,从而提高聚类准确度;传统 DBSCAN 算法无法处理空间分布不均的情况,对此,文献[2]提出结合区域生长与密度聚类的算法,并同时考虑了障碍物几何特征与点云密度特征的关联方法。尽管聚类算法得到了广泛运用,但其高误检率仍不可避免,精度更高、泛化性更强的点云检测方法仍有待探索。

近年来,深度学习逐渐成为环境感知领域的主流技术方案,越来越多基于深度学习的 3D 点云检测算法^[3-6]被提出。文献[7]提出 VoxNet 网络,在深度学习框架中将点云数据转换为体素网格;但会带来计算量大的问题。文献[8]和文献[9]分别提出 PointNet 和 PointNet++ 网络,利用多层感知机(multilayer perceptron,MLP)对每个点进行表征,并设计了编码-解码(Encoder-Decoder)结构,通过上采样和下采样实现局部特征与全局特征的拼接;但因为 PointNet 系列算法是直接对密集点云进行处理,所以算法效率难以保证。文献[10]提出 PointPillars 网络,通过采用一种新的编码方式,将 3D 点云信息转化为伪图像格式,然后利用 2D 卷积网络提取特征,从而避免了计算量较大的 3D 稀疏卷积,提高了网络的运

行效率,是当前少有的能够满足工业落地需求的高实时性算法;但因为 PointPillars 主干网络仅采用自下而上的卷积方式获取语义信息,缺少低层次定位信息对高层次语义信息的融合修饰,并且该主干网络仅通过上采样对特征进行拼接,所以对特征的提取效果并不理想;此外,由于行人和骑行者等体积较小的目标点云数量相对较少、容易被遮挡,PointPillars 主干网络提取的特征难以准确描述小目标语义信息,易造成漏检和误检。

针对上述问题,本文在 PointPillars 算法的基础上提出一种改进的方法。首先,引入特征金字塔模块,将高层的语义特征通过上采样与底层的定位特征进行融合;其次,在主干网络中加入残差模块和卷积注意力模块(convolutional block attention module,CBAM),通过关注不同维度的特征信息来加强小目标的识别能力;最后,通过设置不同尺寸的检测头来完成对不同类别目标的检测。本文通过 KITTI 数据集和实车试验验证所改进的 PointPillars 小目标检测算法的可行性。

1 改进的 PointPillars 算法设计

PointPillars 是一种一阶段端到端的点云检测网络,如图 1 所示。其主要流程可分为 3 个阶段:① 通过柱状特征网络(pillar feature net,PFN)将点云信息转换为伪图像格式;② 主干网络利用 2D 卷积神经网络提取特征信息;③ 将特征信息输入检测头中进行 3D 检测框的回归预测。

图 1 中: C 为一个点柱经过处理后的特征维度; H 、 W 表示网格尺寸。

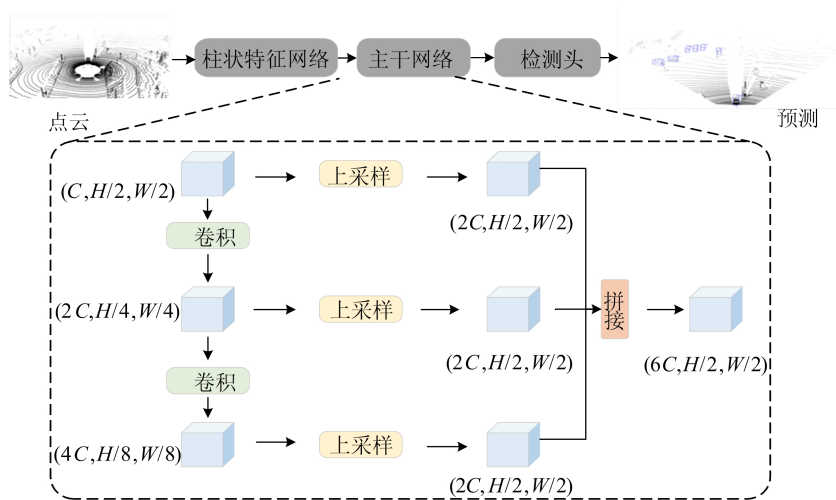


图 1 PointPillars 主干网络结构

1.1 改进的 PointPillars 主干网络

为提高小目标检测精度,本文提出改进的 PointPillars 主干网络,如图 2 所示,其主要包含

提取特征层和预测特征层。改进的 PointPillars 主干网络仍以点云提取网络的输出结果作为输入,即维度为 (C, H, W) 的伪图像。

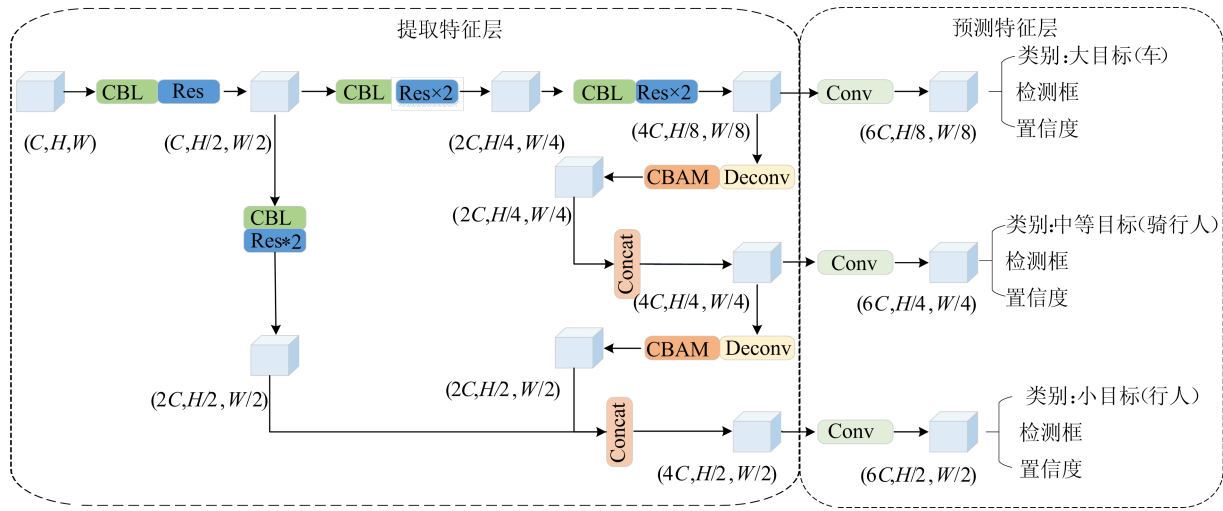


图 2 改进的 PointPillars 主干网络结构

1.1.1 提取特征层

将维度为 (C, H, W) 的特征图连续经过 3 个下采样模块 (CBL) 和残差模块 (Res), 分别得到维度为 $(C, H/2, W/2)$ 、 $(2C, H/4, W/4)$ 、 $(4C, H/8, W/8)$ 的特征图。其中: 下采样模块依次包含 1 个二维卷积层、1 个批归一化层以及 1 个 Leaky ReLU 激活函数层; 设计的残差模块^[11] 如图 3 所示, 先将输入的特征图经过 2 个下采样层, 然后将结果与经过第 1 个下采样层的结果进行拼接, 得到的结果再通过 1 个下采样层后作为残差模块的最终输出。利用残差模块可以增加网络深度, 进而加强网络的特征表达能力, 达到提高模型识别准确度的目的。此外, 残差模块利用内部的拼接操作实现了跳跃连接, 缓解了网络模型深度增加带来的梯度消失问题, 使得整个网络更加鲁棒, 抗过拟合能力更强。

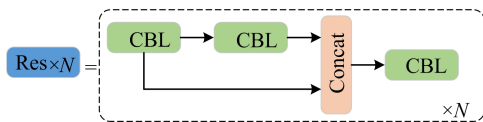


图 3 残差模块结构

的检测精度。

具体而言, 首先将得到的维度为 $(4C, H/8, W/8)$ 的特征图经过转置卷积和卷积块注意力模块 (convolutional block attention module, CBAM)^[12], 得到维度为 $(2C, H/4, W/4)$ 的特征图, 再与维度为 $(2C, H/4, W/4)$ 的特征图进行拼接, 得到维度为 $(4C, H/4, W/4)$ 的特征图, 其中设计的 CBAM 注意力机制模块结构如图 4 所示, 通过将特征依次输入相对独立的通道注意力模块和空间注意力模块, 实现自适应特征修饰。

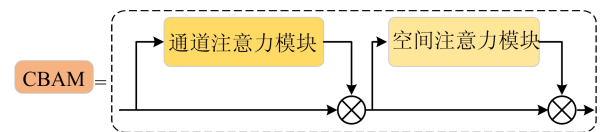


图 4 CBAM 注意力机制模块结构

通道注意力模块将每一个通道都看作一个特征检测器, 关注有用的通道特征信息。

为了提高通道注意力特征的计算效率, 需要对特征图 F 的空间维度进行压缩, 利用平均池化和最大池化以产生不同的空间上下文信息, 计算公式如下:

$$F_{\max}^c = \text{Maxpool}(F) \quad (1)$$

$$F_{\text{avg}}^c = \text{Avgpool}(F) \quad (2)$$

然后将得到的特征输入到多层感知器 (MLP) 中, 通过拼接和 Sigmoid 激活函数处理后输出通道注意力特征图 $M_c(F)$, 计算公式如下:

$$M_c(F) = \text{Sigmoid}\{\text{MLP}(F_{\max}^c) + \text{MLP}(F_{\text{avg}}^c)\} \quad (3)$$

空间注意力模块根据特征图空间内部关系,关注有用的特征信息位置,侧重于特征图中的有效位置信息。首先需要对特征图 F 进行最大池化和平均池化操作,然后将产生的特征图进行拼接,再通过卷积层和 Sigmoid 激活函数处理,最终输出空间注意力特征图 $M_s(F)$,计算公式如下:

$$M_s(F) = \text{Sigmoid}\{\text{Conv}_{7 \times 7}\{\text{Concat}[\text{Maxpool}(F); \text{Avgpool}(F)]\}\} \quad (4)$$

网络消融实验结果见表 1 所列。本文通过消融试验对比,最终选择在转置卷积后引入 CBAM 注意力机制模块,利用网络学习上采样过程中的深层次语义信息,关注网络通道和空间 2 个维度上有意义的特征信息,从而提高卷积网络提取小目标特征的能力。

表 1 网络消融实验结果

特征金字塔	CBAM 注意力机制	平均精度均值/%
		63.40
✓		69.83
✓	✓	77.09

将得到的维度为 $(4C, H/4, W/4)$ 的特征图经过转置卷积和 CBAM 特征注意力机制模块,得到维度为 $(2C, H/2, W/2)$ 的特征图,再将第 1 次下采样后得到的维度为 $(C, H/2, W/2)$ 的特征图经过下采样和残差模块得到维度为 $(2C, H/2, W/2)$ 的特征图,将得到的 2 个维度为 $(2C, H/2, W/2)$ 的特征图进行拼接,最终得到维度为 $(4C, H/2, W/2)$ 的特征图。

1.1.2 预测特征层

根据道路目标的形态、体积差异,本文定义车辆为大目标、骑行人为中等目标、行人为小目标,并设计了 3 种不同尺寸的检测头来完成检测。首先,将维度为 $(4C, H/8, W/8)$ 的特征图经过 1 次卷积,得到维度为 $(6C, H/8, W/8)$ 的第 1 个检测头,因为第 1 个检测头的感受野相对较大,所以利用此检测头检测车辆,并根据车的外观尺寸,设计锚框的长、宽、高分别为 3.90、1.60、1.50 m;其次,将维度为 $(4C, H/4, W/4)$ 的特征图经过 1 次卷积,得到维度为 $(6C, H/4, W/4)$ 的第 2 个检测头,因为第 2 个检测头的感受野处于中等水平,所以利用此检测头检测骑行行人,并根据骑行人的外观尺寸,设计锚框的长、宽、高分别为 1.76、0.60、

1.73 m;最后,将维度为 $(4C, H/2, W/2)$ 的特征图经过 1 次卷积得到维度为 $(6C, H/2, W/2)$ 的第 3 个检测头,因为第 3 个检测头的感受野相对较小,所以利用此检测头检测行人,并根据行人的外观尺寸,设计锚框的长、宽、高分别为 0.80、0.60、1.73 m。

1.2 数据增强

在真实应用场景下,目标点云受到遮挡、距离变化和尺寸变化等因素的影响,点云数据极易受到干扰,导致网络检测效果下降。数据增强可以使数据集包含更加充分的特征信息,避免网络训练过拟合,提高网络训练过程中的抗干扰能力。深度学习中点云数据增强方法^[13-16]主要包括点云平移、点云缩放、点云旋转等。

为了解决道路小目标的漏检、误检等问题,本文采用放大 2 倍的方式对点云原始数据进行增强,如图 5 所示。图 5 中:蓝色点云为原始点云,粉色点云为放大 2 倍后的点云;红色检测框为原始检测框,绿色检测框为放大 2 倍后的检测框。通过放大 2 倍使点云呈现不同的形态和特征,便于网络学习小目标细微的特征,同时可以丰富数据集中道路目标的特征信息和背景信息,提高点云检测网络的鲁棒性和抗干扰能力。

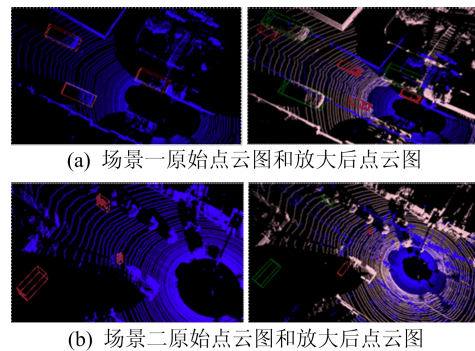


图 5 数据增强前、后点云图

2 网络训练与评估

2.1 试验环境

本文改进的 PointPillars 算法在 Ubuntu18.04 系统中进行可行性分析,利用 Anaconda 搭建算法训练环境,利用 PyTorch1.10.0 深度学习框架和 Python 3.9 进行网络修改;在硬件层面,利用一台高性能工作站进行网络训练,其中央处理器为 Intel Xeon(R) Gold 6226R CPU @2.90 GHz×64,其显卡为 2 张 NVIDIA RTX A6000 48 GiB。

本文搭建的网络框架利用 KITTI 公开数据

集进行训练和评估,最终将训练和评估后的模型进行实车试验。KITTI 3D 目标检测数据集包含 7 481 张训练图像、7 518 张测试图像以及所对应的雷达点云信息。本文将训练集划分为 3 712 张用于训练,3 769 张用于验证。训练过程中,设置批尺寸为 8,初始学习率为 0.001,使用 Adam 优化器,每 30 个迭代轮次进行 1 次衰减系数为 0.1 的更新,总迭代轮次为 300 次。

2.2 评价指标

根据 KITTI 的检测基准,将检测框按照被遮挡程度和检测框高度分为简单、中等和困难 3 种级别。困难级别表示最小检测框为 25 像素,基本难以看见,最大截断为 50%;中等级别表示最小检测框为 25 像素,部分特征被遮挡,最大截断为 30%;简单级别表示最小检测框为 40 像素,完全可以看见,最大截断为 15%。

本文利用验证集来进行网络评价,分别在 2D 层面和 3D 层面对检测结果进行验证。首先,根据匹配结果计算每个类比的查准率 P 和查全率 R ,其中:查准率表示被分类为正样本的样本中真正为正类别的比例;查全率表示真正为正类别的样本中被正确分类为正类别的比例。然后,根据查准率和查全率分别计算每个类别的平均精度 (average precision, AP)。最后对所有类别的 AP

值取平均,得到平均精度均值 mAP,作为网络模型最终的评价指标。根据 KITTI 数据集的检测标准,平均精度使用不同交并比阈值下的检测精度来度量,针对汽车这一类目标,匹配平均精度采用 AP^{70} 来度量;针对骑行人和行人这两类目标,匹配平均精度采用 AP^{50} 来度量。

查准率 P 和查全率 R 表达式如下:

$$P = \frac{T_P}{T_P + F_P} \quad (5)$$

$$R = \frac{T_P}{T_P + F_N} \quad (6)$$

其中: T_P 表示真正类 (True Positive),即实例是正类、预测也为正类,通常计算预测边界框和真实边界框的交并比置信度,大于设定置信度阈值的预测边界框即为 T_P ; F_P 表示假正类 (False Positive),即实例是负类、预测为正类; F_N 表示假负类 (False Negative),即实例是正类、预测为负类。

2.3 检测精度分析

本文选取 PointPillars 和两阶段的 PointRCNN 作为基准网络,验证所提出的改进 PointPillars 网络对小目标识别精度的提升。首先对验证集中车、骑行人、行人 3 类目标的不同检测级别分别计算平均精度 AP,然后求得平均精度均值 mAP,见表 2 所列。

表 2 网络评价指标

%

网络	指标	类别	3D Bbox			2D Bbox		
			简单	中等	困难	简单	中等	困难
PointPillars	AP	车	88.92	74.45	69.66	96.77	90.79	85.90
		骑行人	88.58	74.99	70.73	95.47	84.16	79.55
		行人	59.86	53.32	47.84	74.38	65.76	59.15
	mAP		79.12	67.59	62.74	88.87	80.24	74.87
PointRCNN	AP	车	89.67	76.43	69.83	99.00	90.83	85.92
		骑行人	92.04	77.64	74.60	93.96	82.65	79.77
		行人	57.65	50.17	44.94	71.01	61.30	54.80
	mAP		79.79	68.08	63.12	87.99	78.26	73.50
改进的 PointPillars	AP	车	88.69	76.60	71.99	96.75	92.90	86.06
		骑行人	90.91	76.58	75.58	94.58	86.56	81.53
		行人	62.15	54.18	48.51	76.77	67.52	59.89
	mAP		80.58	69.12	65.36	89.37	82.33	75.83

经分析可知,由于设置了不同尺寸的检测头来分别检测车、骑行人和行人等道路目标,本文提出的改进 PointPillars 算法针对部分类别的检测精度可能存在下降,但在平均精度均值 mAP 上都有提升。其中:对于车这类大目标,检测平均精度均值平均提升约 1%;而对于骑行人和行人这两类小目标,检测平均精度均值均有较大提升,平

均提升 1.74%。

2.4 可视化分析

场景一和场景二可视化效果分析如图 6、图 7 所示。

图 6a 所示为场景一的原始图像信息、原始点云信息以及所包含目标的真值;图 6b 所示为传统 PointPillars 算法的检测结果,但行人小目标被漏

检;图 6c 所示为本文提出的改进 PointPillars 小目标检测算法的检测结果,可以看出,传统 PointPillars 算法检测结果中对于行人小目标的漏检问题得到解决,证明针对行人这类小目标,本文提出的算法具有更好的检测效果。

图 7a 所示为场景二的原始图像信息、原始点云信息以及所包含目标的真值,其中中间道路目标为轨道车辆,本文不予考虑;图 7b 所示为传统 PointPillars 算法的检测结果,检测出的车辆和骑行者分别用红色检测框标出,但图中存在车辆和骑行者这两类目标漏检情况;图 7c 所示为本文提出的改进 PointPillars 小目标检测算法的检测结果,可以看出,传统 PointPillars 算法检测结果中对于车和骑行者的漏检问题得到解决,证明针对车和骑行者这两类目标,本文提出的算法也具有较好效果。

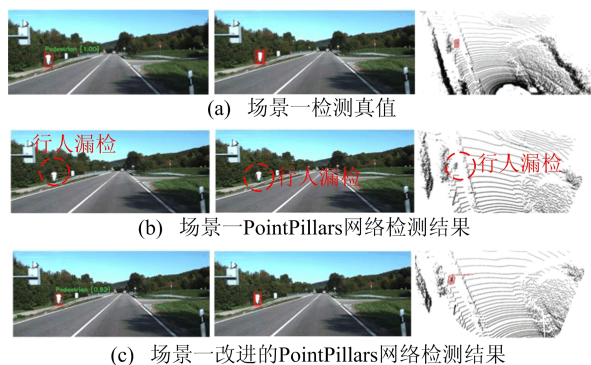


图 6 场景一可视化效果

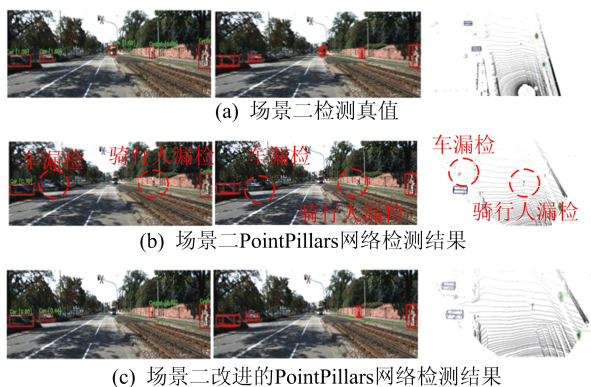


图 7 场景二可视化效果

2.5 实车验证

本文将设计改进的 PointPillars 小目标检测算法模型部署到实车上进行实车验证,如图 8 所示。

利用改进 PointPillars 小目标检测算法对采集到的点云进行实时检测,帧率可达 65 Hz,满足

实时性检测要求。实车可视化结果如图 9 所示。其中:图 9a 所示为场景一中检测到的道路目标的 2D 检测框、3D 检测框以及点云检测结果;图 9b 所示为场景二中检测到的道路目标的 2D 检测框、3D 检测框以及点云检测结果。



图 8 实车试验

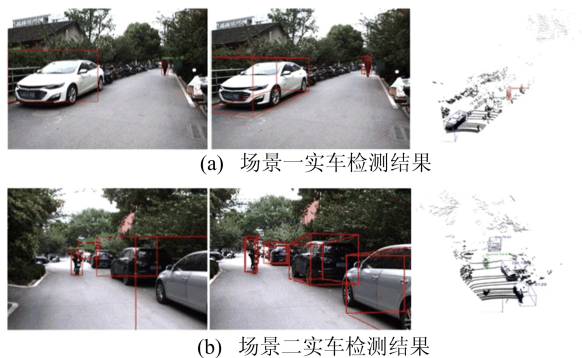


图 9 实车可视化结果

从图 9 可以看出,本文设计的算法对车辆这类大目标、骑行者和行人这两类小目标都有较好的检测效果,可以应用在实际场景中。

3 结 论

针对传统的激光雷达点云检测算法中存在的小目标漏检、误检等情况,本文提出了一种改进的 PointPillars 小目标检测算法。首先,在主干网络中引入自下而上和自上而下的特征金字塔模块,拼接高层特征与低层特征,并设置 3 种检测头来分别检测大、小目标;其次,在主干网络中加入残差模块和 CBAM 注意力机制,通过突出重要语义信息、抑制不重要的信息,提高网络抗干扰能力,

使检测网络更加鲁棒;最后,通过数据增强的方式丰富数据集,为训练过程提供更多的道路目标特征信息和背景信息。

在公开数据集 KITTI 上对本文改进的 PointPillars 小目标检测算法进行了训练和评估。评价指标表明,本文算法对于行人、骑行等小目标的检测平均精度均值平均提升了 1.74%;可视化结果表明,相较于其他算法,本文算法可以减少道路目标的漏检情况;实车试验结果表明,本文算法能够以 65 Hz 的频率完成实时检测。因此,本文设计改进的 PointPillars 算法可以较好地应用在智能驾驶感知领域,为以后更高级别的自动驾驶提供更好的参考。

[参 考 文 献]

- [1] 夏显召,朱世贤,周意遥,等. 基于阈值的激光雷达 K 均值聚类算法[J]. 北京航空航天大学学报, 2020, 46(1): 115-121.
- [2] 汪世财,谈东奎,谢有浩,等. 基于激光雷达点云密度特征的智能车障碍物检测与跟踪[J]. 合肥工业大学学报(自然科学版), 2019, 42(10): 1311-1317.
- [3] Son H, Koh E, Sung S. A study on integrated navigation algorithm using deep learning based lidar odometry and inertial measurement[J]. Journal of Institute of Control, Robotics and Systems, 2020, 26(9): 715-723.
- [4] Alaba S Y, Ball J E. A survey on deep-learning-based lidar 3d object detection for autonomous driving[J]. Sensors, 2022, 22(24): 9577.
- [5] Chen G, Wang F, Qu S, et al. Pseudo-image and sparse points: vehicle detection with 2D LiDAR revisited by deep learning-based methods[J]. IEEE Transactions on Intelligent Transportation Systems, 2020, 22(12): 7699-7711.
- [6] Kim K, Kim C, Jang C, et al. Deep learning-based dynamic object classification using LiDAR point cloud augmented by layer-based accumulation for intelligent vehicles[J]. Expert Systems with Applications, 2021, 167: 113861.
- [7] Maturana D, Scherer S. VoxNet: a 3D convolutional neural network for real-time object recognition[C]//2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). [S. l.]: IEEE, 2015: 922-928.
- [8] 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. [S. l.]: IEEE, 2017: 652-660.
- [9] 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: 5099-5108.
- [10] 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. [S. l.]: IEEE, 2019: 12697-12705.
- [11] Wang J, Zhou J, Liu X, et al. Spectral and spatial residual attention network for joint hyperspectral and lidar data classification[C]//2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). [S. l.]: IEEE, 2021: 278-281.
- [12] Woo S, Park J, Lee J Y, et al. CBAM: convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). [S. l.]: IEEE, 2018: 3-19.
- [13] 董艳秋, 万旺根, 胡文博, 等. 基于可变形卷积和数据增强的三维多目标检测[J]. 工业控制计算机, 2023, 36(3): 22-24.
- [14] Bao W, Xu B, Chen Z. MonoFENet: monocular 3D object detection with feature enhancement networks[J]. IEEE Transactions on Image Processing, 2019, 29: 2753-2765.
- [15] Zhao Q, Gao X, Li J, et al. Optimization algorithm for point cloud quality enhancement based on statistical filtering[J]. Journal of Sensors, 2021, 2021: 1-10.
- [16] Li Z, Yao Y, Quan Z, et al. Spatial information enhancement network for 3D object detection from point cloud[J]. Pattern Recognition, 2022, 128: 108684.

(责任编辑 胡亚敏)