

基于改进 PointPillars 的激光雷达三维目标检测

陈德江, 余文俊*, 高永彬

上海工程技术大学电子电气工程学院, 上海 201620

摘要 针对汽车自动驾驶时通过激光雷达进行三维目标检测时识别结果不准确、目标朝向检测偏差较大的问题, 提出一种基于改进 PointPillars 的激光雷达三维目标检测方法。首先, 基于 Swin Transformer 改进 PointPillars 的二维卷积降采样模块, 使得网络特征提取阶段能够使用自注意力机制来丰富上下文语义并获取全局特征, 增强算法的特征提取能力。其次, 利用点云立柱的特性将点云的地面部分去除, 降低冗余点云的影响, 从而提高三维目标检测的识别精度。在公开数据集 KITTI 上进行的验证实验结果表明: 所提方法具有更高的检测精度, 相较于原 PointPillars, 平均检测精度提升了 1.3 个百分点, 验证了该方法的有效性。

关键词 遥感; 深度学习; 三维目标检测; 激光雷达; Transformer

中图分类号 TP391.4 文献标志码 A

DOI: 10.3788/LOP220840

Lidar 3D Target Detection Based on Improved PointPillars

Chen Dejiang, Yu Wenjun*, Gao Yongbin

School of Electronic and Electrical Engineering, Shanghai University of Engineering Science,
Shanghai 201620, China

Abstract Aiming at the problems of inaccurate recognition results and large deviation of target orientation detection when 3D target detection is carried out by laser radar during auto driving, a 3D target detection method of laser radar based on improved PointPillars is proposed. First of all, based on Swin Transformer's improved two-dimensional convolution downsampling module of PointPillars, the self attention mechanism can be used in the network feature extraction phase to enrich context semantics and obtain global features, and enhance the feature extraction ability of the algorithm. Second, the ground part of the point cloud is removed by using the characteristics of the point cloud column to reduce the impact of redundant point clouds, so as to improve the recognition accuracy of 3D object detection. The experimental results on the public dataset KITTI show that the proposed method has higher detection accuracy. Compared with the original PointPillars, its average detection accuracy is increased by 1.3 percentage points, which verifies the effectiveness of the proposed method.

Key words remote sensing; deep learning; 3D object detection; lidar; Transformer

1 引言

三维目标检测算法是实现自动驾驶路径规划和安全避障的重要研究内容。随着激光雷达制造工艺的进步和深度学习算法的发展, 三维激光雷达输出点云的密度逐年增大, 测量精度越来越高, 而以雷达点云作为输入的深度学习检测器^[1]也逐渐成熟。在这种环境下, 通过激光雷达进行三维目标检测的方法变得更加流行。

近年来, 不仅国外涌现出很多实用的三维目标检

测算法, 国内学者也提出很多有效可行的三维目标检测算法。为了增强点云的特征提取能力, 对点云有更好的表征能力, Charles 等^[2]提出直接从点云中学习特征的 PointNet 算法。为了提高对行人、骑行者等小目标的检测精度, Ku 等^[3]提出基于多传感器输入融合的 AVOD 算法, 该算法的输入为标定好的雷达点云和相机图像, 通过多信息通道增加目标检测的精度。为了实现点对云的直接处理, Zhou 等^[4]提出采用三维卷积的 VoxelNet 算法, 该算法能够保留点云的原始数据特征, 让整个流程更加统一。为了提高点云的三维卷积

收稿日期: 2022-03-01; 修回日期: 2022-03-24; 录用日期: 2022-05-07; 网络首发日期: 2022-05-17

基金项目: 国家自然科学基金(61802253)

通信作者: yuwenjun@sues.edu.cn

运算速度, Yan 等^[5]提出基于三维稀疏卷积的 SECOND 算法,在提升算法精度的同时也提升了其运行效率,让三维目标检测离实际应用更进一步。为了提高目标检测算法的运行速度, Lang 等^[6]提出 PointPillars, PointPillars 采用全新的点云表征方式,将点云立柱化后转成伪图像,在保留点云三维特征的同时采用二维的深度卷积 backbone,极大加快了算法的运行速度,解决了三维目标检测因为运算速度过慢而无法实际应用的问题,但是检测精度较低,并且对于目标方向的判断存在较大误差。目前三维目标检测主要从两个角度进行处理:1)以三维的角度来解决问题,通过三维卷积算子来提取特征,这种方法的精度会更高一些,但是当前计算机对于三维卷积算子的运行仍旧很吃力,更不用说运用在嵌入式设备上;2)将点云表征成伪特征图,将三维降成二维,然后再用二维卷积神经网络进行特征提取,这种方式虽然会损失一些精度,但

在运行效率上完全可以应用于实际。

针对 PointPillars 算法中特征提取能力较弱的问题,本文基于 Swin Transformer (Swin-T)^[7]结构对原算法中的二维卷积降采样结构进行改进,使得特征提取的过程可以使用自注意力机制^[8],增强特征提取能力。同时针对激光雷达地面数据冗余的问题,采用基于点云立柱的地面去除算法降低无用点云的影响,提高检测精度。

2 PointPillars 算法

PointPillars 算法的运行速度很快,可以达到实时效果,甚至超过雷达的扫描频率,但代价就是算法精度相对较低,其将点云转换成立柱 (pillar) 进行表征,进而构成伪图片格式的数据,接着使用二维的 backbone 进行特征提取。PointPillars 的算法流程如图 1 所示。

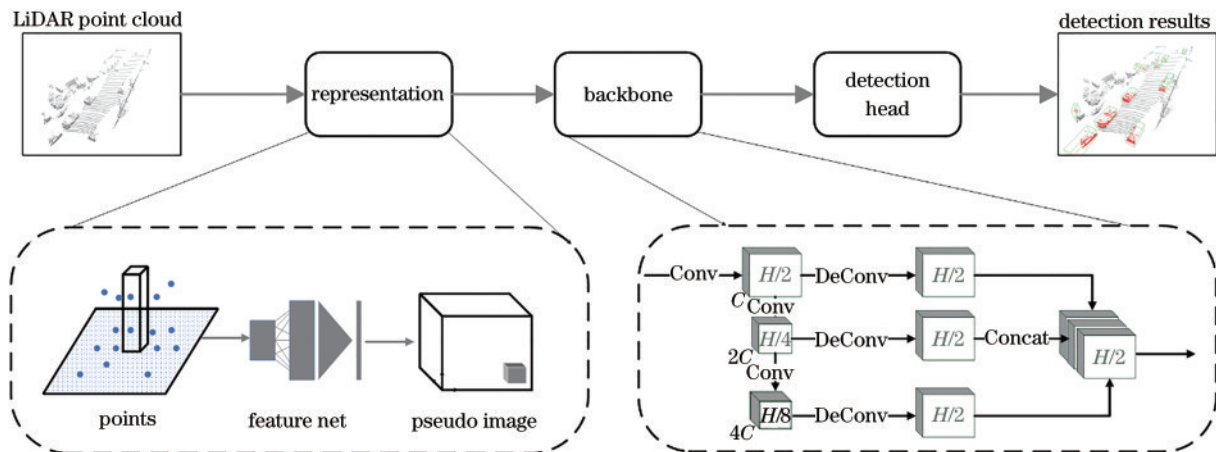


图 1 PointPillars 算法流程图

Fig. 1 Flow chart of PointPillars algorithm

PointPillars 算法大致分成 3 个部分:点云柱状化、特征提取、检测头输出。具体流程如下:

1) 首先将点云转化成为立柱,即将点云按照正方形的格子立柱进行划分,每个立柱中包含多个点,然后通过特征映射网络转换成伪图像,方便接下来使用二维的 backbone 进行特征学习;

2) 利用二维卷积网络 (CNN) 对伪图像进行多次降采样得到不同分辨率的特征图,然后将降采样后的多张特征图再上采样到相同的大小进行拼接,生成最终的特征图;

3) 将特征图输入 SSD^[9] 目标检测头进行包围框 (bounding box) 回归和目标分类,得到物体的位置和种类。

PointPillars 算法相较于其他三维目标检测算法的特点是采用点云立柱化的表征方式,将三维的点云最终转化成二维的伪图像,然后使用二维的 backbone 进行特征提取,这也是它运行速度快的原因。但是 backbone 阶段仅使用普通的二维卷积神经网络进行采

样和特征提取,虽然效率较高,但是特征提取能力相对较弱。而且点柱的划分导致大目标被分割,在检测时容易导致朝向判断误差较大,在 KITTI 数据集上检测可视化效果如图 2 所示,其中实线框为真实值,虚线框为预测值。从图 2 可以明显看出,检测框在朝向判断上存在较大误差。

目前也有一些针对 PointPillars 改进的算法,如 Frustum-PointPillars^[10]、基于稀疏彩色点云的 Color-PointPillars^[11]和基于注意力机制的 Attention-PointPillars^[12]。其中,Frustum-PointPillars 算法采用多模态输入,同时利用相机的二维信息和雷达的三维信息进行目标检测,首先通过图像检测出目标的二维位置,然后利用相机和雷达的联合标定将属于二维目标范围的点云剥离出来送入 PointPillars 中进行三维目标检测,该算法虽然提升了检测精度,但是本身应用时严格依赖传感器的联合标定,在车辆颠簸时存在风险。Color-PointPillars 同样为多模态输入,同时输入标定好的相机图片和雷达点云,通过联合标定将雷达点云和图像



图 2 原版 PointPillars 检测可视化效果

Fig. 2 Visual effect of original PointPillars detection

像素一一对应,这样就可以让点云获得目标的 RGB 色彩,增强点云的特征表达能力,但该算法同样依赖相机和雷达的联合标定,如果标定出现误差,检测效果将会大打折扣。而 Attention-PointPillars 在 PointPillars 算法中引入空间注意力和通道注意力实现对点云伪图像特征信息的放大和抑制,增强了算法的特征建模能力,但是忽略了点云伪图像的全局特征和上下文语义。

3 改进的 PointPillars 算法

3.1 基于 Swin Transformer 的特征图采样

原 PointPillars 算法利用纯粹的二维卷积神经网络进行特征图采样,获取不同分辨率的特征图,然后再将不同大小的特征图上采样到同样尺寸后拼接到一起,这是一种通用的特征提取方式,使用高分辨率但语义信息较弱的特征图来检测小目标,低分辨率但语义信息较强的特征图来检测大目标。

为了增强算法的特征提取能力,克服点云立柱化带来的分裂感,本研究基于 Swin Transformer 实现特征采样模块。Swin Transformer 是 Transformer^[13]的变体,而 Transformer 结构源于自然语言处理(NLP)领域,它最大的特色是自注意力机制^[14],自注意力机制使得特征图能够获取自身特征的全局信息和丰富的上下文语义,并且适用于并行计算。自注意力机制的经典计算方式为

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right), \quad (1)$$

式中: \mathbf{Q} 为注意力查询矩阵; \mathbf{K} 为注意力键矩阵; \mathbf{V} 为注

意力值矩阵; d_k 为 \mathbf{Q} 和 \mathbf{K} 中的向量维度。首先计算矩阵 \mathbf{Q} 和矩阵 \mathbf{K} 的点积,然后除以尺度 $\sqrt{d_k}$ 防止点积结果溢出,计算完点积后再利用 Softmax 函数将点积归一化为概率分布,最后再乘以矩阵 \mathbf{V} 就得到了注意力矩阵。Softmax 函数为

$$\text{Softmax}(z_i) = \frac{\exp(z_i)}{\sum_{c=1}^c \exp(z_c)}, \quad (2)$$

式中: c 为概率输出节点的数量; z_i 为第 i 个概率节点的输出值。通过 Softmax 可以将注意力矩阵归一化到和为 1 的非负数矩阵。

Transformer 结构虽然很好,但是应用到三维目标检测领域存在两个问题:1)三维目标实体变化范围较大,对于不同大小的物体 Transformer 效果并不稳定^[15];2)三维目标检测中特征图的高(H)和宽(W)很大,在转化成为特征单词(tokens)^[16]之后长度范围过大,这样在计算特征单词的注意力矩阵时会导致计算量急剧膨胀^[17],普通的计算机显存无法承受,更无法在嵌入式设备上运行。针对这两个问题,本研究采用 Transformer 结构的变体版本 Swin Transformer 进行代码实现,它通过两种滑动窗口操作:不重叠的 local window 和重叠的 cross window 将注意力计算限制在一定的窗口范围内,实现了类似特征降采样的金字塔操作,在引入 CNN 操作局部性的同时大大降低了 Transformer 在计算注意力矩阵时的计算量,而且滑动窗口的计算方式对于硬件实现更加友好,可以让 Transformer 模块运算速度更快。

参照 Swin Transformer 在视觉领域的应用,所设计的 Swin Transformer 网络结构如图 3 所示,其中 3 个

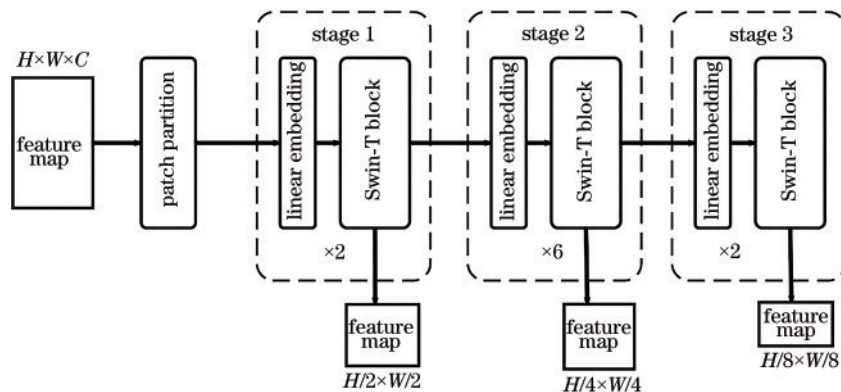


图 3 改进后的特征采样模块

Fig. 3 Improved feature sampling module

阶段的 Transformer 模块分别有自己的深度和注意力头数,降采样的分辨率和原 PointPillars 中采用相同的宽高,方便和原先的二维卷积神经网络替换,甚至一键切换。

3.2 激光雷达地面点云去除

三维激光雷达扫描周围环境后产生的点云较为稀疏,有时候地面的点云数量甚至超出非地面的点云数量,而且三维障碍物必然不属于地面,所以地面点云^[18]不但会影响算法的准确率,甚至会拖慢算法的运行速度。借鉴 PointPillars 本身的点云立柱化操作,可以对

立柱中的点云进行判断,如果该立柱为地面^[19],则可以将该立柱置空,采用零值的张量替代。具体的操作流程如下:判断立柱中最高点和最低点之间的距离差,如果距离差不超过阈值,则可以判断该立柱中的点云为地面。

$$H_{ij} - h_{ij} < s, \quad (3)$$

式中: i 和 j 为立柱的 x 轴和 y 轴坐标; H 为该立柱中的最高点; h 为立柱中的最低点; s 为高度差阈值。如果该式布尔值为真则说明该立柱为地面点云。最终的地面去除效果如图 4 所示:

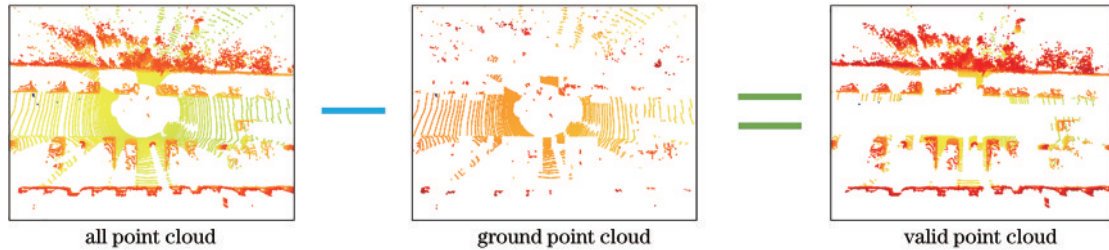


图 4 激光雷达点云去除地面部分
Fig. 4 Lidar point cloud removal of ground parts

3.3 目标朝向损失

在检测结果的回归过程中,三维的 bounding box 采用一个 7 维的向量表示 $(x_c, y_c, z_c, l, w, h, \theta)$,其中 x_c, y_c, z_c 为 bounding box 的中心点, l, w, h 为 bounding box 的长、宽、高, θ 为目标的朝向,采用 SmoothL1 回归损失进行模型训练。针对当方向角的预测方向和真实方向相反时损失剧增的问题,采用 sin 函数进行封装:

$$L_\theta = \text{SmoothL1}[\sin(\theta_p - \theta_t)], \quad (4)$$

式中: θ_p 为预测的朝向角度; θ_t 为真实的朝向角度。这样当 $\theta_p = \theta_t \pm \pi$ 时,损失也不会剧增。

另外针对检测目标的点云数据均匀分布时可能会出现目标朝向反向的问题,本研究参照 SECOND 的损失函数设计,引入 Softmax 分类函数来学习物体的方向,尽量减小目标朝向反向的误差。具体方法为将目标朝向是否正确区分用分类损失函数 Softmax 进行约束,当目标朝向角度大于 0 时为正,目标朝向角度小于

0 时为负,从而构成二分类,通过优化这个二分类损失来降低目标朝向反向的误差。

3.4 数据增强

采用 KITTI 数据集^[20]进行模型训练与测试,作为经典的数据集, KITTI 是当前自动驾驶领域最常用的数据集之一,其点云数据通过 Velodyne-64 线激光雷达获取,可以作为衡量三维目标检测算法效果的基准,但是总的数量相较于二维目标检测数据集仍然较少。而 Transformer 结构相对于其他深度学习模块需要更多数据进行训练才能够拟合,所以对雷达点云数据集采用点云变换的方式进行数据增强。

因为点云所构成的世界是有尺度的三维世界,所以同样可以通过对点云进行镜像翻转来获取新的点云,也可以通过沿着某个轴进行轻微的世界倾斜,还可以通过尺度缩放,比如轻微放大和缩小点云世界。通过这些点云变换,就可以获得更多不同的点云数据,数据增强后同一帧点云衍生出的点云如图 5 所示:

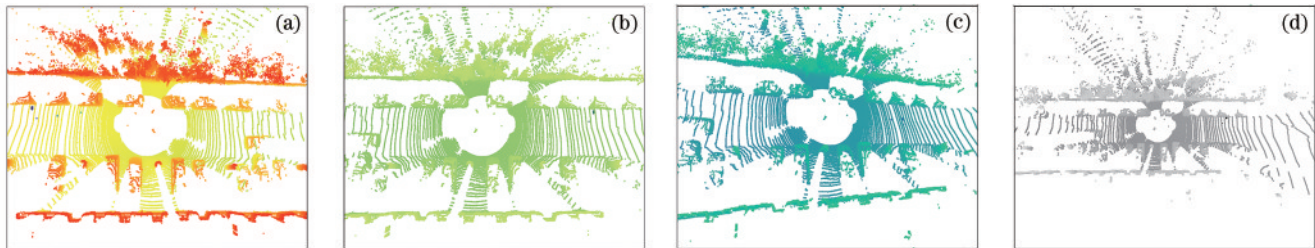


图 5 点云数据增强。(a)原始点云;(b)镜像点云;(c)倾斜点云;(d)缩放点云

Fig. 5 Point cloud data enhancement. (a) Raw point cloud; (b) mirrored point cloud; (c) tilted point cloud; (d) zoomed point cloud

4 实验与结果分析

4.1 实验环境

实验使用 PyTorch 深度学习框架和 OpenPCDet 目标检测框架,模型训练的主机使用的处理器为 Intel(R) Xeon(R) Gold 6240, GPU 为 RTX6000、显存为 48 GB,编程通过 PyCharm 2021.3(Professional Edition) × 64 实现,操作系统为 Ubuntu 18.04。在公开数据集 KITTI 上进行模型训练和测试,训练样本数量为 3769, batch size 设置为 8,使用 Adam 优化器训练,权重衰减值为 0.01,动量值为 0.9,学习率衰减值为 0.1,最大迭代次数设为 100。

4.2 评价指标

本研究的重点为目标朝向,通过平均方向相似度(AOS)指标来判断算法对于目标朝向检测的精度,AOS 的表达式为

$$\begin{cases} S_{AOS} = \frac{1}{11} \sum_{r \in \{0, 0.1, \dots, 1\}} \max_{\tilde{r}: \tilde{r} \geq r} s(\tilde{r}) \\ s(r) = \frac{1}{|D(r)|} \sum_{i \in D(r)} \frac{1 + \cos \Delta_{\theta}^{(i)}}{2} \delta_i \end{cases}, \quad (5)$$

式中: r 为目标检测的召回率;预测结果与真实值之间的余弦距离归一化 $s \in [0, 1]$; $D(r)$ 为召回率 r 下预测为正样本的集合; $\Delta_{\theta}^{(i)}$ 为目标朝向预测值和真实值之间的残差。

4.3 实验分析

首先选择合适的 Swin-T 超参数,包括每个阶段 Transformer 块的深度和注意力头的数量,进行实验的超参数配置如表 1 所示。参考 Swin-T 中的配置,各阶段 block 的深度一般为 2 或者 6,考虑到算法运行的效率和显存大小进行配置。注意力头的数量采用 2 的倍数,并且随着特征采样的降维,对头数进行翻倍。在机器上进行训练和测试,每次训练的时间一般在 8~12 h,测试结果如表 2 所示,条形图对比如图 6 所示。

表 1 Swin-T 模块的超参数配置

Table 1 Hyperparameter configuration of Swin-T module

Model	Depth	Head
M1	(1, 3, 1)	(2, 4, 8)
M2	(2, 6, 2)	(2, 4, 8)
M3	(2, 2, 6)	(2, 4, 8)
M4	(2, 6, 2)	(2, 4, 2)
M5	(4, 8, 4)	(4, 8, 4)

KITTI 数据集包括 3 个难度的场景: Easy、Moderate、Hard,难度越高场景中目标受遮挡的程度越高,点云越稀疏,通常也意味着离雷达越远。M1 的实验结果表明,当 Transformer block 的深度较浅时,效果较差,甚至达不到原版算法的效果,因为准确率没有达到 80%,所以条形图中未予显示。通过 M2 和 M3 的对

表 2 不同 Swin-T 超参数配置准确率对比

Table 2 Comparison of accuracy rates of different Swin-T

Model	hyperparameter configurations				unit: %
	Easy	Moderate	Hard	Average	
M1	50.17	40.42	38.7	43.09	
M2	94.12	89.55	88.48	90.71	
M3	94.23	89.77	88.8	90.93	
M4	94.13	89.23	88.15	90.5	
M5	94.57	89.65	88.75	90.99	

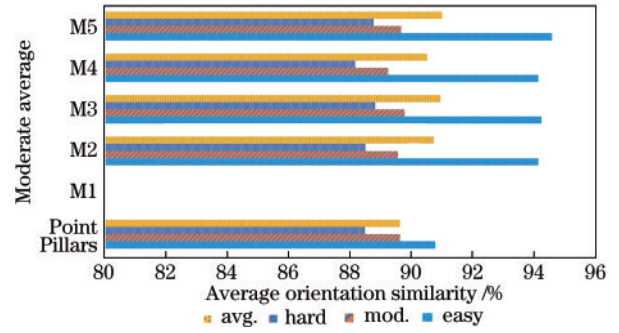


图 6 超参数选取对比实验结果

Fig. 6 Comparison of experimental results for hyperparameter selection

比可以得出,block 在靠后的阶段更深,效果会更好。通过 M2 和 M4 对比可以得出,注意力头数随着特征图的分辨率降低而翻倍是一个正确的选择。最后的 M5 虽然各方面的精度都很高,但是代价是翻倍的显存和降低的运算速度,代价过大并不值得选择,但是也证明了 Transformer block 的叠加和注意力头数的增加是有助于精度提高的,以后随着硬件的发展,算法的精度也可以继续提高。

综上所述,最优的模型是 M3, AOS 精度在 3 个难度场景下的精度均优于原版 PointPillars,说明其更具备鲁棒性,而且显存接近原版算法。改进后的 PointPillars 和现阶段其他激光雷达三维目标检测算法的 AOS 精度对比如表 3 所示,改进前后的 PointPillars 算法检测效果对比折线图如图 7 所示,另外算法改进

表 3 检测结果对比

Table 3 Comparison of test results

Model	Easy	Moderate	Hard	Average
PointPillars ^[6]	90.77	89.61	88.47	89.61
SECOND ^[5]	90.76	89.77	88.82	89.78
SECOND-IoU ^[5]	89.72	88.73	88.33	88.92
PointRCNN ^[21]	90.76	89.58	89.03	89.79
PointRCNN-IoU ^[21]	90.70	89.32	88.87	89.63
Part-A2-Free ^[22]	90.68	89.00	88.64	89.44
AS-PointPillars ^[12]	90.48	88.32	86.51	88.44
AP-PointPillars ^[12]	90.68	88.92	86.90	88.83
Proposed model	94.23	89.77	88.8	90.93

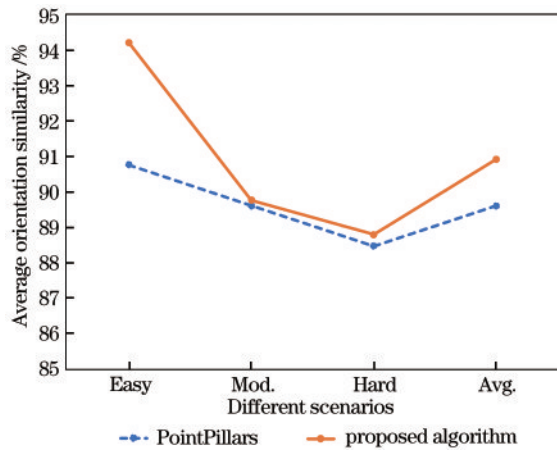


图7 算法改进前后检测精度对比

Fig. 7 Comparison of detection accuracy before and after algorithm improvement

前后 GPU 显存占用和运行速度如表 4 所示。

从表 3 可以看出,改进后的 PointPillars 相较于原版在目标朝向的判断精度上提升了 1.3 个百分点,并且领先于现阶段其他激光雷达三维目标检测方法。通过可视化检测结果也可以看出,改进后的算法和原版在检测效果对比上具有优势。改进前后对比如图 8 所

表 4 算法改进前后的显存占用以及运行速度对比
Table 4 Comparison of GPU memory usage and running speed before and after algorithm improvement

Model	GPU memory /MB	Running speed /s
PointPillars	1267	0.036
Proposed model	1359	0.058

示,其中虚线为预测的 3D 包围框,实线为真实的 3D 包围框,可明显看出改进后的算法对于目标朝向的判断更加准确,误差偏移角度更小,预测框和真实框更加贴合。另外,本研究通过实验证明了利用 Swin-T 进行特征提取优于原版网络单纯使用 CNN 进行特征提取,检测结果更准确,且不同的 Swin-T 超参数配置对于结果有着不同程度的提升。随着 block 深度的加深和注意力头数的增加会提高模型检测的精度,但同时也会降低模型推理速度,因此可以根据不同车辆工控机的配置来选择不同的 Swin-T 模型,对于运算能力较弱的工控机可以选择低配版的 Swin-T,如 M4,对于运算能力较强的工控机可以选择配置更重量级但是检测效果也更好的 Swin-T,如 M5。所以可以根据自动驾驶车辆工控机的算力水平来选择不同的 Swin-T 超参数配置,从而达到检测速度和精度的平衡。

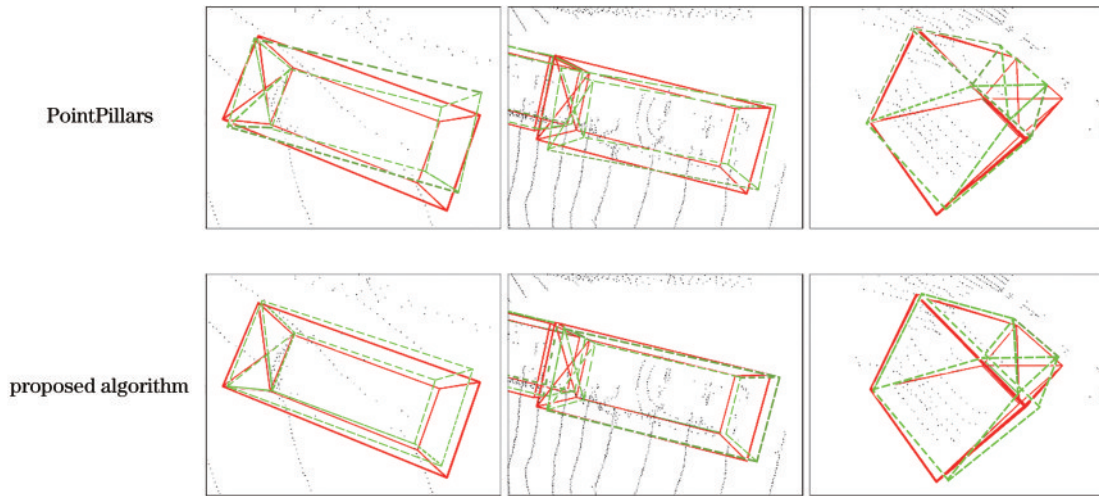


图8 算法改进前后检测效果对比

Fig. 8 Comparison of detection effects before and after algorithm improvement

5 结论

当前自动驾驶和辅助驾驶变得愈加流行和实用,其中路径规划和安全避障都离不开对道路场景的三维感知,只有将道路车辆的位置、大小和朝向精确地检测出来,才能够合理地规划行驶路线和防止车辆碰撞。所提改进算法较好地达成了检测准确率和运行效率的平衡,可以有效应用于道路三维目标检测。将 PointPillars 的特征提取模块采用 Swin Transformer 实现,增强了算法的特征提取能力;通过点云立柱化去除激光雷达点云中的地面部分进行数据增强。实验结果

表明,改进后的算法相比较原算法在精度上提升了 1.3 个百分点,且在各个场景下精度均优于原算法,证明了该算法的有效性和鲁棒性。

本研究主要针对单一帧的激光雷达点云进行检测,但是点云的采集本身属于动态的过程,如果能够利用多帧之间的联系,应该可以进一步增加目标检测的准确率。所以下一步工作是利用多帧点云之间的关系来获取更加丰富的特征信息,从而提高算法的准确率。

参 考 文 献

[1] 赵亮, 胡杰, 刘汉, 等. 基于语义分割的深度学习激光

- 点云三维目标检测[J]. 中国激光, 2021, 48(17): 1710004.
- Zhao L, Hu J, Liu H, et al. Deep learning based on semantic segmentation for three-dimensional object detection from point clouds[J]. Chinese Journal of Lasers, 2021, 48(17): 1710004.
- [2] Charles R Q, Hao S, Mo K C, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 77-85.
- [3] Ku J, Mozifian M, Lee J, et al. Joint 3D proposal generation and object detection from view aggregation [C]//2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 1-5, 2018, Madrid, Spain. New York: IEEE Press, 2018.
- [4] Zhou Y, Tuzel O. VoxelNet: end-to-end learning for point cloud based 3D object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, June 18-23, 2018, Salt Lake City, UT, USA. New York: IEEE Press, 2018: 4490-4499.
- [5] Yan Y, Mao Y X, Li B. SECOND: sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.
- [6] Lang A H, Vora S, Caesar H, et al. PointPillars: fast encoders for object detection from point clouds[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 12689-12697.
- [7] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: hierarchical vision transformer using shifted windows[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. New York: IEEE Press, 2021: 9992-10002.
- [8] 陈小龙, 赵骥, 陈思溢. 基于注意力编码的轻量化语义分割网络[J]. 激光与光电子学进展, 2021, 58(14): 1410012.
- Chen X L, Zhao J, Chen S Y. Lightweight semantic segmentation network based on attention coding[J]. Laser & Optoelectronics Progress, 2021, 58(14): 1410012.
- [9] Liu W, Anguelov D, Erhan D, et al. SSD: single shot MultiBox detector[M]//Leibe B, Matas J, Sebe N, et al. Computer vision-ECCV 2016. Lecture notes in computer science. Cham: Springer, 2016, 9905: 21-37.
- [10] Paigwar A, Sierra-Gonzalez D, Erkent Ö, et al. Frustum-PointPillars: a multi-stage approach for 3D object detection using RGB camera and LiDAR[C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW), October 11-17, 2021, Montreal, BC, Canada. New York: IEEE Press, 2021: 2926-2933.
- [11] 罗玉涛, 秦瀚. 基于稀疏彩色点云的自动驾驶汽车 3D 目标检测方法[J]. 汽车工程, 2021, 43(4): 492-500.
- Luo Y T, Qin H. 3D object detection method for autonomous vehicle based on sparse color point cloud[J]. Automotive Engineering, 2021, 43(4): 492-500.
- [12] 詹为钦, 倪蓉蓉, 杨彪. 基于注意力机制的 PointPillars+三维目标检测[J]. 江苏大学学报(自然科学版), 2020, 41(3): 268-273.
- Zhan W Q, Ni R R, Yang B. An attention-based PointPillars+3D object detection[J]. Journal of Jiangsu University (Natural Science Edition), 2020, 41(3): 268-273.
- [13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12)[2022-02-04]. <https://arxiv.org/abs/1706.03762>.
- [14] 张祥东, 王腾军, 朱劭俊, 等. 基于扩张卷积注意力神经网络的高光谱图像分类[J]. 光学学报, 2021, 41(3): 0310001.
- Zhang X D, Wang T J, Zhu S J, et al. Hyperspectral image classification based on dilated convolutional attention neural network[J]. Acta Optica Sinica, 2021, 41(3): 0310001.
- [15] 史宝岱, 张秦, 李瑶, 等. 基于改进残差注意力网络的 SAR 图像目标识别[J]. 激光与光电子学进展, 2021, 58(8): 0810008.
- Shi B D, Zhang Q, Li Y, et al. SAR image target recognition based on improved residual attention network [J]. Laser & Optoelectronics Progress, 2021, 58(8): 0810008.
- [16] Sun F, Liu J, Wu J, et al. BERT4Rec: sequential recommendation with bidirectional encoder representations from transformer[C]//CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management, November 3-7, 2019, Beijing, China. New York: ACM Press, 2019: 1441-1450.
- [17] Parmar N, Vaswani A, Uszkoreit J, et al. Image transformer[EB/OL]. (2018-02-15)[2022-02-04]. <https://arxiv.org/abs/1802.05751>.
- [18] 邵靖滔, 杜常清, 邹斌. 基于点云簇组合特征的激光雷达地面分割方法[J]. 激光与光电子学进展, 2021, 58(4): 0428001.
- Shao J T, Du C Q, Zou B. Lidar ground segmentation method based on point cloud cluster combination feature [J]. Laser & Optoelectronics Progress, 2021, 58(4): 0428001.
- [19] Postica G, Romanoni A, Matteucci M. Robust moving objects detection in lidar data exploiting visual cues[C]//2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), October 9-14, 2016, Daejeon, Korea (South). New York: IEEE Press, 2016: 1093-1098.
- [20] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11): 1231-1237.
- [21] Shi S S, Wang X G, Li H S. PointRCNN: 3D object proposal generation and detection from point cloud[C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 15-20, 2019, Long Beach, CA, USA. New York: IEEE Press, 2019: 770-779.
- [22] Shi S S, Wang Z, Wang X G, et al. Part-A² net: 3D part-aware and aggregation neural network for object detection from point cloud[EB/OL]. (2019-07-08)[2022-02-04]. <https://arxiv.org/abs/1907.03670v1>.