

doi: 10.19562/j.chinasae.qcgc.2024.07.008

# 面向自动驾驶场景的多目标点云检测算法\*

陶乐<sup>1</sup>,王海<sup>1</sup>,蔡英凤<sup>2</sup>,陈龙<sup>2</sup>

(1. 江苏大学汽车与交通工程学院, 镇江 212013; 2. 江苏大学汽车工程研究院, 镇江 212013)

**[摘要]** 基于点云的三维目标检测算法是自动驾驶系统中关键技术之一。目前基于体素的无锚框检测算法是学术界的研究热点,但是大多数研究都侧重于设计复杂的二阶段修正模块,在牺牲巨大的算法延迟的情况下带来有限的算法性能提升。而单阶段无锚框点云检测算法虽然具有更加精简的检测流程,但其检测性能难以满足自动驾驶场景的需求。对此,本文基于无锚框检测算法 CenterPoint,提出了一种面向自动驾驶场景的单阶段无锚框点云目标检测算法。具体来说,本文通过引入编码解码稀疏模块,极大地促进了三维特征提取器对于空间非连通区域的信息交互,保证了三维特征提取器能够提取到满足各类目标检测的特征。此外,考虑到现有的二维特征融合主干与基于中心点的无锚框检测头的适配存在挑战性,本文通过引入自校正卷积和大核注意力模块,能够有效提取到目标区域的点云特征,并将目标区域的点云特征聚集到中心点区域,从而提升算法对于目标的召回率和检测精度。本文所提出的算法在大规模公开数据集 nuScenes 上进行模型训练和实验验证,与基准算法相比,mAP和NDS分别提升了5.97%和3.62%。同时,本文将所提出的算法在基于自主搭建的实车平台上进行实际道路实验,进一步证明了所提出算法的有效性。

**关键词:** 自动驾驶;深度学习;点云检测;无锚框

## Multi-object Detection Algorithm Based on Point Cloud for Autonomous Driving Scenarios

Tao Le<sup>1</sup>, Wang Hai<sup>1</sup>, Cai Yingfeng<sup>2</sup> & Chen Long<sup>2</sup>

1. School of Automotive and Traffic Engineering, Jiangsu University, Zhenjiang 212013;

2. Automotive Engineering Research Institute, Jiangsu University, Zhenjiang 212013

**[Abstract]** The three-dimensional object detection algorithm based on point cloud is one of the key technologies in the autonomous driving system. Currently, the voxel-based anchor-free detection algorithm is a research hot-spot in academia, but most researches focus on designing complex refinement stage, at the expense of huge algorithm latency, to bring limited performance improvement. Although the single-stage anchor-free point cloud detection algorithm has a more streamlined detection process, its detection performance cannot satisfy the needs of autonomous driving scenarios. In this regard, based on the anchor-free detection algorithm CenterPoint, a single-stage anchor-free point cloud object detection algorithm for autonomous driving scenarios is proposed in this paper. Specifically, the encoding and decoding sparse module is introduced in this paper, which greatly promotes the information interaction of the spatial non-connected areas of the three-dimensional feature extractor, ensuring that the three-dimensional feature extractor can extract features that satisfy various target detection. In addition, considering that it is challenging to adapt the existing two-dimensional feature fusion backbone to the center-based head, in this paper self-calibrated convolution and large kernel attention modules are introduced in to effectively extract point cloud features of the target area, which are then gathered into the center point area, thereby improving the algorithm's recall and accuracy of the target. The proposed algorithm in this article is trained and experimentally verified on the large-

\* 国家重点研发计划项目(2023YFB2504401)资助。

原稿收到日期为2024年01月23日,修改稿收到日期为2024年02月29日。

通信作者:王海,教授,博士,E-mail:wanghai1019@163.com。

scale public dataset of nuScenes. Compared with the benchmark algorithm, mAP and NDS are increased by 5.97% and 3.62% respectively. At the same time, the actual road experiments with the proposed algorithm are conducted on a self-built vehicle platform, further proving the effectiveness of the proposed algorithm.

**Keywords:** autonomous driving; deep learning; point cloud detection; anchor-free

## 前言

环境感知是智能汽车实现安全行驶的重要前提。相较于摄像头,激光雷达作为主动有源传感器,可以直接获得三维空间中的深度信息,对于光照条件和气候条件具有很强的鲁棒性,因此被认为是高阶自动驾驶汽车不可或缺的传感器之一。随着深度学习技术的飞速发展和激光雷达成本的持续降低,基于深度学习的激光雷达三维目标检测算法在近年成为学术界研究热点,相较于传统基于聚类的点云检测算法在性能上有显著提升<sup>[1]</sup>。

目前主流的基于深度学习的点云目标检测算法可以分为3种:基于原始点云的方法、基于体素的方法、基于原始点和体素混合的方法<sup>[2]</sup>。基于原始点的方法的特点是直接使用神经网络对原始点云进行特征提取,其中 PointNet<sup>[3]</sup>和 PointNet++<sup>[4]</sup>奠定了基于原始点的点云检测算法的基础架构,后续的 Point-RCNN<sup>[5]</sup>、IA-SSD<sup>[6]</sup>都是在此基础上进行改进得到的。基于原始点的方法可以实现对原始点云数据的充分利用,但也极大限制了其推理速度,因此难以应用在自动驾驶系统。为了将卷积神经网络应用于点云特征提取,基于体素的方法<sup>[7]</sup>将稀疏的不规则点云数据转换成规则的体素表征形式,使得三维卷积神经网络可以处理体素化后的点云数据。基于体素的点云检测算法在牺牲较少的点云信息的情况下,极大地提升了算法的推理速度,因此在工业界得到广泛应用。基于原始点和体素混合的方法利用基于原始点的方法提出了一个高效的二阶段修正模块,有效弥补了基于体素的方法在体素化过程中会造成的不可避免的点云信息丢失,进一步提升了点云检测算法的性能。PV-RCNN 系列<sup>[8-9]</sup>是最具代表性的基于原始点和体素的混合算法,该算法将原始点云特征和体素特征聚集到通过最远点采样得到的关键点中,随后利用关键点特征对一阶段检测算法得到的检测结果进一步修正。基于原始点和体素的混合方法虽然进一步提升了检测的准确性,但也明显降低了算法的推理速度。

此外,根据是否采用先验锚框,基于深度学习的

点云检测算法也可以划分为有锚框算法和无锚框算法。锚框这个概念最开始是由二维图像检测算法 Faster-RCNN<sup>[10]</sup>提出,即根据先验信息在特征图每个位置预先为各个类别目标设定固定三维尺寸和朝向的边界框,锚框的设定能够极大提升对环境中目标的召回率以及促进算法快速收敛,因此有锚框算法在图像领域具有非常优异的检测性能。受到二维图像领域的启发,很多学者将锚框迁移到点云目标检测领域,同样取得了不错的检测效果。然而,相较于二维图像领域,锚框在点云领域的应用存在明显的局限性。三维空间中的目标具有复杂多变朝向,有限数量的轴对齐锚框难以拟合待检测目标所有的朝向,且锚框的使用会引入额外的计算量。对此, CenterPoint<sup>[11]</sup>和 AFDet<sup>[12]</sup>等算法将图像领域的无锚框算法 CenterNet<sup>[13]</sup>迁移到点云检测算法中,将待检测目标建模成中心点,实现了优于主流有锚框算法的无锚框点云检测算法。同时,用中心点表征目标的方式极大地简化了自动驾驶中的下游任务,如目标跟踪等。

为满足自动驾驶汽车对于算法效率的需求,本文中选择了基于体素的无锚框算法 CenterPoint 进行研究。CenterPoint 沿用 SECOND<sup>[14]</sup>算法中的三维特征提取器,该架构主要由通用稀疏卷积<sup>[15]</sup> (regular sparse convolution, RSC) 和残差稀疏模块 (residual sparse block, RSB) 构成。其中 RSB 通常由两层子流形稀疏卷积<sup>[16]</sup> (submanifold sparse convolution, SSC) 组成。SSC 很好地保证了输入特征图和输出特征图稀疏程度的一致性,但也阻碍了空间非连通区域的特征进行信息交互,从而导致三维特征提取器的感受野不够。后续的一些工作也进行了尝试来提升三维特征提取器的感受野,如采用大核卷积<sup>[17]</sup> 或 Transformer<sup>[18]</sup> 架构,但这些尝试造成了严重的算法延迟。对此,受 HEDNet<sup>[19]</sup> 的启发,本文引入编码解码稀疏模块,能够在保持输入特征图和输出特征图稀疏程度一致的前提下,有效促进空间非连通区域内特征的信息交互,从而保证三维特征提取器能够提取到足够的特征满足各类目标的检测。此外, CenterPoint 也沿用了 SECOND 中的二维特征融合主干,该架构能够很好地适配有锚框算法,但与

CenterPoint中基于中心点的无锚框检测头适配时存在挑战。基于中心点的无锚框检测头须依靠目标的中心点区域的特征来回归目标的类别和边界框属性,这就需要二维特征融合主干在进行特征融合时能将目标区域内的有效特征聚集到中心点区域,但CenterPoint中原有的二维特征融合主干难以满足这一点。对此,本文中首先感受野更大的自校正卷积<sup>[20]</sup>进行特征提取,能够有效提取到目标区域的点云特征,接着利用大核注意力机制将提取到的目标区域特征聚集到目标中心点区域,从而实现与基于中心点的无锚框检测头良好适配。在大规模开源数据集nuScenes<sup>[21]</sup>上进行大量实验验证算法的有效性,同时在实际道路场景上采集的数据上进行算法验证。

# 1 点云检测算法设计

## 1.1 框架概览

单阶段无锚框点云目标检测算法通常由点云体素编码、三维特征提取器、二维特征融合主干以及无

锚框检测头组成。以经典的无锚框三维目标检测算法CenterPoint为基准,围绕网络结构进行优化,所提出的框架如图1所示。具体来说,首先采用动态体素编码方式代替原来的硬体素编码,并采用PointNet-mini<sup>[3]</sup>进行特征提取,得到更具有判别性的体素级特征。考虑到目前主流的三维特征提取器难以捕捉特征之间的长距离依赖关系,本文引入稀疏编码解码模块并重新设计了三维特征提取器的网络架构,能够更加高效地进行点云特征提取。此外,考虑到自动驾驶场景的复杂性,现有的二维特征融合主干存在感受野不足和难以满足基于中心点的无锚框检测头对于各类目标的回归,引入自校正卷积和大核注意力机制,能够促进对所提取特征的充分融合,同时将目标特征聚集到目标中心点。最后,在CenterPoint原有检测头基础上增加IoU预测分支和目标朝向分类分支,能够进一步提升检测器的检测精度。

## 1.2 动态体素编码

原始点云经过体素编码后才能用三维卷积进行

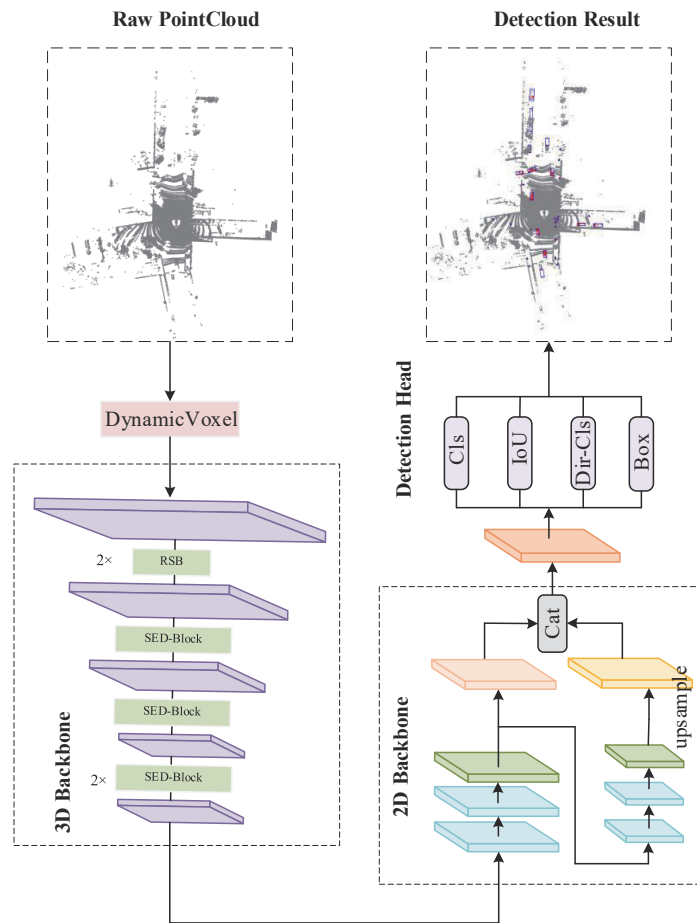


图1 算法网络架构

特征提取,因此体素特征编码对于基于体素的三维目标检测算法至关重要。CenterPoint原采用的硬体素编码,需要预先设定体素数量和体素内点云数量,对于点云数量低于阈值的体素网格需要用零填补,而对于点云数量超过阈值的体素网格需要进行随机舍弃处理,然后直接取非空体素内点云特征的均值作为体素级特征。因此,硬体素编码须预先分配额外的内存,且随机舍弃点云操作给模型训练引入了更多的随机性。此外,硬体素编码没有很好地考虑点云中噪点的存在,不利于后续三维特征提取器进行特征提取。对此,本文中采用动态体素编码方式<sup>[22]</sup>,有效地避免了额外内存消耗和随机舍弃点云的操作。并且本文采用PointNet-mini<sup>[3]</sup>对体素化后的点云进行特征提取,对原始点云中的噪点起到了初步的过滤作用。具体来说,假设  $V = \{p_i = [x_i, y_i, z_i, r_i, t_i] \in R^{N_i \times 5}\}$  为一个非空体素,包含  $N_i$  个点。接着,对非空体素内的点特征进行增广处理,得到  $\hat{V} = \{p_i = [x_i, y_i, z_i, r_i, t_i, x_i^c, y_i^c, z_i^c, x_i^r, y_i^r, z_i^r] \in R^{N_i \times 11}\}$ , 其中  $[x_i^c, y_i^c, z_i^c]$  是相对于当前体素中心点的偏移坐标,  $[x_i^r, y_i^r, z_i^r]$  是体素内所有点云的均值坐标。接着使用PointNet-mini对增广后的点云特征进行提取,得到体素级特征  $F_i$ 。

$$F_i = MLP_2(MLP_1(\hat{V})) \quad (1)$$

其中  $MLP_1$  和  $MLP_2$  均由全连接层、批量归一化层以及 ReLU 激活函数组成。

### 1.3 三维特征提取器

三维特征提取器的作用是对经过体素编码后的体素级特征进一步提取,对于检测器的性能至关重要。目前主流的基于体素的三维目标检测算法通常采用SECOND中稀疏卷积三维主干,主要由通用稀疏卷积和残差稀疏模块堆叠而成。其中每个残差稀疏模块主要由两层卷积核尺寸为3的子流形稀疏卷积构成,子流形稀疏卷积保证了输入特征图和输出特征图稀疏程度的一致性,从而保证了特征提取的高效性,但这也阻碍了特征图中空间不连续特征之间的信息交互。虽然通用稀疏卷积能够在一定程度上促进空间不连续特征之间的信息交互,但是过多使用通用稀疏卷积会显著降低特征图的稀疏程度,从而降低特征提取效率。对此,受到HEDET<sup>[19]</sup>的启发,在三维特征提取器中引入编码解码稀疏模块,如图2所示。该编码解码稀疏模块主要由通用稀疏卷积、残差稀疏模块、稀疏反卷积(inverse submanifold convolution, InverseSubM)组成。在特征编码阶段中,首先使用通用稀疏卷积进行下采样,其目的是将特征图中有效特征进行聚集,随后使用残差稀疏模

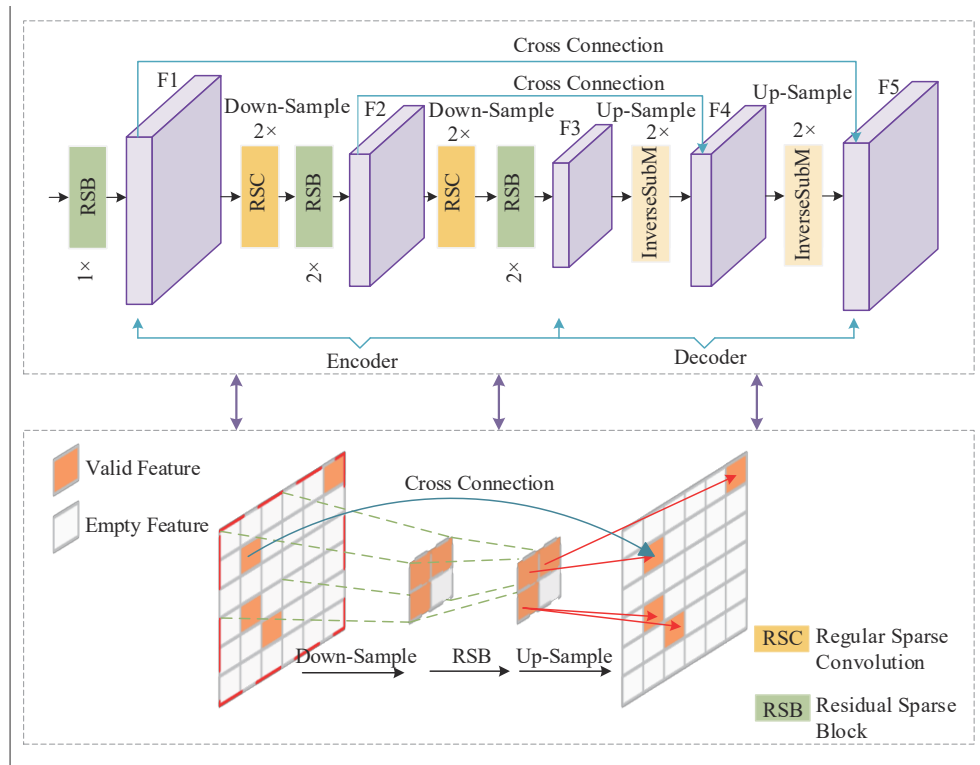


图2 编码解码稀疏模块

块对聚集后的特征进行特征提取,促进有效特征之间的信息交互。随后,稀疏反卷积将编码器输出的特征上采样到与输入特征图一致的分辨率,接着将上采样的特征图与编码器中对应的特征图进行特征叠加,需要注意的是本文仅将上采样的特征图中的有效特征进行特征叠加。在编码器阶段进行两次二倍下采样以保证充分的特征交互,编码解码稀疏模块输出的特征可以表示为

$$F_1 = RSB(X) \quad (2)$$

$$F_2 = RSB(RSC(F_1)) \quad (3)$$

$$F_3 = RSB(RSC(F_2)) \quad (4)$$

$$F_4 = InverseSubM(F_3) + F_2 \quad (5)$$

$$F_5 = InverseSubM(F_4) + F_1 \quad (6)$$

对此,本文设计了更高效的三维特征提取器,其

架构如图3所示。所设计的三维特征提取器仍采用层级的方式进行点云特征提取。考虑到浅层点云几何特征对于最终预测的目标边界框的定位精度非常重要,因此相较于CenterPoint在三维特征提取器的初始阶段增加了特征通道数量,由原来的16增加到32。且在初始阶段本工作没有进行特征下采样操作,仅采用子流形稀疏卷积和残差稀疏卷积模块进行特征提取,有效缓解了浅层几何特征的丢失。在分别进行2倍、4倍、8倍下采样之后,使用编码解码稀疏模块替代残差稀疏卷积模块进行特征提取,能够有效提升空间非连通区域的特征交互。考虑网络的深度提升有利于高层点云特征的提取,在8倍下采样之后,采用了两层编码解码稀疏模块进行特征提取。同时,为防止引入过多的计算量,相较于原始架构,本文中减少了8倍下采样之后的特征通道数,

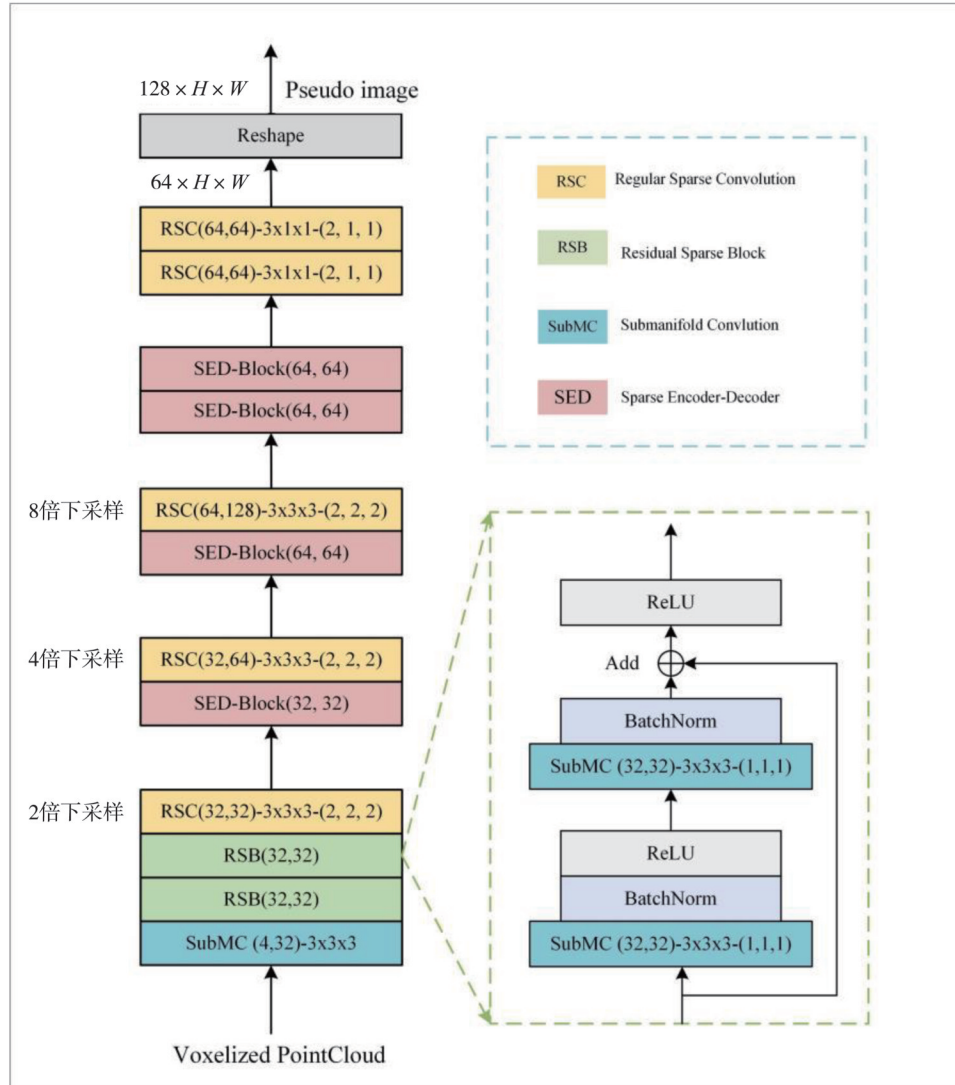


图3 三维特征提取器

由原来的 128 缩减到 64。经过上述改进,所设计的三维特征提取器相较于基准算法更加高效。

#### 1.4 二维特征融合主干

二维特征融合主干以三维特征提取器输出的二维伪鸟瞰图特征为输入,其作用是将提取到的点云特征进一步融合,从而得到满足各类目标回归的特征。以往的基于体素的无锚框通常是沿用有锚框算法 SECOND 中的二维特征融合主干的架构,如图 4

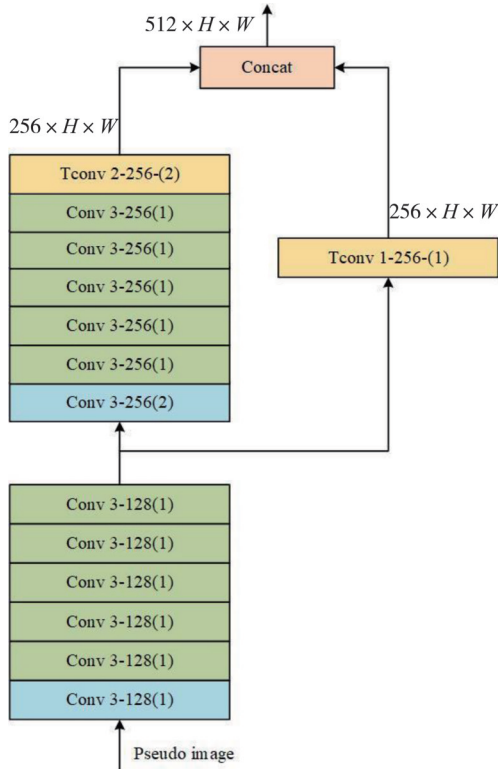


图4 原始二维特征融合主干

所示。该架构以通用二维卷积进行特征提取,其缺点是感受野有限的通用二维卷积难以提取到覆盖整个目标的全局特征<sup>[23]</sup>,尤其是对于尺寸较大的公交车和货车等。此外,该架构与原始的有锚框算法能够很好地适配,但与基于中心点的无锚框算法仍存在一定的适配挑战性,其原因在于基于中心点的无锚框算法是利用目标的中心点特征对目标类别和目标边界框属性进行预测的。而由于点云数据固有的稀疏性和激光雷达扫描的工作特性,大多数目标在中心位置处都不存在点云数据。因此对于基于中心点的无锚框检测算法而言,二维特征融合主干在保证特征融合充分的前提下,须有效地将提取到的目标特征聚集到中心位置,而 CenterPoint 中沿用的二维特征融合主干难以很好地将目标区域的点云特征

聚集到目标的中心点位置,导致对目标的召回率低。对此,本文中对二维特征融合主干进行改进,得到了更加有效的二维特征融合主干,如图 5 所示。首先

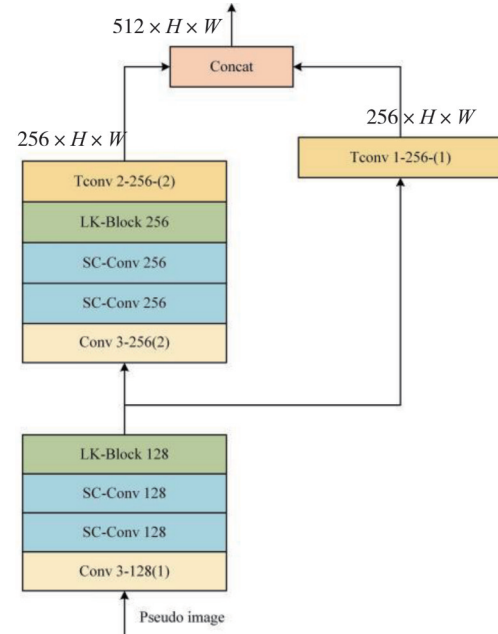


图5 改进后二维特征融合主干

引入具有更大感受野的自校正卷积进行特征提取,相较于通用二维卷积,通过其设计的注意力机制能够有效提取到目标区域的点云特征。为平衡算法的推理速度,本文中分别使用两层自校正卷积进行浅层和深层特征提取,特征通道数分别为 128 和 256。另外,引入了大核注意力机制<sup>[24]</sup>,其架构如图 6 所示。该大核注意力机制模块可以有效地将所提取到的目标区域的点云特征聚集到目标中心点位置。且为提升对各类尺寸目标的识别性能,对大核注意力机制的卷积核尺寸进行调整,采用的卷积核尺寸包括  $1 \times 1$ 、 $5 \times 5$ 、 $7 \times 7$  等。可以发现,本文中的大核注意力机制模块设置在校正卷积特征提取之后,因此可以使网络在提取到目标区域特征后,有效地将目标区域特征聚集到目标中心点位置,能够促进后续基于中心点的无锚框检测头对目标的回归。所设计的二维特征融合主干实现了与基于中心点无锚框检测头的良好适配,进一步提升了检测器的检测性能。

#### 1.5 检测头

检测头以融合后的点云特征为输入,进行最终的目标类别预测和目标边界框属性的回归。本文中仍沿用 CenterPoint 中的基于中心点的无锚框检测头

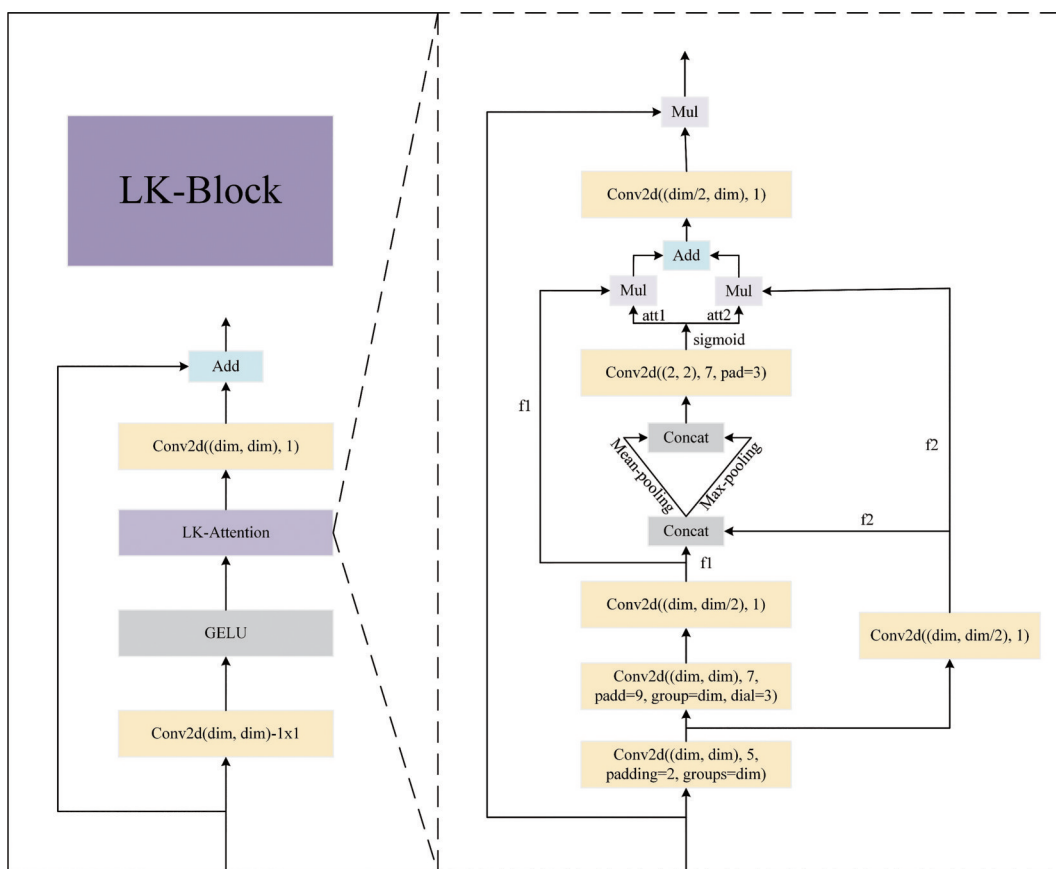


图6 大核注意力模块

设计,原版的 CenterPoint 检测头包括 5 个检测子头,即热力图预测头、XY 平面下中心点位置偏移量回归头、中心点 Z 轴位置回归头、目标三维尺寸回归头以及目标朝向角回归头。本文中为进一步提升模型对于目标朝向角预测的准确性,参照 SECOND 算法,在检测头处增加了一个目标朝向角分类分支。此外,为缓解模型所预测的目标边界框存在置信度和定位精度不匹配的问题,在检测头处增加一个 IoU 分支,并在后处理的非极大值抑制 (non-maximum suppression, NMS) 阶段,用 IoU 分支预测的得分来修正目标的置信度得分,则最终用于 NMS 的置信度得分为

$$confidence = score^{1-\alpha} \cdot iou^{\alpha} \quad (7)$$

式中:  $score$  为初始的置信度得分;  $iou$  为 IoU 分支预测的得分;  $\alpha \in [0, 1]$ , 为超参数, 对  $score$  和  $iou$  进行加权。

### 1.6 损失函数

为能够让模型在训练时快速收敛,应用一系列的损失函数用于各个子检测头的监督。其中热力图预测头采用 focal loss<sup>[25]</sup> 进行监督,XY 平面下中心点

位置偏移量回归头、中心点 Z 轴位置回归头、目标三维尺寸回归头、目标朝向角回归头以及 IoU 分支采用 L1 loss 进行监督。此外,目标朝向角分类分支采用交叉熵损失进行监督。最终的总损失为

$$L_{reg} = \lambda_{center} L_{center} + \lambda_{offset} L_{offset} + \lambda_z L_z + \lambda_{ry} L_{ry} \quad (8)$$

$$L_{total} = L_{reg} + \lambda_{hp} L_{hp} + \lambda_{iou} L_{iou} + \lambda_{dir} L_{dir} \quad (9)$$

式中  $\lambda_{hp}$ 、 $\lambda_{center}$ 、 $\lambda_{offset}$ 、 $\lambda_z$ 、 $\lambda_{ry}$ 、 $\lambda_{iou}$ 、 $\lambda_{dir}$  分别为各个检测子头损失对应的权重,取值范围为  $[0, 1]$ 。

## 2 实验验证

### 2.1 公开数据集介绍

nuScenes 数据集是一个适用于三维目标检测的大规模自动驾驶数据集,其中训练集、验证集、测试集分别包含 28k、6k、6k 帧标注的点云,每帧点云大概有 30k 个点。该数据集标注了具有长尾分布的 10 个类别。nuScenes 数据集采用平均检测精度 (mean average precision, mAP) 和 nuScenes 检测得分 (nuScenes detection scores, NDS) 作为评价指标。mAP 为所有类别目标在 0.5、1、2、4 m 距离阈值下的

平均检测精度。NDS为mAP和其他5个评价指标(分别衡量位移、尺寸、朝向、速度、属性误差)的加权平均值。

## 2.2 实施细节

利用开源点云目标检测框架OpenPCDet<sup>[26]</sup>实现了所提出的算法。在nuScenes数据集上,文中设置检测范围沿着X轴和Y轴均为 $[-54\text{ m}, 54\text{ m}]$ ,Z轴范围为 $[-5\text{ m}, 3\text{ m}]$ 。设置体素尺寸为 $(0.075\text{ m}, 0.075\text{ m}, 0.2\text{ m})$ 。使用8张V100分布式训练模型20轮, batch\_size设置为32。同时,使用AdamW<sup>[27]</sup>作为优化器,采用one-cycle学习率策略,其中最大学习率为 $1 \times 10^{-3}$ ,权重衰减为0.01,动量设置为0.85~0.95之间。

对原始数据集进行合理的数据增广能够促进模型对数据的学习,因此在训练模型时,本文中采用了多种数据增强方法,使用X和Y轴随机翻转、随机因子为 $[0.9, 1.1]$ 的全局缩放、在 $[-\pi/4, \pi/4]$ 之间的随机全局旋转以及在 $[0, 0.5\text{ m}]$ 之间的随机全局平移。此外,还采用真值采样策略来增加训练的点云帧中各类别目标的真值数量。同时,文中也使用了

CBGS<sup>[28]</sup>策略来缓解数据集中目标呈长尾分布的问题。

## 2.3 实验结果

对所提出算法在大规模公开数据集nuScenes上的实验结果进行分析,主要包括与基准算法(Baseline)和与主流基于体素的算法进行对比。表1所示为改进后的算法与Baseline在nuScenes验证集上的实验结果对比。从表中可以直观看到,所提出的算法的mAP和NDS相较于Baseline分别提升了5.97%和3.62%。从单类别检测结果来看,所提出的算法除对于障碍物(Barrier)的检测结果低于Baseline之外,对于其他所有类别的检测性能均优于Baseline,对于货车(Truck)、公交车(Bus)、拖车(Trailer)、工程车(CV)等大型车辆,平均检测精度分别提升了2.48%、3.05%、6.97%、7.24%。且所提出的算法对于自行车(Bicycle)这个类别检测性能提升最为明显,平均检测精度提升了14.6%。为了更加直观地对比改进后的算法与Baseline的检测性能,在nuScenes验证集上进行可视化对比,如图7所示。图中绿色边界框为真值,蓝色边界框为检测框,红色

表1 改进算法与Baseline实验结果对比

算法	Car	Truck	Bus	Trailer	CV	Ped	Motor	Bicycle	TC	Barrier	mAP	NDS
Baseline	0.849	0.574	0.708	0.381	0.169	0.851	0.59	0.420	0.698	0.683	0.592 2	0.664 8
改进算法	0.863 1	0.598 8	0.738 5	0.450 7	0.241 4	0.866 9	0.736 4	0.607	0.750 2	0.666	0.651 9	0.701 0
提升	+1.41%	+2.48%	+3.05%	+6.97%	+7.24%	+1.59%	+8.8%	+18.7%	+5.22%	-1.7%	+5.97%	+3.62%

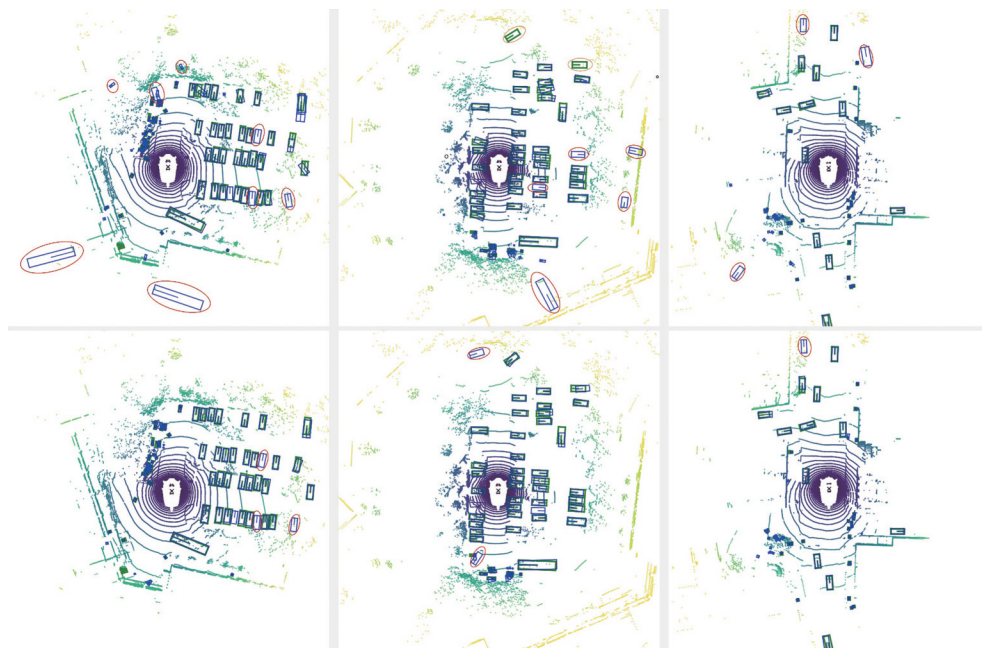


图7 改进算法与Baseline可视化对比

椭圆表示误检,橙色椭圆表示漏检。从上往下,第1行为Baseline的检测结果,第2行为改进后算法的检测结果。从图中可以直观发现,改进后的算法相较于Baseline有更少的误检和漏检,进一步证明改进后的算法检测性能更优。

也将所提出的算法与目前主流的基于体素的点云检测算法在nuScenes验证集上的实验进行对比,如表2所示。从mAP来看,所提出的算法相较于

PointPillar-MultiHead<sup>[29]</sup>、CBGS<sup>[28]</sup>、CenterPoint、VoxelNext<sup>[30]</sup>、CenterFormer<sup>[31]</sup>、PillarNet(res18)<sup>[32]</sup>、Li3DeTr<sup>[33]</sup>、UVTR-L<sup>[34]</sup>、Transfusion-L分别提升20.56%、14.60%、5.97%、4.66%、9.79%、5.29%、3.79%、4.29%、0.62%,NDS分别提升11.87%、7.81%、3.62%、3.45%、4.90%、2.71%、2.50%、2.40%、0.67%<sup>[35]</sup>。从实验结果来看,所提出的算法明显优于主流的基于体素的算法。

表2 改进算法与主流算法性能对比

算法	mATE ↓	mASE ↓	mAOE ↓	mAVE ↓	mAAE ↓	mAP ↑	NDS ↑
PointPillar-MultiHead	0.338 7	0.260 0	0.320 7	0.287 4	0.201 5	0.446 3	0.582 3
CBGS	0.311 5	0.255 1	<b>0.266 4</b>	0.262 6	0.204 6	0.505 9	0.622 9
CenterPoint	0.288 0	0.254 3	0.372 7	0.215 5	0.182 4	0.592 2	0.664 8
VoxelNext	0.301 1	0.252 3	0.405 7	0.216 9	0.185 6	0.605 3	0.666 5
CenterFormer	<b>0.275 0</b>	<b>0.252 0</b>	0.275 0	0.243 0	0.208 0	0.554 0	0.652 0
PillarNet(res18)	0.277 2	<b>0.252 0</b>	0.289 3	0.246 7	0.191 1	0.599 0	0.673 9
Li3DeTr						0.614 0	0.676 0
UVTR-L	0.334 0	0.257 0	0.300 0	0.204 0	<b>0.182 0</b>	0.609	0.677 0
Transfusion-L	0.279 7	0.253 7	0.293 2	0.273 2	0.185 4	0.645 7	0.694 3
改进算法	0.276 9	0.252 7	0.338 4	<b>0.196 0</b>	0.186 0	<b>0.651 9</b>	<b>0.701 0</b>

## 2.4 消融实验

为验证所提出算法中改进模块的有效性,在nuScenes数据集上训练并验证各种模型的变体,实验结果如表3所示。其中算法推理速度FPS是在2080Ti显卡上测得。以CenterPoint初始模型为Baseline,随后依次添加动态体素(DynamicVoxel)、编码解码稀疏三维特征提取器(SED-3D-Backbone)、大核二维特征融合主干(LK-2D-Backbone)、IoU分支(IoU-Branch)以及朝向角分类分支(Direction-Cls)在数据集上进行实验。从表中可以发现,在使用动态体素编码方式之后,算法的mAP和NDS分别提升了0.89%和0.36%,这表明动态体素编码方式相较于硬体素编码方式更加高效,能够初步过滤点云中的噪点,同时为三维特征提取器提供充分的点云特征。紧接着,通过引入编码解

码稀疏三维特征提取器,算法的mAP和NDS相较于新Baseline分别提升了3.88%和1.59%,而算法的推理速度仅下降了0.8FPS。相较于原始的三维特征提取器,在平衡算法推理速度的前提下,检测性能有了非常显著的提升。这表明该算法中设计的三维特征提取器能够有效地促进空间非连通区域内特征的信息交互。随后,使用本文提出的大核融合主干替换原始的融合主干,相较于新的Baseline,mAP和NDS分别提升了0.75%和1.13%,算法推理速度仅由12.3FPS下降到11.6FPS,进一步提升了算法的检测效果。这表明所提出的大核融合主干与基于中心点的无锚框检测头更加适配,能够将目标区域的有效点云特征很好地聚集到目标中心点区域。进一步,通过增加IoU分支并用IoU得分修正目标置信度得分,算法的检测性能也取得了一定的提升,mAP和

表3 改进算法在nuScenes验证集上的消融实验

Baseline	DynamicVoxel	SED-3D-Backbone	LK-2D-Backbone	IoU-Branch	Direction-Cls	mAP	NDS	FPS/Hz
√						0.592 2	0.664 8	13.5
√	√					0.601 1	0.668 4	13.1
√	√	√				0.639 9	0.684 3	12.3
√	√	√	√			0.647 4	0.695 6	11.6
√	√	√	√	√		0.649 5	0.699 6	11.4
√	√	√	√	√	√	0.651 9	0.701 0	11.4

NDS较新 Baseline 分别提升了 0.21% 和 0.40%。最后,在检测头处增加朝向角分类分支,进一步提升算法对于目标朝向角预测的准确性,mAP 和 NDS 分别提升了 0.24% 和 0.14%。可以发现,最终算法推理速度相较于 Baseline 仅下降 2.1 FPS,较好地平衡了算法推理速度和检测精度。

## 2.5 实际场景结果

为进一步验证改进后的算法在实际道路场景下的有效性,将改进后的算法部署到自建的智能车平台,如图 8 所示。该平台配备了一个 80 线速腾激光雷达和一个具有 RTX-2080-Ti 显卡的工控机。在实际道路场景采集的点云数据进行算法验证,检测效果如图 9 所示。可以直观地看出,在真实道路下,改

进后的算法在多种交通场景下表现出优异的检测性能,能够准确地检测各类目标。



图 8 智能车实验平台

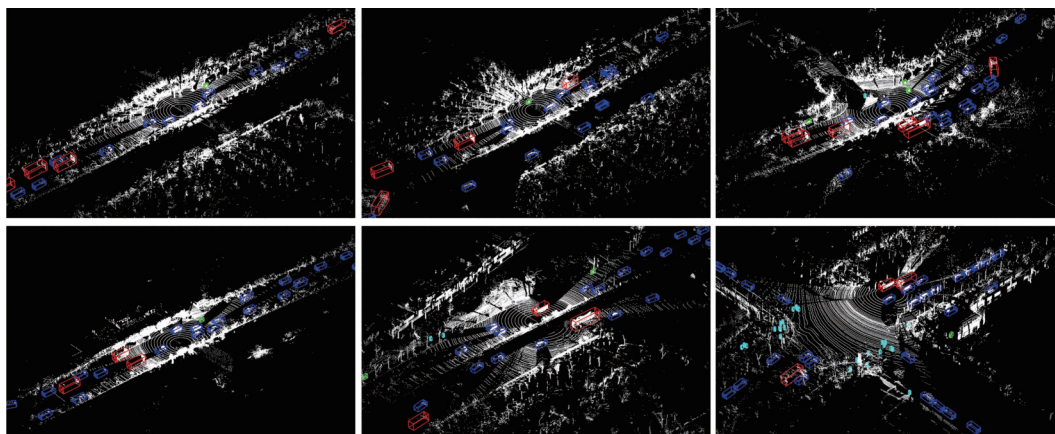


图 9 实际道路场景测试可视化

## 3 结论

以无锚框检测算法 CenterPoint 为基准,提出了一种面向自动驾驶场景的单阶段无锚框点云目标检测算法。具体来说,通过引入编码解码稀疏模块,构建了更加高效的三维特征提取器,保证了三维特征提取器能够提取到满足各类目标检测的特征。此外,在二维特征融合主干中引入的自校正卷积和大核注意力模块,能够有效提取到目标区域的点云特征,并将目标区域的点云特征聚集到中心点区域,从而提升算法对目标的召回率和检测精度。同时,本文还优化体素编码方式和检测头。

所提出的算法在大规模公开数据集 nuScenes 上进行了实验验证,与基准算法相比有了显著的性能提升。最后将所提出的算法在实际道路场景下进行验证,进一步证明了所提出算法的有效性。但是从

实验结果发现,算法对于障碍物的检测效果有所下降,后续将从小目标检测入手,并进一步提升算法的推理速度。

## 参考文献

- [1] 王海,李洋,蔡英凤,等.基于激光雷达的 3D 实时车辆跟踪[J].汽车工程,2021,43(7):1013-1021.  
WANG Hai, LI Yang, CAI Yingfeng, et al. 3D real-time vehicle tracking based on lidar[J]. Automotive Engineering, 2021, 43(7): 1013-1021.
- [2] 娄新雨,王海,蔡英凤,等.采用 64 线激光雷达的实时道路障碍物检测与分类算法的研究[J].汽车工程,2019,41(7):779-784.  
LOU Xinyu, WANG Hai, CAI Yingfeng, et al. A research on an algorithm for real-time detection and classification of road obstacle by using 64-line lidar[J]. Automotive Engineering, 2019, 41(7): 779-784.
- [3] 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,

- 2017: 652–660.
- [4] 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.
- [5] SHI S, WANG X, LI H. PointRCNN: 3D object proposal generation and detection from point cloud[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019: 770–779.
- [6] ZHANG Y, HU Q, XU G, et al. Not all points are equal: learning highly efficient point-based detectors for 3D lidar point clouds [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 18953–18962.
- [7] 龚章鹏, 王国业, 于是. 基于体素网络的道路场景多类目标识别算法[J]. *汽车工程*, 2021, 43(4): 469–477.  
GONG Zhangpeng, WANG Guoye, YU Shi. The algorithm of multi-category object recognition in road scene based on voxel network[J]. *Automotive Engineering*, 2021, 43(4): 469–477.
- [8] 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.
- [9] SHI S, JIANG L, DENG J, et al. PV-RCNN++: point-voxel feature set abstraction with local vector representation for 3D object detection [J]. *International Journal of Computer Vision*, 2023, 131(2): 531–551.
- [10] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 39(6): 1137–1149.
- [11] YIN T, ZHOU X, KRAHENBUHL P. Center-based 3D object detection and tracking[C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021: 11784–11793.
- [12] GE R, DING Z, HU Y, et al. AFDet: anchor free one stage 3D object detection[M]. *arXiv*, 2020.
- [13] DUAN K, BAI S, XIE L, et al. CenterNet: keypoint triplets for object detection [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019: 6569–6578.
- [14] YAN Y, MAO Y, LI B. Second: sparsely embedded convolutional detection[J]. *Sensors*, 2018, 18(10): 3337.
- [15] 夏祥腾, 王大方, 曹江, 等. 基于稀疏卷积神经网络的车载激光雷达点云语义分割方法[J]. *汽车工程*, 2022, 44(1): 26–35.  
XIA Xiangteng, WANG Dafang, CAO Jiang, et al. Semantic segmentation method of on-board lidar point cloud based on sparse convolutional neural network[J]. *Automotive Engineering*, 2022, 44(1): 26–35.
- [16] GRAHAM B, VAN DER MAATEN L. Submanifold sparse convolutional networks[M]. *arXiv*, 2017.
- [17] CHEN Y, LIU J, ZHANG X, et al. LargeKernel 3D: scaling up kernels in 3D sparse CNNs [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 13488–13498.
- [18] MAO J, XUE Y, NIU M, et al. Voxel transformer for 3D object detection [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021: 3164–3173.
- [19] ZHANG G, JUNNAN C, GAO G, et al. HedNet: a hierarchical encoder-decoder network for 3D object detection in point clouds [J]. *Advances in Neural Information Processing Systems*, 2024, 36.
- [20] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 10096–10105.
- [21] CAESAR H, BANKITI V, LANG A H, et al. nuScenes: a multi-modal dataset for autonomous driving [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 11621–11631.
- [22] WANG H, SHI C, SHI S, et al. DSVT: dynamic sparse voxel transformer with rotated sets [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 13520–13529.
- [23] 赵东宇, 赵树恩. 基于级联 YOLOv7 的自动驾驶三维目标检测[J]. *汽车工程*, 2023, 45(7): 1112–1122.  
ZHAO Dongyu, ZHAO Shuen. Autonomous driving 3D object detection based on cascade YOLOv7 [J]. *Automotive Engineering*, 2023, 45(7): 1112–1122.
- [24] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 16794–16805.
- [25] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. *Proceedings of the IEEE International Conference on Computer Vision*, 2017: 2980–2988.
- [26] TEAM O D. OpenPCDet: an open-source toolbox for 3D object detection from point clouds (2020)[Z]. 2020.
- [27] LOSHCHELOV I, HUTTER F. Decoupled weight decay regularization[M]. *arXiv*, 2019.
- [28] ZHU B, JIANG Z, ZHOU X, et al. Class-balanced grouping and sampling for point cloud 3D object detection[M]. *arXiv*, 2019.
- [29] 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.
- [30] CHEN Y, LIU J, ZHANG X, et al. VoxelNeXt: fully sparse VoxelNet for 3D object detection and tracking [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 21674–21683.
- [31] ZHOU Z, ZHAO X, WANG Y, et al. CenterFormer: center-based transformer for 3D object detection [M]//AVIDAN S, BROSTOW G, CISSÉ M, et al. *Computer Vision—ECCV 2022*: Vol. 13698. Cham: Springer Nature Switzerland, 2022: 496–513.