

文章编号:1007-2780(2025)09-1296-12

## 基于多柱特征融合的点云目标检测方法

傅梓寒<sup>1,2</sup>, 李芝贻<sup>3</sup>, 戴 闯<sup>2</sup>, 张丽娟<sup>2\*</sup>

(1. 南京信息工程大学 计算机学院, 江苏 南京 210044;

2. 无锡学院 物联网工程学院, 江苏 无锡 214105;

3. 吉林大学 仪器科学与电气工程, 吉林 长春 130012)

**摘要:**针对现有三维点云目标检测模型在自动驾驶场景中精度与实时性难以兼顾的问题,提出一种高效的点云目标检测方法,旨在提升检测精度,同时满足实时边缘设备部署的需求,以推动自动驾驶和智能交通领域的应用发展。在基于柱体的点云检测框架基础上,设计了一种多柱特征融合模型。模型通过引入点云方差特征和位置编码增强柱体特征表示能力,提出多柱特征融合模块以提升相邻柱体间上下文信息的学习效果,并结合残差连接和Dropout正则化的骨干网络优化特征提取。此外,改进了一种点云数据增强策略,通过混合同类目标点云提升模型对复杂场景的泛化能力。在KITTI数据集上的大量实验表明,本文提出的方法在3D mAP上达到了68.93,在BEV mAP上达到了75.02,模型的推理速度保持在55.65 FPS,同时在DAIR-V2X数据集上也展现出了优秀的性能。所提模型通过多柱特征融合和优化的数据增强策略,实现了检测精度与实时性的平衡,展现出在自动驾驶场景中的应用潜力。未来可进一步优化模型架构,适配嵌入式系统部署,以满足实际应用需求。

**关键词:**3D目标检测;点云;自动驾驶;KITTI;DAIR-V2X

**中图分类号:**TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2025-0096 **CSTR:**32172.14.CJLCD.2025-0096

## Point cloud object detection method based on multi-pillar feature fusion

FU Zhihan<sup>1,2</sup>, LI Zhiyi<sup>3</sup>, DAI Chuang<sup>2</sup>, ZHANG Lijuan<sup>2\*</sup>

(1. School of Computer Science, Nanjing University of Information Science and Technology,  
Nanjing 210044, China;

2. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China;

3. College of Instrument Science and Electrical Engineering, Jilin University, Changchun 130012, China)

**Abstract:** To address the challenge of balancing accuracy and real-time performance in existing 3D point cloud object detection models within autonomous driving scenarios, this paper proposes an efficient point

**收稿日期:**2025-04-25; **修订日期:**2025-07-11.

**基金项目:**国家自然科学基金(No.62206257);无锡市“太湖之光”科技攻关项目(No.K20241044);无锡学院引进人才研究创业基金(No.2023r004, No.2023r006);吉林省科技发展计划重点研发项目(No.20210204050YY)  
Supported by National Natural Science Foundation of China (No.62206257); “Light of the Taihu Lake” Scientific and Technological Research Project for Wuxi Science and Technology Development Fund (No.K20241044); Wuxi University Research Start-up Fund for Introduced Talents (No.2023r004, No.2023r006); Key Research and Development Project of Jilin Province Science and Technology Development Plan (No.20210204050YY)

\*通信联系人, E-mail: zhanglijuan@cwuxu.edu.cn

cloud object detection method. The aim is to improve detection accuracy while meeting the requirements for real-time deployment on edge devices, thereby promoting the application development in autonomous driving and intelligent transportation fields. Based on a pillar-based point cloud detection framework, a multi-pillar feature fusion model is designed. The model enhances the pillar feature representation ability by introducing point cloud variance features and positional encoding. A multi-pillar feature fusion module is proposed to improve the learning of contextual information between adjacent pillars, and a backbone network combining residual connections and Dropout regularization is used to optimize feature extraction. Furthermore, a point cloud data augmentation strategy is improved by mixing point clouds of same-category objects to enhance the model's generalization ability to complex scenes. Extensive experiments on the KITTI dataset demonstrate that the proposed method achieves a 3D mAP of 68.93 and a BEV mAP of 75.02. The model maintains an inference speed of 55.65 FPS. It also demonstrates excellent performance on the DAIR-V2X dataset. The proposed model achieves a balance between detection accuracy and real-time performance through multi-pillar feature fusion and an optimized data augmentation strategy, showcasing its application potential in autonomous driving scenarios. Future work could further optimize the model architecture to adapt to embedded system deployment to meet practical application needs.

**Key words:** 3D object detection; point cloud; autonomous driving; KITTI; DAIR-V2X

## 1 引言

激光雷达是自动驾驶领域最常用的传感器之一,也是机器理解三维场景不可或缺的手段。近年来三维点云物体检测因其在自动驾驶、机器人和智能交通等应用中的重要作用而备受关注<sup>[1]</sup>,3D目标检测模型的性能也随着深度学习技术的进步越来越强。然而,当前许多高精度的3D检测模型往往结构复杂,计算资源消耗大,难以满足实时性和部署效率的双重需求。因此,如何在保障检测精度的同时兼顾部署友好性,成为研究者亟待解决的问题。

当前主流的点云目标检测器大致可以分为3类:基于点的方法、基于体素的方法和基于柱体的方法。基于点的方法(如PointNet<sup>[2]</sup>/PointNet++<sup>[3]</sup>和PointRCNN<sup>[4]</sup>)直接处理原始点云,但其高昂的计算成本限制了其实时应用。基于点的方法需要对每一个点进行处理,并且在全局和局部特征之间进行复杂的聚合操作,因此它们往往存在计算成本过高的问题。基于体素的方法(如VoxelNet<sup>[5]</sup>、SECOND<sup>[6]</sup>)通过将点云划分为三维网格来提升效率,但其依赖的稀疏卷积操作在嵌入式设备上的部署仍存在挑战<sup>[7]</sup>。基于柱体的检测框架具有结构简洁、推理速度快等优势,尤其适合部署于资源受限的边缘设备,

PointPillars<sup>[8]</sup>是这一类的典型代表。该方法通过将点云数据划分为数个柱体,压缩了点云空间的高度,避免了对稠密三维结构的处理,仅需使用高效的二维卷积即可完成特征提取与目标检测。然而,柱体化过程虽提高了效率,却也带来了特征表达能力下降的问题。Geiger等人<sup>[9]</sup>通过知识蒸馏和定位引导分类项来改进PointPillars的检测性能。除了上述几个先驱方法,近年来也不乏一些精度非常高的模型<sup>[10-13]</sup>被提出,但它们的模型结构通常较复杂,受制于当前的硬件水平,难以部署在实际应用之中。这种现状引出了一个关键且具有实际应用价值的研究问题:如何在显著牺牲PointPillars框架所带来的部署友好性和高效率的前提下,针对性地弥补其因柱体化编码导致的特征表达能力不足,从而大幅提升其检测精度。

为了解决上述方法的局限性,本文提出了一种多柱特征融合点云目标检测模型,旨在实现精度与实时性的平衡。模型首先计算点云的方差特征和位置编码<sup>[14]</sup>增强柱体特征编码器,丰富柱体内的空间与统计信息表达。由于体素/柱体的尺寸通常小于目标大小,对点云的体素化会削弱模型捕捉目标整体结构和局部细节的能力,从而影响对该物体的检测效果。因此,针对柱体化削弱了模型捕捉目标整体结构和局部细节能力的核心痛点,本文设计了一个多柱特征融合模块。该模块

旨在提升模型学习相邻柱体间上下文信息的能力,从而生成更稳健的特征表示。在网络设计方面,引入残差连接与 Dropout 正则化以提升模型的泛化能力与稳定性。最后,结合轻量级的 SSD (Single Shot Detector)<sup>[15]</sup> 预测头,实现高效推理。

除了网络本身之外,数据增强是提高检测成功率的重要手段之一。在点云目标检测任务中,有效的数据增强对于提升模型的泛化能力和鲁棒性至关重要。除了常规的全局旋转、缩放等增强手段,一些研究还提出了更精细化的方法以生成高质量的训练样本。本研究采用 Part-Aware Data Augmentation (PA-AUG)<sup>[16]</sup> 作为数据增强策略的研究基础。其主要优势在于,它并非对整个目标进行统一变换,而是利用了“目标内部部件位置 (intra-object part location)”这一先验信息,对目

标不同分区施以增强。这种分区感知的机制能够生成结构上更合理、更多样化的增强样本,尤其适用于处理具有复杂结构的目标。

通过上述一系列有针对性的改进,本文所提出的模型在保持高推理速度的同时,在 KITTI 和 DAIR-V2X 数据集上均取得了具有竞争力的检测精度,证明了所提方法的有效性。

## 2 基于柱体点云检测器的网络结构

基于柱体的方法是一种在点云目标检测中常见的技术,特别是在处理稠密的点云数据时,它能够有效地提高计算效率。PointPillars 是基于柱体方法中的开创性模型。本文提出的方法采用了与 PointPillars 相似的架构。图 1 展示了本文方法的框架图。

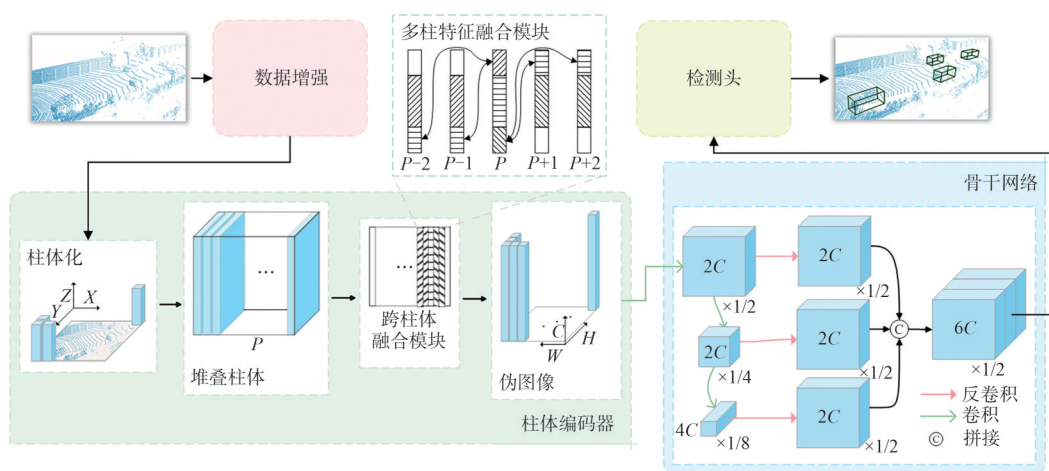


图 1 本文提出的网络框图

Fig. 1 Network architecture proposed in this paper

模型的输入为原始的 3D 点云数据。点云数据包含了每个点的空间坐标  $(x, y, z)$  和反射强度等特征。在预处理过程中,首先通过柱体化将点云空间从鸟瞰视角划分为固定大小的柱状网格,每个网格内包含若干个点。这一步骤的目的是通过将点云划分为规则的、较小的块来减少计算复杂度,并使得每个块内点的数量相对较少,方便后续的特征提取和处理。

在点云数据被划分为柱体之后,接下来是提取每个柱体内点的特征。对于 PointPillars,模型会聚合该点的三维坐标、反射强度以及相对坐标

等特征。为了更好地表示每个柱体的特征,模型采用线性层对这些特征进行进一步编码,将每个柱体的低维特征向量映射到一个更高维的空间,同时使得特征向量被映射到一个固定的长度,以便输入后续网络。至此,原始点云的稀疏表示被转化为更加密集和信息丰富的表示。

经过特征编码后,PointPillars 使用 2D 卷积神经网络 (CNN) 作为骨干网络对所有柱体特征进行进一步的处理。模型将点云转化为一个 2D 伪图像的特征图,从而能够使用高效的 2D 卷积操作进行特征提取。使用 2D 卷积对 3D 点云进

行特征提取令基于柱体的方法具有极大的速度优势。骨干网络通常由多个卷积层和池化层组成,用于提取多尺度的特征。为了增强网络对不同尺度目标的检测能力,通常会引入特征金字塔网络。特征金字塔通过自上而下和横向连接的方式融合不同层级的特征,生成具有丰富语义信息的多尺度特征图。输出的特征图将作为检测头的输入,用于目标检测任务。

检测头预测目标的位置、尺寸和方向。检测头通常包括3个分支:分类分支、回归分支和方向分类分支。分类分支用于预测每个锚框的类别概率,回归分支用于预测每个锚框的中心点偏移、尺寸和方向角,而方向分类分支则用于预测目标的方向。最后通过非极大值抑制去除重叠的检测框,保留置信度最高的检测结果。

### 3 多柱融合点云目标检测的关键算法

本模型采用基于柱体的方法,网络架构如图1所示。整体可以分为4个阶段。首先对输入的点云进行数据增强。这将显著提升模型的泛化能力,尤其是样本量较少的行人和骑行者类别。接下来将点云输入到柱体编码器中,在此,点云首先被划分为一系列柱体。为了高效处理稀疏的点云,所有空的柱都将被舍弃。将余下的柱体堆叠、编码并通过线性层整理成 $P$ 个具有 $C$ 通道数的序列。最后将柱体序列重新映射回初始位置,得到形状为 $[H, W, C]$ 的伪图像。至此,稀疏的点云数据在柱体编码器中转换为容易进行高效特征提取的伪图像。正如多数卷积神经网络那样,将伪图像输入到骨干网络进行进一步的特征提取。本模型的骨干网络设计为3层结构,特征图在这里变换为不同尺度,以处理不同大小和难度的目标。空间特征通过多个改进后的残差块和上采样块进行处理,最终拼接以生成更高维的特征图。最后,检测头部分采用SSD<sup>[15]</sup>预测物体类别、边界框和方向。

#### 3.1 编码阶段

##### 3.1.1 柱体编码器

在模型的这一部分,点云被划分为一系列柱体,然后对柱体进行编码,将不规则的点云数据

转换为规则的网格结构,以便于输入到之后的骨干网络中进行处理。

给定一个点云,将其划分为 $P$ 个大小相同的柱体,每个柱体所包含的点的数量最多为 $M$ 。对于第 $p$ 个柱体中的第 $m$ 个点,其三维坐标为 $\mathbf{v}_{p,m} = (x_{p,m}, y_{p,m}, z_{p,m})$ ,其中 $p = 1, \dots, P, m = 1, \dots, M$ 。不失一般性,下文将对其中一个柱体 $p$ 进行详细的描述。

首先计算柱体中所有点的三维坐标均值 $\mu_p = (\mu_x, \mu_y, \mu_z)$ :

$$\mu_p = \frac{1}{M} \sum_{m=1}^M \mathbf{v}_{p,m}, \quad (1)$$

其中, $\mu_x, \mu_y, \mu_z$ 分别是柱体 $p$ 中点的三维坐标 $x, y, z$ 的均值。

接下来计算 $p$ 内点的三维坐标方差。记 $\sigma_p^2 = (\sigma_x^2, \sigma_y^2, \sigma_z^2)$ 为柱体 $p$ 中 $x, y, z$ 坐标的方差,根据公式(1)得到方差如下:

$$\sigma_p^2 = \frac{1}{M} \sum_{m=1}^M (\mathbf{v}_{p,m} - \mu_p)^2, \quad (2)$$

其中, $\sigma_x^2, \sigma_y^2, \sigma_z^2$ 分别表示 $x, y, z$ 坐标的方差。

位置编码旨在将点的坐标映射到一个更高维的特征空间,这样就可以保留因为柱体化而损失的高度信息。为了实现这一点,假设位置编码后的维度为 $N$ ,且 $N$ 需要是3的倍数,因为通常都为每个坐标轴分配相同数量的特征。对于每个坐标分量 $v \in \{x_{p,m}, y_{p,m}, z_{p,m}\}$ ,其编码公式如式(3)、(4)所示:

$$\mathbf{PE}(v) = [\sin(v \cdot 10000^{\frac{2k}{K}}), \cos(v \cdot 10000^{\frac{2k}{K}})], \quad (3)$$

$$\mathbf{PE}_{p,m} = [\mathbf{PE}(x_{p,m}), \mathbf{PE}(y_{p,m}), \mathbf{PE}(z_{p,m})], \quad (4)$$

公式(3)中 $k = 0, 1, \dots, \frac{K}{2} - 1$ ,编码点的坐标通过交替使用正弦和余弦函数来得到。将 $x_{p,m}, y_{p,m}, z_{p,m}$ 的编码向量按轴向拼接,得到点 $\mathbf{v}_{p,m}$ 的位置编码向量 $\mathbf{PE}_{p,m}$ 。最终,得到如公式(4)所展示的位置编码后的特征:一个形状为 $[P, M, N]$ 的张量。

##### 3.1.2 多柱特征融合模块

点云特征在前面的网络中通过线性层变成统一的64维。本方法在此基础上对相邻柱体的特征进行拆分和选择性融合,以提高检测性能。首先将每个柱体的特征向量 $F_p \in \mathbb{R}^{64}$ 划分为4个

16 维的子特征:

$$F_p = [F_p^{(0:16)}, F_p^{(16:32)}, F_p^{(32:48)}, F_p^{(48:64)}], \quad (5)$$

其中: $p$ 表示柱体的索引, $F_p^{(i:j)}$ 表示柱体特征向量的第*i*到*j*-1维。

该模块通过融合相邻柱的特征增强每个柱体的表示能力。融合的过程如式(6)、(7)所示:

$$F_p^{(0:16)} = \frac{\omega_1 \cdot F_{p-2}^{(0:16)} + \omega_2 \cdot F_{p-1}^{(0:16)}}{\omega_1 + \omega_2}, \quad (6)$$

$$F_p^{(48:64)} = \frac{\omega_1 \cdot F_{p+2}^{(48:64)} + \omega_2 \cdot F_{p+1}^{(48:64)}}{\omega_1 + \omega_2}, \quad (7)$$

其中, $\omega_1$ 和 $\omega_2$ 表示来自不同柱体特征的权重值。具体来说,将当前柱体*p*的第16:48维特征保留,同时对其余特征维度进行相邻特征融合。为了获取更多相邻柱体特征,在本文中,权重0.4和0.6为经验选择,旨在平衡远近柱体的贡献,并略微偏重较近的柱体,反映了局部相关性通常随距离衰减的假设。

为处理序列边界的情况,模块制定了特定的边界规则。对于前两个柱体 $p=0$ 和 $p=1$ ,由于缺乏足够的前驱柱体,其前部子特征 $F_p^{(0:16)}$ 保持不变;对于后两个柱体 $p=P-1$ 和 $p=P-2$ ,由于缺乏足够后继柱体,后部子特征 $F_p^{(48:64)}$ 保持原值。这种边界处理确保了所有柱都能获得有效的特征表示。

通过上述设计,该多柱特征融合模块在保留柱固有特性的同时,利用相邻柱的空间上下文信息增强特征表示能力。固定权重简化了计算复杂度,而从两侧邻居中获取特征则扩展了模型的感知范围。这种方法能够提升点云目标检测模型的鲁棒性和性能,尤其在处理复杂场景时表现出优势。

### 3.2 骨干网络

针对点云数据经过体素化和柱状特征编码后形成的二维伪图像,本文设计了一种基于残差结构并融入Dropout正则化的二维骨干网络。该网络以残差学习为核心,结合下采样、上采样和Dropout正则化机制,旨在有效提取点云的空间特征并增强模型的泛化能力。该设计充分考虑了点云数据的稀疏性和复杂性,通过合理的层次结构和参数配置实现高效的特征处理与融合。

网络的整体架构由多个残差模块和上采样模块组成。输入的空间特征首先经过一系列残差块处理,其结构如图2所示。基础块采用经典

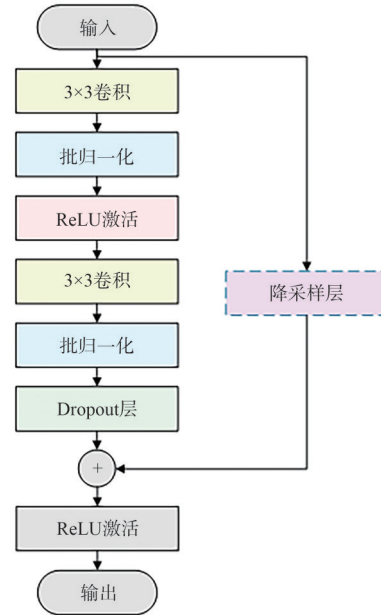


图2 骨干网络基础块的架构

Fig. 2 Structure of backbone network basic block

的残差结构,包含两层 $3 \times 3$ 卷积,每层卷积后接批量归一化和ReLU激活函数。为增强模型的正则化能力,在第二个卷积层后引入了Dropout层,以缓解过拟合风险并提升模型在复杂场景中的鲁棒性。此外,当输入与输出通道数或分辨率不同时,残差连接通过 $1 \times 1$ 卷积和批量归一化进行下采样调整,确保特征维度匹配。

在特征提取后,网络通过上采样模块对不同尺度的特征图进行融合,以生成统一的二维特征表示。具体而言,每个阶段的输出特征图通过转置卷积操作(步幅分别为1、2、4)上采样至一致的 $\frac{H}{4} \times \frac{W}{4}$ 分辨率,并调整通道数至 $2C$ ,最终通过通道拼接生成多尺度特征融合结果。当需要进一步调整分辨率时,网络末端附加一个转置卷积层,确保输出特征图满足后续检测头的需求。最终输出的特征图通道数为上采样通道数之和,为后续的目标检测任务提供了丰富的空间信息。

### 3.3 数据增强

本研究引入并改进了一种高效的基于点云数据的数据增强方法Part-Aware Data Augmentation(PA-AUG)<sup>[16]</sup>。Shi等<sup>[17]</sup>认为点云数据中的前景点包含了有价值的相对位置信息,并将这些前景点相对于其3D框的相对位置称为“intra-object part location”。PA-AUG的核心思想就是利

用这些信息,将目标划分为多个分区,并对每个分区随机应用CutMix、CutMixup、内部点丢弃、稀疏采样和随机噪声生成这5种数据增强方法。通过混合应用这些不同的子方法,能够在训练中生成更加丰富的数据样本。然而,随机噪声在某些低密度点云场景下可能会引入无效特征,从而影响模型的训练效果。因此,本文提出了一种改进策略,用PointMixup<sup>[18]</sup>取代随机噪声生成的方法,以更真实地模拟物体的形变。该方法通过混合来自同一类别的不同目标物体的点来增强训练集,同时保留了物体结构和空间之间的关系。假设数据集中的目标区域为 $B_i$ ,其中 $i=1, \dots, N$ ,表示共有 $N$ 个目标。每个目标被划分为 $K_i$ 个非重叠的区域,可以记为 $B_i = \bigcup_{j=1}^{K_i} P_{i,j}$ ,  $P_{i,j} \cap P_{i,k} = \emptyset (j \neq k)$ ,其中 $P_{i,j}$ 表示目标 $B_i$ 中的第 $j$ 个区域。首先,需要一个来自Beta分布的混合系数 $\lambda \sim \text{Beta}(\alpha, \beta)$ ,其中 $\alpha, \beta > 0$ , $\lambda$ 用于控制两个前景目标在点云和边界框参数混合中的权重。接下来需要对同属于一个类别的目标 $B_i$ 和 $B_{i'}$ 做混合。令 $P_{i,j}$ 和 $P_{i',j}$ 分别表示两个同类别目标的第 $j$ 个区域中的点集。对于每个非空区域,选择两个区域的前 $n = \min(|P_{i,j}|, |P_{i',j}|)$ 个点进行混合:

$$P_{i,j}^{\text{mixed}} = \lambda \cdot P_{i,j}[:n] + (1 - \lambda) \cdot P_{i',j}[:n], \quad (8)$$

其中, $P_{i,j}^{\text{mixed}}$ 是混合后的点集。类似地,目标的边界框参数也需要被混合:

$$b_i^{\text{mixed}} = \lambda \cdot b_i + (1 - \lambda) \cdot b_{i'}, \quad (9)$$

其中, $b_i = (c_i, d_i, r_i)$ 表示中心、尺度和旋转角度等边界框参数。

## 4 实验与结果

### 4.1 实验设置

本文在KITTI数据集<sup>[19]</sup>和DAIR-V2X数据集<sup>[20]</sup>上对所提方法进行了评估。KITTI数据集是自动驾驶领域中最广泛使用的基准之一,包含7 481个训练样本和7 518个测试样本,每个样本包括RGB图像及对应的点云数据。目标物体根据尺寸和遮挡程度被划分为简单、中等、困难3个难度级别。遵循惯例<sup>[21]</sup>,训练样本被划分为训练集(3 712个样本)和验证集(3 769个样本)。DAIR-V2X数据集是清华大学智能产业研究院发布的

大规模多视角数据集,其包含了协同、车端和路端3个子数据集,具有比KITTI数据集更广泛的场景和天气条件。本研究选取了具有22 325帧点云数据的车端数据集进行实验,根据原论文将数据按照5:2:3的比例划分训练集、验证集和测试集。所有模型在训练集上进行训练,并采用更精确的40个召回位置<sup>[22]</sup>进行测试。

所提方法基于OpenPCDet<sup>[23]</sup>代码库实现。在位置编码中,编码特征维度决定每个点生成编码向量的长度。较高的特征维度能够捕捉更丰富的位置信息,但同时增加了计算开销和内存占用。本方法将编码特征维度设置为96。在柱体编码阶段,柱体权重 $\omega_1$ 和 $\omega_2$ 分别设置为0.4和0.6。骨干网络采用3阶段结构,各阶段分别堆叠了1层、2层和2层卷积操作,每个阶段采用步幅为2的卷积以实现空间下采样。对应地,3个阶段的输出特征图通道数依次为128、128和256。为了融合不同尺度的特征信息,网络在每个阶段输出后分别进行上采样操作,上采样步幅分别为1、2和4,并将各阶段特征图统一调整至128个通道。Dropout丢弃率设置为0.2。最终,通过特征融合,形成高分辨率、高语义信息量的输出特征图,为后续的目标检测等任务提供丰富的空间与语义表达能力。在检测头设计上,沿用PointPillars的SSD框架<sup>[15]</sup>,直接从融合特征图预测三维边界框和类别得分。损失函数设置与PointPillars一致。对于PointMixup相关参数,设置如下:控制混合比例 $\lambda$ 的Beta分布参数 $\alpha$ 和 $\beta$ 均设为1,表示两目标点云均匀混合;应用PointMixup增强的概率 $p$ 设为0.5,即半数样本应用该增强方法;设定距离限制为100,以避免增强远距离目标引入的不确定性。其他未提及参数均采用默认设置。模型训练与测试的平台为单块NVIDIA RTX A6000 48G GPU和Intel Xeon w7-2495X。

### 4.2 对比实验

对比实验所用到的方法同样使用OpenPCDet<sup>[23]</sup>代码库复现。根据KITTI评估的标准,将3个类别的IoU分别设置成0.7、0.5和0.5测试所有的模型。如表1所示,所提出的方法设置了0.16和0.12两种柱体尺寸,将它们与其他方法进行比较。当柱体尺寸设置为0.12时,所提方法在平均均值精度(mAP)方面达到了最高的

表 1 不同算法在 KITTI 数据集中的 3D 检测精度对比

Tab. 1 Comparison of 3D detection accuracy of different methods on KITTI dataset

| 算法                           | 模态      | 速度/<br>FPS   | 总体<br>mAP/%  | 汽车 AP/%      |              |              | 行人 AP/%      |              |              | 骑行者 AP/%     |              |              |
|------------------------------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                              |         |              |              | 简单           | 中等           | 困难           | 简单           | 中等           | 困难           | 简单           | 中等           | 困难           |
| MVX-Net <sup>[24]</sup>      | 点云&彩色图像 | 27.10        | 63.57        | 89.13        | 78.88        | 74.37        | 55.97        | 51.38        | 47.29        | 71.06        | 53.75        | 50.28        |
| F-PointNet <sup>[25]</sup>   | 点云&彩色图像 | 5.90         | 62.32        | 82.89        | 68.94        | 62.16        | <b>67.47</b> | <b>57.50</b> | <b>50.52</b> | 69.81        | 52.67        | 48.90        |
| 3DSSD <sup>[26]</sup>        | 点云      | 19.40        | —            | 88.82        | 78.58        | 77.47        | —            | —            | —            | —            | —            | —            |
| SECOND <sup>[6]</sup>        | 点云      | 50.92        | 65.48        | 88.16        | 78.74        | 75.66        | 51.61        | 46.53        | 41.50        | 81.98        | 64.69        | 60.42        |
| MVF <sup>[27]</sup>          | 点云      | 48.60        | 66.02        | 87.57        | 76.77        | 74.23        | 58.76        | 53.48        | 48.02        | 75.15        | 62.05        | 58.13        |
| PillarNet <sup>[28]</sup>    | 点云      | 31.70        | 65.51        | 89.62        | 80.55        | 77.76        | 50.21        | 45.05        | 42.27        | 83.36        | 62.39        | 58.35        |
| SeSame-voxel <sup>[29]</sup> | 点云      | —            | 67.89        | <b>90.58</b> | <b>81.45</b> | <b>78.33</b> | 53.90        | 48.47        | 44.14        | <b>84.01</b> | 66.99        | 63.10        |
| CenterPoint <sup>[30]</sup>  | 点云      | 25.63        | 66.29        | 87.68        | 79.32        | 76.61        | 54.44        | 50.76        | 46.31        | 80.24        | 62.41        | 58.82        |
| VoxelNeXt <sup>[31]</sup>    | 点云      | 40.95        | 67.16        | 86.45        | 77.55        | 75.00        | 59.65        | 53.87        | 48.77        | 80.72        | 63.10        | 59.36        |
| PointPillars <sup>[8]</sup>  | 点云      | <b>82.78</b> | 63.11        | 87.53        | 76.35        | 73.20        | 50.39        | 43.80        | 39.41        | 80.05        | 60.71        | 56.55        |
| Ours (0.16 m)                | 点云      | 61.20        | 65.78        | 88.64        | 78.86        | 75.84        | 54.64        | 48.02        | 43.25        | 78.80        | 63.51        | 60.44        |
| Ours (0.12 m)                | 点云      | 55.65        | <b>68.93</b> | 88.09        | 78.79        | 75.97        | 59.73        | 52.82        | 47.54        | 82.76        | <b>69.46</b> | <b>65.23</b> |

68.93%, 优于 VoxelNeXt<sup>[31]</sup>、CenterPoint<sup>[30]</sup>等主流方法,展示了其综合精确度上的优势。值得注意的是,当使用与 PointPillars<sup>[8]</sup>默认的 0.16 的柱体尺寸时,所提方法同样展示出了很有竞争力的性能。表 1~表 6 中,粗体数字代表最优值。

在汽车类别中,PillarNet<sup>[28]</sup>、CenterPoint<sup>[30]</sup>和 MVX-Net<sup>[24]</sup>等方法表现优异,这可能与它们设计的特点有关。例如,CenterPoint 利用了基于中心的检测方式,这种方法特别适合对称的、规则的目标。虽然所提方法未能在该类别上获得最佳表现,但仍保持了稳定且具有一定竞争力的性能(88.09%,78.79%,75.97%)。在行人类别中,所提方法在简单难度下仅落后于利用 LiDAR 和

RGB 两种模态的 F-PointNet<sup>[25]</sup>,而在中等和困难难度下领先于同样使用基于柱体的方法 PillarNet<sup>[28]</sup>。在骑行者类别中,所提方法相较于其他方法展现出显著的优势,在中等和困难难度下仍保持了 5% 以上的领先。骑行者作为一个复合目标(骑车人+自行车)具有形状多变、结构丰富等特点,体现出所提方法具有更加精细的空间分辨率和更强的局部上下文信息获取能力。此外,在保持了优秀整体性能的同时,所提模型在检测速度方面的表现同样出色,达到了仅次于 PointPillars 的 61.27 FPS 和 55.65 FPS,证明了其在实时设备上部署的可能性。

表 2 展示了模型在鸟瞰视角(BEV)下的性

表 2 不同算法在 KITTI 数据集中的鸟瞰视角(BEV)精度对比

Tab. 2 Comparison of bird's eye view (BEV) accuracy of different methods on KITTI dataset

| 算法                          | 模态      | 速度/<br>FPS   | 总体<br>mAP/%  | 汽车 AP/%      |              |              | 行人 AP/%      |              |              | 骑行者 AP/%     |              |              |
|-----------------------------|---------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|                             |         |              |              | 简单           | 中等           | 困难           | 简单           | 中等           | 困难           | 简单           | 中等           | 困难           |
| MVX-Net <sup>[24]</sup>     | 点云&彩色图像 | 27.10        | 70.72        | 92.83        | 88.17        | 85.68        | 65.29        | 60.22        | 55.88        | 75.99        | 57.84        | 54.59        |
| F-PointNet <sup>[25]</sup>  | 点云&彩色图像 | 5.90         | 68.52        | 87.51        | 81.97        | 73.55        | <b>70.82</b> | <b>61.49</b> | <b>56.48</b> | 74.54        | 56.90        | 53.38        |
| MVF <sup>[27]</sup>         | 点云      | 48.60        | 71.66        | 90.26        | 86.77        | 85.61        | 63.00        | 58.25        | 51.89        | 79.26        | 66.91        | 63.01        |
| SECOND <sup>[6]</sup>       | 点云      | 50.92        | 71.53        | 91.99        | 87.86        | 86.53        | 55.78        | 51.36        | 46.93        | 85.95        | 70.86        | 66.55        |
| PillarNet <sup>[28]</sup>   | 点云      | 31.70        | 71.56        | <b>93.87</b> | 87.74        | 86.63        | 57.56        | 52.43        | 49.39        | 86.28        | 67.18        | 63.00        |
| CenterPoint <sup>[30]</sup> | 点云      | 25.63        | 71.85        | 91.32        | 88.26        | 86.24        | 60.51        | 56.83        | 53.02        | 83.15        | 65.43        | 61.84        |
| VoxelNeXt <sup>[31]</sup>   | 点云      | 40.95        | 73.74        | 91.73        | 87.45        | 85.01        | 65.27        | 59.69        | 54.67        | 86.83        | 68.27        | 64.76        |
| PointRCNN <sup>[4]</sup>    | 点云      | 12.22        | 74.87        | 89.14        | <b>88.60</b> | 86.49        | 65.51        | 58.34        | 51.68        | <b>87.89</b> | 73.90        | <b>72.28</b> |
| PointPillars <sup>[8]</sup> | 点云      | <b>82.78</b> | 69.14        | 92.42        | 87.86        | 85.19        | 55.98        | 49.62        | 45.11        | 82.77        | 63.77        | 59.53        |
| Ours (0.12 m)               | 点云      | 55.65        | <b>75.02</b> | 91.62        | 87.89        | <b>86.89</b> | 65.05        | 58.02        | 53.30        | 86.50        | <b>75.12</b> | 70.77        |

能表现。所提方法在总体 mAP 上超过了两阶段网络 PointRCNN<sup>[4]</sup>,且在各难度级别的 BEV 精度上都实现了优秀的性能,检测速度仅落后于 PointPillars。总的来说,所提方法实现了优秀的速度和精度平衡。

为了进一步检验所提方法的泛化能力和在不同场景下的鲁棒性,我们还在大规模车路协同数据集 DAIR-V2X 上进行了补充对比实验,结果如表 3 和表 4 所示。从表中数据可以看出,本文所

提出的模型在该数据集上取得了 48.73% 的总体 mAP 和 47.25% 的中等难度 mAP,同时在鸟瞰图视角下的性能也保持了相似的竞争力,综合性能优于 SECOND、PillarNet 和 PointRCNN 等多种主流检测方法,并且显著优于基线模型 PointPillars,展现出良好的跨数据集泛化表现。所提方法虽然在精度上略微落后于顶尖方法 CenterPoint 和 VoxelNeXt,但简单的架构设计使其更具有边缘设备可部署性。

表 3 不同算法在 DAIR-V2X 数据集中的 3D 检测精度对比

Tab. 3 Comparison of 3D detection accuracy of different methods on DAIR-V2X dataset

| 算法                          | 总体<br>mAP/% | mAP<br>(中等)/% | 汽车 AP/% |       |       | 行人 AP/% |       |       | 骑行者 AP/% |       |       |
|-----------------------------|-------------|---------------|---------|-------|-------|---------|-------|-------|----------|-------|-------|
|                             |             |               | 简单      | 中等    | 困难    | 简单      | 中等    | 困难    | 简单       | 中等    | 困难    |
| MVF <sup>[27]</sup>         | 42.56       | 40.83         | 62.96   | 50.49 | 47.56 | 41.15   | 37.62 | 36.61 | 39.89    | 34.37 | 32.42 |
| SECOND <sup>[6]</sup>       | 45.72       | 44.46         | 68.66   | 56.87 | 53.89 | 38.74   | 36.81 | 36.12 | 42.62    | 39.70 | 38.04 |
| PillarNet <sup>[28]</sup>   | 44.58       | 43.20         | 68.90   | 57.20 | 54.01 | 37.89   | 35.57 | 35.10 | 41.04    | 36.82 | 34.66 |
| CenterPoint <sup>[30]</sup> | 50.91       | 49.81         | 71.13   | 58.93 | 55.75 | 48.60   | 45.36 | 44.63 | 44.46    | 45.15 | 44.21 |
| VoxelNeXt <sup>[31]</sup>   | 52.64       | 51.41         | 70.00   | 58.07 | 54.88 | 52.84   | 48.76 | 47.51 | 47.78    | 47.41 | 46.49 |
| PointRCNN <sup>[4]</sup>    | 43.65       | 41.54         | 71.37   | 57.54 | 54.30 | 38.93   | 32.77 | 31.51 | 39.64    | 34.30 | 32.52 |
| PointPillars <sup>[8]</sup> | 39.32       | 37.95         | 67.00   | 55.09 | 51.86 | 31.97   | 30.51 | 30.19 | 31.99    | 28.25 | 27.03 |
| Ours                        | 48.73       | 47.25         | 69.23   | 57.49 | 54.63 | 44.34   | 41.49 | 40.46 | 47.11    | 42.76 | 41.02 |

表 4 不同算法在 DAIR-V2X 数据集中的鸟瞰视角(BEV)精度对比

Tab. 4 Comparison of bird's eye view (BEV) accuracy of different methods on DAIR-V2X dataset

| 算法                          | 总体<br>mAP/% | mAP<br>(中等)/% | 汽车 AP/% |       |       | 行人 AP/% |       |       | 骑行者 AP/% |       |       |
|-----------------------------|-------------|---------------|---------|-------|-------|---------|-------|-------|----------|-------|-------|
|                             |             |               | 简单      | 中等    | 困难    | 简单      | 中等    | 困难    | 简单       | 中等    | 困难    |
| MVF <sup>[27]</sup>         | 44.68       | 43.03         | 65.49   | 53.17 | 50.46 | 42.95   | 39.74 | 39.15 | 40.78    | 36.19 | 34.14 |
| SECOND <sup>[6]</sup>       | 48.02       | 46.88         | 71.43   | 59.77 | 57.21 | 39.90   | 38.68 | 38.25 | 43.91    | 42.19 | 40.86 |
| PillarNet <sup>[28]</sup>   | 47.01       | 45.77         | 71.78   | 60.10 | 57.43 | 39.41   | 37.91 | 37.74 | 42.05    | 39.30 | 37.35 |
| CenterPoint <sup>[30]</sup> | 51.85       | 50.76         | 73.46   | 61.41 | 58.82 | 45.14   | 45.43 | 45.00 | 47.15    | 45.44 | 44.78 |
| VoxelNeXt <sup>[31]</sup>   | 55.87       | 54.81         | 73.03   | 61.07 | 58.36 | 55.22   | 52.02 | 51.03 | 50.42    | 51.35 | 50.34 |
| PointRCNN <sup>[4]</sup>    | 44.01       | 42.32         | 72.33   | 60.06 | 55.11 | 38.48   | 33.31 | 32.21 | 38.95    | 33.59 | 32.04 |
| PointPillars <sup>[8]</sup> | 41.98       | 40.97         | 71.21   | 59.81 | 57.15 | 34.16   | 33.00 | 30.19 | 33.37    | 30.10 | 28.87 |
| Ours                        | 50.92       | 49.57         | 71.85   | 60.20 | 57.61 | 46.00   | 43.31 | 42.42 | 48.09    | 45.20 | 43.60 |

### 4.3 消融实验

通过在基线模型 PointPillars<sup>[8]</sup>上系统地集成各模块,本文进行了全面的消融实验,以评估所提出各组件的有效性。实验结果如表 5 所示,展示了各模块在中等难度下对 3 类目标检测性能的影响,同时分析了对整体平均精度(mAP)、推理速度的影响。表 5 中 MF 表示引入的更多特征,XB 表示多柱融合模块,BB 表示骨干网络,DA 表示数据增强;Car 表示汽车类别,Ped. 表示行人类别,Cyc. 表示骑行者类别。

#### 4.3.1 柱体编码器的影响

首先重点评估了柱体编码阶段的改进对整体性能的贡献。将更多特征引入基线模型后,检测性能有了明显提升,mAP 提高至 60.2%。其中,汽车类别的检测精度提升最为显著,AP 达到 77.8%,表明在编码阶段引入的位置编码等手工特征有效丰富了特征表征能力。然而,在行人和骑行者类别上的性能提升较为有限,AP 分别为 42.1% 和 60.7%,说明引入的特征虽然在捕捉较大物体(如汽车)的判别性特征方面具有优势,但

表 5 各模块在 KITTI 数据集上的消融实验

Tab. 5 Ablation study on KITTI dataset

| 算法    |                             | 中等难度 AP/% |    |             | mAP/        | 速度/         |             |      |
|-------|-----------------------------|-----------|----|-------------|-------------|-------------|-------------|------|
| 柱体编码器 |                             | BB        | DA | Car         | Ped.        | Cyc.        | %           | FPS  |
| MF    | XB                          |           |    |             |             |             |             |      |
|       | PointPillars <sup>[8]</sup> |           |    | 76.3        | 43.5        | 59.1        | 59.6        | 87.4 |
| ✓     |                             |           |    | 77.8        | 42.1        | 60.7        | 60.2        | 70.8 |
|       | ✓                           |           |    | 76.4        | 44.7        | 62.1        | 61.1        | 85.8 |
|       |                             | ✓         |    | 76.2        | 43.8        | 63.4        | 61.2        | 73.7 |
|       |                             |           | ✓  | 76.3        | 46.5        | 62.1        | 61.6        | 88.3 |
| ✓     | ✓                           | ✓         |    | 78.4        | 42.8        | <b>63.7</b> | 61.6        | 61.8 |
| ✓     | ✓                           |           | ✓  | 76.7        | 47.5        | 61.6        | 61.9        | 68.7 |
|       |                             | ✓         | ✓  | 76.7        | 46.2        | 66.4        | 63.1        | 73.0 |
| ✓     | ✓                           | ✓         | ✓  | <b>78.9</b> | <b>48.0</b> | 63.5        | <b>63.5</b> | 65.5 |

对体积较小或结构更为复杂的目标提升作用较小。此外,引入新特征后,模型的推理速度小幅下降至 70.8 FPS,反映出处理额外特征所带来的计算开销。

单独引入多柱融合模块同样在基线上带来了性能提升,mAP 达到 61.1%。值得注意的是,该配置在所有类别上均提升了检测性能,其中汽车、行人和骑行者的 AP 分别为 76.4%、44.7% 和 62.1%。尤其在行人和骑行者类别上,AP 相较基线分别提升了 1.2% 和 3.0%,主要归因于其能够通过整合相邻柱体的特征,有效缓解柱状化过程中局部结构信息的损失。通过加强特征交互,该模块提升了模型捕捉物体整体结构及细粒度细节的能力,特别适用于行人和骑行者等小型且结构复杂的目标。此外,该模块设计简洁,计算资源开销极小,推理速度达到 85.8 FPS,与基线模型基本持平。

4.3.2 骨干网络的影响

当引入新设计的骨干网络后,mAP 提升至 61.15%,相较基线提升了 1.5%。该提升在骑行者类别上尤为明显,mAP 从 59.1% 提升至 63.4%,增加了 4.3%;在行人类别上,mAP 也从 43.5% 提升至 43.8%,提升了 0.3%。相较之下,汽车类别的性能略有下降,从 76.3% 降至 76.2%,这是由于基线性能已较高,进一步提升空间有限。

新网络带来的性能提升,尤其在行人和骑行者类别上,主要得益于其增强的特征提取能力。

通过采用残差连接和 Dropout,新骨干网络能够在不显著增加网络复杂度的情况下,高效捕获更多局部特征。尽管推理速度略有下降,但考虑到精度提升,这一折中是合理的,使得新骨干网络成为面向实时激光雷达应用的有效升级方案。该设计在处理如行人和骑行者等点云结构复杂或稀疏的类别时展示出优势,细粒度特征的保持进一步提升了检测精度。而在汽车类别上的微弱性能下降表明,新网络更加适用于小型或复杂目标。在这类任务中,其先进的特征提取机制带来了更显著的收益。

4.3.3 数据增强的影响

在第 3.3 节中,提出并改进了一种新的数据增强方法 Part-Aware Data Augmentation (PA-AUG)<sup>[16]</sup>。针对整体模型及该方法本身,本文进行了消融实验验证。引入该方法后,整体 mAP 提升至 61.6%,在行人和骑行者类别上均实现了 3.0% 的提升。性能的提升主要得益于改进的方法通过在点云样本间插值得到更为真实的增强数据,相比传统的随机噪声生成方式,PointMixup 更能避免在低密度场景下引入无效特征,从而提升了模型的泛化能力。尤其在骑行者类别上,AP 的大幅提升表明新方法在处理稀疏及变化多样的点云结构时效果显著。

表 6 数据增强方法在 KITTI 数据集上的消融实验

Tab. 6 Ablation study of data augmentation on KITTI dataset

| 算法                     | 中等难度 AP/% |       |       | mAP/% |
|------------------------|-----------|-------|-------|-------|
|                        | Car       | Ped.  | Cyc.  |       |
| PA-AUG <sup>[16]</sup> | 76.44     | 47.03 | 62.22 | 61.90 |
| Ours                   | 77.81     | 48.41 | 62.47 | 62.90 |

此外,还对 PA-AUG 方法的改进效果进行了实验验证。如表 6 所示,引入 PointMixup 后,整体 mAP 得到了显著提升,尤其在最具挑战性的行人类别上。由于在点云数据中,行人通常比汽车等目标体积更小且形状更不规则,因此增强方法对于行人类别检测性能的提升尤为重要。

4.4 可视化点云

为全面评估所提模型的性能,超越量化指标的限制,本节可视化了点云,并与基线模型 PointPillars 进行比较,结果如图 3 所示,RGB 图像



图3 检测结果可视化

Fig. 3 Visualization of detection results

(a~c)中GT框、PointPillars和所提模型的预测框颜色分别为绿色、红色和蓝色。该分析涵盖3种场景:(a)狭窄道路、(b)具有高低差的复杂路口、(c)常规道路。在这些场景中,所提模型始终表现出优异性能,能够以更高精度识别所有目标物体,如图3(g,i)所示,所提模型展现出更强的鲁棒性,能有效区分真实目标与环境杂波,而PointPillars在这些复杂场景中常出现对远距离目标和道路中的杆子误识别的情况,如图3(d,f)所示。此外,所提模型在处理具有一定高低差的场景时也展现出更好的性能,如图3(e)中PointPillars对雷达不同水平面的物体容易产生误检,而本文所提模型的表现则更优秀。上述改进凸显了所提模型在特征提取方面的增强能力及其减少误检的能力,这可以归功于其先进的架构设计与优化的数据增强策略。这些因素共同使所提模型在目标与背景区分至关重要的复杂、杂乱环境中表现出色。

## 5 结 论

本文提出了一种基于柱体模型的新型三维物体检测框架,实现了模型速度和精度上的平衡。多柱体特征的融合使模型以较低的计算成本学习更多的上下文信息。在KITTI数据集上的实验结果表明,所提模型对比当前主流模型展现出了具有竞争力的综合性能,在55.65的FPS下,3D mAP达到68.93%,鸟瞰视角mAP达到75.02%。消融实验进一步证明了所提方法的有效性,中等难度mAP对比基线模型提升达到3.9%。在DAIR-V2X数据集上的实验进一步体现了所提模型优秀的泛化性和鲁棒性。本文所做的设计在实现上并不复杂,没有自定义的CUDA运算,因此可以简单地实现TensorRT的加速。未来的工作重点将是在保持高精度和高速度的同时,针对实际应用对模型进行优化,包括在嵌入式系统和自动驾驶平台上的部署。

## 参 考 文 献:

- [1] GUO Y L, WANG H Y, HU Q Y, *et al.* Deep learning for 3D point clouds: a survey [J]. *IEEE Transactions on*

- Pattern Analysis and Machine Intelligence*, 2021, 43(12): 4338-4364.
- [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 (CVPR). Honolulu: IEEE, 2017: 77-85.
  - [3] QI C R, YI L, SU H, *et al.* PointNet++: deep hierarchical feature learning on point sets in a metric space [C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates, Inc., 2017: 5105-5114.
  - [4] 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). Long Beach: IEEE, 2019: 770-779.
  - [5] 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. Salt Lake City: IEEE, 2018: 4490-4499.
  - [6] YAN Y, MAO Y X, LI B. SECOND: sparsely embedded convolutional detection [J]. *Sensors*, 2018, 18(10): 3337.
  - [7] LU L Q, XIE J M, HUANG R R, *et al.* An efficient hardware accelerator for sparse convolutional neural networks on FPGAs [C]. 2019 IEEE 27th Annual International Symposium on Field-Programmable Custom Computing Machines (FCCM). San Diego: IEEE, 2019: 17-25.
  - [8] 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). Long Beach: IEEE, 2019: 12689-12697.
  - [9] 赵晶, 李少博, 郭杰龙, 等. 基于知识蒸馏和定位引导的Pointpillars点云检测网络[J]. *液晶与显示*, 2024, 39(1): 79-88.  
ZHAO J, LI S B, GUO J L, *et al.* Pointpillars point cloud detection network based on knowledge distillation and location guidance [J]. *Chinese Journal of Liquid Crystals and Displays*, 2024, 39(1): 79-88. (in Chinese)
  - [10] WU H, WEN C L, LI W, *et al.* Transformation-Equivariant 3D object detection for autonomous driving [C]// *Proceedings of the 37th AAAI Conference on Artificial Intelligence*. Washington: AAAI, 2023: 2795-2802.
  - [11] LI X, MA T, HOU Y N, *et al.* LoGoNet: towards accurate 3D object detection with local-to-global cross-modal fusion [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Vancouver: IEEE, 2023: 17524-17534.
  - [12] WU X P, PENG L, YANG H H, *et al.* Sparse fuse dense: towards high quality 3D detection with depth completion [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 5408-5417.
  - [13] TIAN Y L, ZHANG X J, WANG X, *et al.* ACF-Net: asymmetric cascade fusion for 3D detection with LiDAR point clouds and images [J]. *IEEE Transactions on Intelligent Vehicles*, 2024, 9(2): 3360-3371.
  - [14] VASWANI A, SHAZEER N, PARMAR N, *et al.* Attention is all you need [C]// *Proceedings of the 31st International Conference on Neural Information Processing Systems*. Long Beach: Curran Associates, Inc., 2017.
  - [15] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: single shot MultiBox detector [C]. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
  - [16] CHOI J, SONG Y J, KWAK N. Part-aware data augmentation for 3D object detection in point cloud [C]. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). Prague: IEEE, 2021: 3391-3397.
  - [17] SHI S S, WANG Z, SHI J P, *et al.* From points to parts: 3D object detection from point cloud with part-aware and part-aggregation network [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, 43(8): 2647-2664.
  - [18] CHEN Y L, HU V T, GAVVES E, *et al.* PointMixup: augmentation for point clouds [C]. 16th European Conference on Computer Vision. Glasgow: Springer, 2020: 330-345.
  - [19] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The KITTI vision benchmark suite [C]. 2012 IEEE Conference on Computer Vision and Pattern Recognition. Providence: IEEE, 2012: 3354-3361.

- [20] YU H B, LUO Y Z, SHU M, *et al.* DAIR-V2X: A large-scale dataset for vehicle-infrastructure cooperative 3D object detection [C]. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans: IEEE, 2022: 21329-21338.
- [21] CHEN X Z, KUNDU K, ZHU Y K, *et al.* 3D object proposals for accurate object class detection [C]// *Proceedings of the 29th International Conference on Neural Information Processing Systems*. Montreal: Curran Associates, Inc., 2015: 424-432.
- [22] SIMONELLI A, BULÒ S R, PORZI L, *et al.* Disentangling monocular 3D object detection [C]. 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul: IEEE, 2019: 1991-1999.
- [23] OPENPCDET DEVELOPMENT TEAM. OpenPCDet: An open-source toolbox for 3D object detection from point clouds [EB/OL]. 2020[2025-02-02]. <https://github.com/open-mmlab/OpenPCDet>.
- [24] SINDAGI V A, ZHOU Y, TUZEL O. MVX-Net: multimodal VoxelNet for 3D object detection [C]. 2019 International Conference on Robotics and Automation (ICRA). Montreal: IEEE, 2019: 7276-7282.
- [25] QI C R, LIU W, WU C X, *et al.* Frustum PointNets for 3D object detection from RGB-D data [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City: IEEE, 2018: 918-927.
- [26] YANG Z T, SUN Y N, LIU S, *et al.* 3DSSD: point-based 3D single stage object detector [C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE, 2020: 11037-11045.
- [27] ZHOU Y, SUN P, ZHANG Y, *et al.* End-to-end multi-view fusion for 3D object detection in LiDAR point clouds [C]// *Proceedings of the 3rd Annual Conference on Robot Learning*. Osaka: PMLR, 2019: 923-932.
- [28] SHI G S, LI R F, MA C. PillarNet: real-time and high-performance pillar-based 3D object detection [C]. 17th European Conference on Computer Vision. Tel Aviv: Springer, 2022: 35-52.
- [29] HAYEON O, YANG C, HUH K. SeSame: simple, easy 3D object detection with point-wise semantics [C]. 17th Asian Conference on Computer Vision. Hanoi: Springer, 2025: 211-227.
- [30] YIN T W, ZHOU X Y, KRÄHENBÜHL P. Center-based 3D object detection and tracking [C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. Nashville: IEEE, 2021: 11779-11788.
- [31] CHEN Y K, LIU J H, ZHANG X Y, *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 (CVPR)*. Vancouver: IEEE, 2023: 21674-21683.

## 作者简介:



傅梓寒,男,硕士研究生,助理研究员,2023年于浙大宁波理工学院获得学士学位,主要从事点云目标检测方面的研究。E-mail: 202312490385@nuist.edu.cn



张丽娟,女,博士,教授,2015年于长春理工大学获得博士学位,主要从事计算机视觉及光学图像处理等方面的研究。E-mail: zhanglijuan@cwuxu.edu.cn