

文章编号:1007-2780(2026)06-0808-15

基于稀疏增强与几何注意力的改进 三维目标检测算法

陈祺东^{1,2,3}, 李家玮^{1,2}, 吴豪^{2,3*}, 姜琼丹²

(1. 南京信息工程大学 计算机学院、软件学院, 江苏 南京, 210044;

2. 无锡学院 物联网工程学院, 江苏 无锡, 214105;

3. 无锡学院 车路多模态感知与控制江苏省高校重点实验室, 江苏 无锡, 214105)

摘要:针对现有基于 PointPillars 的改进框架对行人等小目标检测精度不足的问题,提出一种融合稀疏增强与几何注意力的改进方法。首先,在 Pillar 特征编码阶段引入稀疏 Pillar 增强机制,填补稀疏区域的几何信息空白,为后续特征增强提供有效素材。其次,设计点感知空间注意力机制嵌入编码流程,基于补全后 Pillar 的点云密度与点到中心平均距离两类几何特征,通过轻量级 MLP 生成动态注意力权重,优先放大目标相关 Pillar 的特征信号,解决传统注意力依赖通道统计、不适配点云特性的缺陷。为验证方法的有效性,在公开 KITTI 三维目标检测数据集上开展实验,重点评估小目标的检测性能。实验结果表明,相较于现有的基于 PointPillars 的改进框架,三维模式下的行人目标的平均检测精度提升 2.54%,鸟瞰图模式下的行人目标的平均检测精度提升 5.91%,充分证明了所提创新模块对小目标特征优化的有效性,为自动驾驶长尾场景的感知任务提供了实用解决方案。

关键词:计算机视觉;三维目标检测;注意力机制;点云;激光雷达

中图分类号:TP391.4 **文献标识码:**A **doi:**10.37188/CJLCD.2026-0068 **CSTR:**32172.14.CJLCD.2026-0068

收稿日期:2026-04-22;**修订日期:**2026-05-16.

基金项目:江苏省基础研究计划(No. BK20240315);国家自然科学基金(No. 62406229);无锡市“太湖之光”科技攻关(基础研究)项目(No. K20253044, No. K20253037);江苏省高等学校基础科学(自然科学)研究面上项目(No. 25KJB520046);江苏省高等学校基础科学(自然科学)研究项目(No. 24KJB520039)

Supported by Basic Research Program of Jiangsu Province (No. BK20240315); National Natural Science Foundation of China (No. 62406229); Wuxi Science and Technology Development Fund “Taihu Light” Science and Technology (Fundamental Research) Project (No. K20253044, No. K20253037); Natural Science Foundation of Jiangsu Higher Education Institutions (No. 25KJB520046); Natural Science Foundation of Jiangsu Higher Education Institutions (No. 24KJB520039)

*通信联系人, E-mail: wuhao940917@gmail.com

Improved 3D object detection algorithm based on sparse enhancement and geometric attention

CHEN Qidong^{1,2,3}, LI Jiawei^{1,2}, WU Hao^{2,3*}, LOU Qiongdan²

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

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

3. Jiangsu Provincial University Key Laboratory of Vehicle-Road Multimodal Perception and Control, Wuxi University, Wuxi 214105, China)

Abstract: To address the issue of insufficient detection accuracy for small objects such as pedestrians in existing PointPillars-based frameworks, we propose an improved method that integrates sparse enhancement with geometric attention. First, we introduce a sparse Pillar enhancement mechanism during the Pillar feature encoding stage to fill in the gaps in geometric information in sparse regions, thereby providing effective material for subsequent feature enhancement. Second, we design a point-aware spatial attention mechanism embedded in the encoding process. Based on two types of geometric features of the completed Pillars, namely point cloud density and average point-to-center distance, we generate dynamic attention weights via a lightweight MLP to prioritize the amplification of feature signals from Pillars related to targets, thereby addressing the shortcomings of traditional attention mechanisms that rely on channel statistics and are ill-suited to point cloud characteristics. To validate the effectiveness of the method, experiments are conducted on the public KITTI 3D object detection dataset, with a focus on evaluating detection performance for small objects. The experimental results demonstrate that, compared to existing improved frameworks based on PointPillars, the average detection accuracy for pedestrian targets increases by 2.54% in 3D mode, and 5.91% in bird's-eye-view mode. This results fully validates the effectiveness of the proposed innovative module in optimizing features for small targets and provides a practical solution for perception tasks in long-tail scenarios of autonomous driving.

Key words: computer vision; 3D object detection; attention mechanism; point cloud; light detection and ranging

1 引言

随着自动驾驶技术的快速发展,环境感知系统对于目标识别和定位的准确性要求越来越高,精确的目标检测是保障行车安全的核心要素^[1]。激光雷达作为关键感知传感器,通过发射激光脉冲并测量反射时间获取环境的三维点云数据^[2]。相较于其他传感器,激光雷达获取到的三维空间信息更为精确,因此在自动驾驶系统中的应用愈加广泛。近年来,可重构双模式LiDAR、微梳光源等新型光学感知器件的发展进一步提升了三维主动

感知系统在视场、分辨率和测距效率方面的潜力^[3-4]。

传统的二维目标检测方法主要基于图像或投影平面进行特征提取与目标定位,但其受限于二维信息表达形式,难以全面刻画真实场景中的空间几何结构。相比之下,三维点云数据能够直接反映目标在真实世界中的空间位置、形状和尺度信息,具有更强的表达能力和环境适应性^[5]。因此,基于激光雷达进行三维目标检测的方法成为近年来的研究热点。在精密三维成像领域,自适应多尺度显微成像与多角度偏振光学相干层析成像等方法也表明,多尺度观测和三维形貌重建对于细粒度结构感知具有重要价值^[6-7]。

近年来,国内外已出现许多具有实际应用价值的三维目标检测算法。Maturana 等人^[8]提出的 VoxNet 架构,将体素占用网格表示与监督式三维卷积神经网络相结合。PointNet 算法^[9]和 SECOND 算法^[10]通过直接学习点云特征和引入三维稀疏卷积,实现了高质量特征提取或运行效率的提升,但这两种方式在捕捉局部几何结构或处理不规则数据上仍各有侧重。PointRCNN^[11]基于原始点云的自底向上方法实现了高效的 3D 精细化检测,无需依赖投影。基于体素的方法^[12-14]虽通过栅格化或引入混合卷积解决了点云不规则问题并增强了表征,但也因引入量化误差或复杂网络结构,导致了细节丢失或计算开销的显著增加。

为了克服三维体素带来的计算开销与精度损失之间的矛盾,研究者们开始探索更为高效的投影表示方法。PointPillars^[15]作为一种高效的三维目标检测框架,将点云转换为二维伪图像,在速度与精度间取得了平衡,已广泛应用于自动驾驶感知任务。然而,LiDAR 点云在远距离场景下往往呈现稀疏分布,导致 PointPillars 对小目标的检测性能下降。近年来,研究者针对稀疏点云补全和注意力机制进行了深入探索。学者们尝试通过引入混合池化与改进骨干网络^[16]、构建多维三重注意力模块^[17]、融合通道与空间注意力机制^[18]以及优化体素下采样方法^[19]来强化关键特征表达与噪声抑制。然而,这些策略在提升精度的同时,也面临着降维过程中的信息损失或因改变特征分布而导致的表达不稳定等问题。PillarNeXt^[20]通过引入 ASPP neck 和改进检测头,并结合更充分的训练策略,取得了精度与速度的更优权衡,但也带来了更高的训练开销与调参成本。此外,视觉目标跟踪领域相关研究表明,遮挡、尺度变化与复杂背景也会显著削弱目标定位鲁棒性^[21-22]。

针对这些缺陷,本文提出融合稀疏 Pillar 增强机制 (Sparse-Enhanced Pillar Augmentation, SEPA) 和点感知空间注意力机制 (Point-Aware Spatial Attention, PASA) 的 SP-Pillars 改进框架,其核心贡献包括:(1)融合稀疏 Pillar 增强机制通过邻域真实点借用和虚拟点插值策略,高效填补稀疏 Pillar 的几何信息,并设计基于邻域几何先验的自适应高度校正策略,根据局部场景结构动态调整校正强度;(2)点感知空间注意力机制基

于点云密度和平均距离生成动态权重,优先增强目标相关特征;(3)二者协同形成信息补全与特征聚焦的链路,在 KITTI 数据集上显著提升小目标检测精度,同时维持高效推理速度。

2 相关工作

2.1 基于 PointPillars 的改进方法

PointPillars 通过 3D 点云到 2D 伪图像的维度转换,首次实现了高精度与高实时性的平衡^[15]。PointPillars 作为体素类三维检测方法的高效改进,将点云在 x - y 平面划分为垂直柱状体素,通过柱状特征学习后散射为伪图像,并在其上采用纯 2D 卷积完成特征提取与检测,避免代价高昂的 3D 卷积与手工特征设计。相较于其他三维目标检测算法,PointPillars 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,使用二维的骨干网络 (Backbone) 进行特征提取。但柱状离散化对空间的量化会带来位置信息损失,分辨率降低时定位误差更明显,对行人等小目标的检测精度影响尤为突出。

为进一步提升性能,研究者围绕 PointPillars 的薄弱环节展开改进。Shu 等人^[23]通过在伪图特征中引入通道注意力与空间注意力强化关键信息,并采用轻量化的 FasterNet 替换原下采样网络以减少参数量,同时用 CARAFE 替代反卷积上采样以降低特征模糊、提升多尺度语义融合。Kong 等人^[24]在第一阶段于 Pillar 特征编码中引入改进的 SE 注意力机制并结合动态卷积与空间注意力增强小目标特征表达,在第二阶段利用 Transformer 检测头对候选框进行细化以提升定位精度,但引入 Transformer 与动态卷积后计算开销和参数量明显增加,导致推理速度较基线 PointPillars 明显下降,实时性受到一定影响。张淦等人^[25]提出点云特征增强的改进 PointPillars,通过引入密度感知采样与稀疏化增强,并在网络中加入多种注意力与特征融合模块以强化小目标表达,但稀疏化会削弱目标内部细节,使模型对朝向特征更敏感。MDS-PointPillars^[26]通过在 PFN 中设计多池化编码模块融合最大及平均注意力池化以增强多尺度特征表达,并在骨干网络中引入深度可分离卷积块与无参数注意力 SimAM 以在降低计算量的同时强化关键特征。MAT-

PointPillars^[27]在PointPillars框架中引入多尺度注意力模块优化前两阶段上采样,并在第三阶段上采样中融入Transformer Encoder以增强对小目标与非刚性目标的特征提取能力,但Transformer层数增加会显著提升计算复杂度、影响实时性。赵晶等人^[28]进一步将知识蒸馏与定位引导引入PointPillars,在不增加额外嵌入结构的条件下提升了检测精度,但其对稀疏Pillar内部几何信息缺失的补偿有限。除此之外,研究者们也尝试通过改进训练样本分布策略^[29]和通过密集连接优化特征重用^[30],但这些改进均基于Pillar特征有效的前提,当Pillar稀疏时,优化效果会被严重削弱。

2.2 稀疏点云增强方法

PointPillars框架通过将点云转换为伪图像能够显著提升检测速度,但其在处理室外大规模场景时仍存在局限。激光雷达采集的原始点云通常表现为低密度、离散且不连续的特征^[31],由于扫描轨迹、物体分布以及环境结构差异,点云可能出现局部过采样或欠采样现象^[32],使得数据质量在空间上呈现不均衡性。这种特征稀疏性在Pillar投影过程中往往会导致远端或小目标内部出现大量空置柱体,造成结构信息的断裂与几何特征的缺失。

针对稀疏点云的信息缺失问题,现有方法分别在点云级补全与Pillar级补全两类上进行研究。点云级补全方法聚焦于全局点云的完整重建,通过生成模型填补稀疏区域。缪永伟等人^[33]提出基于生成对抗网络的三维点云自动修复补全方法,其生成器采用编码器-解码器结构,通过输入特征对齐与共享多层感知机(MLP)提取局部形状特征,并利用折叠解码生成缺失部分,最终与原点云融合得到完整形状。判别器对真实与生成点云进行真假判别以持续优化生成器,从而在稠密与稀疏缺失点云上兼顾补全效果与精细结构保持。该方法依赖对抗训练,训练过程可能不稳定且对超参数敏感,同时折叠解码对复杂细节的表达能力有限,遇到大面积缺失时仍可能出现结构偏差。融合图注意力的多分辨率点云补全网络^[34]在GAN框架下利用图注意力层构建点云局部图结构并融合多分辨率特征,结合折叠操作实现缺失结构的逐级重构,从而获得更完整且更均匀的补全点云。但由于图注意力模块参数量

较大,会导致训练时间显著增加,同时非均匀采样点云的局部特征提取上效果不佳,仍依赖更优的采样策略以提升鲁棒性。EF-MSVoteNet框架^[35]能在潜在目标区域内自适应提升点云分辨率,同时降低背景区域的分辨率,结合多尺度投票网络有效增强杂乱环境中目标的特征表征能力。点云级补全能够显式重建连续几何形状,但其通常需要对输入点云进行全局建模并生成大量点,计算与显存开销较大。

Pillar级补全方法为适配PointPillars框架应运而生,以局部特征补全降低计算成本,但补全特征的真实性的提升仍待提升。与点云级方法直接生成三维坐标点不同,Pillar级补全更多是在柱体网格化表示空间内进行结构增强,通过优化柱状特征表示^[36]能够显著增强检测精度。Pillar级补全的核心目标是以较低代价增强稀疏区域的Pillar占用与特征表达,使特征图在空间上更连续、更可学习,从而提升下游检测鲁棒性。在提升效率方面,SPADE^[37]通过动态向量剪枝最大化Pillar级稀疏性,并设计稀疏坐标管理与稀疏感知数据流来高效执行向量稀疏卷积,从而显著减少无效计算并提升速度与能效。后续Pillar级补全研究往往在效率优化之外进一步引入结构先验或跨尺度融合,以在保持稀疏计算优势的同时改善特征的真实性与可迁移性。

3 方法设计

3.1 算法框架概述

为解决PointPillars算法对行人等小目标检测精度低的问题,本文提出了SP-Pillars增强框架,具体框架结构如图1所示。SP-Pillars在原算法的基础上进行了以下改进:(1)改进柱状特征编码模块PFN,在PFN模块中引入稀疏Pillar自适应补全策略SEPA,遍历并识别点数低于阈值 θ 的稀疏Pillar,通过从其 3×3 邻域Pillar借用 k 个有效点或基于邻域点加权平均的线性插值生成虚拟点进行填充,在不改变Pillar网络结构的前提下显著提升稀疏区域点密度与特征多样性;(2)提出点感知空间注意力机制PASA嵌入PFN中,利用Pillar内点到中心的距离与局部密度等空间分布信息为伪图像中每个Pillar生成空间注意力权重,使网络优先关注点分布更密集或更靠

近中心的 Pillar, 以增强伪图像特征表达与目标响应能力。综上, 本文提出的 SP-Pillars 框架通过在

PFN 阶段嵌入两大创新模块的轻量化修改, 精准定位并解决了稀疏 Pillar 特征问题。

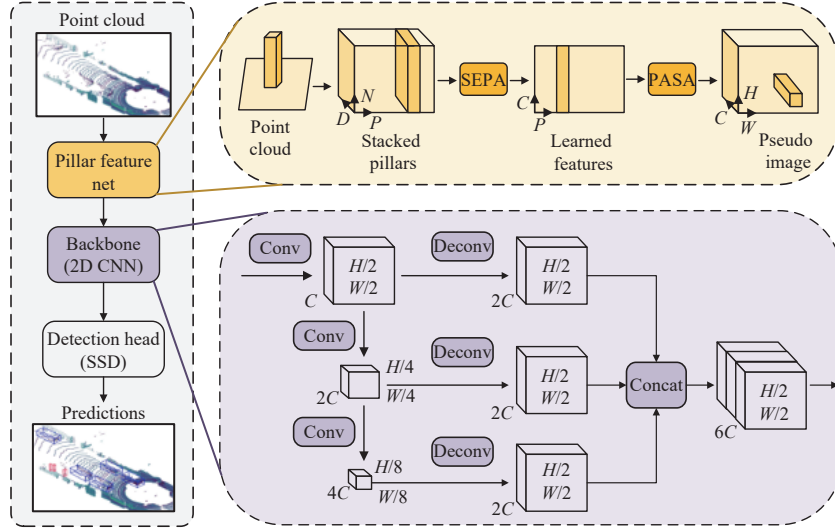


图 1 SP-Pillars 模型框架

Fig. 1 Framework diagram of SP-Pillars model

3.2 稀疏 Pillar 的自适应补全策略

稀疏区域或小目标所在的 Pillar 往往仅包含极少量点云, 在 PFN 聚合时可用几何与语义线索不足, 导致柱状特征表达单一、目标边界模糊, 进而使网络对稀疏 Pillar 的响应偏弱^[38]。针对这一问题, 本文在 PFN 阶段设计了稀疏 Pillar 自适应补全策略, 通过借用邻域 Pillar 的有效点或基于邻域点加权平均进行线性插值生成虚拟点, 对稀疏 Pillar 进行密度填充, 有效缓解了柱状量化带来的信息缺失, 使稀疏区域获得更稳定、更具判别性的柱状特征表示, 从而增强小目标与稀疏场景下的检测鲁棒性与定位精度。SEPA 模块的伪代码如算法 1 所示。

算法 1 稀疏 Pillar 自适应补全方法

输入: 目标 Pillar P , 阈值 k

输出: 处理后的 Pillar 数据 P_{out}

- 1 初始化 P_{out} 为空: $P_{out} \leftarrow \text{empty}$
- 2 检测当前 Pillar 是否稀疏: $\text{is_sparse} \leftarrow \text{CHECK_SPARSE_PILLAR}(P)$
- 3 如果 (is_sparse 为真) 则
 - 4 选择有效邻域: $\text{neighbors} \leftarrow \text{SELECT_EFFEC-TIVE_neighbors}(P, \text{dist}_{xy}, \Delta z)$
 - 5 设置上下文: $\text{context} \leftarrow \text{neighbors}$

- 6 否则
- 7 $\text{context} \leftarrow \text{NULL}$
- 8 结束如果
- 9 确定优先级数据源: $\text{source} \leftarrow \text{SELECT_PRIORITY_SOURCE}(P, \text{context})$
- 10 随机采样: $\text{sampled_points} \leftarrow \text{RANDOM_SAMPLE}(\text{source}, k)$
- 11 计算自适应校正权重: $\alpha \leftarrow \text{COMPUTE_ADAPTIVE_ALPHA}(\sigma_{z, \text{curr}}, \sigma_{z, \text{neigh}})$
- 12 坐标矫正: $\text{sampled_points} \leftarrow \text{COORDINATE_CORRECTION}(\text{sampled_points}, P_{\text{curr}}, P_{\text{neigh}}, \alpha)$
- 13 统计有效点数: $\text{valid_count} \leftarrow \text{COUNT_POINTS}(\text{sampled_points})$
- 14 如果 ($\text{valid_count} < k$) 则
 - 15 执行虚拟点插值: $P_{out} \leftarrow \text{VIRTUAL_POINT_INTERPOLATION}(\text{sampled_points})$
- 16 否则
- 17 $P_{out} \leftarrow \text{sampled_points}$
- 18 结束如果
- 19 返回 P_{out}

传统稀疏 Pillar 识别方法需遍历所有 Pillar 的点数量数组, 再对每个稀疏 Pillar 逐一遍历 3×3 邻域以查询有效邻域 Pillar, 时间复杂度为 $O(N \times 9)$ 。在 SEPA 模块中, 创新性地构建坐标-索引映射表, 将 Pillar 的空间坐标与其在批量数

据中的索引直接关联,实现稀疏 Pillar 及其邻域的 $O(1)$ 时间复杂度查询,显著提升定位效率。

SEPA 优先从有效邻域 Pillar 中借用真实点云,并通过坐标偏移校正确保借用点的空间位置与当前稀疏 Pillar 的几何范围一致,从根本上避免特征几何失真问题。但并非所有非稀疏的邻域 Pillar 都适合作为点云借用源,因此,SEPA 在筛选有效邻域时,额外增加两个约束条件:

(1) 空间距离约束:计算邻域 Pillar 中心与当前稀疏 Pillar 中心的 XY 平面欧氏距离 dist_{xy} ,仅保留 $\text{dist}_{xy} < 0.3 \text{ m}$ 的邻域 Pillar:

$$\text{dist}_{xy} = \left\| \begin{pmatrix} x_{\text{center, neigh}} \\ y_{\text{center, neigh}} \end{pmatrix} - \begin{pmatrix} x_{\text{center, curr}} \\ y_{\text{center, curr}} \end{pmatrix} \right\|, \quad (1)$$

式中: $(x_{\text{center, neigh}}, y_{\text{center, neigh}})$ 是邻域 Pillar 的中心坐标, $(x_{\text{center, curr}}, y_{\text{center, curr}})$ 是当前稀疏 Pillar 的中心坐标。

(2) Z 轴分布约束:计算邻域 Pillar 内点云的 Z 轴均值与当前稀疏 Pillar 的 Z 轴均值的差值 Δz ,仅保留 $\Delta z < 1.5 \text{ m}$ 的邻域 Pillar:

$$\Delta z = \| z_{\text{mean, neigh}} - z_{\text{mean, curr}} \|, \quad (2)$$

式中: $z_{\text{mean, neigh}}$ 是邻域 Pillar 内点云的 Z 轴平均值, $z_{\text{mean, curr}}$ 是当前稀疏 Pillar 内点云的 Z 轴平均值。

空间距离约束直接决定了借用点云的目标归属正确性。一旦该约束失效,跨目标、跨背景的

点云会被错误借用,导致补全后的 Pillar 特征出现严重的几何失真,进而误导后续注意力权重的生成。Z 轴分布约束本质上是空间距离约束在垂直维度上的补充。由于 KITTI 场景中同类小目标的高度分布相对稳定, Δz 约束在大部分情况下处于软过滤状态,仅在地面反射点或悬空结构等少数极端情形下发挥关键作用。因此,通过上述约束筛选后,若存在多个有效邻域 Pillar, SEPA 会按 dist_{xy} 、 Δz 、点数量的优先级排序,优先选择距离更近、Z 轴分布更相似、点云更密集的邻域 Pillar 作为点云的借用源。

确定点云借用源后, SEPA 对真实点云采用随机采样-坐标校正的补全策略,优先从最高优先级的有效邻域 Pillar 随机采样 k 个点;若该邻域点数不足,则取其全部可用点并依次从次优先级邻域补足,直至达到 k 个点。若所有有效邻域的总点数仍小于 k , 则采样全部可用点, 剩余部分由虚拟点生成策略补全。点云补全策略如图 2 所示。

传统方法中的坐标偏移校正是直接将邻域点云的原始坐标复制到当前稀疏 Pillar, 忽略了邻域 Pillar 与当前 Pillar 的中心偏移, 导致借用点的空间位置与当前 Pillar 的几何范围不匹配, 最终造成特征提取时的空间错位。在 SEPA 模块中, 通过计算邻域与当前 Pillar 的 X、Y 偏移量, 确保

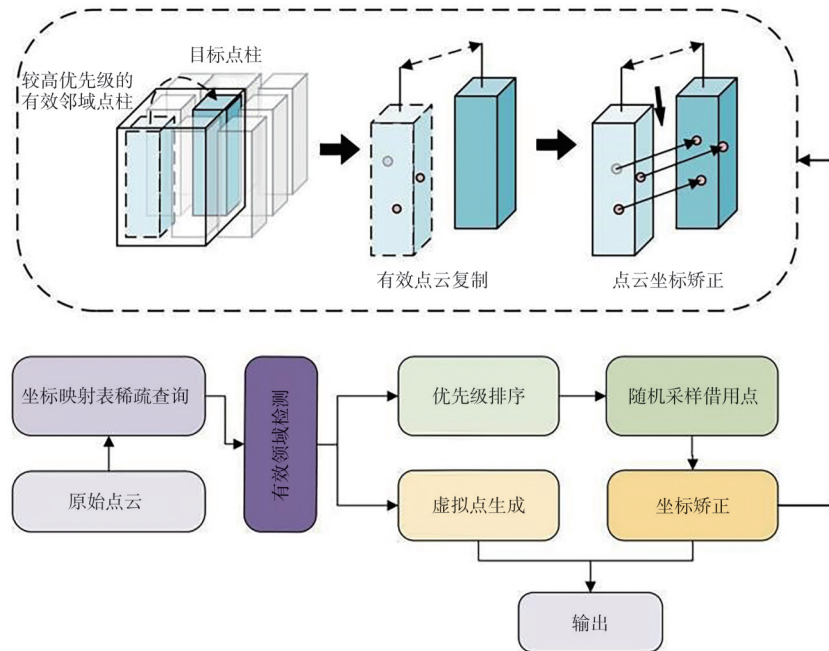


图2 点云补全策略架构

Fig. 2 Architecture of point cloud interpolation strategy

借用点的 X 、 Y 坐标落入当前 Pillar 的网格范围, 解决空间错位问题。同时对 Z 轴也进行校正, 让借用点的高度分布与当前 Pillar 更匹配。通过以下公式对借用点进行坐标校正:

$$\begin{cases} x_{\text{curr}} = x_{\text{neigh}} - (x_{\text{center, neigh}} - x_{\text{center, curr}}) \\ y_{\text{curr}} = y_{\text{neigh}} - (y_{\text{center, neigh}} - y_{\text{center, curr}}) \\ z_{\text{curr}} = z_{\text{neigh}} - (z_{\text{mean, neigh}} - z_{\text{mean, curr}}) \times \alpha \end{cases}, \quad (3)$$

$$\alpha = \alpha_{\min} + (\alpha_{\max} - \alpha_{\min}) \cdot \frac{1}{1 + \gamma \cdot |\sigma_{z, \text{curr}} - \sigma_{z, \text{neigh}}|}, \quad (4)$$

式中: $(x_{\text{neigh}}, y_{\text{neigh}}, z_{\text{neigh}})$ 是邻域点的原始坐标; $(x_{\text{curr}}, y_{\text{curr}}, z_{\text{curr}})$ 是校正后的借用点坐标; α 为基于几何先验得到的自适应权重系数; $\sigma_{z, \text{curr}}, \sigma_{z, \text{neigh}}$ 表示当前 Pillar 与邻域 Pillar 内点云的 Z 轴标准差, 反映各自的垂直结构离散程度; $\alpha_{\min}, \alpha_{\max}$ 用于控制 α 的取值范围; γ 控制 α 对高度差异的敏感度。

在部分极端场景下, 若不存在有效邻域 Pillar, 则需通过虚拟点插值补充剩余点数, 确保每个稀疏 Pillar 的点数量提升至较为均衡且有效的状态。虚拟点基于当前稀疏 Pillar 内现有点的局部几何统计生成, 通过均值插值和微小扰动的方式生成, 以保证补全点符合真实形状分布并抑制噪声引入。其具体流程如下:

(1) 对当前稀疏 Pillar 内的现有点云计算两类局部几何特征: 坐标均值 μ 及反射强度均值 $\bar{r}$, 作为虚拟点生成的约束依据:

$$\mu = \frac{1}{n} \sum_{i=1}^n (x_i, y_i, z_i), \quad (5)$$

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n r_i, \quad (6)$$

式中: (x_i, y_i, z_i) 是点云的坐标, n 是对应 Pillar 内点数量, r 是点云的强度。

(2) 进一步计算协方差矩阵 C , 并作特征分解:

$$C_{v_i} = \lambda_i v_i, \quad \lambda_1 \geq \lambda_2 \geq \lambda_3, \quad (7)$$

式中: v_i 表示主延伸方向; λ_i 刻画各方向的离散尺度, 用于对虚拟点分布施加形状约束。

(3) 以 μ 为中心, 沿 3 条特征方向生成基础虚拟点:

$$\widetilde{p}_{i,1} = \mu + s_{i,1} v_i \quad i \in \{1, 2, 3\}, \quad (8)$$

式中: $s_{i,1}$ 是沿该方向的位移长度, $\widetilde{p}_{i,1}$ 是在方向 v_i

上生成的第 i 个基础虚拟点坐标。

(4) 为避免补全点过于规则, 对每个基础点加入微扰。扰动过大时, 虚拟点易超出当前 Pillar 在 XY 平面的网格边界, 将相邻背景区域的伪几何信号混入目标特征, 削弱目标与背景的可分性; 扰动过小时, 虚拟点几乎与公式 (8) 生成的基础点重合, 沿主方向呈规则排布, 无法为稀疏 Pillar 提供新的几何线索, 特征多样性不足。因此, 扰动幅度随 Pillar 内点云的局部空间分布范围自适应: 对分布较紧凑的目标加入的扰动较小, 避免破坏目标边界; 对分布较宽展的目标加入的扰动较大, 保证特征多样性。公式如下:

$$\widehat{p}_{i,1} = \widetilde{p}_{i,1} + \epsilon, \quad (9)$$

$$\epsilon \sim N(0, \Sigma \epsilon), \quad \Sigma \epsilon = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_z^2), \quad (10)$$

$$\sigma_x = \sigma_y = 0.2 \cdot \min\left(\sqrt{\lambda_1}, \frac{1}{2} \Delta\right), \quad \sigma_z = 0.2 \cdot \sqrt{\lambda_3}, \quad (11)$$

式中: ϵ 是三维零均值各向异性高斯随机向量, Δ 是 Pillar 网格边长, $\widehat{p}_{i,1}$ 是最终虚拟点坐标。

3.3 点感知空间注意力机制

在柱状编码后形成的伪图像中, 各 Pillar 通常被同等对待, 而稀疏背景 Pillar 与真实目标 Pillar 在特征聚合时缺乏显式区分, 导致网络容易受到无效区域干扰, 削弱对目标尤其是点云更密集或结构更清晰区域的关注。针对这一问题, 本文在柱状特征编码模块中提出点感知空间注意力机制并嵌入 PFN 中, 利用 Pillar 内点到中心的距离与局部密度等空间分布信息, 为伪图像中每个 Pillar 自适应生成空间注意力权重, 使模型优先强化点分布更密集或更靠近中心的 Pillar, 并抑制稀疏且噪声占比高的背景 Pillar。该机制在不增加复杂结构的前提下提升了伪图像特征的判别性与目标响应强度, 增强检测鲁棒性并改善目标定位与分类性能。点感知空间注意力机制架构图如图 3 所示。

点感知空间注意力机制模块以 9 维原始特征和 2 维几何先验特征为输入, 基于点云密度与结构紧致性优先放大目标相关 Pillar 的特征信号。PASA 模块从点云的稀疏性本质出发, 筛选出两类与目标判别性强相关的原生几何特征, 确保注意力权重能精准识别并增强目标对应的稀疏 Pillar。

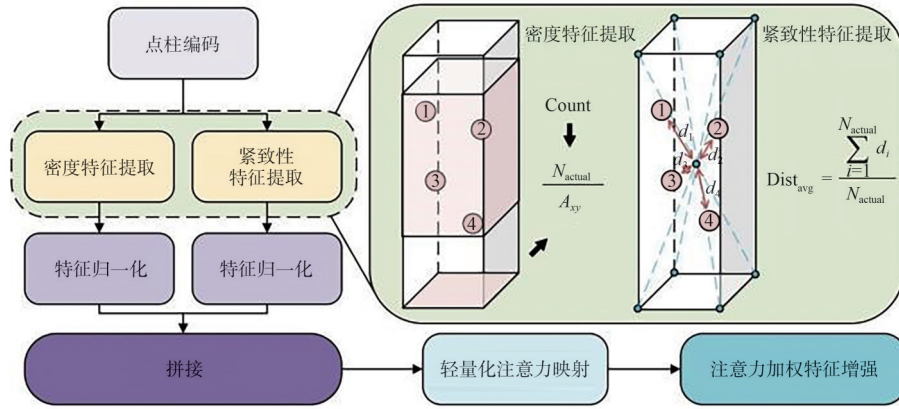


图3 点感知空间注意力机制架构

Fig. 3 Architecture of point-wise spatial attention mechanism

(1)密度特征:

$$d = \frac{n_{\text{points}}}{\text{pillar}_{\text{area}}}, \quad (12)$$

式中: n_{points} 为经 SEPA 补全后 Pillar 内的实际点数量, $\text{pillar}_{\text{area}}$ 为 Pillar 在 XY 平面的投影面积。

(2)距离特征:

$$\text{dist} = \frac{1}{n_{\text{points}}} \sum_{i=1}^{n_{\text{points}}} p_i - c, \quad (13)$$

式中: $p_i = (x_i, y_i, z_i)$ 为 Pillar 内第 i 个点的 3D 坐标, $c = (x_{\text{center}}, y_{\text{center}}, z_{\text{center}})$ 为 Pillar 的 3D 中心坐标, $p_i - c$ 为第 i 个点到中心的欧氏距离。

为避免密度特征 d 与距离特征 dist 的数值范围差异影响 MLP 的训练稳定性, PASA 注意力模块对两类特征进行批量归一化处理:

$$d_{\text{norm}} = \frac{d - \mu_d}{\sigma_d + \epsilon}, \quad (14)$$

$$\text{dist}_{\text{norm}} = \frac{\text{dist} - \mu_{\text{dist}}}{\sigma_{\text{dist}} + \epsilon}, \quad (15)$$

式中: μ_d 、 σ_d 分别为批量内密度特征的均值与标准差, μ_{dist} 、 σ_{dist} 分别为批量内距离特征的均值与标准差, $\epsilon = 10^{-6}$ 为避免分母为零的微小常数。归一化后, 两类特征的数值范围均映射至 $[-1, 1]$, 确保 MLP 各层的梯度稳定, 加速训练收敛。

PASA 通过极简的 MLP 结构实现几何特征到注意力权重的映射, 公式如下:

(1) 将归一化后的密度特征和距离特征拼接成一个局部的几何特征向量:

$$f_{\text{geo}} = [d_{\text{norm}}, \text{dist}_{\text{norm}}]. \quad (16)$$

(2) 通过轻量级 MLP 对几何特征向量进行

变换得到注意力权重 ω :

$$\omega = \sigma(W_2 \cdot \text{ReLU}(W_1 \cdot f_{\text{geo}} + b_1) + b_2), \quad (17)$$

式中: W_1 、 W_2 和 b_1 、 b_2 分别为 MLP 层的权重矩阵和偏置项; ReLU 为非线性激活函数, 用于提取高维特征。

得到注意力权重 ω 后, PASA 通过权重广播及逐元素相乘的方式对 Pillar 编码阶段的特征进行加权增强。具体流程如下:

(1) 权重广播: 将维度为 $[N, 1]$ 的权重 ω 广播至与 Pillar 特征相同的维度 (Pillar 特征的维度为 $[N, C, M]$), 广播后权重维度变为 $[N, 1, 1]$, 确保能与特征逐元素匹配;

(2) 逐元素相乘: 将广播后的权重与 Pillar 特征进行逐元素相乘, 得到增强后的特征 F_{enhanced} :

$$F_{\text{enhanced}} = F_{\text{raw}} \times \omega_{\text{broadcast}}, \quad (18)$$

式中: F_{raw} 为 Pillar 编码阶段的原始特征, $\omega_{\text{broadcast}}$ 为广播后的权重。

PASA 模块在设计上采用仅含 3 层的轻量级 MLP 结构, 总参数量仅为 0.6k, 远低于 CBAM^[39] 的 3.2k 参数量, 在保持特征增强能力的同时显著降低了模型参数开销。在推理速度上, 基于 NVIDIA RTX 4090 平台的测试结果表明, PASA 模块的平均耗时仅为 0.4 ms, 对整体检测框架的实时推理性能影响可忽略不计。

3.4 协同增强机制

通过将 SEPA 与 PASA 注意力模块进行协同融合, 网络在 PFN 阶段对稀疏 Pillar 的特征表达能力得到了系统性增强。首先由 SEPA 针对点数较少、几何信息不足的稀疏 Pillar, 引入自适应补

全策略,利用空间邻域 Pillar 的点分布信息进行点借用或插值生成,有效提高稀疏 Pillar 的点密度和结构完整性。在此基础上,PASA 进一步结合补全后 Pillar 内点的密度特征及其相对于 Pillar 中心的空间分布关系,为每个 Pillar 自适应生成空间注意力权重,使网络能够重点关注包含有效几何信息的 Pillar 区域。该协同增强方式使网络在特征提取过程中既能弥补稀疏区域信息不足的问题,又能够突出潜在目标区域的判别性特征,增强了稀疏点云中目标特征在深层网络中的可传递性与稳定性。通过双机制串行的协同作用,网络在面对小目标以及复杂稀疏场景时,能够更有效地保留和强化关键特征信息,避免稀疏 Pillar 特征在逐层下采样过程中被削弱或淹没,最终实现对三维目标检测精度与鲁棒性的综合提升。

4 实验与结果分析

4.1 数据集

街景图像是目前多项研究的重要数据来源^[40],本实验在公开数据集 KITTI 上进行训练和测试。KITTI 数据集^[41]是自动驾驶研究中常用的重要公开数据集,广泛用于 3D 物体检测任务。该数据集包含来自市区、郊区和乡村等多种场景的 RGB 图像和激光雷达点云数据,数据标注包括二维和三维的真值标签。数据集分为训练集和测试集,分别包含 7 481 个和 7 518 个样本。为了评估模型的性能,训练集进一步划分为 3 712 个训练样

本和 3 769 个验证样本。

4.2 实现细节

本实验使用 PyTorch 深度学习框架,模型训练主机搭载 13th Gen Intel(R) Core(TM) i7-13700F 处理器,GPU 为 RTX4090,操作系统为 Ubuntu20.04。本实验检测目标包括汽车、行人和骑行者,每个目标根据其大小、遮挡程度和截断级别等因素分为容易、中等和困难 3 个难度级别。实验中 x,y,z 的范围分别设置为 $[0,69.12]$ m、 $[-39.68,39.68]$ m、 $[-1,3]$ m,网格大小设置为 $0.16\text{ m}\times0.16\text{ m}\times4\text{ m}$,每个柱中的最大点数设置为 20。自适应权重 α 的取值范围设置为 $\alpha_{\min}=0.2,\alpha_{\max}=0.8$,敏感度系数 $\gamma=2.0$ 。训练 160 轮,batch_size 设置为 8,使用 ADAM 优化器^[42]进行训练,初始学习率设置为 0.001,遵循余弦退火策略进行更新。实验结果包括鸟瞰图(Bird's-Eye View, BEV)、3D 两个维度。

4.3 实验结果

实验通过在 KITTI 数据集中进行验证,将本文模型与其他传统算法模型和最新的方法进行对比。遵循 KITTI 官方测评协议^[43],对于汽车、行人和骑行者的交并比(Intersection over Union, IoU)阈值分别设置为 0.7、0.5、0.5,以平均精度(Average Precision, AP)作为评估指标,对模型的性能进行定量分析。

实验选择基于 PointPillars 的改进算法 PEPillar 作为基准模型,表 1、表 2 分别展示了在 KITTI 测试集中,鸟瞰图模式和 3D 模式下本文

表 1 各模型在鸟瞰图模式的性能对比

Tab. 1 Performance comparison of models in Bird's-Eye View mode

Models	Car AP/%				Pedestrian AP/%				Cyclist AP/%			
	Easy	Mod.	Hard	mAP	Easy	Mod.	Hard	mAP	Easy	Mod.	Hard	mAP
VoxelNet ^[12]	89.35	79.26	77.39	82.00	46.13	40.74	38.11	41.66	66.70	54.76	50.55	57.34
SECOND ^[10]	88.07	79.37	77.95	81.80	55.10	46.27	44.76	48.71	73.67	56.04	48.78	59.50
PointPillars ^[15]	88.35	86.10	79.83	84.76	58.66	50.23	47.19	52.03	79.14	62.25	56.00	65.80
PointRCNN ^[11]	89.14	87.62	86.05	87.60	62.82	58.18	51.46	57.49	84.51	69.11	64.15	72.59
ET-PointPillars ^[19]	88.67	87.91	87.27	87.95	56.07	50.40	47.41	51.29	79.77	61.14	58.46	66.46
EC-PointPillars ^[25]	88.65	86.45	84.32	86.47	61.36	57.84	51.02	56.74	83.25	68.45	63.01	71.57
PE-Pillar ^[36]	89.73	87.31	85.65	87.56	63.74	59.48	53.06	58.76	85.23	70.63	65.52	73.79
PV-RCNN ^[14]	89.89	87.55	86.37	87.94	67.21	58.67	54.50	60.13	87.64	72.50	69.07	76.40
PASS-PV-RCNN++ ^[29]	90.11	88.10	87.60	88.60	66.86	62.00	56.84	61.90	93.76	74.91	70.65	79.77
DensePointPillars ^[30]	92.13	86.31	81.12	86.52	47.80	40.31	37.49	41.87	75.14	63.27	56.70	65.04
Ours	90.82	89.54	88.21	89.52	67.25	62.26	57.56	62.36	93.83	75.05	70.21	79.70

表 2 各模型在 3D 模式的性能对比

Tab. 2 Performance comparison of models in 3D View mode

Models	Car AP/%				Pedestrian AP/%				Cyclist AP/%			
	Easy	Mod.	Hard	mAP	Easy	Mod.	Hard	mAP	Easy	Mod.	Hard	mAP
VoxelNet ^[12]	77.47	65.11	57.73	66.77	39.48	33.69	31.50	34.89	61.22	48.36	44.37	51.32
SECOND ^[10]	83.13	73.66	66.20	74.33	51.07	42.56	37.29	43.64	70.51	53.85	46.90	57.09
PointPillars ^[15]	79.05	74.99	68.30	74.11	52.08	43.53	41.49	45.70	75.78	59.07	52.92	62.59
PointRCNN ^[11]	85.94	75.76	68.32	76.67	49.43	41.78	38.63	43.28	73.93	59.60	53.59	62.37
ET-PointPillars ^[19]	88.67	77.76	77.00	81.14	50.10	44.30	41.22	45.21	73.40	56.19	52.25	60.61
EC-PointPillars ^[25]	85.99	77.66	72.51	78.72	53.47	44.68	42.75	46.97	79.51	62.85	57.11	66.49
PE-Pillar ^[36]	86.69	80.13	77.98	81.60	54.21	44.79	42.84	47.28	82.94	63.29	60.51	68.91
PV-RCNN ^[14]	90.25	81.43	76.82	82.83	52.17	43.29	40.29	45.25	78.60	63.70	57.65	66.65
PASS-PV-RCNN++ ^[29]	87.65	81.28	76.79	81.91	47.66	41.95	38.90	42.84	80.43	68.45	60.93	69.94
DensePointPillars ^[30]	84.60	75.46	68.43	76.16	52.76	45.38	42.63	46.92	75.95	62.36	54.98	64.43
Ours	88.56	81.53	77.75	82.61	55.52	44.37	43.65	47.85	83.64	69.07	61.65	71.45

改进算法与其他算法的精度对比,黑色加粗数据表示最优检测结果。结果表明,在多个目标检测任务当中,相较于其他算法,本文改进算法 SP-Pillars 的检测精度有明显提升,且大部分指标都高于其他算法,尤其是在小目标的检测中提升更明显。其中,在鸟瞰图模式下,汽车、行人、骑行者的 mAP 相较于基准模型 PEPillar 分别提升了 1.96%、3.59%、5.91%;在 3D 模式下,汽车、行人、骑行者的 mAP 相较于基准模型 PEPillar 分别提升了 1.01%、0.57%、2.54%。实验结果说明,将稀疏 Pillar 的自适应补全策略与点感知空间注意力模块的融合改进,有效地保留了更丰富的细粒度信息,实现了检测精度的提升。

三维目标检测的误差主要由分类误差和定位误差构成,前者表现为漏检或误检,后者表现为 3D 框的位置、尺寸或朝向偏差。BEV 模式下 IoU 的计算仅涉及 XY 平面,主要反映水平定位误差;3D 模式下 IoU 需同时满足三维空间的重叠约束,对 Z 轴定位精度和尺寸回归精度更为敏感。对比

SP-Pillars 在两种模式下的 mAP 提升幅度,BEV 与 3D 模式间的提升差异说明 SEPA 通过邻域真实点借用和虚拟点插值策略,在 XY 平面上有效补全了稀疏 Pillar 的水平几何结构,显著降低了水平定位误差和漏检率;而 3D 模式下提升相对受限,主要原因在于 Z 轴坐标校正对高度方向的几何重建精度仍有一定限制。

4.4 消融实验

为了验证所提各模块的有效性,基于 KITTI 数据集的 3 种目标类型进行了消融实验,进而分析它们对模型检测性能的影响。表 3 和表 4 为各模块在 KITTI 数据集集中的结果对比。实验中设置 4 组对照,分别为:基线;基线+SEPA;基线+PASA;基线+SP(本文方案),重点分析小目标 Pedestrian 及 Cyclist 的 AP 变化。

针对 Cyclist,结果表明:SEPA 单独作用时,BEV 模式下 mAP 提升 2.47%,3D 模式下 mAP 提升 1.44%,证明稀疏 Pillar 补全是精度提升的基础,通过真实点借用与虚拟点插值能够有效填

表 3 行人目标上的消融实验结果

Tab. 3 Ablation experiment results on pedestrian

实验设置	Pedestrian mAP/%		BEV 提升/%	3D 提升/%	参数量/M	推理速度/Hz
	BEV	3D				
Baseline	58.76	47.28	—	—	4.83	65.0
Baseline+SEPA	60.25	47.52	1.49	0.24	4.83	61.7
Baseline+PASA	59.61	47.49	0.85	0.21	4.83	63.3
Baseline+SP	62.36	47.85	3.6	0.57	4.83	60.2

表 4 骑行者目标上的消融实验结果

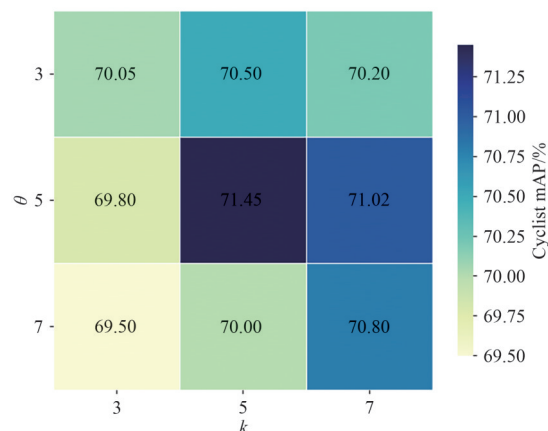
Tab. 4 Ablation experiment results on cyclist

实验设置	Cyclist mAP/%		BEV 提升/%	3D 提升/%	参数量/M	推理速度/Hz
	BEV	3D				
Baseline	73.79	68.91	—	—	4.83	65.0
Baseline+SEPA	76.26	70.35	2.47	1.44	4.83	61.7
Baseline+PASA	75.34	69.12	1.55	0.21	4.83	63.3
Baseline+SP	79.70	71.45	5.91	2.54	4.83	60.2

补信息空白;PASA 单独作用时,BEV 模式下 mAP 仅提升 1.55%,3D 模式下 mAP 提升 0.21%,原因是无补全时,稀疏 Pillar 的密度和距离特征无判别性,PASA 无法生成有效权重,验证无补全的注意力无效;双机制协同时,BEV 模式下 mAP 提升 5.91%,3D mAP 模式下提升 2.54%,协同效果显著,核心原因是 SEPA 为 PASA 提供高密度和短距离的可判别特征,PASA 放大补全后的目标信号,二者形成由补全到增强的闭环。

从效率维度来看,SEPA 单独引入使 FPS 下降 3.3 Hz,开销来源于稀疏 Pillar 的邻域筛选与虚拟点插值;PASA 单独引入使 FPS 下降 1.7 Hz,验证了轻量级 MLP 结构的高效性;二者协同时由于共享的 Pillar 索引计算,总体 FPS 下降仅 4.8 Hz。在模型复杂度方面,SEPA 不引入可学习参数,PASA 的轻量级 MLP 仅含约 0.6k 参数,二者协同后相较基线的参数增量小于 0.001M,在 Mod. 级精度下可视为零增长。最终模型在保持 60.2 Hz 推理速度的同时,精度和效率权衡显著优于单一模块方案,满足自动驾驶感知任务对实时性的要求。

稀疏 Pillar 的核心特征是点云数量较少,难以支撑有效特征的提取,但过少的量化标准需结合数据集特性动态确定。其中,稀疏判定阈值 θ 与最大借用点数 k 是两个重要的核心参数,需验证参数在合理范围内的稳定性。实验基于 KITTI 验证集,测试 $\theta \in \{3, 5, 7\}$ 、 $k \in \{3, 5, 7\}$ 时 Cyclist 的 mAP 变化。结果如图 4 所示: $\theta=5$ 时,mAP 达 71.45%(最优); $\theta=3$ 时,部分中度稀疏 Pillar 被误判,补全后引入冗余; $\theta=7$ 时,部分极度稀疏 Pillar 未被补全,信息缺失; $k=5$ 时,mAP 达 71.45%(最优); $k=3$ 时,补全点数不足; $k=7$ 时,超 Pillar 最大容量导致特征冗余。可见,参数在 [4, 6] 范围内性能波动 $<1.5\%$,方案稳定性强,无需调参

图 4 核心参数 θ 、 k 对 Cyclist 的 mAP 结果影响度。Fig. 4 Influence of core parameters θ and k on Cyclist's mAP results

即可适配不同场景。

为进一步验证自适应高度校正机制的有效性和相关超参数取值的合理性,本文在 KITTI 数据集上对自适应权重 α 的 3 个核心超参数 $\alpha_{\min}$ 、 $\alpha_{\max}$ 与 γ 分别开展敏感性分析实验。实验以行人和骑行者两类小目标为评估对象,考察各超参数在合理区间内变动时模型的 mAP 波动情况。 $\alpha_{\min}$ 控制极端差异下的最小校正强度,确保即使邻域与当前 Pillar 高度差异极大时仍保留基本的对齐能力,防止机制失效; $\alpha_{\max}$ 控制理想情况下的最大校正强度,避免在结构相似时过度对齐而抹平真实细节; γ 控制 α 对 Z 轴差异的敏感度,决定从强校正到弱校正的过渡速率。三者共同界定了 α 的有效作用区间与动态响应特性。

实验结果如表 5 所示,分别固定两个参数的取值,对第 3 个参数的取值进行调整。当 $\alpha_{\min}$ 取值为 0 时,意味着在 Z 轴差异极大情况下完全放弃校正,借用点的高度信息可能与当前 Pillar 完全脱节,导致小目标定位精度下降;当 $\alpha_{\min}$ 取值较大时,即使邻域结构与当前 Pillar 差异显著仍施

表5 α 超参数敏感性分析Tab. 5 Sensitivity analysis of the hyperparameters for α

实验设置		Pedestrian mAP/%		Cyclist mAP/%	
		BEV	3D	BEV	3D
固定 $\alpha_{\max}=0.8$ 、 $\gamma=2.0$	$\alpha_{\min}=0$	58.75	46.36	74.64	69.25
	$\alpha_{\min}=0.1$	60.15	47.23	76.84	70.83
	$\alpha_{\min}=0.2$	62.36	47.85	79.70	71.45
	$\alpha_{\min}=0.3$	61.49	47.61	78.21	71.30
固定 $\alpha_{\min}=0.2$ 、 $\gamma=2.0$	$\alpha_{\max}=0.7$	62.14	47.73	78.13	71.26
	$\alpha_{\max}=0.8$	62.36	47.85	79.70	71.45
	$\alpha_{\max}=0.9$	60.84	47.58	77.68	71.13
	$\alpha_{\max}=1.0$	59.26	46.87	76.32	70.14
固定 $\alpha_{\min}=0.2$ 、 $\alpha_{\max}=0.8$	$\gamma=0.5$	60.35	47.35	76.03	70.05
	$\gamma=1.0$	60.97	47.66	77.54	71.03
	$\gamma=2.0$	62.36	47.85	79.70	71.45
	$\gamma=3.0$	61.04	47.67	76.82	70.67
	$\gamma=5.0$	57.63	45.64	73.49	68.74

加较强校正,会抹平目标的垂直细节,不利于检测。 $\alpha_{\min}=0.2$ 时,模型在两类小目标上均取得较优表现,在保留差异信息与维持基本对齐之间取得了相对平衡。当 $\alpha_{\max}$ 较小时,即使邻域与当前Pillar高度匹配,校正强度仍偏弱,无法充分发挥邻域信息的对齐作用。当 $\alpha_{\max}=1.0$ 时,意味着在邻域可信赖度极高时实施完全对齐,过度依赖均值统计会导致目标内部的真实垂直变化被强制平滑。 $\alpha_{\max}=0.8$ 时,既充分利用了高可信邻域的对齐能力,又能避免过度平滑。 γ 过小时, α 对差异变化不敏感,自适应机制失去动态调节能力; γ 过大时, α 对微小差异即产生剧烈响应,反而损失对齐效果。 $\gamma=2.0$ 时, α 在Z轴差异0~1.5 m的典型区间内实现平滑过渡,与稀疏Pillar的实际几何分布特性相匹配。综合3组实验可见, $\alpha_{\min}$ 、 $\alpha_{\max}$ 与 γ 三者协同界定了自适应 α 的有效响应区间。所选基准配置在多组取值对比中均表现稳定,证明该参数组合能够较好地适配KITTI数据集中稀疏Pillar的几何分布特性。

4.5 可视化分析

为了更直观地显示算法的检测效果,本文开展了检测结果可视化对比实验,图5为不同点云场景下的检测结果,每组从上至下分别为RGB图片、基线模型结果图、本文改进模型结果图。

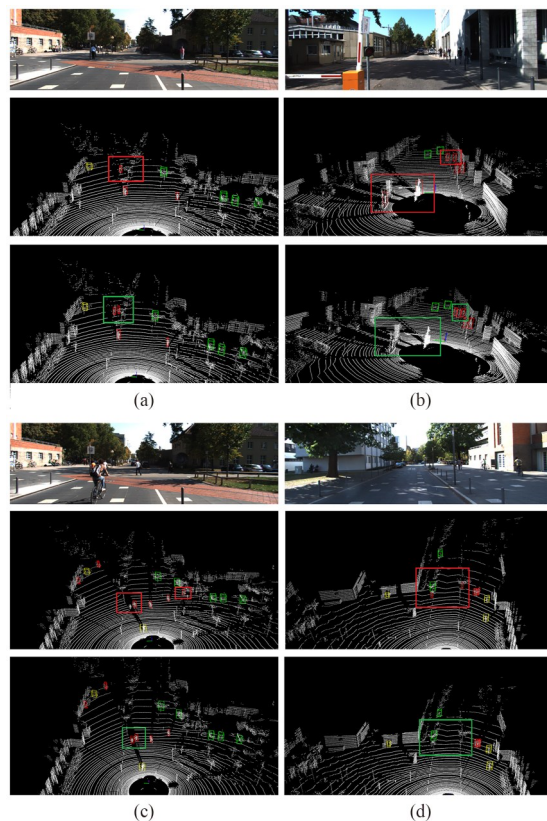


图5 点云检测结果。(a)出现漏检;(b,c)出现漏检和误检;(d)出现误检。

Fig. 5 Point cloud detection results. (a) False negative; (b, c) False negative and false positive; (d) False positive.

对于原基线模型,在图5(a)基线模型中,由于部分行人被其他行人遮挡,导致点云丢失,出现漏检现象。在图5(b)基线模型中,出现了行人的漏检现象,以及由于对比度较低而发生的将栏杆误检成行人的错误。在图5(c)基线模型中,行人因被骑行者遮挡,腿部的点云丢失,导致漏检;另一处因为光线问题出现了将栏杆误检成行人的情况。在图5(d)基线模型中,远处的栏杆因与行人形态相似,导致误检。

在上述4类典型场景中,本文提出的改进模型显著降低了误检与漏检频率。由结果图可见,该模型提升了对行人及骑行者等小目标的捕捉精度,且有效强化了不同类别目标间的辨识度。即便面临点云数据缺失或复杂环境噪声的干扰,本文模型仍能保持稳定的检测性能,有力证明了改进算法对目标特征提取机制的显著增强。

5 结 论

针对自动驾驶 LiDAR 感知任务中稀疏点云导致检测精度下降的核心问题,为解决现有方法补全不真实、注意力不适配的缺陷,本文在 PEPillar 算法基础上,提出了 SP-Pillar 协同框架,通过融合稀疏 Pillar 补全机制与点感知空间注意力机制以提高检测精度,并通过实验验证其有效性。在

KITTI 数据集上的实验结果表明,改进后的算法对行人、骑行者等小目标检测精度上有明显的提升,在鸟瞰图模式下,汽车、行人、骑行者的 mAP 分别提升了 1.96%、3.59%、5.91%;在 3D 模式下,汽车、行人、骑行者的 mAP 分别提升了 1.01%、0.57%、2.54%。同时,可视化结果验证了改进框架的有效性。该方案有效解决了稀疏点云检测的痛点,可为自动驾驶长尾感知提供实用技术支撑。

参 考 文 献:

- [1] ZHAO J Y, ZHAO W Y, DENG B, *et al.* Autonomous driving system: a comprehensive survey [J]. *Expert Systems with Applications*, 2024, 242: 122836.
- [2] 秦静,王伟滨,邹启杰,等. 基于激光雷达点云的 3D 目标检测方法综述[J]. 计算机科学, 2023, 50(S1): 220400214.
QIN J, WANG W B, ZOU Q J, *et al.* Review of 3D target detection methods based on LiDAR point clouds [J]. *Computer Science*, 2023, 50(S1): 220400214. (in Chinese)
- [3] ZHANG L Y, ZHANG C, ZHANG L, *et al.* A dual-mode LiDAR system enabled by mechanically tunable hybrid cascaded metasurfaces [J]. *Light: Science & Applications*, 2025, 14(1): 287.
- [4] YAO B C, WANG W T, XIE Z D, *et al.* Interdisciplinary advances in microcombs: bridging physics and information technology [J]. *eLight*, 2024, 4(1): 19.
- [5] SONG X N, ZHANG H Z, WANG L J, *et al.* Deep learning techniques for point cloud tasks: a review [J]. *Applied Intelligence*, 2025, 55(15): 991.
- [6] ZHENG Y, WANG X, JIANG Z, *et al.* Adaptive multiscale microscope with fast zooming, extended working distance, and large field of view [J]. *Light: Advanced Manufacturing*, 2024, 5(1): 62-74.
- [7] 孟毅儒,吕金光,郑凯丰,等. 采用多角度偏振光学相干层析成像的滑油磨粒三维形貌检测和重建[J]. 中国光学 (中英文), 2025, 18(6): 1449-1462.
MENG Y R, LV J G, ZHENG K F, *et al.* 3-D morphological feature measurement and reconstruction of wear particles using multi-view polarized optical coherence tomography [J]. *Chinese Optics*, 2025, 18(6): 1449-1462. (in Chinese)
- [8] 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. Hamburg, Germany: IEEE, 2015: 922-928.
- [9] QI C R, 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. Honolulu, HI, USA: IEEE, 2017: 77-85.
- [10] YAN Y, MAO Y X, LI B. SECOND: sparsely embedded convolutional detection [J]. *Sensors*, 2018, 18(10): 3337.
- [11] 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. Long Beach, CA, USA: IEEE, 2019: 770-779.
- [12] ZHOU Y, TUZEL O. VoxelNet: end-to-end learning for point cloud based 3D object detection [C]//*Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Salt Lake City, UT, USA: IEEE, 2018: 4490-4499.
- [13] DENG J J, SHI S S, LI P W, *et al.* VoxeL R-CNN: towards high performance voxel-based 3D object detection [C]//*Proceedings of the 35th AAAI Conference on Artificial Intelligence*. Online: AAAI, 2021: 1201-1209.
- [14] SHI S S, GUO C X, JIANG L, *et al.* PV-RCNN: point-voxel feature set abstraction for 3D object detection [C].

- 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2020: 10526-10535.
- [15] 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. Long Beach, CA, USA: IEEE, 2019: 12689-12697.
- [16] 田枫,刘超,刘芳,等. 基于改进 PointPillars 的激光雷达三维目标检测[J]. 激光与光电子学进展, 2024, 61(8): 0812007.
- TIAN F, LIU C, LIU F, *et al.* Laser radar 3D target detection based on improved PointPillars [J]. *Laser & Optoelectronics Progress*, 2024, 61(8): 0812007. (in Chinese)
- [17] LIU Z, ZHAO X, HUANG T T, *et al.* TANet: robust 3D object detection from point clouds with triple attention [C]// *Proceedings of the 34th AAAI Conference on Artificial Intelligence*. New York, NY, USA: AAAI, 2020: 11677-11684.
- [18] 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. (in Chinese)
- [19] LIU Y Y, YANG Z Y, TONG J L, *et al.* ET-PointPillars: improved PointPillars for 3D object detection based on optimized voxel downsampling [J]. *Machine Vision and Applications*, 2024, 35(3): 56.
- [20] LI J Y, LUO C X, YANG X D. PillarNeXt: rethinking network designs for 3D object detection in LiDAR point clouds [C]// *Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver, BC, Canada: IEEE, 2023: 17567-17576.
- [21] 高世严,柳杰,陈文艺,等. 基于深度学习的单目标跟踪算法研究进展[J]. 液晶与显示, 2025, 40(8): 1202-1218.
- GAO S Y, LIU J, CHEN W Y, *et al.* Review of single object tracking algorithm based on deep learning [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(8): 1202-1218. (in Chinese)
- [22] 王伯霄,宋延嵩,董小娜. 难点注意力感知红外小目标检测网络[J]. 中国光学(中英文), 2024, 17(3): 538-547.
- WANG B X, SONG Y S, DONG X N. Indistinguishable points attention-aware network for infrared small object detection [J]. *Chinese Optics*, 2024, 17(3): 538-547. (in Chinese)
- [23] SHU X N, ZHANG L. Research on PointPillars algorithm based on feature-enhanced backbone network [J]. *Electronics*, 2024, 13(7): 1233.
- [24] KONG W W, DU Y S, HE L L, *et al.* Improved 3D object detection based on pointpillars [J]. *Electronics*, 2024, 13(15): 2915.
- [25] 张淦,彭育辉,孙宝哲,等. 基于点云特征增强的 PointPillars 改进算法[J]. 激光与光电子学进展, 2025, 62(6): 0615012.
- ZHANG G, PENG Y H, SUN B Z, *et al.* Improved PointPillars algorithm based on point cloud feature enhancement [J]. *Laser & Optoelectronics Progress*, 2025, 62(6): 0615012. (in Chinese)
- [26] GUAN X L, WAN H, HAN W K, *et al.* MDS-PointPillars: a lightweight obstacle identification method in farmland based on three-dimensional LiDAR for autonomous navigation [J]. *Computers and Electronics in Agriculture*, 2025, 237: 110688.
- [27] YAO X P, LIU P Y, ZHOU J M, *et al.* MAT-PointPillars: enhanced PointPillars algorithm based on multi-scale attention mechanisms and transformer [J]. *PLoS One*, 2025, 20(6): e0325373.
- [28] 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)
- [29] CHEN S T, ZHANG H L, ZHENG N N. Leveraging anchor-based LiDAR 3D object detection via point assisted sample selection [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2025, 26(6): 7939-7952.
- [30] CHANDORKAR A, TERCAN H, MEISEN T. Rethinking backbone design for lightweight 3D object detection in LiDAR [C]// *Proceedings of 2025 IEEE/CVF International Conference on Computer Vision*. Honolulu, HI,

- USA: IEEE, 2025: 1709-1717.
- [31] 李佳男, 王泽, 许廷发. 基于点云数据的三维目标检测技术研究进展[J]. 光学学报, 2023, 43(15): 1515001.
LI J N, WANG Z, XU T F. Three-dimensional object detection technology based on point cloud data [J]. *Acta Optica Sinica*, 2023, 43(15): 1515001. (in Chinese)
- [32] SOHAIL S S, HIMEUR Y, KHEDDAR H, *et al.* Advancing 3D point cloud understanding through deep transfer learning: a comprehensive survey [J]. *Information Fusion*, 2025, 113: 102601.
- [33] 缪永伟, 刘家宗, 陈佳慧, 等. 基于生成对抗网络的点云形状保结构补全[J]. 中国科学: 信息科学, 2020, 50(5): 675-691.
MIAO Y W, LIU J Z, CHEN J H, *et al.* Structure-preserving shape completion of 3D point clouds with generative adversarial network [J]. *Scientia Sinica Informationis*, 2020, 50(5): 675-691. (in Chinese)
- [34] 潘李琳, 邵剑飞. 融合图注意力的多分辨率点云补全[J]. 激光技术, 2023, 47(5): 700-707.
PAN L L, SHAO J F. Multi-resolution point cloud completion fusing graph attention [J]. *Laser Technology*, 2023, 47(5): 700-707. (in Chinese)
- [35] ZHANG X S, HE Y, SONG C L, *et al.* Enhanced frustum multi-scale VoteNet for 3D object detection in cluttered indoor scene [J]. *Applied Intelligence*, 2025, 55(7): 588.
- [36] SUN L B, LI Y F, QIN W H. PEPillar: a point-enhanced pillar network for efficient 3D object detection in autonomous driving [J]. *The Visual Computer*, 2025, 41(3): 1777-1788.
- [37] LEE M, PARK S, KIM H, *et al.* SPADE: sparse pillar-based 3d object detection accelerator for autonomous driving [C]. 2024 IEEE International Symposium on High-Performance Computer Architecture. Edinburgh, United Kingdom: IEEE, 2024: 454-467.
- [38] PARK S, LEE M, CHOI J, *et al.* Selectively dilated convolution for accuracy-preserving sparse pillar-based embedded 3D object detection [C]//*Proceedings of 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle, WA, USA: IEEE, 2024: 8104-8113.
- [39] WOO S, PARK J, LEE J Y, *et al.* CBAM: convolutional block attention module [C]//*Proceedings of the 15th European Conference on Computer Vision*. Munich, Germany: Springer, 2018: 3-19.
- [40] BAJBAA K, USMAN M, ANWAR S, *et al.* Bird eye-view to street-view: a survey [J]. *Applied Intelligence*, 2025, 55(16): 1050.
- [41] 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.
- [42] KINGMA D P, BA J. Adam: a method for stochastic optimization [C]//*Proceedings of the 3rd International Conference on Learning Representations*. San Diego, CA, USA: ICLR, 2015.
- [43] 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, RI, USA: IEEE, 2012: 3354-3361.

作者简介:



陈祺东, 男, 博士, 副教授, 2020 年于江南大学获得博士学位, 主要从事 LLMs 微调及垂直领域落地应用方面的研究。
E-mail: cq_d_jnu@hotmail.com



吴豪, 男, 博士, 讲师, 2024 年于江南大学获得博士学位, 主要从事计算机视觉方面的研究。E-mail: wuhao940917@gmail.com