

DOI: 10.13382/j.jemi.B2508425

改进 PointRCNN 的路侧激光雷达目标检测算法*

常怀清¹ 汪贵平¹ 张凯¹ 赵杉盟² 关丽敏¹

(1. 长安大学电子与控制工程学院 西安 710064; 2. 陕西高速公路工程试验检测有限公司 西安 710086)

摘要:针对 PointRCNN 目标检测算法在复杂场景下特征提取能力不足的问题,首先,设计局部-全局注意力模块(LGAM),利用中心点与其相邻点之间的局部几何关系获得相邻点的注意力权重,实现了局部特征的有效融合,并基于全局双线性正则化方法,获得全局感知特征,后通过融合局部与全局特征实现协同优化;其次,设计多尺度卷积核注意力模块(MKCAM),通过并行标准卷积与膨胀卷积动态聚合多尺度特征,并结合基于通道池化的空间注意力机制,动态聚合多尺度特征;最后,将二者级联嵌入原 PointRCNN 的点云编码网络,增强网络的特征提取能力。此外,针对 PointRCNN 算法中传统非极大值抑制(NMS)固定交并比(IoU)阈值造成的误检问题,引入模糊 NMS,依据目标尺寸与场景密度动态分配 IoU 阈值。整合改进的点云编码网络与模糊 NMS 方法,提出了改进 PointRCNN 目标检测算法。实验结果表明,改进算法在 KITTI 数据集上,车辆、行人及骑行者检测精度分别提升 1.05%、3.43% 和 1.33%;在自建路侧数据集(稀疏点云场景)中,3 类目标精度分别提升 1.3%、2.71% 和 2.9%,验证了方法的有效性与泛化能力。

关键词: 三维目标检测; PointRCNN; 路侧感知; 点云; 特征提取

中图分类号: TP391.41; TN958.98

文献标识码: A

国家标准学科分类代码: 520.6070

Roadside LiDAR object detection based on improved PointRCNN

Chang Huaiqing¹ Wang Guiping¹ Zhang Kai¹ Zhao Shanmeng² Guan Limin¹

(1. College of Electronic and Control Engineering, Chang'an University, Xi'an 710064, China;

2. Shaanxi Expressway Engineering Test and Inspection Co., Ltd., Xi'an 710086, China)

Abstract: To address the insufficient feature extraction capability of the PointRCNN object detection algorithm in complex scenes, we first design a local-global attention module (LGAM). LGAM computes attention weights from the local geometric relationships between each central point and its neighbors, enabling effective fusion of local features. Simultaneously, global contextual features are captured via a bilinear regularization method, and local and global features are then fused for collaborative optimization. Next, we introduce a multi-scale kernel convolutional attention module (MKCAM), which dynamically aggregates multi-scale features by parallelizing standard and dilated convolutions and incorporates a channel-pooled spatial attention mechanism. Both LGAM and MKCAM are cascaded into the original PointRCNN point-cloud encoding network to enhance its feature extraction capacity. Furthermore, to mitigate misdetections caused by the fixed IoU threshold in traditional non-maximum suppression (NMS), we propose a fuzzy NMS that adaptively assigns IoU thresholds based on object size and scene density. By integrating the improved point-cloud encoder with fuzzy NMS, we present an enhanced PointRCNN algorithm. Experimental results on the KITTI dataset show accuracy improvements of 1.05%, 3.43%, and 1.33% for cars, pedestrians, and cyclists, respectively. On our self-collected roadside dataset with sparse point clouds, detection accuracies for the three classes increased by 1.3%, 2.71%, and 2.9%, respectively, confirming the effectiveness and generalization ability of the proposed method.

Keywords: 3D object detection; PointRCNN; roadside perception; point cloud; feature extraction

收稿日期: 2025-05-30 Received Date: 2025-05-30

* 基金项目: 西安市重点产业链技术攻关集群项目(25ZDLJQ00013)、车路一体智能交通全国重点实验室开放基金课题(2025-B009)、云南省交通运输厅科技创新项目(云交科教便[2023]178号)资助

0 引言

随着无人驾驶技术的快速发展,目标检测在智能交通系统中的重要性日益凸显^[1]。路侧感知设备突破单车感知的视距限制和场景局限性,成为解决自动驾驶盲区问题的关键一环^[2]。面对复杂交通场景,激光雷达凭借高精度、全天候感知和三维点云重建能力,成为路侧感知核心载体,其目标检测研究对车路协同体系建设具有重要价值。

近年来,深度学习在3D目标检测网络中的应用较为广泛,大致可分为基于视图、基于点、基于体素和基于点和体素融合4类。基于视图的方法通过投影点云生成2D伪图像,结合2D检测框架实现高效处理。PIXOR^[3]采用鸟瞰视角(BEV)投影与全卷积网络实现目标检测;为缓解投影中的信息损失,LaserNet^[4]通过DLA-Net保留原始数据分布,而Chai等^[5]和Fan等^[6]分别引入图操作符和元卷积核优化特征映射。此类方法虽在计算效率上具有优势,但其投影过程不可避免地损失3D几何信息,且难以解决遮挡、尺度变化与远距离稀疏区域的特征提取问题。基于体素的方法将点云规则化为体素网格,VoxelNet^[7]首创体素特征编码(VFE)层建立基础框架;针对远距离点云特征学习,Spherical^[8]采用球面坐标划分提升稀疏区域聚合能力;工程实践中,PointPillars^[9]通过柱状体素实现105 Hz的检测速度。此类方法通过规则化体素或柱状划分提升计算效率,但应用时难以平衡体素大小,并且在处理体素稀疏性方面计算复杂。基于点云和体素融合的方法结合两者优势:早期融合如PVCNN^[10]通过卷积聚合相邻体素特征,Guan等^[11]扩展多尺度融合;晚期融合以PV-RCNN^[12]为代表,采用体素候选与点云精修的两阶段结构,PV-RCNN++^[13]通过扇区化提案中心采样与VectorPool聚合进一步提升检测效率与精度,刘慧等^[14]和Hu等^[15]分别通过注意力机制与密度感知模块提升特征交互效率。但多阶段特征交互机制显著增加模型复杂度与计算资源消耗。相较而言,基于点的方法直接以原始点云为输入,最大程度保留几何信息,避免了投影或体素化导致的信息损失。以PointRCNN^[16]为代表的两阶段框架通过PointNet++^[17]提取多尺度点云特征生成高质量建议框,并结合局部空间特征优化检测精度。尽管基于点的方法在几何保真度方面具有优势,但其计算复杂度高、特征采样不均匀,在大规模点云场景中易出现特征表达不足与上下文建模能力弱等问题。近年来,射线分组策略^[18]与Transformer全局建模(如Pointformer^[19])在一定程度上缓解了上述缺陷,但对复杂交通环境中多尺度目标的检测仍存在部分问题。本文选择PointRCNN作为研究基础,其模块化结构

便于融合新型算子,但该框架仍存在两方面突出局限:1)在复杂场景下原网络依赖PointNet++结构进行局部特征提取,局部感受野有限,多尺度融合能力较弱,导致在稀疏、遮挡及噪声环境下特征表达不足,小目标与细节特征提取困难;2)原网络中非极大值抑制(non-maximum suppression, NMS)采用固定交并比(IoU)阈值,易造成小目标漏检和大目标误检,降低密集场景下的检测精度。

综上所述,为解决PointRCNN存在的问题,对其做如下改进。1)构建一种融合多尺度动态感受野与局部-全局协同优化的点云编码网络。首先,设计局部-全局注意力模块(local-global attention module, LGAM),利用中心点与其相邻点之间的局部几何关系获得相邻点的注意力权重,实现了局部特征的有效融合,并基于全局双线性正则化的方法,获得全尺度的全局感知特征,在此基础上,通过残差抑制冗余与非线性激活实现局部-全局特征协同优化;其次,设计多尺度卷积核注意力模块(multi-scale kernel convolutional attention module, MKCAM),通过并行标准卷积与膨胀卷积分解大卷积核,结合通道平均池化与最大池化生成空间注意力权重,动态聚合多尺度特征。两者以级联方式嵌入PointRCNN的点云编码网络,实现局部细节增强与全局语义约束的协同优化。2)引入了基于模糊分类的非极大值抑制方法^[20](Fuzzy NMS),通过模糊分类动态分配合适的阈值。最终,在上述两种改进方法的基础上构建改进的两阶段目标检测框架,提高了激光雷达目标检测的精度。

1 PointRCNN 基本原理

PointRCNN是基于原始点云的两阶段三维目标检测框架,网络结构如图1所示,其核心通过自底向上候选框生成策略,利用点云几何特征实现高精度定位。第1阶段采用PointNet++骨干网络提取逐点特征,结合前景点分割生成高质量3D候选框;第2阶段通过池化候选框内点云特征,融合局部坐标编码与全局语义信息,对候选框参数进行精细化修正^[21]。然而,该框架仍有如下局限性:1)复杂场景特征提取能力不足,受点云稀疏性与不规则性影响,原网络依赖PointNet++结构进行局部特征提取,但在面对稀疏区域或背景噪声较多的复杂环境时,局部感受野有限、多尺度融合能力弱,导致特征表达不充分,在小目标、遮挡目标及细节信息提取时存在困难。2)传统NMS采用固定IoU造成的误检问题,传统NMS采用固定IoU阈值,忽略目标尺寸差异与场景密度变化。小目标的轻微位置偏差易被误删,而大目标重叠区域易被误留,同时缺乏对类别差异和置信度分布的适应性,降低密集多目标场景的检测精度。

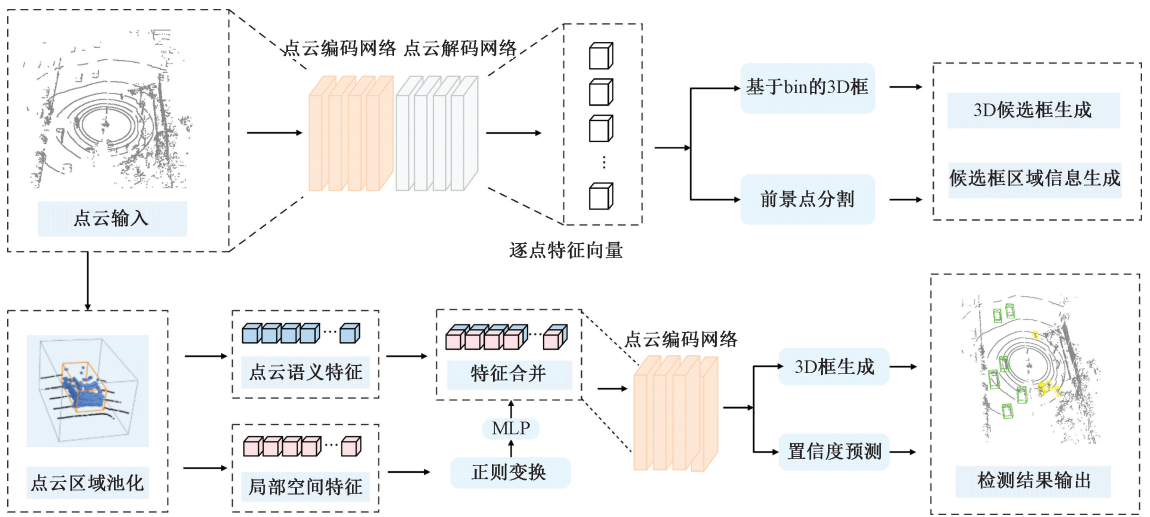


图 1 PointRCNN 网络结构

Fig. 1 PointRCNN network structure diagram

2 改进的 PointRCNN

2.1 网络整体结构

基于原 PointRCNN 算法中存在的问题,构建了一种改进的 PointRCNN 检测框架,其整体结构如图 2 所示。该框架将原始点云输入到融合多尺度动态感受野与局部-全局协同优化的点云编码网络,动态调整感受野并对逐点特征进行上下文感知增强,以更好地捕捉

目标周围的多尺度信息;接着通过前景分割剔除大部分背景点,再对每个前景点进行框回归,即生成初步的候选框;随后对每个候选框内的点云做池化处理,融合第 1 阶段的语义特征与基于相对坐标的局部空间特征,经点云编码网络优化检测框的中心、尺寸和朝向并预测置信度;最后引入基于模糊分类的非极大值抑制方法,通过动态阈值分配有效解决重叠目标分割困难和检测框冗余,最终输出高质量的 3D 检测框及其类别。

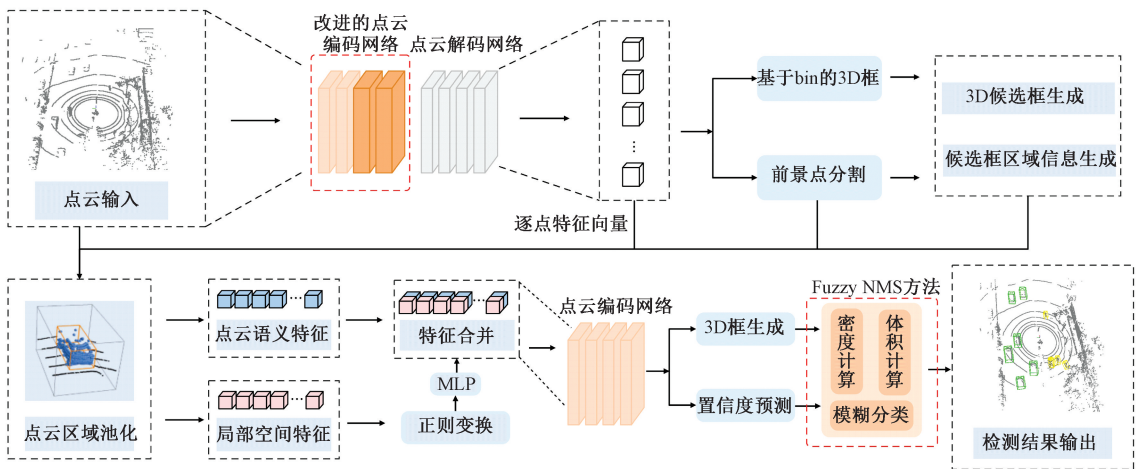


图 2 改进 PointRCNN 网络结构

Fig. 2 Improved PointRCNN network architecture diagram

2.2 改进点云编码网络

原始点云有着丰富的 3D 信息,但现有的特征提取网络难以将点云数据的原始信息进行充分利用,因此设计了 LGAM 模块来在特征提取时融合全局与局部信息。此

外,借鉴 2D 目标检测领域多尺度感受野的成功^[22],提出了一种适用于 3D 目标检测的 MKCAM 模块,通过增大感受野以获得更好的检测性能。

1) LGAM 模块

借鉴 Point Transformer^[23]的基本思想,利用中心点与其相邻点之间的局部几何关系获得相邻点的注意力权重,实现了局部特征的有效融合。在此基础上,引入了全

局双线性正则化方法^[24],通过考虑整个点云的全局感知来改进特征表达。LGAM 模块融合局部细节和全局语义,网络结构如图3所示。

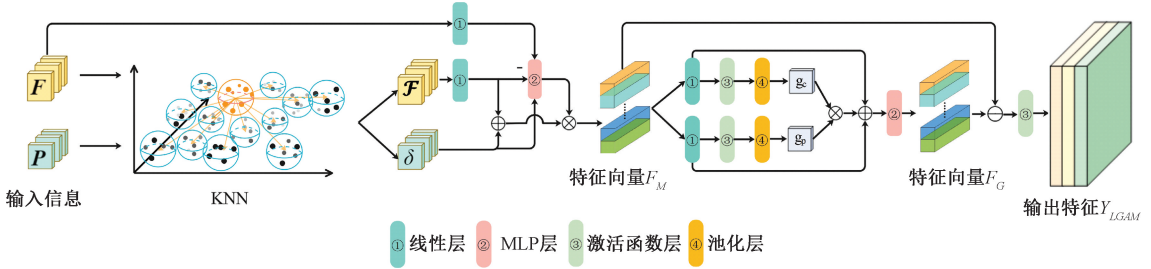


图3 LGAM 模块网络结构

Fig. 3 LGAM module network architecture diagram

对于一个有 N 个点的点云,其中第 i 个点的坐标可以表示为 $\mathbf{p}_i = (x_i, y_i, z_i) \in \mathbf{R}^3$,整个点云可以表示为 $\mathbf{P} = [\mathbf{p}_1, \dots, \mathbf{p}_i, \dots, \mathbf{p}_N] \in \mathbf{R}^{N \times 3}$ 。若将点 \mathbf{p}_i 对应 C 维空间上的特征向量表示为 $\mathbf{f}_i \in \mathbf{R}^C$,那么可以得到整个点云 \mathbf{P} 的特征图 $\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_i, \dots, \mathbf{f}_N] \in \mathbf{R}^{N \times C}$ ^[24]。利用 KNN 算法的思想,可得点云中某个点 \mathbf{p}_i 的 k 个最近邻点 \mathbf{p}_{ik} ,对点 \mathbf{p}_i 周围的最近邻点 \mathbf{p}_{ik} 的位置信息进行编码,得到点 \mathbf{p}_i 的局部几何图 $\mathcal{P}_i = [\mathbf{p}_i; \mathbf{p}_{ik} - \mathbf{p}_i] \in \mathbf{R}^{k \times 6}$,进而得到整个点云 \mathbf{P} 的局部几何图 $\mathcal{P} = [\mathcal{P}_1, \dots, \mathcal{P}_i, \dots, \mathcal{P}_N] \in \mathbf{R}^{N \times k \times 6}$ 。同理,利用 KNN 算法思想可得 \mathbf{f}_i 的局部特征图 $\mathbf{f}'_i = [\mathbf{f}_i; \mathbf{f}_{ik} - \mathbf{f}_i] \in \mathbf{R}^{k \times 2C}$,相应的,可以得到整个点云特征图 \mathbf{F} 的局部特征图 $\mathcal{F} = [\mathbf{f}'_1, \dots, \mathbf{f}'_i, \dots, \mathbf{f}'_N] \in \mathbf{R}^{N \times k \times 2C}$ 。

对局部几何图 \mathcal{P} 进行卷积和批标准化操作,将其转化为更能表达点云局部几何编码信息 δ 。针对局部特征图 \mathcal{F} 和整个点云的特征图 \mathbf{F} ,对其进行线性变换后计算其特征差异并结合局部编码信息,生成关系得分,进一步通过 *Softmax* 转换为注意力权重。在获得注意力权重之后,将局部几何编码信息与特征信息结合,应用注意力权重进行加权聚合,实现特征增强,得到局部上下文融合输出 F_M ,如式(1)所示。

$$F_M = \text{Softmax}(\gamma(\varphi(\mathbf{F}) - \psi(\mathcal{F}) + \delta)) \times (\alpha(\mathcal{F}) + \delta) \quad (1)$$

式中:映射函数 γ 是由两个线性层、一个归一化层和一个 *ReLU* 非线性激活函数构成的多层感知机,而 φ, ψ, α 都是简单的线性层。

为了更好地增强特征的表达性,引入了全局双线性正则化的方法,通过考虑整个点云的全局感知来改进特征映射。在全局双线性正则化的方法中,需获取全局通道特征。如式(2)所示,使用一个卷积层进行降维操作后通过 *ReLU* 激活函数为输出结果提供非线性,之后沿空间轴对 N 个元素进行平均池化操作,得到全局通道特

征 g_c 。

$$g_c = \text{avg}_N(\text{ReLU}(\text{Conv}^{1 \times 1}(F_M))) \quad (2)$$

为了得到全局点特征,使用另一个卷积层进行同样的操作,得到全局点的特征 g_p ,如式(3)所示。

$$g_p = \text{avg}_{C/r}(\text{ReLU}(\text{Conv}^{1 \times 1}(F_M))) \quad (3)$$

式中: r 是降维系数。

全局通道特征 g_c 可以获得通道依赖关系,而全局点的特征 g_p 可以表示全局空间结构信息。为了充分保留这两类全局信息,通过 g_c 和 g_p 外积的平方根来构建全局双线性特征 D ,如式(4)所示。

$$D = \sqrt{g_p \times g_c} \quad (4)$$

应用两个快捷连接和一个多层感知机 \mathcal{M}_F (具体实现为 1×1 的卷积、归一化层和激活层)恢复输出的维度,并且获得了全尺度的全局感知特征 F_G ,如式(5)所示。

$$F_G = \mathcal{M}_F(D + \text{Conv}^{1 \times 1}(F_M) + \text{Conv}^{1 \times 1}(F_M)) \quad (5)$$

为了进一步增强学习到的特征,通过从局部上下文融合 F_M 中减去全局感知特征 F_G 的方法来抑制冗余信息,并最终使用 *mish* 激活函数增强特征映射的非线性能力,LGAM 模块的输出如式(6)所示。

$$Y_{LGAM} = \text{mish}(F_M - F_G) \quad (6)$$

2) MKCAM 模块

MKCAM 模块通过并行标准卷积与膨胀卷积分解大卷积核,结合通道平均池化与最大池化生成空间注意力权重,动态聚合多尺度特征。其网络结构如图4所示,该模块不仅扩展了空间特征的捕获范围,还能通过并行的双路径结构和动态特征聚合机制,有效地处理不同尺度的空间信息。

为能够在进行特征提取时减少参数的数量,降低计算量,首先采用大卷积核分解的方法^[22],使用一个标准卷积核(kernel为5)和一个膨胀卷积(kernel为5,维度为3)代替更大的卷积核,分别获得不同的感受野特征。使

用 K_1 和 K_2 代表网络输入特征经过上述两种不同卷积操作后得到的特征,并将来自不同感受野范围的不同卷积核的特征进行拼接,得到特征 K ,如式(7)所示。

$$K = \text{Concat}[\text{Conv}^{1 \times 1}(K_1), \text{Conv}^{1 \times 1}(K_2)] \quad (7)$$

为获取不同特征空间之间的关联性与差异性,对特征 K 进行池化处理。如式(8)和(9)所示,对特征中同一位置上的各个权重采用通道平均池化和最大池化的策略,实现特征的有效融合,得到两个不同空间注意力图 L_{avg} 和 L_{max} 。

$$L_{avg} = f_{avg}(K) \quad (8)$$

$$L_{max} = f_{max}(K) \quad (9)$$

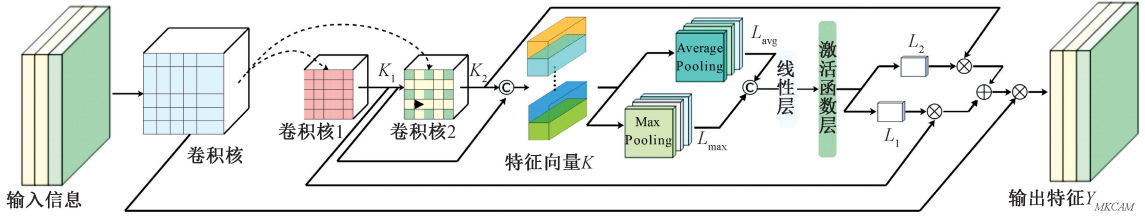


图 4 MKCAM 模块网络结构

Fig. 4 MKCAM module network architecture diagram

3) 点云编码网络设计

基于 LGAM 模块和 MKCAM 模块,构建一种融合多尺度动态感受野与局部-全局协同优化的点云编码网络。两者以级联方式嵌入 PointRCNN 点云编码网络,实现局部细节增强与全局语义约束的协同优化。

对于输入到网络中的点云,PointRCNN 目标检测网络使用具有下采样(set abstraction, SA)层的点云编码网络进行特征提取。虽然低层级的 SA 层更接近原始点云,具有更丰富的原始点云特征信息,但其点云数量过多,直接在此处进行增强特征提取的操作会带来过大的计算负担,SA 层级信息如表 1 所示。因此选择在第 3 层和第 4 层插入模块^[25],保证合理计算复杂度的同时,有效地融合更高维度的特征和语义信息。

表 1 SA 层级信息对比

Table 1 Comparison table of SA level information

层级	SA1	SA2	SA3	SA4
输入点数	16 384	4 096	1 024	256
输出点数	4 096	1 024	256	64
输出特征维度	96	256	512	1 024

改进后的点云编码网络如图 5 所示,主要由多个下采样层构成。针对第 3 与第 4 层下采样层,采用级联架构将输入特征降维后依次输入到两种不同的分支中进行处理,采用“拼接-卷积-残差”的策略进行结果融合。在进行特征拼接时,首先将多尺度特征和多尺度特征在通

为实现 L_{avg} 和 L_{max} 两个不同空间注意力图之间信息交互,对 L_{avg} 和 L_{max} 进行特征拼接,并将拼接后的池化特征进行卷积操作,再使用 Sigmoid 激活函数获得针对特征 K 的空间选择权重 L ,如式(10)所示。

$$L = \text{Sigmoid}(\text{Conv}^{1 \times 1}([L_{avg}, L_{max}])) \quad (10)$$

得到空间选择权重 L 后对来自不同感受野的特征进行对应的空间选择加权,并通过卷积层进行特征融合与维度变换,得到注意力特征。后将注意力特征与输入 F 相乘,获得最终输出,如式(11)所示。

$$Y_{MKCAM} = F \cdot \text{Conv}^{1 \times 1}(L_1 \times K_1 + L_2 \times K_2) \quad (11)$$

道维度上拼接,随后通过卷积层进行特征交互与降维,最后加入残差连接保留原始局部信息。在点云采样层中,首先使用最远点采样(FPS)算法选取其中关键点,以选中的关键点为中心,基于配置文件中的预设半径构建局部邻域,形成局部点群结构,基于局部点群结构进行特征提取操作。为了保证网络能够有效结合多层次特征,在下采样过程中,采用跳跃连接将上采样中的输出与对应维度的下采样层进行级联,增强网络的特征处理能力。

2.3 改进 NMS 方法

传统的 NMS 方法使用固定的 IoU 阈值来过滤重叠检测框,在处理不同尺寸目标时容易出现误检现象,尤其当处理密集检测目标时,往往很难在保留和抑制冗余检测框间取得良好的平衡。针对上述问题,引入了基于模糊分类的非极大值抑制方法,该方法考虑了候选框的体积和密度特征,通过模糊分类动态分配合适的阈值。对于每个候选框计算其体积特征 V_i ,如式(12)所示。

$$V_i = d_{x_i} \cdot d_{y_i} \cdot d_{z_i} \quad (12)$$

式中: d_{x_i} 、 d_{y_i} 、 d_{z_i} 分别表示第 i 个候选框在长宽高 3 个维度上的长度。

对于候选框集合,利用 DBSCAN 聚类算法进行密度聚类,计算第 k 个聚类的归一化密度值 D_k ,如式(13)所示。

$$D_k = \frac{N_k}{\max(N_0, N_1, \dots, N_k)} \quad (13)$$

式中: N_k 表示第 k 个聚类包含聚类框的数量。

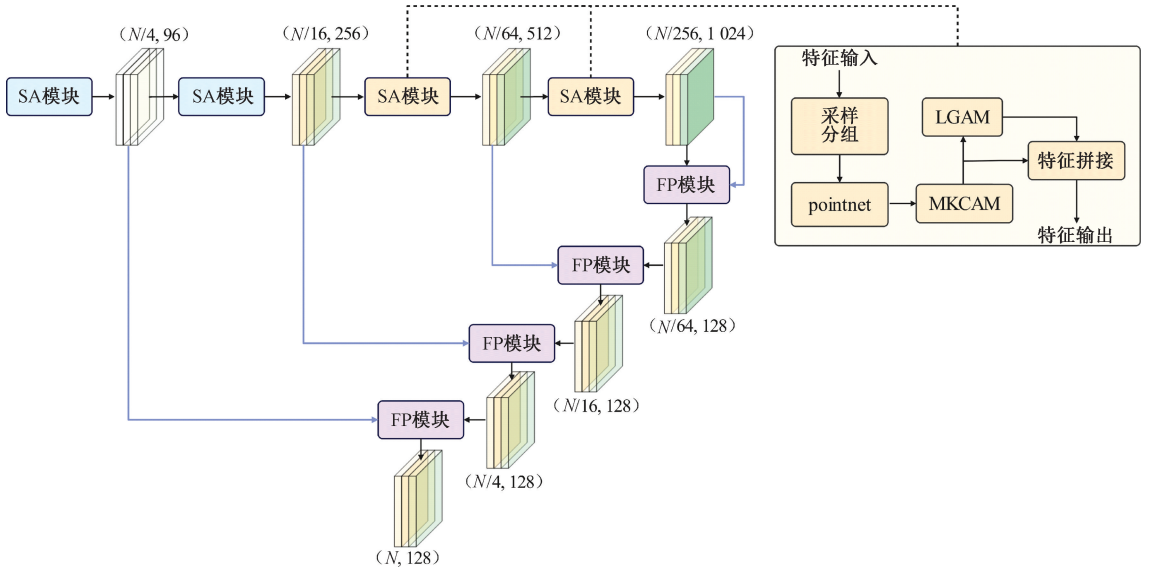


图 5 改进后点云编码网络

Fig. 5 Improved point cloud coding network

基于体积与密度特征,将候选框分为大体积高密度 (large volume high density, LVHD)、小体积高密度 (small volume high density, SVHD) 和低密度 (low density, LD) 3 类。对于 3 类不同的候选框, Fuzzy NMS 根据其不同特征动态调整 IoU 阈值, 具体设置如表 2 所示。这种通过场景特征动态调整 IoU 和置信度分数阈值的方法在处理密集场景时有明显优势, 能更好地提升算法在复杂场景下的检测性能。

表 2 Fuzzy NMS 模糊规则

Table 2 Fuzzy NMS Fuzzy rule table

阈值	LD	LVHD	SVHD
置信度分数	0.1	0.1	0.3
IoU	0.01	0.6	0.1

3 实验数据集

3.1 KITTI 数据集

KITTI 数据集^[26]作为自动驾驶领域的权威基准, 其标注涵盖车辆 (car)、行人 (pedestrian) 及骑行者 (cyclist) 3 类目标的 3D 边界框 (中心坐标、尺寸、航向角)。该数据集点云密度较低 (约 12 万点/帧), 目标尺寸分布不均 (以中型车辆为主), 且存在部分遮挡与截断样本, 适合验证算法在稀疏点云与复杂场景下的鲁棒性。在实验中将原训练集划分为 3 712 个训练样本与 3 769 个验证样本。

3.2 自建路侧数据集

目前基于激光雷达的点云目标检测常用的数据集有

KITTI 公开数据集或 Waymo 公开数据集, 其数据通过移动车辆搭载的激光雷达采集。然而, 路侧感知系统在传感器布设高度、观测视角及场景特征分布等方面与车载环境存在显著差异^[27]。为针对性验证本文方法在路侧应用中的有效性, 本文自主构建路侧场景点云数据集, 实现算法性能的评估。

为适应多场景数据采集, 本文选用速腾 16 线激光雷达作为路侧感知单元, 将其安装在高度为 1.6 m^[28] 的三脚架上。为了保证数据集的全面性, 选取多个具有不同特点的数据采集位置。如图 6 所示, 分别选西安市文艺南路、二环南路、求是路与创新路交叉口进行数据采集。3 个不同场景分别为城市次干路 (机动车、非机动车与行人多种交通参与者共存)、城市主干路 (车速较快且车辆密度大) 与支路 (目标以低密度、低速度为主), 3 种差异化场景的数据集采集能够让构建的数据集更具有广泛的适用性。

在数据采集阶段, 原始点云暂存为 .csv 格式以便初步校验, 后转换为 KITTI 标准 .bin 格式。转换过程中, 剔除雷达坐标系范围外的无效点云, 并对强度值进行线性归一化。标注环节基于 LabelCloud 平台, 严格遵循 KITTI 规范, 对车辆、行人及骑行者进行标注, 最终构建路侧数据集共 4 172 帧 (包括训练集 2 776 帧, 验证集 694 帧, 测试集 702 帧)。

4 实验结果与分析

4.1 实验环境

实验环境配置如表 3 所示。

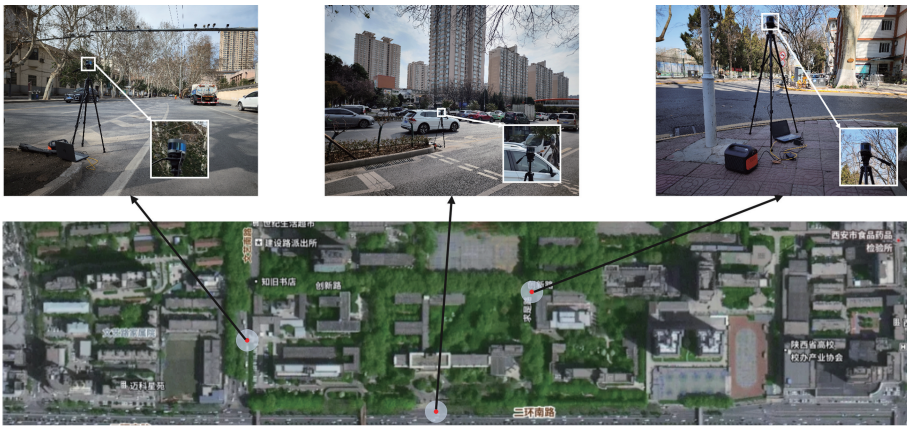


图 6 激光雷达布设

Fig. 6 Lidar layout diagram

表 3 实验环境配置

Table 3 Experimental environment configuration

软/硬件名称	版本配置参数
操作系统	Ubuntu 20. 04
CPU	Intel(R) Core(TM) i9-12900K
GPU	Nvidia RTX3090 GPU
内存	24 GB
CUDA	11. 1
编程语言	Python3. 9
深度学习框架	Pytorch1. 10. 1

4.2 评价指标

在 3D 目标检测任务中,平均精度(average precision, AP)是衡量算法检测性能的核心量化指标,其公式如式(14)所示。

$$AP = \frac{1}{40} \sum_{r \in \{0, \frac{1}{40}, \frac{2}{40}, \dots, 1\}} \max_{r' \geq r} Precision(r') \quad (14)$$

式中: AP 表示采用 R40 插值法计算得到的精度; r 表示召回率的固定采样点; r' 表示精度-召回率(P-R)曲线上某一个实际的召回率。

单个 AP 值是针对特定类别(如车辆、行人或骑行者)在指定 IoU 阈值下的性能度量,而平均精度均值(mAP)则是所有类别、所有 IoU 阈值下的 AP 取平均值,提供了一个更全面的性能评估指标。

4.3 KITTI 数据集实验结果与分析

为验证提出算法的性能和有效性,在 KITTI 数据集下进行评估。在定性评估方面,通过可视化方式展示算法在不同场景下的检测效果。在定量评估方面,设置车辆、行人和骑行者的 IoU 阈值为分别为 0.7、0.5、0.5,将改进后算法与复现的原始 PointRCNN^[16]、Pointpillars^[9]、PV-RCNN++^[13]和 Part-A^{2[29]}算法进行对比。不同算法在 KITTI 数据集上的 3D 模式下目标检测精度如表 4 所示。

表 4 不同算法在 KITTI 数据集上的 3D 目标检测精度

Table 4 3D object detection accuracy of different algorithms on the KITTI dataset

算法	不同算法在 KITTI 数据集上的 3D 目标检测精度 (%)				本文	
	Pointpillars ^[9]	PointRCNN ^[16]	Part-A ^{2[29]}	PV-RCNN++ ^[13]		
车辆	简单	86.03	87.97	87.12	91.42	88.74
	中等	76.36	77.93	75.90	83.28	78.44
	困难	72.01	75.98	75.37	81.12	77.85
	平均	78.13	80.63	79.46	85.27	81.68
行人	简单	52.78	61.18	59.19	58.99	64.42
	中等	47.62	53.68	52.05	52.86	56.11
	困难	42.08	45.97	47.59	48.60	50.59
	平均	47.49	53.61	52.94	53.48	57.04
骑行者	简单	82.62	86.91	84.23	85.75	87.33
	中等	64.79	72.06	71.12	70.42	72.08
	困难	61.58	66.26	65.07	66.06	69.80
	平均	69.66	75.08	73.47	74.08	76.41
mAP	65.09	69.77	68.62	70.94	71.71	

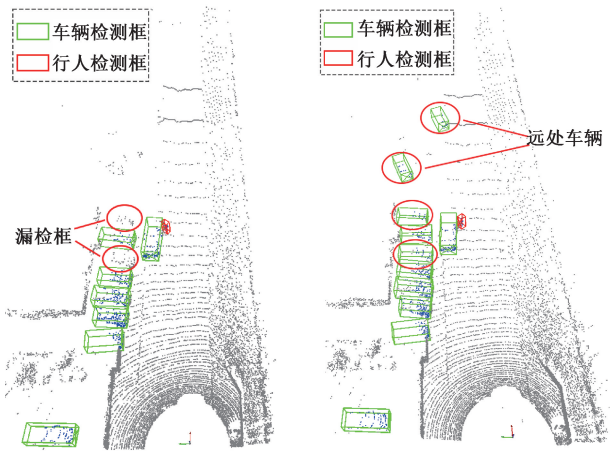
由表 4 可知,改进后算法相较原始 PointRCNN 算法有显著提升,整体 mAP 提升了 1.94%。行人目标通常因为形状不规则且尺寸较小而难以检测,改进后的算法平均检测率提升了 3.43%,说明算法在局部特征提取能力上有所提升。另外,在骑行者与车辆目标检测方面,改进后的算法分别比原算法提高了 1.33%与 1.05%。虽然在车辆检测方面精度略低于 PV-RCNN++,但在骑行者与行人目标检测方面有一定优势。总的来说,改进后的算法在 3D 模式下获得了较好的效果,尤其是在检测小目标方面效果显著。

为了更直观地对改进后算法的性能提升进行展示,对 3D 模式下检测结果进行可视化,如图 7 所示。从图 7 可以看到改进后算法与原算法相比,检测性能有所提升。在车辆停放密集区域,改进后算法不仅得到了更精确的检测框,还能检测出被遮挡车辆。对于远处车辆,虽然改进后算法未能检测出完全正确的车辆朝向,但相比较于

原算法,检测效果也有所改善,说明改进算法在处理远距离目标时的鲁棒性得到提升。



(a) 相机视角真实场景
(a) Camera perspective real scene image



(b) PointRCNN检测结果 (b) PointRCNN detection result chart
(c) 改进后算法检测结果 (c) Improved algorithm detection result chart

图7 3D模式检测结果对比

Fig.7 Comparison chart of 3D mode detection results

BEV模式通过将三维场景投影到地平面上,能够有效消除目标在深度方向上的遮挡,更直观地反映车辆间的相对位置关系,因此BEV模式下的性能评估也有一定意义。不同算法在KITTI数据集上的BEV模式目标检测精度如表5所示。

表5 不同算法在KITTI数据集上的BEV目标检测精度

Table 5 BEV object detection accuracy of different algorithms on the KITTI dataset (%)

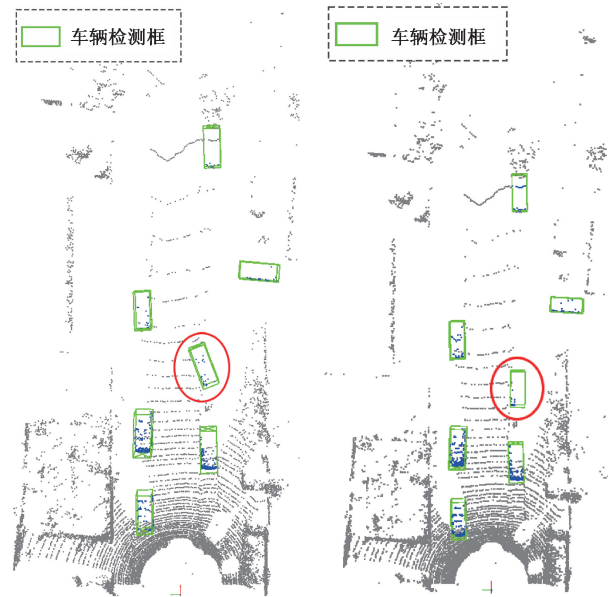
算法	Pointpillars ^[9]	PointRCNN ^[16]	Part-A ^[29]	PV-RCNN++ ^[13]	本文
简单	90.01	89.66	90.15	92.49	90.08
车辆 中等	87.93	86.13	87.46	89.57	87.49
AP 困难	85.80	80.07	86.46	88.75	86.86
平均	87.91	85.29	87.97	90.27	88.14
简单	60.07	63.39	62.64	63.98	68.04
行人 中等	54.94	59.17	57.77	58.21	60.58
AP 困难	48.78	52.33	52.45	53.92	53.18
平均	54.59	58.29	57.62	58.70	60.61
简单	84.07	88.55	84.23	87.56	88.59
骑行者 中等	68.84	74.64	71.12	74.30	74.39
AP 困难	64.62	68.42	65.07	69.86	71.15
平均	72.51	77.20	73.47	77.24	78.04
mAP	71.67	73.59	73.02	75.40	75.60

由表5可知,改进后的算法相较于原始PointRCNN在BEV视角下整体mAP增长了2.01%。其中,车辆、行人、骑行者分别增长了2.85%、2.32%、2.73%。相较于Pointpillars和Part-A²算法,改进后算法在各方面都有优势。虽然在车辆检测方面PV-RCNN++占有一定优势,但是整体mAP改进后的算法优于PV-RCNN++。总的来说,改进后的算法在BEV模式下获得了更好的效果,相较于其他算法也有一定的优势。

对BEV模式下检测结果进行可视化,如图8所示。通过对比发现,改进后的算法在检测框的方向估计上取得了明显进步,对于路边停放车辆,改进后的检测框更准确地反映了实际车辆的朝向,更符合真实场景中车辆的排列方向。



(a) 相机视角真实场景
(a) Camera perspective real scene image



(b) PointRCNN检测结果 (b) PointRCNN detection result chart
(c) 改进后算法检测结果 (c) Improved algorithm detection result chart

图8 BEV模式检测结果对比

Fig.8 Comparison chart of BEV mode detection results

基于KITTI数据集进行消融实验,验证改进特征提取网络与Fuzzy NMS方法的有效性。分别选取车辆、行人、骑行者的IoU阈值为0.7、0.5、0.5的条件下进行评估,实验结果如表6所示。

结果表明,当3个模块分别单独作用时,对检测效果都有一定提升,其中单独使用Fuzzy NMS时,由于行人目

标通常较小且容易出现密集重叠,可以更好地处理这类情况,因此对行人的检测精度提升最大。3 个模块均使用时 mAP 提升最大,说明同时采用 LGAM、MKCAM 和 Fuzzy NMS 时对目标检测网络改进更合理。

表 6 消融实验结果

Table 6 Results of ablation experiment (%)

MKCAM	LGAM	Fuzzy NMS	车辆 AP	行人 AP	骑行者 AP	mAP
			80.63	53.61	75.08	69.77
		✓	81.51	57.86	74.34	71.23
✓			81.83	54.29	76.03	70.93
	✓		81.29	54.43	75.13	70.28
✓	✓	✓	81.68	57.04	76.41	71.71

4.4 自建数据集实验结果与分析

为验证本文提出的方法在路侧场景的检测效果,基于自建路侧数据集对比原 PointRCNN 算法与改进后算法的性能表现,针对车辆、行人、骑行者选取 IoU 阈值分别为 0.5、0.25、0.25,实验结果如表 7 所示。

表 7 不同算法在路侧自建数据集上的 3D 目标检测精度

Table 7 3D object detection accuracy of different algorithms on self built roadside datasets (%)

算法	车辆 AP	行人 AP	骑行者 AP	mAP
PointRCNN ^[16]	86.11	77.99	69.26	77.79
本文	87.41	80.70	72.16	80.09

改进后算法相较于原算法,针对车辆、行人、骑行者分别提升了 1.3%、2.71%、2.9%,平均精度提升 2.3%。改进算法基于自建路侧数据集检测精度提升高于 KITTI 数据集,原因是自建数据集采集时使用 16 线激光雷达,而 KITTI 数据集使用的是 64 线激光雷达,自建路侧数据集点云更稀疏,特征更弱,因此增强特征提取的方法更有效。这一结果也进一步验证了改进后算法在路侧应用上的优势。

为全面评估所提方法在实际场景中的性能,在自建路侧数据集上进行可视化,如图 9~11 所示。

由图 9 可见,在自建路侧数据集下原算法出现了漏检现象,未能检测到红色框内的骑行者对象,改进后的算法成功检测到了目标对象。图 10 则针对重复检测现象进行可视化说明,原算法对同一目标产生了多个检测框,改进后算法无重复检测现象。由图 11 可见,原算法将路侧的草丛误检为骑行者,改进后算法则对环境有更强的鲁棒性,降低了误检率。

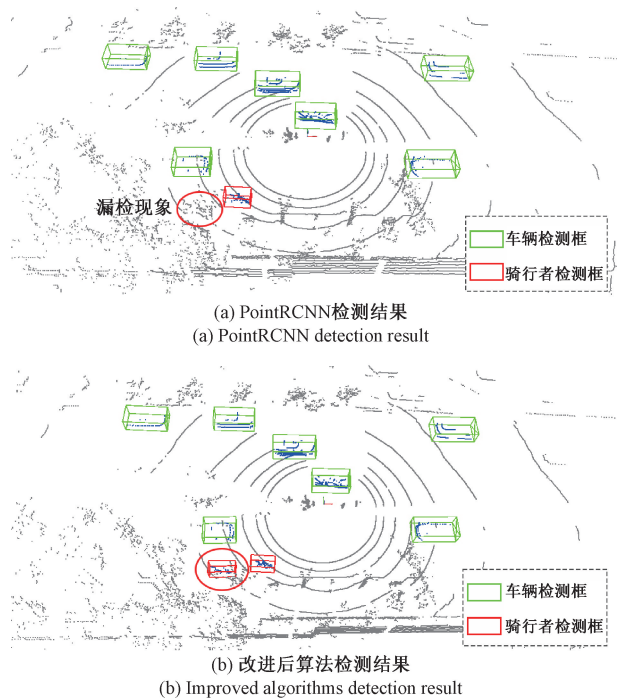


图 9 自建数据集 3D 模式检测漏检现象
Fig. 9 Self built dataset for 3D pattern detection of missed detections

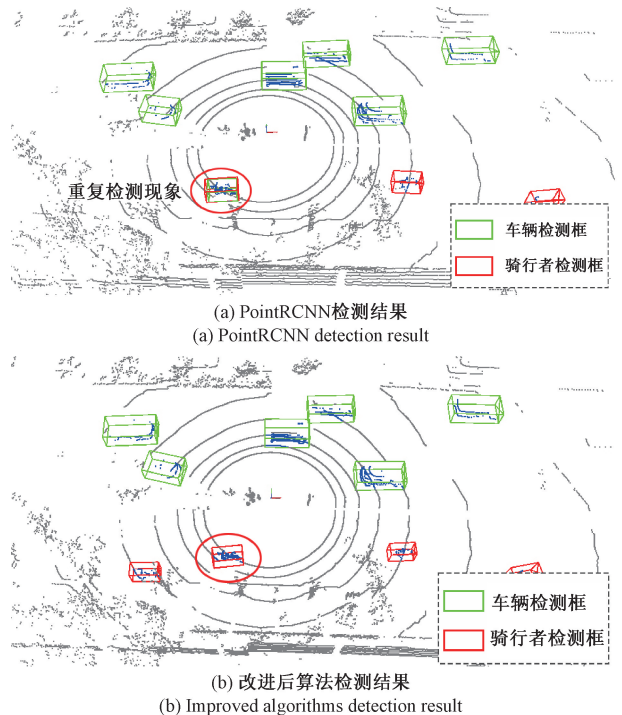


图 10 自建数据集 3D 模式检测重复检查现象
Fig. 10 Self built dataset for 3D pattern detection and repeated checking phenomenon

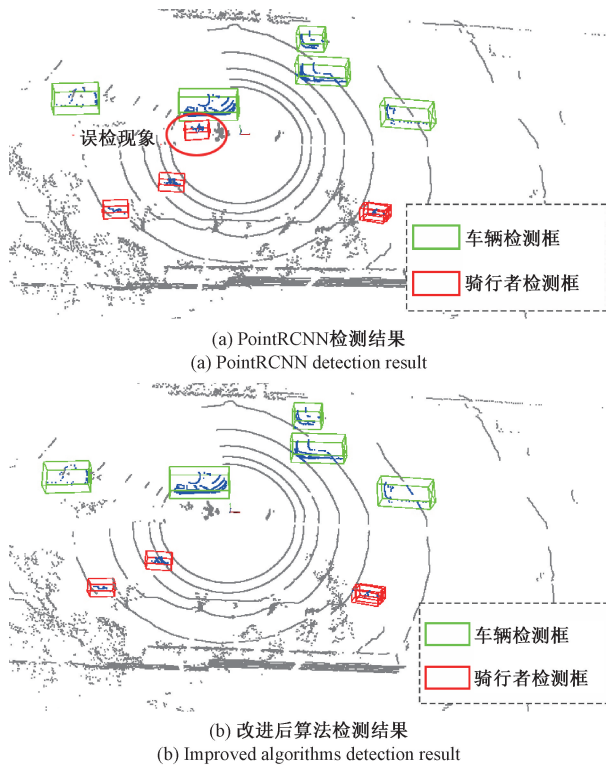


图 11 自建数据集 3D 模式检测误检现象

Fig. 11 Self built dataset 3D pattern detection false detection phenomenon

5 结 论

针对 PointRCNN 算法在复杂场景下特征提取能力不足的问题,提出了 LGAM 和 MKCAM 对 PointRCNN 的点云编码网络进行改进。LGAM 融合局部与全局特征实现局部-全局特征协同优化;MKCAM 扩展了空间特征的捕获范围,还能通过并行的双路径结构和动态特征聚合机制,有效地处理不同尺度的空间信息。二者级联嵌入原 PointRCNN 点云编码网络,增强了网络的特征提取能力。此外,引入了基于模糊分类的非极大值抑制方法,依据目标尺寸与场景密度动态分配 IoU 阈值,解决了 PointRCNN 中传统 NMS 固定阈值的误检问题。为验证改进算法的有效性,基于 KITTI 数据集进行对比实验与可视化分析,设计消融实验验证各个模块对算法的提升效果;并基于自建路侧数据集进一步验证改进算法的有效性。

参考文献

[1] 梁继然,陈壮,董国军,等. 结合注意力机制和密集连接网络的车辆检测方法[J]. 电子测量与仪器学报, 2022,36(3):210-216.
LIANG J R, CHEN ZH, DONG G J, et al. Vehicle detection method combining attention mechanism and

densely connected networks [J]. Journal of Electronic Measurement and Instrumentation, 2022, 36 (3): 210-216.

- [2] 兰浩然,李旭,朱建潇,等. 遮挡环境下基于路侧异源雷达融合的多交通目标鲁棒跟踪方法[J]. 仪器仪表学报,2022,43(11):31-39.
LAN H R, LI X, ZHU J X, et al. Robust tracking method for multiple traffic targets based on roadside heterogeneous radar fusion in occlusion environments[J]. Chinese Journal of Scientific Instrument, 2022, 43(11): 31-39.
- [3] YANG B, LUO W, URTASUN R. Pixor: Real-time 3d object detection from point clouds [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7652-7660.
- [4] MEYER G P, LADDHA A, KEE E, et al. LaserNet: An efficient probabilistic 3d object detector for autonomous driving [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12677-12686.
- [5] CHAI Y, SUN P, NGIAM J, et al. To the point: Efficient 3d object detection in the range image with graph convolution kernels [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 16000-16009.
- [6] FAN L, XIONG X, WANG F, et al. Rangedet: In defense of range view for lidar-based 3d object detection [C]. IEEE/CVF International Conference on Computer Vision, 2021: 2918-2927.
- [7] ZHOU Y, TUZEL O. VoxelNet: End-to-end learning for point cloud based 3d object detection [C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.
- [8] LAI X, CHEN Y, LU F, et al. Spherical transformer for lidar-based 3d recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 17545-17555.
- [9] 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.
- [10] LIU Z, TANG H, LIN Y, et al. Point-voxel CNN for efficient 3d deep learning [J]. ArXiv preprint arXiv: 1907.03739, 2019.
- [11] GUAN T, WANG J, LAN S, et al. M3detr: Multi-representation, multi-scale, mutual-relation 3d object detection with transformers [C]. IEEE/CVF Winter Conference on Applications of Computer Vision, 2022: 772-782.
- [12] SHI S, GUO C, JIANG L, et al. PV-RCNN: Point-voxel feature set abstraction for 3d object detection [C]. IEEE/

- CVF Conference on Computer Vision and Pattern Recognition, 2020; 10529-10538.
- [13] 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.
- [14] 刘慧,董振阳,田帅华.融合点云和体素信息的目标检测网络[J].计算机工程与设计,2024,45(9):2771-2778.
LIU H, DONG ZH Y, TIAN SH H. Object detection network integrating point cloud and voxel information[J]. Computer Engineering and Design, 2024, 45(9): 2771-2778.
- [15] HU J S K, KUI T, WASLANDER S L. Point density-aware voxels for lidar 3d object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 8469-8478.
- [16] SHI S, WANG X, LI H. PointCNN: 3d object proposal generation and detection from point cloud [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 770-779.
- [17] QI C R, YI L, SU H, et al. PointNet ++: Deep hierarchical feature learning on point sets in a metric space[J]. ArXiv preprint arXiv:1706.02413, 2017.
- [18] WANG H, SHI S, YANG Z, et al. RBGNet: Ray-based grouping for 3d object detection [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022; 1110-1119.
- [19] PAN X, XIA Z, SONG S, et al. 3d object detection with pointformer [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021; 7463-7472.
- [20] WANG L, ZHANG X, ZHAO F, et al. Fuzzy-NMS: Improving 3d object detection with fuzzy classification in NMS [J]. IEEE Transactions on Intelligent Vehicles, 2025, 10(1): 130-144.
- [21] 王庆林,李辉,谢礼志,等.基于激光雷达点云的车辆目标检测算法改进研究[J].电子测量技术,2023,46(1):120-126.
WANG Q L, LI H, XIE L ZH, et al. Improvement of vehicle object detection algorithm based on lidar point cloud [J]. Electronic Measurement Technology, 2023, 46(1): 120-126.
- [22] LI Y, HOU Q, ZHENG Z, et al. Large selective kernel network for remote sensing object detection [C]. IEEE/CVF International Conference on Computer Vision, 2023; 16794-16805.
- [23] ZHAO H, JIANG L, JIA J, et al. Point transformer [C]. IEEE/CVF International Conference on Computer Vision, 2021; 16259-16268.
- [24] QIU S, ANWAR S, BARNES N. PNP-3d: A plug-and-play for 3d point clouds [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023, 45(1): 1312-1319.
- [25] BHATTACHARYYA P, HUANG C, CZARNECKI K. SA-Det3D: Self-attention based context-aware 3D object detection [C]. IEEE/CVF International Conference on Computer Vision Workshops, 2021;3022-3031.
- [26] 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.
- [27] 熊文华,胡少鹏,王佩,等.车路协同下路侧交通设施体系及道路分级研究 [J]. 公路, 2022, 67(6): 218-222.
XIONG W H, HU SH P, WANG P, et al. Roadside traffic facility system and road classification under vehicle-road collaboration [J]. Highway, 2022, 67(6): 218-222.
- [28] ZHAO J, XU H, LIU H, et al. Detection and tracking of pedestrians and vehicles using roadside LiDAR sensors [J]. Transportation Research Part C: Emerging Technologies, 2019, 100: 68-87.
- [29] SHI S, WANG Z, SHI J, 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, 2020, 43(8): 2647-2664.

作者简介



常怀清,2019年于河南理工大学获得学士学位,现为长安大学硕士研究生,主要研究方向为自动驾驶感知与轨迹预测。

E-mail: 2024132030@chd.edu.cn

Chang Huaiqing received his B. Sc. degree from Henan Polytechnic University in 2019. He is now a M. Sc. candidate at Chang'an University. His main research interests include autonomous vehicle perception and trajectory prediction.



关丽敏(通信作者),2009年于长安大学获得学士学位,2012年于长安大学获得硕士学位,现为长安大学高级工程师,主要研究方向为智能汽车环境感知、智能交通控制、智能测控技术及应用等。

E-mail: guanlimin@chd.edu.cn

Guan Limin (Corresponding author) received her B. Sc. degree and M. Sc. degree both from Chang'an University in 2009 and 2012. She is now a senior engineer at Chang'an University. Her main research interests include intelligent vehicle environmental perception, intelligent traffic control, and intelligent measurement and control technologies and applications.