DOI:10.19650/j.cnki.cjsi.J2412517

基于 PointPillars 的改进三维目标检测算法*

汤新华,代道文,陈熙源,潘树国
(东南大学仪器科学与工程学院 南京 210096)

摘 要:基于激光雷达的目标检测技术在自动驾驶、机器人导航和无人机等领域得到广泛应用,由于激光雷达点云数据的稀疏性和不均匀分布,目标的检测和分类面临挑战。为此,本文提出一种基于 PointPillars 算法改进的三维目标检测算法,首先设计了更为有效的点云柱状特征编码网络,在编码网络中引入逐点和逐通道的双重注意力编码网络,提高每个 pillar 的特征表示能力。其次,在主干网络部分,融合全局上下文信息网络 GCNet 和 CSPDarknet 网络以提高特征图表征能力,使得网络在特征提取阶段能够更为充分地提取丰富的上下文语义信息。通过 KITTI 数据集进行了实验验证,相较于基准模型,改进方法具有更高的检测精度,在简单、中等和困难 3 种场景下,改进算法平均精度分别提升了 2.12%、2.51% 和 1.84%。同时,改进算法检测速度达到 35.6 FPS,证明了该方法在保持检测算法实时性的同时,有效地提高了检测精度。

关键词:目标检测;激光雷达;注意力机制;全局上下文;残差网络
中图分类号:TH701 **文献标识码**:A **国家标准学科分类代码**:460.40

Improved three-dimensional object detection algorithm based on PointPillars

Tang Xinhua, Dai Daowen, Chen Xiyuan, Pan Shuguo
(*School of Instrument Science and Engineering, Southeast University, Nanjing 210096, China*)

Abstract: LiDAR-based object detection technology is widely used in fields such as autonomous driving, robotic navigation, and drones. However, due to the sparsity and uneven distribution of LiDAR point cloud data, object detection and classification face significant challenges. Aiming at this problem, this paper proposes an improved 3D object detection algorithm based on the PointPillars algorithm. Firstly, a more efficient point cloud pillar feature encoding network is designed, incorporating a dual attention encoding network with point-wise and channel-wise attention, enhancing the feature representation capability of each pillar. Secondly, in the backbone network part, the global context information network (GCNet) and CSPDarknet network are integrated to improve the feature map representation ability, allowing the network to extract rich contextual semantic information more comprehensively during the feature extraction phase. Experiments conducted on the KITTI dataset demonstrate that the proposed method achieves higher detection accuracy compared to the baseline model, with mean Average Precision improvements of 2.12%, 2.51%, and 1.84% in easy, moderate, and hard scenarios, respectively. Additionally, the improved algorithm achieves a detection speed of 35.6 FPS, demonstrating that this method effectively enhances detection accuracy while maintaining real-time performance.

Keywords: object detection; LiDAR; attention mechanism; global context; cross stage partial

0 引 言

随着无人驾驶技术和人工智能的迅猛发展,环境感知算法为自动驾驶汽车后续决策和控制提供了必要的环境信息。激光雷达技术作为一种重要的环境感知手段,可以直接获取深度信息,能够高精度地获取目标物体的三维空间信息,因此在自动驾驶、机器人导航和无人机等领域得到了广泛的应用,成为目标检测和定位的重要传感器之一。然而,激光雷达目标检测面临着一系列的挑战和困难。首先,激光雷达数据呈现出稀疏性[1]和不均匀分布的特点,这对于目标的精确定位和形状恢复造成

收稿日期:2024-02-23 Received Date: 2024-02-23
*基金项目:国家十四五重点研发计划(2021YFB3900804)项目资助

了一定的困难。其次,复杂的背景噪声和遮挡现象对于目标的检测和分类带来了挑战[2]。因此,对于高效、准确的激光雷达目标检测算法的需求变得更加迫切。

目前点云检测主要有两大类方法:基于点的处理方法和基于体素的方法[3]。基于点的检测网络直接对原始点云进行特征提取然后生成 3D 检测框,PointNet[4]、PointNet++[5] 使用多层感知机(multilayer perceptron,MLP)学习点云特征,为基于点云的检测方法提供了可行的思路,但是其效率较低,不适用于实时高效的大规模自动驾驶场景。基于体素的 3D 检测器通常将非结构化点云变换为紧凑形状的规则柱体或体素网格,然后应用 3D 或 2D 卷积神经网络生成 3D 目标候选框[6]。Zhou 等[7] 提出的 VoxelNet 模型,它对输入点云进行密集体素化,首先在三维空间上将点云划分为一个个体素,对划分的每一个非空体素使用体素特征编码器进行局部特征提取。VoxelNet 的主要问题在于数据表示比较低效,中间层的 3D 卷积计算量太大,导致推理速度较慢,无法达到实时性的要求,因此后续很多工作针对其运行效率的问题进行了改进。Yan 等[8] 在 Voxelnet 基础上加以改进,使用稀疏卷积和子流形稀疏卷积进行特征提取,进一步提高了效率[9]。PointPillars[10] 算法进一步将体素简化为柱体,在每个柱体内提取特征并在高度维度进行压缩进一步生成伪图像,然后利用 2D 卷积进行检测,以低延迟实现了不错的性能,可以利用 2D 卷积以有限的成本部署在嵌入式系统上。在上述工作的基础上,Yin 等[11] 则针对三维空间中目标前进的方向变化复杂,物体的方位角难以预测的问题,设计出了一个 anchor-free 的算法 CenterPoint。该算法使用无锚框策略,通过目标中心点回归出目标的大小,方向及速度等信息。

PointPillars 算法能在保证实时性的同时取得较高的检测精度,本文以 PointPillars 算法为基准,提出了一种基于 PointPillars 算法改进的 3D 目标检测算法,进一步提高算法精度。首先在 pillar 特征编码网络引入逐点和通道注意力模块,以提升每个 pillar 的特征表示能力。此外,针对模型中骨干网络在特征提取方面的不足,在 CSPDarknet 网络基础上融合全局注意力模块,对二维卷积降采样模块进行优化,使得特征提取网络能够提取全局特征信息从而使特征图包含更丰富的上下文语义信息,进一步增强算法的特征提取能力,提高目标检测的准确性。

1　理论分析

1.1　网络结构

如图 1 所示为本文改进算法的结构图,主要基于 PointPillars 模型进行改进。算法主要由双重注意力柱状特征编码模块、特征提取网络和目标检测头等模块组成。该方法首先将原始点云转换成由柱体组成的网格,通过简化的 PointNet 网络进行特征提取,并通过双重注意力机制来提高捕获细节的能力,然后将原始点云投影到 X-Y 平面上,产生稀疏的 2D 伪图像,将二维伪图像送入改进的编码器中以学习多尺度空间特征,提高特征图的表征能力。最后,使用三维目标检测头来实现目标分类以及三维检测框的位置、尺寸大小和目标朝向角参数的回归。

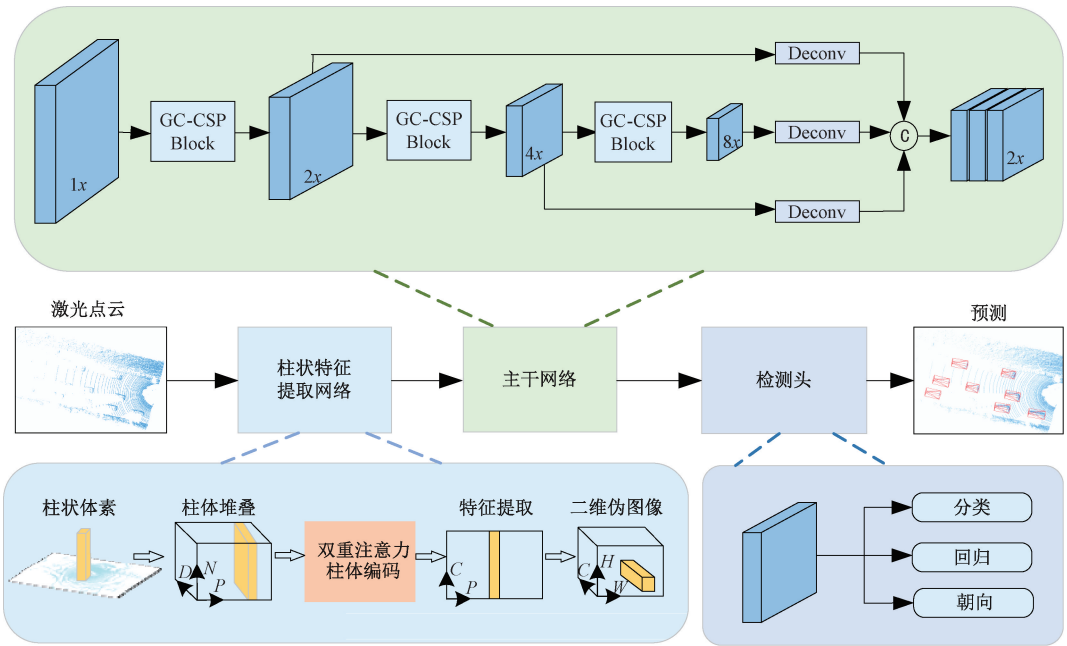


图 1　改进网络结构图

Fig. 1　The diagram of the improved network structure

1.2 双重注意力柱体编码

在点云柱状特征编码模块中,首先利用柱中心和点云范围的信息对原始点进行增强,然后通过多层感知机将增强的点特征映射到高维特征。在最大池化编码模块中,通过对每个 pillar 中的点特征进行最大池化操作,来提取每个体素的紧凑特征表示。最后,根据体素特征在网格中的原始空间位置来排列体素特征,形成大小为 $C \times H \times W$ 的伪图像。为了进一步提高 pillar 编码模块的特征提取能力,引入双重注意力模块,通过逐点和通道注意力机制来增强每个体素内目标的关键信息,同时抑制体素内的噪声点,其结构如图 2 所示。

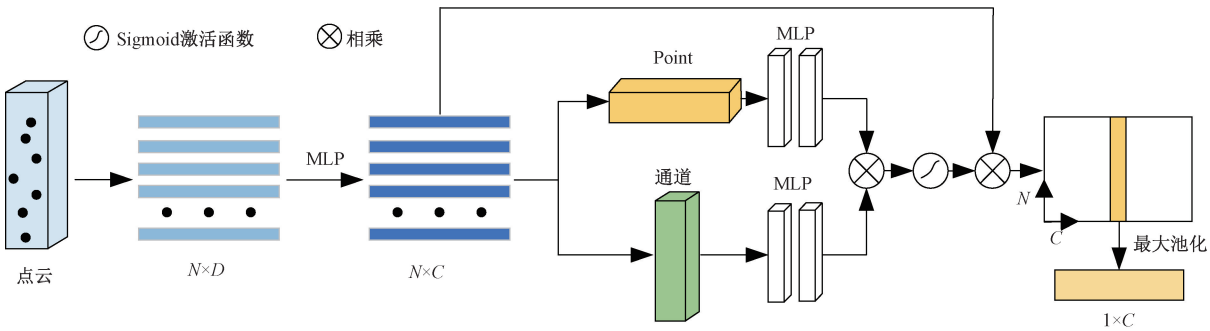


图 2 双重注意力 Pillar 编码

Fig. 2 Double attention pillar encoding

1)点云编码

在点云编码中,首先将点云平均划分为由 1 组体素组成的体素网格,该模块的输入是包含三维坐标和反射强度的原始点云,利用柱体中心和点云范围信息对原始点进行特征增强,然后利用 MLP 网络将扩充后的点云特征映射到特征空间[12]。首先对输入点云进行体素化编码,将 3D 空间划分为大小相同的 pillar 网格,设 $P = \{p_i = [x_i, y_i, z_i, r_i] \in R^{N\times 4}\}$ 是由 N 个点组成的柱体,p_i 为柱体中第 i 个点,x_i, y_i, z_i 分别为点云的三维空间信息,r_i 为点云的反射强度,且每个点的特征维度 $D = 4$。然后对每个柱体内的点云特征进行初步编码,根据每个 pillar 内输入点云的几何分布信息将柱体中每个点的初始特征扩充为较高维度的特征空间,记为 $p_i = \{[x_i, y_i, z_i, r_i, xc_i, yc_i, zc_i, xr_i, yr_i, zr_i] \in R^{N\times 10}\}$,其中 x_i, y_i, z_i 是每个点的三维坐标,r_i 为每个点的反射强度,xc_i, yc_i, zc_i 为该点所处的 pillar 中所有点的几何中心,xr_i, yr_i, zr_i 为每个点与几何中心的相对位置。然后通过 MLP 网络将每个 pillar 扩充后的点云特征映射到 64 维特征。

2)逐点注意力机制

在 PointPillars 的支柱特征网络层中,从 D 维原始数据学得的 C 维特征的感受野有限,导致提取特征的各个单元无法充分利用其局部区域之外的上下文信息。为了应对这一问题,本文侧重于全局空间关系的捕捉,引入逐点注意力模块,通过在点云中建立特征之间的关联,实现对全局依赖性的捕获。对于每个 pillar 内的编码后的点云,采用简化版的 PointNet 提取点云特征,并将其输入逐点注意力模块,借助逐点的注意力机制来捕捉柱体内每个点之间的空间相关性,避免冗余点云或噪声点对特征的影响,并提高对点云覆盖较少的特征描述[13]。通过最大池化层实现跨通道融合点的特征,从而计算出每个点的权重,随后通过 MLP 层进一步增强特征,其数学表达如下:

$$\boldsymbol{S}^k = MLP(MaxPool(\boldsymbol{F})) = \boldsymbol{W}_2\delta(\boldsymbol{W}_1(\boldsymbol{E}^k)) \quad (1)$$

式中:$\boldsymbol{F}$ 为输入特征;$\boldsymbol{W}_1, \boldsymbol{W}_2$ 是两层 MLP 网络的权重参数;δ 是 ReLU 激活函数;$\boldsymbol{E}^k$ 是最大池化后得到的特征;$\boldsymbol{S}^k$ 是计算得到的逐点注意力权重。

3)通道注意力机制

通道注意力模块通过在空间维度上对特征图进行压缩,利用通道注意力机制来学习每个通道内特征的重要性,通过权重分配来引导网络更加关注对目标检测有益的特征,同时抑制无关信息。该模块通过特征向量内通道之间的特征相互作用来生成通道注意力,其主要任务是确定输入特征向量中的“重要”通道部分。首先,对点云编码后的特征进行全局最大值池化操作,以降低空间维度,将特征映射为一维向量 $U^k \in R^{1\times C}$。该向量聚合每个通道上所有点的特征信息,然后将该一维向量输入到一个多层感知机组成的共享网络层。在共享网络层中,通过多层感知机来进一步处理聚合后的特征向量,通过学习适应性权重来处理输入数据并输出所需的通道注意力信息。通道注意力权重的计算如下:

$$\boldsymbol{T}^k = MLP(MaxPool(\boldsymbol{F})) = \boldsymbol{W}_2\delta(\boldsymbol{W}_1(\boldsymbol{U}^k)) \quad (2)$$

式中:$\boldsymbol{F}$ 为输入特征图;δ 是 ReLU 激活函数;$\boldsymbol{W}_1$ 和 $\boldsymbol{W}_2$ 是 MLP 网络的两个权重参数;$\boldsymbol{T}^k$ 是通道注意力权重。

融合逐点和逐通道注意力权重,采用并行注意力机制,则双重注意力权重 $\boldsymbol{M}^k$ 计算如下:

$$\boldsymbol{M}^k = \sigma(\boldsymbol{S}^k \times \boldsymbol{T}^k) \quad (3)$$

式中:σ 表示 Sigmoid 激活函数。

1.3　特征提取网络

PointPillars 的骨干网络通过卷积下采样提取多尺度特征,再经过上采样恢复到相同尺度进行特征融合。针对其主干网络特征提取能力不足的问题,本文采用了 CSPDarknet 网络[14]来强化主干网络,其是在残差网络基础上,通过引入跨阶段局部网络 CSP(cross stage partial)[15]优化网络结构,从而提高准确性和计算效率。此外,为了进一步提升网络的性能,在 CSPDarknet 网络基础上整合了全局上下文信息,将 CSPDarknet 网络和全局上下文模块(global context network, GCNet)[16]融合,其具体结构如图 3 所示。最后,将改进后的复合网络嵌入到 PointPillars 模型的骨干网络下采样模块中,通过自上而下的下采样以及相应的上采样在多个尺度上聚集特征信息,提高骨干网络的特征提取能力,以获得更丰富的特征信息。

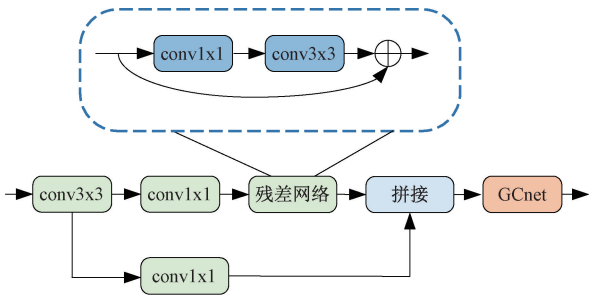


图 3　基于全局上下文的 CSPDarknet 网络

Fig. 3　CSPDarknet network based on global context

CSP 网络是一种跨阶段局部网络,通过将输入特征图进行划分,然后通过跨阶段层次结构将产生的两部分特征图合并,使梯度信息能够在不同网络路径之间传递,从而实现梯度的丰富组合,同时减少计算量[15]。基于 Darknet 网络结构的 CSPDarknet,融合了跨阶段局部结构,旨在实现目标检测模型的轻量化,同时确保计算效率和准确性的平衡。

在特征提取网络中,卷积神经网络通过滑窗方式在整个特征图上进行卷积操作,在捕获全局特征方面存在一定的局限性,相当于只学习到了局部的信息,当目标之间存在较远的关联时,卷积核只能观察到其卷积范围内的部分范围,无法有效捕获长距离依赖关系,从而限制了检测性能。因此,本文在 CSPDarknet 网络基础上融合全局特征信息,提出了一种基于全局上下文注意力网络的 GC-CSPDarknet,通过引入全局上下文网络提高模型对长距离依赖特征的提取能力,突出定位有利的特征并抑制无关噪声。

全局上下文网络属于全局注意力机制的一种应用,能够有效地建模全局上下文信息,并利用全局信息来增强局部特征图的表达能力。GCNet 的结构如下图 4 所示,它是一种残差结构的形式,输入特征首先通过全局上下文建模模块(context modeling)进行处理,接着经过特征转换模块(transform),最终与原始特征进行融合[16]。通过生成全局注意力特征图,GCNet 有助于使模型能够从全局角度关注感兴趣的区域,以实现对整体信息的充分利用,从而在一定程度上减少背景噪声导致的目标漏检和误检。GCNet 充分利用了 Non-local 结构的优势,并对其进行了简化,通过计算与查询位置无关的全局注意力特征图,并将其应用于所有查询位置,从而提高模型特征的多样性。通过计算全局上下文信息,模型更有效地聚焦于场景中的目标,减少背景噪声的干扰,简化后的 Non-local 模块可以表示为:

$$z_i = x_i + \boldsymbol{W}_v \sum_{j=1}^{N_p} \frac{\exp(\boldsymbol{W}_k x_j)}{\sum_{m=1}^{N_p} \exp(\boldsymbol{W}_k x_m)} x_j \tag{4}$$

式中:i 是特征图中查询位置的索引;j、m 枚举特征图中所有可能的位置;N_p 为特征映射中的位置数;x_i,x_j,x_m 表示输入;z_i 为模块输出特征,$\boldsymbol{W}_k$,$\boldsymbol{W}_v$ 表示线性变换矩阵。

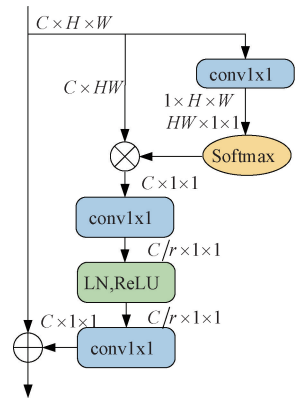


图 4　GCNet 网络结构图

Fig. 4　Structure of GCNet model

GCNet 具体计算过程如下,输入特征维度为 $C \times H \times W$,C 表示通道数,H 表示高度,W 表示宽度,特征一部分经过展平操作,维度变为 $C \times HW$,一部分经 1×1 卷积操作后维度变为 $1 \times H \times W$,继续展平为 $HW \times 1 \times 1$,然后通过 Softmax 函数得到注意力权重值,并将此权重作用于维度为 $C \times HW$ 的特征,从而计算得到全局注意力特征。特征转换模块采用 1×1 卷积进行通道压缩,特征经过激活函数层的处理,随后再次进行 1×1 卷积操作以将其恢复到原始特征维度,通过这种方式大幅减少网络参数数量。最后将得到的全局注意力特征图与原始特征进行充分融合,融合后的特征维度仍然为 $C \times H \times W$。将简化后得到的 Non-local 模块以及特征转换模块结合得到最终的表达式为:

$$z_i = F(x_i, \delta(\sum_{j=1}^{N_p} \alpha_j x_j)) \tag{5}$$

式中：$\sum_j \alpha_j x_j$ 表示全局上下文建模模块；δ 表示用于捕获通道间依赖的特征变换模块；F 表示将全局上下文特征和原始特征进行融合；z_i 表示融合后得到的输出特征。

进一步推导可得 GCNet 网络具体计算公式如下：

$$z_i = x_i + \boldsymbol{W}_{v2}\mathrm{ReLU}\left(LN\left(\boldsymbol{W}_{v1}\sum_{j=1}^{N_p}\frac{\exp(\boldsymbol{W}_k x_j)}{\sum_{m=1}^{N_p}\exp(\boldsymbol{W}_k x_m)}x_j\right)\right) \tag{6}$$

式中：$\alpha_j = \frac{\exp(\boldsymbol{W}_k x_j)}{\sum_m \exp(\boldsymbol{W}_k x_m)}$ 表示全局注意力池化的权重；ReLU 表示 ReLU 激活函数；LN 表示层归一化；$\boldsymbol{W}_k$，$\boldsymbol{W}_{v1}$，$\boldsymbol{W}_{v2}$ 表示线性变换矩阵。

1.4 损失函数

在检测头部分主要进行物体分类、包围框的回归以及目标朝向角的预测，因此损失函数主要包括 3 部分，即物体分类损失、位置回归损失、目标方向损失。每个三维检测框用一个 7 维的向量表示，分别为 (x,y,z,w,h,l,θ)，其中 x,y,z 为中心坐标，w,h,l 为尺寸数据，θ 为方向角。锚框和真值之间的残差定义如式(7) ~ (9) 所示。

$$\Delta x = \frac{x^{gt} - x^a}{d^a}, \Delta y = \frac{y^{gt} - y^a}{d^a}, \Delta z = \frac{z^{gt} - z^a}{h^a} \tag{7}$$

$$\Delta w = \log\frac{w^{gt}}{w^a}, \Delta l = \log\frac{l^{gt}}{l^a}, \Delta h = \log\frac{h^{gt}}{h^a} \tag{8}$$

$$\Delta\theta = \sin(\theta^{gt} - \theta^a) \tag{9}$$

式中：gt 表示物体包围框真值；a 表示锚框(anchor)；且有 $d^a = \sqrt{(w^a)^2 + (l^a)^2}$。

因此得到回归损失函数如式(10)所示，

$$L_{loc} = \sum_{b\in(x,y,z,w,l,h,\theta)} SmoothL1(\Delta b) \tag{10}$$

对于目标检测中正负样本数量极不平衡问题，目标的分类损失采用 Focal Loss 损失函数，如公式(11)所示：

$$L_{cls} = -\alpha_a(1 - p^a)^{\gamma}\log p^a \tag{11}$$

式中：p^a 是锚框的类别概率；α 和 γ 为权重因子，与原论文保持相同，取 $\alpha = 0.25, \gamma = 2$。

此外，为了避免方向判别错误，需对边界框的方向信息进行学习，额外引入方向损失函数 Softmax 损失学习物体的方向，其损失函数定义为 L_{dir}。则总体的损失函数如下：

$$L = \frac{1}{N_{\mathrm{pos}}}(\beta_{\mathrm{loc}} L_{\mathrm{loc}} + \beta_{\mathrm{cls}} L_{\mathrm{cls}} + \beta_{\mathrm{dir}} L_{\mathrm{dir}}) \tag{12}$$

式中：L 为总体的损失函数；N_{pos} 为正样本的数量；L_{loc}、L_{cls}、L_{dir} 分别为回归、分类和朝向角的损失函数，其比重大小分别为 $\beta_{\mathrm{loc}} = 2, \beta_{cls} = 1, \beta_{dir} = 0.2$。

2 公开数据集验证

2.1 实验数据集

本文使用 KITTI 公开数据集[17]进行算法的评估和验证，数据集包含真实驾驶场景的激光雷达点云数据和图像数据，该数据集中相关场景包括 7 481 个训练样本和 7 518 个测试样本，主要类别有车辆、行人和骑行者 3 类，训练样本又被分为训练集和验证集，其中训练集包含 3 712 个样本，验证集包含 3 769 个样本[18-19]，使用训练集进行训练，使用验证集进行实验验证。

实验中采用的评价指标为平均精度(mean average precision, mAP)来评估算法的综合检测性能，该值是由精确率和召回率共同计算得出，其计算公式如式 13、14 所示，设置不同的类别置信度可以绘制检测精度召回率曲线，即精度与召回率关系(precision-recall, PR)曲线，该曲线与坐标轴包围区域的面积可获得目标检测平均精度。

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{13}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{14}$$

式中：TP 表示将正类预测为正类的数量；FP 表示将负类预测为正类的数量；FN 表示将正类预测为负类的数量[20]。

如表 1 所示，数据集中各个对象的大小不同、物体在环境中受遮挡程度不同以及每个物体截断程度均不相同，因此将每个类别的目标划分为简单、中等、困难 3 种难度等级[21]。实验使用 KITTI 验证集来评估汽车、行人和骑行者 3 类的检测结果，汽车的交并比(Intersection over union, IoU)阈值设置为 0.7，行人与骑行者的 IoU 阈值均设置为 0.5。在精度 AP 计算时，采用 KITTI 数据集提供的官方度量标准 R40，R40 指采用 40 点插值法计算 PR 曲线面积，并将结果与 KITTI 鸟瞰(bird′s eye view, BEV)检测基准和 3D 检测基准上的其他模型进行比较。

表 1 3 种场景下的数据划分

Table 1 Data division in three scenarios

	简单	中等	困难
被遮挡程度	完全可见	部分遮挡	难以看见
截断值/%	15	30	50
最小边界框高度/pixels	40	25	25

2.2 实验参数设置

在本文实验中使用 PyTorch 深度学习框架和

mmdetection3d 目标检测框架[22],模型训练与验证的主机使用的处理器型号为 Intel(R) Xeon(R) Platinum 8255C,GPU 为 NVIDIA RTX 3090、显存为 24 G,操作系统为 Ubuntu 20.04。在公开数据集 KITTI 上进行模型训练时,使用 Adam 优化器,并采用单周期学习率策略进行训练,最大学习率为 0.000 3,权重衰减值为 0.01,动量值为 0.85 到 0.95。此外,使用 2 块 RTX 3090 GPU 训练,批量大小 Batch Size 设置为 8,Epoch 设为 80。在训练过程中采用了广泛使用的数据增强策略,包括随机场景旋转,随机场景缩放和随机平移[23]。此外,对每帧点云在 x、y、z 3 个方向上的检测范围分别设置为[0 m,69.12 m]、[-39.68 m,39.68 m]、[-3.0 m,1.0 m],同时使用(0.16 m,0.16 m)作为基本的柱体 Pillar 尺寸[24]。

2.3　实验结果分析

首先在 KITTI 数据集上对 BEV 视角及 3D 视角计算得到的指标进行对比,具体结果如表 2、3 所示。相较于原始的 Pointpillars 算法,本文所提出的改进算法在 KITTI 数据集上对汽车、行人和骑行者的检测结果表现出了较为显著的精度提升。与基准模型 PointPillars 相比,从 3D 视角进行评估,改进后的算法在汽车类别检测方面具有更高的精度,在简单、中等和困难 3 个不同难度的检测等级中,平均精度 AP 分别提高了 0.66%、2.37% 和 0.77%;对于行人检测,在 3 个难度级别上的精度也分别提升了 2.71%、2.82%、2.46%;同样地,在骑行者类别检测方面,在 3 种难度级别上均实现了性能提升,精度分别提升了 2.99%、2.34% 和 2.28%。如表 3 详细展示了在 BEV 视角下各算法平均精度对比情况,相比较于基准算法 Pointpillars,改进后的算法在检测汽车、行人和骑行者 3 类目标时,均表现出更高的精度,尤其对行人小目标效果显著,在简单、中等和困难 3 种等级下平均精度分别提升 2.99%、3.36% 和 2.68%。从上述分析可知,改进后的算法的检测结果不仅优于基准模型,在汽车、行人和骑行者的检测任务上,与其他算法相比也展现出了更优越的性能,这一系列的实验结果验证了改进算法的有效性。

表 2　KITTI 验证集不同检测方法 3D 平均精度

Table 2　3D average precision of different detection methods on the KITTI validation benchmark

	汽车			行人			骑行者		
	简单	中等	困难	简单	中等	困难	简单	中等	困难
SECOND[8]	89.55	79.67	74.34	59.60	52.13	46.60	80.36	63.57	59.65
3DSSD[25]	88.91	79.95	76.96	56.79	51.54	46.61	91.37	70.08	65.80
TANET[13]	88.21	77.85	75.62	70.80	63.45	58.22	85.98	64.95	60.40
Sazan et al[26]	87.44	78.10	75.21	56.55	50.56	46.23	83.94	65.58	61.41
PointPillars[10]	88.57	79.45	76.69	55.54	49.84	45.99	82.38	62.91	58.73
本文	89.23	81.82	77.46	58.25	52.66	48.45	85.37	65.25	61.01

表 3　KITTI 验证集不同检测方法 BEV 平均精度

Table 3　BEV precision of different detection methods on the KITTI validation benchmark

	汽车			行人			骑行者		
	简单	中等	困难	简单	中等	困难	简单	中等	困难
SECOND[8]	93.49	88.92	85.85	64.41	58.28	52.21	85.48	70.10	65.82
3DSSD[25]	93.02	88.97	86.31	62.06	55.98	51.69	94.87	73.24	68.80
TANET[13]	89.90	86.94	86.44	78.55	71.41	66.03	86.20	67.70	63.42
Sazan et al[26]	91.61	87.73	86.41	62.24	56.01	52.08	88.54	70.12	65.44
PointPillars[10]	92.55	88.56	85.93	60.98	55.23	51.38	87.81	67.84	63.91
本文	92.89	89.28	86.40	63.97	58.59	54.06	88.92	68.32	64.36

为了全面评估本章所提出的算法中各个模块的有效性,在 KITTI 数据集的验证集上进行一系列消融实验,即分别验证双重注意力体素编码模块、CSPDarknet 特征提取网络以及全局上下文信息模块对检测网络的性能提升,详细情况如表 4 所示。由表 4 消融实验结果可知,在 Pillar 编码网络中引入双重注意力机制,仅在牺牲极小的时间代价下,在简单、中等和困难 3 种模式下,3 个类别的平均精度分别提升 0.17%、0.74% 和 0.56%,实验说明

表 4 不同模块的消融实验结果

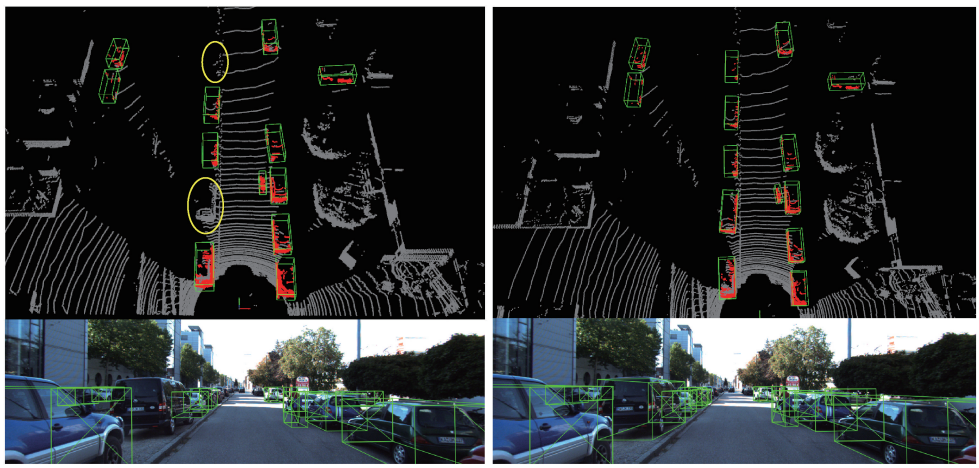
Table 4 The ablation experiment results for each module of the algorithm

双通道注意力	CSPDarknet网络	全局信息	简单	中等	困难	FPS
			75.50	64.07	60.47	50.1
Π			75.67	64.81	61.03	49.5
Π	Π		76.94	65.24	61.43	36.4
Π	Π	Π	77.62	66.58	62.31	35.6

这一方法在一定程度上解决了基于 Pillar 编码点云时可能出现的信息丢失问题。在此基础上,在 Pointpillars 算法的主干网络替换为 CSPDarknet 网络后,网络的特征提取能力进一步增强,3 个类别的平均精度均有提升。在 CSPDarknet 网络基础上进一步融合全局上下文网络,检测精度有了更进一步的提升,3 种模式下分别提升了 0.68%、1.34%和 0.88%,充分说明利用全局信息来增强局部特征图的表达能力,可以提取更丰富的语义信息,模型对特征图的表征能力进一步加强。综上所述,与原始 PointPillars 算法相比,改进后的算法 3 种模式下 3 个类别的平均精度分别提升了 2.12%、2.51%和 1.84%,且改进后算法的检测速度为 35.6 FPS,实验结果说明改进后的模型虽然在检测速度上受到了一定程度的影响,但其仍然能够达到实时检测的要求,较好地平衡了检测精度和检测速度。

2.4 实验结果可视化

图 5~6 所示为 KITTI 数据集上 PointPillars 算法与改进后算法的目标检测可视化结果,实验中使用了 Open3d 在空间中绘制了原始点云及算法检测出的三维目标包围框,图中的点表示激光雷达扫描得到的一帧原始点云,为便于观察仅绘制了前向视角点云,包围框表示目标检测结果。同时为了更好地可视化,将激光雷达检测到的三维检测框结果投影到来自左侧摄像头的图像上。



(a) PointPillars方法
(a) PointPillars method

(b) 本文方法
(b) Our method

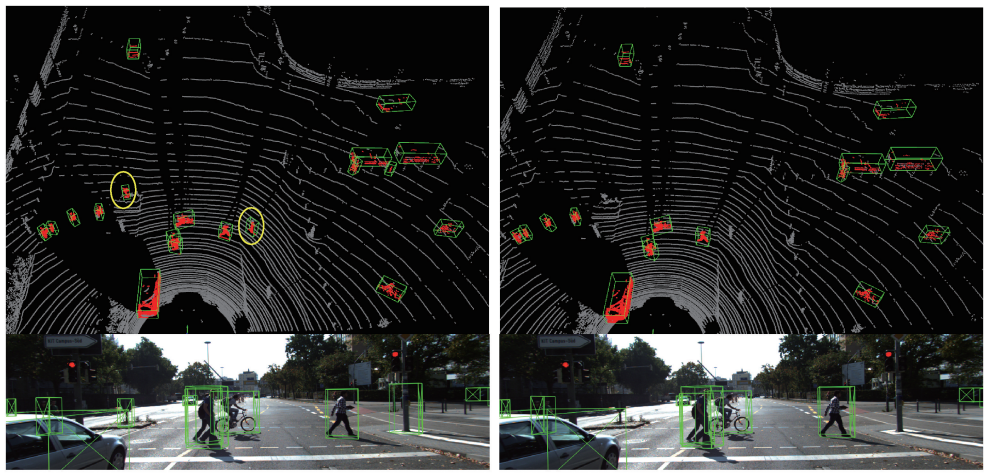
图 5 本文算法与 PointPillars 算法检测效果对比-遮挡场景

Fig. 5 Comparison of detection performance between this algorithm and PointPillars in occluded scenes

如图 5(a)中的可视化结果所示,漏检目标用椭圆框标记,原始的 PointPillars 算法对存在一定遮挡的汽车检测效果较差,存在较多漏检,而改进后的算法结果如图 5(b)所示,对于遮挡的汽车检测效果有所改善,即对漏检情况有一定改善。从图 6(a)可以看出 PointPillars 算法在复杂场景小目标检测上面存在较多的误检情况,误检情况在图 6(a)用椭圆框标识。PointPillars 算法把路边的路灯错误的检测成了行人,而改进后的算法通过全局特征的引入,特征图的表征能力更强,如图 6(b)所示可以在一定程度上避免这种情况。从以上分析可知,与基准 PointPillars 算法相比,改进后的算法在 KITTI 数据集上表现较好,能够对存在遮挡的车辆、行人等目标实现有效检测,改善了 PointPillars 算法目标误检和漏检较多的情况,检测效果显著提高。

3 实际场景实验验证

为了验证本文所提算法在实际场景中的效果,本文采用了无人小车实验平台进行实车试验。如表 5 所示,该平台主要包括 HUNTER 无人车底盘、NVIDIA Jetson AGX Xavier 处理器、RoboSense Helios 32 线激光雷达、XSENS MTi-G-710 惯性测量单元以及华测高精度紧组合



（a）PointPillars方法
（a）PointPillars method

（b）本文方法
（b）Our method

图 6　本文算法与 PointPillars 算法检测效果对比-复杂场景

Fig. 6　Comparison of detection performance between the proposed algorithm and PointPillars in complex scenes

惯导系统 CGI-430,其中激光雷达数据采集频率为 10 Hz,NVIDIA Jetson Xavier 处理器负责点云数据采集。

表 5　无人小车平台配置

Table 5　The configuration of the unmanned vehicle platform

传感器	型号
激光雷达	RoboSense Helios 32
处理器	NVIDIA Jetson AGX Xavier
IMU	XSENS MTi-G-710
RTK	华测 CGI-430

图 7　小车实验平台(左)及实验场景(右)

Fig. 7　Experimental platform(left) and experiment scene(right)

如图 7 所示,采用无人小车平台在实际道路中进行激光点云数据采集,实际行驶轨迹在常规交通道路上。为了进行有效的实验,将算法部署在 Ubuntu18.04 系统中,以机器人操作系统(robot operating system,ROS)作为软件框架进行实车试验,并使用 ROS 附带的可视化工具 rviz 来实现检测结果的可视化。

如图 8 所示为本文 3D 目标检测算法的实验效果,图中检测框为检测算法的输出结果,其主要包括了物体在三维空间中的位置信息及尺寸大小信息,可以看出本文算法能在实车实验中对场景中的目标实现有效的检测识别,为后续的可靠匹配里程计算法提供了很好的基础。

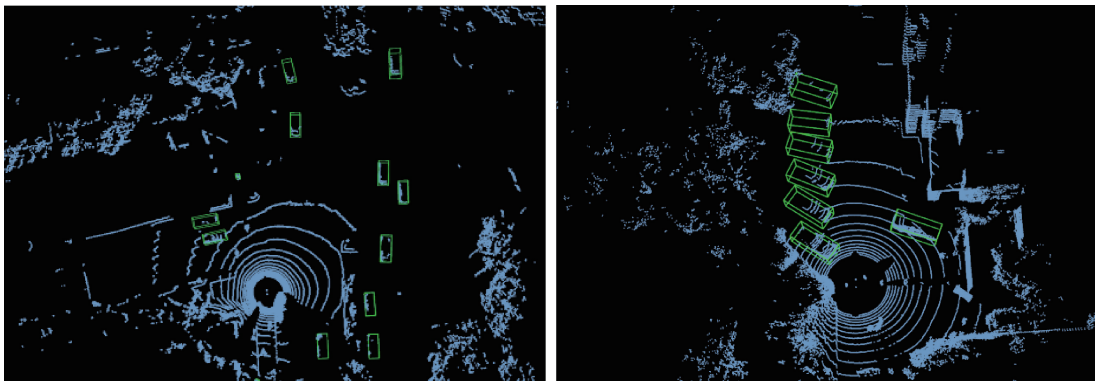


图 8　目标检测可视化结果一(左)和可视化结果二(右)

Fig. 8　Visualization result 1 (left) and visualization result 2 (right) of object detection

4 结 论

针对 PointPillars 算法的不足,本文提出了基于双重注意力和全局上下文信息改进的三维目标检测算法。通过双重注意力柱状特征编码网络,增强了每个体素内目标的关键信息,同时抑制体素内的噪声点。其次,在特征提取阶段,提出了基于全局上下文信息的 GC-CSPDarknet 网络,使得网络获得更加细节的多尺度空间上下文信息,提高了特征图的表征能力。本文通过在 KITTI 数据集上的实验验证了该算法的有效性,在简单、中等和困难 3 种条件下,本文算法平均精度 mAP 分别提升了 2.12%、2.51%和 1.84%,同时改进后的算法检测速度达到 35.6 FPS,满足实时性要求,较好地平衡了检测精度和检测速度。

参考文献

[1] 金宇锋, 陶重犇. 基于 Transformer 的融合信息增强 3D 目标检测算法[J]. 仪器仪表学报, 2023, 44(12): 297-306.
JIN Y F, TAO CH B. Fusion information enhanced method based on transformer for 3D object detection[J]. Chinese Journal of Scientific Instrument, 2023, 44(12): 297-306.

[2] 陈熙源, 戈明明, 姚志婷, 等. 雨雪天气下的激光雷达滤波算法研究[J]. 仪器仪表学报, 2023, 44(7): 172-181.
CHEN X Y, GE M M, YAO ZH T, et al. Research on LiDAR filtering algorithm for rainy and snowy weather[J]. Chinese Journal of Scientific Instrument, 2023, 44(7): 172-181.

[3] 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.

[4] QI C R, SU H, KAICHUN M, et al. Pointnet: Deep learning on point sets for 3d classification and segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 77-85.

[5] QI C R, YI L, SU H, et al. Pointnet++: Deep hierarchical feature learning on point sets in a metric space[J]. Advances in Neural Information Processing Systems, 2017, 30.

[6] LI J Y, LUO CH X, YANG X D. PillarNeXt: Rethinking network designs for 3D object detection in LiDAR point clouds[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 17567-17576.

[7] ZHOU Y, TUZEL O. Voxelnet: End-to-end learning for point cloud based 3D object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4490-4499.

[8] YAN Y, MAO Y X, LI B. Second: Sparsely embedded convolutional detection[J]. Sensors, 2018, 18(10): 3337.

[9] GUO D B, YANG G H, WANG CH H. PillarNet++: Pillar-based 3D object detection with multi-attention[J]. IEEE Sensors Journal, 2023, 23(22): 27733-27743.

[10] 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: 12689-12697.

[11] YIN T W, ZHOU X Y, KRAHENBUHL P. Center-based 3D object detection and tracking[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 11779-11788.

[12] ZHOU S, TIAN Z, CHU X, et al. FastPillars: A deployment-friendly pillar-based 3D detector[J]. ArXiv preprint arXiv:2302.02367, 2023.

[13] LIU ZH, ZHAO X, HUANG T T, et al. Tanet: Robust 3D object detection from point clouds with triple attention[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 11677-11684.

[14] WANG G, WU K. Remote sensing target detection based on improved YoloX [C]. Proceedings of the 2023 International Conference on Computer Vision and Intelligent Technology, 2023: 1-4.

[15] WANG C Y, LIAO H Y M, WU Y H, et al. CSPNet: A new backbone that can enhance learning capability of CNN[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 390-391.

[16] CAO Y, XU J, LIN S, et al. GCNet: Non-local networks meet squeeze-excitation networks and beyond [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, 2019: 1971-1980.

[17] GEIGER A, LENZ P, URTASUN R. Are we ready for autonomous driving? The kitti vision benchmark suite[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2012: 3354-3361.

[18] ZHENG W, TANG W L, CHEN S J, et al. Cia-ssd: Confident iou-aware single-stage object detector from point cloud[C]. Proceedings of the AAAI conference on artificial intelligence, 2021, 35(4): 3555-3562.

[19] 胡杰, 安永鹏, 徐文才, 等. 基于激光点云的深度语

义和位置信息融合的三维目标检测[J]. 中国激光, 2023, 50(10): 200-210.
HU J, AN Y P, XU W C, et al. 3D object detection based on deep semantic and positional information fusion of LiDAR point clouds[J]. Chinese Journal of Lasers, 2023, 50 (10): 200-210.

[20] 童小钟, 魏俊宇, 苏绍璟,等. 融合注意力和多尺度特征的典型水面小目标检测[J]. 仪器仪表学报, 2023, 44(1): 212-222.
TONG X ZH, WEI J Y, SU SH J, et al. Typical small target detection on water surfaces fusing attention and multi-scale features[J]. Chinese Journal of Scientific Instrument, 2023, 44(1):212-222.

[21] JHONG S Y, CHEN Y Y, HSIA C H, et al. Density-aware and semantic-guided fusion for 3D object detection using LiDAR-camera sensors[J]. IEEE Sensors Journal, 2023, 23(18): 22051-22063.

[22] LIS K, KRYJAK T. PointPillars backbone type selection for fast and accurate LiDAR object detection [C]. Proceedings of the International Conference on Computer Vision and Graphics. Cham: Springer Nature Switzerland, 2022: 99-119.

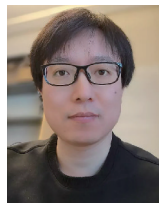
[23] GAN X, SHI H, YANG S, et al. MANet: End-to-end learning for point cloud based on robust pointpillar and multiattention[J]. Wireless Communications and Mobile Computing, 2022, 2022(1): 6909314.

[24] ZHANG L, MENG H, YAN Y B, et al. Transformer-based global PointPillars 3D object detection method[J]. Electronics, 2023, 12(14): 3092.

[25] YANG Z T, SUN Y N, LIU SH, et al. 3DSSD: Point-based 3D single stage object detector[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 11037-11045.

[26] MOHAMMED S, AB RAZAK M Z, ABD RAHMAN A H. Using efficient IoU loss function in PointPillars network for detecting 3D object[C]. Proceedings of the Iraqi International Conference on Communication and Information Technologies (IICCIT), 2022: 361-366.

作者简介



汤新华(通信作者),2007 年于南京航空航天大学获得学士学位,2010 年于东南大学获得硕士学位,2014 年于意大利都灵理工大学获得博士学位,现为东南大学副教授,主要研究方向为 GNSS 导航系统、多源融合定位系统、无人驾驶系统等。

E-mail:xinhua. tang@ seu. edu. cn

Tang Xinhua (Corresponding author) received his B. Sc. degree in 2007 from Nanjing University of Aeronautics and Astronautics, received his M. Sc. degree in 2010 from Southeast University, received his Ph. D. degree in 2014 from Polytechnic University of Turin. Now he is an associate professor in Southeast University. His main research interests include GNSS Navigation, Multi-Source integrated Navigation and Autonomous vehicle.