基于注意力机制的PointPillars三维目标检测方法研究

于亚涛[1]　翟宏瑞[2]

（1.浙江西图盟数字科技有限公司，宁波 315327；2.重庆工贸技师学院，重庆 400050）

【摘要】为解决现有的基于点云的三维目标检测算法对小目标检测效果较差的问题，在传统点柱（PointPillars）模型的基础上提出了一种基于注意力机制的PointPillars三维目标检测算法。首先，对PointPillars模型中的支柱（Pillar）特征网络进行改进，在编码网络中引入一组高效通道注意力（ECA）机制，使每个Pillar模块能够更充分地考虑其局部详细几何信息，从而显著提高每个Pillar模块的特征表示能力；其次，在主干网络中的二维卷积下采样模块中引入混洗注意力（Shuffle Attention）模块，通过提升特征表达的多样性和有效性，使模型网络聚焦于重要特征并抑制不重要或冗余特征，从而显著增强了算法的特征提取能力。在KITTI数据集上的测试结果表明，改进算法在三维行人检测的简单、中等和困难3种难度下的检测精度分别提升了4.04百分点、3.43百分点和1.93百分点，显著优于原始的PointPillars网络，行人小目标检测性能得到提升。

关键词：3D目标检测　PointPiilars　注意力机制　小目标检测

中图分类号：U495　　**文献标志码：**A　　**DOI：** 10.20104/j.cnki.1674-6546.20240438

Research on a PointPillars 3D Object Detection Method Based on Attention Mechanism

Yu Yatao[1], Zhai Hongrui[2]

(1. Zhejiang Xitumeng Digital Technology Co., Ltd., Ningbo 315327; 2. Chongqing Industrial & Trade Technician College, Chongqing 400050)

【Abstract】To address the issue that existing point cloud-based 3D object detection algorithms perform poorly on small object detection, a PointPillars 3D object detection algorithm based on an attention mechanism is proposed on the basis of the traditional PointPillars model. Firstly, the pillar feature network in the PointPillars model is improved by introducing a set of Efficient Channel Attention (ECA) mechanisms in the encoding network, allowing each Pillar module to better consider its local detailed geometric information, thereby significantly enhancing the feature representation capability of each Pillar module. Secondly, a Shuffle Attention module is introduced into the 2D convolutional downsampling module of the backbone network. By improving the diversity and effectiveness of feature representation, the model network focuses on important features while suppressing less important or redundant features, significantly enhancing the algorithm′s feature extraction ability. Test results on the KITTI dataset show that the improved algorithm improves the 3D pedestrian detection accuracy by 4.04 percentage points, 3.43 percentage points and 1.93 percentage points for easy, moderate, and hard difficulties, respectively, significantly outperforming the original PointPillars network and enhancing the performance of small pedestrian object detection.

Key words: 3D object detection, PointPillars, Attention mechanism, Small object detection

【引用格式】于亚涛, 翟宏瑞. 基于注意力机制的PointPillars三维目标检测方法研究[J]. 汽车工程师, 2026(5): 20-25.

YU Y T, ZHAI H R. Research on a PointPillars 3D Object Detection Method Based on Attention Mechanism[J]. Automotive Engineer, 2026(5): 20-25.

1　前言

激光雷达技术和深度学习算法的进步显著提升了三维目标检测的性能，不仅增强了自动驾驶系统的路径规划能力，也大幅提高了其在复杂环境下的安全避障能力[1]。激光雷达点云处理方法主要分

为基于原始点云和基于体素两类，点柱（PointPillars）模型[2]是一种基于体素的方法，它将点云立柱转换为伪图像，在保留三维特征的同时用二维特征提取高维特征，运行效率较高，但对小目标和较远目标的识别效果不佳。

为提升小目标的识别效率，有学者对PointPillars进行了改进。陈德江等[3]使用Swin Transformer改进PointPillars的二维卷积降采样结构，增强了网络的特征提取能力，但其仅对目标朝向的识别精度有所提升，并未涉及三维检测框的准确率。詹为钦等[4]引入串行和并行注意力机制，通过对伪图像的特征信息进行放大和抑制，减少了误检的情况，但没有考虑各支柱（Pillar）点云间的相互关联。贾远鹏等[5]在网络中引入通道注意力机制和行、列注意力机制，取代传统聚类方法，在TuSimple数据集上进行训练和测试，检测精度得到提高。王盈丰等[6]通过在PointPillars中引入双重注意力模块、全局注意力模块等机制，使改进后的算法获得了良好的检测精度和鲁棒性。

针对PointPillars现有改进方法的不足，本文提出基于改进PointPillars的激光雷达点云三维目标检测算法，在Pillar特征网络中引入高效通道注意力（Efficient Channel Attention，ECA）模块，提高Pillar模块的特征表示能力，在骨干网络下采样模块中引入混洗注意力（Shuffle Attention，SA）模块，使其聚焦于更重要的特征，进一步提高算法的检测能力。

2 基于深度学习的目标检测网络模型

2.1 网络模型

PointPillars采用体素化方法将原始点云转换成结构化的伪图像，以实现目标检测，其由3个部分组成：支柱特征网络（Pillar Feature Net，PFN）、主干网络（Backbone）和检测头（Detection Head，DH）。支柱特征网络负责提取点云的基本特征，主干网络进一步处理和提炼这些特征，检测头最终输出检测结果。具体操作流程如图1所示：首先，将点云按照正方形网格进行划分，使其被转换成支柱；接着，通过特征映射网络将支柱转换为伪图像；然后，使用二维卷积网络对伪图像进行多次降采样，获得不同分辨率的特征图，并将这些特征图上采样到相同的大小进行拼接，从而生成最终的特征图；最后，将生成的特征图输入单次多框检测器（Single Shot MultiBox Detector，SSD）[7]目标检测头，进行包围框（Bounding Box）回归和目标分类。

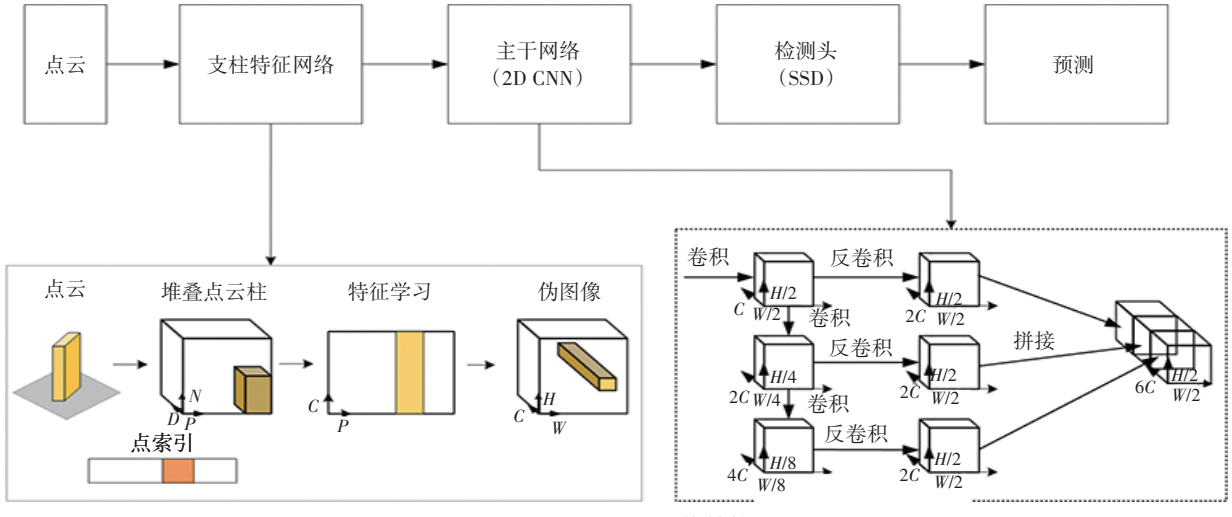


图1 PointPillars网络结构

2.2 网络模型的局限性

PointPillars通过采用支柱表征的方法，提高了检测速度和数据处理效率。然而，这种方法在减少数据处理量的同时，也导致了检测精度的降低。在实际应用时，PointPillars的检测精度还有一定的提升空间。为了解决这一问题，本文从支柱特征提取和主干网络两个方面对现有的PointPillars进行优化。

3 基于PointPillars的目标检测算法改进

PointPillars算法仅使用2D卷积神经网络（Convolutional Neural Network，CNN）来提取特征，提高了算法运行速度。然而，这也导致主干网络的特征提取能力受限，同时，在该算法支柱特征网络最

大池化过程中会丢失部分细粒度信息，从而影响PointPillars的检测精度。因此，为提高该模型的检测性能，本文改进了PointPillars的Pillar编码网络，并优化了主干网络中的二维卷积下采样模块，如图2所示。

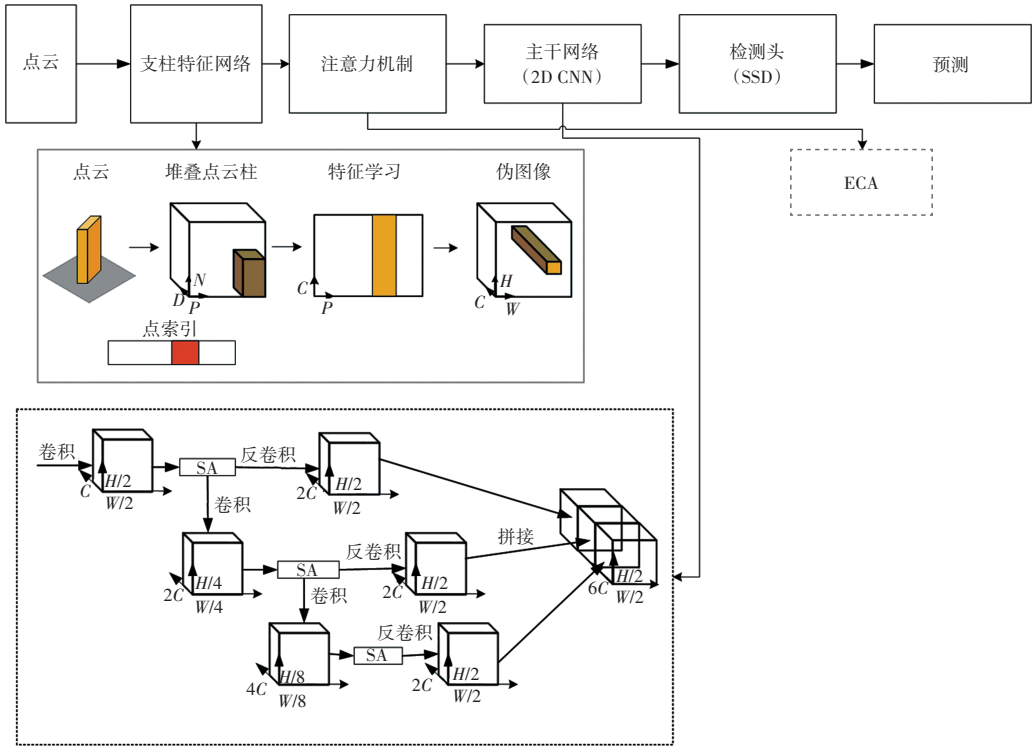


图2　改进的PointPillars网络结构

3.1　基于注意力机制的支柱特征提取网络

在PointPillars的原始支柱特征网络中，点云特征提取被限制在划分好的网格中。这种限制使得网络无法有效提取邻域点云的全局上下文信息，从而导致对于小目标物体的特征提取较为困难。为了解决这一问题，研究人员在编码网络中引入了ECA模块[8]。该机制充分考虑了每个Pillar模块的局部详细几何信息，进而增强了每个Pillar模块的特征表示能力。这一改进显著提升了模型在小目标检测方面的性能，ECA模块的结构如图3所示。

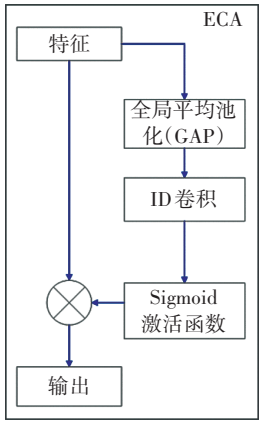


图3　ECA模块结构

在ECA机制中，ECA模块能够捕获本地跨通道交互，这意味着它可以识别和利用通道之间的关系来增强特征表示。为了实现这一目标，ECA采用卷积核为k的快速1D卷积。这种方法不仅高效，还能够准确捕捉到跨通道的交互信息。参数k表示参与一个通道注意力预测的邻居数量，即在预测一个通道的注意力时考虑的相邻通道的数量。ECA使用带状矩阵$\boldsymbol{W}_k$来学习通道注意力，这种矩阵结构使通道之间的交互能够更加有效地进行：

$$\boldsymbol{W}_k=\begin{bmatrix} w^{1,1} & \cdots & w^{1,k} & 0 & 0 & \cdots & \cdots & 0 \\ 0 & w^{2,2} & \cdots & w^{2,k+1} & 0 & \cdots & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & w^{C,C-k+1} & \cdots & w^{C,C} \end{bmatrix} \tag{1}$$

式中：C为网络中所处理的特征图的通道数；k为用于捕捉跨通道交互信息的卷积核的大小，即在进行卷积操作时所覆盖的范围；w^{pq}为带状矩阵$\boldsymbol{W}_k$中第p行第q列元素，具体表示局部跨通道交互的起始通道和所覆盖的通道范围。

对于未降维的聚合特征向量$y\in\mathbf{R}^C$，ECA通道共享相同的学习参数，考虑y_i与其k个相邻通道的相互作用，可以计算出y_i的权值ω_i为：

$$\omega_i=\sigma\left(\sum_{j=1}^{k}\omega_i^j y_i^j\right),\ y_i^j\in\Omega_i^k \tag{2}$$

式中：σ为群卷积块，是一种特定类型的卷积操作，用于处理通道间的信息；y_i^j为特征图$\boldsymbol{y}$中与第i个通道相邻的第j个通道的特征；Ω_i^k为特征图$\boldsymbol{y}$中第i个通道所对应的k个相邻通道的集合。

ECA的跨通道交互通过卷积核为k的快速1D卷积实现：

$$\boldsymbol{\omega}=\sigma(C_{1Dk}(\boldsymbol{y})) \tag{3}$$

式中：$\boldsymbol{\omega}$为注意力权重向量，C_{1D}为1D卷积。

3.2 改进的主干网络

PointPillars框架的主干网络依赖于传统的CNN来提取特征。由于传统CNN的特征提取能力有限，为了提高特征提取的效果，本文引入SA模块进行改进。特征图$\boldsymbol{X}\in\mathbf{R}^{C\times H\times W}$的尺寸定义为$C$（通道数）、$H$（高度）、$W$（宽度）。SA首先将特征图按照通道维度划分为$G$组，修改后每组的通道数变为$C/G$。划分后的特征图$\boldsymbol{X}_k\in\mathbf{R}^{(C/G)\times H\times W}$进一步分成2个分支，通道数变为$C/(2G)$，特征图变为$\boldsymbol{X}_{k1}\in\mathbf{R}^{[C/(2G)]\times H\times W}$，分别输入通道注意力模块和空间注意力模块中进行处理。对经通道注意力模块和空间注意力模块处理后的特征图进行输出连接，并使用Shuffle算子[9]来实现跨组信息的传输。SA模块最终输出的特征图的尺寸与输入的特征图$\boldsymbol{X}$相同，确保信息的完整传递。SA模块如图4所示。

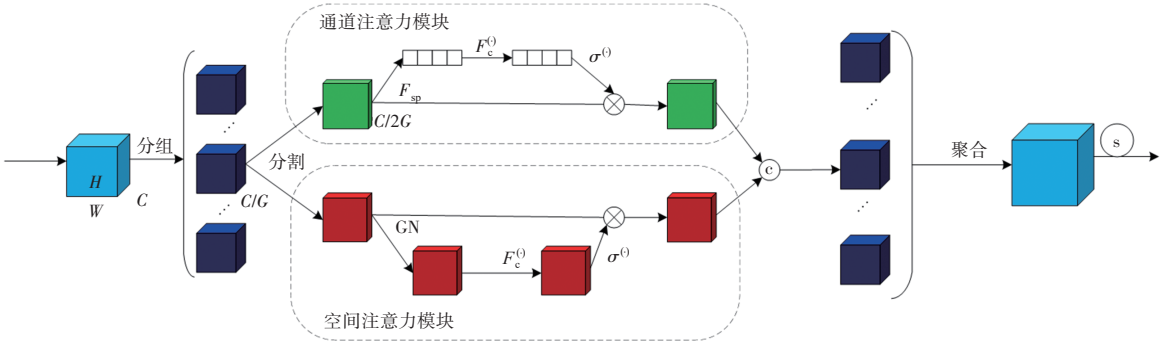


图4 SA模块示意

4 试验结果

4.1 数据集

KITTI数据集是自动驾驶领域广泛使用的数据集，包含点云和图像数据，覆盖市区、乡村等多种场景。为了进行训练和测试，本文将数据集划分为7 481个训练样本和7 518个测试样本，并进一步将训练集划分为3 712个训练样本和3 769个验证样本，以确保模型的有效性和泛化能力。在模型训练过程中，仅使用了KITTI数据集提供的点云数据，这有助于模型更好地学习和理解场景中的三维信息。本文所用硬件环境为NVIDIA RTX 4060 GPU、12th Gen Intel Core i7-12650H CPU，所用软件环境为Ubuntu 18.04、Python 3.8、Pytorch 1.10、Cuda 11.1等。使用自适应矩估计（Adam）优化器实现端到端的训练，设置批量大小（BatchSize）为4、权重衰减值为0.01、动量值为0.9、学习率衰减值为0.1、最大迭代次数为80次。

4.2 定量分析

KITTI官方数据集分为简单、中等和困难3种难度。三维模式验证结果如表1所示，均采用单步目标检测算法。平均精度（Average Precision，AP）的计算公式为：

$$P_A=\sum_{r\in 0,0.01,0.02,\cdots,1} P_{int}(r)\times\Delta r \tag{4}$$

式中：$P_{int}(r)$为在召回率$\geqslant r$时的最大精度，Δr=0.01为采样步长。

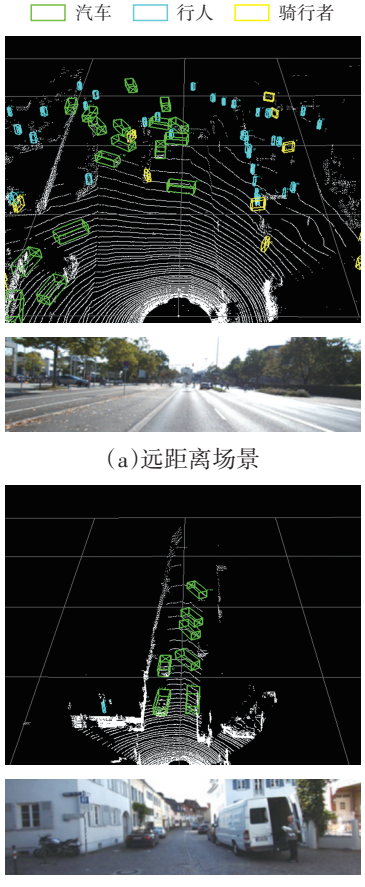
表1 KITTI数据集三维模式平均精度评估结果 %

算法	汽车（Car）			行人（Pedestrian）			骑行者（Cyclist）		
	简单	中等	困难	简单	中等	困难	简单	中等	困难
体素网络（VoxelNet）[10]	77.47	65.11	57.73	39.48	33.69	31.50	61.22	48.36	44.37
稀疏嵌入式卷积检测（SECOND）[11]	83.13	73.66	66.20	51.07	42.56	37.29	70.51	53.85	46.90
PointPillars	83.78	75.64	71.78	50.76	44.38	41.26	80.94	63.62	60.82
PointPillars+ECA	85.61	75.89	71.85	50.92	44.95	41.36	82.62	64.48	61.18
PointPillars+Shuffle	84.77	75.30	71.66	53.33	47.45	42.76	79.17	62.08	59.18
PointPillars+ECA+Shuffle	84.09	75.72	71.88	54.80	47.81	43.19	79.45	61.26	57.54

从表1中可以看出：改进的PointPillars算法对各目标的检测结果均较好，相对于PointPillars算法，行人的检测精度明显提高，汽车和骑行者的检测精度变化不大。注意力模块的引入能够增加对支柱特征局部几何信息的关注度，使算法对行人的检测精度得到有效提升。上述结果证明了改进后PoinPillars的有效性，以及引入模块后的有效性。

4.3 定性分析

在KITTI数据集上使用改进算法的检测效果如图5所示。



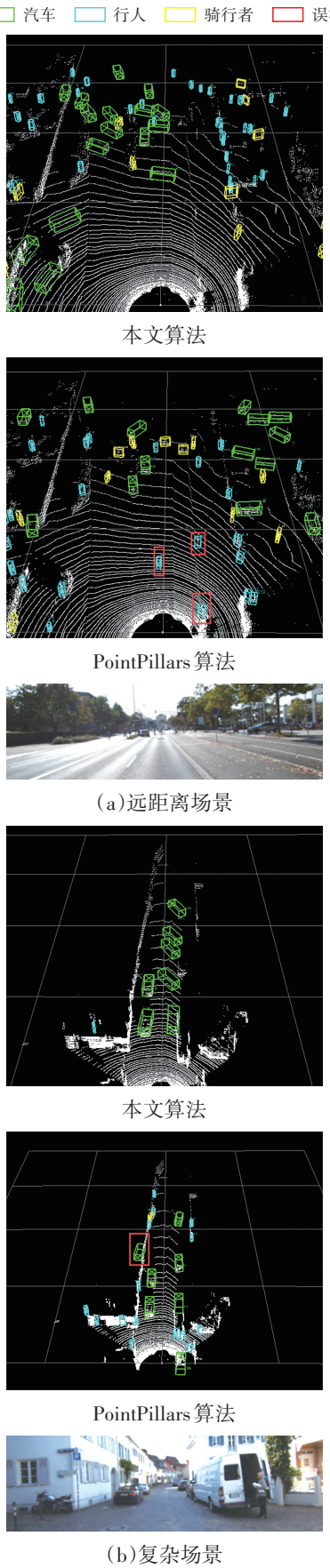
(a)远距离场景

(b)复杂场景

图5 本文算法对不同场景的3D目标检测效果

从图5a中可以看出，对于含有较为稀疏的远距离小目标的场景，本文算法具有较好的检测效果；从图5b中可以看出，对于目标种类多且较为密集的场景，本文算法仍可进行较为准确的检测。改进算法也存在一定的误判情况，如图5a中，路边的小树被误检为骑行者或者行人，但相较于PointPillars算法中小目标误检和漏检较多的情况，改进的PointPillars检测精度有所提升。

改进算法和PointPillars算法对KITTI数据集的远距离场景和复杂场景的检测效果对比如图6所示。



本文算法

PointPillars算法

(a)远距离场景

本文算法

PointPillars算法

(b)复杂场景

图6 本文算法与PointPillars算法的检测效果对比

从图6中可以看出，PointPillars算法在远距离场景小目标检测中存在较多的误检情况，其中：路边的树木等目标被错误地检测成了行人；连续且距离较近的枝干被误识别为骑行者；在复杂场景中，方形的物体被误识别为汽车。本文算法也存在一定的误检情况，如远距离场景中路边的树干被识别为行人，复杂场景标志牌被识别成骑行者。但本文算法避免了PointPillars算法检测时小目标误检和漏检较多的情况，取得了较好的可视化效果，总体检测效果有所提升。

5 结束语

为解决小目标物体识别精度不高的问题，本文对PointPillars模型进行改进，在点云编码方面加入ECA模块，提高了每个Pillar模块的特征表示能力，从而提升了模型的小目标检测性能，并在主干网络的下采样部分加入SA模块，使模型在网络特征提取阶段能够提升特征表达的多样性和有效性，聚焦于更重要的特征，同时抑制不重要特征和冗余特征，从而增强了算法的特征提取能力。试验结果表明，本文方法提升了小目标物体的检测准确率，使三维模式行人平均检测精度提升了3.13百分点，优于原始的PointPillars网络。在未来的研究中，将考虑采用轻量级网络降低模型的复杂度，提升算法的实时性。

参 考 文 献

[1] 杨庆鑫, 孔德明, 陈晶, 等. 基于密度聚类和双重注意力机制的PointPillars改进方法[J]. 激光与光电子学进展, 2025, 62(10): 114-122.
YANG Q X, KONG D M, CHEN J, et al. PointPillars Improvement Method Based on Density Clustering and Dual Attention Mechanism[J]. Progress in Laser and Optoelectronics, 2025, 62(10): 114-122.

[2] LANG A H, VORA S, CAESAR H, et al. PointPillars: Fast Encoders for Object Detection from Point Clouds[C]// Proceedings of the 2019 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2019: 12697-12705.

[3] 陈德江, 余文俊, 高永彬. 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, 2023, 60(10): 447-453.
CHEN D J, YU W J, GAO Y B. Advances in Laser and Optoelectronics, 2023, 60(10): 447-453.

[4] 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. PointPillars+3D Object Detection Based on Attention Mechanism[J]. Journal of Jiangsu University (Natural Science Edition), 2020, 41(3): 268-273.

[5] 贾远鹏, 陈学文, 哈瑞峰. 双注意力机制下自动驾驶汽车车道线深度感知研究[J]. 重庆理工大学学报(自然科学), 2023, 37(7): 44-50.
JIA Y P, CHEN X W, HA R F. Research on Depth Perception of Lane Lines of Autonomous Vehicles under Dual Attention Mechanism[J]. Journal of Chongqing University of Technology (Natural Science), 2023, 37(7): 44-50.

[6] 王盈丰, 吴俭, 宋佳, 等. 基于全局注意力机制的Robust-PointPillars三维目标检测[J]. 舰船电子对抗, 2024, 47(2): 86-92.
WANG Y F, WU J, SONG J, et al. Robust-PointPillars Three-Dimensional Target Detection Based on Global Attention Mechanism[J]. Ship Electronic Countermeasures, 2024, 47(2): 86-92.

[7] CHU H C, WANG T H, MIAO Q N, et al. Ship Detection Based on Improved SDD Algorithm[J]. Instrumentation, 2024, 11(4): 35-43.

[8] WANG Q, WU B, ZHU P, et al. ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks [C]// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, 2020: 11531-11539.

[9] ZHANG Q L, YANG Y. SA-NET: Shuffle Attention for Deep Convolutional Neural Networks[C]// ICASSP IEEE International Conference on Acoustics. Toronto, Ontario, Canada: IEEE, 2021: 2235-2239.

[10] 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. Piscataway, USA: IEEE, 2018: 4490-4499.

[11] YAN Y, MAO Y X, LI B. SECOND: Sparsely Embedded Convolutional Detection[J]. Sensors, 2018, 18(10).

（责任编辑 斛 畔）

修改稿收到日期为2025年3月19日。