

融合多尺度特征和自适应 NMS 的 3D 目标检测^{*}张李辉¹ 刘紫燕^{1,2}

(1. 贵州大学大数据与信息工程学院 贵阳 550025; 2. 贵州大学公共大数据国家重点实验室 贵阳 550025)

摘 要: 3D 目标检测是自动驾驶感知系统的关键技术之一,能准确检测驾驶环境的状态从而保证行车安全。针对自动驾驶场景中行人和骑行者等小目标的 3D 检测精度较低的问题,提出一种基于多尺度特征和自适应非极大值抑制的 3D 目标检测算法。首先,设计多尺度特征提取器,以获取大、中、小尺度的特征。其次,设计多尺度检测头以生成不同尺寸目标的候选框,从而补充小目标候选框。为了平衡多尺度候选框的数量,设计一种基于 ANMS 的候选框筛选算法,提高了对不同尺寸目标的检测精度。在 KITTI 数据集上的结果表明,改进算法在确保汽车类目标检测精度的同时,对行人和骑行者的检测精度达到 62.57% 和 73.30%,比基线算法高 2.04% 和 1.33%,验证了改进算法在小目标检测方面具有较好的 3D 检测性能。

关键词: 3D 目标检测;多尺度特征;自适应非极大值抑制

中图分类号: TN958.98 **文献标识码:** A **国家标准学科分类代码:** 510.4

3D object detection on fusing multi-scale features and adaptive NMS

Zhang Lihui¹ Liu Ziyang^{1,2}

(1. College of Big Data and Information Engineering, Guizhou University, Guiyang 550025, China;

2. State Key Laboratory of Public Big Data, Guizhou University, Guiyang 550025, China)

Abstract: 3D object detection is a critical technology for automatic driving perception systems, which accurately detects the state of the driving environment and ensures the safety of driving. Aiming at the problem of low 3D detection accuracy of small objects such as pedestrians and cyclists, a 3D object detection algorithm based on multi-scale features and adaptive non-maximum suppression (ANMS) is proposed. Firstly, a multi-scale feature extractor is designed to obtain large, medium, and small-sized features. Secondly, a multi-scale detection head is constructed to generate the candidate boxes of objects of different sizes, thereby supplementing the candidate boxes of small objects. To balance the number of multi-scale candidate boxes, a candidate box screening algorithm on ANMS is designed, enhancing the detection accuracy of objects of different sizes. The results on the KITTI dataset indicate that the improved algorithm achieves 62.57% and 73.30% detection accuracy for pedestrians and cyclists, which is 2.04% and 1.33% higher than the baseline while ensuring the detection accuracy of car-type objects, which verifies that the improved algorithm has preferable 3D detection performance in small object detection.

Keywords: 3D object detection; multiscale features; adaptive non-maximum suppression

0 引 言

近年来,自动驾驶系统因提升行车安全的优点而受到广泛关注。智能感知是自动驾驶系统的重要组成部分,旨在准确检测驾驶环境的状况,为自动驾驶的决策和控制提供可靠的环境观察结果^[1]。感知系统通过激光雷达和相机等传感器采集三维(three-dimensional, 3D)世界中的信息,通过检测算法对汽车、行人和骑行者等目标进行精准识别

和定位。检测结果的精度直接影响到自动驾驶系统的可靠性和准确性^[2]。

3D 目标检测算法包括基于图像的检测算法和基于点云的检测算法^[3]。为解决图像缺乏深度信息的问题, Park 等^[4]引入深度估计器,把深度估计器预测的深度图转换为伪激光雷达点云,并用基于点云的算法实现 3D 目标检测。郑自立等^[5]将含有深度信息的图像特征转换到体素空间且投影到鸟瞰图(bird's eye view, BEV)网格来完成检测任

务。在基于图像的 3D 检测算法中,由于难以获取精确的深度信息,该类算法的检测性能不佳^[6]。与图像不同,激光雷达采集的点云含有准确的深度信息,使基于点云的算法具备较高的精度且对低光照环境有良好的鲁棒性^[7]。基于点云的 3D 目标检测算法包括基于点的检测算法和基于体素的检测算法^[8]。由于点云中点的数量庞大,直接从点云中提取特征需要的计算成本相对较高^[9]。为提高点云的特征提取效率,PV-RCNN^[10]和 Voxel RCNN^[11]等算法将点云转化为有序的体素,从而降低算法的计算开销。Voxel RCNN 算法是经典的基于体素的 3D 目标检测算法之一,其在细化阶段对体素特征进一步处理并预测 3D 框,略微降低推理速度换取了整体检测精度的提升。金宇峰等^[12]在 Voxel RCNN 算法中引入图像特征,设计 Transformer 结构的区域建议模块来融合特征,提高了算法的鲁棒性。张林杰等^[13]在 Voxel RCNN 的细化阶段引入局部特征,将全局体素特征与局部特征进行融合,提高了算法的检测精度。韩建栋等^[14]提出的 Myvoxel R-CNN 算法在 Voxel RCNN 的 3D 主干网络和 2D 主干网络中增加注意力机制,提升算法的特征学习能力。然而,Voxel RCNN 算法及现有改进的 Voxel RCNN 算法仅使用一个固定尺度的检测头,使大尺寸目标因特征含量高而取得良好的检测效果,而对于行人、骑行者等小目标因特征含量低而存在错检漏检的问题。

针对现有算法的不足,本文以 Voxel RCNN 网络为基础设计多尺度特征提取器和多尺度检测头来获取不同尺寸目标的候选框。为了解决候选框数量增加导致单个非极大值抑制(non-maximum suppression, NMS)难以筛选有效候选框的问题,设计基于自适应非极大值抑制(adaptive non-maximum suppression, ANMS)的候选框筛选算法,以均衡不同尺寸目标的候选框数量。

1 Voxel RCNN 算法

Voxel RCNN 包括输入端、3D 主干网络、区域建议网络和细化阶段,如图 1 所示。

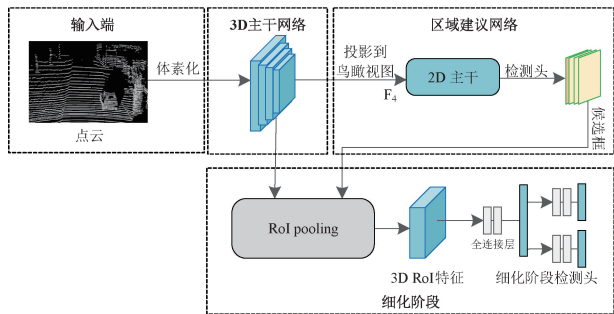


图 1 Voxel RCNN 网络结构图

Fig. 1 Network structure of Voxel RCNN

在输入端,Voxel RCNN 把无序点云量化为规则的体素。在主干网络部分,Voxel RCNN 使用 3D 主干网络提取

体素特征。在区域建议网络中,Voxel RCNN 把 3D 主干网络提取的 3D 特征沿高度压缩到鸟瞰视图,将三维特征投影到二维平面。然后,一个固定尺度的检测头从 2D 主干提取的前景特征中预测目标候选框。在细化阶段,感兴趣区域池化模块(region of interest pooling, RoI pooling)基于候选框来查询前景目标的 RoI 特征。最后,细化阶段的检测头从 RoI 特征中生成目标的 3D 框。

2 改进 Voxel RCNN 算法

2.1 改进 Voxel RCNN 算法架构

Voxel RCNN 在区域建议网络部分使用一个检测头从单一尺度的特征图中预测目标候选框,未对不同感受野的多尺度特征进行差异化处理,导致算法对不同尺度目标的检测性能存在较大差异,容易错检漏检小目标。针对这个问题,对 Voxel RCNN 进行改进,如图 2 所示。

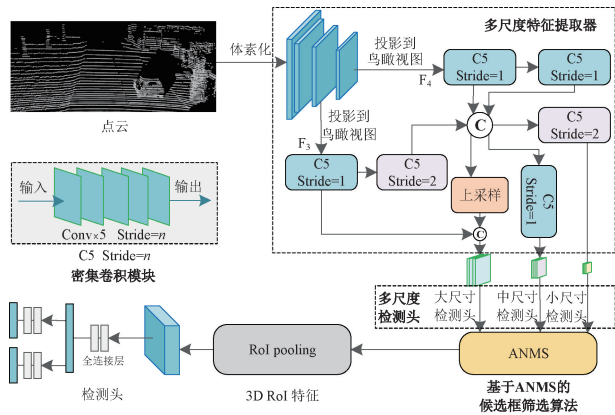


图 2 改进 Voxel RCNN 算法架构

Fig. 2 Network structure of improved Voxel RCNN

首先,设计多尺度特征提取器、多尺度检测头和密集卷积模块来生成不同尺度目标的候选框。其中,多尺度特征提取器用于提取和融合投影到鸟瞰视图的深层特征 F_4 和浅层特征 F_3 ,生成大、中、小尺度的特征图。多尺度检测头包含大、中、小 3 个尺度的检测头,用于生成不同尺寸目标的候选框,以增加小尺寸目标的候选框数量占比,从而提高小目标的检测精度。密集卷积模块在特征提取时,可根据卷积步幅改变特征图的尺寸。

其次,设计基于 ANMS 的候选框筛选算法来滤除冗余候选框,均衡不同尺寸目标的候选框数量。3 个不同尺度的检测头生成的候选框的数量较原 Voxel RCNN 算法增加数倍,导致候选框中存在大量冗余,影响检测精度,耗费计算资源。为了解决这个问题,设计基于 ANMS 的候选框筛选算法,使模型在训练过程中自适应查询多个 ANMS 的最优 NMS 阈值组合,使不同尺寸目标的候选框的数量占比保持最优。其中,本文对 NMS 进行改进,使其阈值在训练过程中自适应寻优,即 ANMS。

最后,细化阶段的 RoI pooling 模块基于多尺度候选框

查询更均衡的 3D RoI 特征。检测头从 3D RoI 特征中预测目标的最终 3D 框。

2.2 多尺度特征提取器

本文设计的多尺度特征提取器有两个功能:一是提取和融合上下文信息;二是生成小、中、大 3 种尺度的特征图。不同尺度的特征图将输入到对应尺度的检测头,以预测和补充不同尺寸目标的候选框,从而提高小目标候选框的数量占比。多尺度特征提取器的结构如图 3 所示。

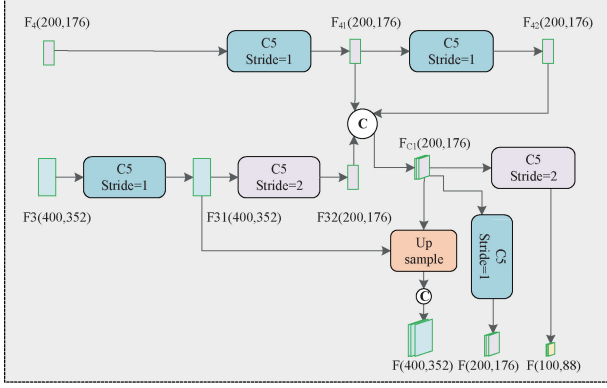


图 3 多尺度特征提取器

Fig. 3 Multi-scale feature extractor

首先,将 3D 主干网络输出的第 4 层特征和第 3 层特征投影到鸟瞰视图,得到特征图 F_4 和 F_3 。然后,使用两个步幅均为 1 的 C5 模块对第 4 层特征 F_4 进行特征提取,分别得到特征图 F_{41} 和 F_{42} 。其中,特征图 F_{41} 和 F_{42} 的尺度分别为 (200,176) 和 (200,176),计算公式为:

$$\begin{aligned} F_{41} &= C5(F_4, stride = 1) \\ F_{42} &= C5(F_{41}, stride = 1) \end{aligned} \quad (1)$$

式中:C5 为密集卷积模块,用于提取特征,详细过程见 2.3 节。 F_{41} 是 F_4 的浅层特征, F_{42} 是 F_4 的深层特征,卷积步幅均为 1 可确保特征图尺度不改变。

此外,两个步幅分别为 1 和 2 的 C5 模块从第 3 层特征 F_3 中提取浅层特征,分别得到特征图 F_{31} 和 F_{32} 。其中, F_{31} 的尺度为 (400,352),含有丰富的尺寸小目标特征。 F_{32} 是 F_3 下采样后的特征,尺度为 (200,176),用于解决深层网络缺乏浅层特征的问题。计算公式为:

$$\begin{aligned} F_{31} &= C5(F_3, stride = 1) \\ F_{32} &= C5(F_{31}, stride = 2) \end{aligned} \quad (2)$$

然后拼接特征图 F_{32} 、 F_{41} 、 F_{42} 得到特征图 F_{c1} 。 F_{c1} 聚合了浅层特征和深层特征,尺度为 (200,176),计算公式为:

$$F_{c1} = \text{Concat}(F_{32}, F_{41}, F_{42}) \quad (3)$$

同时,一个步幅为 2 的 C5 模块对 F_{c1} 下采样和特征融合,得到尺度为 (100,88) 的小尺度特征图,计算公式为:

$$F_{(100,88)} = C5(F_{c1}, stride = 2) \quad (4)$$

为获取中尺度特征图,使用一个步幅为 1 的 C5 模块对 F_{c1} 特征提取,生成尺度为 (200,176) 的中尺度特征图,计算公式为:

$$F_{(200,176)} = C5(F_{c1}, stride = 1) \quad (5)$$

大尺度特征图由 F_{c1} 上采样后的特征图和浅层特征 F_{31} 拼接而成,计算公式为:

$$F_{(400,352)} = \text{Concat}(\text{Upsample}(F_{c1}), F_{31}) \quad (6)$$

最后,不同尺度的特征图将馈送到对应尺度的检测头,以预测不同尺寸目标的候选框。

2.3 密集卷积模块

为了增强网络的特征提取能力,构建密集卷积模块 C5,结构如图 4 所示。

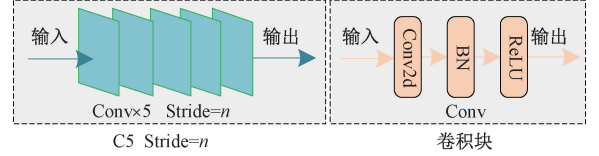


图 4 密集卷积模块 (C5)

Fig. 4 Dense convolutional module (C5)

C5 模块由 5 个卷积块堆叠组成,前一个卷积块的输出作为后一个卷积块的输入,每个卷积块由卷积层、归一化层和整流线性单元(rectified linear unit, ReLU)构成。C5 模块的计算公式为:

$$\begin{aligned} f_1 &= \text{ReLU}(\text{BN}(\text{conv2d}(x_{in}, n))) \\ f_{i+1} &= \text{ReLU}(\text{BN}(\text{conv2d}(f_i))) \end{aligned} \quad (7)$$

式中: x_{in} 为 C5 模块的输入, f_1 是第一个卷积块的输出。 n 是第一个卷积块的卷积步幅,当 n 为 1 时,卷积块不改变特征图的尺度;当 n 为 2 时,卷积块进行下采样使特征图的尺度缩小为原来的一半。 i 为卷积块的数量,取值从 1~4。

2.4 多尺度检测头

为实现从多尺度特征图中预测不同尺寸目标的候选框,设计了多尺度检测头,由小、中、大 3 个不同尺度的检测头组成,其输入分别是多尺度特征提取器所获取的尺度为 (100,88)、(200,176)、(400,352) 的特征图。鉴于特征图的尺度不同,使得各个检测头需要独立地预测候选框。而且,每个检测头对不同尺寸目标的候选框的预测精度不同,使得各个检测头在训练时的收敛程度和速度不同。所以各个检测头需要设置不同的损失函数。

在锚框数量为 6 的情况下,小尺度、中尺度和大尺度检测头分别生成 $100 \times 88 \times 6 = 52\ 800$ 、 $200 \times 176 \times 6 = 211\ 200$ 和 $400 \times 352 \times 6 = 844\ 800$ 个候选框。由此可知,候选框的数量比原 Voxel RCNN 增加数倍。

在改进的 Voxel RCNN 网络中,多尺度检测头的损失函数组成多尺度区域提议的损失函数 L_{MRP} 。 L_{MRP} 由小尺度检测头损失 L_{rpn_100} 、中尺度检测头损失 L_{rpn_200} 和大尺度检测头损失 L_{rpn_400} 组成,计算公式为:

$$L_{MRP} = \lambda_1 \cdot L_{rpn_100} + \lambda_2 \cdot L_{rpn_200} + \lambda_3 \cdot L_{rpn_400} \quad (8)$$

式中: λ_1 、 λ_2 和 λ_3 为平衡各尺度检测头损失的权重超参数,在实验中均设置为 1。

2.5 自适应非极大值抑制

在筛选多尺度检测头生成的 3 组候选框时,常用的处理方法是 将 3 组候选框直接拼接,然后用 NMS 筛选并保留 512 个候选框。然而,NMS 难以从数倍数量的候选框中筛选保留有效候选框,不仅耗费计算资源,还会降低检测性能,如表 3(b)所示。另外一种方法是使用 3 个 NMS 分别筛选候选框,如表 3(c)所示。然而,这 3 组候选框对不同尺寸目标的置信度不同,故每个 NMS 的阈值均需要针对性的调整和寻优。使用多个 NMS 筛选候选框时,不仅需要为每一个 NMS 设置阈值,还需要考虑多个 NMS 阈值之间的相关性,使得人工设置最优的 NMS 阈值尤为困难。

为了解决这个问题,设计基于 ANMS 的候选框筛选算法,如图 5 所示。

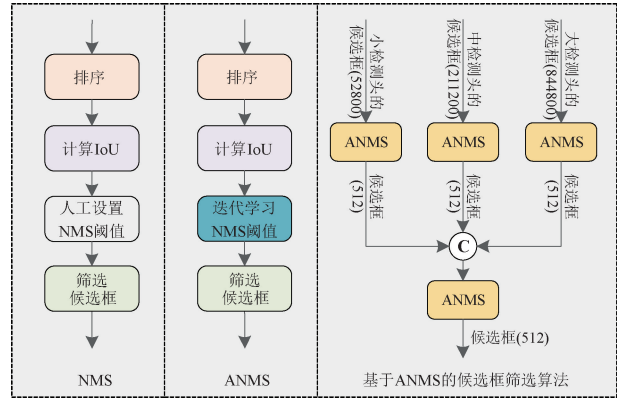


图 5 基于 ANMS 的候选框筛选算法

首先,设计 ANMS 模块,使模型在训练过程中迭代学习最优的 NMS 阈值,从而自适应保留最优候选框、过滤冗余候选框。与 NMS 算法把 NMS 阈值作为超参数人工调整的方法不同,所提 ANMS 将 NMS 阈值设置为可学习的参数,使模型在训练过程中自动寻找合适的 NMS 阈值,如表 3(d)所示。然后,设计基于 ANMS 的候选框筛选算法,其使用 3 个 ANMS 模块对 3 个检测头生成的候选框独立地筛选,分别保留 512 个候选框。随后,拼接筛选后的 3 组候选框,再使用一个 ANMS 进一步筛选拼接后的候选框,以获取细粒的多尺度候选框。在损失函数的约束下,基于 ANMS 的候选框筛选算法使用的 4 个 ANMS 在训练过程中查询适合该尺寸目标的最优 NMS 阈值,迭代优化四个 NMS 阈值,查询最优的阈值组合,使算法最大限度的保留有效候选框、滤除冗余候选框,如表 3(e)所示。

3 实验结果与分析

3.1 数据集与实验配置

使用自动驾驶领域的公开数据集 KITTI 进行实验,该数据集共有 7 481 帧点云数据,其中 3 712 帧用于训练模型,3 769 帧用于模型验证评估。对数据集中的汽车、行人

和骑行者进行检测,检测难度根据遮挡和截断的程度分为简单、中等和困难 3 个级别。

实验在操作系统为 Ubuntu20.04, CPU 为 AMD Ryzen 9 5900HX,显卡为 RTX 3080 laptop GPU 的实验平台上进行训练和评估。模型的训练参数设置如表 1 所示。

表 1 训练参数设置

参数名称	数值
Batch size	2
Epochs	30
Lr	0.01
Lr decay	0.1
Weight decay	0.01
Momentum	0.9

3.2 评价指标

本文以 3D 框的平均精度作为评价指标,对每个类别和不同遮挡等级计算 11 点插值平均精度(11-point interpolated average precision, AP_R11),计算公式为:

$$P = \frac{TP}{TP + FP}$$
 (9)

$$R = \frac{TP}{TP + FN}$$
 (10)

$$AP_R11 = \frac{1}{11} \sum_{i=1}^{11} P_{interp}(R_i)$$
 (11)

式中:TP 是真正例,FP 是假正例,FN 是假反例。对于每个召回率点 R_i , 计算或插值得到相应的最大精度 $P_{interp}(R_i)$ 。

计算 3D 框的平均精度时,汽车、行人和骑行者在 3 个难度级别下的交并比阈值(intersection over union, IoU)均按照通用标准分别设置为 0.7、0.5 和 0.5。IoU 的计算公式为:

$$IoU = \frac{V_{pred} \cap V_{gt}}{V_{pred} \cup V_{gt}}$$
 (12)

式中: V_{pred} 是预测框的体积, V_{gt} 是真实框的体积。

3.3 消融实验

为了验证本文算法的有效性,设计并完成消融实验,如表 2 所示。实验(a)为基线模型 Voxel RCNN,其使用一个固定尺寸的检测头和一个 NMS。实验(b)使用多尺度特征提取器和多尺度检测头共同构成的多尺度候选区域提议(multi-scale regional proposal, MRP)网络和一个 NMS,其检测效果较差。具体而言,使用一个 NMS 时,大尺寸目标的候选框因置信度高而被冗余保留,小尺寸目标的候选框因置信度低而被滤除,导致小目标的检测精度下降。实验(c)使用了 MRP 网络和基于 ANMS 的候选框筛选算法,检测精度显著提升。尤其在中等难度下,行人和骑行者的平均精度较基线(a)分别提升 2.04% 和 1.33%,较实验(b)

分别提升 1.11%和 11.84%，证明了基于 ANMS 的候选框筛选算法能够查询到最优的 NMS 阈值组合来保留有效候选框，平衡了不同尺寸目标的候选框数量。

表 2 不同模块对改进算法 3D AP 值的影响

Table 2 Influence of different modules on the 3D AP of the improved algorithm

实验	MRP	ANMS	汽车 AP _{3D} /%			行人 AP _{3D} /%			骑行者 AP _{3D} /%		
			简单	中等	困难	简单	中等	困难	简单	中等	困难
(a)			89.19	83.88	78.64	66.74	60.53	55.64	86.54	71.97	68.53
(b)	✓		89.28	79.14	78.51	66.38	61.46	57.06	86.44	61.46	57.06
(c)	✓	✓	89.18	83.82	78.72	67.94	62.57	57.25	92.01	73.30	69.97

表 3 对比了不同的候选框筛选策略对所提改进算法的影响。从表 3(b)和(c)可知,仅使用 1 个 NMS 或使用 3 个 NMS 对 3 组候选框进行筛选均无法平衡各类目标的检测精度,证明其无法平衡各类目标的候选框占比。从表 3(d)可知,使用 1 个 ANMS 较(b)和(c)有所改善。从表 3(e)可知,使用基于 ANMS 的候选框筛选算法能更好的平衡各类目标的候选框占比,同时提高行人和骑行者的检测精度。

表 3 不同的候选框筛选策略

Table 3 Different candidate box screening strategies

实验	策略	汽车 AP _{3D} /%			行人 AP _{3D} /%			骑行者 AP _{3D} /%		
		简单	中等	困难	简单	中等	困难	简单	中等	困难
(a)	Baseline	89.19	83.88	78.64	66.74	60.53	55.64	86.54	71.97	68.53
(b)	1 个 NMS	89.28	79.14	78.51	66.38	61.46	57.06	86.44	61.46	57.06
(c)	3 个 NMS	89.29	79.05	78.44	64.47	57.00	53.60	85.83	72.31	69.49
(d)	1 个 ANMS	89.16	79.05	78.42	66.81	61.95	57.00	87.63	74.03	71.45
(e)	基于 ANMS 的候选框筛选算法	89.18	83.82	78.72	67.94	62.57	57.25	92.01	73.30	69.97

为了验证改进算法在不同数量的训练样本下的检测精度,分别用样本数量为 316、625、1 256 个和 3 712 个的训练集训练改进的算法,实验结果如图 6 所示。随着训练样本的增加,改进算法对 3 类目标的检测精度随之提高,说明改进算法在训练数据增多时能取得更好的检测性能。图 6 所示为改进算法在中等难度下对汽车、行人、骑行者的 3D 平均精度。

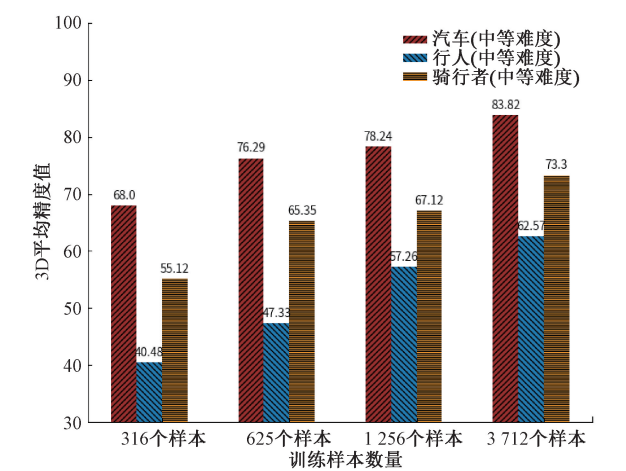


图 6 不同数量训练样本对检测性能的影响

Fig. 6 Influence of different numbers of training samples

3.4 对比实验

在 KITTI 验证集上对比了本文算法、基线算法(Voxel RCNN)、基于点云的算法(Pointpillars^[15]、Part A²^[16]、PV RCNN、SAR-SECOND^[17])以及基于 Voxel RCNN 改进的算法(Myvoxel R-CNN 和文献[12]),如表 4 所示。与其他算法相比,本文算法在汽车、行人和骑行者 3 个类别均取得出色的检测效果,在中等难度级别分别达到 83.82%、62.57%和 73.30%的检测精度。尤其是行人和骑行者两个类别的检测精度优于其他算法,在中等难度级别下较基线算法分别提升 2.04%和 1.33%。总的来说,改进算法在 KITTI 验证集上性能表现优异,尤其对小目标如行人和骑行者具备良好的检测性能。

将所提改进 Voxel RCNN 算法与基线算法 Voxel RCNN 和 Pointpillars 算法进行可视化对比,如图 7 所示。在第 1 个场景,改进算法正确检测到 Pointpillars 漏检的汽车。在第 2 个场景,Voxel RCNN 漏检被汽车遮挡的行人,Pointpillars 算法漏检右侧的汽车和骑行者,而改进算法检测到远距离的汽车和被遮挡的行人,检测效果有所改善。在第 3 个场景,Voxel RCNN 把远距离非目标物体误检为行人,Pointpillar 算法漏检了目标,而改进算法能够准确检测到骑行者和行人。

表 4 本文算法与经典算法、改进的 Voxel RCNN 算法的 3D 检测精度对比

Table 4 Comparison of 3D AP of this paper's algorithm, classic algorithms, and other improved Voxel RCNN algorithms

算法	输入	汽车 AP _{3D} /%			行人 AP _{3D} /%			骑行者 AP _{3D} /%		
		简单	中等	困难	简单	中等	困难	简单	中等	困难
PV RCNN ^[10]	点云	89.34	79.26	78.71	63.10	54.84	51.77	86.06	70.42	64.55
文献[12]	多模态	91.35	85.89	78.41	54.77	44.23	40.17	83.94	64.53	57.46
Myvoxel R-CNN ^[14]	点云	92.65	83.34	82.49	63.07	55.39	50.44	—	—	—
Pointpillars ^[15]	点云	86.46	77.28	75.98	57.49	50.65	43.93	75.63	57.15	55.97
Part A ^{2[16]}	点云	89.55	79.40	78.84	65.68	60.05	55.44	85.50	69.90	65.48
SAR-SECOND ^[17]	点云	89.12	79.09	78.55	55.81	50.33	48.21	83.71	69.61	63.92
基线算法	点云	89.19	83.88	78.64	66.74	60.53	55.64	86.54	71.97	68.53
本文算法	点云	89.18	83.82	78.72	67.94	62.57	57.25	92.01	73.30	69.97

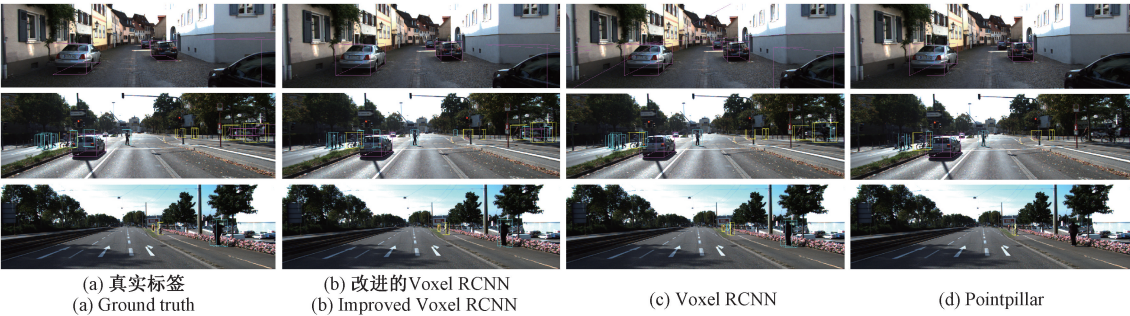


图 7 改进的 Voxel RCNN、Voxel RCNN 和 Pointpillars 算法的可视化效果图

Fig. 7 Visualization of the improved Voxel RCNN、Voxel RCNN and Pointpillars

3.5 应用分析

图 8 展示了所提算法应用于复杂场景的可视化效果图。每一列的上下两幅子图分别对应同一场景下的真实标签和预测结果。场景一是汽车密集分布、目标数量多、遮挡较为严重的复杂场景。场景二是小尺度的行人和骑行者同时分布在远距离位置和近距离位置的复杂场景。

场景三是行人大量密集分布且存在遮挡的复杂场景。场景四是汽车、行人和骑行者同时存在的多类别且多目标的复杂场景。所提改进算法在目标类别多、数量多、遮挡严重的复杂场景中准确检测了行人、骑行者和汽车目标，证明了所提算法的有效性。

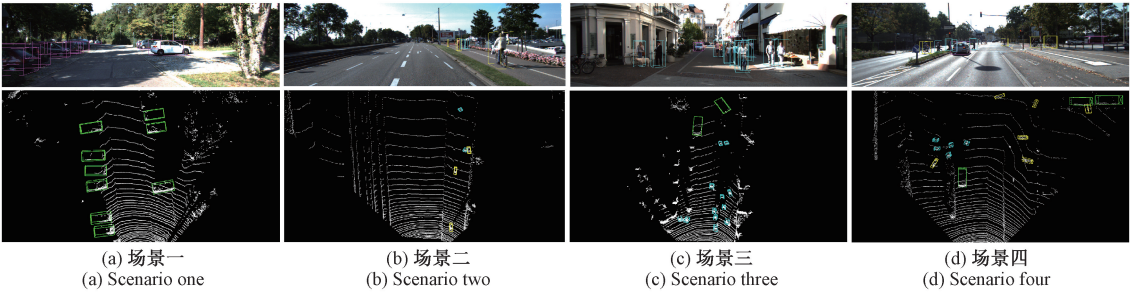


图 8 复杂场景下的可视化效果图

Fig. 8 Visualization in complex scenarios

4 结 论

针对三维场景中行人和骑行者等小目标的 3D 检测精度较低的问题，设计了融合多尺度特征和自适应 NMS 的 3D 目标检测算法。首先，设计了多尺度特征提取器和多尺度检测头来生成不同尺寸目标的候选框。为了平衡各

类目标的候选框数量，设计基于 ANMS 的候选框筛选算法，以保留有效候选框、降低计算开销、提高算法的鲁棒性。改进算法在 KITTI 数据集对中等难度的行人和骑行者的检测精度达到 62.57% 和 73.30%，较基线 Voxel RCNN 提高 2.04% 和 1.33%。实验结果表明改进算法提升了对小目标的 3D 检测精度，为自动驾驶汽车定位小目

标提供依据,对自动驾驶环境感知有重要意义。后续研究将把模型应用到更复杂的检测环境中。

参考文献

- [1] 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. Research on improving vehicle target detection algorithm based on lidar point cloud [J]. Electronic Measurement Technology, 2023, 46(1): 120-126.
- [2] 程鑫,王宏飞,周经美,等.基于体素柱形的激光雷达点云车辆目标检测算法[J].中国公路学报,2023,36(3): 247-260.
CHENG X, WANG H F, ZHOU J M, et al. Vehicle detection algorithm based on voxel pillars from LiDAR point clouds [J]. China Journal of Highway and Transport, 2023, 36(3): 247-260.
- [3] 陈慧娟,吴一全,张耀.基于深度学习的三维点云分析方法研究进展[J].仪器仪表学报,2023,44(11): 130-158.
CHEN H X, WU Y Q, ZHANG Y. Research progress of 3D point cloud analysis methods based on deep learning [J]. Chinese Journal of Scientific Instrument, 2023, 44(11): 130-158.
- [4] PARK D, AMBRUS R, GUIZILINI V, et al. Is pseudo-lidar needed for monocular 3d object detection[C]. IEEE/CVF International Conference on Computer Vision, 2021: 3142-3152.
- [5] 郑自立,徐健,刘秀平,等.联合多注意力和 C-ASPP 的单目 3D 目标检测[J].电子测量与仪器学报,2023, 37(8):241-248.
ZHENG Z L, XU J, LIU X P, et al. Combined multi-attention and C-ASPP network for monocular 3D object detection [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(8): 241-248.
- [6] 梁振明,黄影平,宋卓恒,等.自动驾驶中基于深度学习的 3D 目标检测方法综述[J].上海理工大学学报, 2024,46(2):103-119.
LIANG ZH M, HUANG Y P, SONG ZH H, et al. Survey on deep learning-based 3D object detection methods in autonomous driving [J]. Journal of University of Shanghai for Science and Technology, 2024, 46(2): 103-119.
- [7] 李昌财,陈刚,侯作勋,等.自动驾驶中的三维目标检测算法研究综述[J].中国图象图形学报,2024,29(11): 3238-3264.
LI CH C, CHEN G, HOU Z X, et al. Survey of 3D object detection algorithms for autonomous driving[J]. Journal of Image and Graphics, 2024, 29(11): 3238-3264.
- [8] 鲁斌,孙洋,杨振宇.基于点云稀疏空间特征聚合激励的单阶段 3D 目标检测模型[J].计算机辅助设计与图形学学报,2024,36(5):721-733.
LU B, SUN Y, YANG ZH Y. Aggregate and excitate sparse spatial feature: single-stage 3D object detector from point clouds [J]. Journal of Computer-Aided Design & Computer Graphics, 2024, 36(5): 721-733.
- [9] 霍威乐,荆涛,任爽.面向自动驾驶的三维目标检测综述[J].计算机科学,2023,50(7):107-118.
HUO W L, JING T, REN SH. Review of 3D object detection for autonomous driving [J]. Computer Science, 2023, 50(7): 107-118.
- [10] DENG J, SHI S, LI P, et al. Voxel R-CNN: Towards high performance voxel-based 3d object detection [C]. AAAI Conference on Artificial Intelligence, 2021, 35(2): 1201-1209.
- [11] 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.
- [12] 金宇锋,陶重彝.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.
- [13] 张林杰,柴志雷,王宁.基于局部信息融合的点云 3D 目标检测算法[J].浙江大学学报(工学版),2024, 58(11):2219-2229.
ZHANG L J, CHAI ZH L, WANG N. Point cloud 3D object detection algorithm based on local information fusion[J]. Journal of Zhejiang University (Engineering Science), 2024, 58(11): 2219-2229.
- [14] 韩建栋,范学媛.Myvoxel R-CNN:基于体素的三维点云目标检测模型[J].小型微型计算机系统,2024, 45(8):1908-1913.
HAN J D, FAN X Y. Myvoxel R-CNN: 3D point cloud object detection model based on voxelization[J]. Journal of Chinese Computer Systems, 2024, 45(8): 1908-1913.
- [15] LANG A H, VORA S, CAESAR H, et al. Pointpillars: Fast encoders for object detection from point clouds[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 12697-12705.

[16] 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.

[17] 彭颖,张胜根,黄俊富,等. 基于自注意力机制的两阶段三维目标检测方法[J]. 科学技术与工程, 2024, 24(25):10825-10831.

PENG Y, ZHANG SH G, HUANG J F, et al. Two-

stage 3D object detection method based on self-attention mechanism [J]. Science Technology and Engineering, 2024, 24(25): 10825-10831.

作者简介

张李辉, 硕士研究生, 主要研究方向为目标检测。

E-mail: zhanglihui57@163.com

刘紫燕(通信作者), 教授, 硕士生导师, 主要研究方向为无线通信系统、移动机器人、大数据挖掘分析。

E-mail: Leizy@sina.com