

DOI:10.19651/j.cnki.emt.2416110

FSSD-DETR 自动驾驶场景实时目标检测算法*

沈焯^{1,2} 王泉² 吴昊²

(1.南京信息工程大学计算机学院 南京 210044; 2.无锡学院物联网工程学院 无锡 214105)

摘要: 目标检测技术作为自动驾驶技术的关键环节,对于车辆实现自主导航和决策功能至关重要。现有算法在同时满足检测精度和检测速度上仍有困难。对此,文中提出一种基于 RT-DETR 的实时目标检测算法 FSSD-DETR。该算法在主干网络中引入 FADC 模块,对特征提取过程进行优化;通过引入小目标检测层提升对远处车辆行人小目标的检测性能;基于 SSFF 模块和 TFE 模块对颈部网络进行了重新设计,进而提升检测的准确性;采用 DySample 上采样算子改善最近邻插值法可能出现的细节丢失、锯齿状边缘和图像失真等问题。实验结果表明,相比原始的 RT-DETR 模型,改进算法在 SODA10M 和 BDD100K 数据集上 mAP 分别提升了 3.6% 和 2.1%。实验说明,FSSD-DETR 在保证实时性的同时显著提升了检测精度,具有应用价值。

关键词: RT-DETR;自动驾驶;道路环境;目标检测;多尺度特征融合

中图分类号: TP391;TN791 **文献标识码:** A **国家标准学科分类代码:** 520.60

FSSD-DETR real-time object detection algorithm for autonomous driving scenarios

Shen Ye^{1,2} Wang Quan² Wu Hao²(1. School of Computing, Nanjing University of Information Science and Technology, Nanjing 210044, China;
2. School of Internet of Things Engineering, Wuxi University, Wuxi 214105, China)

Abstract: As a key component of autonomous driving technology, object detection technology is crucial for vehicles to achieve autonomous navigation and decision-making functions. Existing algorithms still face difficulties in meeting both detection accuracy and detection speed simultaneously. In this regard, a real-time object detection algorithm FSSD-DETR based on RT-DETR is proposed. This algorithm introduces FADC module into the backbone to optimize the feature extraction process. A small object detection layer is introduced to improve the detection performance of small targets for distant vehicles and pedestrians. Based on the SSFF module and TFE module, the neck network has been redesigned to improve the accuracy of detection. The DySample upsampling operator is used to replace Nearest Neighbor Interpolation to improve possible issues such as detail loss, jagged edges and image distortion. The experimental results show that compared to the original RT-DETR model, the improved algorithm has increased mAP by 3.6% and 2.1% on the SODA10M and BDD100K datasets respectively. The experiment demonstrates that FSSD-DETR significantly improves the detection accuracy while ensuring real-time performance, which has application value.

Keywords: RT-DETR; autonomous driving; road environment; object detection; multi-scale fusion

0 引言

近年来,随着城市的快速发展,交通车辆的日益增长,道路环境变得更加复杂多变,这对自动驾驶的目标检测技术提出了更高的要求^[1]。在这样的场景中,算法需要能够准确识别和定位各种动态和静态的目标,如车辆、行人、自行车、交通标志以及其他可能影响交通流的障碍物。

为了应对这些挑战,目标检测技术正朝着更加智能化和自动化的方向发展^[2]。目标检测算法通过学习大量数据来提取目标的特征,从而在多变的环境中保持高准确率。这些算法不仅要能够处理日常的交通场景,还要能够应对极端天气、复杂光照条件以及交通高峰时段的挑战。同时,算法的实时性和效率也是研究的重点。在实际的车辆感知系统中,目标检测算法需要快速准确地处理连续的图像数

收稿日期:2024-05-25

* 基金项目:道路交通安全公安部重点实验室开放课题基金(2024ZDSYSKFKT01-2)项目资助

据,以支持实时的决策和响应。这就要求算法不仅要有强大的处理能力,还要满足实时要求。

目标检测作为一项基本的计算机视觉任务,涉及识别和定位图像中的物体^[3]。目前,目标检测器有两种典型的架构:基于 CNN 的和基于 Transformer 的。在过去的几年里,人们对基于 CNN 的检测器进行了广泛的研究并将其应用于自动驾驶目标检测中。Fang 等^[4]通过优化 Mask R-CNN 将其应用于自动驾驶环境感知任务并且展现出良好的性能。然而基于 R-CNN 的方法却牺牲了算法的实时性能。张蕊等^[5]通过引入深度可分离卷积和多种注意力机制改进 YOLOv5s 从而提升夜间检测能力。胡森等^[6]通过在主干网络中引入双层路由注意力机制来提高模型对小目标物体的检测性能。然而,YOLO 系列算法在处理小目标以及密集目标场景时仍面临一定的挑战,这在一定程度上限制了其在某些复杂场景中的应用。

Carion 等^[7]摒弃了传统检测器中的锚点和非极大值抑制,提出了基于 Transformer 架构的端到端的检测器 DETR。DETR 采用二分匹配,将目标检测视为一个直接集合预测问题,避免复杂的后处理过程,大大简化了检测流程。尽管 DETR 具有明显的优势,但它存在两个主要问题:训练收敛缓慢和查询难以优化。对此,Deformable-DETR^[8]通过优化注意力机制,加快了模型对多尺度特征的学习。DAB-DETR^[9]引入了 4D 参考点,并采用分层优化的方法,逐步改进预测的边界框。DN-DETR^[10]通过查询去噪技术,有效缩短了训练时间。然而,DETR 的高计算成本问题尚未得到有效解决。这一问题导致 DETR 很难实现实时目标检测,限制了其在实际部署中的广泛使用。针对上述问题,Zhao 等^[11]提出了一个实时端到端对象检测器 RT-DETR。

目前,RT-DETR 在多个领域中展现出其强大的适应性和泛化能力。段嘉明等^[12]通过在主干网络中融合可变形卷积(deformable convolutional networks version 2, DCNv2)及在颈部网络中引入可学习位置编码(learning position encoding, LPE)使得 RT-DETR 能够适应扣件损伤的检测任务。Yu^[13]通过在主干网络中引入反向残差移动模块(inverted residual mobile block, iRMB)并且将 RT-DETR 的回归损失函数替换为基于最小点距离的 IoU (minimum point distance based IoU, MPDIoU),从而提升对于无人机小目标的检测性能。然而,RT-DETR 在自动驾驶领域尚未应用,并且已有的基于 RT-DETR 改进模型并不适用于自动驾驶检测任务。

针对此现状,本文提出了一种基于 RT-DETR 的改进模型 FSSD-DETR。本文主要工作如下:

1) 在主干网络中,通过引入频率自适应扩张卷积(frequency-adaptive dilated convolution, FADC)对传统卷积操作进行了优化,提升了模型在应对复杂场景时对于数据的提取能力。

2) 引入小目标检测层(small object detection layer, SODL)以利用浅层特征,帮助网络准确检测和定位小目标。

3) 在颈部网络中,基于尺度序列特征融合(scale sequence feature fusion, SSFF)模块和三重特征编码(triple feature encoding, TFE)模块改进特征融合过程,充分利用 SSFF 模块在融合全局多尺度特征方面的优势和 TFE 模块在处理局部多尺度特征信息上的能力。

4) 通过采用 DySample 上采样算子改善最近邻插值法可能出现的问题。

1 RT-DETR 算法

RT-DETR 是一个实时的端到端的目标检测模型,由主干网络、编码器和解码器三部分组成。RT-DETR 的关键创新之一是其高效的混合编码器,该编码器通过解耦尺度内的特征交互和尺度间的特征融合,优化了对多尺度特征的处理方式,从而提高了模型的运行效率。此外,RT-DETR 引入了 IoU 感知查询选择机制,该机制能够根据预测框与目标框之间的 IoU 来选择和优化查询,确保解码器接收到的初始对象查询具有更高的质量。

2 FSSD-DETR 整体架构

为了提升 RT-DETR 对车辆及行人等小目标检测性能的不足,加强对于复杂交通场景中的多尺度问题处理能力,同时保证模型的实时检测性能,本文在主干网络中引入 FADC 模块和小目标检测层。在颈部网络基于 SSFF 模块和 TFE 模块改进特征融合过程。最后采用 DySample 上采样算子提高图像质量。FSSD-DETR 的总体架构如图 1 所示。

2.1 FADC 模块

RT-DETR 模型的主干网络为深度残差网络(residual network18, ResNet18),其主要由多个 BasicBlock 堆叠而成。BasicBlock 主要由两个相对简单的卷积层构成。考虑到需要处理复杂的行车场景,为了在不增加大量参数的前提下更好地提取自动驾驶场景中的各种尺度物体的特征,本文在主干网络的 P5 层引入 FADC 模块^[14],从而帮助模型更好地捕获图像中物体的边缘和纹理等细节信息。

FADC 模块原先用于语义分割任务中处理像素级的图像特征。不同于传统扩张卷积将全局扩张率固定为超参数的做法,其能够根据图像的局部频率组件动态地调整扩张率,从而更灵活地捕获不同尺度的特征。FADC 模块如图 2 所示,该模块融合了 3 个关键组件:频率选择、自适应卷积核和自适应扩张率。频率选择通过动态重加权机制平衡特征表示中的高低频分量使网络能同时捕捉细节和全局信息。自适应卷积核将卷积核参数分解为低频和高频分量,并动态调整这些分量的比例,从而捕获更多的高频信息。自适应扩张率基于局部频率分量动态地为每个像素点

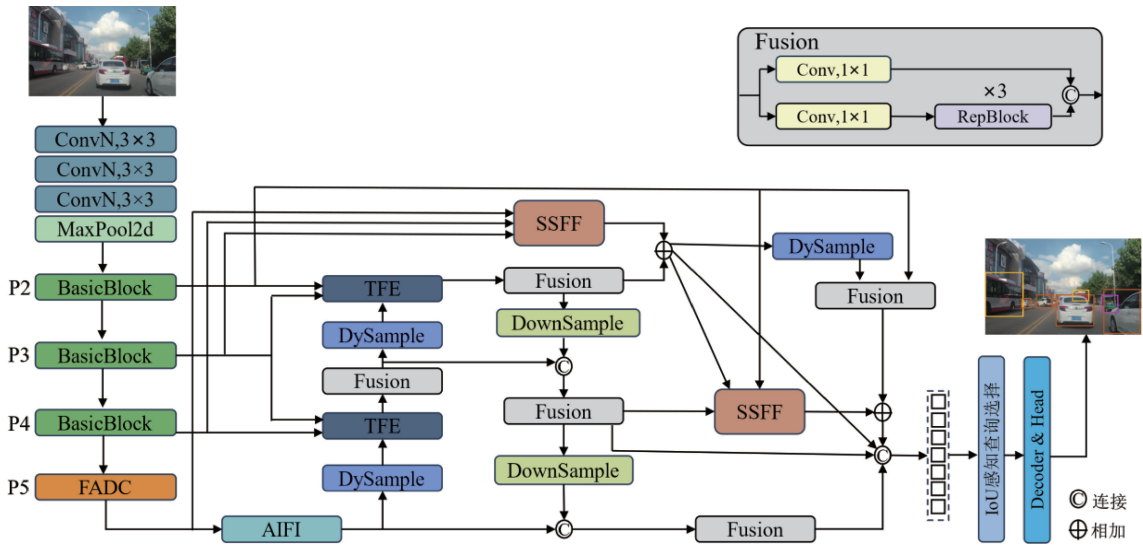


图 1 FSSD-DETR 总体架构

Fig. 1 The overall architecture of FSSD-DETR

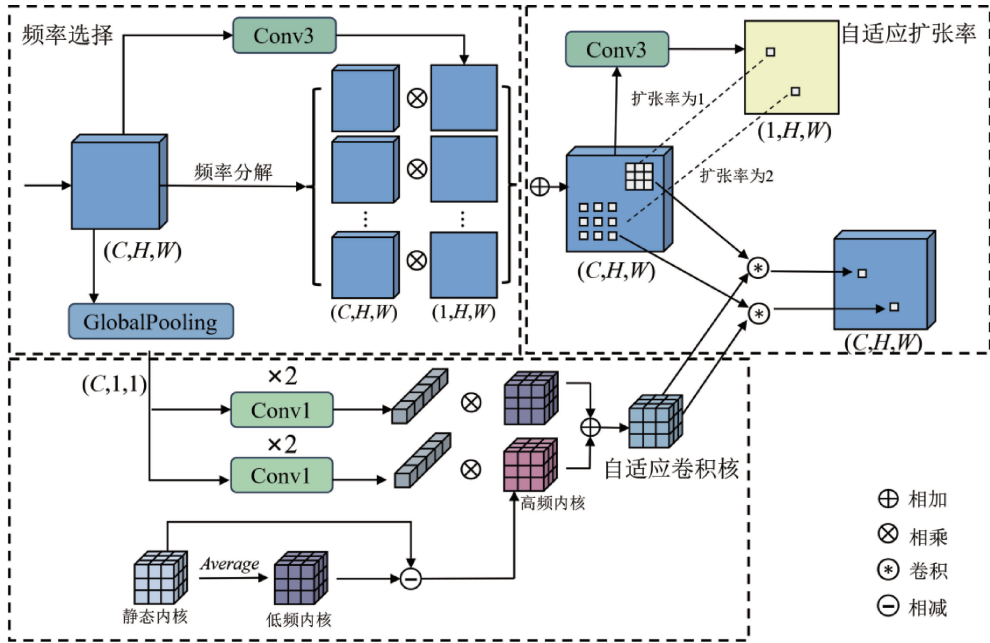


图 2 FADC 模块

Fig. 2 FADC module

独立预测扩张率,以实现有效带宽和感受野之间的平衡。

2.2 小目标检测层

在自动驾驶的实际场景中,需要考虑高速场景和不同光照场景对于小目标的影响,由于图像分辨率、物体距离和物体本身的大小等因素,一些物体在图像中可能只占据很小的一部分。RT-DETR 采用 80×80 大小的特征图用于检测小尺寸目标。对于输入大小为 640×640 的图像,特征图中一个网格的感受野大小为 8×8 。网络很难学习原始图像中小于 8 像素高或宽的目标的特征信息,导致过多的漏检率,而漏检一个行人或车辆可能会导致严重的

后果。

如图 1 所示,在改进后的模型中,对于大小为 160×160 的图像进行一次卷积操作形成 P2,有效利用了大小为 160×160 的图像以利用更为丰富的底层特征信息,从而提升对于小目标的敏感度。随后,将其加入到颈部网络的特征融合过程,以此来强化对小目标的检测能力。通过这一设计,能够有效避免在网络深层处理过程中小目标信息逐渐丢失的问题,从而显著提升对小目标的检测性能。

2.3 ST 模块

自动驾驶场景中,目标的大小和尺度可能存在巨大差

异。例如,从小型车辆到大型货车,从远处的行人到近处的障碍物。为提升模型对于上述场景的处理能力,本文在颈部网络中引入 SSFF 模块和 TFE 模块^[15]改进特征融合过程。

1) SSFF 模块

传统的特征金字塔通常只是通过简单的求和或连接来融合不同尺度的特征,RT-DETR 同样采用简单的拼接或求和进行特征融合。对此,SSFF 模块则通过将不同尺度的特征图堆叠为 4D 张量,引入 3D 卷积来进行特征提取,并且在尺度维度上进行特征融合。SSFF 模块如图 3 所示。首先,在尺度序列特征融合前调整特征图的通道数和大小,使得不同尺度的特征图具有相同的通道数和大小。然后,使用 Unsqueeze 操作增加特征图的维度,将它们从三维张量[高度,宽度,通道]转换为四维张量[深度,高度,宽度,通道]。之后沿深度维度拼接 4D 特征图,形成用于后续卷积的 3D 特征图。最后,使用 3D 卷积、3D 最大池化和 Squeeze 操作对拼接后的 3D 特征图进行操作完成尺度序列特征提取。SSFF 模块结合了来自多个图像尺度的全局语义信息,使模型能够同时利用深层特征图的高维信息和浅层特征图的细节信息,增强模型对于不同大小目标的检测能力。

2) TFE 模块

为了识别密集重叠的小目标,例如远处密集的车辆或行人,一种有效的方法是将图像进行放大处理,以便比较这些小目标在不同尺度下的形态变化。然而,常规的特征金字塔融合方式只是将较小尺寸特征图进行上采样与大尺寸特征图进行融合比较,而忽略了大尺寸特征图中蕴含

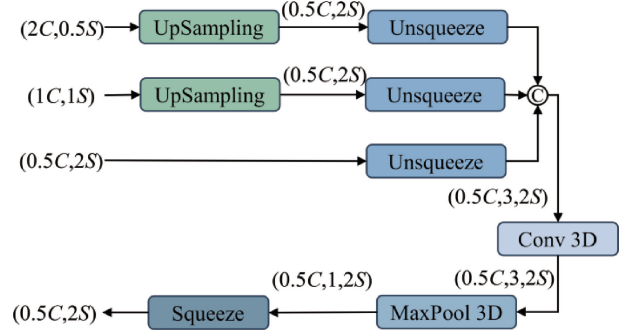


图 3 SSFF 模块

Fig. 3 SSFF module

的丰富细节。对此,TFE 模块将大、中、小 3 种不同尺寸的特征进行拆分,并在融合过程中特别突出了大尺寸特征图的重要性,从而使模型能够捕获小目标对象的局部细节,在识别密集重叠小物体时展现出了更高的准确性和鲁棒性。TFE 模块如图 4 所示,通过处理将大尺寸特征图的通道数调整为 1C,然后使用最大池化加上平均池化的混合结构进行下采样,从而保留了高分辨率特征。对于小尺寸特征图使用卷积来调整其通道数,然后进行上采样,从而保持低分辨率图像局部特征的丰富性,防止小目标特征信息的丢失。最后,将具有相同维度的大、中、小 3 个尺寸的特征图在通道维度上进行拼接。该过程可用式(1)表示:

$$\mathbf{F}_{TFE} = \text{Concat}(\mathbf{F}_l, \mathbf{F}_m, \mathbf{F}_s) \quad (1)$$

其中, \mathbf{F}_{TFE} 表示 TFE 模块输出的特征图。 \mathbf{F}_l 、 \mathbf{F}_m 和 \mathbf{F}_s 分别表示大尺寸、中等尺寸和小尺寸的特征图。 \mathbf{F}_{TFE} 由 \mathbf{F}_l 、 \mathbf{F}_m 和 \mathbf{F}_s 连接得到。

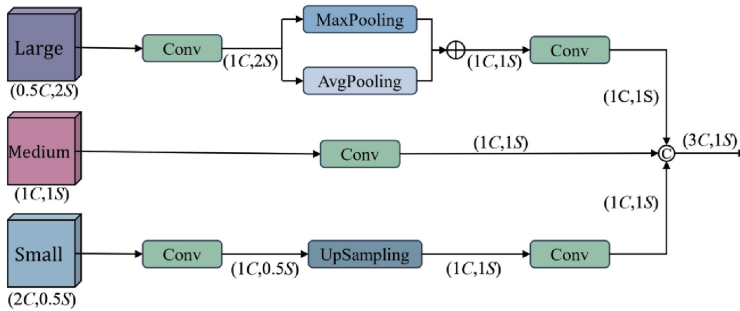


图 4 TFE 模块

Fig. 4 TFE module

2.4 DySample 上采样算子

RT-DETR 在进行图像上采样时,默认采用最近邻插值法。这种方法操作简便,通过直接令变换后像素的灰度值等于距它最近的输入像素的灰度值来实现图像放大。不过,这种方法仅关注像素点的空间位置关系,而没有充分考虑到图像内容的语义信息,因此在处理图像时可能无法完全保留图像的原始细节。对于远处的车辆及行人等对象,可能会出现细节丢失、锯齿状边缘和图像失真等

问题。

针对最近邻插值法的不足,本文提出使用 DySample^[16]上采样算子替换最近邻插值法。DySample 采用点采样的方法实现动态上采样,规避了传统的基于内核的动态上采样方法中常见的复杂动态卷积和子网络构建过程,显著减少了参数量。DySample 结构如图 5 所示。首先,从给定大小为 $C \times H \times W$ 的输入特征图通过采样点生成器转化成大小为 $2 \times sH \times sW$ 的采样集 \mathbf{S} , \mathbf{S} 中第一维

度的 2 表示 x 和 y 坐标。然后利用 *grid sample* 函数对输入特征进行重新采样,得到大小为 $C \times sH \times sW$ 的上采样特征图。在 DySample 的设计中,采样点生成器是实现高效上采样的关键部分。该部分通过结合线性层,动态范围因子和 pixel shuffle^[17]子像素卷积共同生成偏移量 O ,随

后与原始网格位置 G 相加得到最终采样集 S 。其中动态范围因子的引入能够提高偏移量的灵活性,缓解偏移量之间的重叠问题。该过程可用式(2)表示:

$$O = 0.5 \text{sigmoid}(\text{linear}_1(\mathbf{X})) \cdot \text{linear}_2(\mathbf{X}) \quad (2)$$

其中, σ 表示 *sigmoid* 函数,用于生成动态范围因子。

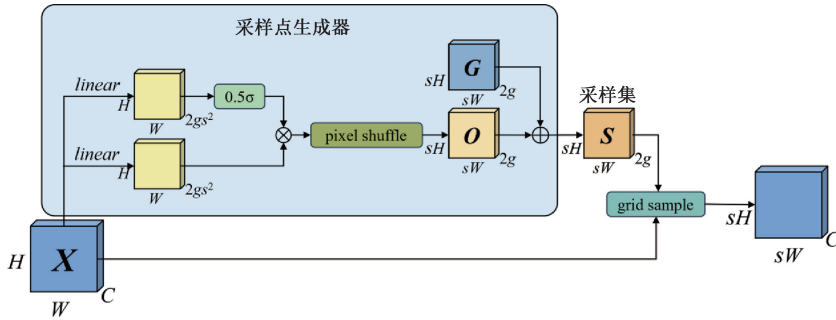


图 5 DySample 结构

Fig. 5 The structure of DySample

3 实验与结果分析

3.1 数据集

本文选取华为诺亚方舟实验室联合中山大学发布的行车场景数据集 SODA10M^[18]。该数据集包含了多种不同的道路场景、天气和时间段。官方提供了带标签的 5 000 张训练集,5 000 张验证集和 10 000 张测试集。由于测试集的标签并未公开,因此本文在验证集上验证本文模型性能。

同时为验证本文改进算法的泛化能力,选取经典的 BDD100K^[19]数据集进行验证。该数据集规模庞大、内容多样,贴近实际驾驶场景。BDD100K 数据集共有 70 000 张训练集、10 000 张验证集和 20 000 张测试集。为加快训练速度,本文从中随机选用带标签的 10 000 张数据集并按照 8 : 1 : 1 的比例重新划分为训练集、验证集和测试集。

3.2 实验环境

本文所有实验均在 Ubuntu20.04 操作系统下进行, GPU 为 NVIDIA GeForce GTX 4090,显存为 24 GB,CPU 为 Intel (R) Xeon (R) Gold 5418Y,深度学习架构为 Pytorch1.11.0,编程语言为 Python3.8,加速计算架构为 CUDA11.3。在训练过程中,将初始学习率设置为 0.000 1,使用权重衰减为 0.000 1 的 AdamW 优化器,并将 batch

size 设置为 8。所有训练都并未使用预训练权重。其他策略和超参数遵循基准模型。

3.3 评估指标

为了评估本文改进算法的有效性,本文选用精确率 (precision, P)、召回率 (recall, R)、平均精度 (mean average precision, mAP)、参数数量和每秒传输帧速 (frames per second, FPS) 作为评估指标。相关计算公式如下:

$$P = \frac{TP}{TP + FP} \quad (3)$$

$$R = \frac{TP}{TP + FN} \quad (4)$$

$$mAP = \frac{\sum_1^N \int_0^1 p(r) dr}{N} \quad (5)$$

其中, TP (true positive) 是模型正确识别为正样本的样本数量, FP (false positive) 是模型错误识别为正样本的样本数量。 FN (false negative) 是模型未能识别为正样本的真正正样本数量。 N 表示类别的数量, $p(r)$ 表示精确率和召回率曲线。

3.4 对比实验

1) 与 RT-DETR 对比

为了验证本文算法的有效性和科学性,将本文网络与 RT-DETR 进行比较,实验结果如表 1 所示。

表 1 RT-DETR 和 FSSD-DETR 在 SODA10M 数据集上的比较结果

Table 1 Comparison results of RT-DETR and FSSD-DETR on the SODA10M dataset

模型	mAP/%	Pedestrian	Cyclist	Car	Truck	Tram	Tricycle	FPS
RT-DETR	55.9	45.3	65.6	84.3	65.5	51.2	23.5	100.1
本文网络	59.5	51.3	68.6	85.7	68.6	54.7	28.1	49.9

从表 1 中可以看出,在关键指标 mAP 上,本文网络达到了 59.5%,相比基准模型 RT-DETR 的 55.9% 有了明显

的增长。在 Pedestrian、Cyclist、Car、Truck、Tram 和 Tricycle 六个类别中, FSSD-DETR 的检测精度均有一定

程度的提升。考虑到数据集本身的复杂性与难度,主要类别 Car 提升的 1.4% 精度能够反映模型在应对复杂交通场景时能够保持较高的检测精度和稳定性。值得一提的是, Pedestrian 和 Tricycle 类别分别显著上涨了 6% 和 4.6%。由于这两个类别通常涉及到小目标检测,这一显著的提升充分表明 FSSD-DETR 在面对小目标时,其检测能力得到了大幅增强。尽管 FSSD-DETR 在 FPS 上有所降低,但仍

然能够满足实时检测的需求。

训练过程中 RT-DETR 和 FSSD-DETR 的 mAP、精确率和召回率的变化曲线如图 6 所示。如图 6(a) 所示,随着 epoch 的增加,本文算法的 mAP 明显高于 RT-DETR,可见 FSSD-DETR 的优越性。如图 6(b) 和 (c) 所示,随着 epoch 的增加,两种算法的精确率和召回率都有所波动,但本文算法始终保持在较高的水平。

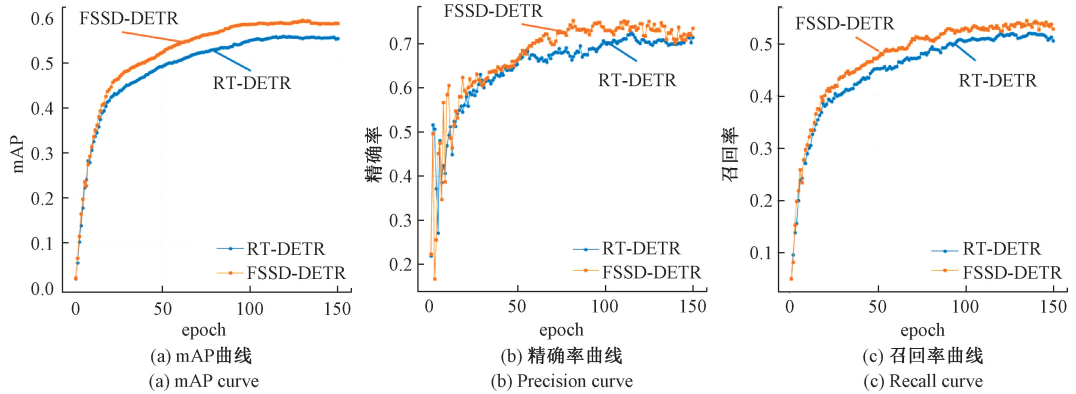


图 6 mAP、精确率和召回率的变化曲线

Fig. 6 The variation curves of mAP, precision and recall

RT-DETR 与 FSSD-DETR 在不同天气和复杂道路环境下的效果对比如图 7 (左为改进前效果,右为改进后效果) 所示。通过对比,可以清晰地看到 FSSD-DETR 能够在多变的天气条件和多样化的道路场景中保持稳定的表现。

相比之下,RT-DETR 在处理小目标检测时经常出现漏检的情况。这一对比充分证明了 FSSD-DETR 在降低漏检率方面的显著成效,其高效的检测能力能够提高自动驾驶系统的安全性。

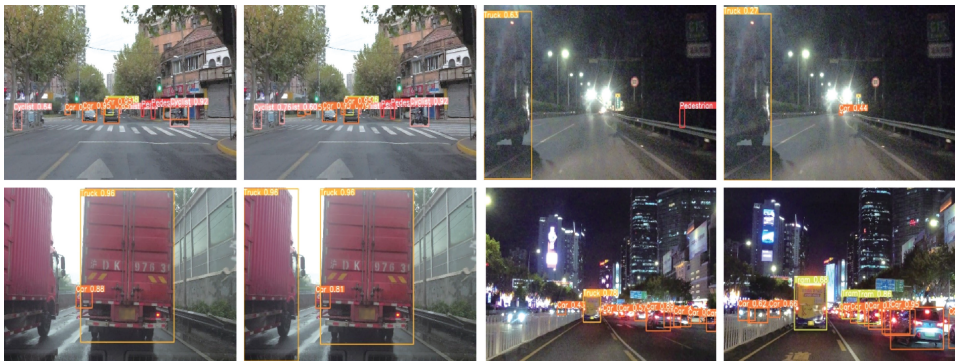


图 7 改进效果对比图

Fig. 7 Comparison chart of improvement effect

为了进一步验证本文算法的泛化能力,本文在 BDD100k 数据集上比较 RT-DETR 和 FSSD-DETR 的性能。比较结果如表 2 所示。可以看出,在除了 Train 类别的各个类别上 mAP 均有提升。在 Person 和 Car 两个主要类别上分别提升了可观的 4.1% 和 2.1%。在 Truck 类别

上提升了最高的 4.5%,这也印证了本文算法在面对复杂的场景时展现出很强的泛化能力。至于 Train 类别,由于数据集中该类别的样本数量相对较少,导致算法在该类别上的性能表现不理想。最终,改进模型的 mAP 比 RT-DETR 高出了 2.1%。

表 2 RT-DETR 和 FSSD-DETR 在 BDD100K 数据集上的比较结果

Table 2 Comparison results of RT-DETR and FSSD-DETR on the BDD100K dataset

模型	mAP/%	Person	Rider	Car	Bus	Truck	Bike	Motor	Traffic Light	Traffic Sign	Train	FPS
RT-DETR	46.3	59.0	33.9	79.5	40.9	47.5	33.8	35.1	65.3	67.6	0.6	102.7
本文网络	48.4	63.1	34.3	81.6	45.3	52.0	35.2	35.7	66.5	69.7	0.0	61.1

2)与检测算法对比

本文在保持相同输入尺寸(640×640)的前提下,与不同模型进行对比实验。比较结果如表 3 所示。本文算法的 mAP 达到了 59.5%,超越了所有参与比较的 YOLO 系列模型和一些基于 RT-DETR 的改进模型。与 YOLO 系列最高的 YOLOv9c 相比,本文算法在更少参数量的情况下高出了

6.7%。此外,本文算法达到了最高的 73.4%精确率和最高的 54%召回率,减少了漏检的可能性。然而,在追求高精度的过程中,本文算法在参数量方面相比基准模型有所上升,导致 FPS 有所下降。总体而言,本文算法通过增加少量参数量实现了精度的显著提升,并且满足实时性。在实际应用中,这种精度与速度之间的平衡具有很高的实用价值。

表 3 不同模型在 SODA10M 数据集上的比较结果

Table 3 Comparison results of different models on the SODA10M dataset

模型	输入尺寸	主干网络	mAP/%	精确率/%	召回率/%	参数量/ 10^6	FPS
YOLOv5m	640	CspDarkNet	50.7	63.5	45.6	25.0	173.2
YOLOv5l	640	CspDarkNet	52.1	64.8	47.2	53.1	140.3
YOLOv6m	640	EfficientRep	50.4	64.2	45.6	52.0	165.8
YOLOv8m	640	CspDarkNet	51.3	64.9	46.4	25.8	185.4
YOLOv8l	640	CspDarkNet	52.5	64.9	46.5	43.6	142.0
YOLOv9c	640	RepNCSPNELAN4	52.8	69.7	46.0	25.3	101.7
YOLOv10b	640	CSPNet	52.7	67.3	46.2	20.4	112.2
YOLOv10l	640	CSPNet	52.6	66.1	46.8	25.7	95.8
RT-DETR	640	ResNet18	55.9	70.0	51.9	19.9	100.1
RT-DETR-iRMB	640	ResNet18+iRMB	53.6	66.4	49.9	16.3	65.9
IR-DETR	640	ResNet18	53.5	69.6	49.5	20.2	97.6
本文网络	640	ResNet18+FDAC	59.5	73.4	54.0	21.3	49.9

3.5 消融实验

为分析本文引入的改进模块对模型性能的影响,本文进行了消融实验。实验结果如表 4 所示。根据表中内容可以看出在引入了小目标检测层后, mAP50 上涨了 2.7%,但参数量也随之上涨了一些。引入的 ST 模块在减少参数量的同时,实现了精度的轻微提升。然后,对 DySample 和 FADC 模块的引入进行了实验比较。当只引入了 DySample 模块时,精度略有上涨 0.2%。但当只引入

了 FADC 模块时,精度却下降了 0.6%。这主要是因为最近邻插值法丢失了 FADC 模块提取的细节信息从而导致性能下降。当同时添加了 DySample 和 FADC 模块时, DySample 能够更好地保留图像的细节信息使模型达到了最好的性能。这些结果表明, DySample 和 FADC 模块对模型的性能具有一定的提升作用。最终,改进模型相较基准模型 mAP 上涨了 3.6%,精确率上涨了 3.4%,召回率上涨了 2.1%。

表 4 消融实验

Table 4 Ablation study

SODL	ST	DySample	FADC	mAP/%	精确率/%	召回率/%	参数量/ 10^6
				55.9	70.0	51.9	19.9
✓				58.6	74.0	52.7	21.8
✓	✓			58.9	71.2	55.0	21.1
✓	✓	✓		59.1	74.3	54.0	21.2
✓	✓		✓	58.3	70.9	52.9	21.3
✓	✓	✓	✓	59.5	73.4	54.0	21.3

4 结 论

为解决自动驾驶中精度与速度的平衡问题,本文提出一种基于 RT-DETR 的实时检测模型 FSSD-DETR。通过将 FADC 模块加入模型的主干网络从而提升模型提取特征的能力。通过加入小目标检测层提升模型对于小目标的检测性能。利用 SSFF 模块和 TFE 模块的优势重新改

进特征融合过程,使模型能够更全面地捕捉图像中的关键信息。通过采用 DySample 上采样算子改进最近邻插值法的不足。实验结果表明,本文算法在 SODA10M 数据集上 mAP 达到 59.5%,49.9 的 FPS 满足实时检测需求。同时,精确率和召回率均有显著提升。在 BDD100K 数据集上,本文算法 mAP 达到 48.4%,61.1 的 FPS 同样满足实时检测需求。相较于其他主流算法,本文所提出的算法在

自动驾驶的目标检测任务中展现出了优秀性能。在未来的工作中,研究重点将放在轻量化设计。在各项性能不降低的情况下,保证精度与速度的平衡,实现轻量化模型。

参考文献

- [1] WANG Y. Overview on key technology of perceptual system on self-driving vehicles[J]. *Autoelectr*, 2016, 57(12):12-16.
- [2] JANAI J, GUNNEY F, BEHL A, et al. Computer vision for autonomous vehicles: Problems, datasets and state of the art[J]. *Foundations and Trends in Computer Graphics and Vision*, 2020, 12(1-3): 1-308.
- [3] ZOU Z, CHEN K, SHI Z, et al. Object detection in 20 years: A survey[J]. *Proceedings of the IEEE*, 2023, 111(3): 257-276.
- [4] FANG S, ZHANG B, HU J. Improved mask R-CNN multi-target detection and segmentation for autonomous driving in complex scenes[J]. *Sensors*, 2023, 23(8): 3853.
- [5] 张蕊,高诗博,赵霞,等. 基于改进 YOLOv5s 的无人驾驶夜间车辆目标检测算法[J]. *电子测量技术*, 2023, 46(17):87-93.
ZHANG R, GAO SH B, ZHAO X, et al. Autonomous nighttime vehicle object detection algorithm based on improved YOLOv5s [J]. *Electronic Measurement Technology*, 2023, 46(17): 87-93.
- [6] 胡森,姜麟,陶友凤,等. 改进 YOLOv7 的自动驾驶目标检测算法[J]. *计算机工程与应用*, 2024, 60(11): 165-172.
HU M, JIANG L, TAO Y F, et al. Improved YOLOv7 autonomous driving object detection algorithm [J]. *Computer Engineering and Applications*, 2024, 60(11):165-172.
- [7] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers [C]. *European Conference on Computer Vision*. Cham: Springer International Publishing, 2020: 213-229.
- [8] ZHU X, SU W, LU L, et al. Deformable detr: Deformable transformers for end-to-end object detection [J]. *ArXiv preprint arXiv*: 2010.04159, 2020.
- [9] LIU S, LI F, ZHANG H, et al. Dab-detr: Dynamic anchor boxes are better queries for detr[J]. *ArXiv preprint arXiv*: 2201.12329, 2022.
- [10] LI F, ZHANG H, LIU S, et al. Dn-detr: Accelerate detr training by introducing query denoising [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022: 13619-13627.
- [11] ZHAO Y, LYU W, XU S, et al. Detsr beat yolos on real-time object detection[J]. *ArXiv preprint arXiv*: 2304.08069, 2023.
- [12] 段嘉明,白堂博,许贵阳,等. 基于 IR-DETR 的轨道图像增强及扣件损伤检测方法[J/OL]. *铁道建筑*, 1-5 [2024-09-05]. <http://kns.cnki.net/kcms/detail/11.2027.U.20240622.2100.002.html>.
DUAN J M, BAI T B, XU G Y, et al. Track image enhancement and fastener damage detection method based on IR-DETR[J/OL]. *Railway architecture*, 1-5 [2024-09-05]. <http://kns.cnki.net/kcms/detail/11.2027.U.20240622.2100.002.html>.
- [13] YU Z. RT-DETR-iRMB: A lightweight real-time small object detection method [C]. *2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC)*. IEEE, 2024, 6: 1454-1458.
- [14] CHEN L, GU L, FU Y. Frequency-adaptive dilated convolution for semantic segmentation [J]. *ArXiv preprint arXiv*: 2403.05369, 2024.
- [15] KANG M, TING C M, TING F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation [J]. *Image and Vision Computing*, 2024, 147: 105057.
- [16] LIU W, LU H, FU H, et al. Learning to upsample by learning to sample [C]. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023: 6027-6037.
- [17] SHI W, CABALLERO J, HUSZAR F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network [C]. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016: 1874-1883.
- [18] HAN J, LIANG X, XU H, et al. Soda10m: Towards large-scale object detection benchmark for autonomous driving[J]. *ArXiv preprint arXiv*: 2106.11118, 2021.
- [19] YU F, CHEN H, WANG X, et al. Bdd100k: A diverse driving dataset for heterogeneous multitask learning [C]. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 2636-2645.

作者简介

沈焯, 硕士研究生, 主要研究方向为计算机视觉, 车联网。
E-mail: 202312490351@nuist.edu.cn

王泉(通信作者), 博士, 正高级工程师, 主要研究方向为通信工程, 智能制造。

E-mail: wangquan@cxwu.edu.cn

吴昊, 博士, 主要研究方向为人工智能, 气象预测。

E-mail: wuhao@cxwu.edu.cn