DOI:10.19651/j.cnki.emt.2415792

基于注意力机制融合特征的车辆目标检测方法

过鑫炎[1] 朱 硕[1,2] 孙佳豪[2] 梁吉丰[2] 汪宗洋[3]
(1.南京信息工程大学 南京 210044; 2.无锡学院 无锡 214105; 3.无锡汐沅科技有限公司 无锡 214000)

摘 要: 为了解决道路监控下的车辆目标检测精度低的问题,本文提出一种改进 YOLOv7 的车辆检测方法。首先引入跨空间学习的高效多尺度注意机制 EMA 来提高对特征信息的关注;其次将颈部网络中的 SPPCSPC 模块替换为 SPPFCSPC 模块,裁剪 CBS 层,引入 EMA 注意力机制,以强化对小目标区域的关注,获取更准确的车辆特征;同时,将 EMA 注意力引入 MP 模块中,使网络融合更多重要的特征信息;最后,采用 MPDIoU 损失函数,加快模型收敛速度并提高检测精度。实验结果表明,改进后的 YOLOv7 检测精度为 86.69%,相比原始 YOLOv7 网络提高了 2.83%,可以有效地提升车辆目标检测精度,为道路视频监控等提供保证。

关键词: 车辆检测;YOLOv7;注意力机制;MPDIoU loss
中图分类号: TN919.8 **文献标识码**: A **国家标准学科分类代码**: 510.1040

Vehicle object detection method based on attention mechanism integrated features

Guo Xinyan[1] Zhu Shuo[1,2] Sun Jiahao[2] Liang Jifeng[2] Wang Zongyang[3]
(1. Nanjing University of Information Science and Technology, Nanjing 210044, China;
2. Wuxi University, Wuxi 214105, China; 3. Wuxi Xiyuan Co., Ltd. of Technology, Wuxi 214000, China)

Abstract: To address the issue of low vehicle detection accuracy in road surveillance, this paper proposes an improved vehicle detection method based on YOLOv7. Firstly, we introduce the Efficient Multi-Scale Attention Mechanism (EMA) for cross-space learning to enhance attention to feature information. Secondly, we replace the SPPCSPC module in the neck network with the SPPFCSPC module, trim the CBS layer, and introduce the EMA attention mechanism to strengthen attention to small target areas, thereby obtaining more accurate vehicle features. Additionally, we incorporate the EMA attention into the MP module to fuse more important feature information. Finally, employing the MPDIoU loss function accelerates model convergence and enhances detection accuracy. Experimental results show that the improved YOLOv7 achieves a detection accuracy of 86.69%, which is a 2.83% improvement over the original YOLOv7 network. This enhancement effectively boosts the accuracy of vehicle object detection, providing assurance for applications such as road video surveillance.

Keywords: vehicle detection; YOLOv7; attention mechanism; MPDIoU loss

0 引 言

随着经济的不断发展,截至 2023 年底,我国机动车保有量达 4.35 亿辆。面对如此庞大的交通体系,提供实时的交通信息,以及确保交通安全和效率成为一项至关重要的任务。在此背景下,车辆目标检测技术作为解决方案的关键组成部分之一,在全球范围内受广泛关注。

现如今,为实现高效的车辆目标检测,学者们已经着手研究智能化的检测方法,这些研究涵盖了机器学习以及与深度学习相关的识别技术。其中,机器学习作为传统方法在车辆检测中因步骤繁琐而导致实时性不佳,并且这种复杂性会引发检测精度不高、泛化能力不强等问题[1]。

在深度学习中,车辆检测可以分为两阶段检测模型和单阶段检测模型。前者分为预测车辆区域和从预测区域中检测车辆目标两个阶段,代表算法包括区域卷积神经网络(region-convolutional neural networks, R-CNN)[2]、快速区域卷积神经网络(fast region-convolutional neural networks, Fast R-CNN)[3]、更快区域卷积神经网络(faster region-convolutional neural networks, Faster R-CNN)[4]和区域全卷积网络(region-based fully convolutional networks,

收稿日期:2024-04-09

R-FCN)[5],该模型通常能够实现较高精度但具有相对较慢的检测速度。后者把车辆识别和检测整合到同一个网络中,不先预测车辆区域,代表算法包括 YOLO (you only look once, YOLO)系列[6]和单步多框检测器(single shot multibox detector,SSD)[7],该模型通常具有较低精度但检测速度高。随着单阶段检测模型的不断改进,近年来又发展出了视网膜网络(RetinaNet)[8]、中心点检测网络(CenterNet)[9]和实时 Transformer 目标检测和跟踪器(real-time detection transformer, RT-DETR)[10]等目标检测算法,其中 RetinaNet 过引入焦点损失(focal loss),重点关注难以分类的样本,从而改善单阶段检测器在处理前景与背景不平衡问题时的性能;CenterNet 是通过预测物体中心点的位置以及尺寸和偏移量,直接从热力图中回归出物体的边界框,从而实现高效的目标检测;RT-DETR 是通过结合 Transformer 架构进行特征提取和全局上下文建模,实现高效的端到端目标检测。因此,单阶段检测模型逐渐适用于要求高检测精度的车辆目标识别场景。

近年来单阶段检测模型的算法研究中,朱茂桃等[11]在 YOLOv3 的基础上,将网络中 3 个平行分支网络进行了参数共享来改善小尺度目标的检测情况;郭宇阳等[12]分别在 YOLOv4 的主干网络和颈部网络使用了幽灵网络(GhostNet)和深度可分离卷积进行轻量操作,并使用 Fire Module 加深网络深度,从而提高网络检测速度;Li 等[13]在 YOLOv5 中加一个额外的预测头来检测较小比例的物体,同时引入了双向特征金字塔网络(bidirectional feature pyramid network,BiFPN)来融合来自不同尺度的特征信息;蔡刘畅等[14]采用 GhostNet 与 YOLOv7 相结合轻量化网络,并针对有效特征层增加通道注意力机制以减少车辆漏检;薛震等[15]在 YOLOv7 中引入 BoT(bag of tricks, BoT)结构使网络更加关注整体图像信息,并更换 SIoU 损失函数加速网络收敛,使得模型可以满足微光环境下的多目标探测;许晓阳等[16]在 YOLOv7-tiny 主干网络中采用深度可分离卷积设计的轻量级 ELAN-DW 模块和 Head 层引入 GhostNet V2 模块来降低参数量,并根据残差结构设计全新跨尺度网络来提高模型性能;杜娟等[17]在 YOLOv7 网络中添加了小目标检测层来增加对小目标的特征学习能力,进而在颈部和检测头引入协调坐标卷积(CoordConv)来更好地感受特征图信息,并采用 P-ELAN 结构对骨干网络进行轻量化处理,使得网络在检测精度和检测速度上达到平衡;张利丰等[18]在 YOLOv8 网络中采用多尺度融合的方式对主干网络进行重构,并在 Neck 部分输出的特征图之后加入轻量型注意力机制(triplet attention,TA),提升模型的特征提取能力;Zhai 等[19]在 SSD 的基础上,设计了密集卷积网络(DenseNet-S-32-1)代替了 SSD 的主干网络,同时引入多尺度特征层的融合机制,并在目标预测之前建立残差块,提高了模型性能。

目前,虽然基于深度学习的目标检测方法已取得了一定进展,但这些方法主要集中在小目标检测或模型轻量化改进方面。此外,这些改进方法所使用的数据集大多是基于车载设备、平视或小角度俯视采集的图像,而这与道路监控摄像头采用的俯视视角存在显著差异。由于俯视角度可能导致部分车辆被遮挡或出现模糊现象等,从而减少了图像中可用的信息量,因此,现有模型难以在大角度俯视视角下有效地执行车辆检测任务。

针对上述问题,本文提出一种基于 YOLOv7 的改进算法,改进主要包括以下 3 方面:1)在 YOLOv7 中将改进的多尺度空间特征金字塔池化结构(spatial pyramid pooling feature concatenation and spatial pyramid combination, SPPFCSPC)模块替换原多尺度空间金字塔池化结构(spatial pyramid pooling concatenation and spatial pyramid combination, SPPCSPC)模块;2)对 SPPFCSPC 和最大池化(maxpool,MP)模块中的卷积层进行裁剪,并引入高效多尺度注意力(efficient multi-scale attention,EMA)注意力机制进行优化,来加强道路监控下车辆检测;3)采用基于最小点距离的边界框相似度比较度量(minimum point distance based IoU,MPDIoU)作为损失函数,获取更准确的回归结果,以及加快模型收敛速度。通过以上改进,该算法能够提高车辆检测效果,从而有效解决车辆检测时出现的问题。

1 YOLOv7 算法概述

YOLOv7 算法是在 YOLOv5 的基础上进行了显著改进,通过优化网络结构和损失函数,提升了目标检测性能,其性能相较于其他 YOLO 模型有了显著提升。因此,本文采用 YOLOv7 作为基准模型,其网络结构如图 1 所示。

YOLOv7 的网络结构由 Input 输入端、Backbone 网络、Neck 网络与 Head 输出端 4 个部分组成。输入图像在进入 Input 输入端后会进行一系列的预处理操作,如数据增强,将图像尺寸缩放至 640×640 大小等[20]。Backbone 网络主要由 CBS 模块、ELAN 模块和 MP 模块组成,其结构如图 2 所示。CBS 模块由 Conv 层、BN 层和 SiLU 激活函数组成。ELAN 由多个 CBS 模块堆叠组成,通过不同长短的梯度路径会产生丰富的特征信息,从而提高网络的学习能力,具有更强的鲁棒性。MP 模块融合了 CBS 模块和最大池化层,专注于下采样操作,有效降低特征丢失,提升模型对输入信息的捕获和表示能力。其次,Neck 网络主要由 SPPCSPC 模块、ELAN-w 模块、CBS 模块、MP 模块和 UPSample 模块组成。其中,ELAN-W 模块相比于 ELAN 增加了两个拼接操作,SPPCSPC 模块是用来增大感受野,其模块结构图如图 3 所示。最后,将特征融合出的 3 个多尺度特征图(大小分别为 80×80×128,40×40×256,20×20×512)送入 Head 检测头,通过 RepConv 模块生成目标的边界框,并进行目标类别的预测,以获得准确的预测结果。

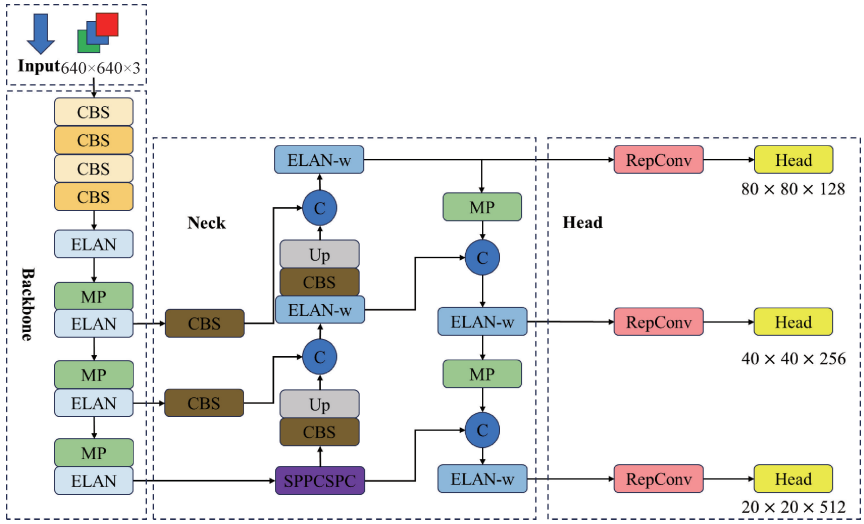

图 1　YOLO v7 网络结构图

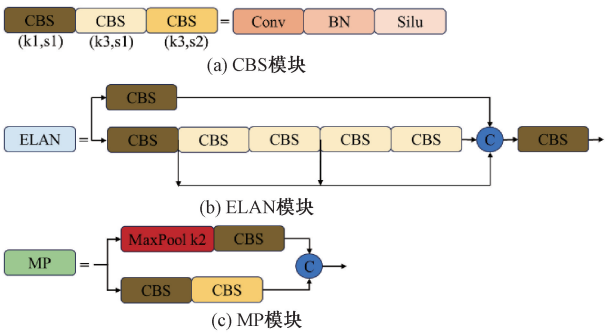

图 2　Backbone 网络各模块结构图

2　改进的 YOLOv7 算法

针对在道路监控场景下车辆目标检测所面临的问题，本文基于 EMA 注意力机制对 YOLOv7 网络进行了优化，改进后的网络结构如图 4 所示。

2.1　EMA 注意力机制

注意力机制是一种模拟人类视觉系统的技术，它在计算机视觉领域引起了广泛关注。在 YOLO 系列网络中，引入注意力机制的目的是通过动态调整模型对输入图像不同区域的关注程度，从而更有效地捕捉目标特征。这种机制使得网络能够更加聚焦于图像中对目标检测最关键的

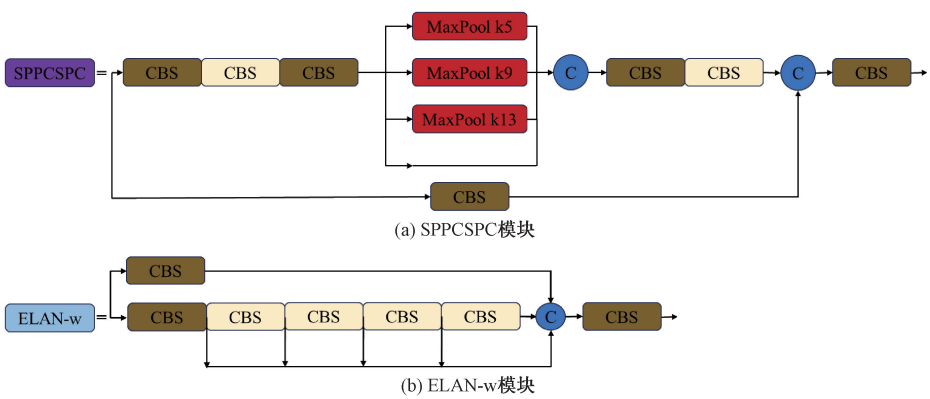

图 3　Neck 网络各模块结构图

部分，提高了模型对复杂场景的适应性。通过对注意力机制的细致调整，YOLO 网络在车辆目标检测任务中表现出更高的准确性和鲁棒性。因此，本文将在 YOLOv7 网络中引入 EMA 注意力机制[21]，如图 5 所示为 EMA 注意力模块。

EMA 注意力机制是基于跨空间学习的高效多尺度注意力模块，首先，对于任意给定的输入特征图 $X \in R^{C\times H\times W}$，EMA 为了能够学习到图像中不同的语义信息，会使输入特征图 X 沿着通道维度方向划分为 G 个子特征图，其中特征分组由 $X=[X_0,X_1,\cdots,X_{G-1}]$，$X_i \in R^{C//G\times H\times W}$ 提供。其次，EMA 为了获取分组特征图的注意力权重描述符，会将子特征图沿着 3 条平行路径同时进行提取，其中第 1 和 2 条路径都在 1×1 分支上，第 3 条路径在 3×3 分支上。一方面，在 1×1 分支中，特征图分别沿 X、Y 方向对

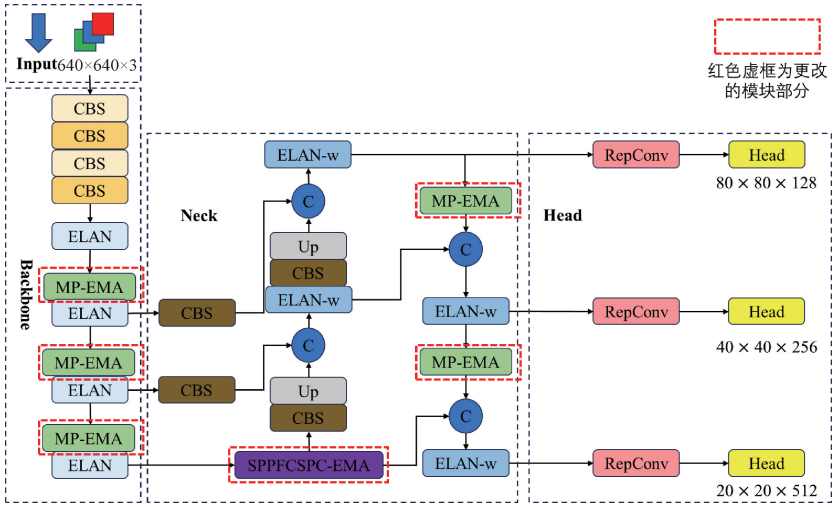

图 4 改进后的 YOLOv7 网络结构图

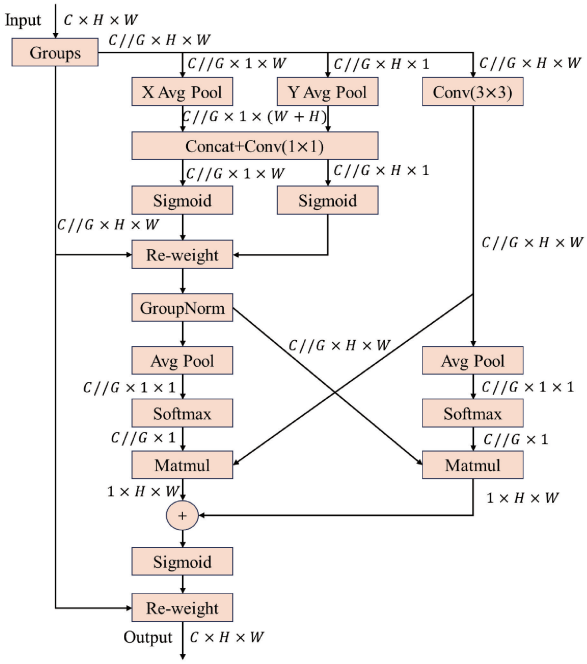

图 5 EMA 注意力模块

通道进行 1D 全局平均池化操作,然后将获取到两个特征图沿图像高度方向进行拼接,使其共同进行 1×1 卷积运算,并将卷积后的输出特征图分解为两个向量,采用两个非线性 Sigmoid 函数对线性卷积进行拟合,以适应 2D 二项分布。接着,为了在 1×1 分支的两条平行路径之间实现不同的跨通道交互特征,通过简单的乘法将每个组内的两个通道注意力图聚合在一起。另一方面,在 3×3 分支中,子特征图通过 3×3 卷积捕获局部区域之间的语义信息交流,以扩展特征表示的范围。这样,EMA 可以根据信道间的语义信息来调整不同信道的重要性,从而进行编码将精确的空间结构信息嵌入到每一个信道中。紧接着,EMA 利用二维全局平均池化对 1×1 支路的输出特征图进行全局空间信息编码,在通道特征联合激活机制之前,将最小支路的输出直接转化为对应的维度形状,即 $R_1^{1\times1\times C//G}\times R_3^{C//G\times H\times W}$。其中 2D 全局池化操作公式如下:

$$z_c=\frac{1}{H\times W}\sum_j^H\sum_i^W x_c(i,j) \tag{1}$$

为了提高计算效率,在 2D 全局平均池化的输出处采用二维高斯映射的自然非线性函数 Softmax 来拟合以上的线性变换。在上述一系列的处理后,将 1×1 分支和 3×3 分支并行处理出的输出特征图进行矩阵点积运算相乘,得到了第 1 个 1×1 分支处的空间注意力特征图。此外,在 3×3 分支处同样进行 2D 全局平均池化获取全局空间信息,在两个分支的通道特征联合机制激活前,1×1 分支处的特征图会直接转换为相应的维度形状,即 $R_3^{1\times1\times C//G}\times R_1^{C//G\times H\times W}$。在与 1×1 分支处的相同处理后,在 3×3 分支处生成了第 2 个空间注意力特征图,在此同样嵌入了精确的空间位置信息。最后,将 1×1 分支与 3×3 分支处的输出特征图相结合,组成一个由两个空间注意力权重值拼接出来的集合,再次使用 Sigmoid 函数对其进行拟合处理。最终,EMA 的输出特征图与输入特征图 X 的尺寸大小相同,同时可以捕获像素级的成对关系,并突出显示所有像素的全局上下文信息。

2.2 SPPFCSPC-EMA 改进

引入了基于空间金字塔池化(spatial pyramid pooling,SPP)结构的 SPPCSPC 模块[22],其中包含跨阶段部分(cross stage partial,CSP)结构,通过添加大的残差边对输入图像进行辅助优化和特征提取,进而提高了模型的精度[23]。然而,这也导致了模型参数和计算量的显著增加,进而减缓了模型的速度。为解决这一问题,借鉴了 YOLO v5 中提出的快速空间金字塔池化(spatial pyramid pooling-fast,SPPF)结构的思想[24],对 SPPCSPC 模块进行改进,得到 SPPFCSPC 模块。主要改进是使用 3 个池化核 k 为 5 的 Maxpool 替代 SPPCSPC 模块中的池化核 k 为 5、9、13

的 3 个 Maxpool，并且将原来的并行结构改变为串并结构，使池化核 k 为 5 的感受野能够与小型车辆的目标尺寸相匹配，从而提取到更多小目标特征，在提升小目标检测精度的同时，也能有效提升检测速度[25]。

SPPFCSPC-EMA 是在 SPPFCSPC 结构的基础上进行了重构的空间金字塔池化模块，旨在提取被遮挡、模糊的车辆特征，以进一步提高识别精度。其结构通过裁剪 CBS 模块和引入 EMA 注意力机制的重构方法对 SPPFCSPC 结构进行优化。相比于 SPPCSPC，以及重构前的 SPPFCSPC，重构后的 SPPFCSPC-EMA 在结构上经过上述调整，旨在更有效地捕捉小目标特征信息。详细结构示意图如图 6 所示。

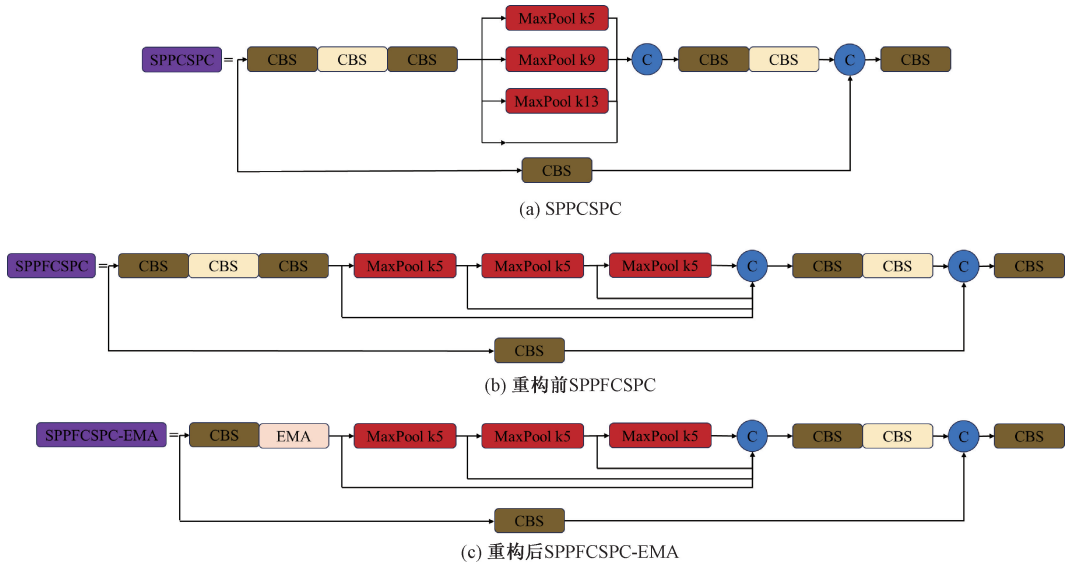

图 6　SPPFCSPC-EMA 重构前后对比图

SPPFCSPC-EMA 重构过程如下：

1）裁剪 CBS 层。为了保留小目标的边缘信息，减少卷积层的过滤效果，本文在最大池化层前裁剪掉 2 个 CBS 层。在最开始只使用一层 CBS 层进行平滑特征，从而降低计算复杂度和参数规模，使此模块能够更有效地处理小目标特征信息。

2）引入 EMA 注意力机制。在池化层之前嵌入基于跨空间学习的高效多尺度注意力模块 EMA，通过融合不同尺度的上下文信息使得特征图产生更好的像素级关注，从而强化小目标区域的关注，降低有效特征的损失。

在 SPPCSPC 重构为 SPPFCSPC-EMA 的过程中，对 CBS 层输出的特征图进行裁剪，从而加快收敛速度，并引入 EMA 注意力机制，有助于聚焦被遮挡、模糊的车辆检测区域。此外，通过缩小池化核，可以有效地匹配上小目标的感受野，从而更精确地提取车辆特征。这一系列的优化措施旨在提升对上述目标的检测效果，以降低车辆错检率和漏检率。

2.3　MP-EMA 改进

由于道路上的监控摄像头通常以俯视视角观察路况，拍摄到的部分车辆目标较小，含有的图像关键信息较少，为了更好地提取车辆的关键特征信息，本文在 MP 结构中，通过在含有两个 CBS 层的支路中裁剪掉第 1 个 CBS 层，再引入 EMA 注意力机制的方法，重构出 MP-EMA 模块，其结构如图 7 所示。在输入特征图 $X \in R^{C\times H\times W}$ 进入 MP-EMA 模块后，存在两个不同的分支同时对其进行处理。其中，第 1 个分支中包括了 1 个池化核为 2 的最大池化层和 1 个 1×1 的 CBS 层，输入特征图首先经过最大池化层进行下采样，使得特征图的长宽变为原来的 1/2，再通过 1×1 的 CBS 层改变特征图经过此分支后的通道数，从而得到第 1 个分支的特征图 $X_1 \in R^{C\times\frac{H}{2}\times\frac{W}{2}}$；第 2 个分支中包括了一个 EMA 注意力机制模块和 1 个步长为 2 的 3×3 的 CBS 层，输入特征图首先经过 EMA 注意力模块进行特征增强，得到增强后的特征图 $\tilde{X}_2$，再通过步长为 2 的 3×3 的 CBS 层进行 2 倍下采样，使得特征图的长宽同样变为原来的 1/2，并且改变特征图经过此分支后的通道数，得到第 2 个分支的特征图 $X_2 \in R^{C\times\frac{H}{2}\times\frac{W}{2}}$。最后，将两个分支输出的特征图进行拼接，得到最终的输出结果 $Y \in R^{2C\times\frac{H}{2}\times\frac{W}{2}}$。其中，输出结果的公式如下：

$$Y = X_1 + X_2 \tag{2}$$

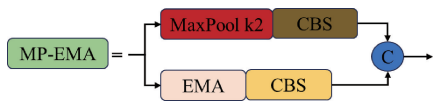

图 7　MP-EMA 结构图

2.4　损失函数替换

YOLOv7 使用 CIoU 作为损失函数，综合考虑了预测框与真实框之间的长宽比、中心点距离以及重叠面积等因

素[26]。利用式(3)～(7)计算 CIoU 为:

$$L_{CIoU}=1-IoU+\frac{\rho^2(b,b^{gt})}{c^2}+\alpha\nu \tag{3}$$

$$\alpha=\frac{\nu}{(1-IoU)+\nu} \tag{4}$$

$$\nu=\frac{4}{\pi^2}\left[\arctan\left(\frac{w_{gt}}{h_{gt}}\right)-\arctan\left(\frac{w}{h}\right)\right]^2 \tag{5}$$

$$\frac{\partial\nu}{\partial w}=\frac{8}{\pi^2}\left[\tan^{-1}\left(\frac{w}{h}\right)-\tan^{-1}\left(\frac{w_{gt}}{h_{gt}}\right)\right]^2\frac{h}{h^2+w^2} \tag{6}$$

$$\frac{\partial\nu}{\partial h}=\frac{8}{\pi^2}\left[\tan^{-1}\left(\frac{w}{h}\right)-\tan^{-1}\left(\frac{w_{gt}}{h_{gt}}\right)\right]^2\frac{w}{h^2+w^2} \tag{7}$$

式中:IoU 表示代表预测框和真实框的交并比,b 和 b^{gt} 分别表示两框的中心点坐标,ρ 代表两者之间的欧氏距离,c 表示其最小外接矩形的对角线长度。α 是作为权重函数,v 是用于衡量预测框和真实框宽高比的一致性的评价指标。式中:w_{gt} 和 h_{gt} 表示真实框的宽和高,w 和 h 表示预测框的宽和高。CIoU 考虑的是预测框与真实框之间的比值,但是存在一些问题,如预测框的宽和高与真实框的宽和高形成一定比值时,惩罚函数会失去效果。此外,式(6)和(7)表明,在 CIoU 中,宽高 w,h 是梯度符号相反的参数,不能同时增大或减小。

因此,本文引入 MPDIoU loss,MPDIoU 是一种基于最小点距的边界盒相似度度量,考虑预测框与真实框的相交比率、中心点距离、宽度和高度偏差[27]。其中,MPDIoU 损失函数以预测框的左上、右下坐标为输入,以特定的方式定义一个矩形,使预测框中两点与真实框的距离最小,该方法可以加速网络预测框的回归收敛,获得更准确的预测结果。MPDIoU 的计算利用式(8)～(10)进行:

$$d_1^2=(x_1^{ped}-x_1^{gt})^2+(y_1^{ped}-y_1^{gt})^2 \tag{8}$$

$$d_2^2=(x_2^{ped}-x_2^{gt})^2+(y_2^{ped}-y_2^{gt})^2 \tag{9}$$

$$MPDIoU=IoU-\frac{d_1^2}{w^2+h^2}-\frac{d_2^2}{w^2+h^2} \tag{10}$$

其中,式(8)和(9)中的 ped 和 gt 分别代表预测框和真实框,(x_1^{ped},y_1^{ped}) 和 (x_2^{ped},y_2^{ped}) 分别表示预测框的左上点和右下点坐标,(x_1^{gt},y_1^{gt}) 和 (x_2^{gt},y_2^{gt}) 表示真实框的左上点和右下点坐标,d_1 和 d_2 分别为预测框和真实框的左上点和右下点之间的距离,式(10)中的 w 和 h 分别代表输入图像的宽度和高度。

3 实验结果与分析

3.1 数据集

本实验所采用的数据集为 SAKSHAM JAIN 采集到的车辆类别目标检测数据集[28],其中本文选择了 6 类,分别是 Motorcycle、Auto、Car、Bus、Truck、Tractor。数据集共有 8 219 张图片,按照 7∶2∶1 的比例划分成训练集、验证集、测试集。

3.2 实验环境

实验环境如表 1 所示。

表 1 实验环境

名称	配置
操作系统	Windows10
CPU	12th Gen Intel ® Core(TM)i7-12700K
GPU	NVIDIA GeForce RTX 3060(12 G 显存)
编程语言	Python3.9
算法框架	Pytorch1.13
加速环境	CUDA11.7

3.3 性能指标评价

本文采用的算法评价指标,包括参数量(parameters)、浮点运算次数(GFLOPs)、准确率(precision)、召回率(recall)、平均精度(mAP)和每秒传输帧数(FPS),以上指标为全面评估算法性能提供了综合的度量标准。相关的计算公式如下:

$$Percison=\frac{TP}{TP+FP} \tag{11}$$

$$Recall=\frac{TP}{TP+FN} \tag{12}$$

$$mAP=\frac{\sum AP}{N} \tag{13}$$

$$FPS=\frac{\text{Frames}}{Time} \tag{14}$$

式中:TP 表示正确检测框数,FP 表示错误检测框数,FN 表示漏检测框数,$\sum AP$ 指所有检测类别的总 AP 值,N 指所有检测类别总数,$Frames$ 代表帧数,$Time$ 代表检测时间。

3.4 消融实验结果分析

在 SAKSHAM JAIN 采集到的车辆类别目标检测数据集上进行消融实验验证本文改进方法的有效性,其中共设置了 6 组实验。表 2 中"√"表示引入的改进方法,其中包括改进的 SPPFCSPC-EMA、MP-EMA 和 MPDIoU 模块。评价指标中通过采用 mAP 来对比每组改进后的目标检测的精度,采用 Params、GFLOPS 来对比每组改进后的参数量和计算量的大小,采用 FPS 来对比每组改进后的实时性情况。消融实验结果如表 2 所示。

从表 3 可以得出,第 1 组为初始的 YOLOv7 算法,mAP 为 83.86%,FPS 为 40.65,作为评价指标基线;第 2 组为加入改进的 SPPFCSPC-EMA 模块,mAP 上升了 0.53%,参数量下降了 6.94%,计算量下降了 1.77%,FPS 略微上升到 41.02,说明加入改进的 SPPFCSPC-EMA 模块能够强化对小目标区域的关注,取更准确的车辆特征,同时也可以降低模块的参数量和计算量,提升推理速度;第 3 组为加入改进的 MP-EMA 模块,mAP 上升了 2.15%,参数量上升了 6.86%,计算量上升了 11.5%,FPS 下降到 32.46,说明 MP-EMA 模块能够很有效获取车辆的关键信息特征,但会使模块的参数量和计算量都得到上

表 2　消融实验结果

组别	SPPFCSPC-EMA	MP-EMA	MPDIoU	mAP/%	Params/M	GFLOPs/G	FPS
1				83.86	37.222	105.199	40.65
2	√			84.39	34.639	103.339	41.02
3		√		86.01	39.774	117.291	32.46
4			√	85.34	37.222	105.199	40.65
5	√	√		86.29	37.192	115.430	32.82
6	√	√	√	**86.69**	**37.192**	**115.430**	**32.82**

升，降低推理速度；第 4 组为使用 MPDIoU 损失函数，可以具有更高的边界框回归效率和精度，mAP 上升了 1.48%，FPS 保持在 40.65；第 5 组为在第 2 组的基础上加入改进的 MP-EMA 模块，mAP 上升了 2.43%，FPS 下降到 32.82，说明改进的 SPPFCSPC-EMA 与 MP-EMA 模块在组合起来后可以提升对车辆检测效果；第 6 组为本文所提最终算法，算法包括了改进的 3 个模块，mAP 较原 YOLO v7 算法上升了 2.83%，达到了 86.69%，参数量下降了 0.08%，计算量上升了 9.7%，FPS 下降到 32.82，证明了本文各个改进方法对道路监控下车辆目标检测的有效性，但是降低了少量的参数量却使计算量得到上升，使得模型推理速度下降。

3.5　对比实验结果分析

为了全面验证本文算法的有效性，本文对提出的最终算法与其他一系列的主流目标检测算法进行了综合对比实验，其中包括了 Faster R-CNN、SSD、YOLOv4、YOLOv5_l、YOLOv8_l、RetinaNet、CenterNet 和 RT-DETR。通过在数据集上进行 mAP、Params、GFLOPS 和 FPS 的比较，本文能够全面了解各算法在目标检测精度、参数量、计算量和实时性的性能表现。实验结果详如表 3 所示。

从表 3 的实验结果可以得出，本文方法在该车辆类别目标检测数据集上的 mAP 相较于两阶段检测网络 Faster R-CNN[29] 提高了 7.96%，相较于单阶段模型 SSD[30] 提高了 9.08%。通过对比 YOLO 系列的 YOLOv4[31]、YOLOv5_l[32] 和 YOLOv8_l[33]，本文方法分别提高了 20.76%、4.33% 和 1.21%。而本文方法相较于近年发展出新单阶段算法 RetinaNet[34] 和 CenterNet[35]，分别提高了 1.62% 和 0.69%。最后，通过对比最新的实时目标检测算法 RT-DETR[36]，本文方法提高了 0.56%，达到了 86.69%。但可以发现，本文方法相较于其他方法显示出 FPS 相对较低，推理速度上较低。综上所述，本文所提算法与其他主流算法相比，在牺牲了一点检测速度的前提下，可以满足在车辆检测上准确性的需求。

表 3　不同网络对比实验结果

算法	mAP/%	Params/M	GFLOPS/G	FPS
Faster R-CNN	78.73	136.791	401.813	16.23
SSD	77.61	24.280	275.372	24.80
YOLOv4	65.93	63.965	59.989	57.52
YOLOv5_l	82.36	46.658	114.645	42.27
YOLOv8_l	85.48	43.634	165.425	34.04
RetinaNet	85.07	36.434	146.906	18.57
CenterNet	86.00	32.665	70.217	72.13
RT-DETR	86.13	19.882	114.000	50.86
本文方法	**86.69**	**37.192**	**115.430**	**32.82**

3.6　可视化分析

为了直观地展示本文方法在不同场景下的检测效果，本文在数据集中选取了白天道路场景下和晚上夜晚道路场景下的两张图片进行测试，检测效果如图 8～10 所示。从图像中可以明显看出，在白天和夜间道路场景下，本文算法在道路监控下的车辆方面漏检现象有所降低，同时车辆检测精度也相对更高，因此改进后的 YOLOv7 模型能够有效地对道路监控场景下的车辆进行检测和识别。

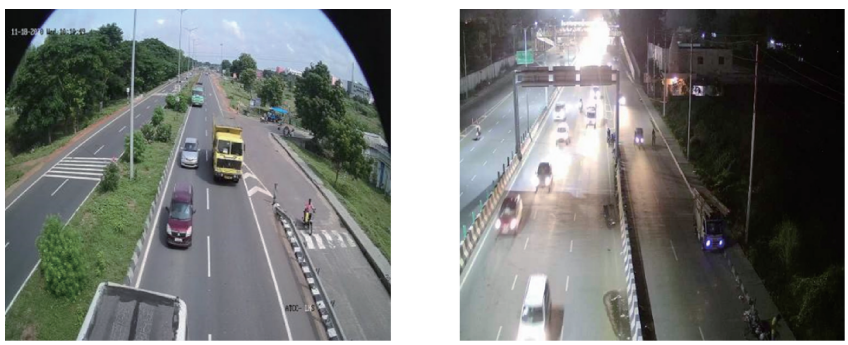

图 8　原始图像

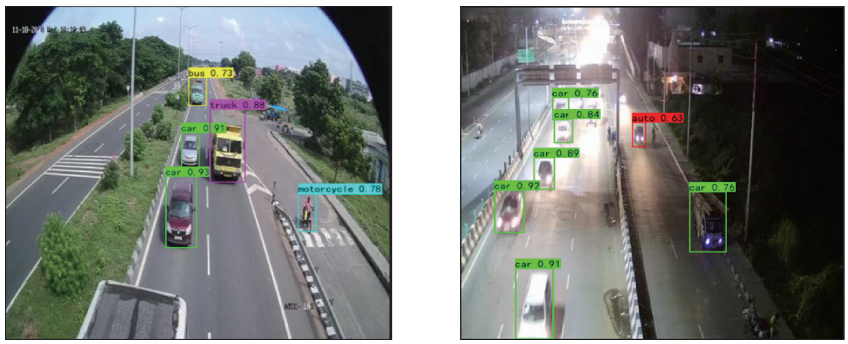

图 9 YOLO v7 检测效果图

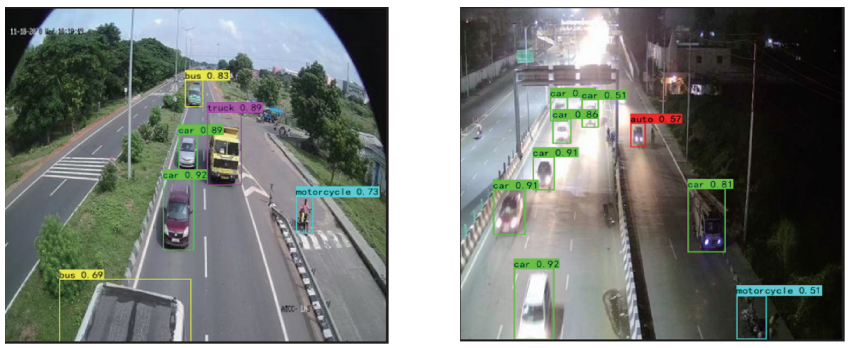

图 10 本文算法检测效果图

4 结 论

为了使计算机可以准确地对车辆进行目标识别,解决在道路监控下的车辆目标检测出现的各种问题,本文提出一种改进的 YOLOv7 车辆检测方法。通过裁剪部分卷积层和引入 EMA 注意力机制的方法,改进出 SPPFCSPC-EMA 和 MP-EMA 模块,分别用来提高对车辆的检测以及强化车辆关键的特征信息,并且更换损失函数 MPDIoU 来加速网络预测框的回归收敛,获得更准确的预测结果。实验结果表明,改进后的算法在道路监控下的车辆目标检测上,可以提高整体的平均精度。但仍存在模型参数量较大、检测速度相对较低等不足,在未来的工作中,将从轻量化网络方面进行深入研究,使网络能拥有更好的有效性和实时性。

参考文献

[1] 张茹.基于深度学习的车型检测研究[D].兰州:西北师范大学,2022.

[2] XIE X, CHENG G, WANG J, et al. Oriented R-CNN for object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 3520-3529.

[3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 38(1): 142-158.

[4] FAN Q, BROWN L, SMITH J. A closer look at Faster R-CNN for vehicle detection[C]. 2016 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2016: 124-129.

[5] DAI J, LI Y, HE K, et al. R-FCN: Object detection via region-based fully convolutional networks[C]. Proceedings of the 30th International Conference on Neural Information Processing Systems, 2016: 379-387.

[6] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.

[7] JEONG J, PARK H, KWAK N. Enhancement of SSD by concatenating feature maps for object detection[J]. ArXiv preprint arXiv:1705.09587, 2017.

[8] RAJENDRAN S P, SHINE L, PRADEEP R, et al. Fast and accurate traffic sign recognition for self driving cars using retinanet based detector[C]. 2019 International Conference on Communication and Electronics Systems(ICCES), IEEE, 2019: 784-790.

[9] DUAN K, BAI S, XIE L, et al. Centernet: Keypoint triplets for object detection[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 6569-6578.

[10] ZHAO Y, LYU W, XU S, et al. Detrs beat YOLOs on real-time object detection[J]. ArXiv preprint arXiv:2304.08069,2023.

[11] 朱茂桃,邢浩,方瑞华.基于 YOLO-TridentNet 的车辆检测方法[J].重庆理工大学学报(自然科学),2020,34(11):1-8.

[12] 郭宇阳,胡伟超,戴帅,等.面向路侧交通监控场景的轻量车辆检测模型[J].计算机工程与应用,2022,58(6):192-199.

[13] LI S, YANG X, LIN X, et al. Real-time vehicle detection from UAV aerial images based on improved YOLOv5[J]. Sensors, 2023, 23(12): 5634.

[14] 蔡刘畅,杨培峰,张秋仪.基于 YOLOv7 的道路监控车辆检测方法[J].陕西科技大学学报,2023,41(6):155-161,175.

[15] 薛震,张亮亮,刘吉.基于改进 YOLOv7 的融合图像多目标检测方法[J].兵器装备工程学报,2023,44(6):166-172.

[16] 许晓阳,高重阳.改进 YOLOv7-tiny 的轻量级红外车辆目标检测算法[J].计算机工程与应用,2024,60(1):74-83.

[17] 杜娟,崔少华,晋美娟,等.改进 YOLOv7 的复杂道路场景目标检测算法[J].计算机工程与应用,2024,60(1):96-103.

[18] 张利丰,田莹.改进 YOLOv8 的多尺度轻量型车辆目标检测算法[J].计算机工程与应用,2024,60(3):129-137.

[19] ZHAI S, SHANG D, WANG S, et al. DF-SSD: An improved SSD object detection algorithm based on DenseNet and feature fusion[J]. IEEE Access, 2020, 8: 24344-24357.

[20] 吴旭红,赵清华.基于改进 YOLOv7 的无人机航拍图像目标检测[J].电光与控制,2024,31(2):35-40,111.

[21] OUYANG D, HE S, ZHANG G, et al. Efficient multi-scale attention module with cross-spatial learning[C]. ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), IEEE, 2023: 1-5.

[22] HE K, ZHANG X, REN S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 37(9): 1904-1916.

[23] WANG C Y, BOCHKOVSKIV A, LIAO H Y M. Scaled-YOLOv4: Scaling cross stage partial network[C]. Proceedings of the IEEE/cvf Conference on Computer Vision and Pattern Recognition, 2021: 13029-13038.

[24] 邱天衡,王玲,王鹏,等.基于改进 YOLOv5 的目标检测算法研究[J].计算机工程与应用,2022,58(13):63-73.

[25] 胡淼,姜麟,陶友凤,等.改进 YOLOv7 的自动驾驶目标检测算法[J/OL].计算机工程与应用,1-11[2024-01-31]. http://kns.cnki.net/kcms/detail/11.2127.TP.20230922.1630.004.html.

[26] ZHENG Z, WANG P, LIU W, et al. Distance-IoU loss: Faster and better learning for bounding box regression[C]. Proceedings of the AAAI Conference on Artificial Intelligence, 2020, 34(7): 12993-13000.

[27] SILIANG M, YONG X. MPDIoU: A loss for efficient and accurate bounding box regression[J]. ArXiv preprint arXiv:2307.07662,2023.

[28] JAIN S. Vehicle detection 8 classes[DS]. (2020)[2024-05-18]. https://www.kaggle.com/datasets/sakshamjn/vehicle-detection-8-classes-object-detection.

[29] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.

[30] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, 2016: 21-37.

[31] BOCHKOVSKIV A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. ArXiv preprint arXiv:2004.10934,2020.

[32] JOCHER G, STOKEN A, BOROVEC J, et al. Ultralytics/YOLOv5:v3.1-bug fixes and performance improvements[EB/OL]. (2020-10-29)[2024-05-21]. https://zenodo.org/record/4154370.

[33] TERVEN J R, CORDOVA-ESPARAZA D M, et al. Ultralytics/YOLOv8:YOLOv8 docs[EB/OL]. (2023-01-10)[2024-05-21]. https://ultralytics.com/YOLOv8.

[34] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.

[35] ZHOU X, WANG D, KRAHENBUHL P. Objects as points[J]. ArXiv preprint arXiv:1904.07850,2019.

[36] ZHAO Y, LYU W, XU S, et al. Detrs beat YOLOs on real-time object detection[J]. ArXiv preprint arXiv:2304.08069,2023.

作者简介

过鑫炎,硕士,主要研究方向为图像处理。

E-mail:guoxy1023@163.com

朱硕(通信作者),讲师,博士,主要研究方向为数字媒体信息处理、机器视觉、光电信息处理。

E-mail:zshuo2011@163.com