

复杂场景下的交通标志小目标检测算法^{*}

王 浩 张其猛 龚德成

(上海应用技术大学计算机科学与信息工程学院 上海 201418)

摘要: 在交通标志识别应用中,待检测目标多为小目标,易出现漏检、误检等问题。针对这些问题,基于YOLOv8s算法设计了一种改进的交通标志识别算法,FKDS-YOLOv8s。使用FasterBlock重构C2f模块,形成新的轻量化模块C2f-Faster,既提升模型特征提取能力,又降低了计算开销;基于SENet和ResNeXt模型设计一种新的检测头Detect_SR,使模型能够有效地聚焦于小目标的关键特征;融入轻量且高效的动态上采样器DySample,显著减少了GPU内存消耗;通过增加上采样和Prediction输出层次,模型能够捕捉丰富的位置信息,有效解决了YOLOv8s模型在处理小目标时信息不足的问题;引入Shape-IoU损失函数,优化了原CIoU在边框回归中的不足;此外,在Neck部分融入了本文新设计的注意力机制DKN-Attention,在上采样和下采样过程中定位微小物体场景的注意力区域,提升了远处小型交通标志的特征提取和识别能力。实验在中国交通标志数据集TT100K上进行,结果表明,FKDS-YOLOv8s相比基准模型,在查准率(P)、查全率(R)和mAP50上分别提升了5.9%、4.2%和6.3%。较传统方法,FKDS-YOLOv8s在性能上表现出显著优势。

关键词: 交通标志识别;动态上采样;注意力机制;轻量化;Shape-IoU

中图分类号: U491; TP391; TN919.82 文献标识码: A 国家标准学科分类代码: 520.20

Small target detection algorithm for traffic signs in complex scenes

Wang Hao Zhang Qimeng Gong Decheng

(School of Computer Science and Information Engineering, Shanghai Institute of Technology, Shanghai 201418, China)

Abstract: In the application of traffic sign recognition, most of the targets to be detected are small targets, which are prone to problems such as missed detection and false detection. In order to solve these problems, an improved traffic sign recognition algorithm, FKDS-YOLOv8s, was designed based on the YOLOv8s algorithm. FasterBlock is used to reconstruct the C2f module to form a new lightweight module C2f-Faster, which not only improves the feature extraction ability of the model, but also reduces the computational overhead. Based on the SENet and ResNeXt models, a new detection head Detect_SR was designed to enable the model to effectively focus on the key features of small targets. DySample, a lightweight and efficient dynamic upsampler, significantly reduces GPU memory consumption. By increasing the output level of upsampling and prediction, the model can capture rich position information, which effectively solves the problem of insufficient information when the YOLOv8s model processes small targets. The Shape-IoU loss function is introduced to optimize the shortcomings of the original CIoU in the border regression. In addition, the newly designed attention mechanism DKN-Attention is integrated into the Neck part, which locates the attention area of the scene of small objects during the upsampling and downsampling process, and improves the feature extraction and recognition ability of small traffic signs in the distance. The experimental results were carried out on the Chinese traffic sign dataset TT100K, and the results showed that FKDS-YOLOv8s improved the accuracy (P)、recall rate (R) and mAP50 by 5.9%、4.2% and 6.3%, respectively, compared with the benchmark model. Compared with the traditional method, FKDS-YOLOv8s shows significant advantages in performance.

Keywords: traffic sign recognition; dynamic upsampling; attention mechanism; lightweight; Shape-IoU

0 引言

在自动驾驶和驾驶辅助等智慧交通技术的发展过程

中,交通标志识别检测技术已成为交通控制的重要组成部分,通过实时图像检测,该技术能够识别限速、警告、禁止等交通标志,为驾驶员和智能驾驶系统提供实时的导航与道

路信息。交通标志检测技术已取得了卓越进展,但在真实交通应用环境中,有效且快速地检测小型交通标志仍面临诸多挑战:复杂环境下容易出现误检和漏检;交通标志尺寸较小且特征不明显,增加了检测难度;难以在保证检测精度的同时降低计算复杂度。本文研究的基于 YOLOv8s 的交通标志识别算法,重点提升了对远处微小交通标志的识别准确率,确保搭载此算法的智能辅助系统在车辆行驶中能够提前发现远处的交通标志,对保障交通安全具有重要意义。

目前,交通标志检测的方法研究主要分为两大类:基于传统手工特征的方法和基于深度学习的方法。早期基于传统手工特征的交通检测方法主要有基于形状特征的方法、基于颜色特征的方法以及多特征融合的方法。文献[1]提出了一种基于定向梯度直方图(histogram of oriented gradients, HOG)检测方法,通过在 CIELab 和 YCbCr 颜色空间中引入颜色信息来增强特征向量的判断能力,从而获得了较高的检测性能。杨斐等^[2]通过改进基于颜色的图像快速分割算法,对交通标志检测信息进行分割压缩处理,可以确保交通标志信息分割的完整性和时效性。传统的交通标志检测方法虽然在实时性和鲁棒性方面表现较好,但对于小型交通标志的检测,手工设计的特征无法充分表示标志的细微差别,这些方法常常无法满足实际需求。

基于深度学习的目标检测算法主要分为两类:一种是二阶段(Two-Stage)检测方法,代表性算法包括 Faster R-CNN^[3]、SPPNet^[4]等,这类方法通过候选区域网络在特征层生成候选框,然后利用这些候选框中的卷积特征进行分类和精确定位。二阶段检测算法因计算量大,检测速度较慢,不适合部署到车载设备当中,且无法满足实时性要求。单阶段(One-Stage)检测方法直接将整张图像作为网络输入,在输出层同时进行边界框的位置回归和类别预测,因此具有更高的推理速度。Liu 等^[5]提出了单次多框目标检测方法 SSD(single shot multibox detector, SSD),通过金字塔结构对图像进行下采样,捕捉不同尺度的特征信息,但是 SSD 在训练时对预训练模型依赖较大。YOLO(you only look once, YOLO)系列^[6-13]的核心思想是将输入图像分割成固定大小的网格,每个网格直接预测边界框的位置以及所属类别的置信度,该方法的优势在于极快的检测速度,但精度上往往不如 R-CNN 系列的检测方法;其中 YOLOv8 是较新的 YOLO 系列算法,兼顾了识别精度和识别效率,通过将 YOLOv5 中的 C3 模块替换成 C2f 模块,YOLOv8 在保证模型轻量化的基础获得了丰富的梯度流信息,并且 YOLOv8 中引入了无锚框预测方法,直接预测对象中心位置,加速了推理过程。

近年来的深度学习交通标志检测研究中,陈朋弟等^[14]提出了一种基于 Mask R-CNN 的交通标志识别方法,该方法通过改进锚框的宽高比,大幅提升了检测精度,可达到 98.33%,但每秒仅能处理约 5 张图片;双阶段目标检测算

法在精度上表现优异,但推理速度较慢,实时性能相对较差。侯卓成等^[15]提出在 YOLOv4 网络中引入 Ghost Net 作为基础网络来对彩色数字仪表进行识别,检测速度提高到 37.2 FPS;李志刚等^[16]提出了一种轻量化 YOLOv5 交通标志识别方法,采用 K-means 聚类算法优化锚框计算,并结合 Stem 模块与 ShuffleNetV2 模块,达到了 95.9% 的检测精度和 20 FPS 的检测速度,尽管检测精度有所提升,但检测速度仍显得过慢;赵会鹏等^[17]提出了一种改进的 Ghost-YOLOv8 轻量化检测模型,该模型设计了全新的 GhostC2f 模块取代原有的 C2f 模块,从而实现了模型轻量化,此外在 Neck 部分还引入了 GAM 注意力机制模块,以增强特征中的语义和位置信息,最终将检测速度提升至 294 FPS,检测速度虽然得到了大幅提升,但模型的平均检测精度仅为 71.8%。在驾驶过程中,目标识别速度和检测精度至关重要,通常情况下,人眼在低速下的连贯图像识别速度为 30 FPS,但随着驾驶速度的提升,道路场景的变化加速,检测速度至少需要达到 50 FPS,才能满足高速行驶中的实时检测需求。目前的检测算法多适用于低速行驶下的交通标志识别,尚未能满足高速行驶中的检测要求。因此,未来的研究仍需进一步优化模型,实现模型轻量化与检测精度之间的平衡。

在 TT100K^[18]数据集中,约 81% 的交通标志在图像中所占像素不到 1%,且常伴随褪色、遮挡等现象,因此现有的一阶段检测算法在交通标志识别与检测中,仍面临误检、漏检等问题。考虑到 YOLOv8 采用 C2f 模块能够更有效地捕捉小目标特征信息,并且无锚框预测显著减少了误差积累,本文针对自然场景下交通标志检测存在漏检和虚警问题,在 YOLOv8s 模型的基础上设计了一种用于交通信号小目标检测的 FKDS-YOLOv8s(faster dika dysample senet-yolov8s, FKDS-YOLOv8s)模型,用于提高交通标志的检测准确率。

1 YOLOv8s 网络模型结构

YOLOv8 提供了 5 个不同版本:YOLOv8-n、s、m、l、x。其中,YOLOv8s 具有较快的推理速度和较小的网络参数量,为了平衡检测精度和速度,本文选用 YOLOv8s 作为交通标志检测的基线模型。YOLOv8 的模型结构可以分为 4 个主要部分,YOLOv8 的网络模型结构如图 1 所示。

1) 输入端:负责将输入图片调整为训练所需的尺寸,并进行 Mosaic 等相关的数据增强操作。

2) 骨干网络:负责从输入的图片中提取对网络模型训练所需的特征。Backbone 采用了跨阶段局部网络结构,增强特征提取能力,该结构主要由卷积模块(Conv)、C2f 模块和 SPPF 模块组成。SPPF 模块属于空间金字塔池化结构,通过扩展感受野,实现了局部特征与全局特征的融合。

3) 颈部:负责对 Backbone 中提取到的特征信息进行融合。Neck 分优化了 YOLOv5 的设计,去除了上采样阶段

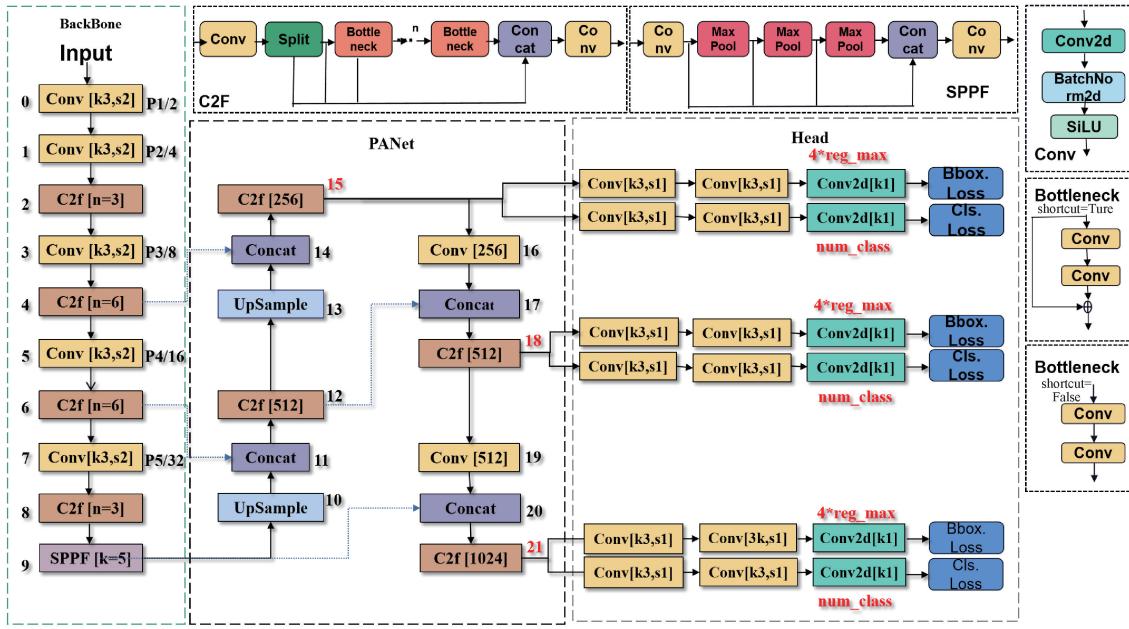


图 1 YOLOv8 网络模型结构图

Fig. 1 Structure diagram of the YOLOv8 network model

的卷积操作,并将 C3 模块替换为 C2f 模块,显著提升了模型在不同尺寸目标上的检测性能。

4) 检测头:采用了解耦头结构,将检测任务和分类任务分离开来,增强了模型的泛化能力和鲁棒性。

2 改进的 YOLOv8s 算法

2.1 改进的 YOLOv8s 网络模型结构

为了降低现有检测算法对交通标志的误检率和漏检率,文献[19]提出了一种多尺度交通标志检测模型 CR-YOLOv8,在特征提取阶段,引入 CBAM 注意力模块来增强通道和空间特征,使网络更容易地学习到小目标的关键信息,在特征融合阶段引入 RFB 模块,以更少的计算开销提高了特征多样性,通过引入 WIoU 损失函数,模型在训练过程中能够有效平衡多尺度目标,提高了模型泛化能力和检测精度,但是由于 CBAM 需要额外计算来生成注意力权重,增加了模型的计算复杂度,因此对资源受限的嵌入式设备并不友好,此外,CBAM 的设计主要基于通道和局部空间信息,难以在复杂背景或多目标场景中充分捕捉全局上下文信息,并且其注意力机制依赖于局部卷积操作,无法有效建模特征图中的长距离依赖关系。而在交通标志检测任务中,目标往往较小且距离较远,这进一步限制了 CBAM 的表现,导致在该模型中增加 CBAM 注意力机制,参数量大幅度增加的情况下,检测精度提升的并不是太理想。

本文基于以上研究从以下几个点进行改进:1) 构造轻量化模块 C2f-Faster,降低主干网络的模型参数,加快识别速度;2) 引入由 ResNeXt^[20]和 SENet^[21]结合而设计出的

一种新颖的检测头 Detect_SR (detect with senet and resnext, Detect_SR),聚焦于小目标关键特征;3) 融入 DySample^[22]超轻量动态上采样算子,使网络模型在上采样过程中具备极高的灵活性和效率;4) 增添小目标检测头,提升网络模型的全局感知能力;5) 使用 Shape-IoU 损失函数优化原损失函数 CIoU 在边框回归中存在的不精确问题;6) Neck 中引入本文新设计的 DKN-Attention (deformable large kernel attention with denv3, DKN-Attention) 注意力机制,使网络模型具备从大量信息中有选择地筛选出少量重要信息并聚焦到这些重要信息上,改进后的网络模型结构如图 2 所示。

2.2 C2f-Faster 模块设计

在真实道路场景中对交通标志检测时,要求检测设备具备实时性和便携性,这就需要嵌入检测设备中的检测算法具备较高的检测速度和较少的计算量。目前卷积神经网络中常用的轻量化网络有 EfficientNet^[23]、MobileNet^[24]、ShuffleNet^[25]和 GhostNet^[26],这些轻量化的卷积神经网络都是利用组卷积(GConv)或者深度卷积(DWConv)来进行空间特征的提取。文献[27]针对 PCB 缺陷检测方法存在效率低、误检率高、通用性低、实时性差等问题,引用 MobileNetV3 作为特征提取网络,提升对 PCB 小目标缺陷的检测性能,同时使用改进二分 K-means 聚类结合交并比(IoU)损失函数确定锚框,解决预设锚框不适用 PCB 小目标缺陷检测的问题,提升检测精度,设计出基于改进 YOLOv4 算法的 PCB 缺陷检测方法。这些轻量化的研究通过减少浮点数的运算量(FLOPs),以取得每秒更少的浮点运算次数,但是由于这些卷积的算子会

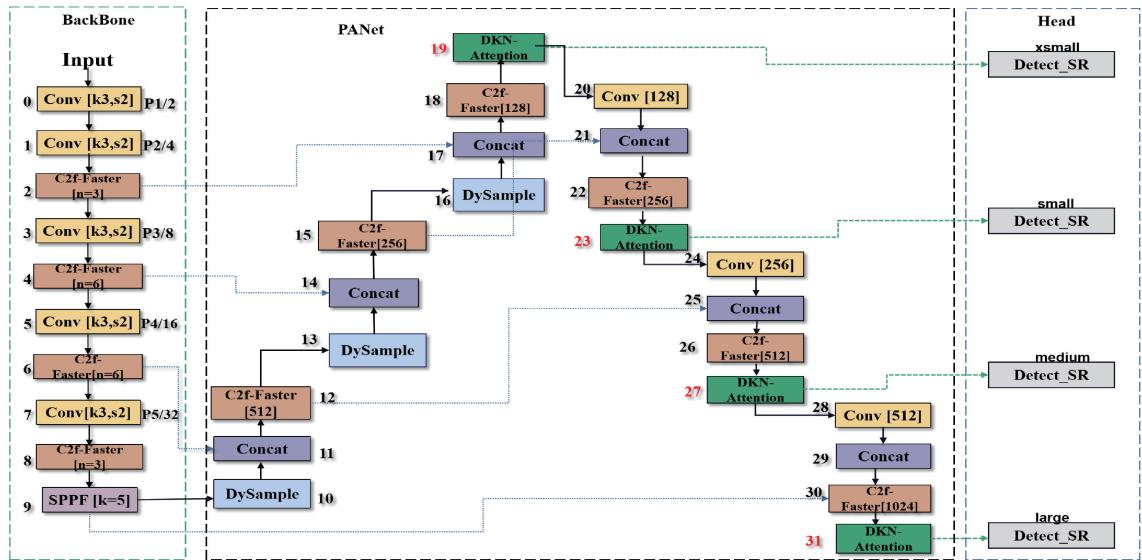


图2 FKDS-YOLOv8s模型结构图

Fig. 2 Structure diagram of the FKDS-YOLOv8s model

频繁的进行访问内存,导致在深度卷积中,模型的计算效率并不理想。FLOPs 和延时时间之间的关系为:

$$Latency = \frac{FLOPs}{FLOPS} \quad (1)$$

式中:FLOPs 代表浮点运算次数,可以用来衡量模型和算法的复杂度;FLOPS 为每秒浮点运算次数的缩写,可以作为有效计算速度的度量。

为避免轻量化网络出现的上述问题并实现更快的神经网络,使用一种新颖的部分卷积(partial convolution, PConv),在减少冗余的计算和内存访问的同时实现更高效地提取空间特征。PConv 与其他卷积对比如图 3 所示。

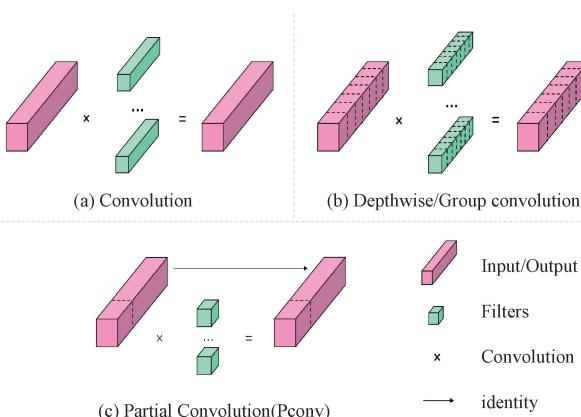


图3 卷积对比

Fig. 3 Convolution comparison

在 PConv 的基础上进一步衍生出 FasterNet^[28]框架,可以在各种设备上高速运行,FasterNet 网络结构图如图 4 所示。FasterNet 的整体架构包含 4 个 stage,每个 stage 都由 1 个 FasterNet block 组成,前面是一个嵌入层(步长为 4

的常规卷积 4×4)或合并层(步长为 2 的常规卷积 2×2),用扩展空间下的采样和信道数,最后 3 个图层用于特征分类。在每个 FasterNet block 中, PConv 层后面与两个 PWConv 层或 1×1 的常规卷积层相连,形成倒置残差快。

基于 FasterNet 具备较高的检测速度和较少的计算量,将 C2f 中的 Bottleneck 模块替换为 FasterNet 中的 FasterBlock 模块,构造一个新的轻量化模块,命名为 C2f-Faster,以此来减少模型的参数量,网络结构如图 5 所示。

2.3 Detect_SR 检测头设计

YOLOv8 目标检测算法的检测头采用了解耦检测头设计,将分类任务和回归任务独立处理。TT100K 交通标志数据集中交通标志类别众多,且每种标志的外观形状差异显著。在这种多分类且小目标密集的检测场景中,传统的解耦检测头会出现漏检情况。

ResNeXt 是一种简单、高度模块化的网络结构,在图像分类应用领域中尤为出色。ResNeXt 通过使用多个相同的低维特征图,在卷积操作之后,将多组结构进行求和,使用平衡堆叠相同拓扑结构的 blocks,通过在同一层内使用多个并行路径提升了模型在分类任务下的准确率。

SENet 考虑了特征通道之间的关系,在特征通道上加入了注意力机制,通过学习的方式自动获取每个特征通道的重要程度,利用得到的重要程度来提升特征并抑制对当前任务不重要的特征,网络结构如图 6 所示。回归任务需要准确预测目标的边界框,模型需要精确提取目标的形状、大小和位置信息,在目标边界不清晰或被遮挡的情况下,SENet 的特征加权机制能够强化与目标形状和边缘相关的重要特征。

基于 SENet 与 ResNeXt 模块的优点,本文通过 SENet 和 ResNeXt 融合形成 SEResNeXt 模块,如图 7 所示,通过

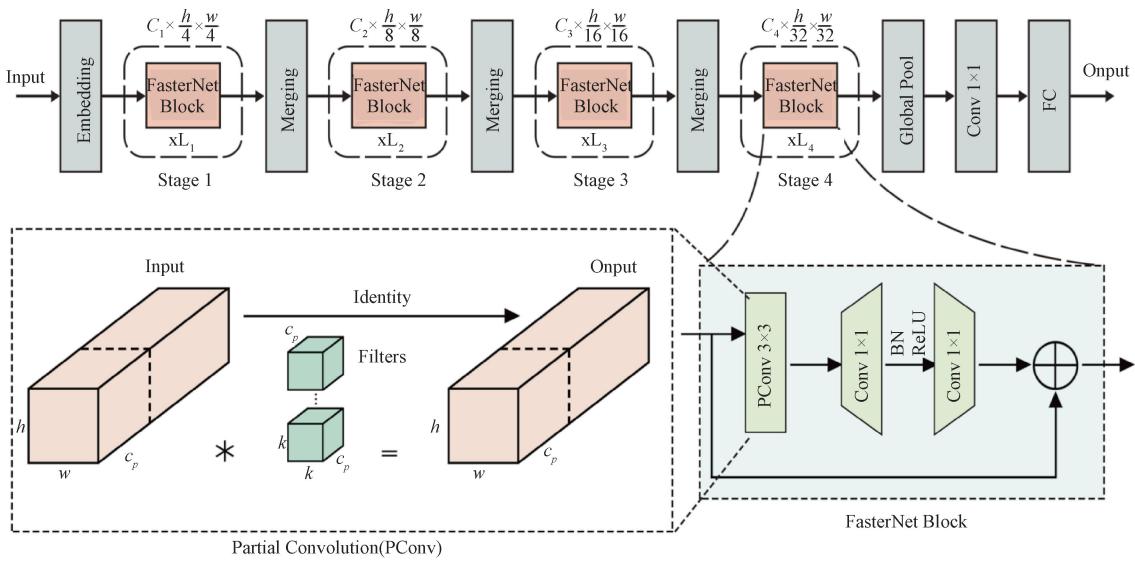


图 4 FasterNet 网络结构图

Fig. 4 FasterNet network structure

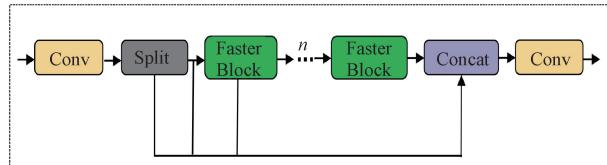


图 5 C2f-Faster 模块结构图

Fig. 5 Structure diagram of the C2F-FASTER module

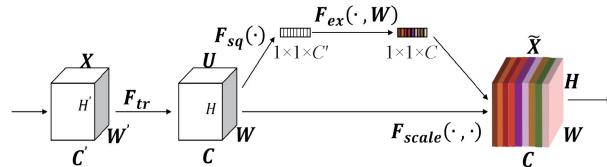


图 6 SENet 网络模型结构图

Fig. 6 Structure diagram of SENet network model

将 SEResNeXt 模块融入原解耦头, 形成新的检测头 Detect_SR, 网络结构如图 8 所示。通过使用新的检测头使得分类分支更加侧重于确定与目标类别最匹配的特征, 定位分支专注于优化边界框的中心点坐标、宽度和高度, 即分类任务主要关注目标的核心区域内容, 而回归任务则更多地关注边界信息的精确定位, 有效降低模型在小目标检测时的漏检率和误检率。

2.4 动态上采样器 DySample

上采样器(Upsample)是用于提高特征图分辨率的重要组成部分,通常在检测小物体时发挥重要作用。YOLOv8的上采样方法,遵循固定的规则来插值上采样,在处理小物体或细微边缘时,精度可能有所下降。DySample是一种超轻量的动态上采样器,从点采样的角度制定了上采样的规则,可以根据输入特征的空间分布动

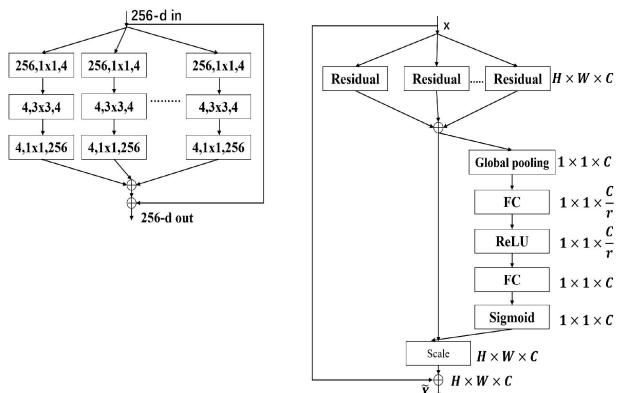


图 7 ResNeXt 模块(左)与 SEResNeXt 模块(右)

Fig. 7 ResNeXt module (left) and SEResNeXt module (right)

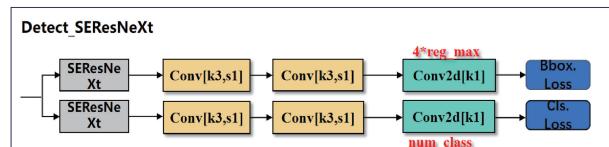


图 8 Detect SR 检测头结构图

Fig. 8 Structure diagram of Detect_SR detection head

态调整插值权重,这比传统的固定插值方法更加灵活,可以更好地捕捉复杂的图像结构和细节,有效保持目标的形状和边缘。由于交通标志属于小目标检测,而 DySample 在处理复杂细节和小物体时表现优异,因此将其引入到 Neck 结构中,替代传统的上采样方法,采样时优先选择那些对模型训练更有贡献的正样本,抑制负样本对模型训练过程中产生的干扰,极大优化了上采样效率。

2.5 添加小目标检测头

在 YOLOv8 系列模型中, Neck 部分通过两次上采样

后结合C2f模块进行特征融合,并利用尺寸分别为 20×20 、 40×40 和 80×80 的特征图进行目标检测,由于交通标志的尺寸比较小,而YOLOv8s的下采样倍数较大,随着网络层数的加深,小目标的特征信息很容易被忽略^[29]。当输入图像尺寸过大时,模型难以精确捕捉图像中的细微信息,导致小目标检测出现误检或漏检的现象。根据国际组织SPIE的定义,小目标是指在 256×256 分辨率图像中,目标面积小于80个像素的目标,即占比小于0.12%的目标。在TT100K数据集中,约有81%的交通标志属于小目标,因此,本文在Neck部分额外增加了一次上采样,并新增了一个预测输出层。改进后,预测特征图的尺寸扩展为 20×20 、 40×40 、 80×80 和 160×160 ,增加了 160×160 的检测头,专门用于小目标检测,使模型能够获取更多、更丰富的位置信息。改进后的模型从主干网络开始进行初步特征提取,通过引入注意力机制与上下文信息进行融合,将第19层输出层作为模型的小目标检测头。图9展示了改进后的检测层结构。

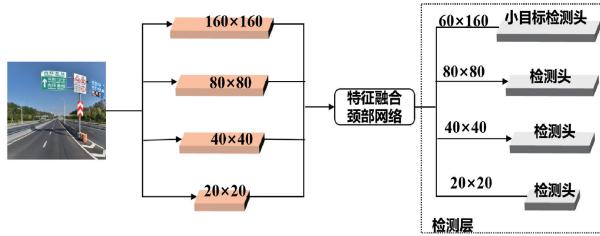


图9 改进后检测层图

Fig. 9 Improved detection layer diagram

2.6 损失函数改进

目标检测的性能评估与损失函数的设计紧密相关,YOLOv8采用了CIoU(complete-iou,CIoU)损失函数。假设 w,h 分别表示预测框的宽度和高度, w^{gt} 和 h^{gt} 为真实框的宽度和高度, b,b^{gt} 分别代表预测框和真实框的中心点位置, c 表示预测框和真实框外接框的对角线长度, p 表示中心点之间的欧氏距离, IoU (交并比)表示预测框与真实框的重叠区域比例。CIoU损失函数由以下三部分组成:

1) IoU 部分:表示预测框和真实框的重叠区域, IoU 越大,表示两个框的重叠程度越高。

$$IoU = \frac{\text{area of overlap}}{\text{area of union}} \quad (2)$$

2) 中心点距离: CIoU引入了预测框和真实框的中心点之间的欧氏距离,用于度量框位置的偏差。

$$d = \frac{p^2(b, b^{gt})}{c^2} \quad (3)$$

3) 宽高比一致性:表示预测框和真实框的宽高比差异度量。

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{gt}}{h^{gt}} - \arctan \frac{w}{h} \right)^2 \quad (4)$$

CIoU损失函数的完整公式为:

$$a = \frac{v}{(1 - IoU) + v} \quad (5)$$

$$L_{\text{CIoU}} = 1 - IoU + \frac{p^2(b, b^{gt})}{c^2} + a \times v \quad (6)$$

式中: a 为权重参数,用于平衡边界框的大小和边界框中心点的位置。

CIoU损失函数通过引入中心点距离的惩罚项,使得预测框更准确地趋向真是框的中心位置,有效减少位置误差。虽然CIoU在宽高比和位置方面表现良好,但它的重点在预测框与真实框的重叠区域,忽略了某些物体的边缘特征,在处理目标物体边缘时,CIoU可能存在误差,Shape-IoU是一种边框回归损失方法,能够有效减少预测框和真实框在边缘区域的偏差和模糊。文献[30]在模型中通过引入Inner-ShapeIoU损失函数,使得边界框回归更加准确,增强检测效果,有效解决了现有安全帽和反光衣检测模型对小目标和复杂天气中目标检测精度低等问题。因此,对于交通标志这种形状各异、轮廓不清晰的数据集检测来说,Shape-IoU更加适合这种复杂的检测场景,Shape-IoU公式如下:

$$ww = \frac{2 \times (w^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (7)$$

$$hh = \frac{2 \times (h^{gt})^{scale}}{(w^{gt})^{scale} + (h^{gt})^{scale}} \quad (8)$$

$$distance^{shape} = hh \times (x_c - x_c^{gt})^2 / c^2 + ww \times (y_c - y_c^{gt})^2 / c^2 \quad (9)$$

$$\Omega^{shape} = \sum_{t=w,h} (1 - e^{-w_t})^\theta, \theta = 4 \quad (10)$$

$$\begin{cases} w_w = hh \times \frac{|w - w^{gt}|}{\max(w, w^{gt})} \\ w_h = ww \times \frac{|h - h^{gt}|}{\max(h, h^{gt})} \end{cases} \quad (11)$$

式中: $scale$ 是缩放因子,与数据集中目标大小有关, ww 和 hh 分别表示水平方向和垂直方向上的权重系数。边界框回归损失如下:

$$L_{\text{Shape-IoU}} = 1 - IoU + distance^{shape} + 0.5 \times \Omega^{shape} \quad (12)$$

2.7 融入DKN-Attention注意力机制模块

本文设计了一种新的注意力机制,称为DKN-Attention,它结合了可变形大卷积核注意力(deformable large kernel attention, D-LKA)和DCNv3可变形卷积。在D-LKA中,大卷积核与可变形卷积的结合,进一步增强了模型对复杂图像模式的适应能力,D-LKA模块的结构如图10所示,该模块由多个卷积层组成,主要包括以下部分:1) 标准2D卷积(Conv2D);2) 带偏移量的可变形深度卷积(Deformable-DW Conv2D),允许网络根据输入特征自适应地调整感受野;3) 偏移场(Offsets Field),用于指导可变形卷积层如何动态调整采样位置;4) 激活函数

GELU, 增强模型的非线性表达能力。

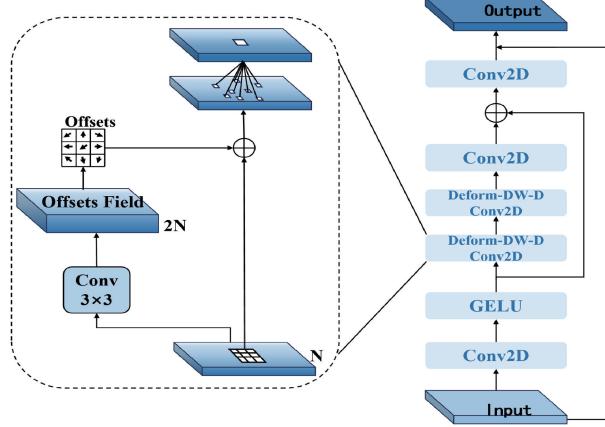


图 10 D-LKA 模块结构图

Fig. 10 Structure diagram of the D-LKA module

DCNv3 是一种应用于图像目标检测的可变形卷积模块。DCNv3 能够适应各种几何变换, 例如尺度变换、长宽比的变化、旋转等, 从而增强模型对目标形变的建模能力, 这种特性尤其适合处理背景复杂、形状多样的小目标交通标志场景的检测。DCNv3 模块的三维原理如图 11 所示, 其卷积核大小为 3×3 , 将输入特征的 3×3 区域映射为输出特征中的 1×1 。在传统卷积中, 3×3 区域的选取是规则的形状, 而可变形卷积则通过加入偏移量, 动态调整每个点的选取位置。

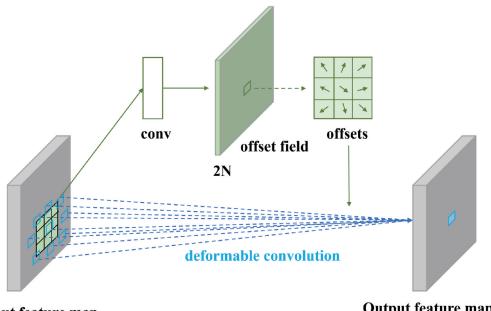


图 11 DCNv3 结构三维原理图

Fig. 11 Three-dimensional schematic diagram of DCNv3 structure

在 Neck 部分的 4 个输出层模块之后添加 DKN-Attention 模块, 利用大卷积核捕捉图像中丰富的上下文信息, 并通过可变形卷积实现灵活的特征图采样。通过在上采样和下采样过程中有效定位小物体场景中的关键注意力区域, DKN-Attention 注意力机制显著提升了对远距离微小交通标志的特征提取与识别能力。DKN-Attention 注意力机制结构如图 12 所示。

3 实验与结果分析

3.1 实验环境及参数配置

实验在 ubuntu22.04 操作系统上进行, CPU 为 AMD

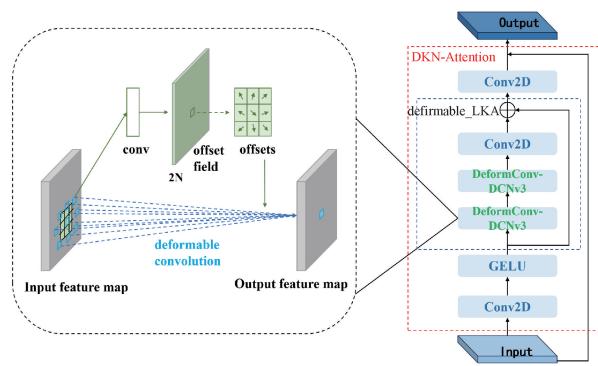


图 12 DKN-Attention 注意力机制结构图

Fig. 12 Structure diagram of DKN-Attention attention mechanism

EPYC 9754, GPU 为 NVIDIA GeForce RTX 4090D 24 GB 显存, 深度学习框架为 PyTorch 2.3.0, Cuda 版本为 12.1, 采用 Python 3.12 编程语言。实验中的 epochs 设置为 200, batchsize 设置为 24, 图像输入端分辨率设置为 640×640 , 使用 SGD 优化器, 初始学习率为 0.01, 其他参数使用 YOLO 系列的默认参数, 为了避免模型过分依赖预训练权重, 本文所有实验均未使用额外预训练, 后续消融实验和对比实验中的参数设置与实验环境与上述保持一致。

3.2 数据预处理

文中用的数据集为腾讯与清华大学联合开发并公开的交通标志数据集 TT100K。该数据集包含近 1 万张来自各种自然场景的交通标志图像, 覆盖 151 类不同的交通标志, 与其他交通标志数据集相比, TT100K 的特点在于类别更多、目标更小, 且识别难度更大。由于数据集中各类别的标志分布不均衡, 会影响模型的收敛速度, 导致模型在学习某些类别特征时表现不足, 影响泛化能力。针对这一问题, 本文对数据集进行了以下处理步骤: 1) 筛选出交通标志实例数量超过 100 个的类别, 共 45 类, 包含 9 738 张图片; 2) 将筛选后的图片划分为训练集、验证集和测试集, 其中 6 793 张用于训练, 1 949 张用于验证, 996 张用于测试。TT100K 交通标志数据集主要包括 3 类交通标志: 警告标志、提示标志和禁止标志。以“w”开头的为警告标志, 如“w13”表示十字交叉路口; 以“i”开头的为提示标志, 如“i3”表示环岛行驶, “i12”表示向左转弯; 以“p”开头的为禁止标志, 这类标志在 TT100K 数据集中最为常见, 如“p13”表示禁止某两种车驶入, “pm10”表示限重 10 吨, “pg”表示减速让行, “pl100”表示限速 100。训练过程中使用了 Mosaic 数据增强方法, 以增加训练样本的多样性, 提升模型的泛化能力。

3.3 评价指标

实验评价指标主要包括查准率 Precision(P)、查全率 Recall(R)、平均精度 (mAP)、参数量、模型大小以及推理速度 FPS。定义如下:

$$Precision = \frac{TP}{TP + FP} \quad (13)$$

$$Recall = \frac{TP}{FN + TP} \quad (14)$$

$$AP = \int_0^1 P(R) dR \quad (15)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP(i) \quad (16)$$

式中: TP 表示正确预测为正样本的数量, FP 表示错误预测为正样本的负样本数量, FN 表示错误预测为负样本的正样本数量。 mAP 值是所有类别的平均精度, n 为类别总数, i 表示第 n 个类别, AP 值表示某一类别的平均精度, 定义为在不同召回率下的平均精度。

每秒帧数(frames per second, FPS)是衡量模型推理速度的一个重要指标, 计算公式为:

$$FPS = \frac{1}{P + I + N} \times 100 \quad (17)$$

式中: P 为图像预处理时间; I 为图像输入模型到模型输出结果的时间; N 代表后处理时间。

图 13 是 TT100K 数据集样本各类别所含交通标志数量、预测框以及对数据集中目标框位置和大小分布进行可视化分析, 分别展示了预测框的大小、目标框中心点在图像中的坐标分布, 以及目标框宽高占图像的比例。统计结果表明, 大多数交通标志位于图像的中间偏右侧, 且目标

框所占的图像比例大多不足 0.05, 并且交通标志的预测框都相对较小, 进一步给交通标志检测带来了困难。

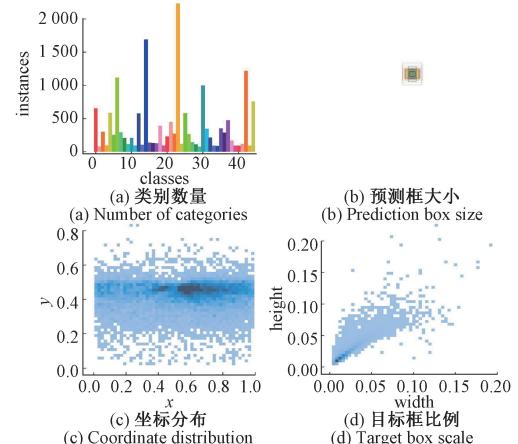


图 13 TT100K 数据集目标框分布情况

Fig. 13 Distribution of target boxes in the TT100K dataset

3.4 实验结果分析

1) 消融实验

为了验证本文对 YOLOv8s 的改进各模块对于交通标志目标识别精度提升的有效性, 本文共设计了 7 组环境一致、参数一致的消融实验, 实验结果如表 1 所示。

表 1 消融实验

Table 1 Ablation experiment

编号	YOLOv8s	C2f-Faster	Detect_SR	DySample	small head	Shape-IoU	DKN-Attention	P/%	R/%	mAP50/%	参数/M	模型大小/M	FPS
1	√							80.1	75.5	82.1	11.1	21.5	104.5
2	√	√						82.3	70.5	80.5	8.3	16.1	122.9
3	√	√	√					85.3	75.3	84.6	8.6	16.9	82.8
4	√	√	√	√				85.6	75.7	85.1	7.7	16.4	91.4
5	√	√	√	√	√			87.4	76.9	86.4	8.4	17.2	64.1
6	√	√	√	√	√	√		85.8	79.2	87.7	8.6	16.9	64.7
7	√	√	√	√	√	√	√	86	79.7	88.4	11.3	22.1	54.1

编号 2 中, 将轻量化模块 C2f-Faster 融入 YOLOv8s 网络结构后, 模型计算复杂度大幅降低, 导致查准率(P)、查全率(R)和平均精度(mAP)有所下降, 但参数量显著减少, 推理速度(FPS)大幅提升; 编号 3 用 Detect_SR 检测头替换了原有检测头, 使得查准率、查全率和平均精度显著提升, 弥补了编号 2 的不足; 编号 4 通过融入 DySample, 使各项评价指标均有小幅提升, 以及参数量和模型大小均有所减少; 编号 5 在编号 4 的基础上增加了一个小目标检测头, 增强了模型对图像全局信息的理解, 虽然该检测头增加了参数量, 导致推理速度有所下降, 但查准率、查全率和平均精度得到进一步提升; 编号 6 更新损失函数, 由查全率可见 Shape-IoU 损失函数能够有效减少预测框和真实框

在边缘区域的偏差和模糊, 在目标定位方面, 有显著的性能提升, 但同时也导致了查准率有所下降; 编号 7 通过在 Neck 部分中融入 DKN-Attention 注意力机制, 使模型能够捕捉到丰富的上下文信息, P、R、Map50 这 3 个指标在编号 6 的基础上分别提高了 0.2%、0.5% 和 0.7%。与原始的 YOLOv8s 模型相比, FKDS-YOLOv8s 模型在查准率、查全率和 mAP50 这 3 个精度指标上分别提高了 5.9%、4.2% 和 6.3%, 充分证明了本文所提出的 FKDS-YOLOv8s 模型在提高小目标识别精度上具有显著效果; 在提高各项精度评价指标的同时, 参数量和模型大小均有小幅度上升, 导致推理速度 FPS 下降, 但是相对于常规高清视频处理所需要的 25~30 fps 的标准, 本文方法 54.1 fps

的值仍满足该应用标准。

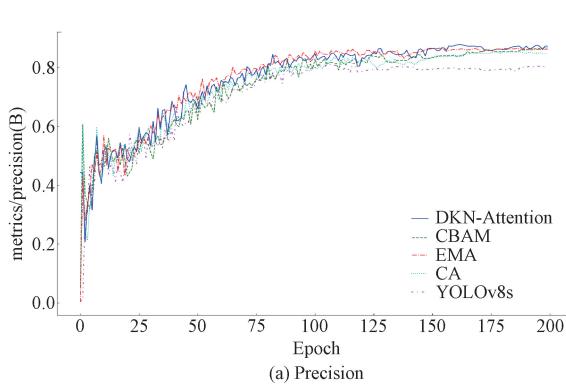
为了进一步验证本文设计的注意力机制模块对检测小目标的有效性,在上述的实验中分别用经典注意力机制(CA、CBAM、EMA)替换本文模型相同位置所使用的注意力机制模块,模块进行横向比较,实验结果如表 2 和图 14 所示。由图 14 可知,随着迭代次数的增加,模型在训练到 200 轮时均已达到收敛状态,并且模型增加了 DKN-

Attention 注意力机制后,在 precision、map 等关键精度中均处于最优状态,改进模型的表现优于基准模型。由表 2 可知,相较于 CA^[31]、CBAM^[32]、EMA^[33] 等经典注意力机制,DKN-Attention 模块虽然增加了参数量,但在 mAP50/% 与 mAP50-95/% 上分别提升了 1.6%、1.5%、0.7% 与 1.6%、1.6%、0.6%,且 F1 分数最优,这说明本文设计的 DKN-Attention 注意力机制模块取得了最好的效果。

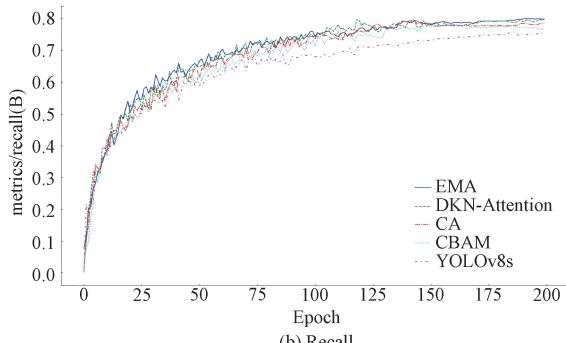
表 2 注意力机制对比实验

Table 2 Comparative experiments on attention mechanisms

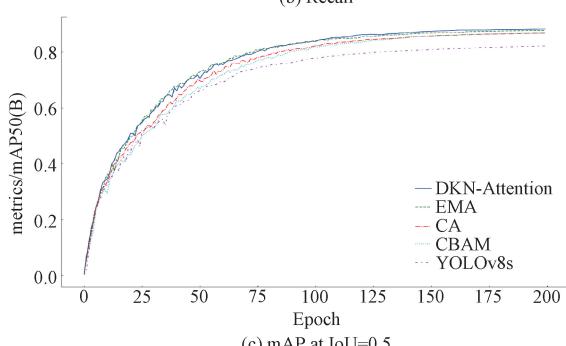
注意力机制模块	P/%	R/%	mAP50/%	mAP50-95/%	参数/M	F1/%
CA	84.7	78.5	86.8	67.9	8.7	81.2
CBAM	86.5	76.6	86.9	67.9	12.1	81.5
EMA	8680.0	80.0	87.7	68.9	8.6	81.8
DKN-Attention	8680.0	79.7	88.4	69.5	11.3	82.6



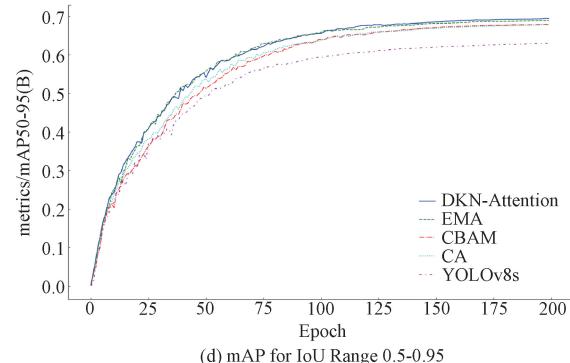
(a) Precision



(b) Recall



(c) mAP at IoU=0.5



(d) mAP for IoU Range 0.5-0.95

图 14 注意力机制可视化对比图

Fig. 14 Visualization of attention mechanism

2) 对比实验

为验证本文算法对交通标志识别和检测的有效性,选取几种常用检测算法以及当前优秀检测算法(CR-YOLOv8, DMC-YOLO, YOLOv5s-BNC)进行对比实验,实验参数与本文模型参数保持一致,差异参数则保持默认设置,对比实验结果的结果如表 3 所示。与先进的二阶段目标检测算法 Faster-RCNN 相比,本文模型在各项指标上均表现更优,且 Faster-RCNN 的计算量庞大,不适合部署在车载设备上。而与单阶段目标检测算法 YOLO 系列相比,本文模型在查准率(P)、查全率(R)和 mAP50/% 等关键精度指标上都取得了最佳表现,这与本文重点提升模型检测精度目标相一致,尽管本文模型在参数量和推理速度(FPS)上未占据明显优势,但仍保持了相对较高的水平。

3) 可视化实验

本文对多种交通标志图像进行了可视化对比实验,涵盖了远距离小目标、单一目标、多目标以及遮挡目标的情况,并将其识别效果与原始 YOLOv8s 算法进行了比较。图 15 为模型识别效果对比图,第 1 排展示了原始 YOLOv8s

表3 检测算法性能对比

Table 3 Performance comparison of detection algorithms

模型	P/%	R/%	mAP50/%	参数/M	模型大小/M	FPS
Faster-RCNN	67.1	70.5	70.3	17.1	108.9	23.0
YOLOv5s	70.4	60.9	66.8	7.1	14.0	112.2
YOLOv5m	77.7	71.4	77.6	21.0	40.6	89.5
YOLOv7-tiny	50.2	48.9	49	6.1	11.9	78.1
YOLOv8n	71.1	64.4	70.5	3.0	6.0	131.0
YOLOv8s	80.1	75.5	82.1	11.1	21.5	104.5
YOLOv10n	65.2	60.4	65	2.7	5.5	160.7
YOLOv10s	77.2	69.2	77.8	8.1	15.8	154.2
CR-YOLOv8 ^[19]	—	—	86.9	14.6	—	103.0
DMC-YOLO ^[34]	—	—	84.8	5.2	—	30.58
YOLOv5s-BNC ^[35]	81.3	78.5	67.5	—	17.1	74.0
本文	86.0	79.7	88.4	11.3	22.1	54.1



图15 识别效果对比

Fig. 15 Comparison of recognition effects

算法的识别结果,第2排则展示了本文算法的识别效果。原始YOLOv8s算法在识别过程中存在漏检(如第1排第1张漏检禁止通行pne类)和误检(如第1排第4张将“限制质量30t”pm30类误判为“限速80”pl80类,将“限速60”pl60类误判为“限速80”pl80类,第5张将“禁止掉头”p5类误判为“禁止机动车驶入”p10类),且检测精度较低的情况。相比之下,FKDS-YOLOv8s算法不仅能够准确识别pne、pm30、pl60与p5等交通标志,并且在置信度方

面表现优异,对多目标和遮挡目标也能有效识别。

如图16所示,通过使用热力图对比了YOLOv8s与改进后模型的特征图可视化图像。从图中可以看出YOLOv8s的特征图仅关注了少量的交通标志,而改进后模型的特征图中,大多数目标交通标志的中心都受到了广泛的关注,表明改进后的模型可以更加全面的关注不同尺度下交通标志的特征和位置信息,增强交通标志的检测的效果。

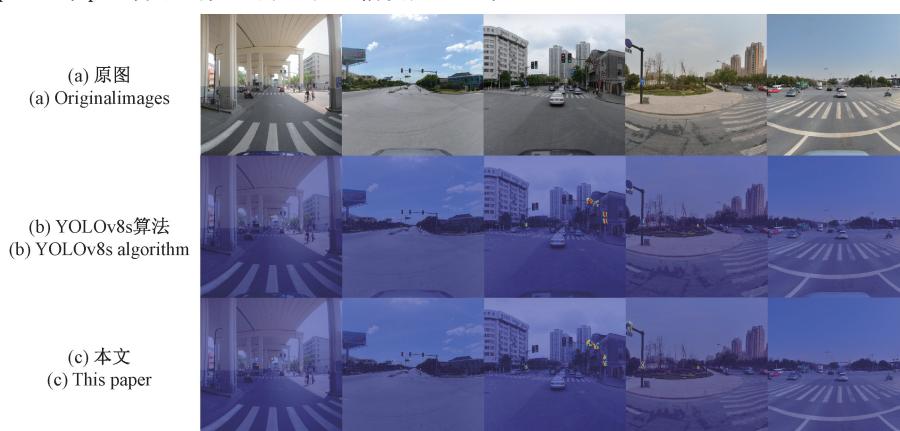


图16 YOLOv8s与KFDS-YOLOv8s的热力图对比

Fig. 16 Comparison of heat maps of YOLOv8s and KFDS-YOLOv8s

4 结 论

本文提出了一种改进的 YOLOv8s 城市交通标志目标检测算法, 旨在解决传统目标检测算法存在的误检、漏检以及精度不高的问题。本文提出了以下改进措施: 首先采用改进的 C2f-Faster 模块, 融合超轻量的动态上采样器和引入 Shape-IoU 损失函数, 在实现网络轻量化的同时, 显著提升了模型对小目标的精准捕捉能力; 其次结合 DKN-Attention 注意力机制, 并引入 160×160 尺寸的检测头, 通过缩小预测误差, 使模型在密集城市交通中对被遮挡的小目标更加敏感, 从而有效解决漏检和精度不高的问题; 此外, 新的检测头能够更好地完成回归任务和分类任务。实验结果显示, 与 YOLOv8s 基准模型相比, 改进后的模型在 TT100K 数据集上的 P、R 和 mAP50 分别提升了 5.9%、4.2% 和 6.3%, 验证了其良好的泛化性和鲁棒性, 充分表明该改进模型具备在实际场景中应用的潜力。

参考文献

- [1] CREUSEN I M, WIJNHOVEN R G J, HERBSCHLEB E, et al. Color exploitation in hog-based traffic sign detection [C]. IEEE International Conference on Image Processing, 2010: 2669-2672.
- [2] 杨斐, 秦勃, 李伟. 基于多帧视频图像的交通标志实时检测 [J]. 计算机科学与应用, 2017, 7(5): 10.
- YANG F, QING B, LI W. Real-time traffic sign detection based on multi-frame video images [J]. Computer Science and Applications, 2017, 7(5): 10.
- [3] GIRSHICK R. Fast R-CNN [C]. International Conference on Computer Vision (ICCV), 2015.
- [4] HE K M, ZHANG X Y, REN SH Q, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition [J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2014, 37 (9): 1904-1916.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision, Spring, Cham, 2016.
- [6] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection [J]. ArXiv preprint arXiv: 2004.10934, 2020.
- [7] GLENN J, ALEX S, JIRKA B, et al. Ultralytics/yolov5 [EB/OL]. (2022-06-18) [2022-06-22]. <http://github.com/ultralytics/yolov5>.
- [8] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- [9] TERVEN J, ESPARZA D M, ROMERO-GONZALEZ J. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [10] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection [C]. Computer Vision & Pattern Recognition, 2016: 779-788.
- [11] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger [C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.
- [12] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.
- [13] YU J, YE X J, TU Q. Traffic sign detection and recognition in multiimages using a fusion model with yolo and vgg network [J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23 (9): 16632-16642.
- [14] 陈朋弟, 黄亮, 夏炎, 等. 基于 Mask R-CNN 的无人机影像路面交通标志检测与识别 [J]. 国土资源遥感, 2020, 32(4): 61-67.
- CHEN P D, HUANG L, XIA Y, et al. Detection and recognition of road traffic signs in UAV image based on mask R-CNN [J]. Remote Sensing for Land and Resources, 2020, 32(4): 61-67.
- [15] 侯卓成, 欧阳华, 胡鑫, 等. 基于改进的 YOLOv4 彩色数字仪表读数识别方法 [J]. 电子测量技术, 2022, 45(6): 124-129.
- HOU ZH CH, OUYANG H, HU X, et al. Color digital instrument reading recognition method based on improved YOLOv4 [J]. Electronic Measurement Technology, 2022, 45(6): 124-129.
- [16] 李志刚, 张娜. 一种轻量型 YOLOv5 交通标志识别方法 [J]. 电讯技术, 2022, 62(9): 1201-1206.
- LI ZH G, ZHANG N. A lightweight YOLOv5 traffic sign identification method [J]. Telecommunications Technology, 2022, 62(9): 1201-1206.
- [17] 赵会鹏, 曹景胜, 潘迪敬, 等. 改进 YOLOv8 算法的交通标志小目标检测 [J]. 现代电子技术, 2024, 47(20): 141-147.
- ZHAO H P, CAO J SH, PAN D J, et al. Improved YOLOv8 algorithm for small target detection of traffic signs [J]. Modern Electronic Technology, 2024, 47(20): 141-147.
- [18] ZHU ZH, LIANG D, ZHANG S H, et al. Traffic-sign

- detection and classification in the wild[C]. 2016 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016:2110-2118.
- [19] ZHANG L J, FANG J J, LIU Y X, et al. CR-YOLOv8: Multiscale object detection in traffic sign images[J]. IEEE Access, 2024, 12:219-228.
- [20] XIE S N, GIRSHICK R, DOLLAR P, et al. Aggregated residual transformations for deep neural networks[C]. 2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016:5987-5995.
- [21] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks [C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [22] LIU W Z, LU H, FU H T, et al. Learning to upsample by learning to sample [C]. IEEE/CVF International Conference on Computer Vision, 2023: 6004-6014.
- [23] TAN M X, LE Q V. Efficientnet: Rethinking model scaling for convolutional neural networks[J]. ArXiv preprint arXiv: 1905.11946, 2019.
- [24] HOWARD A, SANDLER M, CHEN B, et al. Searching for mobilenetv3 [C]. 2019 IEEE/CVF International Conference on Computer Vision(ICCV), 2019:1314-1324.
- [25] ZHANG X Y, ZHOU X Y, LIN M X, et al. Shufflenet: An extremely efficient convolutional neural network for mobile devices[C]. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2017: 6848-6856.
- [26] HAN K, WANG Y H, TIAN Q, et al. GhostNet: More features from cheap operations[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2020: 1577-1586.
- [27] 伍济钢,成远,邵俊,等.基于改进 YOLOv4 算法的 PCB 缺陷检测研究[J].仪器仪表学报, 2021(10): 170-177.
- WU J G, CHENG Y, SHAO J, et al. PCB defect detection based on improved YOLOv4 algorithm[J]. Chinese Journal of Scientific Instrument, 2021 (10): 170-177.
- [28] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2023:12021-12031.
- [29] 杨辉羽,李海明.单聚合 YOLO 航拍小目标检测算法[J].国外电子测量技术,2023,42(4):131-140.
- YANG H Y, LI H M. Small target detection algorithm for single aggregation YOLO aerial photography [J]. Foreign Electronic Measurement Technology, 2023, 42(4):131-140.
- [30] 许迪,张淑卿,葛超.面向复杂环境的 YOLOv8 安全装备检测[J].电子测量技术,2024,47(7):121-129.
- XU D, ZHANG SH Q, GE CH. YOLOv8 safety equipment detection for complex environment [J]. Electronic Measurement Technology, 2024, 47(7):121-129.
- [31] HOU Q B, ZHOU D Q, FENG J SH. Coordinate attention for efficient mobile network design[C]. 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2021:13708-13717.
- [32] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. Computer Vision-ECCV 2018, Springer, Cham, 2018:3-19.
- [33] OUYANG D L, HE S, ZHANG G ZH, et al. Efficient multi-scale attention module with cross-spatial learning[J]. ArXiv preprint arXiv: 2305.13563, 2023.
- [34] 栾孟娜,郑秋梅,王风华.基于 DMC-YOLO 的交通标志实时检测算法[J/OL].计算机工程, 1-10 [2024-11-27]. <https://doi.org/10.19678/j.issn.1000-3428.0069409>.
- LUAN M N, ZHENG Q M, WANG F H. Real-time traffic sign detection algorithm based on DMC-YOLO [J/OL]. Computer Engineering, 1-10 [2024-11-27]. <https://doi.org/10.19678/j.issn.1000-3428.0069409>.
- [35] 郑娟毅,郭梦月,杨溥江,等.改进 YOLOv5s 的交通标志检测算法[J].中国电子科学研究院学报, 2023, 18(11):1029-1041.
- ZHENG J Y, GUO M Y, YANG P J, et al. Improved traffic sign detection algorithm for YOLOv5s [J]. Journal of China Academy of Electronic Science, 2023, 18(11):1029-1041.

作者简介

王浩,教授,硕士生导师,博士,主要研究方向为智能交通信息与控制、计算机应用。

张其猛(通信作者),硕士研究生,主要研究方向为深度学习、计算机视觉。

E-mail: zhang_qimeng1998@163. com

龚德成,硕士研究生,主要研究方向为深度学习。