

改进YOLOv8的道路交通标志目标检测算法

田 鹏, 毛 力

江南大学 人工智能与计算机学院, 江苏 无锡 214122

摘 要:虽然,当前检测技术日趋成熟,但对于复杂环境下的小目标检测仍然是研究的重难点。针对道路交通场景中普遍存在的交通标志小目标比例较高,环境干扰因素较大的问题,提出了一种基于YOLOv8改进的道路交通标志目标检测算法。由于小目标检测中容易出现漏检的现象,利用BRA(bi-level routing attention)注意力机制提高网络对小目标的感知能力。此外,还利用可形变卷积模块DCNv3(deformable convolution v3),针对特征图中的不规则形状具有更好的特征提取能力,使骨干网络更好地适应不规则的空间结构,更精准地关注重要目标,从而提高模型对遮挡重叠目标的检测能力。DCNv3和BRA模块均在基本不增加模型权重大小的情况下提高模型准确性。同时引入基于辅助边框的Inner-IOU损失函数。在RoadSign、CCTSDb、TSDD、GTSDB四个数据集上,分别进行了小样本训练、大样本训练、单目标检测和多目标检测,实验结果均有所提高。其中,在RoadSign数据集上的实验结果最佳,YOLOv8改进模型的均值平均精度mAP50与mAP50:95分别达到了90.7%和75.1%,相较于基线模型,mAP50与mAP50:95分别提升了5.9和4.8个百分点。实验结果表明,YOLOv8改进模型有效地实现了在复杂道路场景下的交通标志检测。

关键词:YOLOv8;小目标检测;可形变卷积;注意力机制;复杂道路场景

文献标志码:A **中图分类号:**TP391.4 **doi:**10.3778/j.issn.1002-8331.2309-0415

Improved YOLOv8 Object Detection Algorithm for Traffic Sign Target

TIAN Peng, MAO Li

School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi, Jiangsu 214122, China

Abstract: Although the current testing technology is becoming increasingly mature, the detection of small targets in complex environments is still the most difficult point in research. Aiming at the problem of high target proportion of traffic signs in road traffic scenarios, the problem of high target proportion of small targets and large environmental interference factors, it proposes a type of road traffic logo target test algorithm based on YOLOv8 improvement. Due to the prone to missed inspection in small target testing, the bi-level routing attention (BRA) attention mechanism is used to improve the network's perception of small targets. In addition, it also uses a shape-changing convolutional module deformable convolution V3 (DCNV3). It has a better feature extraction ability for irregular shapes in the feature map, so that the backbone network can better adapt to irregular space structures, and pay more accurately to important attention, objectives, thereby improving the detection ability of the model to block the overlapping target. Both DCNV3 and BRA modules improve the accuracy of the model without increasing the weight of the model. At the same time, the Inner-IOU loss function based on auxiliary border is introduced. On the four data sets of RoadSign, CCTSDb, TSDD, and GTSDB, small sample training, large sample training, single target detection, and multi-target detection are performed. The experimental results are improved. Among them, the experiments on the RoadSign data set are the best. The average accuracy of the improved YOLOv8 model mAP50 and mAP50:95 reaches 90.7% and 75.1%, respectively. Compared with the baseline model, mAP50 and mAP50:95 have increased by 5.9 and 4.8 percentage points, respectively. The experimental results show that the improved YOLOv8 model effectively implements the traffic symbol detection in complex road scenarios.

Key words: YOLOv8; small target detection; deformable convolution; attention mechanism; complex road scenes

道路交通标志目标检测算法是一种先进的计算机视觉技术,旨在自动识别道路上的交通标志并提供准确

的定位和分类^[1]。随着智能交通系统的快速发展,道路交通标志的准确检测对于实现智能驾驶、交通管理和驾

作者简介:田鹏(1999—),男,硕士研究生,主要研究方向为模式识别、人工智能等;毛力(1967—),通信作者,男,硕士,教授,硕士生导师,主要研究方向为人工智能、机器学习等,E-mail:wxmaoli@163.com。

收稿日期:2023-09-25 **修回日期:**2024-01-05 **文章编号:**1002-8331(2024)08-0202-11

驶员辅助等应用至关重要^[2]。

道路交通标志的目标检测一直是计算机视觉领域的研究热点,交通标志检测常见的方法有基于颜色的方法^[3]、基于形状的方法^[4]、基于多特征融合的方法^[5]和基于深度学习的方法。传统方法通常依赖于手工设计的特征提取器和分类器,面临着复杂背景、遮挡和尺度变化等问题。通过引入基于深度学习的目标检测技术,取得了突破性的进展。基于深度学习的检测算法主要分为双阶段和单阶段检测算法,双阶段算法通过提取候选框进行二次修正得到检测结果,代表算法有R-CNN^[6]、Fast R-CNN^[7]、Faster R-CNN^[8]等。单阶段算法可以一次性完成定位和分类,代表算法有SSD系列算法^[9-11]和YOLO系列算法^[12-14]。

YOLO^[15-16]算法采用了一种单阶段的目标检测方法,通过将整个图像划分为网格单元,并在每个单元中预测多个边界框和类别概率,实现对交通标志的快速检测。相较于传统方法,YOLO^[17]算法具有更高的检测速度和准确性。

该算法的核心思想是通过深度卷积神经网络(CNN)从原始图像中学习特征表示,然后使用预测器生成边界框和类别概率。YOLOv8采用了Darknet的深度神经网络结构,结合不同层次的特征提取,有效地捕捉到交通标志的形状、纹理和上下文信息。此外,YOLOv8算法还引入了一系列优化策略,如多尺度训练、数据增强和损失函数的优化,进一步提升了交通标志检测的性能。

相对于其他目标检测任务而言,小目标检测的历史较短。主要通过数据增强,尺度感知,特征融合,超分辨率等方法提高检测准确率,但大多并不适用于类似复杂道路场景这样目标变化较大的场景。

为解决上述问题,本文提出了一种改进的YOLOv8模型,融合可形变卷积^[18]模块Deformable convolution v3(DCNv3)和Bi-level routing attention(BRA)^[19]模块,以适应不规则状态,获取目标重要特征,减少漏检误检测,使得算法具备了对不同尺度、角度和光照条件下交通标志的鲁棒性。

基于YOLOv8改进的道路交通标志目标检测算法具有快速、准确和鲁棒的特点,为实现智能交通系统的高效运行提供了有力的支持。随着技术的不断发展和改进,相信该算法将在道路交通标志识别领域发挥越来越重要的作用,并为交通安全和智能出行提供更加可靠的保障。

1 相关工作

1.1 YOLO与注意力机制

针对Backbone,2021年Yan等人^[20]将SE Block模块嵌入至YOLOv5中,以突出特征图中的重要特征,抑制不相关特征,使mAP提升了1.44%,从而提高网络检

测的性能。

Li等人^[21]分别将SE Net(squeeze and-excitation networks)和CBAM(convolutional block attention module)引入YOLOv3的Backbone中,自主学习每个通道的权重,增强关键特征,抑制冗余特征。与YOLOv3相比,引入SE Net和CBAM使mAP分别提升了8.00%和8.50%。

针对Neck,有一部分工作引入了通道注意力。2021年Ma等人^[22]提出将特征选择模块(feature select module,FSM)引入Neck层,对每个通道进行评分,过滤掉不相关的特征,关注高贡献的特征,避免特征融合阶段引入过多的不相关特征。该模块再分别对YOLOv3、YOLOv4以及YOLOv5-L进行改进,提升效果分别为0.60%、1.10%和1.50%。

另有一部分工作同时将空间和通道注意力进入至Neck层。如,2020年Ju等人^[23]利用全局注意机制和空间位置注意机制,将基于注意力机制的自适应特征融合方法(adaptive feature fusion with attention mechanism,AFFAM)与YOLOv3结合,同时自适应的学习通道特征的相关性和空间特征在不同尺度上的重要性。该方法在PASCAL VOC07、KITTI和Smart UVM数据集上分别获得了5.08%、2.30%和7.41%的mAP提升。

总体而言,在Backbone中引入注意力机制,可使网络关注目标所在的通道和区域,提高特征提取的效度;在Neck层引入注意力机制,可选择性地、自适应地完成特征融合,改善特征融合的有效性。只是,无论在哪个阶段引入注意力机制,都会带来网络复杂度的增加,如何有效平衡代价及其性能增益,如果针对目标检测的特定任务构建更加针对性的注意力机制都还值得进一步研究。

1.2 YOLOv8模型

YOLOv8是最新的最先进的YOLO模型,它在以前成功的YOLO版本基础上,引入了新的功能和改进,进一步提升了其性能和灵活性,能够更好地适应各种目标检测任务。

整个网络分为两部分:特征提取网络和检测网络。特征提取网络负责从输入图像中提取高级语义特征,通常采用多个卷积层和池化层组成的结构。这些卷积层逐渐减小特征图的尺寸,同时增加通道的数量,以捕获不同尺度和抽象级别的特征。

检测网络是YOLOv8的核心部分,它通过在不同尺度的特征图上进行检测来实现目标检测。在检测网络中,通过引入多个边界框和类别概率的预测器,可以同时检测多个目标。每个预测器负责预测一组边界框的位置和对应的类别概率。

为了获得更好的检测结果,YOLOv8还使用了特征融合技术,将不同尺度的特征图进行融合,以提高对不

同尺度目标的检测能力。

在网络的 Head 部分, YOLOv8 选择解耦头(decoupled head)和 Anchor Free 策略, 相比 YOLOv5 采用耦合头(coupled head)和 Anchor Based 策略不再有之前的 Objectness 分支, 只有解耦的分类和回归分支。

在网络的 Backbone 部分, YOLOv5 中的 C3 模块被替换成了 C2f 模块, 实现了进一步的轻量化, 还能获得更加丰富的梯度流信息, 同时, YOLOv8 选择将上采样之前的 1×1 卷积去除, 将 Backbone 不同阶段输出的特征直接送入了上采样操作。此外, YOLOv8 依旧使用了 YOLOv5 等模型中使用的 SPPF 模块。

YOLOv8 抛弃了以往的 IOU 匹配或者单边比例的分配方式, 而是使用了 Task aligned assigner 匹配方式。YOLOv8 的分类损失为 VFL Loss, 回归损失为 CIOU Loss, 使得预测结果更接近真实。

1.3 可形变卷积

1.3.1 可形变卷积简介

最初, 深度学习中使用的是常规卷积操作, 它在图像中的每个位置应用相同的卷积核。然而, 常规卷积在处理具有形变或局部区域变化的特征时可能存在一定的局限性。为了处理更大感受野和更丰富的特征信息, 空洞卷积^[24](dilated convolution)被引入。空洞卷积在卷积核内部引入空洞或间隔, 使卷积核能够跳过一些像素点, 从而增加感受野的范围。空洞卷积可以提供更广阔的上下文信息, 但仍然受限于固定的采样模式。为了克服常规卷积和空洞卷积的限制, 可形变卷积被提出。可形变卷积通过在卷积过程中对卷积核进行形变, 使其能够适应不同位置的局部特征变化。形变卷积通过引入可学习的形变参数, 使卷积核的采样位置和形状能够根据特征图中的局部特征进行自适应调整, 能够更好地适应目标的非刚性形变、遮挡和背景复杂等情况。

从图 1 可以看出, 可形变卷积的采样位置是可变的, 或者说是可学习的, 因此可形变卷积可以更好地考虑到物体形状变化。

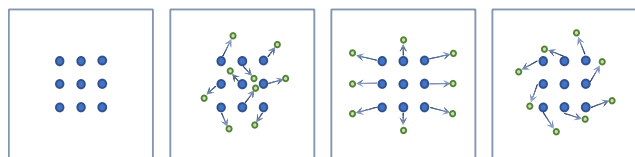


图1 可形变卷积像素点采样

Fig.1 Deformable convolutional pixel sampling

1.3.2 可形变卷积原理

可形变卷积的原理^[25]是基于一个网络学习 offset, 使得卷积核在 input feature map 的采样点发生偏移, 集中于本文感兴趣的区域或者目标。如图 2 所示。

标准卷积的操作过程: 对于输入的二维 feature map 上的每一个位置 P_0 , 进行以下卷积操作:

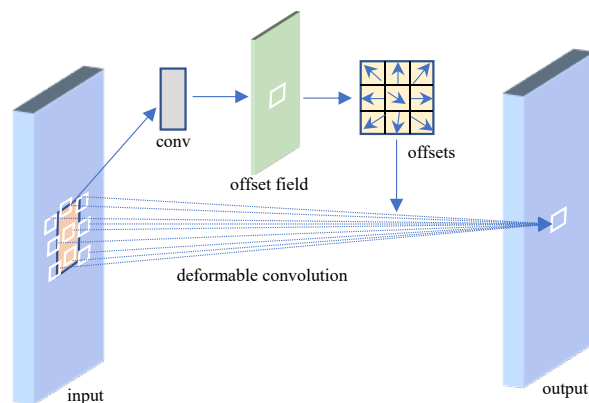


图2 卷积核偏移

Fig.2 Convolutional kernel offset

$$y(P_0) = \sum_{P_n \in R} \omega(P_n) \cdot x(P_0 + P_n) \quad (1)$$

其中, P_n 是卷积核的每一个位置, ω 是卷积核, $R = \{(-1, -1), (-1, 0), \dots, (0, 1), (1, 1)\}$ 。

可形变卷积的操作过程:

$$y(P_0) = \sum_{P_n \in R} \omega(P_n) \cdot x(P_0 + P_n + \Delta P_n) \quad (2)$$

其中, $\{\Delta P_n | n = 1, 2, \dots, N\}$, $N = |R|$, 对应着图中 offsets 的每一个位置。

在可形变卷积的操作中, 在原来的偏移量 R 的基础上又加入了一个二维偏移 ΔP_n (x 、 y 轴上的偏移), 这个 ΔP_n 的值对应图中 offsets 对应位置的值。由于 offsets 是一个浮点值, 因此对应的不是特征图上一个真实的位置, 因此该位置的值是通过计算周围四个真实值的双线性差值得到的。双线性差值通过下述公式得到:

$$x(p) = \sum_q G(q, p) \cdot x(q) \quad (3)$$

$$G(q, p) = g(q_x, p_x) \cdot g(q_y, p_y) \quad (4)$$

其中, $g(a, b) = \max(0, 1 - |a - b|)$ 。

在原始特征图上使用了一个 3×3 , $\text{padding} = 3$ 的卷积, 得到一个长宽与原特征图相同的, 通道数 $2N$ 的特征图。该图上的每一个特征点, 就有 $2N$ 个值: 2 对应 x 、 y 轴的偏移量, N 对应 N 个 ΔP_n 的值。

对每一个 P_0 , P_n 有 N 个值, 对应着卷积核的大小, ΔP_n 同样也有 N 个值, 对应上图中 offset field 特征图的 N 个通道, 对于输出的特征图上的每个点, 可以单独决定在原图上采样的 3×3 的特征点的空间位置。

通过以上的可形变卷积方式自适应提取特征, 采样位置更符合目标本身的形状和尺度, 更利于密集遮挡目标及小目标的特征提取。

2 改进 YOLOv8 算法

针对深度学习网络的改进和使用对于众多复杂实际场的检测任务都具有现实意义^[26-28]。本文希望在 YOLOv8 的基础上进行相关改进, 实现对于复杂场景下

交通标志小目标的高效检测。

在使用YOLOv8模型完成复杂场景下的道路交通标志目标检测任务时,发现针对小目标检测,YOLOv8存在漏检现象,且针对重叠、遮挡目标检测效果较差。本文针对检测中存在的不足进行讨论,并提出解决方案。

在YOLOv8的Backbone部分,新增加一个可形变卷积Deformable convolution v3(DCNv3)模块,便于卷积在训练过程中扩展到更大的模块,获取更多,更全面的特征信息。将YOLOv8Head中的C2F模块替换为在分割领域使用的Bi-level routing attention(BRA)模块,引导网络更好地利用不同尺度的特征信息,使得网络能够更好地理解图像的上下文关系和细节信息。通过全局和局部的特征交互,Bi-level routing attention能够提高目标检测的准确性和鲁棒性,尤其在处理具有尺度变化、遮挡和复杂背景的图片场景时具有优势。在损失函数阶段,引入Inner-IOU代替CIOU损失函数,提高了模型的泛化能力和检测精度。改进后的网络结构图如图3所示。

2.1 Bi-level Routing Attention(BRA)

为了缓解多头自注意力(multi-head self-attention, MHSA)的可扩展性问题,先前的一些小目标检测方法采用了不同的稀疏注意力机制,其中每个查询只关注少量的键值对,而非全部。然而,这些方法有两个共性问题:

要么使用手工制作的静态模式,无法自适应;要么在所有查询中共享键值对的采样子集,无法做到互不干扰。

因此,BRA注意力机制对于小目标检测的最大功效就在于其是一种动态稀疏注意力机制,其关键思想是引入了两个层级的注意力机制来建立任务之间的关联和信息交互。在粗糙区域级别过滤掉大部分不相关的信息交互,以便只保留一小部分路由区域,减少了无效任务之间产生的干扰,增强了有效信息之间的交互。其次,在这些路由区域的联合中应用细粒度的token-to-token注意力,通过相关联任务之间的深度交互,获取更多的有效特征信息。与传统注意力机制相比,能够更加灵活地针对输入图像的内容来动态调整注意力的分布,从而对于不同尺度和复杂度的目标有更好的适应效果,因此能够更加精准地捕捉小目标的特征。

如图4所示,首先计算任务内注意力,用于在任务内部对特征进行加权。帮助模型选择在每个任务中最重要的特征。具体实现过程如下:

对于一个给定的2D特征图 $X \in \mathbb{R}^{H \times W \times C}$,将其划分为 $S \times S$ 个不重叠的区域,使得每个区域内包含 $\frac{HW}{S^2}$ 个

特征向量。确定 $Q, K, V \in \mathbb{R}^{S^2 \times \frac{HW}{S^2} \times C}$, 对应的线性预测如下:

$$Q = X^r W^q, K = X^r W^k, V = X^r W^v \quad (5)$$

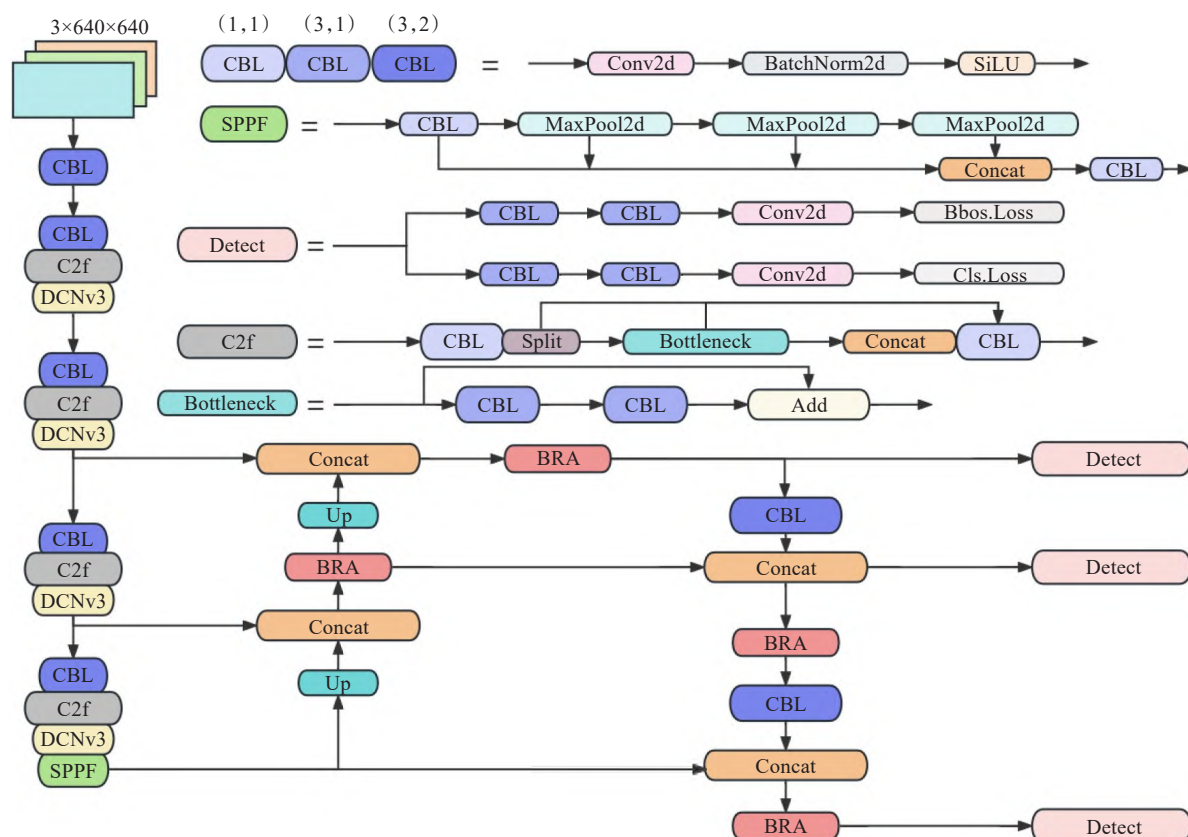


图3 网络模型结构

Fig.3 Network model structure

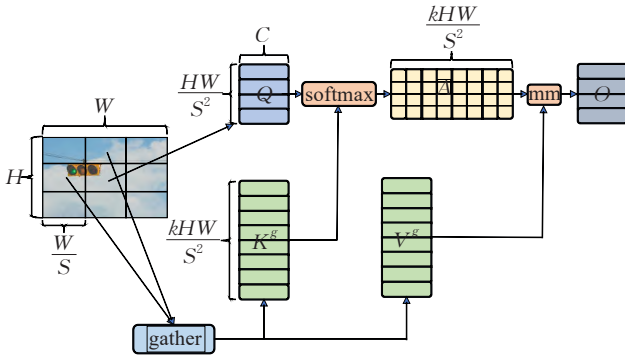


图4 BRA注意力机制作用过程

Fig.4 Action process of BRA attention mechanism

其中, $W^q, W^k, W^v \in \mathbb{R}^{C \times C}$, 分别是 Q, K, V 的投影权重。

然后, 通过构造一个有向图确定区域间的参与关系, 首先通过对 Q 和 K 取平均值, 推导出 $Q', K' \in \mathbb{R}^{S^2 \times C}$, 再通过以下公式得到邻接矩阵 A^r :

$$A^r = Q'(K')^T \quad (6)$$

代表着输入信息到内容区域之间的关联程度, 接下来通过以下索引矩阵为每个区域保留高关联度来修改关联图:

$$I^r = \text{topkIndex}(A^r) \quad (7)$$

其中, I^r 的第 i 行包含了与第 i 个区域最相关的 K 个特征图区域。这个过程实现了不同目标任务之间的信息交互, 成功筛选出与检测目标关联性较高的信息, 使得这些信息在模型学习过程中占据主导地位, 而其他的无关信息则被抑制。实现了注意力的有效分配。

得到任务内的关联矩阵 I^r 后, 可以进行任务间的注意力计算, 即:

$$K^g = \text{gather}(K, I^r), V^g = \text{gather}(V, I^r) \quad (8)$$

其中, $K^g, V^g \in \mathbb{R}^{S^2 \times \frac{kHW}{S^2} \times C}$, 然后将注意力集中到收集到的信息上:

$$O = \text{Attention}(Q, K^g, V^g) \quad (9)$$

通过 Bi-level routing attention 双层路由的机制, 模型可以学习到不同任务之间的相关性和依赖关系, 并自适应地分配注意力权重, 从而提高多任务学习的效果和泛化能力。它帮助模型关注任务特定的重要特征, 促进任务间的信息交互, 减少过拟合, 并提升整体的学习性能。

2.2 Deformable Convolution v3(DCNv3)

基础的可形变卷积虽然能够更多地提取到关键目标的特征信息, 但在进行小目标检测的时候, 也会提取到很多无用的上下文信息, 会干扰特征提取的过程, 降低算法的表现。因此, 在这里本文采用最新的 DCNv3 算子, 能够弥补传统卷积在长距离建模能力和自适应空间聚合能力上的不足。DCNv3 在可形变卷积基础上具有以下扩展:

(1) 在神经元之间共享权值。原始的卷积之间具有

独立的线性投影权值, 因此参数和复杂度与采样点的数量呈线性关系, 这极大地限制了模型的效率。为了解决这一问题, 将原始卷积权值拆分为深度采样点两部分, 深度方向部分由原始位置感知调节, 采样点方向设置采样点之间共享同一投影权值。

(2) 引入多组机制。将空间聚合过程分为 N 个组, 每个组都有独立的采样偏移量和调节尺度, 因此单个卷积层上的不同组可以具有不同的空间聚合模式, 从而获取更多的特征信息。

(3) 归一化处理。原始的卷积用 Sigmoid 进行处理, 会导致训练不稳定的问题。这里使用 SoftMax 进行归一化处理, 使得整个训练过程更加稳定。

扩展后的 DCNv3^[29] 可表示为:

$$y(p_0) = \sum_{g=1}^G \sum_{k=1}^K w_g m_{gk} x_g(p_0 + p_k + \Delta p_{gk}) \quad (10)$$

在实现了稀疏全局建模的同时, 又适当地保留了归纳偏置, 可以说是在计算量和准确度之间进行了更好的权衡。

DCNv3 不仅具有检测所需的大范围有效感受野, 而且具有以输入和任务信息为条件的自适应空间聚合能力, 从而能够更好地完成复杂环境下的小目标检测任务。

2.3 Inner-IOU 损失函数

IOU 损失函数在计算机视觉任务中得到广泛的使用。原 YOLOv8 采用的基于 IOU 的损失函数是与 YOLOv5 一致的 CIOU Loss, 一种在分类问题中常用的损失函数。它衡量目标检测模型在检测目标位置和大小方面的准确性, 能够更好地处理目标框之间的重叠和错位情况, 提高了模型的性能评估。

CIOU Loss 的计算方式如下:

$$L_{\text{CIOU}} = 1 - \text{IOU} + \frac{\rho^2(b, b^{\text{gt}})}{C^2} + \alpha v \quad (11)$$

其中, v 仅包含了要预测的横纵比, 由于在小目标检测中检测目标的 w 和 h 相对较小, 导致 CIOU 无法反应真实情况, 可能会以不合理的方式优化相似性。

因此, 在实际的目标检测应用场景当中, 由于物体的尺度可能会发生很大的变化, 尤其在小目标检测中样本受环境因素影响较大, CIOU Loss 具有一定的局限性。一方面, 其对于长宽比例较大的物体及不规则物体的判断能力较弱; 另一方面, 无法针对不同的检测任务和检测目标进行自适应调整, 没有考虑不同样本之间的平衡性, 当样本差异性较大时, 也会制约模型的收敛速度, 所以 CIOU Loss 在小目标检测中不具备较好的判断能力和泛化能力。

为解决以上局限性, 本文引入基于辅助边框的 Inner-IOU 损失函数, 通过辅助边框计算 IOU 损失。使用较小尺度的辅助边框计算 IOU 损失将有助于高 IOU 样本回归, 达到加速收敛的效果。与之相反使用较大尺

度的辅助边框计算IOU损失能够加速低IOU样本回归过程。基于此,针对不同的数据集与检测器,Inner-IOU引入尺度因子 ratio 控制辅助边框的尺度大小用于计算损失,能够更快更有效的获得回归结果。

如图5所示,分别为较小尺度和较大尺度情况下Inner-IOU计算方式的示意图,将Inner-IOU应用至CIOU损失函数中的计算方式如下所示:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \quad (12)$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \quad (13)$$

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \quad (14)$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \quad (15)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (16)$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \quad (17)$$

$$IOU^{inner} = \frac{inter}{union} \quad (18)$$

$$L_{Inner-CIOU} = L_{CIOU} + IOU - IOU^{inner} \quad (19)$$

与传统的IOU计算方法考虑预测边界框和整体边界框间的整体重叠区域相比,Inner-IOU通过关注边界框的核心部分,提供了对重叠区域更加精确的判断,通过尺度因子ratio控制辅助边框的大小,实现对于不同检测任务和检测目标的动态调整,使得模型具有更好的泛化能力。

Inner-IOU是一种更细致、更专注于目标中心的性能评估指标,它通过辅助框的尺度调整提高了目标检测任务的精确度和效率。在小目标检测中,Inner-IOU的特性被发挥到极致,能够在小范围能准确判断出检测目标。

3 实验与结果分析

3.1 数据集

选用四个数据集作为文本的实验数据,均为公开数据集,且均为真实世界图像包含的交通标志,分别是RoadSign、CCTSDb、TSDD、GTSDB。数据集相关样例及标注如图6所示,标注的交通标志目标均满足小目标检测的条件,其中RoadSign和GTSDB为小样本目标检测,TSDD为单目标检测,便于检测在不同条件下模型

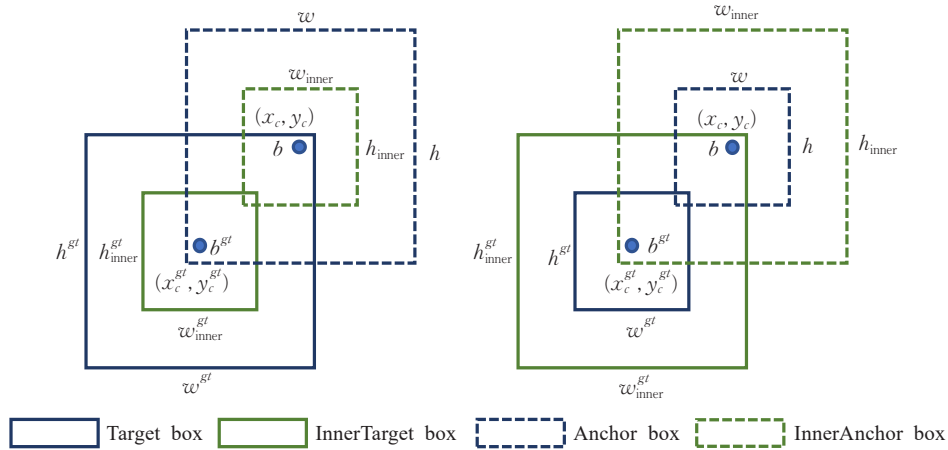


图5 Inner-IOU示意图

Fig.5 Description of Inner-IOU



图6 数据集样例图

Fig.6 Sample dataset diagram

的适应能力,实验表明,均取得了较好的结果。

RoadSign 是 Paddle 官网用于目标检测的公开数据集,是一款双目相机数据集,采集多个城市,包括了春夏秋冬四个季节的街区场景,是在不同的时间,天气条件,照明条件以及移动模糊下收集的,存在遮挡、模糊等目标信息,可以用于小目标检测。数据尺寸统一为 $2\,048 \times 1\,024$ 。将数据集分为训练集和测试集,共 877 张图片,其中训练集 701 张图片,验证集 176 张图片,包括 Speed-limit、Crosswalk、Trafficlight、Stop 四类城市街道场景中常见的类别作为检测对象。

CCTSDb 数据集是由长沙理工大学的相关学者及团队制作而成的,其有交通标志样本图片 17 856 张,共含交通标志近 40 000 个,标注了常见的指示标志(Mandatory)、禁令标志(Prohibitory)及警告标志(Warning)三大类交通标志。其中部分图片在原始图片基础上进行横向改变尺寸的图片,添加椒盐噪音,调整亮度等操作转换而来,另有部分图片是对行车视频进行逐帧抽取的图片。分辨率由 $1\,000 \times 350$ 至 $1\,024 \times 768$,且包含不同天气,环境,路况,时间下收集到的真实街景交通标志图。将数据集分为训练集、测试集和验证集,其中训练集 13 084 张图片,测试集 1 500 张图片,验证集 3 272 张图片。

TSDD 数据集是由国家自然科学基金发布的中国交通标志数据集,包括 10 000 张交通场景图像,包含多种交通标志近 30 000 个,图像是在不同的时间,天气条件,照明条件以及移动模糊下收集的,全部为原始图像,分辨率由 $5\,184 \times 3\,456$ 至 640×320 。将数据集分别训练集,测试集和验证集,其中训练集 7 000 张图片,测试集 2 000 张图片,验证集 1 000 张图片,将所有交通标志作为一个分类(Traffic sign)进行单目标检测。

GTSDb 数据集为德国交通标志数据集,前面三个数据集均为国内数据集,选取一个国外数据集进行测试,证明本文模型具有较好的普适性。GTSDb 数据集选自德国不同条件下的街道场景图像,包含图片共 900 张,标注交通标志近 2 000 个,分为指示标志(Mandatory)、禁令标志(Prohibitory)、危险标志(Danger)及其他标志(Others),分辨率为 $1\,360 \times 800$,将数据集分为训练集和测试集,其中训练集 600 张图片,测试集 300 张图片。

3.2 模型实验环境设置与训练结果

(1) 实验环境设置

本实验的硬件环境为 Intel i5-10400K CPU、1 张 RTX3060 GPU 显卡和 2 根 8 GB 内存条。操作系统为 windows11 的 64 位操作系统,版本号 22H2,采用 PyTorch1.11 作为深度学习框架,CUDA 版本 11.3, cuDNN 版本 8.2, Python 版本 3.8。

(2) 实验参数设置

在城市街道场景识别训练时,模型的输入样本尺寸转换为 640×640 ,批量大小为 16,采用 SGD 优化器,初试

学习率设置为 0.01,动量设置为 0.937。Epoch 设置为 300,训练时长大约 6~24 h。具体参数设置如表 1 所示。

表 1 实验参数设置

Table 1 Experimental parameter settings

参数名称	参数值
输入尺寸	640×640
初始学习率	0.01
最小学习率	0.000 1
迭代次数	300
批量大小	16
优化器	SGD
动量	0.937
权重衰减	0.000 5
线程数量	8
ratio	0.75

3.3 模型性能与目标检测结果分析

本文采用均值平均精度(mean average precision, mAP)mAP50, mAP50:95 作为模型的评价指标,其值的大小与模型呈正相关。计算方式如下:

$$P = \frac{TP}{TP + FP} \quad (20)$$

$$R = \frac{TP}{TP + FN} \quad (21)$$

$$mAP = \frac{\sum AP}{N(class)} \quad (22)$$

其中, TP 为检测正确的目标数量, FP 为检测错误的目标数量, FN 表示漏检数量, P 表示准确率, R 表示召回率, AP 值为 P 、 R 曲线下包围的面积, AP 表示所有类别的 AP 值总和, $N(class)$ 表示类别总数。mAP50 表示将预测框与真实边界框的交并比(intersection over union, IoU)设置为 0.50 时,计算每个类别的 AP 的均值。mAP50:95 表示 IoU 从 0.50 取到 0.95,间隔为 0.05 时 AP 的均值。 P 和 R 均在 IoU 阈值为 0.50 时统计。

为评估本文模型在城市街道场景中的检测性能,在四个数据集上对模型进行了测试。表 2~5 给出了验证集图片上 YOLOv8 模型与本文提出的模型进行的检测对比,由实验可知,在 RoadSign 数据集上本文模型的 mAP50 和 mAP50:95 分别达到了 90.7% 和 75.1%,分别比 YOLOv8 模型高出了 5.9 和 4.8 个百分点;在 CCTSDb 数据集上, mAP50 和 mAP50:95 分别提高了 2.3 和 1.6 个百分点;在 TTSD 数据集上, mAP50 和 mAP50:95 分别提高了 3.5 和 2.1 个百分点;在 GTSDb 数据集上, mAP50 和 mAP50:95 分别提高了 4.7 和 3.4 个百分点。

图 7 中给出了不同场景的测试对比图,依次为边框标注的图像、YOLOv8 模型检测结果及本文模型的检测结果,实验结果表明对小目标密集目标和遮挡目标的检测,本文模型具有明显优势。具体实验效果如表 2~5 所示。

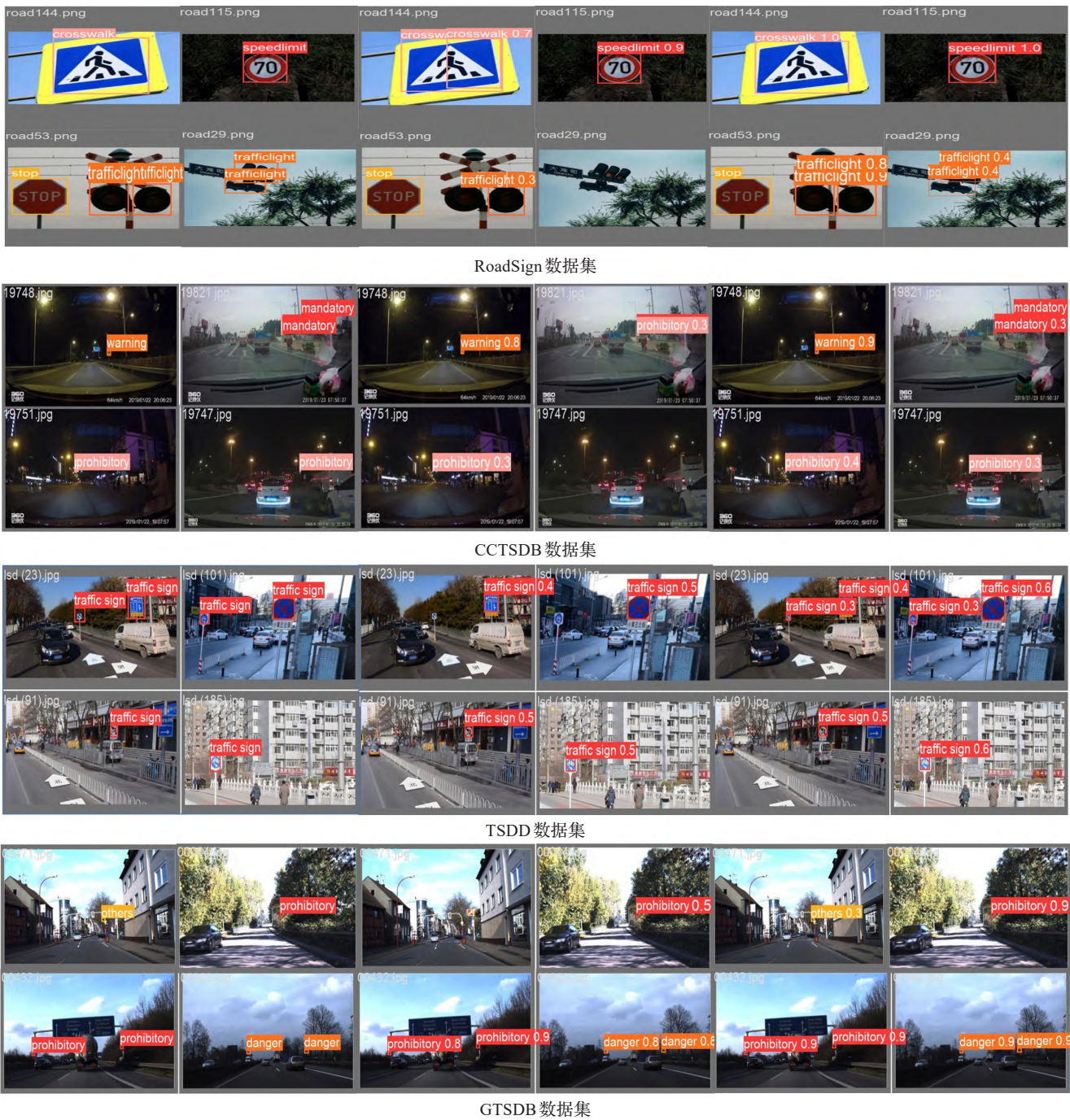


图7 四个不同数据集检测效果对比

Fig.7 Comparison of detection effects between four different datasets

表2 RoadSign数据集上改进前后模型效果对比

Table 2 Comparison of model effects before and after improvement on RoadSign dataset 单位: %

类别	YOLOv8n		本文模型	
	mAP50	mAP50:95	mAP50	mAP50:95
Speedlimit	96.5	84.1	99.5	87.9
Crosswalk	92.4	80.2	98.4	85.8
Trafficlight	55.3	39.5	66.2	44.5
Stop	94.8	77.5	98.7	82.3
All	84.8	70.3	90.7	75.1

除了在mAP50和mAP50:95指标对模型改进前后进行对比外,还在四个不同数据集上对改进前后模型权

表3 CCTSDB数据集上改进前后模型效果对比

Table 3 Comparison of model effects before and after improvement on CCTSDB dataset 单位: %

类别	YOLOv8n		本文模型	
	mAP50	mAP50:95	mAP50	mAP50:95
Mandatory	71.3	47.0	72.6	48.4
Prohibitory	81.4	52.7	83.2	53.4
Trafficlight	82.8	49.0	86.7	51.8
All	78.5	49.6	80.8	51.2

重大小及检测速度FPS进行了对比,如表6所示。实验结果表明模型权重大小在只有少量增加的情况下具有

表4 TSDD数据集上改进前后模型效果对比

Table 4 Comparison of model effects before and after improvement on TSDD dataset 单位: %				
类别	YOLOv8n		本文模型	
	mAP50	mAP50:95	mAP50	mAP50:95
Traffic sign	84.3	38.6	87.8	40.7
All	84.3	38.6	87.8	40.7

表5 GTSDB数据集上改进前后模型效果对比

Table 5 Comparison of model effects before and after improvement on GTSDB dataset 单位: %				
类别	YOLOv8n		本文模型	
	mAP50	mAP50:95	mAP50	mAP50:95
Mandatory	83.0	61.9	88.5	62.3
Prohibitory	94.7	73.0	98.8	75.5
Danger	90.8	63.9	96.4	72.0
Others	81.4	59.0	85.0	61.6
All	87.5	64.5	92.2	67.9

更快的图片处理速度,满足实时性的需求,提高了模型的检测效率。

表6 四个数据集改进前后模型权重大小及速度对比

Table 6 Comparison of model weight size and speed before and after improvement of four datasets				
数据集	YOLOv8n		本文模型	
	权重/MB	FPS	权重/MB	FPS
RoadSign	5.8	238	6.4	277
CCTSDb	5.9	265	6.2	314
TSDD	5.9	227	6.2	250
GTSDB	6.2	167	6.5	217

3.4 消融实验

为验证改进方法对YOLOv8模型的影响,对上述的改进(可形变卷积模块DCNv3、BRA注意力机制和Inner-IoU损失函数)在RoadSign数据集、CCTSDb数据集和TSDD数据集上进行了消融实验以证明改进方法的有效性和必要性。“√”表示在模型中使用了该方法,“—”表示未使用该方法,使用YOLOv8原模块。由表7~9可知,每一项改进都对模型的检测精度有一定的提升,由此证明了改进方法的科学性和有效性。

表7 RoadSign数据集消融实验对比

Table 7 Comparison of ablation experiments on RoadSign dataset 单位: %				
BRA	DCNv3	Inner-IoU	mAP50	mAP50:95
—	—	—	84.8	70.3
√	—	—	87.3	73.4
—	√	—	88.1	74.1
—	—	√	85.6	71.8
√	√	√	90.7	75.1

从消融实验看,可形变卷积DCNv3和BRA注意力机制两个重要模块对检测结果都有着一定的提升,将改进前后的模型在复杂环境下的检测结果进行定性对比分析,如图8、9所示。

表8 CCTSDb数据集消融实验对比

Table 8 Comparison of ablation experiments on CCTSDb dataset 单位: %				
BRA	DCNv3	Inner-IoU	mAP50	mAP50:95
—	—	—	78.5	49.6
√	—	—	78.8	49.5
—	√	—	79.2	50.4
—	—	√	79.0	50.2
√	√	√	80.8	51.2

表9 TSDD数据集消融实验对比

Table 9 Comparison of ablation experiments on TSDD dataset 单位: %				
BRA	DCNv3	Inner-IoU	mAP50	mAP50:95
—	—	—	84.3	38.6
√	—	—	85.6	39.8
—	√	—	86.2	40.1
—	—	√	85.6	39.6
√	√	√	87.8	40.7



图8 改进前后在夜晚条件下的检测结果对比

Fig.8 Comparison of detection results under night conditions before and after improvement



图9 改进前后在逆光及遮挡条件下的检测结果对比

Fig.9 Comparison of detection results under backlight and occlusion conditions before and after improvement

如图8(a)所示,在夜晚环境下图像识别清晰度和亮度降低,对于小目标的定位难度更高,在图8(b)中,改进前的模型并没有检测到交通标志,在图8(c)中,改进后的模型定位并检测到了目标。

如图9(a)所示,在逆光条件且多个交通标志存在相互遮挡的情况下,图像清晰度较差,目标难以捕捉,对于交通标志的检测有一定难度,在图9(b)中,改进前的模型并没有检测到交通标志,在图9(c)中,改进后的模型定位并检测到了目标。

3.5 对比实验

为进一步验证模型的有效性,在RoadSign数据集和CCTSDb数据集上,将本文提出的模型与SSD、Faster R-CNN、CenterNet、YOLOv3、YOLOv5、YOLOv5-ghost-Net、YOLOv7、YOLOv7-tiny、YOLOX、YOLOv8-EMA、YOLOv8-CBAM进行检测对比,图像输入尺寸统一设

置为640×640,进行对比的超参数和训练参数都设置为默认值。实验结果如表10、11所示,实验结果证明了本文模型的有效性。

表 10 RoadSign数据集上与其他模型效果对比

Table 10 Comparison of performance with other models on RoadSign dataset			
模型	骨干网络	mAP50	单位:% mAP50:95
SSD	VGG16	76.2	62.8
Faster R-CNN	Resnet50	77.6	63.5
YOLOv3	CPSDarknet53	79.2	65.7
YOLOv5	CPSDarknet53	80.4	66.4
YOLOv5-ghostNet	GhostNet	86.8	71.6
CenterNet	Resnet50	81.2	66.7
YOLOv7	CPSDarknet53	83.6	69.2
YOLOv7-tiny	EfficientNet	86.6	71.8
YOLOX	CPSDarknet53	87.2	72.5
YOLOv8-EMA	CPSDarknet	86.3	72.2
YOLOv8-CBAM	CPSDarknet	86.6	72.0
本文模型	本文网络	90.7	75.1

表 11 CCTSDB数据集上与其他模型效果对比

Table 11 Comparison of performance with other models on CCTSDB dataset			
模型	骨干网络	mAP50	单位:% mAP50:95
SSD	VGG16	72.5	44.6
Faster R-CNN	Resnet50	74.6	45.3
YOLOv3	CPSDarknet53	75.8	47.5
YOLOv5	CPSDarknet53	77.6	48.2
YOLOv5-ghostNet	GhostNet	79.8	50.0
CenterNet	Resnet50	77.8	49.2
YOLOv7	CPSDarknet53	78.2	49.0
YOLOv7-tiny	EfficientNet	80.2	49.8
YOLOX	CPSDarknet53	80.1	50.3
YOLOv8-EMA	CPSDarknet	78.7	49.5
YOLOv8-CBAM	CPSDarknet	79.0	49.4
本文模型	本文网络	80.8	51.2

4 结语

本文针对道路复杂场景下的交通标志目标检测任务,在YOLOv8模型的基础上提出了具体的改进方案。在backbone中增加可形变卷积DCNv3模块,提高模型对于遮挡目标的关注能力,在复杂环境下高效提取小目标的有效特征信息。在head中使用BRA注意力机制模块替换原有的C2f模块,在YOLO中融入注意力机制模块,引导网络更好地关注到不同尺度上的感受野信息,利用注意力机制对融合前的特征图进行更新,将更多注意力聚焦于关键特征,使融合后用于目标检测的特征图包含更多有效信息,加强对于小目标的学习能力,实现了模型对微小目标的平衡学习,同时使用Inner-IOU代替CIOU作为损失函数的计算方式,能够区分不同的回归样本并使用不同尺度的辅助边框来计算损失,有效

加速边界框回归过程,有效提高了模型的泛化能力。在四个不同的数据集上分别进行了小样本训练、大样本训练、单目标检测和多目标检测,模型的mAP50和mAP50:95值相较于原始模型都获得了不同程度的提升。文模型仍有改进空间,未来将考虑在保证模型检测精度的条件下,优化模型结构,提高模型的检测速度和泛化能力,减少过拟合,使模型能更好地应用于复杂环境下的城市街道场景检测。

参考文献:

[1] 刘海斌,张友兵,周奎,等.改进YOLOv5-S的交通标志检测算法[J].计算机工程与应用,2024,60(5):200-209.
LIU H B, ZHANG Y B, ZHOU K, et al. Traffic sign detection algorithm based on improved YOLOv5-S[J]. Computer Engineering and Applications, 2024, 60(5): 200-209.

[2] BARODI A, BAJIT A, ZEMMOURI A, et al. Improved deep learning performance for real-time traffic sign detection and recognition applicable to intelligent transportation systems [J]. International Journal of Advanced Computer Science and Applications, 2022, 13(5): 712-723.

[3] AGRAWAL S, CHAURASIYA K R. Ensemble of SVM for accurate traffic sign detection and recognition[C]//Proceedings of the 1st International Conference on Graphics and Signal Processing, 2017: 10-15.

[4] REN X Y, MIN Z. An overview of traffic sign detection and recognition algorithms[C]//Proceedings of the International Journal of Multimedia Information Retrieval, 2022: 193-210.

[5] YAZDAN R, VARSHOSAZ M. Improving traffic sign recognition results in urban areas by overcoming the impact of scale and rotation[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 171: 18-35.

[6] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.

[7] GIRSHICK R. Fast R-CNN[C]//Proceedings of the IEEE International Conference on Computer Vision, 2015: 1440-1448.

[8] REN S, HE K, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[C]//Advances in Neural Information Processing Systems, 2015: 1137-1149.

[9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Proceedings of the European Conference on Computer Vision, 2016: 21-37.

[10] 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.

- [11] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 6517-6525.
- [12] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [13] REDMIN J, FARHADI A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6517-6525.
- [14] REDMON J, FARHADI A. YOLOv3: an incremental improvement[J]. arXiv:1804.02767, 2018.
- [15] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. Yolov4: optimal speed and accuracy of object detection[J]. arXiv: 2004.10934, 2020.
- [16] LI C, LI L, JIANG H, et al. YOLOv6: a single-stage object detection framework for industrial applications[J]. arXiv: 2209.02976, 2022.
- [17] WANG C Y, BOCHKOVSKIY A, LIAO H. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real time object detectors[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 7464-7475.
- [18] DAI J F, QI H Z, XIONG Y W, et al. Deformable convolutional networks[C]//Proceedings of the 2017 IEEE International Conference on Computer Vision, 2017: 764-773.
- [19] ZHU L, WANG X J, KE Z H, et al. BiFormer: vision transformer with bi-level routing attention[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 10323-10333.
- [20] YAN F, XU Y. Improved target detection algorithm based on YOLO[C]//Proceedings of the 2021 4th International Conference on Robotics, Control and Automation Engineering (RCAE), 2021: 21-25.
- [21] LI J, WANG H, XU Y, et al. Road object detection of YOLO algorithm with attention mechanism[J]. Frontiers in Signal Processing, 2021, 5: 9-16.
- [22] MA Y, ZHANG S. Feature selection module for CNN based object detector[J]. IEEE Access, 2021, 9: 69456-69466.
- [23] JU M, LUO J, WANG Z, et al. Adaptive feature fusion with attention mechanism for multi-scale target detection[J]. Neural Computing and Applications, 2021, 33: 2769-2781.
- [24] YU F, KOLTUN V. Multi-scale context aggregation by dilated convolutions[C]//Proceedings of the International Conference on Learning Representations, 2016.
- [25] ZHU X Z, HU H, LIN Z F, et al. Deformable ConvNets V2: more deformable, better results[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.
- [26] 张红民, 庄旭, 郑敬添, 等. 优化 YOLO 网络的人体异常行为检测方法[J]. 计算机工程与应用, 2023, 59(7): 242-249.
- ZHANG H M, ZHUANG X, ZHENG J T, et al. Optimize the detection method of human abnormal behavior of YOLO network[J]. Computer Engineering and Applications, 2023, 59(7): 242-249.
- [27] 何儒汉, 熊捷繁, 熊明福. 基于背景自适应学习的行人重识别算法研究[J]. 计算机工程与应用, 2023, 59(7): 126-133.
- HE R H, XIONG J F, XIONG M F. Study of pedestrian re-identification algorithm based on background adaptive learning[J]. Computer Engineering and Applications, 2023, 59(7): 126-133.
- [28] 何坚, 郭泽龙, 刘乐园, 等. 基于滑动窗口和卷积神经网络的可穿戴人体活动识别技术[J]. 电子与信息学报, 2022, 44(1): 168-177.
- HE J, GUO Z L, LIU Y Y, et al. The wearable human activity identification technology based on sliding windows and convolutional neural networks[J]. Journal of Electronics and Information, 2022, 44(1): 168-177.
- [29] WANG, W H, DAI J F, CHEN Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 14408-14419.