

融合坐标注意力和 BiFPN 的 YOLOv5s 交通标志检测方法^{*}

韩岩江 王 伟 王峰萍

(西安工程大学 西安 710048)

摘 要:针对不同光照条件下的小目标交通标志检测存在的不易检测、错检等问题,提出了一种融合坐标注意力机制和双向加权特征金字塔(BiFPN)的 YOLOv5s 交通标志检测方法。首先,在特征提取网络中融入坐标注意力机制,提升网络对重要特征的关注意度,增加模型在不同光照条件下的检测能力;其次,在特征融合网络中使用 BiFPN,提升模型的特征融合能力,改善对小目标交通标志的检测能力;最后,考虑到真实框与预测框之间的方向匹配问题,将 CIOU 损失函数改为 SIOU 损失函数,进一步提升模型的检测性能。在 GTSDb 数据集上进行验证,与原始模型相比,平均精度均值(mAP)提升了 3.9%,推理时间为 2.5 ms,能够达到实时检测的标准。

关键词:交通标志检测;YOLOv5s;注意力机制;特征融合;损失函数

中图分类号: TP391 **文献标识码:** A **国家标准学科分类代码:** 520.6040

YOLOv5s traffic sign detection method based on coordinate attention and BiFPN

Han Yanjiang Wang Wei Wang Fengping

(Xi'an Polytechnic University, Xi'an 710048, China)

Abstract: Aiming at the problems of difficult detection and misdetection in the detection of small target traffic signs under different lighting conditions, a YOLOv5s traffic sign detection method that combines coordinate attention mechanism and BiFPN is proposed. First, the coordinate attention mechanism is integrated into the backbone network of the network to improve the network's attention to important features and increase the detection ability of the model under different lighting conditions; secondly, the BiFPN is used in the feature fusion network. Improve the feature fusion ability of the model and improve the detection ability of small target traffic signs; finally, considering the direction matching problem between the real frame and the predicted frame, the CIOU loss function is changed to the SIOU loss function to further improve the detection performance of the model. Validated on the GTSDb dataset, compared with the original model, the mAP is increased by 3.9%, and the inference time is 2.5 ms, which can meet the standard of real-time detection.

Keywords: traffic sign detection; YOLOv5s; attention mechanism; feature fusion; loss function

0 引 言

交通标志检测是计算机视觉应用于道路交通领域的重要研究方向,也是无人驾驶系统的基础。交通标志检测结果的实时性、准确性与驾驶人员的人身安全息息相关。受光照、距离等环境因素的影响,交通标志检测难度高于

一般目标检测,所以提出实时性与准确性兼顾的交通标志检测算法对驾驶安全至关重要。

传统交通标志检测方法可分为基于颜色的和基于形状两种类型^[1-2]。基于颜色的交通标志检测方法,利用不同颜色空间模型来对交通标志进行检测,但由于自然场景下不同光照照射交通标志会产生不同颜色,导致其检测

收稿日期:2022-08-20

^{*} 基金项目:陕西省教育厅一般专项(20JK0660)、2021 年中国高校产学研创新基金(2021ALA02002)、2021 年“纺织之光”中国纺织工业联合会高等教育教学改革研究项目(2021BKJGLX004)、西安工程大学 2020 年高等教育研究项目(20GJ05)资助

准确性不高。基于形状的交通标志检测最常用的方法是使用霍夫变换^[3],形状检测方法虽然避免了自然条件下的光照影响,但是由于标志老化、变形等问题,不能取得很好的检测效果,存在错检、漏检等问题。

近年来,卷积神经网络(convolutional neural network, CNN)作为一种主流的深度学习方法,其在计算机视觉方面有着广泛的应用。当前主流的目标检测方法被分成两类,单阶段检测方法和两阶段检测方法,其主要区别是是否使用区域候选框。两阶段检测方法先生成区域候选框,后进行分类和位置的判定,使得网络分割成为两个阶段,典型的两阶段检测方法有 R-CNN、Fast R-CNN、Faster R-CNN、Mask R-CNN 等;单阶段检测则没有生成区域候选框这一过程,典型的单阶段检测方法以 YOLO^[4-7]系列和 SSD^[8]为代表,与两阶段目标检测方法相比,单阶段目标检测方法在检测速度上有了显著提升,同时检测精度也能满足使用要求。

目前,针对交通标志检测方向存在的问题,很多学者开始探索使用深度学习方法进行研究。Xiong 等^[9]提出了一种基于 CNN 的交通标志检测算法,使用 RPN 网络来检测交通标志,提升了模型的检测精度,但是其模型体积较大,在使用中易受到限制。张上等^[10]使用剪枝操作缩小模型体积,有效解决了模型体积过大的问题,但检测精度有所下降。吕禾丰等^[11]通过改进非极大值抑制算法,通过提高生成检测框的准确率,提升检测精度。Khan 等^[12]针对于复杂环境下交通标志中存在着较暗区域的问题,提出了具有照明预处理能力的新型智能交通标志识别系统,通过亮度增强技术增强图像较暗区域的光线,在 GTSDb 数据集上取得了良好的效果。赵坤等^[13]基于 YOLOv3 提出了一种实时自适应图像增强方法,主要针对弱光照条件下进行自适应增强,提升了模型在弱光条件下的检测性能。文献^[12-13]虽然对于弱光条件下的交通标志检测效果有所改善,但是不能很好的兼顾小目标交通标志的检测问题。为了解决小目标交通标志检测精确度较低的问题,李烨等^[14]通过使用融合空间注意力机制的倒金字塔网络来进行交通标志检测,提升了小尺度交通标志的检测性能,但是其检测速度相对较低。郭璠等^[15]基于 YOLOv3,将通道注意力机制和空间注意力方法融入到网络中,改善了模型对小目标检测的性能。Wang 等^[16]基于 YOLOv5,提出了一种新型的特征融合网络 AF-FPN 来用于交通标志检测,对于小目标标志的检测精度提升明显,但是实时检测速度有所下降。文献^[14-16]改善了小目标交通标志的检测精度,但其都存在着自身局限性,且在不同光照下模型的泛化性较差,不能很好的用于实际应用。

YOLO 模型以其速度快、精度高的特点在很多领域都体现了优秀的检测性能^[17-19],因此,针对不同光照条件下的小目标交通标志检测存在的检测、错检等问题,本文基于最新的 YOLOv5s 模型,提出了一种融合坐标注

意力机制(coordinate attention, CA)和双向加权特征金字塔(BiFPN)的交通标志检测方法。首先,在主干提取网络中引入坐标注意力机制,使得网络更加关注感兴趣区域,增加网络的特征提取能力;其次,使用 BiFPN 作为特征融合网络,增强网络的特征融合能力,使得位置信息与语义信息充分融合;最后使用 SIoU 损失函数来替代原始模型中的 CIoU 损失函数。本文在 GTSDb 数据集上进行实验,实验结果表明改进后的模型对强光、弱光等条件下的检测准确率提升明显,同时解决了小目标交通标志检测困难的问题,提高了模型的鲁棒性和可靠性。

1 本文方法

1.1 YOLOv5s 结构

YOLO 是单阶段目标检测的代表,其核心思想是将图像分为多个网格,如果目标物体的中心落在某个网格之内,就由该网格负责预测该目标物体。YOLOv5s 模型于 2020 年由 Jocher 等提出,其具有检测速度快、模型体积小、准确率高等优点。

YOLOv5s 由 4 个部分组成,包括输入端(Input)、主干网络(Backbone)、颈部网络(Neck)和预测网络(Prediction)。输入端的作用是图像预处理至输入分辨率大小以及数据增强等。主干网络的主要作用是进行特征提取,其采用 CSPDarkNet53 接收不同尺寸的特征图,并通过 CSP1_X 模块提取图片主要信息。颈部网络采用特征金字塔网络(feature pyramid networks, FPN)^[20]和路径聚合网络(path aggregation network, PAN)^[21],其作用是融合图像的浅层位置特征与深层语义特征,增强网络的特征融合能力,并生成不同尺寸特征图。预测网络根据生成不同尺寸的特征图来获取不同大小目标物体的类别和位置信息。YOLOv5s 网络结构如图 1 所示。

1.2 改进网络结构

为了解决不同光照条件下小目标交通标志存在的检测、错检等问题,本文提出了一种融合坐标注意力机制和 BiFPN 的 YOLOv5s 交通标志检测方法。本文所提出方法的网络结构如图 2 所示,该网络结构包括 4 个部分,输入端、特征提取网络、特征融合网络和预测网络。所提出模型的主要有如下改进。

1)将 CA 融合到主干特征提取网络当中,增加网络对重要特征的兴趣,提高网络的特征提取能力。

2)使用 BiFPN 结构进行跨层特征融合,将语义信息和位置信息充分结合,增强网络的特征融合能力。

3)将原始模型中 CIoU 损失函数改进为 SIoU 损失函数,充分考虑真实框与预测框之间的方向匹配问题,提升模型收敛性能。

1.2.1 CA 模块

CA 由 Hou 等^[22]提出,其将位置信息嵌入到通道注意力机制中,使得该注意力机制会同时对位置信息与通道信息进行关注,取得了良好的效果。本文将 CA 引入网络

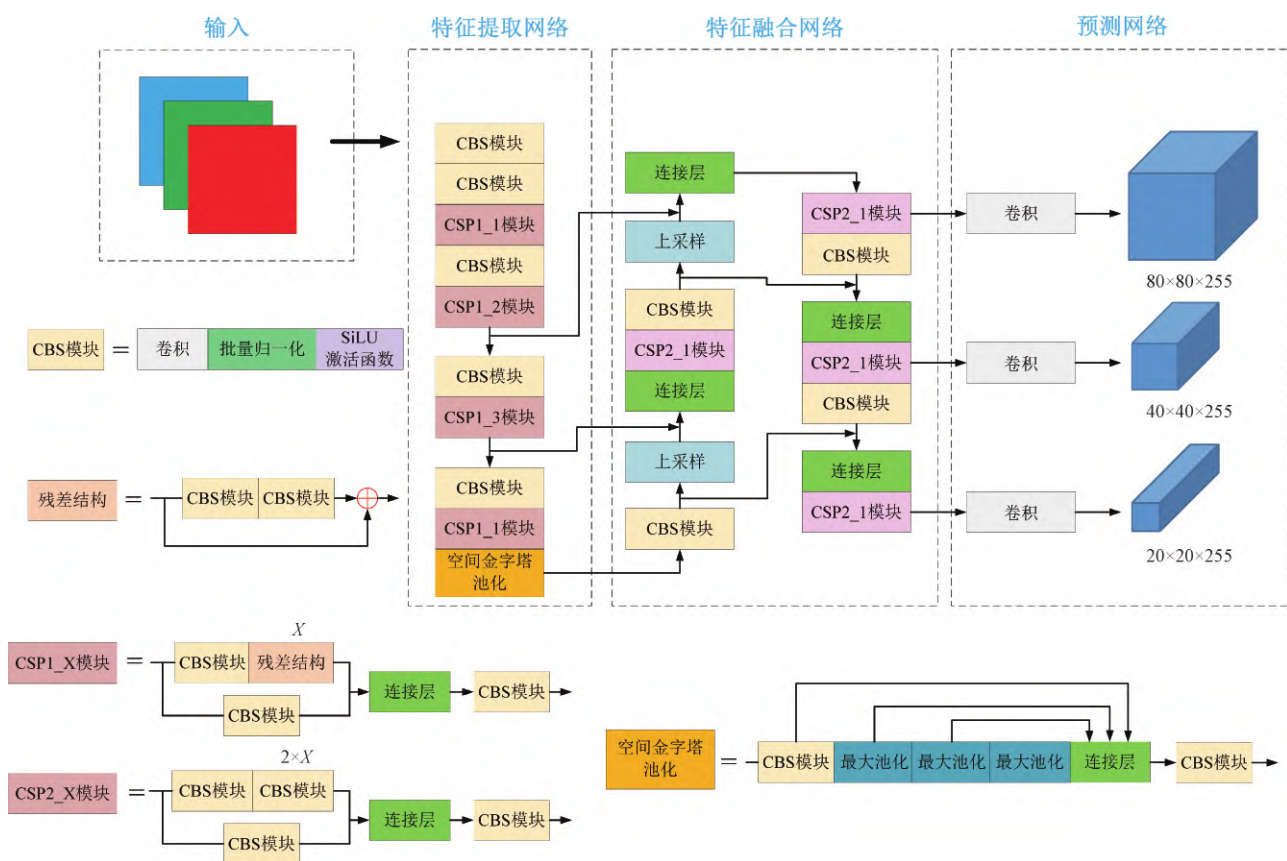


图1 YOLOv5s 网络结构

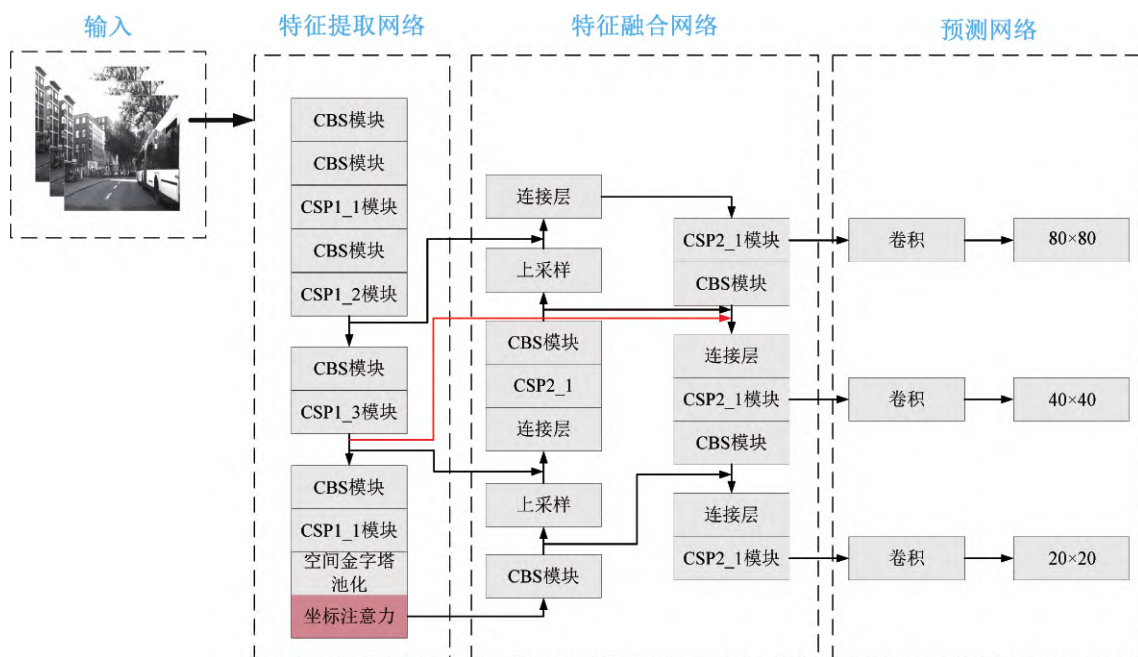


图2 本文模型网络结构

的主干特征提取网络,增强模型对重点特征的注意力,为重要特征添加权重,增强网络的特征提取能力。CA 模块

采用坐标信息嵌入和坐标注意力生成捕获特征的通道和位置之间的关系。

1) 坐标信息嵌入

坐标信息嵌入将输入特征 X 进行全局平均池化操作,用两个空间范围内的池化核 $(H, 1)$ 和 $(1, W)$ 分别沿着水平坐标和垂直坐标对每个通道进行编码。

高度 h 处的第 c 个通道输出可以表示为:

$$z_c^h(h) = \frac{1}{W} \sum_{0 \leq i < W} x_c(h, i) \quad (1)$$

宽度 w 处的第 c 个通道输出可以表示为:

$$z_c^w(w) = \frac{1}{H} \sum_{0 \leq j < H} x_c(j, w) \quad (2)$$

这种编码方式分别沿着两个方向进行聚合,形成了一对方向感知特征图,每张特征图都包含着输入特征在一个空间方向的远程特征依赖关系以及在另外一个空间方向精确的位置信息。坐标信息嵌入增加了网络输入信息的特征,加深了网络对于感兴趣目标区域的关注程度。

2) 坐标注意力生成

坐标注意力生成将坐标信息嵌入在两个方向上生成的特征图进行拼接,使用共享的 1×1 卷积变换函数 F_1 对其进行变换:

$$f = \delta(F_1([z^h, z^w])) \quad (3)$$

式中: δ 为非线性激活函数; r 为缩减率,作用是降低模型的复杂度。生成的特征图 $f \in \mathbb{R}^{\frac{C \times (H+W)}{r}}$ 是在水平方向和垂直方向对空间信息编码所形成的特征图。对特征图进行批量归一化和非线性处理,沿空间维度将 f 分割为 $f^h \in \mathbb{R}^{\frac{C \times H}{r}}$ 与 $f^w \in \mathbb{R}^{\frac{C \times W}{r}}$ 。另外使用两个 1×1 的卷积变换将 f^h 和 f^w 变换为与输入 X 具有相同通道数量的特征图 F_h 和 F_w ,最后使用 Sigmoid 激活函数进行处理,生成注意力权重 g^h 和 g^w 。

$$g^h = \sigma(F_h(f^h)) \quad (4)$$

$$g^w = \sigma(F_w(f^w)) \quad (5)$$

3) 坐标注意力输出

经过如上两个步骤,CA 模块结构如图 3 所示。输出的坐标注意力 Y 为:

$$y_c(i, j) = x_c(i, j) \times g_c^h(i) \times g_c^w(j) \quad (6)$$

CA 考虑了对空间信息的编码,式(6)最终输出的 Y 包含了来自于水平方向的注意力和来自垂直方向的注意力。两个注意力图中的每个元素都反应了感兴趣的对象是否存在于相应的行和列中。CA 能够更加准确的定位到感兴趣的区域,同时包含了通道信息与位置信息,减少了光照对于模型检测效果的影响,增加了模型在不同光照下的检测能力。

1.2.2 双向加权特征金字塔网络

特征金字塔结构 FPN 的目的是融合浅层的位置信息和深层的语义信息,如图 4(a)所示,原始的金字塔结构采用从上到下的信息融合,提升了网络的信息提取能力,但是融合过程也会导致信息的丢失。YOLOv5s 采用 PANet 结构,如图 4(b)所示, PANet 基于 FPN 图像特征

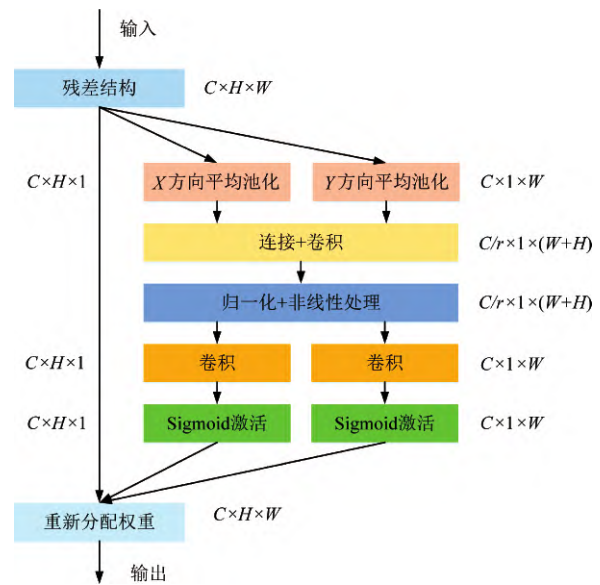


图3 CA 模块结构

金字塔的思想,不仅进行从上到下的特征融合,而且增加了从下到上的特征融合,从而达到减少信息丢失的目的,取得了较好的检测效果,但是网络训练的参数数量有所增加。对于交通标志检测而言,由于远距离时其目标太小,存在着较多的小目标对象,导致原始模型存在着检测精度低的问题。BiFPN^[23]如图 4(c)所示, BiFPN 增强了网络的信息提取能力,使其低层次的位置信息能够和高层次对的语义信息更好的结合,从而进一步提升网络对于目标的检测性能。原始网络的 PANet 结构只是在通道上进行叠加,而 BiFPN 将权重信息考虑在内,同时进行双向跨尺度特征融合。

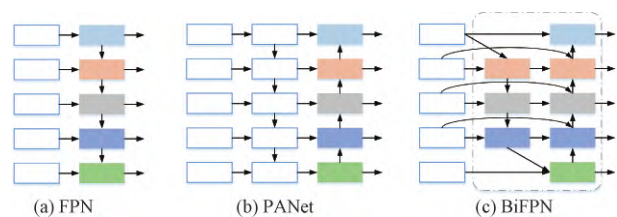


图4 不同特征金字塔结构

YOLOv5s 首先提取到不同下采样倍数的特征图,不同特征图用于检测不同大小的目标。网络首先将输入图像分别进行 8 倍下采样,16 倍下采样以及 32 倍下采样,所得到的特征图分别用于检测大目标、中等目标和小目标。其次将提取到的特征图输入特征融合网络进行特征融合,将低层的位置信息将高层的语义信息相融合。最终在检测网络中输出结果。

本文基于 BiFPN 的思想,使用 16 倍下采样后形成的 40×40 特征图与后面的特征图进行跨层连接,使得位置信息提取的更加充分,减少了特征信息丢失,增加网络对小目标的检测能力。同时其几乎不增加成本,对模型参数

量大小影响不大,改进后的结构如图5所示。

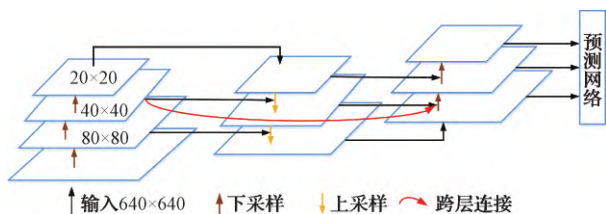


图5 BiFPN 特征融合网络

BiFPN 使用加权特征融合的方式为每个特征添加一个额外的权重。快速归一化融合式(7)用来约束每个权重的大小,使权重大小保持在 $0 \sim 1$,提高了模型在 GPU 上的运算速度。

$$O = \sum_{i=0} \frac{\omega_i}{\sum_{j=0} \omega_j + \varepsilon} \cdot I_i \quad (7)$$

式中: ω_i 代表可以学习的权重大小; I_i 表示输入特征。

1.2.3 SIoU 损失函数

交并比(intersection over union, IoU)^[24]代表预测框与真实框的重叠率,即两框之间交集与并集的比值,公式如下:

$$IoU = \frac{B \cap B^{GT}}{B \cup B^{GT}} \quad (8)$$

式中: B 为预测框; B^{GT} 为真实框。

IoU 损失存在不能反映两框之间的距离大小以及无法精确反应两框之间重合度大小的问题。GIoU^[25]改善了这一问题,能够很好的反应两框之间的重合度,但其对 IoU 的依赖较大,在垂直方向存在较大误差且模型不容易收敛。DIoU^[26]通过计算两个预测框中心的欧氏距离,取得了较好的效果,加速了模型的收敛。YOLOv5 模型采用 CIoU 作为损失函数,CIoU 在 DIoU 的基础上,将预测框的宽高比考虑在内,添加了惩罚项,使得预测框更加接近于真实框。

$$L_{CIoU} = 1 - IoU + \frac{\rho^2(b, b^{gt})}{c^2} + \alpha v \quad (9)$$

式中: α 为权重系数; v 用来度量长宽比的特性。

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w}{h} - \arctan \frac{w^{gt}}{h^{gt}} \right)^2 \quad (10)$$

Gevorgyan^[27]提出 SIoU 损失,除了考虑到预测框之间的距离、重叠面积和宽高比之外,还考虑了预测框与真实框之间的方向匹配问题,添加了角度惩罚项。本文采用性能更加高效的 SIoU 替代原始模型中的 CIoU,增加了网络的收敛性能,从而提高模型的整体性能。

SIoU 的思想是将模型在 X 轴或者 Y 轴上做预测,其次沿着相关轴进行逼近。首先尝试对角度进行最小化,如图6所示,如果 $\alpha \leq \frac{\pi}{4}$,则对 α 进行最小化,否则最小化

$$\beta = \frac{\pi}{2} - \alpha。$$

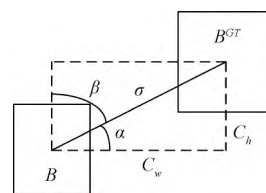


图6 SIoU 角度成本计算示意图

由此引入角度成本计算公式:

$$\Lambda = 1 - 2 \cdot \sin^2 \left(\arcsin x - \frac{\pi}{4} \right) \quad (11)$$

角度成本计算被应用于距离成本中,将角度成本引入距离成本,对距离成本进行重新定义:

$$\Delta = \sum_{l=x,y} (1 - e^{-\gamma \rho_l}) \quad (12)$$

形状损失的定义为:

$$\Omega = \sum_{l=w,h} (1 - e^{-\omega_l})^\theta \quad (13)$$

最终 SIoU 损失函数定义为:

$$L_{SIoU} = 1 - IoU + \frac{\Delta + \Omega}{2} \quad (14)$$

2 实验结果及分析

2.1 数据集与实验平台

实验数据集采用 GTSDB 数据集^[28],数据集包含各种不同天气如雨天、雾天、晴天等,涵盖了不同光照下的交通场景,该数据集被分为强制、禁止、警告、其他等4类,本文选择前3类图像进行实验。模型在训练中使用 Mosaic、水平翻转、平移等数据增强方式,以增强模型的泛化能力。

原始数据集图像为 ppm 格式,标签为 txt 格式,但是其与 YOLO 的 txt 格式不符合,需要进行格式转化。图像转化为.jpg 格式,将标签首先转化为 json 格式,再转化为 voc 格式,最后转化为所需要的 YOLO 标签格式。

实验使用 Ubuntu 18.04 操作系统,采用 Tesla V100-SXM2-16GB 进行训练,Python 版本为 3.8,使用 Pytorch 11.0 框架,CUDA 版本为 10.2。实验训练轮次为 300 轮,初始学习率设置为 0.01。

2.2 评测指标

实验采用评价指标包含准确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)。

$$P = \frac{TP}{TP + FP} \times 100\% \quad (15)$$

$$R = \frac{TP}{TP + FN} \times 100\% \quad (16)$$

$$AP = \int_0^1 p(r) dr \quad (17)$$

$$mAP = \frac{1}{c} \sum_{i=1}^c AP_i \quad (18)$$

式中: TP 表示被预测为正类的正样本; FP 表示被预测

为正类的负样本; FN 表示被预测为负类的正样本; c 表示类别数量。

2.3 实验结果

1) 检测效果

为了证明本文所提出模型的效果,选取了部分测试集

图片来进行说明,且选择的图像中均包含了警告、强制、禁止三类交通标志牌,不同光照下检测效果如图7所示。从图7可以看出,在强光正常目标、强光小目标、弱光正常目标、弱光小目标、自然光正常目标、自然光小目标等情况下,本文模型均取得了良好的检测结果。



图7 不同光照下检测效果

2) 对比实验

为了解决不同光照条件下交通标志检测准确度不高的问题,本文引入了CA注意力机制。为了验证CA注意力机制在本文模型上的效果,因此与在同样位置添加SE^[29]、ECA^[30]、CBAM^[31]等主流注意力机制进行对比如表1所示。根据表1可以看出,在模型浮点运算次数几乎相同的情况下,本文所使用的坐标注意力机制CA在mAP指标上均高于当前主流的SE、ECA以及CBAM注意力机制,能够达到93.6%。

为了验证本文使用BiFPN作为特征融合网络的性

表1 不同注意力机制比较

方法	mAP/%	FLOPs/($\times 10^9$)
SE	92.5	15.8
ECA	93.3	15.9
CBAM	92.8	15.8
CA	93.6	15.8

能,对主流特征融合网络FPN、PANet、BiFPN的mAP、模型参数量以及模型权重进行比较如表2所示。根据表2可知,采用自顶向下单方向的FPN网络检测精度不高。在FPN的基础上增加自底向上路径的PANet网络检测

性能有所提高。在 PANet 基础上增加跨层连接对的 BiFPN 网络的检测性能最好。mAP 较 PANet 提升了 2.1%, 同时其参数量和权重大小几乎无增长, 在不增加开销的情况下提升了网络的检测性能。

表 2 特征融合网络比较

方法	mAP/%	参数量/($\times 10^6$)	权重/MB
FPN	90.0	6.0	12.3
PANet	90.5	7.1	13.7
BiFPN	92.6	7.1	13.9

为了进一步验证本文提出模型对不同光照条件下交通标志的检测效果, 本文与 YOLOv3、YOLOv4、YOLOv5m、YOLOX^[32]、YOLOv6 以及原始网络 YOLOv5s 等主流目标检测模型进行对比, 网络训练均不使用预训练模型, 以突出本文所提出模型效果的优越性, 实验结果如表 3 所示。

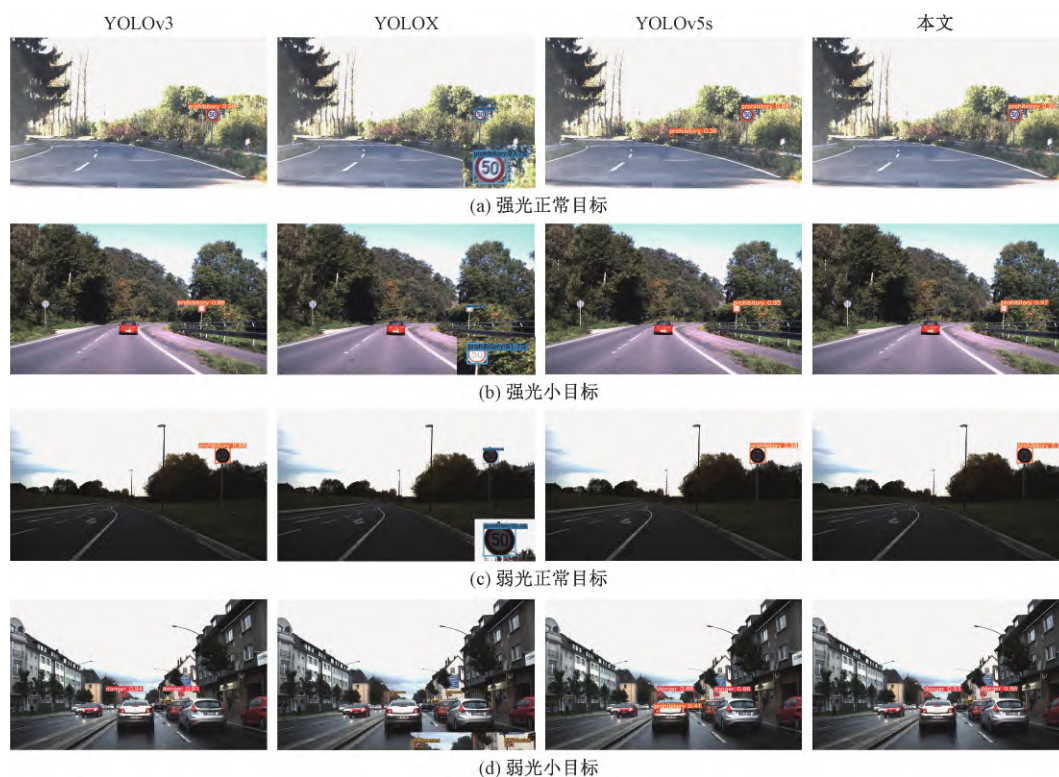
表 3 本文模型与其他主流模型对比

方法	mAP/%	t/ms	参数量/ ($\times 10^6$)	权重/ /MB	FLOPs/ ($\times 10^9$)
YOLOv3	94.0	6.1	61.5	123.5	154.7
YOLOX	89.7	2.9	8.9	68.51	26.6
YOLOv6	94.6	6.0	—	36.3	—
YOLOv5s	90.5	1.9	7.0	13.7	15.8
YOLOv5m	92.7	4.4	20.9	42.2	48.1
本文	94.4	2.5	7.1	14.6	16.1

根据表 3 可以看出, YOLOv3、YOLOv4、YOLOv6、

YOLOv5m 以及 YOLOX 的权重大小均远远大于本文模型, 为本文模型的 3~9 倍。与其相比, 本文所提出模型体积较小, 属于轻量级网络模型。虽然 YOLOv6 的 mAP 比本文模型略高 0.2%, 但是其推理时间, 模型权重都过大, 不适宜实际使用。与原始网络 YOLOv5s 相比, 本文模型虽然由于使用了 BiFPN 结构导致模型大小增加了 0.9×10^6 、推理速度上增加了 0.6 ms, 但是模型的 mAP 提升了 3.9%。本文模型在 mAP 值、推理时间、模型权重以及模型复杂程度上都优于 YOLOv3、YOLOX、YOLOv5m 等主流模型, 且检测速度能够达到实时检测的标准。

为了充分说明本文模型的有效性, 选取了部分测试集图像与 YOLOv3、YOLOX、YOLOv5s 模型进行对比, 以便于直观的展现本文模型的检测性能。如图 8 所示, 图 8(a)、(b) 在强光条件下对正常标志和小目标标志进行检测的效果。本文模型能够准确检测出交通标志牌, 且检测准确率高于其他 3 个模型。图 8(c)、(d) 为在弱光条件下对正常标志和小目标交通标志进行检测的结果。由图 8(d) 可以看出, 虽然本文模型检测准确率略低于 YOLOv3, 但是与原始网络相比, 本文模型提升了检测准确率, 改善了原始 YOLOv5s 网络存在的错检问题。图 8(e)、(f) 为在自然光条件下模型对正常标志和小目标标志的检测结果, 在自然光照条件下, 本文模型仍然能够保证较高的检测准确率。与改进之前的网络相比, 本文模型在各种光照条件下的检测准确率均有所提升, 同时对错检问题有所改善。



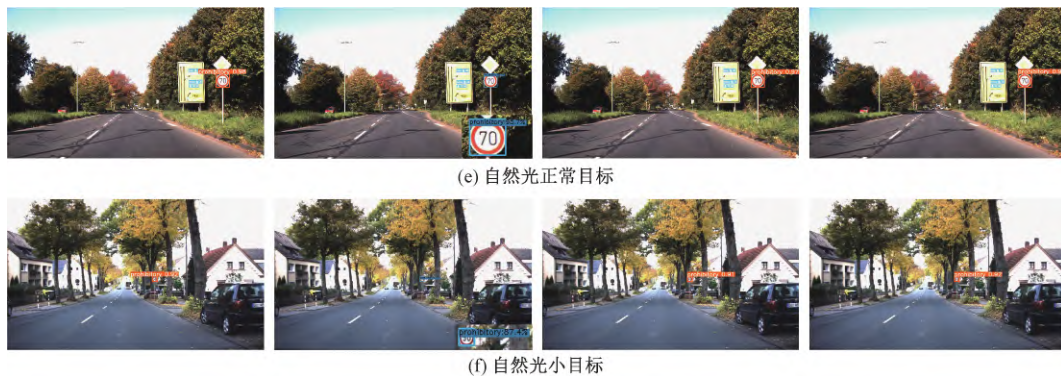


图8 不同模型检测效果对比

从对比实验结果可以看出,在不同光照条件下,本文模型在交通标志检测方面的性能高于其他主流模型,与原始模型 YOLOv5s 相比,本文模型对交通标志的检测效果更加优越,提升了交通标志的检测性能,且对于小目标标志检测精度提升明显,同时对于错检问题有所改善,更适

合在复杂场景下进行应用。

3) 消融实验

为了验证本文提出各模块在 YOLOv5 网络模型中的作用,本文进行了消融实验。消融实验的结果如表 4 所示。

表4 添加各模块后性能比较

方法	mAP/%	AP ₁ (Danger)	AP ₂ (Mandatory)	AP ₃ (Prohibitory)	t/ms	参数量 /($\times 10^6$)	FLOPs /($\times 10^9$)
YOLOv5s	90.5	92.2	82.8	96.5	1.9	7.0	15.8
YOLOv5s+SIoU	91.5	91.8	85.5	97.1	1.8	7.0	15.8
YOLOv5s+BiFPN	92.6	94.4	86.4	96.9	2.3	7.1	16.0
YOLOv5s+CA	93.6	95.1	87.6	98.2	1.9	7.1	15.8
本文	94.4	96.1	89.0	98.2	2.5	7.1	16.1

从表 4 可以看出,本文模型在参数量几乎没有增加的情况下,mAP 提升了 3.9%且推理时间仅为 2.5 ms。在使用 SIoU 损失函数以后,mAP 较使用 CIoU 时候相比提升了 1%。在特征融合网络部分使用 BiFPN,减少了特征损失,mAP 提升了 2.1%,增强了网络对目标的检测能力。在主干特征提取网络中引入了 CA 注意力,使得模型对目标的位置信息与通道信息紧密结合,增加了网络对重要特征的关注,mAP 提升了 3.1%。各种改进方式对警告、强

制、禁止类别的 AP 值均有所增加,本文模型对禁止类标志的检测精度最高,为 98.2%。本文模型与原始模型相比,mAP 提升 3.9%,对警告、强制、禁止类别标志的 AP 值分别提升 3.9%、6.2%、1.7%,取得了优异的检测效果。

改进了各个模块之后,模型在训练集上的训练效果如图 9 所示。图 9(a)为当 $IoU=0.5$ 时的平均精度,可以直观看出模型在改进之后,性能逐渐加强。从图 9(b)可以

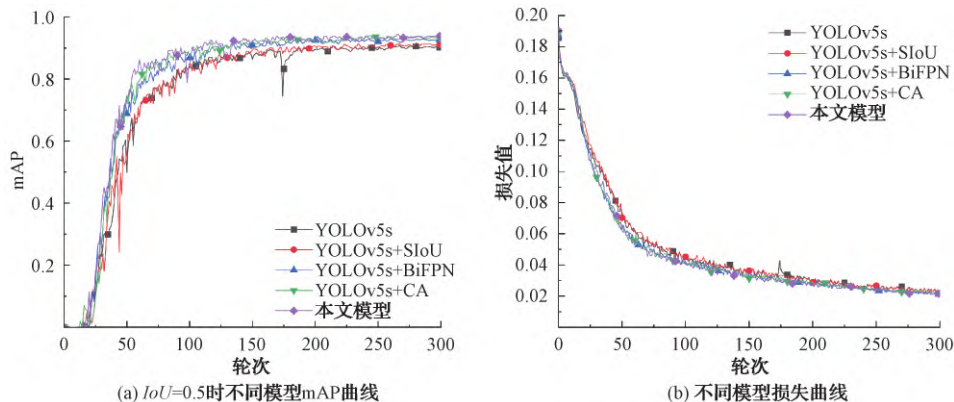


图9 添加各模块训练曲线

看出,在进行改进之后,模型的收敛速度加快,损失值越来越小,模型收敛能力加强。

3 结 论

本文提出了基于 YOLOv5s 的交通标志检测模型,为了解决不同光照情况下的小目标交通标志不易检测且存在错检等问题,本文在主干网络中引入了 CA 坐标注意力机制,将位置信息与通道信息相结合,提升了网络对感兴趣区域的获取能力;其次,在特征融合网络使用 BiFPN 替代 PANet 结构,提升了网络的特征融合能力;最后,使用 SiOU 替代 CiOU 解决了预测框与真实框之间的方向匹配问题,增加了网络的收敛性。

实验结果表明,主干网络中引入的坐标注意力机制性能好于其他主流注意力机制,网络中融合的 BiFPN 结构减少了特征的损失,充分保留了图像的位置信息,增强了对小目标的检测能力,与 YOLOv3、YOLOv5m、YOLOX、YOLOv6 等主流目标检测模型相比,本文模型的模型体积小、检测精度高、且推理速度较快,能够满足实时的交通标志检测。与原始模型 YOLOv5s 相比,本文模型在体积增加不大的情况下,将 mAP 值提升了 3.9%。本文模型在兼顾检测速度的同时提升了检测精度,改善了不同光照下小目标交通标志的检测性能,但是在复杂场景下仍然存在着漏检问题,未来将针对漏检问题做出改善。

参 考 文 献

- [1] LI H, SUN F, LIU L, et al. A novel traffic sign detection method via color segmentation and robust shape matching[J]. Neurocomputing, 2015, 169: 77-88.
- [2] BERKAYA S K, GUNDUZ H, OZSEN O, et al. On circular traffic sign detection and recognition[J]. Expert Systems with Applications, 2016, 48: 67-75.
- [3] ZAKLOUTA F, STANCIULESCU B. Real-time traffic sign recognition in three stages[J]. Robotics and Autonomous Systems, 2014, 62(1): 16-24.
- [4] 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.
- [5] REDMON J, FARHADI A. YOLO9000: Better, faster, stronger[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 7263-7271.
- [6] REDMON J, FARHADI A. YOLOv3: An incremental improvement[J]. Computer Science, 2018, arXiv:1804.02767.
- [7] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: Optimal speed and accuracy of object detection[J]. Computer Science, 2020, arXiv:2004.10934.
- [8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. European conference on computer Vision. Cham: Springer, 2016: 21-37.
- [9] XIONG C Z, WANG C, MA W X, et al. A traffic sign detection algorithm based on deep convolutional neural network[C]. 2016 IEEE International Conference on Signal and Image Processing (ICSIP), 2016: 676-679.
- [10] 张上,王恒涛,冉秀康.基于 YOLOv5 的轻量化交通标志检测方法[J].电子测量技术,2022,45(8):129-135.
- [11] 吕禾丰,陆华才.基于 YOLOv5 算法的交通标志识别技术研究[J].电子测量与仪器学报,2021,35(10):137-144.
- [12] KHAN J A, YEO D, SHIN H. New dark area sensitive tone mapping for deep learning based traffic sign recognition[J]. Sensors, 2018, 18(11): 3776.
- [13] 赵坤,刘立,孟宇,等.弱光照条件下交通标志检测与识别[J].工程科学学报,2020,42(8):1074-1084.
- [14] 李烨,顾晨峰.基于注意力机制的多尺度小目标交通标志检测[J].小型微型计算机系统,2022,43(2):381-386.
- [15] 郭璠,张泳祥,唐璠,等.YOLOv3-A:基于注意力机制的交通标志检测网络[J].通信学报,2021,42(1):87-99.
- [16] WANG J, CHEN Y, GAO M, et al. Improved YOLOv5 network for real-time multi-scale traffic sign detection[J]. Computer Science, 2021, arXiv:2112.08782.
- [17] 马丽萍,贫鑫,马文哲,等.基于改进 YOLOv3 模型的道路车辆多目标检测方法[J].西安工程大学学报,2021,35(5):64-73.
- [18] 安胜彪,姜慧儒,白宇.一种改进 YOLOv4 的半导体芯片表面字符识别算法[J].国外电子测量技术,2022,41(4):77-82.
- [19] 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,2021,42(10):171-178.
- [20] LIU T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2117-2125.
- [21] LIU S, QI L, QIN H, et al. Path aggregation network for instance segmentation[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8759-8768.

- [22] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 13713-13722.
- [23] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [24] JIANG B, LUO R, MAO J, et al. Acquisition of localization confidence for accurate object detection[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 784-799.
- [25] REZATOFIGHI H, TSOI N, GWAK J Y, et al. Generalized intersection over union: A metric and a loss for bounding box regression[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 658-666.
- [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] GEVORGYAN Z. SIoU Loss: More powerful learning for bounding box regression[J]. Computer Science, 2022, arXiv:2205.12740.
- [28] HOUBEN S, STALLKAMP J, SALMEN J, et al. Detection of traffic signs in real-world images: The German traffic sign detection benchmark[C]. International Joint Conference on Neural Networks, IEEE, 2013: 1-8.
- [29] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 7132-7141.
- [30] WANG Q L, WU B, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. Proceedings of the 2020 IEEE Conference on Computer Vision and Pattern Recognition, Piscataway: IEEE, 2020: 11531-11539.
- [31] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [32] GE Z, LIU S, WANG F, et al. YOLOX: Exceeding YOLO series in 2021[J]. Computer Science, 2021, arXiv:2107.08430.

作者简介

韩岩江, 硕士研究生, 主要研究方向为目标检测。

E-mail: 2621232368@qq.com

王伟, 博士, 副教授, 主要研究方向为网络信息安全、网络智能化应用。

E-mail: 17445643@qq.com

王峰萍, 博士, 讲师, 主要研究方向为图形图像处理、人工智能与模式识别、交通信息处理。

E-mail: 467202822@qq.com