

基于改进 YOLOv5 算法的钢材表面缺陷检测

燕雨, 荆宇超, 史孟翔, 杨朵

(郑州大学 电气与信息工程学院, 河南 郑州 450001)

摘要:为了解决钢铁缺陷检测效率低下和因误检造成的经济损失问题,提出了用于钢材缺陷检测的 YOLOv5 改进算法。在保持原 YOLOv5 检测层不变的情况下,新增加 3 条自适应权重辅助分支,用于提取 YOLOv5 网络的浅层信息,同时辅助分支也可以增强整体网络的梯度流动,使得训练效果更好;在网络的主干部分加入 EMA 注意力机制,经过 EMA 模块加权后的特征信息可以帮助模型更好地关注和理解重要的目标特征;使用 SIoU 代替了 CIoU 损失函数,SIoU 引入的角度损失和形状损失可以使锚框在回归过程更加快速准确,提高检测的稳定性和鲁棒性。通过对 NEU-DET 数据集的实验,所提的改进算法相比原 YOLOv5s 精确率提升了 3.7 个百分点,相比其他主流算法也拥有更好的检测性能。

关键词:YOLOv5; 自适应辅助分支; 注意力机制; 损失函数

中图分类号:TP391

文献标志码:A

doi: 10.13705/j.issn.1671-6833.2025.01.007

对钢材表面缺陷的识别主要包括 3 个功能:缺陷的检测、分类和定位^[1]。漏磁检测法、人工目测抽检法、涡流检测法、红外线检测法等都属于传统钢铁检测方法。漏磁检测法无法检测到一些对磁场影响不大的细小裂纹;人工目测抽检法效率较低,且长时间呆在检测车间会对工人身体健康产生不利影响;涡流检测法对钢材厚度有一定要求;由于钢铁自身对红外线的吸收特性存在差异,使用红外线检测方法可能会受到影响,从而导致无法对钢材的缺陷进行精确的分类^[2]。

基于深度学习的目标检测算法则表现出了比传统检测方法更加优良的性能。在深度学习用于目标检测的前期阶段,出现了很多卷积神经网络模型,如 AlexNet^[3]、VGG^[4]、GoogLeNet^[5]、ResNet^[6]、MobileNets^[7]等。经过不断的迭代更新,卷积神经网络模型发展出了多种目标检测算法。不同的神经网络模型结构和层次不同,可以适用于不同的数据和任务。根据算法的流程,可以将其分为两类:单阶段算法和两阶段算法。单阶段算法直接在输入的图像上进行预测,直接生成用于定位和分类的检测框,检测速度快,但是精度差,代表性的有 SSD^[8]和 YOLO^[9-11]系列算法。两阶段算法先生成候选框,再经过分类器

对候选框进行分类,检测速度慢,检测精度高,典型的两阶段算法有 Faster R-CNN^[12]和 R-FCN^[13]等。针对钢材缺陷检测存在的问题,阎馨等^[14]基于 SSD 算法,在 SSD 网络结构中使用 Transformer 多头注意力机制替换了原本的 Conv5_1 层,提高了小目标的检测精度的同时显著增加了参数量。院老虎等^[15]通过引入卷积块注意力机制(convolutional block attention module,CBAM)加强了特征提取能力,但是实时检测性能不强。

为了满足钢铁缺陷检测中实时性和精度的要求,降低因误检造成的损失,本文提出了改进的 YOLOv5-AES 算法。在 YOLOv5s 的基础上,加入了具有自适应功能的辅助分支、EMA 注意力机制(efficient multi-scale attention module)^[16]和 SIoU^[17]损失函数,增加了角度损失和更合理的距离损失,具有更快的收敛速度和更好的收敛效果。

1 YOLOv5 算法简介

YOLOv5 算法是经典的单阶段算法,可以直接生成目标框,具有检测速度快的特点,同时集成了众多改良策略,具有良好的综合检测性能。YOLOv5 网络首先经过 Focus 模块对输入图像进行切片和重

收稿日期:2024-05-25;修订日期:2024-06-14

基金项目:国家自然科学基金资助项目(62303424);中国博士后科学基金资助项目(2020M722871)

作者简介:燕雨(1994—),男,河南郑州人,郑州大学讲师,博士,从事计算机视觉研究,E-mail:yuyan@zzu.zzu.cn。

新排列,目的是增强感受野和特征表达能力,同时将大尺寸图片转化为拥有小尺寸和多通道数的特征图。YOLOv5 的主干网络是 Darknet 架构演化而来的 CSPDarknet,其特点就是在阶段连接处使用 CSP 模块进行卷积。CSP 模块有两个卷积分支并行迭代:一个进行深度卷积,一个只进行单次卷积。最后将两个卷积分支结果进行拼接,在增强特征的同时,有效地减小了参数量和计算量。CSP 模块通过不同阶段的信息交叉连接来提高信息流动效率。CSP-Darknet 还引入 SPP 模块,先通过一个标准卷积将通道数减小为原来的一半,然后对其分别进行 3 个核分别为 5、9、13 的池化操作,将 3 次池化结果和未池化数据进行连接,最后经过一个标准卷积将通道数降为 1/2,此时通道数和输入特征通道数相等以便进行后续操作。SPP 模块通过多尺度信息的引入来增强模型的特征表示能力。YOLOv5 算法的 Neck 部分采用了 FPN(feature pyramid net-

work)结构,通过从高到低和从低到高的方式来融合多尺度信息,最后在 3 个不同尺度的特征信息上进行预测。

在数据预处理方面,为防止数据集过少引起的过拟合,YOLOv5 提出了 Mosaic 数据增强技术,即将 4 张或 9 张图片拼接起来送入网络训练,扩充数据集的同时也提高了网络的鲁棒性和普适性。损失函数方面,YOLOv5 使用 CIoU 损失函数。

2 YOLOv5 改进方案

本文基于改进 YOLOv5 算法的钢材表面缺陷检测网络结构如图 1 所示。在原先 3 个预测层的基础上,添加了 3 个自适应辅助分支,用来提取不同尺度的信息。在 YOLOv5 的主干网络添加了 EMA 注意力机制,用来聚焦关键特征信息,减小背景信息的权重。损失函数使用了 SIoU,与原损失函数 CIoU 相比收敛速度更快、效果更好。

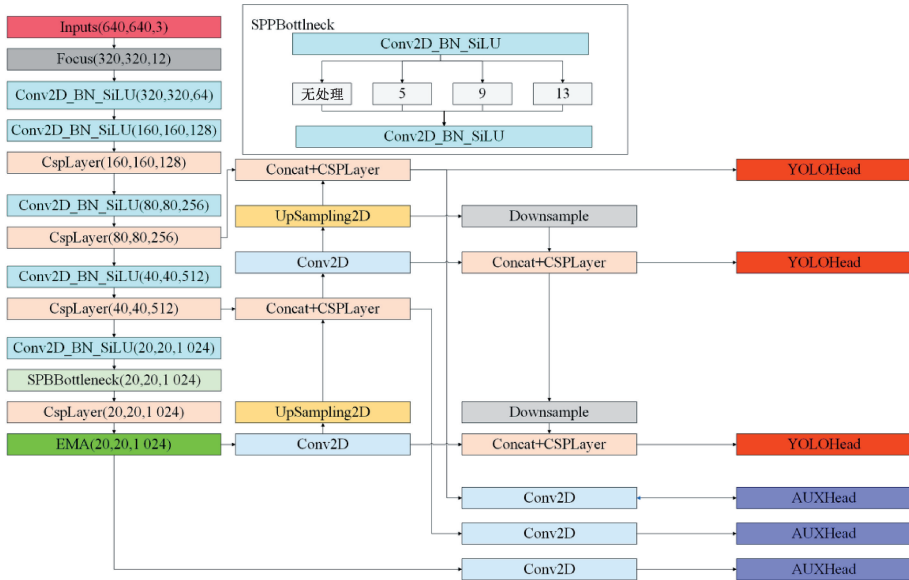


图 1 改进后的 YOLOv5 算法网络结构图

Figure 1 Improved YOLOv5 algorithm network structure diagram

2.1 自适应辅助分支

YOLOv5 网络通过 3 个预测层来进行目标回归,但是随着神经网络的深入,得到的特征矩阵包含的信息更多且抽象,往往忽视了包含底层特征的浅层信息。为了提取浅层信息,本文添加了 3 条和原 YOLOv5 预测层具有相同尺度的辅助分支,如图 2 所示。

辅助分支是提取浅层信息的常用手段,增加分支的反向传播还可以提高梯度的流动,有利于模型的回归迭代。辅助检测头的设计还可以使网络更好地处理目标在不同尺度和不同位置上的特征。它能够提供额外的监督信号,帮助网络更好

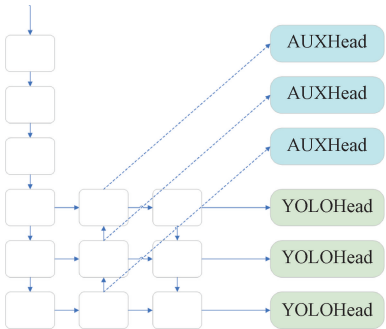


图 2 辅助分支示意图

Figure 2 Diagram of auxiliary branches

地学习不同目标的特征表示。同时,辅助检测头还可以加速网络的收敛,且得到的特征往往更加

本地化和精细化,具有较高的空间分辨率。但是浅层信息定位目标准确率较低,无法作为主要预测特征信息。因此,辅助分支往往给予较小的权重来确保辅助分支得到的无关信息不会影响主干的优先性,但是过小的权重会直接影响辅助分支的有效性。本文赋予辅助分支自适应的权重来保证最终模型结构的合理性:

$$P_i = m_i^2 / \left(\sum_{k=1}^3 m_k \right)^2 + \alpha; \quad (1)$$

$$Loss = \sum_{i=1}^3 Loss_i^{\text{lead}} + \sum_{i=1}^3 P_i \times Loss_i^{\text{aux}}. \quad (2)$$

式中: P_i 为第 i 个辅助分支的权重; m_i 为从第 i 个辅助分支所得到的特征信息上筛选出的正锚框的数量; α 为最小权重,本文设置为 0.05; $Loss$ 为总损失; $Loss_i^{\text{lead}}$ 为第 i 个主检测头的损失; $Loss_i^{\text{aux}}$ 为第 i 个辅助检测头的损失。3 个辅助头的尺寸和 3 个主检测头的尺寸分别相等,因此辅助头之间的重要程度也各不相同。辅助检测头以正锚框的数量来判断特征信息的重要程度,检测出的正锚框数量越多则分配其权重越多,同时降低检测正锚框数量较少的辅助头的权重,使得模型可以合理分配资源给对应的分支。自适应权重的辅助分支可以大大提高模型的泛化能力,更加关注那些难以分类的样本,且可以更好地适应样本不均衡问题,提高模型的鲁棒性。

2.2 EMA 注意力机制

在对钢材缺陷数据的深度学习中,随着 YOLOv5 网络的深化,会产生大量的冗余数据,大大降低了有效数据的优先性。注意力机制作为图像处理领域的关键技术,具有强化重要特征的功能,同时还具有随插随用的便捷性,因此引入了 EMA 注意力机制到 YOLOv5 网络中。

首个通道注意力机制 SE(squeeze and excitation module)通过对通道进行加权处理来增强有效特征的优先性,但是 SE 注意力机制只关注了通道特征,对空间特征没有进行有效利用。CBAM 注意力机制通过卷积来提取空间特征,但难以获取到远距离的依赖关系。CA 注意力机制(coordinate attention)为了解决这个问题,利用两个 1D 的池化操作,分别沿着垂直和水平方向进行特征的提取,具有这两个方向特征的特征图被单独编码,最后和输入特征相乘得到更加优质的输出特征。EMA 注意力机制对 CA 注意力机制进行了改进,添加了 1 个并行子网来捕捉多尺度特征关系,并且用跨空间学习方法来融合两个子网之间的特征信息。

如图 3 所示,从 CA 注意力机制中选取 1×1 卷积的共享分量,在 EMA 注意力机制中,把它称作 1×1 分支。为了聚合多尺度关系,添加了 3×3 卷积的并行子网,称为 3×3 分支。

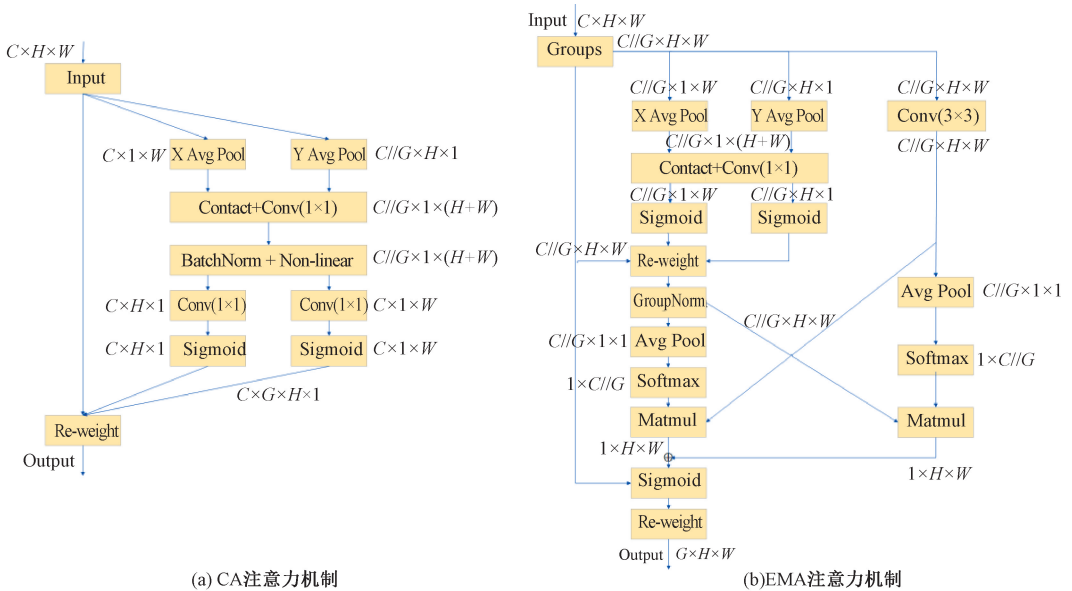


图 3 注意力机制结构图

Figure 3 Structure diagram of attention mechanism

对于输入特征 $X \in \mathbf{R}^{C \times H \times W}$, EMA 将其分成 G 个子特征 $X = [X_0, X_1, \dots, X_{G-1}]$, $X_i \in \mathbf{R}^{C // G \times H \times W}$ 。为了不失一般性,让 G 远小于 C , 本文基于先前的研究经验将 G 设定为 8。为了收集多尺度注意力权

重,将处理后的输入特征进行两个 1×1 分支和 1 个 3×3 分支的编码。在 1×1 分支的输出分解为两个向量后,采用两个非线性 Sigmoid 函数来拟合线性卷积的 2D 二态分布。然后利用 2D 全局池化来编码

全局空间信息:

$$z_c = \frac{1}{H \times W} \sum_j^H \sum_i^W x_c(i, j)。 \quad (3)$$

为了高效计算,在 2D 池化后通过非线性函数 Softmax 来拟合,得到的输出与并行的 3×3 分支进行矩阵点积运算,就取得了第一个空间注意力矩阵。此外 3×3 分支进行与 1×1 分支相同的池化和 Softmax 操作来对全局信息编码,并且 1×1 分支在通道特征的联合激活机制之前被变换为对应的维度形状,点积运算之后就得到了第 2 个空间注意力矩阵。两个矩阵相加,最后通过 Sigmoid 函数形成注意力权重矩阵。输入特征与权重矩阵相乘即可得到经过 EMA 注意力机制优化过后的输出特征。

2.3 SIoU 损失函数

损失函数是深度学习回归的重要参数,是对预期结果准确性的一种衡量指标。YOLOv5 中使用 CIoU 作为损失函数。CIoU 在 DIoU 的基础上引入了宽高比进行回归,解决了在真实框和预测框的中心重叠时 DIoU 等同与 IoU 的问题。但引入的宽高比具有一定的不合理性,同时无法考虑到每次的回归方向,导致回归速度缓慢。SIoU 由此提出了角度损失、距离损失和形状损失。其中,角度损失为

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

$$x = \frac{c_h}{\sigma} = \sin \alpha; \quad (5)$$

$$\sigma = \sqrt{(b_{c_x}^{gt} - b_{c_x})^2 + (b_{c_y}^{gt} - b_{c_y})^2}; \quad (6)$$

$$c_h = \max(b_{c_y}^{gt}, b_{c_y}) - \min(b_{c_y}^{gt}, b_{c_y})。 \quad (7)$$

式中: Δ 为角度损失; b 为预测框; b^{gt} 为目标框; b_{c_x} 为框中心点的 x 坐标; b_{c_y} 为框中心点的 y 坐标; σ 为目标框和预测框中心点的距离。

如图 4 所示,SIoU 的角度损失模型先将预测框带到最接近的坐标轴,然后沿着该坐标轴继续进行回归。为了达到这一目的,在回归的过程中要使 α 最小化,如果 α 已经小于 $\pi/4$,则对 β 进行最小化。

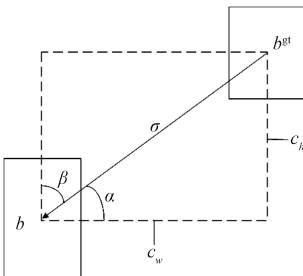


图 4 锚框回归示意图

Figure 4 Schematic diagram of anchor frame regression

考虑到新增加的角度损失,SIoU 重新定义了距离损失:

$$\Delta = \sum_{t=x,y} (1 - e^{-\gamma p_t}); \quad (8)$$

$$p_x = \left(\frac{b_{c_x}^{gt} - b_{c_x}}{c_w} \right)^2; \quad (9)$$

$$p_y = \left(\frac{b_{c_y}^{gt} - b_{c_y}}{c_h} \right)^2; \quad (10)$$

$$\gamma = 2 - \Delta。 \quad (11)$$

式中: Δ 为距离损失; c_w 和 c_h 分别为由两框中心点确立的矩形的宽和高。可以看出,当 α 趋近于 0 时,距离损失显著减小,当 α 等于 $\pi/4$ 时,距离损失达到最大,拥有最大优先级。

形状损失为

$$\Omega = \sum_{t=w,h} (1 - e^{-\theta})^\theta。 \quad (12)$$

$$\begin{cases} \omega_w = \frac{|w - w^{gt}|}{\max(w, w^{gt})}; \\ \omega_h = \frac{|h - h^{gt}|}{\max(h, h^{gt})}。 \end{cases} \quad (13)$$

式中: Ω 为形状损失成本; θ 为对形状的关注程度,当 θ 设置为 1 时,会优先关注形状损失,从而减小了距离损失和角度损失,因此定义 θ 为 2~6,根据实验和研究分析发现将 θ 设定为 4 效果最好。最终得到 SIoU 损失函数 $Loss_{SIoU}$ 为

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

3 实验结果及分析

3.1 数据来源

本文使用的数据集为 NEU-DET 公开数据集,是由东北大学发布的钢材表面缺陷数据集,包括裂纹(crazing)、夹杂(inclusion)、斑块(patch)、麻点(pitted_surface)、压入氧化铁皮(rolled-in scale)以及划痕(scratches)。每种缺陷包含尺寸为 200×200 像素的图片 300 张。由于裂纹和压入氧化铁皮缺陷图像模糊且包含少量标注错误,会对检测任务造成负面影响,因此删除这两类。最终将训练集、验证集和测试集以 7:2:1 的比例进行实验。

3.2 实验环境及设备

本文使用 Python 作为编程语言,初始学习率为 0.01,权重衰减为 0.000 5,epoch 设置为 800,batch-size 设置为 8,使用的 GPU 为 3050Ti-4 G,CPU 为 I5-12500 H。

3.3 实验评价指标

为衡量本次实验的有效性,本文使用了精确率 P 和召回率 R ,并使用精确率和召回率的 P - R 曲线计算平均精度 mAP 来对目标检测模型进行整体评估。精确率体现了模型对物体预测的正确程度;召回率体现了模型对找全物体的全面程度;平均精度是用来衡量各个类检测好坏结果的平均值。具体公式如下所示:

$$P = \frac{TP}{TP + FP}; \tag{15}$$

$$R = \frac{TP}{TP + FN} = \frac{TP}{GT}。 \tag{16}$$

$$mAP = \frac{1}{k} \sum_{k=1}^n AP_k。 \tag{17}$$

$$AP = \int_0^1 P(R) dR; \tag{18}$$

式中: TP 为模型预测为正例且预测正确的数量; FP 为模型预测为正例且预测错误的数量; FN 为模型预测为反例且预测错误的数量; k 为类别数量; AP_k 为第 k 个类别分别以精确率和召回率为纵轴和横轴组成的 P - R 曲线的线下面积。本文采用 $mAP@0.5:0.95$,表示在 IoU 阈值为 0.5 的情况下,以 0.05 的步长从 0.5 到 0.95 下的 mAP 的平均值。

3.4 实验结果

本文实验模型在原 YOLOv5 的基础上进行了网络结构的改进。为保证模型对比的公平性,所有模型均未载入预训练权重。为了验证改进措施的有效性,本文设置了 3 组消融实验,具体如表 1 所示。

表 1 改进算法消融实验

Table 1 Improved algorithmic ablation experiments					
AUX	EMA	SIoU	P	R	$mAP@0.5:0.95$
			0.834	0.825	0.490
✓			0.854	0.802	0.492
✓	✓		0.861	0.812	0.493
✓	✓	✓	0.871	0.810	0.494

由表 1 可知,加入自适应辅助分支后,精确率提高了 2 百分点,说明添加辅助分支可以有效提取浅层信息,提高梯度流动。在主干网络引入 EMA 注意力机制在精确率上增加了 0.7 百分点,召回率上提升了 1 百分点,能有效帮助模型处理目标信息,减少无关信息的干扰以提高模型性能。使用 SIoU 替换原 YOLOv5 的 CIoU 损失函数后,提高了 1 百分点的精确率,说明 SIoU 在对锚框的回归中可以应对更加复杂情况的同时,回归效果也更加优异。

为了验证自适应权重的辅助分支相比固定权重的辅助分支拥有更加优良的性能,本文以 0.05 的步长从 0.05 到 0.3 对不同权重的辅助分支进行了实验,如表 2 所示。可以看出,当辅助分支固定权重时,太低或太高都会使得效果不理想:权重太低会使 3 条辅助分支中最优异的分支难以发挥作用;权重太高会使辅助分支提取到的无关信息占用了大量资源,当权重为 0.15 时,精确率达到最高的 0.85。而提出的自适应辅助分支可以帮助辅助分支确定更加适合的权重,使得 3 条辅助分支发挥出对应的作用且减少了对无关信息的关注。

表 2 不同权重辅助头实验

Table 2 Experiments with different weighted auxiliary heads			
模型	P	R	$mAP@0.5:0.95$
YOLOv5s	0.834	0.825	0.490
AUX-0.05	0.841	0.801	0.488
AUX-0.10	0.836	0.821	0.490
AUX-0.15	0.850	0.808	0.490
AUX-0.20	0.834	0.808	0.485
AUX-0.25	0.846	0.805	0.491
AUX-0.30	0.836	0.830	0.490
AUX-self	0.854	0.802	0.492

图 5 为 YOLOv5 和本文算法的检测效果图,检测框使用不同颜色进行标注,检测框的左上方为其检测类别和置信度。从图 5 可以看出,本文算法相对于 YOLOv5 算法,在大部分情况下检测效果都更加优异。

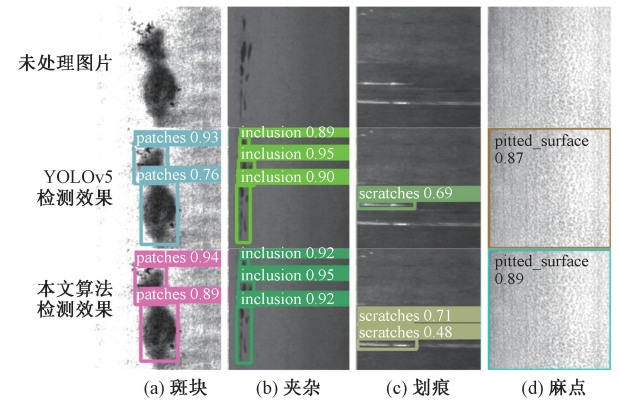


图 5 检测效果图

Figure 5 Detection renderings

由表 3 可知,YOLOv7、YOLOv8 和 SSD 的精确率均逊于本文的 YOLOv5-AES 算法;YOLOv5 和 YOLOv7-tiny 算法的计算量虽小于本文算法,但是用一部分计算量来换取精确率在实际应用中是更加有效的;Faster R-CNN 的召回率最高,但其较低的精

确率和庞大的计算量难以应用在实际设备中。

表 3 对比实验

Table 3 Comparative experiments

模型	<i>P</i>	<i>R</i>	<i>mAP@0.5:0.95</i>	计算量/ GFLOPs
YOLOv5s	0.834	0.825	0.490	16.5
YOLOv7-tiny	0.820	0.754	0.458	13.2
YOLOv7	0.836	0.763	0.483	105.2
YOLOv8	0.773	0.675	0.430	28.7
YOLOv5-AES	0.871	0.810	0.494	22.0
SSD	0.849	0.689		1.5
Faster R-CNN	0.590	0.890		369.8

4 结论

本文提出的 YOLOv5 改进方法增加了 3 条自适应权重的辅助分支,可以有效提取浅层信息,还可以大大提高梯度的流动;在主干网络引入了 EMA 注意力机制,对输入信息进行加权处理,使得模型更加关注有效信息,减小了无关信息的干扰;使用 SIoU 替换原 YOLOv5 的 CIoU 损失函数,在锚框回归过程中可以应对更加复杂的情况,回归效果更好。本文提出的改进算法相比原 YOLOv5s 精确率提升了 3.7 百分点,相较于 YOLOv7、YOLOv8 等有一定优势。在下一步工作中可以考虑对模型轻量化处理,提高检测速度,使其能在移动平台上部署。

参考文献:

[1] WEN X, SHAN J, HE Y, et al. Steel surface defect recognition: a survey[J]. Coatings, 2023, 13(1):1-30.

[2] 刘杰. 探究钢铁表面缺陷检测的研究与实现[J]. 中国设备工程, 2023(6): 172-174.

LIU J. Research and implementation of steel surface defect detection[J]. China Plant Engineering, 2023(6): 172-174.

[3] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.

[4] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [EB/OL]. (2015-04-10) [2024-04-19]. <https://arxiv.org/abs/1409.1556>.

[5] SZEGEDY C, LIU W, JIA Y Q, et al. Going deeper with convolutions [C]//2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2015: 1-9.

[6] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition

(CVPR). Piscataway: IEEE, 2016: 770-778.

[7] HOWARD A G, ZHU M L, CHEN B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications [EB/OL]. (2017-04-19) [2024-04-19]. <http://arxiv.org/abs/1704.04861>.

[8] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//14th European Conference on Computer Vision. Cham: Springer, 2016: 21-37.

[9] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08) [2024-04-19]. <http://arxiv.org/abs/1804.02767>.

[10] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2017: 6517-6525.

[11] 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). Piscataway: IEEE, 2023: 7464-7475.

[12] REN S Q, HE K M, 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, 2017, 39(6): 1137-1149.

[13] DAI J F, LI Y, HE K M, et al. R-FCN: object detection via region-based fully convolutional networks[EB/OL]. (2023-12-11) [2024-04-19]. <http://arxiv.org/abs/1605.06409>.

[14] 阎馨, 杨月川, 屠乃威. 基于改进 SSD 的钢材表面缺陷检测[J]. 现代制造工程, 2023(5): 112-120.

YAN X, YANG Y C, TU N W. Steel surface defect detection based on improved SSD[J]. Modern Manufacturing Engineering, 2023(5): 112-120.

[15] 院老虎, 常玉坤, 刘家夫. 基于改进 YOLOv5s 的雾天场景车辆检测方法[J]. 郑州大学学报(工学版), 2023, 44(3): 35-41.

YUAN L H, CHANG Y K, LIU J F. Vehicle detection method based on improved YOLOv5s in foggy scene[J]. Journal of Zhengzhou University (Engineering Science), 2023, 44(3): 35-41.

[16] OUYANG D L, HE S, ZHANG G Z, 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). Piscataway: IEEE, 2023: 1-5.

[17] GEVORGYAN Z. SIoU loss: more powerful learning for bounding box regression [EB/OL]. (2022-05-25) [2024-04-19]. <https://arxiv.org/abs/2205.12740>.

Steel Surface Defect Detection Based on Improved YOLOv5 Algorithm

YAN Yu¹, JING Yuchao¹, SHI Mengxiang¹, YANG Duo¹

(School of Electrical and Information Engineering, Zhengzhou University, Zhengzhou 450001, China)

Abstract: In order to solve the problem of low efficiency of steel defect detection and economic loss caused by false detection, an improved YOLOv5 algorithm was proposed for steel defect detection. Under the condition of keeping the original YOLOv5 detection layer unchanged, the improved algorithm added three auxiliary branches with adaptive weights to extract the shallow information of the YOLOv5 network, and the auxiliary branches could also enhance the gradient flow of the whole network, which made the training effect better. The EMA attention mechanism was added to the main part of the network, and the weighted feature information of the EMA module could help the model better focus on and understand the important target features. SIoU was used instead of the CIoU loss function, and the angle loss and shape loss introduced by SIoU could make the anchor frame more fast and accurate in the regression process to improve the stability and robustness of the detection. Through experiments on the NEU-DET dataset, the proposed algorithm had an accuracy improvement of 3.7 percentage points compared with the original YOLOv5s, and had better detection performance than other mainstream algorithms.

Keywords: YOLOv5; adaptive auxiliary branches; attention mechanism; loss function