文章编号:1674-2974(2025)04-0135-14 DOI:10.16339/j.cnki.hdxbzkb.2025273

基于位置感应卷积与注意力机制的钢材缺陷检测

解妙霞[1],程照中[1†],李嘉乐[2],李玲[1],贺宁[1]
(1. 西安建筑科技大学 机电工程学院,陕西 西安 710055;
2. 西安建筑科技大学 信息与控制学院,陕西 西安 710055)

摘 要:为了提高钢材缺陷检测精度,提出一种基于YOLOv5s的缺陷检测算法YOLOv5s-FNCE.首先,在骨干特征提取网络中加入新型NAMAttention注意力机制,提高对目标的感知和区分能力;并提出新型的C3-Faster,通过减小内存访问和冗余计算更有效地提取特征;在特征融合网络和输出端引入位置卷积CoordConvs,增强目标的语义感知能力和全局感知能力;最后,引入新的损失函数Focal-EIoU,以加快收敛速度,提高回归精度.实验结果表明,YOLOv5s-FNCE算法在钢材表面缺陷数据集上的均值平均精度达到了75.1%,比原始YOLOv5s提高了1.7个百分点,检测速度则提升了20.5%,证明了该算法在钢材缺陷检测中能够有效提升检测速度和检测精度.

关键词:目标检测;YOLOv5;位置感应;损失函数;注意力机制;钢材缺陷
中图分类号:TP391.41;TG146 **文献标志码**:A

Steel Defect Detection Based on Position-sensitive Convolution and Attention Mechanisms

XIE Miaoxia[1], CHENG Zhaozhong[1†], LI Jiale[2], LI Ling[1], HE Ning[1]
(1. School of Electromechanical Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China;
2. School of Information and Control, Xi'an University of Architecture and Technology, Xi'an 710055, China)

Abstract: To improve the accuracy of steel defect detection, a defect detection algorithm YOLOv5s-FNCE based on YOLOv5s is proposed. Firstly, a novel NAMAttention attention mechanism is added to the backbone feature extraction network to improve the perception and differentiation of the target; and a new C3-Faster is proposed to extract the features; the positional convolutional CoordConvs is introduced in the feature fusion network and at the output to enhance the semantic perception ability and global perception ability of the target; and finally, a new loss function Focal-EIoU is introduced to accelerate the convergence speed and improve the regression accuracy. Experimental results show that the mean average accuracy of the YOLOv5s-FNCE algorithm on the steel surface defects dataset reaches 75.1%, which is 1.7% higher than that of the original YOLOv5s, the detection speed is increased by 20.5%, which proves that the algorithm can effectively improve the detection speed and accuracy in

* **收稿日期**:2024-04-25
基金项目:陕西省重点研发计划项目(2024GX-YBXM-178, 2022NY—094), Key R&D Program Projects of Shaanxi Province(2024GX-YBXM-178, 2022NY—094)
作者简介:解妙霞(1977—),女,陕西韩城人,西安建筑科技大学教授,博士
† 通信联系人,E-mail:632530936@qq.com

steel defect detection.

Key words: target detection; YOLOv5; position sensor; loss function; attention mechanism; steel defect

在钢材的检测中,常用的检测手段有频闪检测法[1]、红外检测法[2]、漏磁检测[3]、涡流检测[4]以及激光扫描.也可用 CCD 相机采集影像资料,借助电脑视觉及讯号处理程式,对瑕疵相关信息进行预估与撷取.以上几种检测手段均存在着各自的不足,无法适应现代高通量检测的需求(目标小、特征分离困难、环境变量复杂).研究开发高精度、高效率的自动测试技术,是实现钢铁工业自动化的一个重要环节.

近年来,以深度学习为代表的机器视觉技术的快速发展,为钢材的缺陷检测提供了全新的技术手段,其主要分为两类:

一种是Two-Stage两阶段对象探测方法,常见的有 R-CNN[5]、SPPNet[6]、快速 R-CNN[7]等.该方法在扩展性、稳定性和准确性方面都优于常规方法,但是其运算时间长,一般要首先产生可探测的图像,然后再经过几次的融合、镶嵌等运算.由于设计的复杂性,需要反复地进行特性计算,所以对于钢材结构来说是不合适的.另一种单阶段检测方法(One-Stage)是采用深度卷积神经网络(DCNN)对物体进行定位与分类,该方法能够直接获取物体的空间坐标,并一次性生成物体类别的预测概率.当前,常用的One-Stage 单阶段探测方法有 YOLO[8-9]、SSD[10]以及 FSSD[10]三种,而YOLO具有快速、简单、高效、不需要过度参考等优点.

目前,很多学者采用基于深度学习的目标检测算法进行钢材缺陷检测.针对易遗漏的低维靶,程婧怡等[11]对 YOLOv3 增加 104×104 特征层,利用 DIoU 损失函数及 K-Means++聚类方法来提升对低维靶的探测准确率.沈希忠等[12]对 YOLOv5 进行了改进,提出了一种新的定位坐标系算法,用于微小损伤的探测,扩展了特征信息的抽取范围,并有效克服了微小损伤的探测困难.沈春光等[13]将SDD网状结构应用于钢带裂纹的诊断中,获得了87.8%的mAP值.王艳玲等[14]将SSD方法应用于钢带裂纹的诊断,并在NEU-CLS上实现了40FPS的检出速度和85.58%的mAP.上述研究虽然在一定程度上针对不同的场景进行了优化,但仍有较大的优化空间,主要体现在两个方面.一方面,在钢材生产环境中需要满足实时检测的要求,这就对检测速度提出了更高要求.另一方面,钢材表面的缺陷类型较多,且存在形状不规则、不规律、界限不明确等表面特征,这就使得目标位置识别困难,且背景干扰较多,亟需一套能够同时提高目标类型、位置和界限检测精度的缺陷检测算法.

YOLO 算法系列已经迭代到了 YOLOX[15],但对于钢材缺陷检测而言,YOLOv5 网络[16]相对于 YOLOv3[17]更加稳定、更加精确、更加快速,相比于 YOLOv4[18],拥有更低的复杂度和更快的速度,相对于 YOLOv7[19]和 YOLOX 具有更快的实时响应和更少的资源消耗.综上所述,本文选取了检测速度快、计算模型小的 YOLOv5 作为基网络,并在此基础上进行了改进.本文的主要工作如下:

为解决在复杂背景下特征难提取问题,在骨干特征提取网络 Backbone 中加入新型的 NAMAttention[20]注意力机制,以区分不同通道和空间之间的重要性,减少无关信息的干扰,提高模型对目标特征的感知和区分的能力.

原有模型含有大量卷积参数导致计算量烦琐复杂和整体计算速度缓慢,针对这个问题本文基于 FasterNet[21]思想提出新型的 C3-Faster,以替换原有的 C3 模块,通过减小内存访问和冗余计算进行更高效的提取特征.

原有的网络对于浅层特征的融合不是很充分,对位置信息的感应匮乏,导致小目标缺陷会被漏检错检,比如细小裂纹或麻点等小目标缺陷.本文在特征融合网络(Bottleneck)和输出端(head)部分中引入位置卷积(CoordConvs)[22],可以对图像信息进行多通道处理,重新聚类进一步增强小目标缺陷的语义感知能力和全局感知能力,提升检测精度.

本文在 Focal[23]损失函数的思想上引入新的损失函数 Focal-EIoU,替换了 YOLOv5 原有的 CIoU[24]损失函数,解决了目标边界框回归的过程中训练样本不平衡的问题,加快了收敛速度,提高了回归精度.

1　YOLOv5 目标检测算法

1.1　YOLOv5 网络概述

YOLO 算法的基本思路是把物体探测问题转换成一种回归问题,利用卷积神经网络对被探测物体进行分类概率和待探测物体的位置坐标进行预测.

YOLOv5 架构由四大模块构成:输入端(Input)、骨架特征提取网络(Backbone)、特征融合网络(Bottleneck)、预测报头(Prediction).该算法在输入端使用 Mosaic 数据增强技术,通过随机缩放、随机裁剪和随机排列等方法,在增加数据集复杂性的同时,能够添加更多的小型物体,从而提高训练结果的鲁棒性.本项目拟采用 Focus 结构与 CSPDarknet53 相结合的方法,以 CSPDarknet[24] 为核心,通过残差网络(Residual),实现特征图谱的快速降维,在保证检测准确率的前提下,提高检测速度与特征抽取能力.在特征融合方面,拟利用空间棱锥池(SPP)与路径聚集(PANet),借鉴 YOLOv4 自上而下与自下而上的多尺度特征融合方法.在输出端,利用 CIoU-Loss 损耗函数,在 80×80,40×40,20×20 等三种输出尺度下,分别对大、小目标进行分类与定位预报;最后,利用 DIou-NMS 剔除冗余信息,得到具有最大置信度的待预测对象.

YOLOv5 根据修改网络深度和宽度两个缩放的参数,可以得到五个不同大小的模型,分别是:YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x,其模型大小和精度也随之递增.考虑到便于实验,本文采用了检测速度和精度兼顾的 YOLOv5s 作为改进的基础模型.

1.2　YOLOv5s 网络结构

YOLOv5s 的网络结构如图 1 所示.其中 k 表示卷

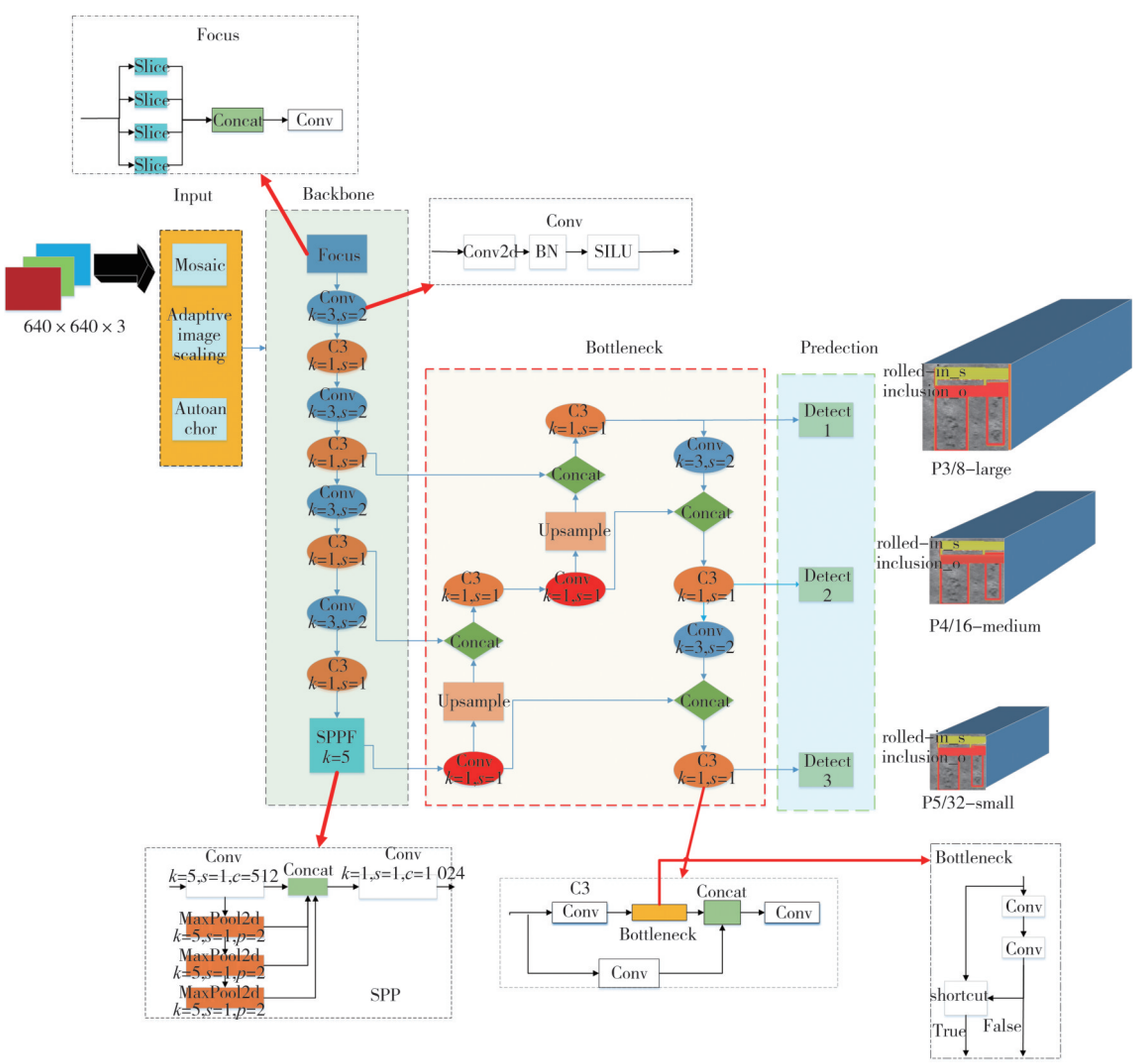


图 1　YOLOv5s 网络结构图

Fig.1　YOLOv5s network structure

积核尺寸,s表示卷积步长,p表示填充大小,n表示模块重复次数.

2 模型改进

2.1 改进网络概述

本文是在YOLOv5的基础上提出了改进网络,首先引入规范化注意力机制减少背景干扰,然后引入C3-Faster来替代原网络中的C3模块,减小计算量和内存访问,最后使用位置卷积CoordConvs和损失函数Focal-EIou来提高目标提取速度和收敛精度.结合以上改进关键点,将改进之后的网络命名YOLOv5-FNCE(Faster+NAMAttention+CoordConvs+Focal-EIou),改进后的模型网络结构如图2所示,其中k、s、p、n与上文含义一致.

2.2 NAMAttention规范化注意力机制模块

Hu等[25]在2018年提出通道注意力模块SENet,其重点是获得每一个通道输入进来的特征层的权值,并对其进行池化,能够对不同的特征分配不同的权重.Hu等在SENet[25]的基础上提出了新的注意力机制模块CBAM[26],能够在空间和通道两个维度上同时调整捕捉的权重信息.这些方法成功利用了不同维度的交互信息,但是他们缺乏对权重影响因素的考虑,因此本文引入了基于规范化的注意力机制模块(normalization-based attention module, NAM),它可以抑制较不显著的权重,并利用权重稀疏惩罚,从而在保持性能的同时更加高效地计算.

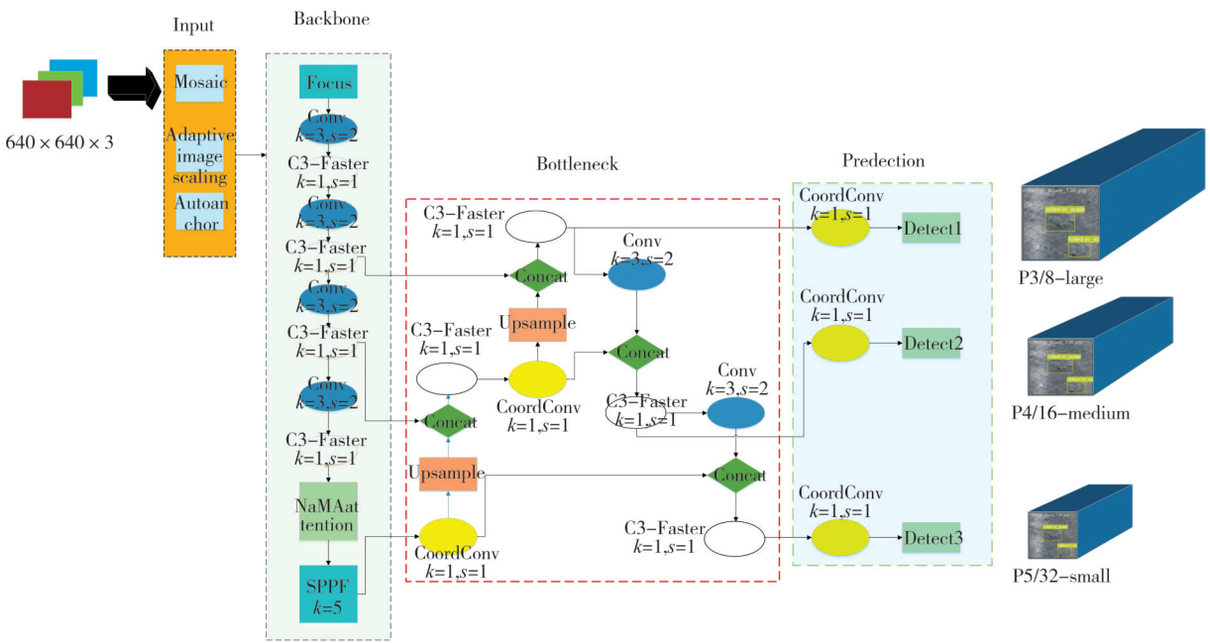


图2 YOLOv5-FNCE网络结构图

Fig.2 YOLOv5-FNCE network structure

NAM基于CBAM的注意力模块整合,对信道与空间的子注意力机制进行了重构,并在各块的末端分别嵌入NAM模块.针对残差网络,将此模型嵌入模型的末尾.通道模型采用准规范化(BN)缩放系数,缩放系数反映各通道的变化幅度,同时也反映信道的重要程度.比例系数就是网络的变异系数,变异系数越大,说明通道的改变程度越大,通道所蕴含的信息量和重要程度也就越高;而那些没有太大改变的渠道,则是单纯的,不重要的.如式(1)所示,比例因子测试通道的方差表示其重要性.

$$B_{\text{out}} = \text{BN}(B_{\text{in}}) = \gamma \frac{B_{\text{in}} - \mu_\beta}{\sqrt{\sigma_\beta^2 + \varepsilon}} + \beta \tag{1}$$

式中:μ_β和σ_β分别为均值和标准差,是可训练的仿射变换参数.

$$M_c = \text{Sigmoid}(W_\gamma(\text{BN}(F_1))) \tag{2}$$

通道注意子模块如图3和式(2)所示,其中M_c表示输出特征,γ为各通道的标度因子,权重为$W_\gamma = \gamma_i / \sum_{j=0} \gamma_j$.

此外对空间维度应用缩放因子BN来衡量像素的统一性,称为像素归一化,对应的空间注意子模块如图4和式(3)所示,其中输出记为M_S,λ为比例因子,权重为$W_\lambda = \lambda_i / \sum_{j=0} \lambda_j$.

$$M_S = \text{Sigmoid}(W_\lambda(\text{BN}_S(F_2))) \tag{3}$$

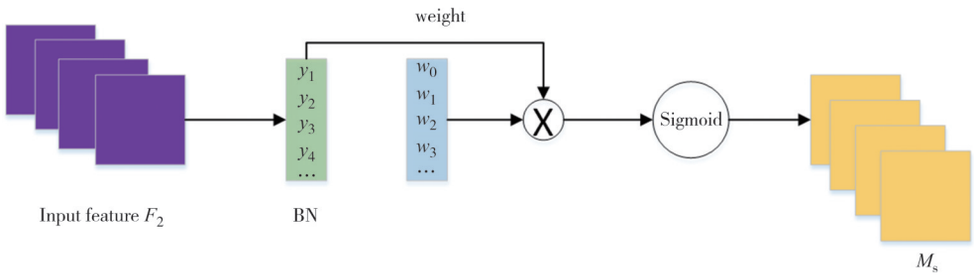


图 3　通道注意子模块

Fig.3　Channel attention submodule

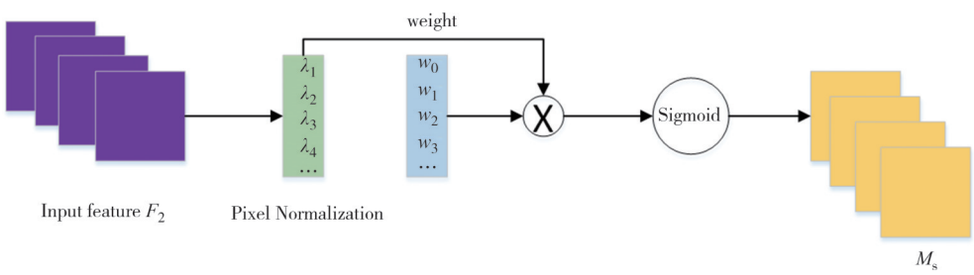


图 4　空间注意子模块

Fig.4　Spatial attention submodule

为了抑制不太显著的权重,在损失函数中引入了正则化项,如式(4)所示.

$$\text{Loss} = \sum_{(x,y)} l(f(x,W),y) + \rho \sum g(\gamma) + p \sum g(\lambda) \tag{4}$$

本文在特征提取网络的 SPP 模块之前加入了 NAM 注意力机制,有助于提升特征的表达能力,抑制无关信息,提高检测性能,并促进特征之间的有效融合.YOLOv5 中加入不同的注意力机制(如表 1 所示),发现 NAM 模块能够在不增加模型复杂度和浮点数计算量的情况下,显著提升检测精度.

表 1　YOLOv5s 加入三种不同注意力机制对比

Tab.1　Comparison of YOLOv5s with three different attention mechanisms

模型	参数量	浮点数	mAP@0.5
YOLOv5s	7 026 307	15.8	0.734
YOLOv5s-SE	7 059 075	15.8	0.732
YOLOv5s-CBAM	7 550 693	16.6	0.723
YOLOv5s-NAM	7 027 331	15.8	0.740

2.3　基于 FasterNet Block 的 C3 网络重构

YOLOv5 网络模型中包含了很多的 C3 模块,从而对输入的图像进行特征提取、特征融合等操作,但是由于传统的卷积中存在着大量冗余,包含着大量的网络参数,这消耗了大量的计算资源.为了提升模型计算速度,本文引入了部分卷积(partial convolution, PConv)来对原有的 C3 模块进行重构.普通卷积层的算法和计算量如式(5)和式(6)所示.

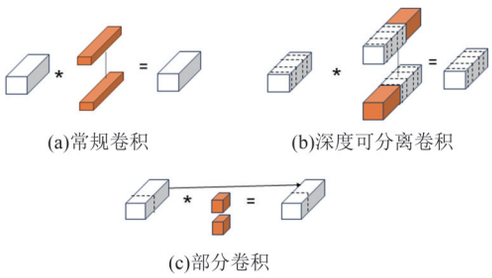
$$Y = X * f + b \tag{5}$$

$$\text{FLOPS} = n \times h' \times w' \times c \times k \times k \tag{6}$$

给定输入 $X \in R^{c \times h \times w}$,其中通道数量为 c、通道的高度为 h、通道宽度为 w,其中 * 代表的是卷积的运算,b 是偏差项,R 代表特征图,$Y \in R^{n \times h' \times w'}$ 则表示输出了通道数为 n 的特征图,并且输出的通道高度为 h'、通道宽度为 w',$f \in R^{c \times k \times k \times n}$ 表示大小为 $k \times k$ 的 $c \times n$ 个卷积操作.卷积神经网络的时间复杂度可以用浮点运算量(floating point of operations, FLOPS)表示,如式(6)所示,在卷积的过程中,卷积核的数量和通道数的数量很多,导致一般的 FLOPS 都非常大.

深度可分离卷积(depthwise separable convolution, DWConv)是指将卷积操作分解为深度卷积(depthwise convolution)和逐点卷积(pointwise convolution)两步的卷积操作.相对于传统卷积,深度可分离卷积具有计算量更小、模型参数更少、性能更优越的特点,它将传统卷积的计算量从 $h \times w \times k^2 \times c^2$ 缩小到了 $h \times w \times k^2 \times c$.虽然 DWConv 可以有效降低计算量 FLOPS,但是不能简单地替换普通卷积,因为这样做会导致精度严重下降.

PConv 卷积来自 FasterNet Block 部分,它受深度卷积的影响,目的是同时减小内存访问和计算冗余.其和常规卷积、DWConv 的区别如图 5 所示.



(a)常规卷积 (b)深度可分离卷积

(c)部分卷积

图5 卷积对比图

Fig.5 Convolution comparison plots

PConv对部分输入通道应用常规卷积来进行空间特征提取,而对其余的通道保持不变.对于连续或者规则的内存访问,则将第一个或最后一个连续的c_p通道作为整个特征图的代表进行计算,这样可以在不丧失一般性的情况下认为输入和输出特征图具有相同数量的通道.

因此,PConv的FLOPS:$h \times w \times k^2 \times c_p^2$,其中,$h$和$w$是特征图的高和宽,$k$是卷积核的大小,$c_p$是常规卷积作用的通道数;这相当于之前普通卷积的$c_{in}$变成了$c_p$,而在实际实现时一般$r = \frac{c_p}{c} = \frac{1}{4}$.所以,PConv的FLOPS只有常规卷积的1/16.而PConv的内存访问情况:$h \times w \times 2c_p + k^2 \times c_p^2 \approx h \times w \times 2c_p$,PConv的内存访问量仅仅为常规卷积的1/4,所以无须进行内存的访问.

本文以此为基础,使用FasterNet Block替代了YOLOv5s的原有C3卷积模块,并对其PConv进行重构,采用1×1卷积,轻量化的模型在提升检测精度的同时,计算量和参数量也大大减少.检验了FasterNet Block新模块对网络的整体运行速度的提升作用,结果如表2所示.

表2 YOLOv5s与YOLOv5-C3Faster对比

Tab.2 Comparison of YOLOv5s and YOLOv5-C3Faster

模型	参数量	浮点数	时间/ms
YOLOv5s	7 026 307	15.8	13.39
YOLOv5s-C3Faster	6 004 861	13.5	12.97

由表2可以看出,用C3Faster替换原来YOLOv5s的C3模块之后,参数量减少了14.54%,计算量减少了14.56%,推理速度加快了3.1%.实验结果表明,重构C3结构之后确实能够满足轻量化的需求,能够有效降低模型复杂程度和推理计算的成本,从而降低了对于硬件的需求,也能弥补加入注意力机制NAM和之后的CoordConv卷积模块所带来的参数量和计算量的提升.

2.4 基于位置感应卷积的特征提取改进

YOLOv5中含有很多传统的卷积,无法将空间表示转化为直角空间和像素空间的坐标.传统的卷积算法不能满足定位信息的要求,甚至在进行局部操作时,也不知道当前卷积核的具体位置;他只能感知到一小部分的信息,却不知道具体的信息在什么地方.卷积法为等变量,也就是说,在将各滤波器加到输入端来生成输出时,各滤波器的具体位置是未知的.然而,对于本文研究的钢铁表面缺陷,其缺陷尺度丰富,边缘形状复杂,单尺度下的缺陷往往会损失大量细节.因此可以通过让滤波器了解自己的位置来帮助卷积,为此在卷积输入中加入了两个通道:一个i坐标通道,一个j坐标通道.将由其产生的卷积称为CoordConv(位置感应卷积),如图6所示.

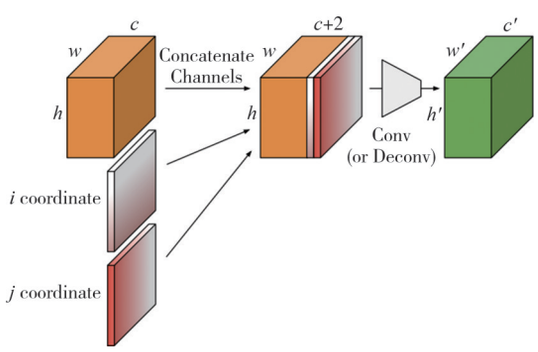


图6 卷积结构图

Fig.6 Convolutional structure diagram

CoordConv可以实现为标准卷积的简单扩展,其中实例化了额外的通道并填充了坐标信息,然后将它们沿通道连接到输入表示,并应用到了标准卷积层.具体来说,i坐标通道是一个$h \times w$的矩阵,其第一行填充是0,第二行填充为1,第三行填充为2,以此类推.j坐标通道类似,但是列填充值为常数值,行的填充值却不是常数值.然后对i和j坐标进行最终的线性缩放,以使它们落在[-1,1]的区间内.对于二维卷积,两个(i,j)坐标足以完全指定一个输入像素,但是如果需要,还可以添加坐标作为第三条通道,其表达式如式(7)所示.

$$r = \sqrt{(i - h/2)^2 + (j - w/2)^2} \tag{7}$$

因此首先在特征融合网络中引入了CoordConv(位置感应卷积),并将其加在了上采样之前,通过将局部特征和全局融合相结合,从不同的分辨率聚合信息来提高特征图表达能力,便于更好地提取特征.其次在输出端预测头之前,加入了三个CoordConv卷积,这样做的好处是可以通过感知坐标信息,对已经

特征融合的信息进行再次融合,有利于多尺度多目标的复杂物体检测.改进之后的 YOLOv5s+CoordConv 和 YOLOv5s 的计算结果如表 3 所示.

表 3　YOLOv5s 与 YOLOv5s+CoordConv 对比

Tab.3　Comparison of YOLOv5s and YOLOv5s+Coord Conv

模型	参数量	mAP@0.5	时间/ms
YOLOv5s	7 026 307	0.734	13.39
YOLOv5s+CoordConv	10 140 675	0.747	13.79

从表 3 可以看出,在原有网络模型中加入该卷积,虽然参数量大幅提升,但并不会导致模型检测速度的降低,还能够有效增强特征融合,从而提升检测的精度.

2.5　损失函数的改进

YOLOv5 损失函数主要由定位损失函数 CIoU-Loss、分类损失函数和目标置信度损失函数 BCE Loss 三者组成.其中 CIoU Loss 如式(8)所示,虽然考虑了边界框回归的重叠面积、中心点距离、高宽比,但是其公式中的 v 反映的是高宽的差异,而不是高宽与其置信度分别的差异.因此,有时会阻碍模型的优化相似性.

$$L_{\mathrm{CIoU}} = 1 - \mathrm{IoU} + \frac{\rho^{2(b,b^{\mathrm{gt}})}}{c^2} + av \tag{8}$$

v 关于 h 和 w 的梯度计算如下:

$$\frac{\partial v}{\partial w} = \frac{8}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}} - \arctan\frac{w}{h}\right)\frac{h}{w^2 + h^2} \tag{9}$$

$$\frac{\partial v}{\partial h} = -\frac{8}{\pi^2}\left(\arctan\frac{w^{\mathrm{gt}}}{h^{\mathrm{gt}}} - \arctan\frac{w}{h}\right)\frac{w}{w^2 + h^2} \tag{10}$$

针对该问题,在 CIoU 的基础上将高宽比拆开,从而提出了 EIoU Loss 损失函数.EIoU Loss 公式如式(11)所示,其中包含了 IoU 损失、距离损失、高度损失(重叠面积、中心点距离、高宽比)三个部分.同时,针对目标边界盒回归预测中存在的不均衡问题,也就是同一张图像中,具有较小回归误差的优质框架数目远远小于具有较大误差的劣质框架数目.

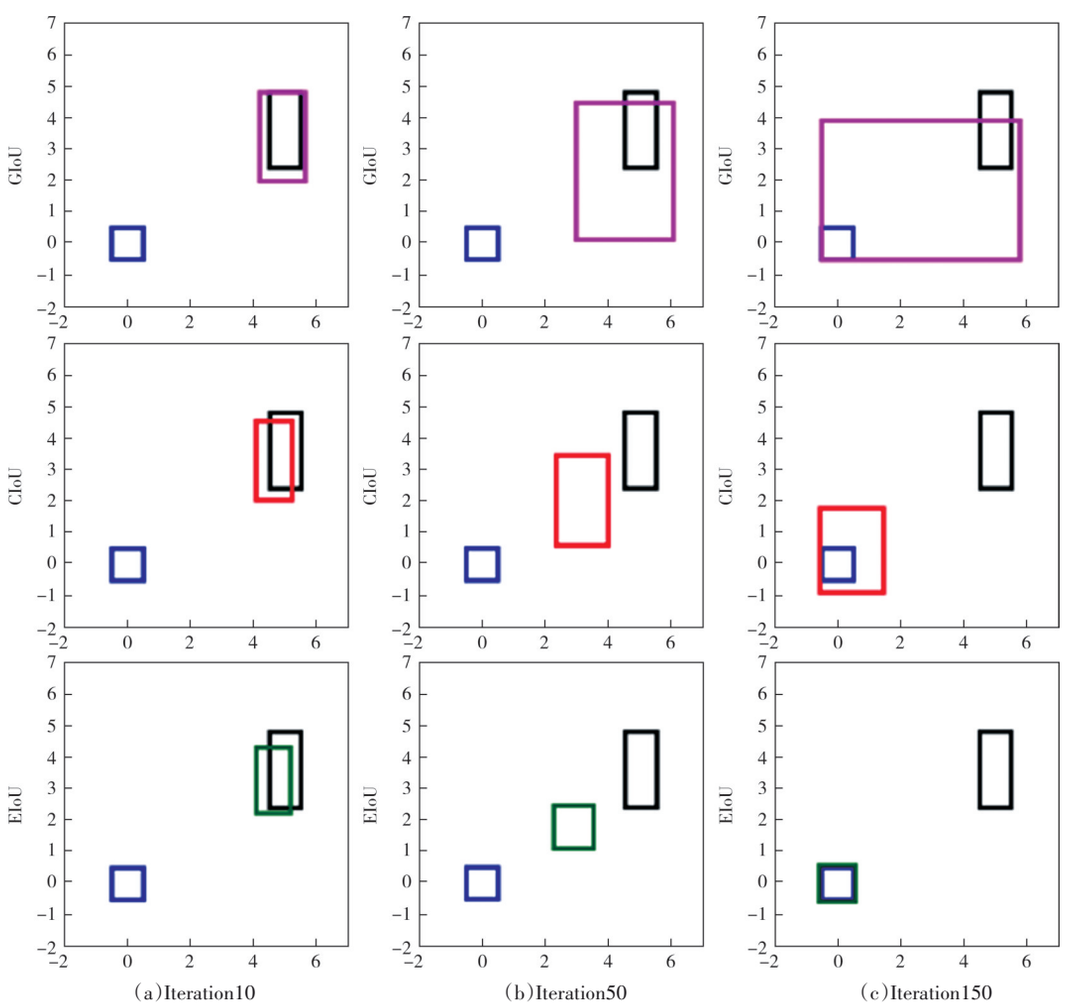


图 7　三种损失函数的对比

Fig.7　Comparison of three loss functions

因此,拟从梯度的视角,将优质锚框区与劣质锚框区分离,并在此基础上,通过权衡优质与劣质两种类型锚框架对损失函数的贡献,构建高品质与低品质锚框架的数学模型.因此,在不同质量的锚框架上,梯度的变化是适中的.为了均衡不同质量锚的作用,一般采用(0,1)的梯度函数.在增加一个超参数值后,尽管在很大程度上减少了困难样品的梯度系数,但是在高质量样品中,其梯度系数将下降,这是不理想的.可以用以下表达式对梯度系数进行(0,1)标准化:

$$g(x)=\frac{\partial L_{\mathrm{f}}}{\partial_x}= \begin{cases}-\alpha\ln(\beta), x>1;\ 1/\mathrm{e}\leqslant\beta\leqslant 1\\ -\alpha x\ln(\beta x), 0<x\leqslant 1;\ 1/\mathrm{e}\leqslant\beta\leqslant 1\end{cases} \tag{11}$$

得到能够对边界框进行回归监督的 Focal L1 Loss 如下:

$$L_{\mathrm{f}(x)}= \begin{cases}-\dfrac{\alpha x^2(2\ln(\beta x)-1)}{4}, 0<x\leqslant 1;\ 1/\mathrm{e}\leqslant\beta\leqslant 1\\ -\alpha ln(\beta)x+C, x>1;\ 1/\mathrm{e}\leqslant\beta\leqslant 1\end{cases} \tag{12}$$

结合 Focal Loss 提出了 Focal-EIoU Loss,公式如下所示:

$$L_{\mathrm{Focal-EIoU}}=\mathrm{IoU}^{\gamma}L_{\mathrm{EIoU}} \tag{13}$$

其中当γ=0.5时可以实现最佳的权衡,传统的 Focal Loss 更加注重训练困难的样本,越困难的样本损失越大.而 Focal-EIoU Loss 则是 IoU 越高的样本损失越小,相当于加权作用,有助于提高回归精度.

从表4可以看出,改进后的损失函数能够达到更高的收敛速度,CIoU 经过 195 轮 epoch 才依次收敛,而 Focal-EIoU 经过 117 轮 epoch 便可收敛,mAP@0.5 更是达到了 0.75,这说明采用 Focal-EIoU 能使得训练流程更加稳定和准确,提升了预测框的精准程度,加速了模型收敛.

表4 YOLOv5s不同损失函数对比

Tab.4 Comparison of different loss functions for YOLOv5s

模型	参数量	mAP@0.5	Convergence epoch
YOLOv5s+CIoU	7 026 307	0.734	195
YOLOv5s+EIoU	7 026 307	0.75	184
YOLOv5s+Focal-EIoU	7 026 307	0.75	117

3 实验结果与分析

为了检验本文提出的 YOLOv5s-FNCE 算法的效果,做以下钢材表面缺陷检测的实验.

3.1 数据集

本实验采用的是东北大学公开的钢材缺陷数据集,包含有常见的六种钢材表面缺陷,每个类别包含 300 张照片,总共 1 800 张.如图 8 所示,包括裂纹(crazing)、麻点(pitted surface)、划痕(scratches)、夹杂(inclusion)、斑块(patches)以及轧制氧化皮(rolled-in scale).以 7∶2∶1 的比例随机抽取划分为训练集、验证集和测试集.

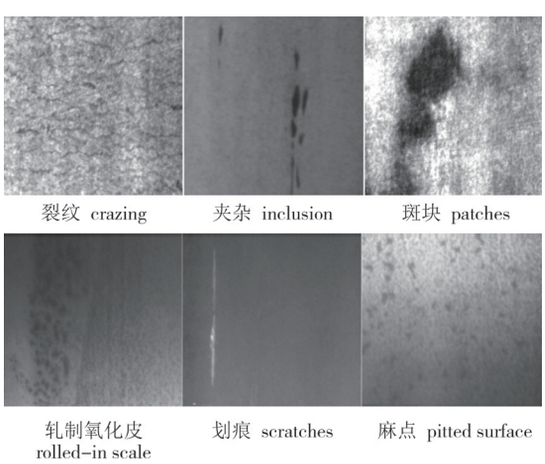


图 8 数据集图像

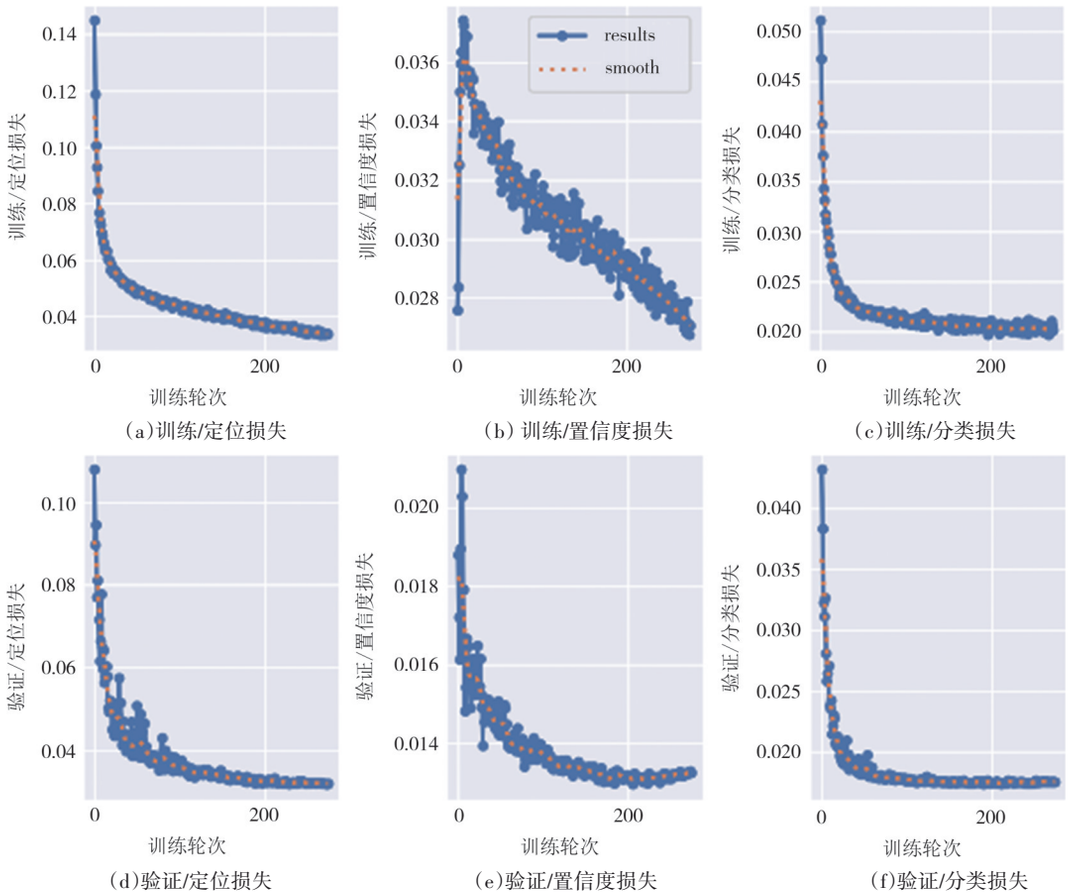
Fig.8 Dataset images

3.2 实验参数设定

实验环境如下:AMD Ryzen 9 7945HX with Radeon Graphics 高性能处理器,Nvidia GeForce GTX 4060 显卡,16 GB 内存,操作系统为 Windows11,Python 版本为 3.11.4,深度学习框架为 PyTorch2.1.1+CPU,优化器为 SGD,初始学习率为 0.01,权重衰减因子为 0.005,Batchsize 为 8,输入图像大小为 640×640,动量为 0.937.整个训练过程如图 9 所示,经过 275 轮迭代,分类损失(cls_loss)、模型回归框定位损失(box_loss)、置信度损失(obj_loss)都趋于收敛.

3.3 实验指标

本实验使用了平均均值精度 mAP(mean average precision)、检测帧率 FPS(frames per second)、召回率 Recall、计算量 FLOPS、精准率 Precision、参数量 Parameters 等作为评估模型的各项性能的指标.



(a)训练/定位损失　(b) 训练/置信度损失　(c)训练/分类损失

(d)验证/定位损失　(e)验证/置信度损失　(f)验证/分类损失

图 9　训练损失曲线

Fig.9　Training loss curve

精准率(Precision)是真正例(TP)占所有被模型预测为正例(TP+FP)的比例;召回率(Recall)是真正例(TP)占所有实际正例(TP+FN)的比例,其表达式如下:

TP(True Positive)为真正例,就是实际为正例并且被模型正确检测为正例的数量;FP(False Positive)为假正例,即实际为负例却被模型错误检测为正例的数量.

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{14}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{15}$$

FLOPS表示浮点运算次数,理解为计算量(计算时间复杂度),可以用来衡量算法的复杂度,常用做神经网络模型速度的间接衡量标准,GFLOP即每秒十亿次的浮点运算数;FLOPS的计算公式如下:

$$F_{\mathrm{FLOPS}} = h \times w\left[C_{\mathrm{in}} \times k^2 \times \left(C_{\mathrm{in}} \times k^2 - 1\right) + 1\right] \times C_{\mathrm{out}} \tag{16}$$

参数量(Parameters)指的是模型所包含的参数的数量,模型中使用到的卷积、全连接里面的权值矩阵对应的每一个数字,都是参数量的组成.对于卷积层,参数量的计算公式如下:

$$P_{\mathrm{Conv}} = (k_w \times k_h \times C_{\mathrm{in}} + 1) \times C_{\mathrm{out}} \tag{17}$$

式(16)、式(17)中:k_w和k_h分别是卷积层的宽和高;k为卷积核的大小;h和w分别为输入特征图的高和宽;C_{in}和C_{out}分别是输入和输出特征图的通道数.

以召回率R(Recall)作为横坐标,精准率P(Precision)作为纵坐标,在不同阈值下形成的曲线称为P-R曲线.由于召回率和精确率是一组矛盾的指标,需要找一个指标同时兼顾,所以以P-R曲线下方围成的面积即平均精确度(AP)作为新的指标,该指标数值越大说明精度越高.单类别的精确度可以使用积分的方式来计算;而多类别的精确度,可以通过分别求得各个类别单独的AP值,再除以类别总数m来计算得到均值平均精度mAP值.二者表达式如下所示:

$$\mathrm{AP}=\int_0^1 P(r)\mathrm{d}r \tag{18}$$

$$\mathrm{mAP}=\frac{\sum_{i=1}^{m}\mathrm{AP}_i}{m} \tag{19}$$

3.4 消融实验

为了验证改进算法对钢材表面缺陷检测的有效性,分别对各个改进模块进行了消融实验,实验结果如表5所示.

表5中"√"代表实验算法中使用了对应的模块,"×"代表算法中没有使用对应的模块.由表5可得第一组为原始YOLOv5s算法,其mAP值为73.4%;第二组在YOLOv5s的SPP模块之前加入注意力模块NAM,可以看到在对Parameters几乎没有什么影响的情况下,模型的mAP提升了0.6个百分点;第三组仅加入CoordConv模块时,模型的Parameters显著增多,但模型的mAP提升了1.3个百分点,进一步说明了位置卷积增强特征提取的可行性;第四组仅加入了C3-Faster模块,模型的参数量显著减少,mAP值提升了0.5个百分点,验证了该模块不但可以减少原有模型的内存访问和冗余计算,更能进一步减少加入位置卷积模块而造成参数量上涨的影响;第五组仅加入了Focal-EIoU损失函数时,模型参数量不变,mAP值提升了1.6个百分点;第六组为本文结合上述改进模块的优点,加入了各个改进模块后,召回率达到了71.7%,mAP值提升了1.7个百分点,对背景复杂,特征难提取的缺陷达到精准检测的效果,解决了微小目标错检漏检的问题.进一步验证了本文算法对钢材表面缺陷检测的有效性.

表5 消融实验结果表

Tab.5 Ablation experiment result graph

YOLOv5s	NAM	CoordConv	C3-Faster	Focal-EIoU	mAP/%	召回率/%	参数量	浮点数
√	×	×	×	×	73.4	71.2	7 026 307	15.8
√	√	×	×	×	74.0	70.9	7 027 331	15.8
√	×	√	×	×	74.7	70.6	10 140 675	21.5
√	×	×	√	×	73.9	71.9	5 795 779	12.7
√	×	×	×	√	75.0	72.4	7 026 307	15.8
√	√	√	√	√	75.1	71.7	9 267 523	18.7

3.5 与主流目标检测算法的对比实验

为了验证本文提出的针对钢材表面缺陷检测的改进算法的优势和性能,与相关文献模型和主流模型YOLOv3、YOLOv7、YOLOX,及近两年比较先进的YOLOv8、YOLO11[27]等进行横向的实验对比,数据集都为NUE-DET数据集,主要采用mAP、FPS和AP三种指标进行评估,实验结果如表6所示.

从表6可以看出,FasterRcnn和DDN作为两阶段two-stage检测算法模型,相对于YOLOv3能够保证较高的检测精度,DDN(RestNet50)的mAP值达到了最高的82.3%,但其检测速度FPS过低,并不适用于快速精准的检测任务;而SSD检测算法不仅FPS低,mAP值也只有54.7%,无法满足检测要求.

而本文提出的YOLOv5s-FNCE检测算法,相较于文献[30]和文献[31]的改进模型而言,拥有最高的检测精度和检测速度,各类缺陷检测精度也有了较大提升;与主流算法对比,相较于YOLOv4、YOLOv7、YOLOX-S、YOLOv6,不但有最高的FPS帧率,mAP值更是分别提高了5.5个百分点、0.6个百分点、0.6个百分点和4.1个百分点;和近两年最新提出的YOLOv8、YOLO11对比,mAP值分别提高了1.6个百分点和2.2个百分点.

相较于YOLOv5s,加入模块之后,网络结构复杂度提升,但FPS并没有因此下降,甚至提升了20.5%,mAP值更是提高了1.7个百分点,各类缺陷的检测精度均有提升;特别是对于裂纹(Cr)和麻点(Ps)的检测精度更是所有对比算法中的最优值.因此,上述对比可进一步验证本文提出的模型相较于其他主流模型,分别在检测精度和速度方面具有优势,尤其是对于细小缺陷的检测.

目标检测真值图通常指的是在计算机视觉和深度学习领域中用于训练和评估目标检测算法的图像标注.这些真值图通常包含了标注的目标位置、类别和其他相关信息,用于训练模型和评估检测算法的性能.这些真值图对于评估目标检测算法的准确性和鲁棒性非常重要,因为它们提供了模型训练和验

表6　对比实验结果

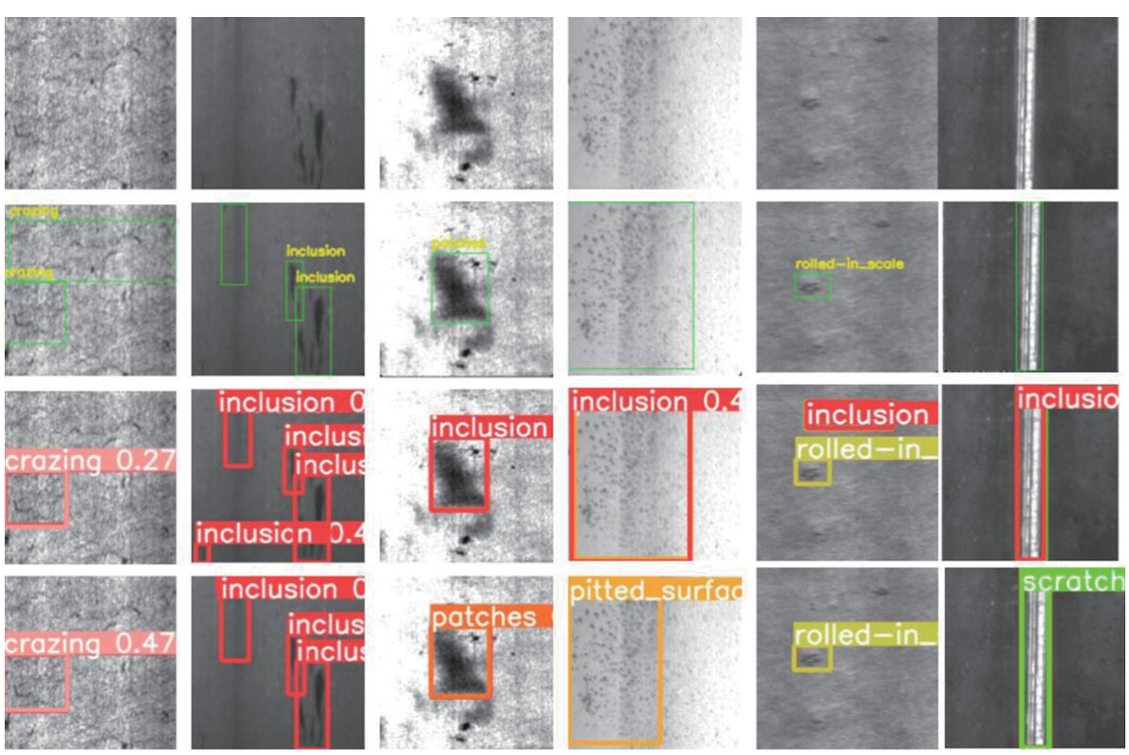
Tab.6　Comparison experiment result

算法模型	FPS/(帧·s^{-1})	mAP/%	AP/%					
			裂纹	夹杂	斑块	麻点	轧制氧化皮	划痕
YOLOv3	13	68.4	45.7	60.8	84.4	74.5	61.1	87.2
FasterRcnn	<20	70.2	46.7	61.3	82.8	76.5	**70.7**	83.3
SSD	62	54.7	21	54.4	76.5	78.2	31.4	80.5
DDN(RestNet50)[28]	<10	**82.3**	46.7	61.3	82.8	76.5	**70.7**	83.3
YOLOv4	59	69.6	39.5	68.3	85.2	73.8	59.3	81.7
YOLOv7	83.3	74.5	37.5	**85.3**	92.6	82.6	61.0	87.8
YOLOv5s	125.1	73.4	43.3	73.4	90.1	83.2	62.1	88.6
YOLOX-S[29]	52.4	74.5	35.1	82.1	92.1	84.1	53.6	87.6
YOLOv6[30]	121	71	31.9	69.7	**94.1**	83.2	54.7	82.0
YOLOv8	310	73.5	43.1	76.1	92.5	74.7	61.1	**93.3**
YOLO11	**344.8**	72.9	42.3	73.6	93	80.6	54.4	**93.3**
文献[31]	75	74.1	32.3	87.3	89.6	82.7	64.3	88.4
文献[32]	35.7	73.3	42.0	80.0	89.4	81.5	57.0	89.6
YOLOv5s-FNCE	150.8	75.1	**47.4**	72.9	90.2	**85.1**	64.4	90.5

注:粗体表示最优值.

证所需的标准参照.

如图 10 所示,第一行为原始缺陷照片,第二行为原始缺陷的真值图,第三行为 YOLOv5s 模型的检测结果,和真值图对比可以看出,会出现漏检和错检,比如将斑块(patches)、划痕(scratches)和麻点(pitted surface)错误检测成了夹杂(inclusion),并且多检测出了夹杂错误;第四行为本文提出的 YOLOv5s-FNCE 模型检测结果图,和真值图检测结果吻合较好,与原模型相比,各类缺陷并未出现错检、漏检、多检等情况,检测精度有较大提升.



(a)crazing　(b)inclusion　(c)patches　(d)pitted surface　(e)rolled-in scale　(f)scratches

图 10　六种缺陷对比图

Fig.10　Comparison chart of the six defects

3.6 模型泛化性实验

为了进一步验证改进模型的泛化性,除了在东北大学公开的钢材缺陷数据集验证本算法,另外采用印刷电路板(PCB)瑕疵数据集(图11)进行验证;该数据集由北京大学发布,其中包含1 386张图像以及6种缺陷:包括漏孔(missing hote)、鼠咬(mouse bite)、开路(open circuit)、短路(short)、毛刺(spur)和杂铜(spurious copper)),用于检测、分类和配准任务;训练集、验证集和测试集按7∶2∶1的比例划分.

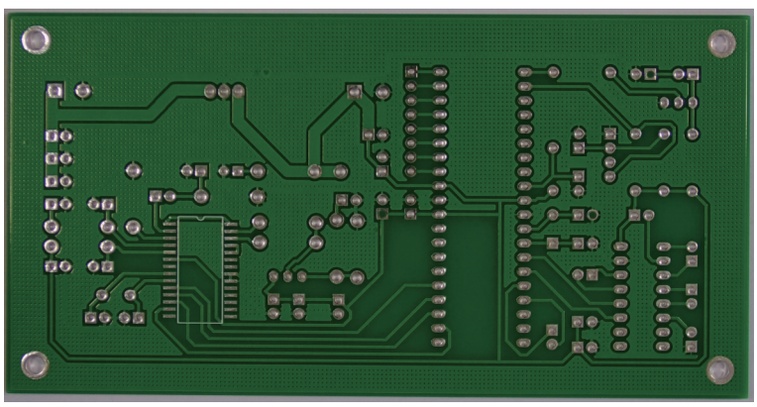


图11 PCB数据集示例

Fig.11 PCB dataset sample

各算法在参数量、检测速度、准确率等不同指标的对比结果如表7所示.

从表7可以看出,相较于原算法,改进算法的FPS帧率和mAP值分别提高了25.7和2.3个百分点;开路、毛刺、鼠咬三类缺陷也有明显的精度提升;证明了改进检测算法的泛化性.

表7 对比实验结果

Tab.7 Comparison experiment result

算法模型	FPS/(帧·s^{-1})	mAP/%	AP/%					
			漏孔	杂铜	开路	短路	毛刺	鼠咬
YOLOv5s	125.1	88.2	**99.5**	**93.1**	81.2	**97.1**	83.6	68
YOLOv5-FNCE	**150.8**	**90.5**	**99.5**	85.7	**92.3**	95.9	**90**	**86**

注:粗体表示最优值.

4 结 论

本文针对钢材表面缺陷检测中,环境变量复杂和个别缺陷类型因目标小而引起的特征难提取的问题,提出了改进后的YOLOv5检测模型.通过对比实验和消融实验可以得到以下结论:

1)引入了高效注意力模块NAM,对空间和通道子模块进行特征提取加强,并对特征信息进行归一化处理,能够更加关注重点区域,减少复杂背景的影响.

2)模型使用C3-Faster模块替换原有C3模块,减少模型不必要的计算量;在颈部网络和预测头引入CoordConv位置感应卷积,增强对于浅层特征的融合,提升特征位置信息的感应能力,赋予钢材缺陷目标更高的比例权重.

3)引入Focal-EIoU损失函数,加快收敛速度,有效降低误检.

上述结论表明,本文模型在NUE-DET数据集上精度相对于原有的YOLOv5s模型提升了1.7个百分点.相对于其他常用模型,本文模型在综合性能上有较大优势,可以实现对钢材表面缺陷的高效检测.后续研究将进一步挖掘特征融合之间的关联,采用更有效快速的特征融合模块来提高算法检测性能.

参考文献

[1] 李少波, 杨静, 王铮, 等. 缺陷检测技术的发展与应用研究综述[J]. 自动化学报, 2020, 46(11): 2319-2336.
LI S B, YANG J, WANG Z, et al. Review of development and

application of defect detection technology[J]. Acta Automatica Sinica,2020,46(11):2319-2336.(in Chinese)

[2] 康爽,陈长征,罗园庆,等. 基于微分形态学梯度风力发电机叶片缺陷边缘增强的红外检测研究[J]. 太阳能学报,2021,42(6):432-437.
KANG S, CHEN C Z, LUO Y Q, et al. Study on infrared detection edge enhancement of wind turbine blade defects based on differential morphology gradient [J]. Acta Energiae Solaris Sinica,2021,42(6):432-437.(in Chinese)

[3] 许鹏,耿明,方舟,等. 基于复合电磁的高速轨道缺陷检测方法研究[J]. 机械工程学报,2021,57(18):57-65.
XU P, GENG M, FANG Z, et al. Study on high-speed rail defect detection method based on combination of EC and MFL testing[J]. Journal of Mechanical Engineering, 2021, 57(18): 57-65.(in Chinese)

[4] 朱玉龙,赵迎松,方阳,等. 孔边裂纹的旋转涡流检测[J]. 中国机械工程,2023,34(8):883-891.
ZHU Y L, ZHAO Y S, FANG Y, et al. Rotating eddy current testing for inspection of cracks at hole edge[J]. China Mechanical Engineering, 2023, 34(8): 883-891.(in Chinese)

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

[6] HE K M, ZHANG X Y, REN S Q, 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.

[7] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision (ICCV). Santiago, Chile. IEEE, 2015: 1440-1448.

[8] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA. IEEE, 2016: 779-788.

[9] 彭艺,涂馨月,杨青青,等. 基于改进YOLOv5的轻量化无人机检测算法[J]. 湖南大学学报(自然科学版),2023,50(12):28-38.
PENG Y, TU X Y, YANG Q Q, et al. Lightweight UAV detection algorithm based on improved YOLOv5 [J]. Journal of Hunan University (Natural Sciences), 2023, 50 (12): 28-38. (in Chinese)

[10] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[C]// Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.

[11] 程婧怡,段先华,朱伟. 改进YOLOv3的金属表面缺陷检测研究[J]. 计算机工程与应用,2021,57(19):252-258.
CHENG J Y, DUAN X H, ZHU W. Research on metal surface defect detection by improved YOLOv3[J]. Computer Engineering and Applications, 2021, 57(19): 252-258.(in Chinese)

[12] 沈希忠,吴迪. 基于YOLO的铝型材料表面小缺陷检测[J]. 浙江工业大学学报,2022,50(4):372-380.
SHEN X Z, WU D. Detection of small defects on aluminum profile surface based on YOLO [J]. Journal of Zhejiang University of Technology, 2022, 50(4): 372-380.(in Chinese)

[13] 沈春光,李虎威,荆涛,等. 基于深度学习的带钢表面缺陷检测在小样本数据集的应用[J]. 轧钢,2022,39(2):82-86.
SHEN C G, LI H W, JING T, et al. Application of deep learning-based surface defect recognition of plate in small sample dataset[J]. Steel Rolling, 2022, 39(2): 82-86.(in Chinese)

[14] 王艳玲,苏盈盈,罗妤,等. 基于SSD模型的钢带表面缺陷检测系统设计[J]. 重庆科技学院学报(自然科学版),2021,23(3):95-98.
WANG Y L, SU Y Y, LUO Y, et al. Design of steel strip surface defect detection system based on SSD model [J]. Journal of Chongqing University of Science and Technology (Natural Sciences Edition), 2021, 23(3): 95-98.(in Chinese)

[15] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. 2021: 2107.08430. https://arxiv. org/abs/2107.08430v2.

[16] ZHU X K, LYU S C, WANG X, et al. TPH-YOLOv5: improved YOLOv5 based on transformer prediction head for object detection on drone-captured scenarios [C]//2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada. IEEE, 2021: 2778-2788.

[17] FARHADI A, REDMON J. Yolov3: An incremental improvement [C]//Computer Vision and Pattern Recognition. Berlin/Heidelberg, Germany. Springer, 2018: 1804-2767.

[18] BOCHKOVSKIY A, WANG C Y, LIAO H M. YOLOv4: optimal speed and accuracy of object detection [EB/OL]. 2020: 2004.10934.https://arxiv.org/abs/2004.10934v1.

[19] WANG C Y, BOCHKOVSKIY A, LIAO H 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). Vancouver, BC, Canada. IEEE, 2023: 7464-7475.

[20] LIU Y C, SHAO Z R, TENG Y Y, et al. NAM: normalization-based attention module[EB/OL]. 2021: 2111.12419.https://arxiv. org/abs/2111.12419v1.

[21] CHEN J R, KAO S H, HE H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks [EB/OL]. 2023: 2303.03667. https://arxiv. org/abs/2303.03667v3.

[22] LIU R, LEHMAN J, MOLINO P, et al. An intriguing failing of convolutional neural networks and the CoordConv solution [EB/OL]. 2018: 1807.03247. https://arxiv. org/abs/1807.03247v2.

[23] ZHANG Y F, REN W Q, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.

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

[25] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT. IEEE, 2018.

[26] SONG Q S, LI S B, BAI Q, et al. Object detection method for grasping robot based on improved YOLOv5[J]. Micromachines, 2021, 12(11): 1273.

[27] VARGHESE R, M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness [C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India. IEEE, 2024: 1-6.

[28] 郭龙源,段厚裕,周武威,等. 基于Mask R-CNN的磁瓦表面缺陷检测算法[J]. 计算机集成制造系统,2022,28(5):1393-1400.
GUO L Y, DUAN H Y, ZHOU W W, et al. Surface defect detection algorithm of magnetic tile based on Mask R-CNN[J]. Computer Integrated Manufacturing Systems, 2022, 28(5): 1393-1400. (in Chinese)

[29] 熊聪,于安宁,高兴华,等. 基于改进YOLOX的钢材表面缺陷检测算法[J]. 电子测量技术,2023,46(9):151-157.
XIONG C, YU A N, GAO X H, et al. Steel surface defect detection algorithm based on improved YOLOX[J]. Electronic Measurement Technology, 2023, 46(9): 151-157. (in Chinese)

[30] 黄硕清, 黄金贵. 基于RFB和YOLOv5特征增强融合改进的钢材缺陷检测方法[J/OL]. 计算机工程, 2024: 1-15. (2024-06-03). http://kns.cnki.net/KCMS/detail/detail.aspx? filename=JSJC20240530001&dbname=CJFD&dbcode=CJFQ.
HUANG S Q, HUANG J G. Improved steel defect detection method based on RFB and YOLOv5 feature enhancement fusion[J/OL]. China Industrial Economics, 2024: 1-15. (2024-06-03). http://kns.cnki.net/KCMS/detail/detail.aspx? filename=JSJC20240530001&dbname=CJFD&dbcode=CJFQ. (in Chinese)

[31] 曹义亲, 伍铭林, 徐露. 基于改进YOLOv5算法的钢材表面缺陷检测[J]. 图学学报, 2023, 44(2): 335-345.
CAO Y Q, WU M L, XU L. Steel surface defect detection based on improved YOLOv5 algorithm[J]. Journal of Graphics, 2023, 44(2): 335-345. (in Chinese)

[32] 王浩然. 基于YOLOv5的钢材表面缺陷检测研究[D]. 桂林: 广西师范大学, 2022.
WANG H R. Research on steel surface defect detection based on YOLOv5 [D]. Guilin: Guangxi Normal University, 2022. (in Chinese)