DOI: 10.13382/j.jemi.B2407618

融合 HGnetv2 和注意力机制的钢材表面缺陷检测方法*

张 航[1] 周 毅[1] 邱宇峰[2]
(1. 武汉科技大学信息科学与工程学院 武汉 430081;2. 宝信软件(武汉)有限公司 武汉 430080)

摘 要:针对多尺度、多类型和复杂背景的钢材表面缺陷检测精度低的问题,设计一种融合 HGnetv2 和注意力机制的改进 YOLOv5 算法。首先,基于 HGnetv2 网络结构引入注意力机制作为骨干层,提升对小目标缺陷的特征提取能力。然后,在特征融合层中,将注意力机制和 Involution 操作结合,实现对浅层边缘信息和深层语义信息的有效聚合。其次,采用 CBME_C2f 替换了原模型的 C3_Bottleneck,提供了更丰富的梯度流信息。此外,使用一种新的预测框损失 VCIoU,通过计算预测框和目标框顶点和两者中心点之间的位置信息特征,提高了边界框回归精度。最后,引入了 MetaAconC 激活函数,自适应地调整每个特征图通道激活的非线性程度,提高从复杂背景中提取特征信息的性能。实验结果表明,此方法在 NEU-DET 数据集上平均精度 mAP50 指数达到了 81.4%,相较于 YOLOv5s 算法提高了 5.4%,mAP@50:95 指数达到了 44.1%,相较于 YOLOv5s 算法提高了 2.8%。除此之外,针对此数据集中的微小缺陷 Crazing,平均检测精度达到了 55.4%,相比原 YOLOv5s 提高了 18.1%,同时检测速度为 80.6 fps。与其他主流的缺陷检测算法对比,该算法在满足对钢材表面检测实时性要求的条件下,提高了检测精度。

关键词:钢材缺陷;缺陷检测;YOLOv5;注意力机制;深度学习;人工神经网络
中图分类号:TP391.4;TN98 **文献标识码**:A **国家标准学科分类代码**:520.2060

Detection method of steel surface defects with fusion of HGnetv2 and attention mechanism

Zhang Hang[1] Zhou Yi[1] Qiu Yufeng[2]
(1. School of Information Science and Engineering, Wuhan University of Science and Technology, Wuhan 430081, China; 2. Baosight Software(Wuhan) Co., Ltd, Wuhan 430080, China)

Abstract: Addressing the low accuracy problem in detecting multi-scale, multi-type steel surface defects within complex backgrounds, this paper designs an improved YOLOv5 algorithm that integrates HGnetv2 with an attention mechanism. First, the HGnetv2 network incorporates an attention mechanism as a backbone layer to enhance feature extraction capabilities for small target defects. Second, in the feature fusion layer, attention mechanisms and involution operations are combined to achieve effective aggregation of edge features in shallow layers and semantic information in deep layers. Besides, CBME_C2f replaces the C3_Bottleneck module to improve gradient flow. Additionally, a new bounding box loss function, VCIoU, is used to calculate positional features between the vertices and center points of the prediction and target boxes, enhancing bounding box regression precision. Finally, MetaAconC is introduced to adaptively adjust the non-linearity of activation for each feature map channel, improving the ability to extract feature information from complex backgrounds. Experimental results on the NEU-DET dataset show that the proposed method achieves an mAP50 of 81.4% and an mAP@50:95 of 44.1%, which is 5.4% and 2.8% better than YOLOv5s respectively. For the small defects such as crazing in this dataset, the detection accuracy reaches 55.4%, representing an 18.1% improvement over YOLOv5s, while maintaining a detection speed of 80.6 fps. Compared to other mainstream defect detection algorithms, this algorithm improves accuracy while meeting the real-time demands of steel surface defect detection.

Keywords: steel defects; defect detection; YOLOv5; attention mechanism; deep learning; artificial neural network

收稿日期:2024-06-24 Received Date:2024-06-24
*基金项目:国家自然科学基金(62372343)项目资助

0　引　言

钢铁生产是各个国家的基础建设非常重要的一环，钢铁材料在我国使用场所非常广泛，比如建筑工程、交通运输，机械制造、能源开采、电力工业等等涵盖了经济建设、基础设施建设、军事安全、能源资源开发、生活用品等多个领域，是支撑国家经济发展和社会进步的重要材料之一。在钢铁厂中，所生产的钢铁表面由于生产、运输或者存储过程中会使得钢铁表面出现细裂纹、杂质物、斑块、凹陷与坑洞、表面氧化铁锈、划痕等 6 种常见缺陷，这其中一些缺陷可能导致重大的事故，所以确保钢铁表面质量是一个很值得重视的问题。然而采用人工判定钢铁表面缺陷的方法不仅低效，还容易产生漏检，从而造成质量隐患。因此，探索出一种对于精度和速度都很高的表面缺陷检测算法有着重要意义。基于深度学习的计算机视觉进行目标检测现如今是一种高精度、高效率、较低成本的方法[1]。

视觉目标检测作为计算机视觉领域的一个重要研究方向，旨在从图像或视频中识别并定位出感兴趣的目标或物体。对于钢铁表面缺陷检测而言，视觉目标检测的关键在于构建一个能够自动学习和识别各种缺陷模式的模型。通过训练深度学习模型，能够让计算机学会从复杂的背景中区分出缺陷区域，并对其进行准确的分类和定位。目前深度学习在目标检测领域的应用非常广泛，基于深度学习的目标检测算法主要分为两类，一阶段算法(one-stage)和两阶段算法(two-stage)。其中一阶段算法直接从图像中对目标进行检测分类，代表算法为 SSD、YOLO。二阶段算法进行检测分为两步，第一步先生成候选框，第二步从候选框中筛选出具体的目标和分类，代表算法为 Faster R-CNN。至今为止，许多深度学习领域学者对目标检测开展了大量的研究，提出了许多的检测方法。赵佰亭等[2]基于 YOLOv7 模型提出了 ECC-YOLO 算法，该算法通过引入 ConvNeXt 特征增强模块、C2fFB 模块降低计算量、MPCE 模块削弱背景干扰，在钢表面缺陷检测中实现了 77.2%的 mAP，较 YOLOv7 提高了 10.1%的检测精度。马志程等[3]针对光学元件划痕缺陷检测，改进了 Mask R-CNN 模型，采用 CSPResNet 和 ESE 注意力机制，通过 K-means 聚类优化 anchor boxes，并引入 Focal Loss，从而显著提升了检测精度，且几乎不影响推理速度。石欣等[4]基于 YOLOv4 提出了一种改进的远距离行人小目标检测方法，通过引入浅层特征和特征自适应融合来增强多层语义关联，同时结合 ESRGAN 提升小目标检测精度。然而，尽管精度有所提高，方法在复杂背景下的小目标检测效果仍有待优化。姜媛媛等[5]针对印刷电路板表面小目标缺陷检测，提出了 Multi-CR YOLO 网络，通过多尺度特征融合和增强模块，有效提高了检测精度，同时保持了实时检测速度，模型轻量且高效，满足了小目标缺陷的实时检测需求。张福豹等[6]提出一种结合 RRDNet 弱光增强和改进 YOLOv3 算法的锯链缺陷检测方法，通过自适应增强图像亮度和改进网络结构，显著提高了检测精度，降低了漏检率和过检率，实现了锯链缺陷的在线检测。Zheng 等[7]基于 YOLOv7 网络模型改进了一种能够在具有复杂背景的输电线路图像中实现小目标的检测，通过引入 Coordinate Attention 模块和 HorBlock 模块提升了模型的检测精度，以及利用 SIoU 损失和焦点损失函数加速模型收敛，有效的实现复杂背景下小目标的精确检测。

尽管现有方法在小目标检测和复杂背景处理方面取得了一定进展，但仍然面临诸多挑战。首先，由于钢铁表面缺陷种类繁多、背景复杂，特别是小缺陷的特征信息较少且易受噪声干扰，现有方法难以精确定位和识别这些细节，导致漏检率较高。其次，在提升检测精度的同时，如何满足工业检测对实时性的严格要求，依然存在难以平衡的矛盾。针对这些问题，本研究提出了一种改进的 YOLOv5 算法，在保证实时性检测需求的基础上显著提升了模型的检测精度。主要创新点如下：

1)将模型的骨干网络替换为一种改进的沙漏形特征提取网络(hourglass netv2, HGnetv2)网络，保证准确性的前提下，使用深度可分离卷积(depthwise separabl convolution, DWConv)和轻量级卷积(LightConv)模块降低了骨干网络的计算量，并且堆叠多个沙漏形卷积模块(hourglass block, HGBlock)和卷积块注意力模块(convolutional block attention module, CBAM)提高了模型对多尺度缺陷的感知能力。

2)颈部网络使用改进卷积(convolution-batchnormalization-metaaconc, CBME_Conv)替换部分原始卷积，在 FPN 层嵌入反卷积(Involution)算子和坐标注意力机制(coordinate attention, CA)，在捕捉到特征空间中远距离像素点关系之外，也增强算法对不同方向关键信息的关注，提高模型对复杂背景缺陷的辨析能力，提高对小缺陷的检测精度降低漏检率。

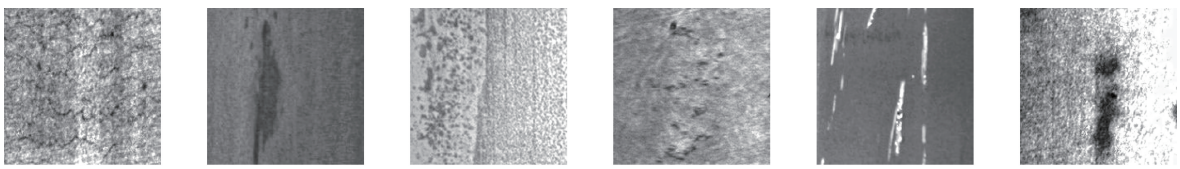
3)用改进跨阶段部分层模块(conv-bn-me cross stage partial layer_2conv, CBME_C2f)结构替换颈部网络的原始 C3 结构，在保证轻量化的同时获得更加丰富的梯度流信息，并根据模型尺度来调整通道数，动态的自适应的调整激活函数的线性/非线性，有助于提高泛化能力并且大幅提升了模型性能。

4)使用边界框 VCIoU 损失，在不增加计算量的同时提升了模型检测精度。经过实验，在 NEU-DET 数据集上验证了改进 YOLOv5 算法的有效性。

1 背景知识

1.1 缺陷检测问题

缺陷检测作为目标检测的一个重要分支,在工业自动化和质量控制领域扮演着至关重要的角色。其核心目标是在各种复杂背景下准确识别并定位出产品表面或内部的异常部分,如裂痕、污点、凹陷等。这些缺陷不仅影响产品的美观,还可能降低其性能和使用寿命,因此,高效的缺陷检测方法对于提升产品质量具有重要意义。缺陷的种类繁多,形态各异,这要求检测方法必须具备高度的灵活性和准确性。以钢铁表面缺陷为例,常见的缺陷包括 Crazing(龟裂)、Inclusion(夹杂)、Patches(斑块)等,如图 1 所示。这些缺陷在形态、大小和分布上都有所不同,增加了检测的难度。特别是 Crazing 缺陷,其细微的裂纹网络在初期可能并不明显,且容易与钢材表面的纹理混淆,导致检测难度极大。



(a) 细裂纹 (a) Crazing
(b) 杂质物 (b) Inclusion
(c) 凹陷与坑洞 (c) Pitted_surface
(d) 氧化铁锈 (d) Rolled-in_scale
(e) 划痕 (e) Scratches
(f) 斑块 (f) Patches

图 1 钢铁表面 6 种缺陷

Fig. 1 Six defects on the surface of steel

1.2 YOLOv5 算法

近年来,深度学习技术的快速发展为缺陷检测领域带来了新的解决方案。在目标检测领域,一阶段和二阶段检测器是两种主要的方法。一阶段检测器,如 YOLO,因其速度快、效率高的特点,在实时性要求较高的场景中应用广泛。而二阶段检测器,如 Faster R-CNN,则以其高精度检测能力在需要精细识别的场景中占据优势。在钢铁表面缺陷的检测中,考虑到实时性和准确性的平衡,选择了 YOLO 作为检测模型。YOLO 模型能够直接在图像上预测边界框和类别概率,实现了快速而有效的缺陷检测。通过调整 YOLO 模型的网络结构和参数设置,可以针对钢铁表面缺陷的特点进行优化,提高检测的精确度和召回率。

YOLOv5 是 YOLO 算法中的一个非常经典并且性能非常优良的一个版本,具有检测精度高、检测速度快的特点。网络一共有 4 个版本,分别是 YOLOv5s、YOLOv5m、YOLOv5l、YOLOv5x 四个模型。工业缺陷检测旨在实现一个轻量化并且检测精度高的检测算法,因此选择 YOLOv5s 版本为基线。结构是由骨干网络、特征融合层网络、检测头组成。骨干网络进行特征的提取,由标准卷积块 Conv 模块和 C3 模块以及空间金字塔池化层 SPPF 模块组成。颈部网络采用 FPN[8]+PAN[9] 的结构,FPN 层自顶向下传达强语义特征,PAN 塔自底向上传达定位特征,对不同尺度的特征进行融合,提高模型对不同尺度缺陷的检测能力。检测头部分主要是对颈部网络所输出的 3 个尺寸的特征图进行检测,特征图的每个像素点网格都存在 3 个先验框,用来预测和回归目标。首先对先验框的中心点坐标进行偏移得到预测框的中心点坐标,以及对预测框的宽高进行调整得到预测框的宽高,通过非极大值抑制算法(NMS)过滤掉多余的预测框,上面的步骤也称为预测结果的解码操作,得到最终的检测结果。

2 改进的 YOLOv5 算法

基线网络模型 YOLOv5s 尽管在工业检测领域以其高效的检测速度、优秀的准确性和相对较小的模型大小而广受好评,但是在实际的应用中,依旧存在一些挑战和问题,这些问题可能包括在某些复杂背景下的检测精度不足,对于小目标检测的困难、或者是在处理特定类型物体时的泛化能力有限,比如钢板表面缺陷裂纹 Crazing 等,基于此所提出的算法在骨干网络、特征融合层以及交并比计算等几个方面进行了相应的改进,以解决上述问题。

2.1 骨干网络设计

缺陷检测任务中普遍存在微小特征的检测,但是 YOLOv5 骨干网络在提取小目标特征时存在以下两个问题。首先,定位精度虽然能够满足一般应用的需求,但是达不到精确检测的要求,容易出现漏检或误检的情况;其次,由于采用了少量的卷积结构,这对于提取缺陷特征是不利的。本研究引入 RT-DETR[10] 模型的主干网络 HGnetv2 去替换 YOLOv5 的主干网络,RT-DETR 是由百度推出的第一款实时的端到端检测模型,其利用两个主干,一个是 HGNet,另外一个是 ResNet,其中 HGnetv2 即为引入的主干网络,它具有强大的特征提取能力,能够提

取更加丰富和深层次的特征,从而更好地刻画目标缺陷的细节和语义信息,采用了轻量化的设计,在保持高精度的同时,也具有较高的计算效率。改进 YOLOv5 结构如图 2 所示。

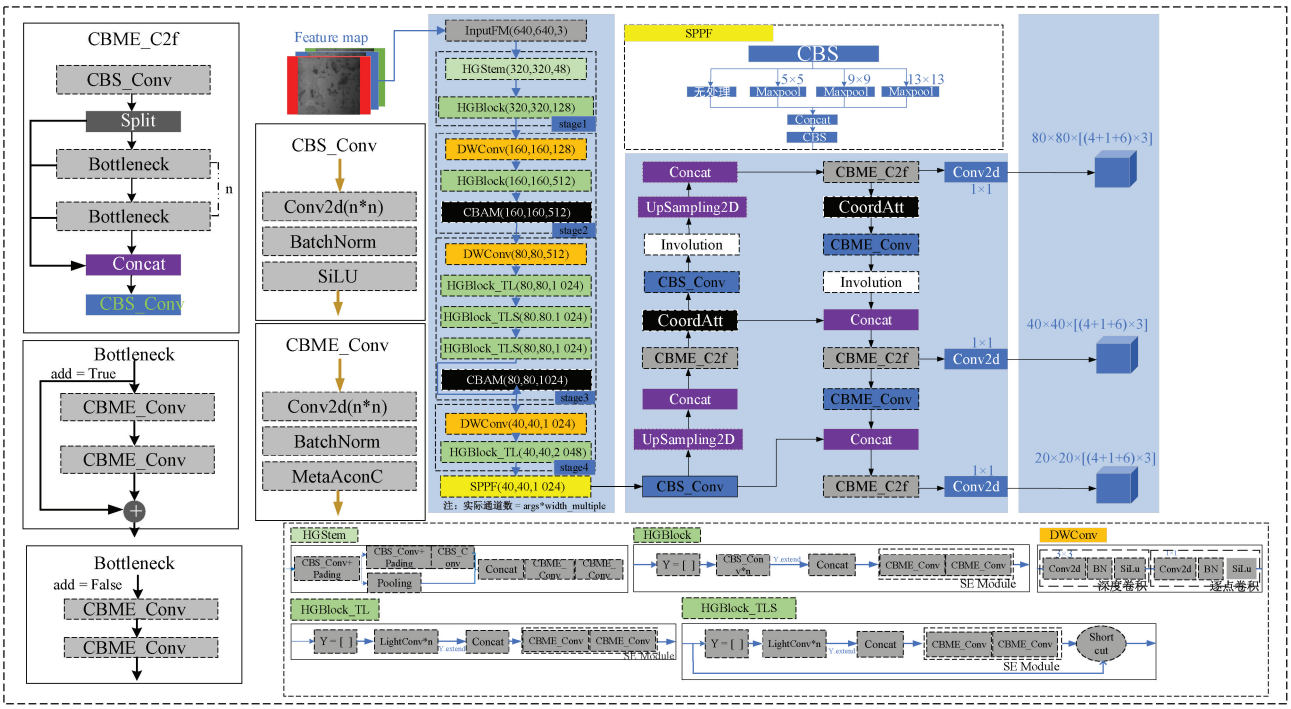


图 2　改进 YOLOv5 模型结构

Fig. 2　Improved YOLOv5 model structure

HGnetv2 是一种用于目标检测的神经网络模型,经过训练可以通过输入图像来检测图中的目标物体。HGnev2 针对 GPU 设备,尽可能多的使用 3×3 的标准卷积对特征进行提取,且标准卷积的数量随层数深度增加而增加,从而得到一个有利于 GPU 推理的骨干网络,同样速度下,精度也远远超越其他的卷积神经网络。HGnetv2 网络结构是由多个 HGBlock 组成,相对于 HGnet,在 HGnetv2 中引入了 LightConv 和 DWConv[11] 模块,目的也是为了在增强特征提取效果的前提下减少计算量。

所提方法的部分卷积中采用自适应控制激活[12](MetaAconC)作为激活函数。MetaAconC 是一种基于 Swish[13] 和 Maxout[14] 系列激活函数的扩展,钢材表面缺陷的形态多样,包括细小划痕、凹坑和纹理变化等。这些缺陷往往具有高度的非线性,并且不同类型的缺陷在特征空间中的分布差异显著。MetaAconC 通过动态 β 调整,在特征图的每一通道上自适应地选择激活函数的非线性程度,更好地捕捉这些复杂特征。AconA 激活函数是深度学习领域中常用的 Swish 激活函数,是 ReLU 函数的一种平滑近似、AconB 即推广到 PReLU[15]、Leaky-ReLU[16] 等函数的平滑形式、AconC 通过学习参数 P1 和 P2,能够获得性能更好的激活函数。而 MetaAconC 激活函数是基于 AconC 对其的一种扩展和改进,参数 β 是通过全局池化和两层全连接层自适应生成的。这种自适应机制使得每个神经元都可以根据输入数据的变化,灵活选择激活函数的形态。这一特点使得 MetaAconC 适合处理钢材表面缺陷检测这类复杂和多变的特征。Acon 公式和 Maxout 系列部分激活函数对比如表 1 所示。改进后的 HGnetv2 总体结构如图 3 所示。

表 1　Acon 激活函数与 Maxout 激活函数

Table 1　Acon activation function and Maxout activation function

		Maxout activation	Acon activation
x	0	max(x,0):ReLU	Acon-A(Swish): $x \cdot \sigma(\beta x)$
p_1x	p_2x	max(p_1x,p_2x)	Acon-C: $(p_1 - p_2)x \cdot \sigma(\beta(p_1 - p_2)x) + p_2x$

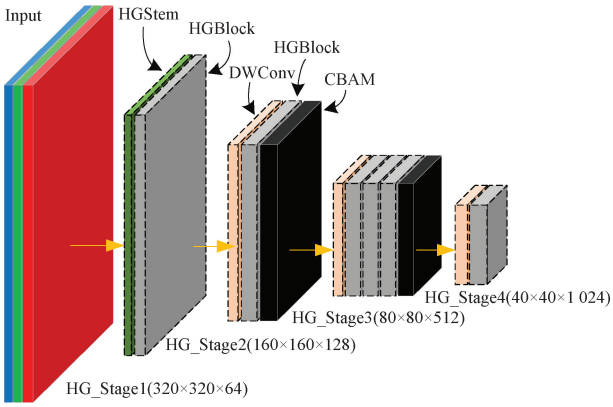


图 3 改进后的 HGnetv2 结构

Fig. 3 Improved HGnetv2 architecture

2.2 注意力模块

在缺陷检测中,复杂背景信息可能会影响算法的识别能力。当缺陷位于具有纹理、颜色和光照变化的复杂背景中时,缺陷特征容易与背景混淆,导致检测算法难以区分。制造业产品表面常常存在这些干扰因素,从而降低了检测的准确性。骨干层网络在提取特征时,可能会因为网络结构、参数设置等原因,导致对某些缺陷的特征提取精度不高。特征融合层在将不同层级的特征进行融合时,可能会因为融合方式、融合比例等原因,导致某些关键信息的丢失或混淆。例如,低层特征分辨率更高、包含更多细节信息,但语义性较低、噪声较多;高层特征则具有更强的语义信息,但分辨率较低、对细节的感知能力较差。首先,针对骨干层网络在提取特征时对于小缺陷 Crazing 检测精度不高的问题,在骨干网络中引入了卷积块注意力模块 CBAM[17] 模块。如图 4 所示。其次,如图 6 所示,针对特征融合层在检测带有复杂背景信息的缺陷时效果不佳的问题,采用了坐标注意力机制模块 CA[18],在融合多个尺度特征层时,通过坐标分解,可以高效捕捉小缺陷与全局之间的依赖关系。

1) 卷积块注意力模块

CBAM 由两个独立的子模块组成,分别为图 5(a)中的通道注意力模块和图 5(b)中的空间注意力模块。混合注意力模块 CBAM 就是将通道注意力模块和空间注意力模块的输出特征逐元素相乘,得到最后的注意力增强特征,此特征将继续用作后续网络层的输入,注意到关键信息的同时,抑制了噪声和无关信息。此模块嵌入到骨干网络的 Stage2 和 Stage3 部分,如图 2 和 3 所示,提高骨干网络对特征的感知能力,以便于在多尺度融合特征时有更好的表现。经过通道注意力模块和空间注意力模块处理后的特征表示如式(1)和(2)所示。

$$F' = M_c(F) \otimes F \tag{1}$$

$$F'' = M_s(F') \otimes F' \tag{2}$$

其中,通道注意力 $M_c(F)$ 和空间注意力 $M_s(F')$ 如式(3)和(4)所示。

$$M_c(F) = \sigma(MLP(\text{AvgPool}(F)) + MLP(\text{MaxPool}(F))) \tag{3}$$

$$M_s(F') = \sigma(F_c^{7\times7}([\text{AvgPool}(F'); \text{MaxPool}(F')])) \tag{4}$$

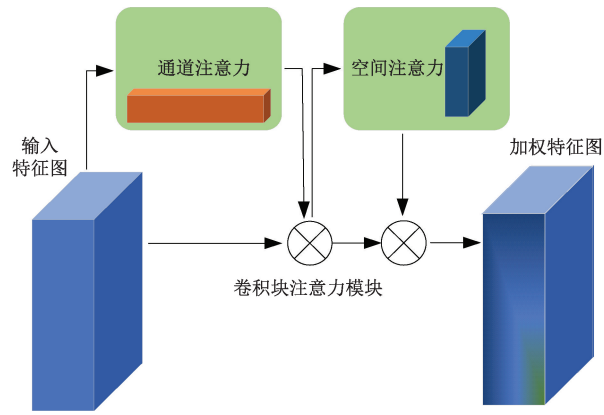


图 4 CBAM 模块结构

Fig. 4 CBAM module structure

2) 坐标注意力机制模块

CA 主要由两个部分构成,主干为坐标注意力生成,支干为两个并行阶段,是对特征坐标信息进行嵌入,即使用精确的位置信息捕捉到空间维度的远程交互信息。将特征图分别在宽度和高度方向分别进行全局平均池化,得到全局信息分解分别将特征映射到了高维度和宽维度上,输出特征表示如式(5)和(6)所示。

$$Z_c^h(h) = \frac{1}{W_0}\sum_{0\leqslant i<W} x_c(h,i) \tag{5}$$

$$Z_c^w(w) = \frac{1}{H}\sum_{0\leqslant j<H} x_c(j,w) \tag{6}$$

这两个变换沿着两个空间方向进行特征聚合,返回沿高和宽方向的感知注意力图。相较于 CBAM,CA 模块更适合应用于特征融合,因为 CA 不仅可以捕捉通道间的信息,还能有效聚合空间维度的长距离依赖,并保留精确的位置信息。这两种变换也允许注意力模块捕捉到沿着一个空间方向的长距离依赖,并保存沿着另一个空间方向的精确位置信息,这有助于网络更准确地定位感兴趣的目标。将式(5)和(6)所产生的聚合特征 $Z_c^h(h)$ 和 $Z_c^w(w)$ 沿一个方向进行 Concat 拼接,然后使用一个共享的 1×1 卷积 F_1,经过激活函数处理后得到中间特征表示 $f \in R^{\frac{C}{r}\times(H+W)}$,这里的 r 表示下采样比率。如式(7)所示。

$$f = \delta(F_1([z^h, z^w])) \tag{7}$$

沿着空间维度,再将 f 进行 split 操作,分成 $f^h \in R^{C\times H\times1}$ 和 $f^w \in R^{C\times1\times W}$,然后分别利用 1 × 1 卷积进行升维度操作,再结合 sigmoid 激活函数得到和输入 x 相同的

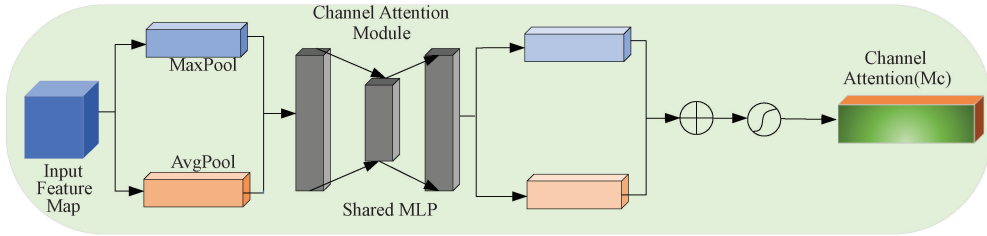


图 5　CBAM 模块分解结构

Fig. 5　CBAM module decomposition structure

通道数的注意力权重参数 $g^h \in R^{C\times H\times 1}$ 和 $g^w \in R^{C\times 1\times W}$。如式(8)和(9)所示。

$$g^h = \delta(F_h(f^h)) \tag{8}$$

$$g^w = \delta(F_w(f^w)) \tag{9}$$

最后对 g^h 和 g^w 扩展作为注意力权重,对原始特征图像进行乘法加权得到特征映射,CA 模块最终输出表达式为如式(10)所示。

$$y_c(i,j) = x_c(i,j) \times g_c^h(i) \times g_c^w(j) \tag{10}$$

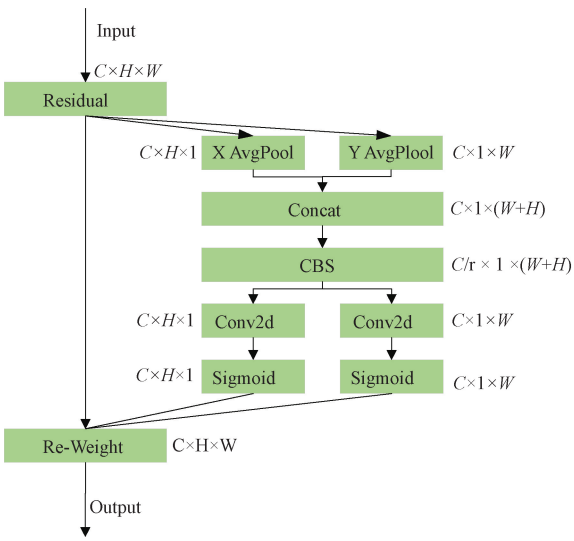


图 6　Coordinate Attention 结构

Fig. 6　Coordinate Attention structure

2.3　CBME_C2f

为优化模型性能,使用改进跨阶段部分层模块 CBME_C2f 替换 YOLOv5 中的 C3 模块,结构如图 7 所示。C2f 通过分割输入特征并并行处理梯度,有效提高了梯度流的丰富性和模型的检测精度。以下通过 3 个方面分析 C2f 在梯度流动和性能提升中的优势:

1) 跳跃路径保留梯度

C2f 模块通过 Split 操作,将输入特征分成两部分。一部分特征通过跳跃路径直接传递到后续层,不经过深层卷积或降维操作。这条路径在反向传播时能够保持完整的梯度流动,不会发生梯度衰减。这相当于引入了一条附加的“梯度通道”,如式(11)所示。

$$\frac{\partial L}{\partial X_{Skip}} = \frac{\partial L}{\partial X_{out}} \tag{11}$$

确保梯度不会因为深层网络的复杂操作而消失。这与 C3 模块的设计不同,C3 依赖于残差路径,但其大部分梯度仍需要经过多层 BottleNeck 进行特征学习,导致随着网络深度的增加,梯度逐渐衰减。

2) 分散梯度衰减风险

C2f 的优势不仅在于跳跃路径的完整梯度保留,还在于它通过两条路径分散了梯度的流动。跳跃路径中的梯度不受卷积层的影响,保持完整,而 BottleNeck 路径中的梯度虽然会衰减,但通过并行设计,这种衰减只影响一部分梯度,如式(12)所示。

$$\frac{\partial L}{\partial X_{C2f}} = 0.5 \cdot \frac{\partial L}{\partial X_{Skip}} + 0.5 \cdot \frac{1}{r^n} \cdot \frac{\partial L}{\partial X_{out}} \tag{12}$$

其中,r 为降维比,n 为 BottleNeck 的层数。

相比之下,C3 模块中虽然有残差连接,但大部分梯度仍然依赖于经过多层 BottleNeck,C3 模块的梯度衰减风险更集中,而 C2f 模块通过分散设计更好地避免了这

种集中衰减。

3）更均匀的梯度分布

C2*f* 模块的并行路径设计使梯度分布更加均匀。通过两条路径的协同工作，避免了单一路径对梯度流动的过度依赖。这种设计在深层网络中特别有效，减少了梯度消失问题的发生，而 C3 模块由于主要依赖于 BottleNeck 路径的梯度，容易在深度增加时出现梯度衰减。因此，C2*f* 模块在深层网络中能够更好地平衡梯度流动，确保每层特征都能得到有效的梯度反馈。

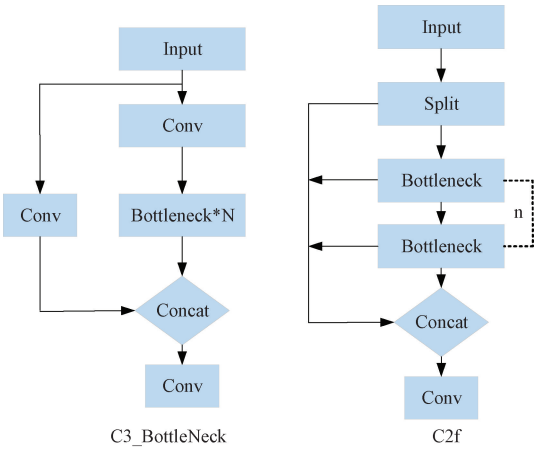


图 7 C3 模块与 C2*f* 模块结构

Fig. 7 C3 module and C2*f* module structure

2.4 Involution 动态卷积

传统卷积和分组卷积在空间上具有空间未知性，即在所有空间位置上共享相同的卷积核，导致无法有效捕捉长距离空间依赖关系。而其通道特异性则指每个输出通道都拥有独立的卷积核，聚合所有输入通道的信息。这使得传统卷积在处理复杂背景和多尺度缺陷时能力受限。

为解决这一问题，网络中引入一种 Involution[19] 算子，Involution 操作如图 8 所示，具体引入后的结构如图 2 所示。Involution 算子反其道而行，在通道维度上共享卷积核，而在空间维度上使用空间特异的卷积核。这种设计能够灵活学习不同位置间的空间依赖，尤其擅长处理长距离信息交互。将 Involution 算子嵌入特征融合层后，能够有效增强模型对复杂背景和多尺度缺陷的建模能力，提高检测精度。

Involution 卷积核的大小为 $H \times W \times K \times K \times G$，其中 $G << C$，表达所有通道共享 G 个卷积核，Involution 的操作表达式如式(13)所示。

$$Y_{i,j,k} = \sum_{(u,v)\in\Delta_k} H_{i,j,u+K/2,v+K/2,u+KG/C} X_{i+u,j+v,k} \tag{13}$$

其中，$H \in \boldsymbol{R}^{H\times W\times K\times K\times G}$ 是 Involution 卷积核。

在 Involution 中，没有像传统卷积一样采取固定的权重参数作为可学习的参数，而是根据输入的特征图生成对应的 Involution kernel，从而能够确保卷积核和输入特征图尺寸在空间维度上自动对齐。Involution kernel 生成的通用形式如式(14)所示。

$$H_{i,j} = \Phi(X_{\psi_{i,j}}) = \boldsymbol{W}_1\sigma(\boldsymbol{W}_0 X_{i,j}) \tag{14}$$

其中，$\psi_{i,j}$ 是坐标(i,j) 领域的一个索引集合，所以 $X_{\psi_{i,j}}$ 表示输入特征图上包含 $X_{i,j}$ 的某个部分，$\boldsymbol{W}_0 \in \boldsymbol{R}^{\frac{C}{r}\times C}$ 和 $\boldsymbol{W}_1 \in \boldsymbol{R}^{(K\times K\times G)\times\frac{C}{r}}$，两者是线性变换矩阵，$r$ 是通道缩减比率，σ 是中间的 BN 层和 ReLU 激活函数。

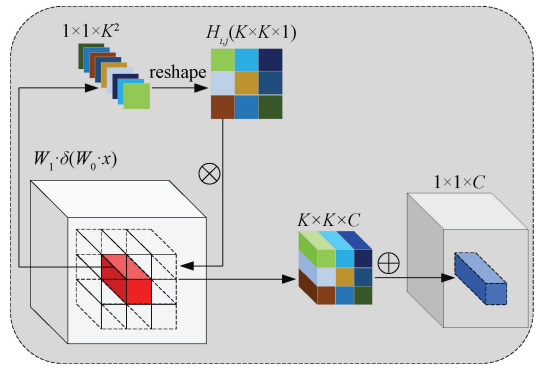


图 8 Involution 结构

Fig. 8 Involution structure

生成卷积核的过程为：

$$X_{i,j}:1\times 1\times C \xrightarrow{FC\ 全连接层} 1\times 1\times\frac{C}{r} \xrightarrow{FC\ 全连接层} 1\times 1\times(K^2) \xrightarrow{reshape} K\times K\times 1$$

获得卷积核之后的卷积过程为：

$$Input\ feature\ map X_{\omega_{i,j}}:K\times K\times C$$

$$\xrightarrow{involution\ kernel:K\times K\times GK\times K\times K\ CH、W\ 求和} 1\times 1\times C$$

其中，$\omega_{i,j}$ 为坐标(i,j) 附近的 $K\times K\times C$ 的领域。

2.5 VCIoU Loss

在 YOLOv5 网络中，特征图经过骨干网络和特征融合层后，输出 3 个加强特征层，经过头部网络得到预测结果。预测过程中涉及 3 类损失：边界框回归损失(RegLoss)，用于衡量预测框与真实框的位置误差；分类损失(ClsLoss)，用于衡量物体分类的准确性；置信度损失(ObjLoss)，用于衡量特征点是否包含物体的置信度差异。YOLOv5 通过 9 个不同大小的先验框匹配真实框，优化预测框的位置，使其更接近真实框。

为进一步提升预测框的准确性，传统的 L1[20] 损失和 IoU[21] 损失虽被广泛使用，但它们无法全面反映预测框与真实框的空间关系。为此，GIoU[22] 引入了对重叠和非重叠区域的关注，而 DIoU[23] 则加入了框的中心点距离，改善了回归稳定性。最终，CIoU[23] 在 DIoU 的基础上增加了对框长宽比的优化，结合位置、距离和尺度，使目标

框的回归更加精确和稳定。如式(15)所示。

$$L_{CIoU} = 1 - IoU(A,B) + \frac{\rho^2(A_{ctr},B_{ctr})}{c^2} + \alpha v \tag{15}$$

并且，$\alpha = \frac{v}{(1-IoU)+v}, v = \frac{4}{\pi^2}(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h})^2$。$A$ 代表真实框，B 代表预测框，A_{ctr} 和 B_{ctr} 代表两个框的中心点距离 $\rho^2(A_{ctr},B_{ctr})$ 为中心点的欧氏距离，c 为外接最大矩阵的对角线长度。

但是如果预测框和真实框的长宽比是相同的，其中 v 恒为 0，则长宽比的惩罚项 α 也恒为 0。在 $v = \frac{4}{\pi^2}(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h})^2$ 中分别对 w 和 h 求偏导数得到结果如下：

$$\frac{\partial v(w,h)}{\partial w} = \frac{-8h^2}{\pi^2(h^2+w^2)}(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h})$$

$$\frac{\partial v(w,h)}{\partial h} = \frac{8h^2}{\pi^2(h^2+w^2)}(\tan^{-1}\frac{w^{gt}}{h^{gt}} - \tan^{-1}\frac{w}{h})$$

发现这两个梯度是一对相反数，即预测框的 w 和 h 不能同时增大或减小，这可能会导致难以适应真实框的比例变化。

因此使用一种新的边界框 VCIoU 函数[24]，如式(16)所示。

$$VCIoU = IoU - \left|\ln(\frac{d_1+d_2}{c})\right| \tag{16}$$

VCIoU 损失函数如式(17)所示。

$$Loss_{VCIoU} = 1 - VCIoU = 1 - IoU + \left|\ln\left(\frac{d_1+d_2}{c}\right)\right| \tag{17}$$

VCIoU 总结了 DIoU 中提到的回归损失中需要考虑的 3 个几何指标：预测框与目标框的重叠、中心点距离、宽高比的一致性，使用目标框的左上顶点和预测框的中心点的欧式距离以及目标框的中心点和预测框的右下顶点的欧式距离作为基准代替了 DIoU 中的中心点之间的距离，VCIoU 可以缩短预测框与目标框之间的距离，另一方面可以调整预测框的大小，实现预测框与目标框长宽比的一致。如图 9 所示 VCIoU 示意图，黑边实线的为真实框，黑边虚线的为预测框。

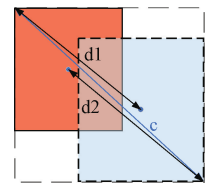


图 9　VCIoU 中的欧式距离

Fig. 9　Euclidean distance in VCIoU

图 10 为在相同 IoU 下，不同位置和不同的长宽比的 VCIoU 损失值，可以看出，VCIoU 可以有效区分预测框与目标框的位置、重叠和纵横比。

由图 10(a)和(b)可知，GIoU 对于两个框重叠时，无法反映真实框的位置。由图 10(b)和(c)可知，DIoU 对于两个框重叠时，无法反映真实框的长宽比。由图 10(d)和(e)可知，对于真实框和预测框中心点重合的情况，GIoU 和 DIoU 均为 IoU 的值，无法反映真实框的长宽比。然而，引入的 VCIoU 可以有效区分预测框与目标框的位置、重叠情况及纵横比。

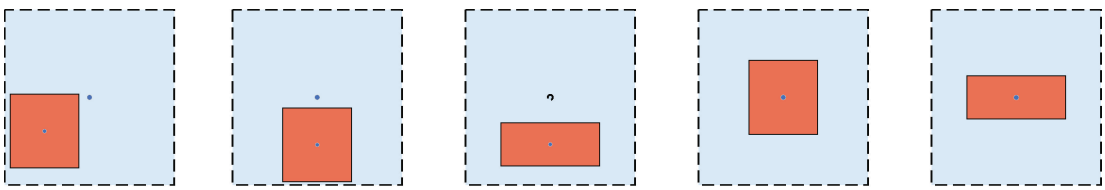


图 10　不同情况下的不同损失函数值

Fig. 10　Different loss function values in different cases

3　实验与结果分析

3.1　训练环境和数据集

实验运行环境为 Linux 操作系统，CPU 为 AMD EPYC 7302，内存为 64 GB；GPU 为 NVIDIA GeForce RTX 3090，显存为 24 GB；深度学习框架 Pytorch 版本为 2.0.0，CUDA 版本为 11.8，Python 版本为 3.8。参数设置如表 2 所示。

表 2 训练参数设置

Table 2 Training parameter setting

Parameters	Setting
Epoch	200
Batch_size	6
Optimizer	SGD
Input_size	640×640
Lr0	0.01

实验所选用的数据集是东北大学宋克臣团队制作的 NEU-DET 数据集,此数据集总共 1 800 张钢材表面缺陷图像,在模型训练过程中将数据集分为训练集、验证集、测试集,比例为 8 ∶ 1 ∶ 1。其中的缺陷包含有裂纹(crazing, Cr)、夹杂(inclusion, In)、斑块(patches, Pa)、划痕(scratches, Sc)、麻点(pitted_surface, Ps)、氧化铁皮(rolled-in_scale, Rs)等 6 类缺陷类型。

3.2 模型评估指标

为了评价改进 YOLOv5s 算法在 NEU-DET 数据集上的缺陷检测效果,决定选用的评价指标包括检测精度(precision, P)、召回率(recall, R)、平均精度(average precision, AP)、平均精度均值(mean average precision, map)、每秒帧数(frames per second, FPS)等。其计算公式如(18)~(21)所示。

$$Precision = \frac{TP}{TP + FP} \times 100\% \tag{18}$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \tag{19}$$

$$AP = \int_0^{Recall} (Precision)\, d_{Recall} \tag{20}$$

$$map = \frac{\sum_{i=1}^{k} AP_i}{k} \tag{21}$$

如图 11 所示,在训练的前 25 个 Epochs 中,改进 YOLOv5s 模型检测平均精度 mAP 快速上升,迭代轮次达到 125 轮后,模型 mAP 达到 0.8 左右,随后模型的 mAP 值趋于稳定,经多次实验验证,模型训练 200 个 Epoch 能够表现出最佳的置信度,所以在实验中设置 200 轮次训练。

改进模型效果明显,由图 12 损失函数变化所示,模型在 25 轮后达到稳定的收敛区间,其中,图 12(a)边界框损失稳定在 0.04 以下,图 12(b)分类损失稳定在 0.004 以下,经历 150 轮训练后,边界框损失稳定在 0.03 以下,分类损失稳定在 0.001 左右。所以边界框以及分类回归精度均优于原 YOLOv5s 模型。

3.3 消融实验

为了分析本算法各个改进点对钢铁表面缺陷检测的有效性,在 YOLOv5s 的基础上设计了消融实验,即将改进的模块分别加入到 YOLOv5s 中,对每个模型得到的实验结果进行分析,如表 3 所示。

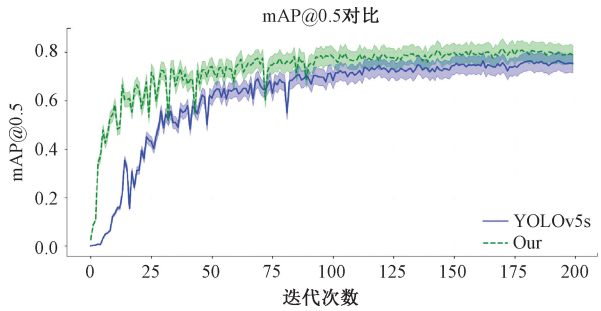
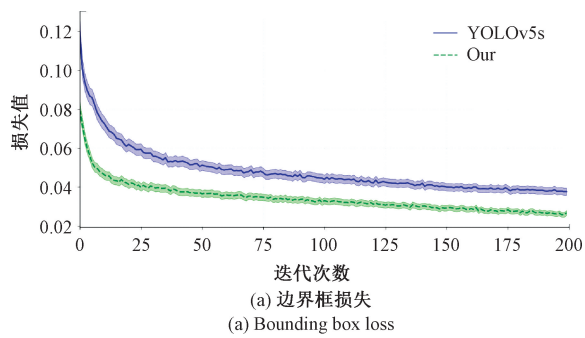


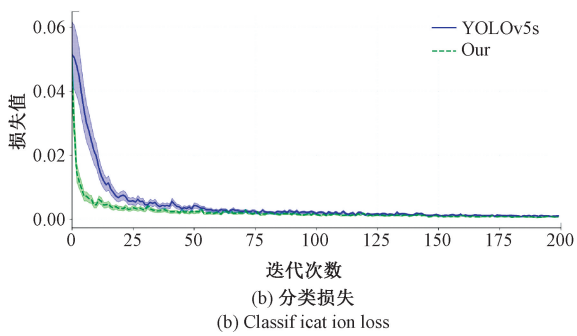
图 11 改进模型与原模型 mAP 曲线对比图

Fig. 11 mAP curve comparison between the proposed model and the original model



(a) 边界框损失

(a) Bounding box loss



(b) 分类损失

(b) Classif icat ion loss

图 12 损失对比图

Fig. 12 Loss comparison plot

表 3 中“√”表示实验算法中使用了对应的模块,“×”表示算法中没有使用对应的模块。消融实验结果表明,采用改进后的网络模型,mAP 从 76.0%提升至了 81.4%。但是模型检测速度下降,说明模型精度有大幅度的提升,但是所改进的模块增加了网络的计算量。在替换了骨干网络后,模型的 mAP 值由 76.0%提升至 76.5%,并且相比于原模型的检测速度 112.8 fps,模型的检测速度达到了 121.4 fps。在特征融合层引入 CBME_C2f 模块后,丰富了梯度流信息,模型能够根据反向传播更好的学习特征,mAP 值达到了 77.7%。在模型的骨干层和特征融合层引入注意力机制,mAP 值从 77.7%提升

至 78.1%,说明注意力机制提升了模型对空间、通道以及位置的感知能力。Involution 新型卷积结构的引入进一步提升了模型的检测精度,增强了模型的表达能力。在激活函数改变为 MetaAconC 后,检测精度由 78.3%提升至 79.6%,虽然使网络趋于复杂,但是相比于原始 SiLU 激活函数能够在卷积操作时,灵活选择线性或非线性形态,使模型表达能力增强。在引入一种新的边界框损失函数 VCIoU,提升了边界框回归精度,同时 mAP 的值达到了 81.4%。

表 3　消融实验结果(注:粗体部分为最优值)

Table 3　Results of ablation experiment (Note: the black part is the optimal value)

HGnetv2	CBME_C2f	Attention	Involution	MeAconC	VCIoU	*AP*/%						*mAP*@0.5/%	*FPS*/fps
						CR	IN	PA	PS	RS	SC		
×	×	×	×	×	×	37.3	91.5	82.0	87.0	68.7	89.6	76.0	112.8
√	×	×	×	×	×	42.4	92.8	80.6	87.2	64.6	91.4	76.5	**121.4**
√	√	×	×	×	×	45.8	92.1	82.9	85.3	67.2	92.9	77.7	117.5
√	√	√	×	×	×	38.5	91.9	82.0	91.4	71.3	93.2	78.1	110.2
√	√	√	√	×	×	41.1	93.2	80.6	**92.5**	69.9	92.2	78.3	111.7
√	√	√	√	√	×	46.9	91.7	**83.4**	87.7	72.5	**95.2**	79.6	83.5
√	√	√	√	√	√	**55.4**	**93.9**	81.8	84.7	**78.5**	94.1	**81.4**	80.6

3.4　对比实验

为了验证改进 YOLOv5 算法在检测精度的优势,在此章节开展一系列不同算法在 NEU-DET 数据集上的检测效果对比实验。

首先,为了进一步验证 CBME_C2f 模块对特征提取的改进效果,实验中采用了 GradCAM[25] 进行热力图可视化分析。首先比较了 C3 和 CBME_C2f 模块在特征融合层的检测表现,如表 4 所示。实验结果显示,将特征融合层的 C3 模块替换为改进后的 CBME_C2f 模块后,在精度上有 1.1%的提升,优于原模型。

表 4　特征提取模块对比表(注:粗体部分为最优值)

Table 4　Feature extraction module comparison table (Note: the best value is in bold)

Model	Backbone	*AP*/%						*mAP*@0.5/%	*FPS*/fps
		CR	IN	PA	PS	RS	SC		
YOLOv5s+C3_BottleNeck	CSPDarkNet53	**40.8**	91.4	81.3	85.8	70.5	90.5	76.7	**108.6**
YOLOv5s+CBME_C2f	CSPDarkNet53	36.1	**93.0**	**83.1**	**85.9**	**75.1**	**93.5**	**77.8**	105.2

实验通过 GradCAM 生成的热力图 13 显示,CBME_C2f 模块在处理 3 种缺陷(inclusion、scratches、patches)时的激活区域更广泛,尤其是在高贡献区域(红色和黄色)上比 C3 模块覆盖更多图像区域。在 Inclusion 缺陷上,CBME_C2f 不仅在核心区域表现出高激活,还扩展至边缘和周围区域。在 Scratches 和 Patches 缺陷上,该模块能够捕捉划痕和复杂形状的细节和边缘,展示了更强的特征捕捉能力。相比之下,C3 模块的激活集中在核心区域,梯度流动受限,导致模型更多依赖局部特征。实验表明,CBME_C2f 模块通过更强的梯度传递和特征捕捉能力,提升了模型的检测精度。

在钢铁缺陷检测中,Crazing(细裂纹或龟裂)通常被视为小缺陷类型。这是因为 Crazing 通常表现为表面上的微小裂纹,这些裂纹的尺寸较小且密集。基于 HGnetv2 和注意力机制相结合的模型结构,旨在提升对微小缺陷的检测能力。HGnetv2 通过堆叠多个 HGBlock 来逐层提取多尺度特征,能够同时捕捉图像中的全局结构信息和细节纹理。为了进一步增强模型对微小特征的感知能力,网络中引入了 CBAM 和 CA 注意力模块。表 5 显示了针对小缺陷检测的对比实验结果,实验表明,将 YOLOv5s 的骨干网络替换为 HGnetv2 后,A 模型在钢铁缺陷数据集上对小缺陷 Crazing 的检测精度和召回率都有提升。B 模型在此基础上融入了 CBAM 模块,通过通道注意力增强了对小目标相关特征的关注,空间注意力则精细定位目标区域。C+CA 模型在特征融合层引入 CA 注意力模块,CA 模块通过坐标分解,能更高效地捕捉全局依赖关系,尤其是小缺陷与全局上下文的联系,提升检测效率和精度。

实验中选用了 YOLOv5s、SSD[26]、Faster R-CNN[20]、YOLOv3[27] 以及基于 Transformer 的端到端检测算法 RT-DETR[10]、GE-YOLOv5s[28]、YOLOv5-CD[29] 等共 7 种目标检测主流算法在相同数据集和相同训练环境下进行对比实验,结果如表 6 所示。

表 5 小缺陷检测对比表(注:粗体部分为最优值)

Table 5 Small defect detection comparison table (Note: the best value is in bold)

Model	Backbone	*AP*/%	*Precision*/%	*Recall*/%	*mAP*@0.5:0.95/%
		Crazing			
YOLOv5s	CSPDarkNet53	37.3	45	21.5	14.2
A	HGnetv2	42.4	56.1	**45.3**	16.1
B	HGnetv2+CBAM	46.6	58.8	41	16.9
C+CA	HGnetv2+CBAM	**47.6**	**61.9**	31.9	**17.4**

(注:模型 C+CA 是在模型 B 的基础上在特征融合层添加 CA 注意力机制)

(Note: Model C+CA is based on model B by adding CA attention mechanism in the feature fusion layer.)

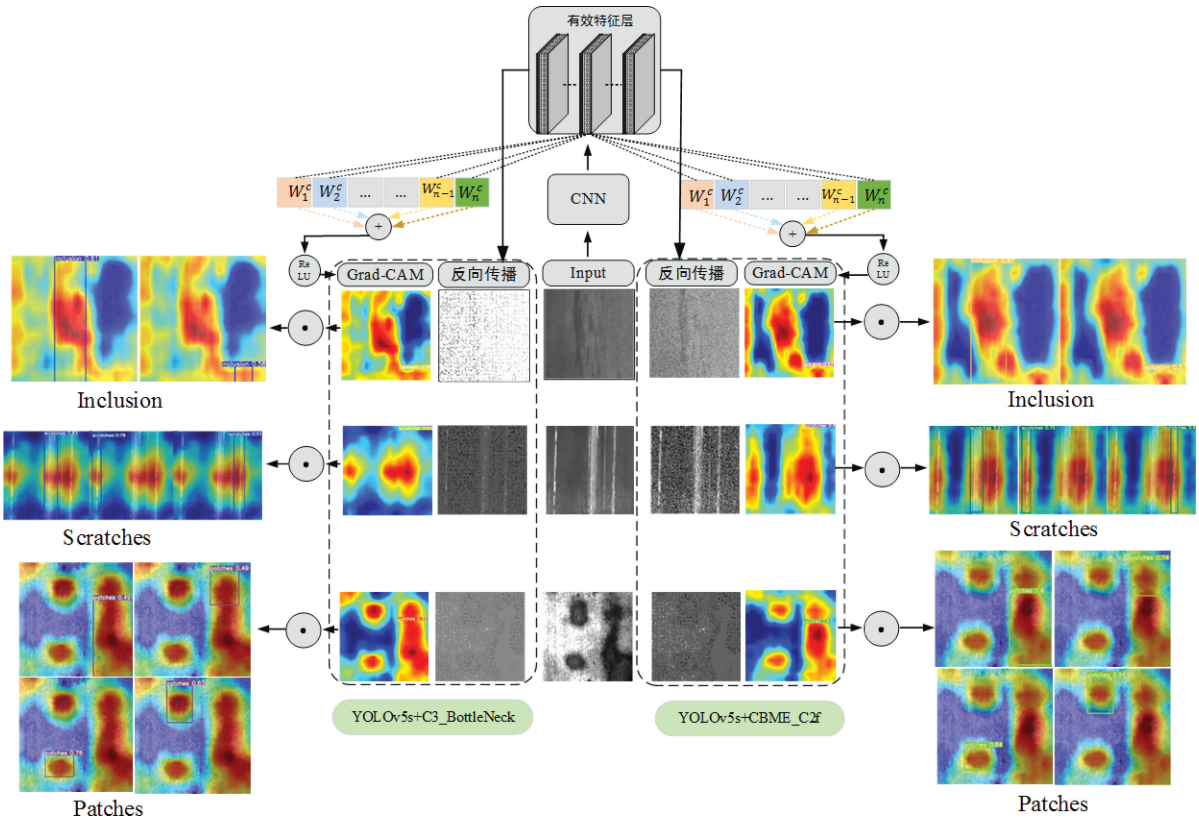


图 13 C3_BottleNeck 与 CBME_C2f 模块热力图分析

Fig. 13 Thermal diagram analysis of C3_BottleNeck and CBME_C2f module

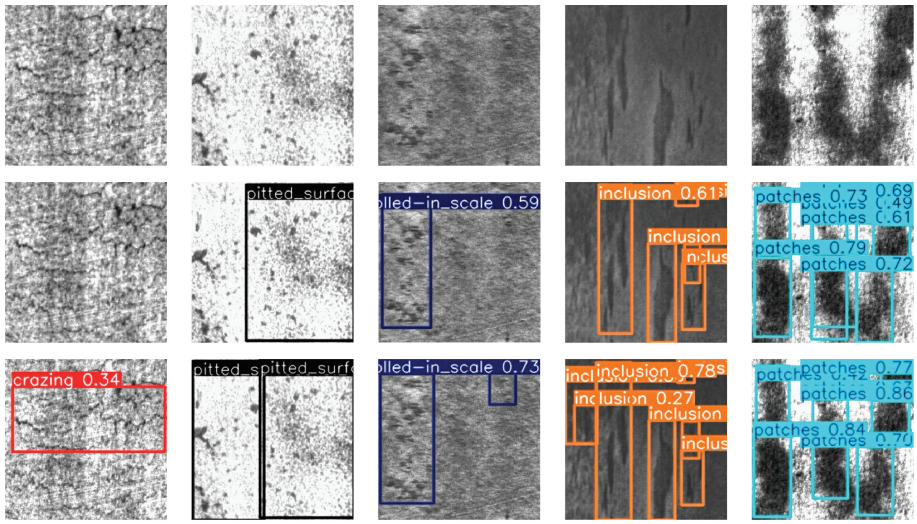
表 6 各模型检测性能对比(注:粗体部分为最优值)

Table 6 Comparison of the detection performance of each model (Note: The bolded part is the optimal value)

Model	Backbone	*AP*/%						*mAP*@0.5/%	*FPS*/fps
		CR	IN	PA	PS	RS	SC		
Faster R-CNN	ResNet-50	39.2	82.3	80.7	85.6	71.2	90.6	74.9	28.1
SSD	VGG16	29.8	73.7	80.8	88.3	61.5	70.8	67.5	**133.8**
YOLOv3	DarkNet53	33.8	75.4	77.8	80.6	65.2	91.2	70.7	80.2
YOLOv5s	CSPDarkNet53	37.3	91.5	82.0	87.0	68.7	89.6	76.0	112.8
RT-DETR-L	HGnetv2	28.8	81.4	91.8	80.0	60.4	89.9	72.0	118.9
GE-YOLOv5s	GhostCSPDarkNet53	49.1	89.7	**92.1**	82.9	73.9	88.6	79.4	109.5
YOLOv5-CD	改进 CSPDarkNet53	47.9	90.3	89.6	**90.1**	65.3	**95.6**	79.8	104.7
Our	改进 HGnetv2	**55.4**	**93.9**	81.8	84.7	**78.5**	94.1	**81.4**	80.6

由表 6 可以知道,改进后模型的平均精确度评估指标 mAP@0.5(%)在所有模型中为最高值,高达 81.4%,其中缺陷 Crazing、Inclusion、Rolled-in_scale 等 3 种缺陷的检测精度均是最优。相较于模型 Faster R-CNN、SSD、

YOLOv3、YOLOv5s、RT-DETR-L、GE-YOLOv5s、YOLOv5-CD 的平均精确度指标 mAP@0.5，分别提高了 6.5%、13.9%、10.7%、5.4%、9.4%、2%、1.6%。其中模型 GE-YOLOv5s 和 YOLOv5-CD 的检测精度分别排名第 3 和第 2，说明改进 YOLOv5s 算法相比于目前流行的检测算法更优，有更高的检测精度。改进算法的检测速度虽低于 Faster R-CNN 和 YOLOv3 外的其他算法，但仍满足工业实时检测的要求。此外，改进的 YOLOv5s 算法在小缺陷 Crazing 的检测上相较于其他算法有显著提升，AP 值达到 55.4%，表明该算法能有效缓解小缺陷检测精度不足的问题。将原始 YOLOv5 算法和改进算法作检测效果对比，具体如图 14 所示。其中第 1 行为原始图像，第 2 行为原 YOLOv5s 模型的检测效果，第 3 行为改进模型的检测效果，每一列对应不同类型的缺陷。



(a) 细裂纹 (a) Crazing　(b) 凹陷与坑洞 (b) Pitted_surface　(c) 氧化铁锈 (c) Rolled-in_scale　(d) 杂质物 (d) Inclusion　(e) 斑块 (e) Patches

图 14　检测结果对比

Fig. 14　Comparison of test results

4　结　论

钢材表面缺陷具有多尺度、多类型和复杂背景的特点，当前主流缺陷检测算法存在易受背景干扰、检测精度低、小目标缺陷易漏检等不足。针对上述问题，改进的 YOLOv5 算法首先融合 HGnetv2 骨干网络与注意力机制，显著提升了对小缺陷特征的识别能力。其次，引入改进的 CBME_C2f 模块替代原 C3 模块，通过提升梯度流信息的丰富性来增强模型的学习能力。进一步地，使用 MetaAconC 激活函数，自适应调整激活函数的形状及选择性地激活特征，以提升模型在复杂背景中准确识别缺陷的能力。然后，引入 Involution 操作，使特征空间中的每个像素点成为可训练参数，增强了模型的空间感知能力。最后，采用新的边界框损失函数 VCIoU，提高了边界框回归的效率和检测精度。将改进后的 YOLOv5 算法在 NEU-DET 数据集上与原算法进行对比实验，平均检测精度提升 5.4%。并且，该算法相较于 Faster R-CNN、SSD、YOLOv3、RT-DETR 算法以及 GE-YOLOv5s、YOLOv5-CD 算法都有更加高的检测精度，对于难以检测的 Crazing 缺陷相比于原 YOLOv5s 模型的 AP 值提升了 18.1%。但是该算法的检测速度仍有进一步优化空间，未来将在缺陷检测模型轻量化方面开展研究，力求在模型参数缩减的情况下，保持检测精度和提高检测速度。

参考文献

[1] 费佳杰，李宏胜，任飞，等. 基于深度学习的钢表面缺陷检测方法综述[J]. 现代信息科技，2023，7(19)：107-112.
FEI J J，LI H SH，REN F，et al. Steel surface defect detection method based on the deep learning review[J]. Journal of Modern Information Technology，2023，7(19)：107-112.

[2] 赵佰亭，张晨，贾晓芬. ECC-YOLO：一种改进的钢材表面缺陷检测方法[J]. 电子测量与仪器学报，2024，38(4)：108-116.
ZHAO B T，ZHANG CH，JIA X F. ECC-YOLO：An improved method for detecting steel surface defects[J]. Journal of Electronic Measurement and Instrumentation，2024，38(4)：108-116.

[3] 马志程，李丹，张宝龙. 基于改进 Mask R-CNN 的光学元件划痕缺陷检测研究[J]. 电子测量与仪器学报，2023，37(4)：231-239.

MA ZH CH, LI D, ZHANG B L. Research on optical element scratch defect detection based on improved Mask R-CNN [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(4): 231-239.

[4] 石欣, 卢灏, 秦鹏杰, 等. 一种远距离行人小目标检测方法[J]. 仪器仪表学报, 2022, 43(5): 136-146.
SHI X, LU H, QIN P J, et al. A long-distance pedestrians small target detection method [J]. Chinese Journal of Scientific Instrument, 2022, 43(5): 136-146.

[5] 姜媛媛, 蔡梦南. 轻量化的印刷电路板缺陷检测网络 Multi-CR YOLO[J]. 电子测量与仪器学报, 2023, 37(11): 217-224.
JIANG Y Y, CAI M N. Lightweight printed circuit board defect detection network Multi-CR YOLO[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(11): 217-224.

[6] 张福豹, 吴婷, 赵春峰, 等. 基于弱光增强与 YOLO 算法的锯链缺陷检测方法[J]. 电子测量技术, 2024, 47(6): 100-108.
ZHANG F B, WU T, ZHAO CH F, et al. Saw chain defect detection method based on low light enhancement and YOLO algorithm [J]. Electronic Measurement Technology, 2024, 47(6): 100-108.

[7] ZHENG J, WU H, ZHANG H, et al. Insulator-defect detection algorithm based on improved YOLOv7 [J]. Sensors, 2022, 22(22): 8801.

[8] LIN 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.

[9] LI H, XIONG P, AN J, et al. Pyramid attention network for semantic segmentation [J]. ArXiv preprint arXiv: 1805.10180, 2018.

[10] ZHAO Y AN, LYU W Y, XU SH L, et al. Detrs beat yolos on real-time object detection [J]. ArXiv preprint arXiv:2304.08069, 2023.

[11] LIU ZH, MAO H Z, WU CH Y, et al. A convnet for the 2020s [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 11976-11986.

[12] MA N, ZHANG X, LIU M, et al. Activate or not: learning customized activation [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 8032-8042.

[13] RAMACHANDRAN P, ZOPH B, LE Q V. Searching for activation functions [J]. ArXiv preprint arXiv: 1710.05941, 2017.

[14] GOODFELLOW I, WARDE-FARLEY D, MIRZA M, et al. Maxout networks [C]. Proceedings of the 30th International Conference on Machine Learning, PMLR, 2013, 28(3):1319-1327.

[15] HE K M, ZHANG X, REN SH Q, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification [C]. Proceedings of the IEEE International Conference on Computer Vision, 2015: 1026-1034.

[16] MAAS A L, HANNUN A Y, NG A Y. Rectifier nonlinearities improve neural network acoustic models [C]. Proceedings of the International Conference on Machine Learning, 2013, 30(1): 3.

[17] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision, 2018: 3-19.

[18] 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.

[19] LI D, HU J, WANG C, et al. Involution: Inverting the inherence of convolution for visual recognition [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 12321-12330.

[20] REN SH 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, 2016, 39(6): 1137-1149.

[21] YU J, JIANG Y, WANG Z, et al. Unitbox: An advanced object detection network [C]. Proceedings of the 24th ACM International Conference on Multimedia, 2016: 516-520.

[22] 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.

[23] ZHENG ZH H, 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.

[24] YANG H, FANG Y, LIU L, et al. Improved YOLOv5 based on feature fusion and attention mechanism and its application in continuous casting slab detection [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-16.

[25] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: Visual explanations from deep networks via gradient-based localization [J]. International Journal of Computer Vision, 2020, 128: 336-359.

[26] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[27] REDMON J, FARHADI A. YOLOv3: An incremental improvement [J]. ArXiv preprint arXiv: 1804.02767, 2018.

[28] 叶志宇, 贾晓芬, 王天奇. 面向煤矸识别的目标检测算法[J]. 电子测量与仪器学报, 2024, 38(8): 145-152.
YE ZH Y, JIA X F, WANG T Q. Target detection algorithm for coal waste identification [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 145-152.

[29] WANG B, WANG M, YANG J, et al. YOLOv5-CD: strip steel surface defect detection method based on coordinate attention and a decoupled head [J]. Measurement: Sensors, 2023, 30: 100909.

作者简介

张航,2023 年于武汉东湖学院获得学士学位,现为武汉科技大学硕士研究生,主要研究方向为缺陷检测、机器视觉。
E-mail: 2692574962@qq.com

Zhang Hang received his B. Sc. degree from Wuhan Donghu University in 2023. Now he is a M. Sc. candidate of Wuhan University of Science and Technology. His main research interests include defect detection and machine vision.

周毅(通信作者),2016 年于武汉大学获得博士学位,曾为英国考文垂大学的访问学者,目前是武汉科技大学副教授,主要研究方向为缺陷检测、深度强化学习与过程控制系统应用。
E-mail: zhouyi83@wust.edu.cn

Zhou Yi (Corresponding author) received his Ph. D. degree from Wuhan University in 2016. He was a visiting scholar at Coventry University in UK. He is now an associate professor at Wuhan University of Science and Technology. His main research interests include defect detection, deep reinforcement learning, and process control system applications.

邱宇峰,2002 年于武汉科技大学获得学士学位,2008 年于武汉科技大学获得硕士学位,现为宝信软件(武汉)有限公司工程师,主要研究方向为缺陷检测、故障诊断与过程控制系统应用。
E-mail: 54821556@qq.com

Qiu Yufeng received his B. Sc. degree from Wuhan University of Science and Technology in 2002, and M. Sc. degree from Wuhan University of Science and Technology in 2008, respectively. Now he is an engineer at Baosight Software (Wuhan) Co., Ltd. His main research interests include defect detection, fault diagnosis, and process control system applications.