

基于自注意特征融合的钢材表面小目标缺陷检测<sup>\*</sup>冯夫健<sup>1,2</sup> 罗太维<sup>1,2</sup> 谭 棉<sup>1,2</sup> 汪小梅<sup>1,2</sup> 王岳继<sup>1,2</sup>

(1. 贵州民族大学数据科学与信息工程学院 贵阳 550025;

2. 贵州民族大学贵州省模式识别与智能系统重点实验室 贵阳 550025)

**摘 要:** 针对钢材表面小目标缺陷占比小,对比度低,导致钢材表面小目标缺陷检测模型提取丰富缺陷特征失效的问题。基于联系上下文信息和增强特征融合之间的关系,对钢材表面小目标缺陷检测问题提出以下解决方案:首先,结合滑动窗口机制 Swin Transformer,利用分层结构和局部窗口整合不同特征块的特征信息,以降低卷积操作密集性的基础上增强小目标缺陷特征信息的对比度;其次,采用坐标注意力机制使模型获得更多的位置信息,以增强小目标缺陷特征信息的多样性;最后结合具有丰富梯度流信息的特征融合模块 CSP-FCN,提出了基于自注意特征融合的钢材表面小目标缺陷检测模型 SFNet,该模型将不同尺度特征融合以产生丰富的语义信息,增强钢材表面小目标缺陷的特征表达能力。实验结果表明,SFNet 在 NEU-DET 和 GC10-DET 公开数据集上的检测性能优于目前经典的目标检测模型。此外,所提模型在参数量减少为原来 1/2 的基础上平均精度值分别提升了 3% 和 3.7%。

**关键词:** 钢材表面缺陷;小目标缺陷检测;Swin Transformer;位置信息;特征融合

**中图分类号:** TP391.41; TN919.8 **文献标识码:** A **国家标准学科分类代码:** 460.4030

Defect detection of small targets on steel surface based on  
self-attention feature fusionFeng Fujian<sup>1,2</sup> Luo Taiwei<sup>1,2</sup> Tan Mian<sup>1,2</sup> Wang Xiaomei<sup>1,2</sup> Wang Yueji<sup>1,2</sup>

(1. School of Mathematical Sciences and Information Engineering, Guizhou Minzu University, Guiyang 550025, China;

2. Key Laboratory of Pattern Recognition and Intelligent Systems of Guizhou Province, Guizhou Minzu University, Guiyang 550025, China)

**Abstract:** In addressing the issue of ineffective extraction of rich defect features for small-scale surface defects on steel due to their low contrast and small proportion, this paper proposes a solution for small-target defect detection. Leveraging the relationship between contextual information integration and enhanced feature fusion, we introduce the following approaches: incorporating the sliding window mechanism Swin Transformer, which integrates feature information from different blocks hierarchically and through local windows to enhance the contrast of defect features while reducing convolutional operation density; the model employs Coordinate Attention to obtain more positional information, enhancing the diversity of features related to small-target defects. Additionally, we propose the steel surface small-target defect detection model SFNet based on self-attention feature fusion, integrating features with richer semantic information across different scales using the CSP-FCN feature fusion module. Experimental results demonstrate that SFNet achieves superior detection performance on the NEU-DET and GC10-DET public datasets compared to current classic object detection models. Furthermore, the proposed model achieves an average precision improvement of 3% and 3.7%, respectively, while reducing the parameter count to half of its original size.

**Keywords:** steel surface defects; small target defect detection; Swin Transformer; position information; feature fusion

## 0 引 言

随着我国工业水平的不断发展,钢材作为航空航天、机

械制造和国防工业等领域的重要原材料,其生产技术取得了巨大进步。然而,在钢材的生产制造过程中,由于原料质量、制造设备和生产环境等因素的影响,钢材表面可能会出

收稿日期:2024-07-26

<sup>\*</sup> 基金项目:贵州省科技计划项目(QKHJCZK2022YB195, QKHJCZK2023YB143, QKHPTRCZCKJ2021007, QKHJCZK2022YB197)、贵州省教育厅自然科学研究项目(QJJ2022015, QJJ2023061, QJJ2023012, QJJ2022047, QJJ2024063)、贵州民族大学博士研究启动项目(GZMUZK[2024] QD04)、贵州民族大学基金科研项目(GZMUZK[2023]YB14)、贵州省模式识别与智能系统重点实验室开放课题(GZMUKL[2022]KF01)资助

现裂纹、凹坑、划痕等各类小目标<sup>[1]</sup>缺陷。这些小目标缺陷因其占比小、对比度低,往往在人工质检过程中被忽略,导致钢材的强度和耐磨性降低,影响钢材的正常使用,甚至引发安全事故<sup>[2]</sup>。此外,钢材表面小目标缺陷的特征通常较为微小且不明显,在复杂背景下容易被忽略,导致现有的表面缺陷检测模型难以应对此类缺陷,因此,工业表面小目标缺陷检测日益成为国内外研究人员关注的热点。

目前,钢材表面缺陷检测方法可归纳为基于传统的图像处理和基于深度学习的机器视觉检测方法。在传统检测方法方面,图像处理技术在简单场景中获得了良好的效果,例如 Sun 等<sup>[3]</sup>提出了一种基于小波的各向异性扩散滤波算法,该算法可以有效地过滤掉冷轧带钢中不需要的纹理背景,并保留有价值的详细信息。Cao 等<sup>[4]</sup>提出了一种改进的最小二乘法,该方法利用状态转移和迭代算法实现镀镍孔带钢的缺陷检测。虽然上述基于传统的检测方法在缺陷面积较大、前背景易区分等简单场景下获得了较高的检测精度,但通常使用低级特征<sup>[5]</sup>(如边缘信息)进行小目标缺陷检测,其性能较差,而且依赖于人工设计特征提取规则。

近几年来基于深度模型的机器视觉检测方法由于其强大的特征自动提取能力受到了国内外研究者的广泛关注,并在先进制造、航空航天、军事科技等与国计民生紧密相关的诸多领域有了一定的研究成果。随着钢材表面缺陷所呈现的占比小、对比度低的趋势日益显著,许多基于深度学习的表面缺陷检测方法也在这样的背景下不断涌现,具体可分为两类:单阶段目标检测算法和两阶段目标检测算法。

两阶段目标检测算法主要包括 Faster R-CNN 和 Mask R-CNN,而单阶段目标检测算法主要包括 SSD 和 YOLO 系列。前者在检测过程中会生成一系列候选框,后者则将目标检测直接转化为回归问题进行处理。因此,两阶段目标检测算法虽然具有相对较高的检测精度,但其检测速度相对于单阶段目标检测算法较慢。所以,单阶段目标检测算法更贴近于钢材表面缺陷检测任务。如 Yang 等<sup>[6]</sup>在 SSD 中引入了扩展卷积和残差注意网络,以改善模型的感受野,使模型更加关注目标。Li 等<sup>[7]</sup>提出一种改进的 YOLO 算法,用于端到端的表面缺陷检测,极大提升了模型的速度与准确度。Cheng 等<sup>[8]</sup>改进了 YOLOv3 算法,采用 K-means++ 聚类算法优化聚类中心,并融合新增特征图层,提高模型对金属表面缺陷的检测精度,但算法需预先设置锚框尺寸参数,影响了检测速度。Kou 等<sup>[9]</sup>基于 YOLOv3 进行改进,使用无锚框(anchor-free)的方式提升模型的速度,设计密集卷积块(dense block)用于提取更丰富的特征信息,从而提升模型的准确率与鲁棒性。Xu 等<sup>[10]</sup>在 Mask R-CNN 中引入了特征金字塔和边缘检测分支,有效地提高了缺陷检测的准确率。Song 等<sup>[11]</sup>基于改进的 U-net 设计了一种检测金属部件表面弱划痕的方法,该方法将特征通道增加了一倍,简化了原有的网络卷积结构。该网络是一个典型的编码器-解码器体系结构,通过连续的

池化操作来改善空间上下文。但是,过多的下采样操作会导致详细的空间信息丢失,不利于获得完整的划痕区域。Zou 等<sup>[12]</sup>提出了一种基于 SegNet 的裂缝检测方法,该方法将编码器和解码器的多阶段特征图结合起来,生成不同比例尺的特征图。同时,将各比例尺的特征图进行连接,得到裂纹信息。综上所述,基于深度学习的小目标缺陷检测方法虽然实现了特征的自动提取,但通常单一利用上下文信息<sup>[13-14]</sup>、特征融合<sup>[15]</sup>、生成增强特征<sup>[16-17]</sup>等方式对小目标缺陷进行检测,尽管性能得到一定的增强,但大部分忽略了彼此之间的联系,导致其面对钢材表面小目标缺陷存在占比小、对比度低时,造成检测性能不佳。

因此,针对钢材表面小目标缺陷因占比小、对比度低导致提取丰富缺陷特征失效的问题,本文提出了基于自注意特征融合的钢材表面小目标缺陷检测模型(self-attention feature fusion, SFNet)。该模型在 YOLOv5 的基础上,利用滑动窗口机制在全局范围内处理输入图像,在降低卷积操作密集性的基础上通过减少缺陷特征信息的丢失以增强对比度,并采用坐标注意力机制(coordinate attention, CA)使模型有效关注小目标缺陷的位置信息。此外,该模型结合特征融合模块(faster convolutions of CSP bottleneck, CSP-FCN)将不同尺度特征融合,以产生丰富的语义信息,增强特征表达能力。该模型通过联系上下文信息和增强特征融合之间的关系,在面对钢材表面缺陷检测任务时能够加强模型对小目标缺陷的有效关注,以进一步提高检测精度和效率,为钢材生产质检自动化设备提供有效的算法支撑。本文的主要贡献如下:

1)提出了基于自注意力特征融合的钢材表面小目标缺陷检测模型 SFNet,该模型通过联系上下文信息和增强特征融合之间的关系,以强化小目标缺陷的特征表达能力。通过对比实验发现该模型能够有效检测小目标缺陷。

2)结合滑动窗口机制,使 SFNet 在有效减少特征信息丢失的基础上增强对比度,并通过坐标信息准确定位缺陷位置。通过消融实验发现该模型在检测精度和效率上实现了有效平衡。

## 1 SFNet 模型结构

针对钢材表面缺陷检测模型在捕捉钢材表面小目标缺陷时出现提取丰富的特征失效问题,其本质在于钢材表面小目标缺陷占比小、对比度低,导致模型提取特征时无法准确区分缺陷和背景信息。受滑动窗口启发,假设降低卷积操作的密集性可以减少缺陷特征信息的丢失以增强对比度,以及增强缺陷特征信息的多样性可以提高小占比缺陷的特征表达能力,基于上述假设,本文结合滑动窗口机制 Swin Transformer、坐标注意力机制 CA、梯度流信息更加丰富的特征融合模块 CSP-FCN 设计了基于自注意特征融合的钢材表面小目标缺陷检测模型 SFNet,模型的网络结构如图 1 所示。

由图 1 可知,在主干网络中应用滑动窗口机制(Swin

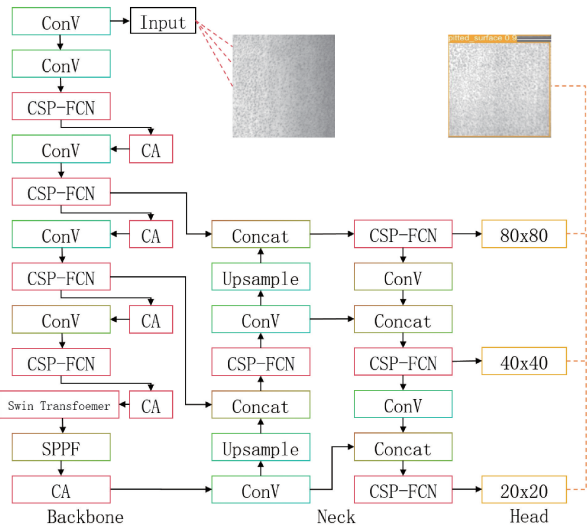


图 1 SFNet 结构

Fig. 1 SFNet structure

Transformer)<sup>[18]</sup>,通过这种机制实现图像特征的全局建模和局部关联计算,能够更有效地增强钢材表面小目标缺陷信息的对比度,并显著降低计算复杂度和模型参数量;然后

采用坐标注意力机制(CA)<sup>[19]</sup>增强模型对小目标缺陷位置的有效关注;此外,结合具有更加丰富的梯度流信息的特征融合模块(CSP-FCN)生成多尺度特征,以综合利用各级特征图的信息,增强特征表达能力,提高模型的检测精度。

### 1.1 滑动窗口机制(Swin Transformer)

通常情况下,使用传统的单阶段目标检测算法进行缺陷检测时,需要对整个图像进行密集的卷积操作,而过多的卷积计算在钢材表面缺陷检测任务中会因为缺陷占比小而丢失特征信息,还会导致计算复杂度和参数量较大。因此,基于降低卷积操作的密集性可以减少小目标缺陷特征信息的丢失以增强对比度的假设,本文在单阶段目标检测算法 YOLOv5s 的主干中结合 Swin Transformer,利用其滑动窗口多头自注意力机制在全局范围内处理图像,这样能够使模型更好地捕捉图像中的长距离依赖关系,降低模型中卷积操作的密集性,从而在减少计算复杂度和参数量的基础上减少特征信息的丢失以增强对比度。Swin Transformer 通过滑动窗口计算进行表示,经过滑动窗口方案将自注意力计算限制在不重叠的局部窗口,同时允许跨窗口连接,从而提高了效率。Swin Transformer 网络结构如图 2 所示。

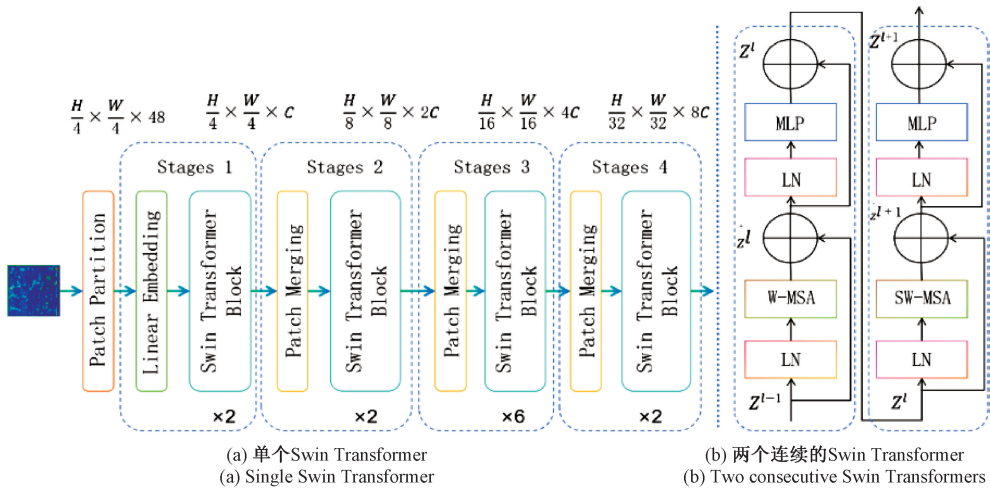


图 2 Swin Transformer 结构

Fig. 2 Swin Transformer structure

由图 2(a)可知,首先将图片输入图像块分层进行分块,然后在通道方向展平,输入大小由  $H \times W \times 3$  变为  $\frac{H}{4} \times \frac{W}{4} \times 48$ ,然后利用线性嵌入层将每个像素的通道数做据线性变换,其中图像块分层、线性嵌入层本质是直接通过一个卷积层实现,其功能和图像块合并层一致,都是对输入特征矩阵进行下采样和通道调整。然后经过 4 个阶段构建大小不一的特征图,除了 1 阶段中先通过一个线性嵌入层外,其余 3 个阶段都是先通过一个图像块合并层进行下采样,然后重复堆叠 Swin Transformer 块,其中 Swin Transformer 块有两种结构,如图 2(b)所示,这两种

结构的差异仅在于一个使用了 W-MSA 结构,而另一个使用了 SW-MSA 结构。W-MSA 即窗口多头自注意力机制模块,主要原理是将特征图划分成一定大小的窗口,然后在每一个窗口内部单独进行自注意力计算,这样做有利于减少计算复杂度,其具体计算公式如下:

$$\Omega(W-MSA) = 4h\omega C^2 + 2M^2h\omega C \quad (1)$$

其中,  $h$  表示特征图的高度,  $\omega$  表示特征图的宽度,  $C$  表示特征图的深度,  $M$  则表示每个窗口的大小。

而 SW-MSA 则表示滑动窗口多头自注意力机制模块,它主要是为了解决采用 W-MSA 结构时只能在每个窗口内进行自注意力计算而导致窗口之间无法进行信息传递的问题,即



通过偏移的 W-MSA 进行不同窗口之间的信息交流,采用上述两种结构,有效减少了全局计算导致像素数量的二次复杂度,使模型更适合钢材表面小目标缺陷检测任务。

1.2 坐标注意力 (coordinate attention, CA)

基于增强缺陷特征信息的多样性可以提高小目标缺陷的特征表达能力这一假设,结合坐标注意力机制 CA,使模型有效关注缺陷位置信息。在钢材表面缺陷检测中,缺陷通常是局部区域,并且钢材表面存在噪声和背景干扰,这些因素可能会干扰模型对缺陷检测的准确性。因此,通过坐标注意力机制 CA 对输入图像中的缺陷位置信息进行学习和利用,使网络重用低层特征,其结构如图 3 所示。

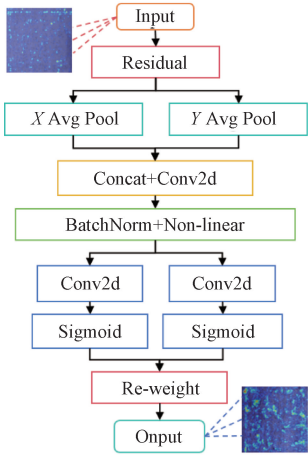


图 3 CA 结构

Fig. 3 CA structure

由图 3 可知,坐标注意力机制是在通道注意力中嵌入了位置信息,使模型能够在更大的区域上进行注意力计算。因此,它除了捕获跨通道的信息外,还捕获方向感知和位置感知的信息,以此提高缺陷特征信息的多样性来帮助模型更加精准地定位和识别钢材表面小目标缺陷。

1.3 特征融合模块 (CSP-FCN)

为使钢材表面小目标缺陷检测模型充分考虑上下文

信息和增强特征融合之间的联系,以保持轻量化的同时拥有更加丰富的语义信息,在网络中将 C3 结构替换为 CSP-FCN 模块。如图 4 所示,CSP-FCN 模块主要用于跨阶段融合特征图,通过将上一阶段的特征图与下一阶段的特征图进行融合,使模型获得更加丰富的梯度流信息并减少计算复杂度。

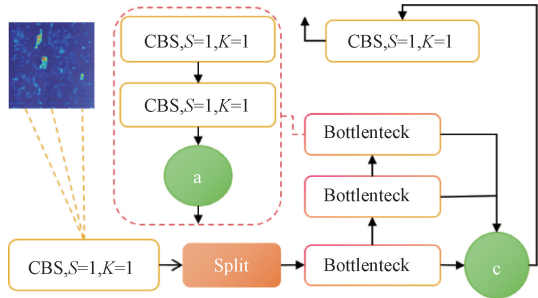


图 4 CSP-FCN 结构

Fig. 4 CSP-FCN structure

CSP-FCN 模块主要是将两个特征图进行拼接,然后通过卷积操作来融合信息。将低分辨率且语义丰富的特征图与高分辨率且语义信息较弱的特征图进行融合,生成多尺度特征图,以此提高模型对钢材表面小目标缺陷的检测性能。

2 实验设置及分析

2.1 数据集

实验所用数据集为东北大学宋克臣老师团队公开的钢材表面缺陷数据集 NEU-DET 以及 GC10-DET。图 5 所示为 NEU-DET 数据集,其中有 6 类缺陷,每类 300 张,共 1 800 张图片,从左至右依次为裂纹(cr)、杂质(in)、斑块(pa)、麻点(ps)、压入氧化皮(rs)及划痕(sc)。实验时,按 8:1:1 的比例随机将数据集分为训练集、验证集和测试集,即 1 440 张图像作为训练集,180 张图像作为验证集,180 张图像作为测试集。

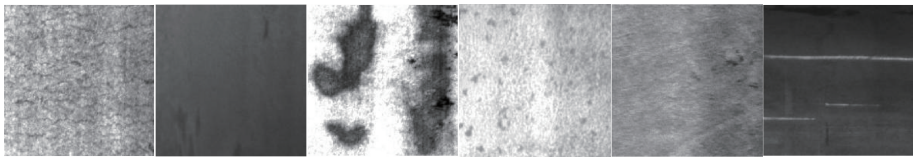


图 5 NEU-DET 数据集钢材表面缺陷

Fig. 5 Steel surface defects of NEU-DET dataset

GC10-DET 数据集共有 2 294 张图片,如图 6 所示,从左至右依次为月牙弯(Cg)、水斑(Ws)、丝斑(Ss)、焊缝(Wl)、折痕(Cr)、油斑(Os)、异物(In)、腰折(Wf)、冲孔(Pu)、轧痕(Rp) 10 类缺陷,并按照上述比例进行随机划分,即有 1 835 张图像作为训练集,229 张图像作为验证集,230 张图像作为测试集。

2.2 参数设置

在 Linux 操作系统下,使用 Jupyterlab-pytorch 在 NVIDIA GeForce RTX 3090 GPU 上,针对 NEU-DET 和 GC10-DET 两种钢材数据集进行表面小目标缺陷检测研究,验证本文所提方法的有效性。模型训练过程中,参数设置均保持原模型参数不变,即学习率为 0.01,动量为



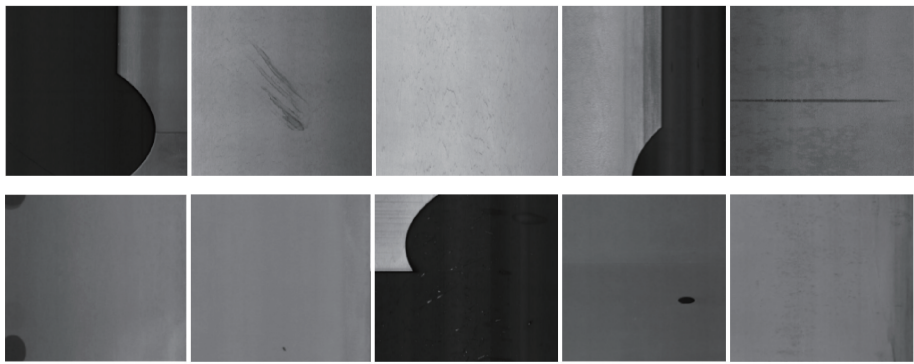


图 6 GC10-DET 数据集钢材表面缺陷

Fig. 6 Steel surface defects of GC10-DET dataset

0.937,权重衰减为 0.000 5,epoch 为 300 次,Batch size = 8,图片大小为  $640 \times 640$ ,并启用余弦退火学习率衰减策略。

2.3 评价指标

本文在模型比较时,从以下几个方面进行了评估:模型参数量 (Params)、浮点运算量 (GFLOPs)、模型体积 (Weight)、精确率  $P$ 、召回率  $R$  以及平均精度值  $mAP$ 。Params 指训练过程中模型的总参数数量。GFLOPs 表示进行浮点运算的次数。模型体积指在训练完成后,生成模型权重的大小。精确率  $P$  是指在所有被模型预测为正例的样本中,实际上是正例的概率。召回率  $R$  是指在所有实际正例样本中,被模型成功预测为正例的概率。 $AP$  代表在不同召回率下的精确率的均值。具体而言,可以画出  $P$ - $R$  曲线,并计算该曲线与坐标轴之间的面积,这个面积即为平均精度  $AP$ ,它反映了模型准确率和召回率的综合性能。最后, $mAP$  是对各个类别的  $AP$  求均值所得到的指标,代表了所有类别的平均精度。具体计算公式如下:

1) 准确率

$$P = \frac{TP}{TP + FP} \tag{2}$$

2) 召回率

$$R = \frac{TP}{TP + FN} \tag{3}$$

3) 平均精度

$$AP = \int_0^1 p(r)dr \tag{4}$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \tag{5}$$

其中,  $TP$  表示模型检测正确的正样本;  $FP$  为检测为错误的正样本;  $FN$  为检测为错误的负样本;  $N$  表示类别数量,  $AP_i$  表示第  $i$  类的检测精度。

2.4 模型检测精度对比

为验证 SFNet 的有效性,将其与目前经典的目标检测模型进行了对比,相关的实验参数设置均保持一致。表 1、2 分别为各模型在 NEU-DET 和 GC10-DET 上的检测精度对比,表 3 为各模型复杂度对比。

表 1 NEU-DET 上的检测精度对比

Table 1 Comparison of detection accuracy on NEU-DET

%

模型	$mAP$	cr	in	pa	ps	rs	sc
EfficientDet <sup>[20]</sup>	70.1	45.9	62.0	83.5	<b>85.6</b>	70.7	73.1
Faster-RCNN <sup>[21]</sup>	72.3	42.9	67.9	84.9	79.1	68.8	89.9
YOLOv3	72.3	36.9	80.2	90.0	71.5	63.3	92.7
YOLOv5	74.4	37.1	84.7	93	75	67	89.3
YOLOv8	76.1	45.7	<b>90.6</b>	85.5	82.5	<b>76.8</b>	<b>95.8</b>
SFNet	<b>77.4</b>	<b>50.8</b>	88.2	<b>96.7</b>	77.8	60.7	90.3

从表 1 可知,EfficientDet 的  $mAP$  只有 70.1%,整体性能表现不佳,说明该模型无法有效识别小目标缺陷;Faster-RCNN 和 YOLOv3 相对于 EfficientDet 而言, $mAP$  值增加了 2.2%,检测精度有一定的提升,但在面对占比小、对比度低这类小目标缺陷时依然无法达到较高的检测

精度。SFNet 的  $mAP$  值较原始模型提升了 3%,表明本文所提方法是有效的。YOLOv8 虽然在杂质、压入氧化皮和划痕上取得了最高的检测精度,但其参数量和模型体积远远大于 SFNet,并且 YOLOv8 的  $mAP$  值比 SFNet 降低了 1.3%,此外,SFNet 在裂纹和斑块两种缺陷类别上有较大

表 2 GC10-DET 上的检测精度对比

Table 2 Comparison of detection accuracy on GC10-DET

模型	<i>mAP</i>	pu	wl	cg	ws	os	ss	in	rp	cr	wf
YOLOv2	43.3	72.5	32.8	81.9	47.6	40.3	47.3	9.6	1.8	19.2	80.5
Faster-RCNN <sup>[22]</sup>	62.7	89.9	55.4	87.2	59.9	65.3	57.9	19.4	<b>36.4</b>	<b>73.6</b>	81.8
EDDN <sup>[23]</sup>	65.2	90.0	88.5	84.8	55.8	62.2	65.0	25.6	<b>36.4</b>	52.1	<b>91.9</b>
YOLOv8	65.3	88	86.3	94.2	77.4	70.3	50.8	<b>39.7</b>	20.7	34.4	91.8
YOLOv5	66.1	<b>97.6</b>	80.1	<b>96.2</b>	78.7	73.5	65.2	22.7	1.32	62.6	83.5
SFNet	<b>69.8</b>	96.8	<b>91.1</b>	95.7	<b>81.1</b>	<b>78.3</b>	<b>70</b>	26.2	1.68	66.1	90.8

表 3 模型复杂度对比

Table 3 Comparison of model complexity

模型	<i>Params</i> /M	<i>GFLOPs</i> /G	<i>Weight</i> /MB
SSD	24.01	63.1	91.6
Faster R-CNN	28.39	77.2	108.4
YOLOv3	61.60	155.2	234.8
YOLOv4	64.03	141.52	244.0
YOLOv5	7.03	15.8	14.5
YOLOv8	11.12	28.4	22.5
SFNet	<b>3.93</b>	<b>13.2</b>	<b>8.3</b>

提升,整体上也获得了最高的 *mAP* 值,表明了本文所提方案能够有效识别钢材表面存在的占比小、对比度低这类小目标缺陷,优于对比的检测模型。

表 2 为各模型在 GC10-DET 上的检测精度对比。根据表 2 显示,在真实工业场景中收集的表面缺陷数据集 GC10-DET 上,SFNet 依然获得了最好的检测效果,其 *mAP* 值为 69.8%,比 YOLOv5 提高了 3.7%;尽管 YOLOv8 在焊缝和夹杂物上表现最好,但 *mAP* 比 SFNet 减少了 4.5%,并且其参数量和模型体积远大于 SFNet;而 EDDN 的 *mAP* 仅为 65.2%。虽然 Faster-RCNN 在轧坑和折痕上表现最佳,但 *mAP* 比 SFNet 降低了 7.1%,并且 Faster-RCNN 为两阶段目标检测模型,一般来说,两阶段目标检测模型在精度上要优于单阶段目标检测模型,但很显然基于单阶段目标检测算法的 SFNet 整体性能比 Faster-RCNN 更好,在检测精度得到有效提升的同时模型更加轻量化,其中 YOLOv2 检测效果较差,难以应对钢材表面小目标缺陷检测任务。

表 3 所示为各模型复杂度对比,SFNet 相比原模型的参数量减少了 3.1,运算量减少了 2.6,模型体积减少了 6.2。与 YOLOv3、YOLOv4、YOLOv8、SSD 等单阶段目标检测模型相比,SFNet 更适用于钢材表面缺陷检测任务,而相比二阶段检测算法 Faster R-CNN, SFNet 无论是在检测精度还是模型轻量化上都有了很大的提升。实验结果表明,SFNet 实现了检测精度提升和模型轻量化的较好平衡,在工业生产和便携式移动设备上更有优势。

2.5 结果可视化

为直观展示 SFNet 在 NEU-DET 和 GC10-DET 上的检测性能,将其 *P-R* 曲线及检测效果可视化如下。

图 7 是根据模型的精确率 *P* 和召回率 *R* 所绘制的 *P-R* 曲线图,图中的 4 条曲线分别表示本文所提模型在 2 个数据集上的 *P-R* 曲线以及基线模型在两个数据集上的 *P-R* 曲线,显然,SFNet 在 2 个数据集上的检查精度较原模型均有较大提升。

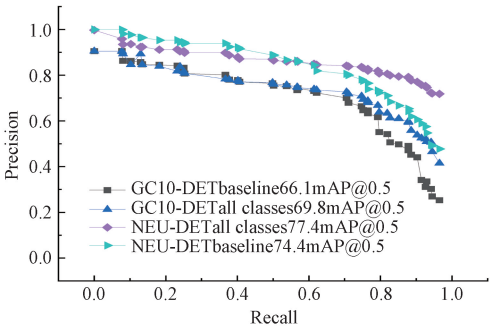


图 7 *P-R* 曲线  
Fig. 7 *P-R* curves

图 8、9 所示分别为 SFNet 在 NEU-DET 和 GC10-DET 两个数据集上的局部检测效果图,图中展示了模型预测时生成的边界框、置信度分数以及缺陷类别等信息。从图中可以看到,SFNet 在面对两种公开数据集检测任务时

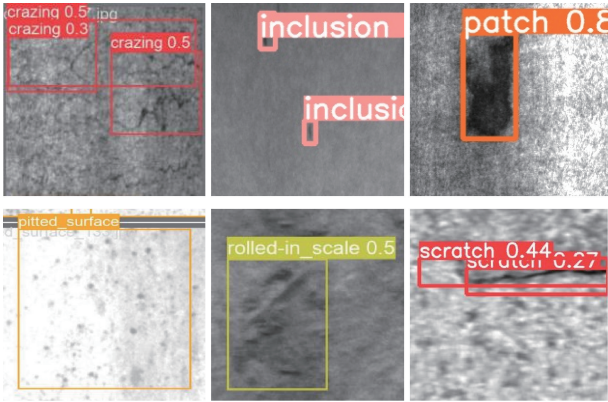


图 8 NEU-DET 上的局部检测效果  
Fig. 8 Effect of local detection on NEU-DET

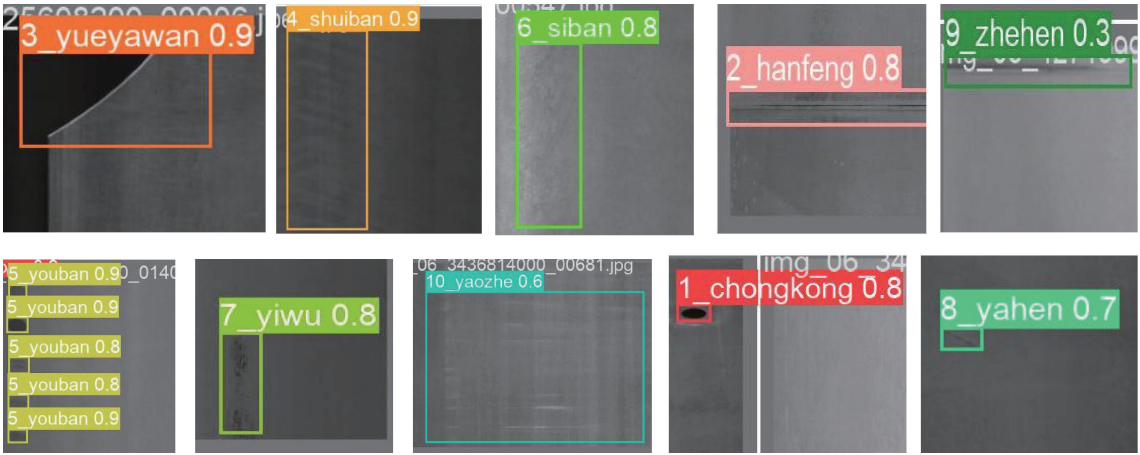


图 9 GC10-DET 上的局部检测效果  
Fig. 9 Effect of local detection on GC10-DET

均能检测出各类缺陷,但对其中折痕类缺陷的预测分数较低,主要原因是样本数据不均衡导致,总体来看,模型能较好的检测到缺陷样本,应对真实工业生产任务具有良好的稳定性。

2.6 消融实验  
为探究本文所提方案对钢材表面缺陷检测模型的增益程度,在 NEU-DET 上进行消融实验,实验环境和相关参数均保持不变,实验结果如表 4 所示。

表 4 在 NEU-DET 上的消融实验  
Table 4 Ablation experiments on NEU-DET

Swin-Transformer	CA	CSP-FCN	$P/\%$	$R/\%$	$mAP/\%$	$Params/M$	$GFLOPs/G$	$Weight/MB$
—	—	—	69.7	70.5	74.4	7.02	15.8	14.5
✓			72.0	71.7	75.3	3.34	11.7	7.1
	✓		67.5	73.9	75.2	2.09	5.1	4.6
		✓	74.2	71.9	75.5	9.32	21.7	19.1
✓	✓	✓	<b>73.0</b>	<b>74.0</b>	<b>77.4</b>	<b>3.93</b>	<b>13.2</b>	<b>8.3</b>

如表 4 所示,在原模型中采用 Swin Transformer 模块后,模型在轻量化上得到了很大的提升,参数量和模型体积减少为原来的 1/2,这得益于使用了分层结构和滑动窗口多头自注意力机制,能够在全局范围内处理图像,同时减少密集的卷积操作。而引入坐标注意力机制后,模型在整体性能上得到了提升,说明模型能够有效关注缺陷区域。虽然在模型中采用梯度流信息更加丰富的特征融合模块 CSP-FCN 后,模型的参数量、计算复杂度以及模型体

积较原模型均有所增加,但检测精度  $P$  却提高了 2.2%,这主要是因为 CSP-FCN 特征融合模块将低级特征图和高级特征图进行拼接,获得了更丰富的特征信息。结合上面 3 种方案后,SFNet 获得了最高的检测精度,并且模型参数量、计算复杂度和体积大小均得到有效降低,实现了检测精度和模型轻量化的有效平衡。  
为了进一步探究本文所提模型的泛化性,在 GC10-DET 数据集上进行消融实验,实验结果如表 5 所示。

表 5 在 GC10-DET 上的消融实验  
Table 5 Ablation experiments on GC10-DET

Swin-Transformer	CA	CSP-FCN	$P/\%$	$R/\%$	$mAP/\%$	$Params/M$	$GFLOPs/G$	$Weight/MB$
—	—	—	66.4	63.2	66.1	7.03	15.8	14.4
✓			71.8	65.4	68.0	3.35	11.7	7.0
	✓		69.9	66.3	69.5	2.10	5.1	4.5
		✓	65.7	64.8	67.0	9.30	21.6	18.9
✓	✓	✓	<b>68.9</b>	<b>67.4</b>	<b>69.8</b>	<b>3.93</b>	<b>13.2</b>	<b>8.2</b>



从表5可知,本文所提方法对GC10-DET数据集依然有效,3个方案都使模型产生了增益,其中引入坐标注意力机制后对模型的影响最大,但检测精度并没有达到最优,而结合3个方法后,模型在检测精度和轻量化方面都获得了最佳性能。

### 3 结 论

针对钢材表面小目标缺陷因占比小、对比度低导致提取丰富缺陷特征失效的问题,本文提出了一种基于自注意特征融合的钢材表面小目标缺陷检测模型SFNet。该模型为降低卷积操作密集性的同时减少小目标缺陷特征信息的丢失,在主干中结合分层结构和滑动窗口机制Swin Transformer以增强对比度,使模型参数数量和计算复杂度得到了有效降低,并采用坐标注意力机制CA加强模型对缺陷位置信息的有效关注;此外,为增强特征表达能力,在模型中使用具有丰富梯度流信息的特征融合模块CSP-FCN,将低级特征与高级特征相结合,以产生丰富的缺陷特征信息,有效提高了模型的特征表达能力。与目前经典的目标检测模型相比,本文所提方法在NEU-DET和GC10-DET数据集上实现参数数量减少为原来1/2的基础上平均精度值分别提升了3%和3.7%。

由于样本不均衡的原因,本文在异物和轧痕类缺陷检测精度上仍存在不足,未来工作将进一步探索弱监督轻量化模型在此类缺陷上的应用,以提升钢材表面小目标缺陷检测的精度和效率。

### 参考文献

[1] 翁俊辉,成乐,黄曼莉,等.基于CS-YOLOv5s的无人机航拍图像小目标检测[J].电子测量技术,2024,47(7):157-162.  
WENG J H, CHENG L, HUANG M L, et al. Small target detection in UAV aerial images based on CS-YOLOv5s[J]. Electronic Measurement Technology, 2024,47(7):157-162.

[2] 赵林熔,甄国涌,储成群,等.基于CBE-YOLOv5的钢材表面缺陷检测方法[J].电子测量技术,2023,46(15):73-80.  
ZHAO L R, ZHEN G Y, CHU CH Q, et al. Steel surface defect detection method based on CBE-YOLOv5[J]. Electronic Measurement Technology, 2023,46(15):73-80.

[3] SUN J, LI CH, WU X J, et al. An effective method of weld defect detection and classification based on machine vision[J]. IEEE Transactions on Industrial Informatics, 2019, 15(12): 6322-6333.

[4] CAO B, LI J, LIU CH F, et al. Defect detection of nickel-plated punched steel strip based on improved least square method[J]. Optik-International Journal

for Light and Electron Optics, 2020, 206: 164331.

[5] ZUO B, WANG F. Surface cutting defect detection of magnet using Fourier image reconstruction [J]. Computer Engineering and Applications, 2016, 52(3): 256-260.

[6] YANG L, WANG ZH, GAO S. Pipeline magnetic flux leakage image detection algorithm based on multiscale SSD network[J]. IEEE Transactions on Industrial Informatics, 2019, 16(1): 501-509.

[7] LI J Y, SU ZH F, GENG J H, et al. Real-time detection of steel strip surface defects based on improved YOLO detection network[J]. IFAC-Papers OnLine, 2018, 51(21): 76-81.

[8] 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.

[9] KOU X P, LIU SH J, CHENG K Q, et al. Development of a YOLO-V3-based model for detecting defects on steel strip surface [J]. Measurement, 2021, 182: 109454.

[10] XU Y, LI D, XIE Q, et al. Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN[J]. Measurement, 2021, 178: 109316.

[11] SONG L, LIN W, YANG Y G, et al. Weak micro-scratch detection based on deep convolutional neural network[J]. IEEE Access, 2019, 7: 27547-27554.

[12] ZOU Q, ZHANG ZH, LI Q, et al. Deepcrack: Learning hierarchical convolutional features for crack detection [J]. IEEE Transactions on Image Processing, 2018, 28(3): 1498-1512.

[13] DONG H, SONG K, HE Y, et al. PGA-Net: Pyramid feature fusion and global context attention network for automated surface defect detection[J]. IEEE Transactions on Industrial Informatics, 2019, 16(12): 7448-7458.

[14] HU H, GU J, ZHANG ZH, et al. Relation networks for object detection[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3588-3597.

[15] CHENG X, YU J. RetinaNet with difference channel attention and adaptively spatial feature fusion for steel surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 70: 1-11.

[16] LIU S, HUANG D, WANG Y. Learning spatial fusion for single-shot object detection [J]. ArXiv preprint arXiv:1911.09516, 2019.

[17] 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 defects on steel surfaces[J]. Journal of Electronic Measurement and Instrumentation, 2024,38(4):108-116.

[18] LIU Z, LIN Y, CAO Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]. IEEE/CVF International Conference on Computer Vision,2021: 10012-10022.

[19] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition,2021: 13713-13722.

[20] TAN M, PANG R, LE Q V. Efficientdet: Scalable and efficient object detection [ C ]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.

[21] 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.

[22] CARION N, MASSA F, SVNNAEVE G, et al. End-to-end object detection with transformers [ C ]. European Conference on Computer Vision. Cham: Springer International Publishing, 2020: 213-229.

[23] LYU X, DUAN F, JIANG J, et al. Deep metallic surface defect detection: The new benchmark and detection network[J]. Sensors, 2020, 20(6): 1562.

作者简介

**冯夫健**(通信作者),博士,教授,主要研究方向为智能计算、微计算及其应用。  
E-mail:fujian\_feng@gzmu.edu.cn

**罗太维**,硕士研究生,主要研究方向为工业缺陷检测。  
E-mail:luotaiwei.xuting@qq.com

**谭棉**,硕士,副教授,主要研究方向为自然图像抠图、微计算。  
E-mail:tanmian@gzmu.edu.cn

**汪小梅**,硕士研究生,主要研究方向为自然图像抠图。  
E-mail:3146088016@qq.com

**王岳继**,硕士研究生,主要研究方向为工业缺陷检测。  
E-mail:2503108165@qq.com