

YOLOv8n-CSG: 轻量化钢材表面缺陷检测算法 *

赵佰亭^{1,2} 张 敏² 贾晓芬¹

(1. 安徽理工大学煤炭无人化开采数智技术全国重点实验室 淮南 232001;
2. 安徽理工大学电气与信息工程学院 淮南 232001)

摘要:为解决钢材表面缺陷检测中因缺陷类型繁多、尺寸差异显著造成检测精度低,以及现有模型复杂度高等问题,提出了一种改进YOLOv8n的轻量化检测算法YOLOv8n-CSG。首先,引入上下文引导模块(context guided block,CG block)设计C2f_CG模块增强对周围特征的捕捉能力,增强信息关联性;其次,加入星型网络模块(Star Block)设计出C2f_Star模块,将输入数据映射到高维的非线性特征空间,生成丰富的特征表示,使得模型在处理细微缺陷时更加有效;最后,设计了集成分组混洗卷积(grouped and shuffled convolution,GSConv)和高效多尺度注意力机制(efficient multi-Scale attention,EMA)的轻量化检测头GSE_Detect,保持了原检测头的高效的同时降低复杂度。在NEU-DET数据集上进行多组实验,结果表明,改进后的YOLOv8n-CSG网络模型平均精度均值(mAP)mAP@0.5达到了76.8%,相较于YOLOv8n,mAP@0.5提升了6.9%、精度提升了11.3%、计算量降低了37%、参数量降低了35.2%,展现出对钢材表面缺陷更佳的检测能力,且平衡了模型的性能和复杂度。

关键词:缺陷检测;轻量化 YOLOv8n;C2f_CG;C2f_Star;GSE_Detect

中图分类号: TN911.73 文献标识码: A 国家标准学科分类代码: 520.2060

YOLOv8n-CSG: Lightweight steel surface defect detection algorithm

Zhao Baiting^{1,2} Zhang Min² Jia Xiaofen¹

(1. State Key Laboratory of Digital Intelligent Technology for Unmanned Coal Mining, Anhui University of Science and Technology, Huainan 232001, China; 2. China Institute of Electrical and Information Engineering, Anhui University of Science and Technology, Huainan 232001, China)

Abstract: In order to solve the problems of low detection accuracy and high complexity of existing models due to the variety of defect types, significant size differences, and high complexity of existing models in the detection of steel surface defects, a lightweight detection algorithm YOLOv8n-CSG with improved YOLOv8n was proposed. Firstly, the design of the CG Block module was introduced C2f_CG which enhanced the ability to capture the surrounding features and enhance the information relevance. Secondly, a C2f_Star module is designed by adding the Star Block module, which maps the input data to the high-dimensional nonlinear feature space and generates rich feature representations, which makes the model more effective in dealing with subtle defects. Finally, a lightweight detector GSE_Detect integrating GSConv and EMA attention mechanisms was designed to maintain the high efficiency of the original detector and reduce the complexity. Multiple sets of experiments on the NEU-DET dataset show that the improved YOLOv8n-CSG network model mAP@0.5 reaches 76.8%, compared with YOLOv8n, mAP@0.5 is improved by 6.9%, the accuracy is increased by 11.3%, the calculation cost is reduced by 37%, and the parameter quantity is reduced by 35.2%, showing a better detection ability for steel surface defects, and balancing the performance and complexity of the model.

Keywords: defect detection; lightweight YOLOv8n; C2f_CG; C2f_Star; GSE_Detect

0 引言

在钢材的铸造、轧制、热处理和使用过程中,常常会出现夹杂物、裂纹、划痕、氧化皮、斑块等各种表面缺陷。当应用在汽车外壳、家电外观、建筑装饰材料等需要高质量表面的应用中,这些缺陷会影响其美观。当应用在工业领域中,钢材表面的缺陷会加速其老化,缩短使用寿命,还可能引发严重的安全事故。因此,钢材缺陷的检测在工业生产中至关重要。

传统的钢材缺陷检测方法主要依赖于人工目视检查和简单的物理测试手段。这些方法存在效率低下、主观性强、检测精度不高等问题。随着计算机视觉和人工智能技术的快速发展,基于深度学习^[1]的自动化检测技术逐渐成为钢材缺陷检测^[2]的主流趋势。这些方法能够通过图像处理和特征提取,实现对缺陷的高效自动化检测,具有更高的准确性。基于深度学习的目标检测算法可分为两类。

1) 以 R-CNN 系列(包括 R-CNN^[3]、Fast R-CNN^[4] 和 Faster R-CNN^[5-6])为代表的基于滑窗或候选区域生成的两阶段方法,此类方法先产生可能包含目标物体的候选区域,再进行精细分类与边界框回归。此类方法算法较为复杂,检测速度较慢,不易于应用在实际的工业生产检测中。

2) 以 SSD^[7]、EfficientDet^[8] 与 YOLO^[9-10] (you only look once)为代表的单阶段方法,此类方法通过一次前向传播直接从输入图像中预测出目标物体的位置及其类别,简化检测流程的同时提高了运行速度,更符合工业级实时检测的需求。但是在处理复杂目标时效果仍然不理想。

为了解决钢材表面缺陷检测精度低、算法复杂度高等问题,许多研究者提出了其优化的检测模型,Chu 等^[11]提出了一种基于改进 YOLOv8 的轻量化钢表面缺陷检测网络 YOLO-SDS。通过替换 YOLOv8 的骨干网络为 StarNet,实现轻量化优化,同时引入轻量级模块可扩张残差(DWR)和 C2f 特征提取模块,增强多尺度特征提取能力;并将遮挡感知注意力机制 SEAM(spatially enhanced attention module)集成到检测头中,提升在复杂场景下的检测性能。Gao 等^[12]提出了 CDN-YOLOv7 算法,通过加入 CARAFE(content-aware ReAssembly of FEatures)上采样算子、级联注意力机制和解耦头,提升网络的特征融合和表征能力,同时引入归一化 Wasserstein 距离和 NFEIoU 损失函数,减少漏检率。徐洪俊等^[13]针对 YOLOv5 在钢材缺陷检测中的不足,提出了一种基于 YOLOv5s 的改进算法,通过在特征提取模块中加入卷积块注意力模块(CBAM)机制、使用 CARAFE 上采样、以及替换

SPPF(spatial pyramid pooling faster) 为 SPPCPSC(SPPF cross stage partial channel),提升了检测性能和抗干扰能力。

本文设计了一种基于深度学习的钢材表面缺陷检测算法,通过结合卷积神经网络^[14](convolutional neural network, CNN)等先进算法,有效识别钢材表面的多种缺陷,提高检测精度^[15],提升钢材产品的整体质量和安全性。对于 YOLOv8n 算法对钢材表面缺陷检测的精度低,小目标检测^[16]困难,对不规则的细微缺陷难以检测的问题。是因为 C2f 模块不能充分挖掘特征之间的语义信息,并且钢材表面存在微小缺陷,造成提取的特征图比较模糊,不能有效的融合特征。而 YOLOv8n 的检测头复杂度太高,影响检测的效率。基于以上问题本文对 YOLOv8n 做出改进。

1 YOLOv8n 算法改进

YOLOv8^[17]延续了 YOLO 系列实时目标检测的优势,并在网络架构以及整体性能等方面进行了显著改进。除了完成传统的目标检测任务, YOLOv8 还支持图像分割和目标跟踪等多种任务,展现出更高的灵活性与扩展性。

为了应对检测钢材表面缺陷的挑战,本文提出了一种专门为此目的设计的检测算法,其基于 YOLOv8,命名为 YOLOv8n-CSG,其网络架构如图 1 所示。

本文引入 CGNet^[18]中的上下文引导模块 Context Guided Block(CG Block)和 Backbone 层的 C2f 融合,以增强特征之间的语义信息,增强相邻信息之间的相关性。引入 StarNet^[19]中的星型模块(Star Block)和 Neck 层的第 3 个和第 4 个 C2f 融合,将提取的特征映射到高维非线性特征空间,以增强特征融合的高效,增强网络学习缺陷信息的能力。将分组混洗卷积(grouped and shuffled convolution, GSConv)和高效多尺度注意力机制(efficient multi-scale attention, EMA)注意力机制集成到检测头上,在降低参数量和复杂度的同时,保持网络的高效性。

1.1 C2f_CG 模块

C2f 模块是 YOLOv8 中的一个关键模块,它通过引入跨阶段部分连接和特征融合的设计,显著提升了网络的性能和效率。C2f 模块主要由 CBS、Split 和 Bottleneck 组成。尽管 C2f 模块能够有效减少计算需求和内存使用,但其基础卷积操作未能充分挖掘输入特征之间的相互关系和语义信息,限制了其在复杂特征提取中的表现。

改进后的 C2f_CG 模块通过用 CG Block 模块替代原有的 Bottleneck 模块,增强了特征提取网络对周围特征信息的结合能力,从而提升了邻近特征之间的相关性。考虑到数据集相对简单,过度的归一化操作可能导致模型

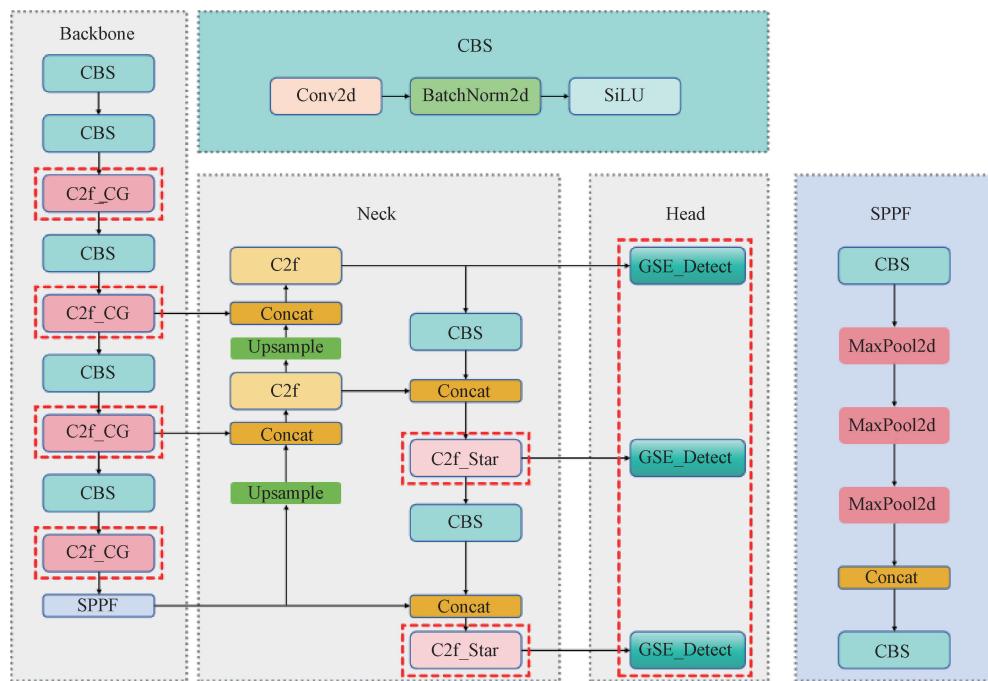


图 1 YOLOv8n-CSG 的网络结构

Fig. 1 YOLOv8n-CSG network structure

的拟合能力降低,同时钢材缺陷类内差异较小,不需要过多的非线性表达。在这种情况下,去除 1×1 卷积后的BN和SiLU激活函数,有助于避免网络过度复杂化,使模型能够更自由地拟合数据,从而取得更好的结果。C2f和改进后的C2f_CG模块的网络结构如图2所示。

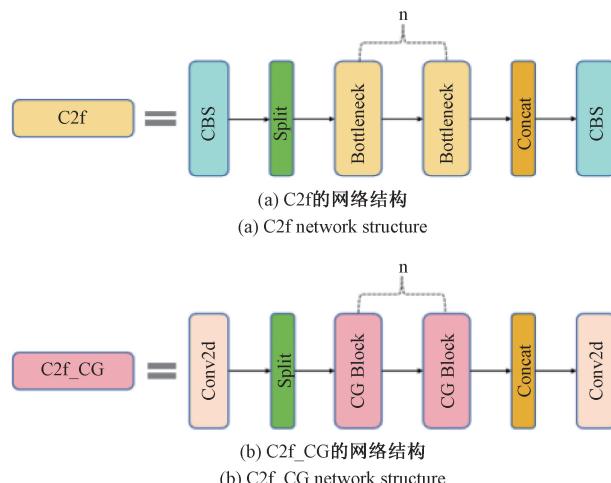


图 2 C2f 和 C2f_CG 的网络结构

Fig. 2 C2f and C2f_CG network structure

CG Block的基本原理是模拟人类视觉系统依赖上下文信息来理解场景。CG block用于捕获局部特征、周围上下文和全局上下文,并将这些信息融合起来,这一模块包含如下4部分。

1)将特征图经过 1×1 卷积然后分别输入到 f_{loc} 和 f_{sur} 中;

2) f_{loc} 提取局部特征,使用 3×3 的普通卷积, f_{sur} 提取周围上下文特征,使用 3×3 的扩张卷积;

3) f_{joi} 提取联合特征,将 f_{loc} 和 f_{sur} 的输出进行Concat操作,再进行BN(Batch Normalization)和PReLU(Parametric ReLU);

4) f_{glo} 提取全局上下文特征,将输入进行全局平均池化(GAP)和多层感知机,将得到的权重和输入按元素相乘。

CG Block网络结构如图3所示, f_{loc} 和 f_{sur} 采用通道卷积(Channel-wise Convolutions)减少参数量。

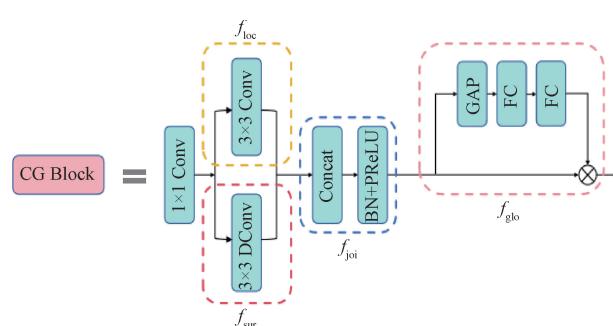


图 3 CG Block 的网络结构

Fig. 3 CG Block network structure

1.2 C2f_Star 模块

由于钢材表面缺陷的大小差异较大且某些缺陷非常微小,捕获的图像往往较为模糊。这使得特征的有效融合变得困难,进而影响了缺陷的边缘、纹理和形状的区分,导致检测效率低下。为了解决这个问题,引入 Star Block 模块来替换 YOLOv8 网络中 Neck 层 C2f 的 Bottleneck 模块。设计了 C2f_Star 模块,其网络结构如图 4 所示。

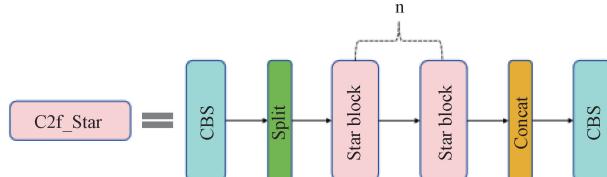


图 4 C2f_Star 的网络结构

Fig. 4 C2f_Star network structure

Star Block 是一种新颖且高效的网络架构,其核心创新在于引入了“星型操作”(Star Operation)来实现特征映射。星型操作通过元素乘法将来自两个不同特征空间的特征进行交互,类似于传统机器学习中的内核技巧。其基本原理是通过元素乘法将输入映射到高维、非线性的特征空间,从而有效增强特征表示能力。Star Operation 的独特优势在于,它能够显著扩展特征空间的维度,而无需增加网络宽度(即不增加通道数),这使得模型在提升表达能力的同时,保持了计算效率。在单层网络中,星型操作通常可以表示为:

$$(\mathbf{W}_1^T \mathbf{X} + \mathbf{B}_1) \otimes (\mathbf{W}_2^T + \mathbf{B}_2) \quad (1)$$

式中: \mathbf{W}_1^T 和 \mathbf{W}_2^T 是两个权重矩阵; \mathbf{X} 表示输入要素; \mathbf{B}_1 和 \mathbf{B}_2 表示两个偏置; \otimes 表示元素乘法运算。

Star Block 是一个以 Star Operation 为核心概念构建的高效神经网络,其特点在于结构简洁而性能强大,由深度卷积层 (DW-Conv)、全连接层 (FC)、激活函数 (ReLU6) 和星型操作组成。DW-Conv 层在每个通道上进行独立卷积操作,保持空间分辨率不变。FC 层将输入特征映射到不同的子空间,从而增强表达能力。ReLU6 激活函数引入非线性,进一步提升模型的表示能力。作为核心运算,星型操作通过元素乘法融合两个分支的特征,从而实现高效的信息整合。Star Block 的框架结构如图 5 所示。

1.3 GSE_Detect 检测头

YOLOv8 共有 3 个检测头,每个检测头由双分支构成,每个分支都有两个 3×3 卷积,共使用了 12 个 3×3 卷积进行构建,导致其参数量和计算复杂度非常大,接近整个模型大小的 $1/2$ 。Li 等^[20]结合 DW-Conv 和普通卷积 (CBS) 的特点,提出一种轻量化的 GSConv,该卷积不

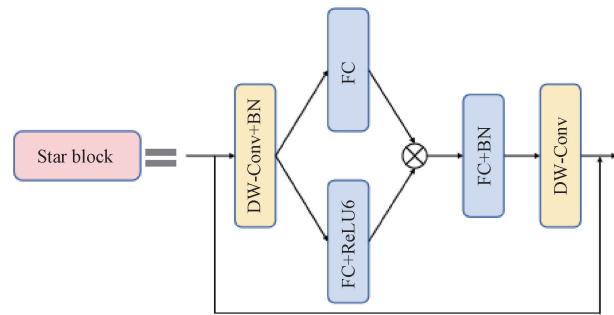


图 5 Star Block 的网络结构

Fig. 5 Star Block network structure

仅保留了 CBS 高效的特点,且有效降低参数量和计算复杂度。本文受文献[20]启发,将其 3 个检测头每个分支的第 1 个 3×3 卷积替换为 3×3 的 GSConv,大大降低参数量和复杂度;将另一个 3×3 卷积替换成 EMA^[21]注意力机制,以维持网络整体的性能,将此检测头命名为 GSE_Detect。Detect 和 GSE_Detect 网络结构如图 6 所示。

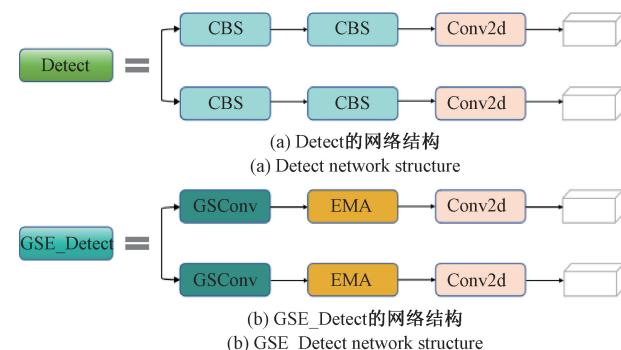


图 6 Detect 和 GSE_Detect 的网络结构

Fig. 6 Detect and GSE_Detect network structure

GSConv 由 CBS、DW-Conv 和通道混洗 (Shuffle) 模块组成,其结构如图 7 所示。为了减少计算量,GSConv 通过压缩通道数的方式进行优化。假设输入和输出通道数分别为 $C1$ 和 $C2$,该卷积首先对输入进行 CBS 操作,生成信息 A,此时通道数被压缩至 $C2/2$;接着,经过 DW-Conv 操作得到信息 B,通道数保持不变;最后,信息 A 和信息 B 被拼接在一起,并通过 Shuffle 操作,最终输出通道数为 $C2$ 的信息。

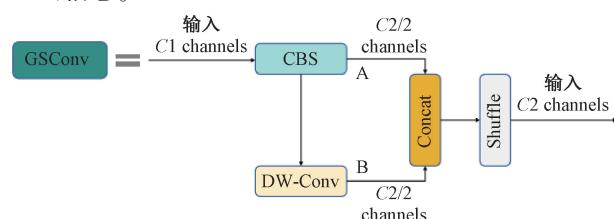


图 7 GSConv 网络结构

Fig. 7 GSConv network structure

为了提升模型的性能和效率,在检测头中引入 EMA 注意力机制。EMA 采用了 CA 模块中的共享 1×1 卷积分支,并命名为 1×1 分支。为了更好地聚合多尺度空间结构信息,EMA 在 1×1 分支并行添加了一个 3×3 卷积,命名为 3×3 分支。通过结合特征分组和多尺度结构,EMA

能够有效地建立短程和长程依赖,从而显著提升模型的性能。在该结构中,“ g ”表示分组,“X Avg Pool”表示一维水平全局池化,而“Y Avg Pool”表示一维垂直全局池化。其完整网络结构如图 8 所示。

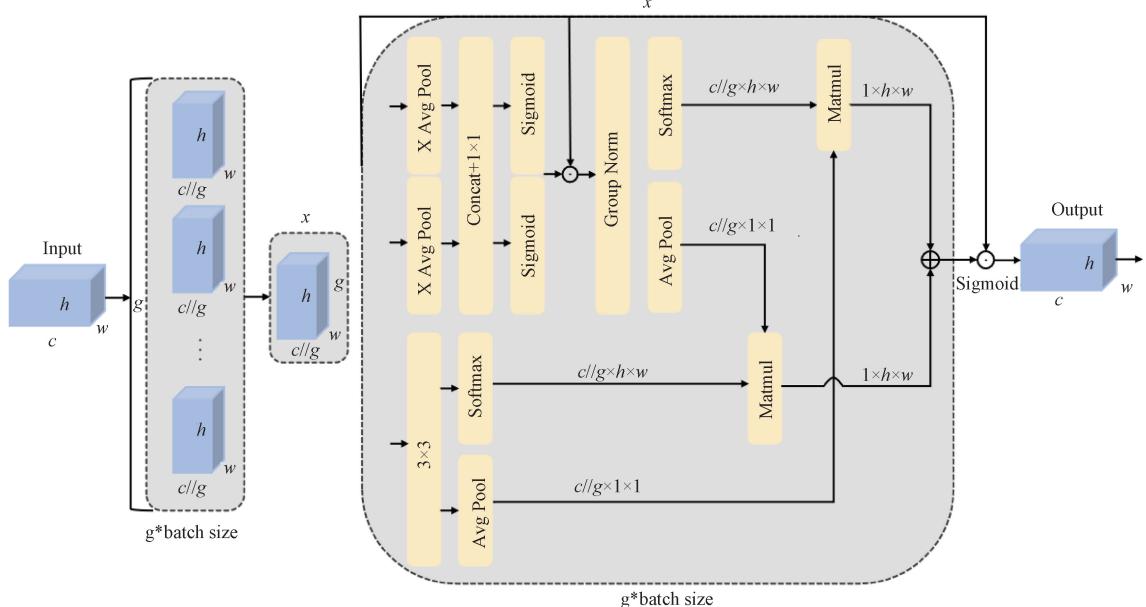


图 8 EMA 网络结构

Fig. 8 EMA network structure

2 实验及结果分析

2.1 实验环境与参数配置

本文实验环境的软硬件配置如表 1 所示。

表 1 实验环境

Table 1 Experimental environment

名称	参数
操作系统	Windows 11
CPU	Intel Core i5-13400F
GPU	NVIDIA GeForce RTX 4060Ti 8 GB
内存	32 GB
解释器	Python 3.9
深度学习框架	Pytorch 1.12.0 + Cuda 11.6

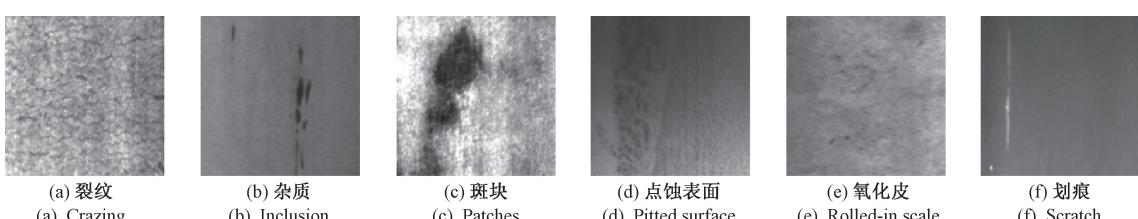


图 9 各类缺陷示例

Fig. 9 Examples of various types of defects

训练参数如下:输入图像尺寸为 640×640 , 批大小(batch_size)设为 16, 迭代次数(epoch)为 200, 非最大值抑制中的交并比(IoU)阈值为 0.7, 起始学习率为 0.01, 权重衰减因子为 0.0005, 动量为 0.937, 其余参数为默认值。

2.2 实验数据集介绍

本文采用东北大学钢材表面缺陷数据集 NEU-DET, 该数据集涵盖了 6 种常见的钢材表面缺陷, 包括裂纹(crazing, Cr)、杂质(inclusion, In)、斑块(patches, Pa)、点蚀表面(pitted Surface, PS)、氧化皮(rolled-in Scale, RS)和划痕(scratches, SC), 如图 9 所示。其中, 每种缺陷类型包含 300 张图片, 共 1800 张。对数据集进行划分, 使得训练集、验证集和测试集比例为 8:1:1。

2.3 实验评价指标

为了评估模型性能增强功能的优越性,采用了常用的指标^[22]精度(precision, P)、召回率(recall, R)、平均精度(AP)、平均精度均值(mAP)。其中 mAP@0.5 表示 IoU 为 0.5 阈值处的 mAP。mAP@0.5:0.95 则表示 IoU 为 0.5~0.95 阈值的平均 mAP。TP 表示为正确检测到的实例,相反,FP 表示为错误检测到的实例。FN 表示漏检的实例。通过评估 TP、FP 和 FN,可以计算关键性能指标,如精度和召回率。AP 表示 PR 曲线围成的面积。

2.4 改进网络实验结果与分析

1) 改进算法与基准模型对比实验

在 NEU-DET 数据集上分别对原始 YOLOv8n 模型与改进 YOLOv8-CSG 模型进行训练与验证,可以看出改进后的 YOLOv8n-CSG 模型在 6 类钢材表面缺陷的平均精度的值相较于 YOLOv8n 检测数据的值均有一定提升,且改进后的模型在参数量、模型计算量上都下降较多。对比结果如表 2 所示。

表 2 算法改进前后对比

Table 2 Comparison before and after algorithm improvement

模型	参数量/ ($\times 10^6$)	浮点数/ GFLOPS	AP/%					mAP@0.5/%
			Cr	In	Pa	PS	RS	
YOLOv8n	3.01	8.1	44.4	76.8	92.8	76.5	55.1	73.9
YOLOv8n-CSG	1.95	5.1	52.5	84.8	93.3	79.0	66.2	85.2

对比结果表明,本文的 YOLOv8-CSG 模型对各类钢材表面缺陷展现出优异的检测效果,改进后的模型较原模型参数量降低了 35.2%,计算量降低了 37%,mAP@0.5 提升了 6.9%,YOLOv8-CSG 模型在轻量化的同时能更精准识别各类缺陷,具备卓越的检测能力。

2) C2f_CG 模块实验

为了验证 C2f_CG 模块的有效性,在 NEU-DET 数据集上分别对原始 YOLOv8n 模型,模型 A 为将 Bottleneck 模块替换成 CG Block 模块,模型 B 为在 A 的基础上继续

将 C2f 中两个 1×1 卷积的 BN 和激活函数移除,模型 C 为在 B 的基础上将原网络的 4 个 C2f 重复次数全部改成 1,进行训练与验证,其中加粗的数值表示最优。

由表 3 可以看出,上述 3 种 CG Block 和 C2f 的融合方式,在 mAP 上均得到了提升,参数量和计算量都有一定的下降,但选择融合方式 C 后,对网络的性能提升贡献最大,检测精度和模型复杂度都达到最优。除此之外,与基线模型相比较,参数量和计算量分别降低了 0.48×10^6 和 1.4 GFLOPs,mAP@0.5 提升了 3.5%。

表 3 C2f_CG 模块有效性验证

Table 3 C2f_CG module validity verification

模型	参数量/ $(\times 10^6)$	浮点数/GFLOPS	P/%	R/%	mAP@0.5/%	mAP@0.5:0.9/%
YOLOv8n	3.01	8.1	60.4	68.4	69.9	36.2
A	2.54	6.8	58.8	73.0	72.6	37.2
B	2.54	6.8	61.1	70.2	73.0	37.3
C	2.53	6.7	61.4	70.7	73.4	38.4

3) C2f_Star 模块实验

为了验证 C2f_Star 模块的有效性,在 NEU-DET 数据集上分别对原始 YOLOv8n 模型,模型 D 为将 Neck 网络中的 4 个 C2f 模块全部替换成 C2f_Star 模块,模型 E 为将 Neck 网络中的后 2 个 C2f 模块替换成 C2f_Star 模块,进行训练与验证,其中加粗的数值表示最优。

由表 4 可知,上述两种 C2f_Star 模块和 C2f 的替换

方式,在 mAP 上均得到了提升,参数量和计算量都有一定的下降,但选择替换方式 E 后,对网络的性能提升贡献最大,检测精度和 mAP@0.5 达到最优。除此之外,与基线模型相比较,参数量和计算量分别降低了 0.17×10^6 和 0.2 GFLOPs,mAP@0.5 提升了 3.4%。表明 C2f_Star 模块能够在不增加复杂度的情况下将特征映射到高维、非线性的特征空间,让特征融合更加高效。

表 4 C2f_Star 有效性验证

Table 4 C2f_Star module validity verification

模型	参数量/ $(\times 10^6)$	浮点数/GFLOPS	P/%	R/%	mAP@0.5/%	mAP@0.5:0.9/%
YOLOv8n	3.01	8.1	60.4	68.4	69.9	36.2
D	2.81	7.7	70.4	66.1	73.0	39.2
E	2.84	7.9	71.3	67.3	73.3	38.2

4) GSE_Detect 检测头实验

为了验证 GSE_Detect 检测头的有效性, 在 NEU-DET 数据集上分别对 YOLOv8n 模型、GSConv 与不同的注意力机制集成的检测头进行训练与验证, 其中加粗的数值表示最优。

由表 5 可知, 上述 5 种不同的注意力机制与 GSConv 集成的检测头, 在 mAP 上均得到了提升, 参数量和计算

量都有一定的下降, 但选择 EMA 注意力机制和 GSConv 集成后, 对网络的性能提升贡献最大, mAP@0.5 和 mAP@0.5:0.9 都达到最优。除此之外, 与基线模型相比较, 参数量和计算量分别降低了 0.42×10^6 和 1.4 GFLOPs, mAP@0.5 提升了 2.2%。GSConv 通过压缩通道有效降低了参数量和计算复杂度, EMA 注意力机制让模型更加高效。实验证明了 GSE_Detect 检测头的合理性。

表 5 GSE_Detect 有效性验证

Table 5 GSE_Detect module validity verification

模型	参数量/($\times 10^6$)	浮点数/GFLOPS	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv8n	3.01	8.1	60.4	68.4	69.9	36.2
SE	2.54	6.1	62.4	66.6	70.7	37.9
SEAM	2.57	6.2	64.7	68.0	71.3	35.7
ECA	2.53	6.1	64.4	64.4	71.0	36.5
SimAM	2.53	6.1	67.2	63.8	71.3	36.7
EMA	2.59	6.7	63.7	69.0	72.1	38.5

2.5 消融实验

为了评估所提出的网络的性能和效率, 对网络架构的组件进行了各种实验, M1 表示 C2f_CG 模块, M2 表示

C2f_Star 模块, M3 表示 GSE_Detect 模块, \checkmark 表示加入该模块。其中加粗的数值表示最优。

表 6 消融实验

Table 6 Ablation experiments

M1	M2	M3	参数量/($\times 10^6$)	浮点数/GFLOPS	P/%	R/%	mAP@0.5/%	mAP@0.5:0.95/%
YOLOv8n			3.01	8.1	60.4	68.4	69.9	36.2
	\checkmark		2.53	6.7	61.4	70.7	73.4	38.4
		\checkmark	2.84	7.9	71.3	67.3	73.3	38.2
			\checkmark	2.59	6.7	63.7	69.0	72.1
	\checkmark	\checkmark	2.36	6.5	68.3	70.6	75.7	39.9
	\checkmark		2.11	5.2	68.3	68.2	75.4	39.4
		\checkmark	2.43	6.5	65.5	70.4	73.7	38.4
	\checkmark	\checkmark	1.95	5.1	71.7	73.0	76.8	40.4

由表 6 可见, 当将 Backbone 中的 C2f 模块替换为 C2f_CG 模块时, 模型的参数量和计算量分别减少了 0.48×10^6 和 1.4 GFLOPs, 同时 mAP@0.5 提升了 3.5%, 证明 C2f_CG 模块有效减少冗余信息并增强特征提取能力。同时替换 C2f_CG 和 C2f_Star 模块时, 模型复杂度和检测精度进一步优化, 参数量和计算量减少 0.65×10^6 和 1.6 GFLOPs, mAP@0.5 提升了 5.8%。将 C2f_CG 和 GSE_Detect 检测头同时替换时, 参数量和计算量分别减少了 0.9×10^6 和 2.9 GFLOPs, mAP@0.5 提升了 5.5%。替换 C2f_Star 模块和 GSE_Detect 检测头时, 参数量和计算量分别减少 0.58×10^6 和 1.6 GFLOPs, mAP@0.5 提升了 3.8%。最终, 本文提出的 YOLOv8n-CSG 算法在各项指标上均优于 YOLOv8n, 且达到了最佳性能: 参数量和计算量分别降低了 1.06×10^6 和 3 GFLOPs, 精度和召回率分别提升 11.3% 和 4.6%, mAP@0.5 和 mAP@0.5:

0.9 分别提升了 6.9% 和 4.2%。该模型在保持较低参数量和计算负担的同时, 显著提升了性能, 证明 YOLOv8n-CSG 在检测性能和模型复杂性之间达到了更优的平衡。

2.6 对比实验

为了验证本文算法的优越性, 在 NEU-DET 数据集和相同实验条件下, 将 YOLOv5n、YOLOv5s、YOLOv7、YOLOv7-tiny、YOLOv8n、YOLOv8s、YOLOv10n、YOLOv11n、YOLOv8n-CSG 算法进行实验验证, 以及参考文献[23-25]的算法进行对比。表 7 为了其对比实验的结果, 其中加粗表示最优值。

由表 7 可知, 本文提出的 YOLOv8n-CSG 算法相对于传统算法表现出了优越性, mAP@0.5 达到 76.8% 的同时参数量和计算量仅为 1.95×10^6 和 5.1 GFLOPs。相对于文献[23]来说, 表现得更加优异。而文献[24]相对本

表 7 对比实验

Table 7 Comparative experiments

模型	参数量/($\times 10^6$)	浮点数/GFLOPS	P/%	R/%	mAP@ 0.5/%	mAP@ 0.5; 0.95/%
YOLOv5n	1.77	4.2	63.4	71.8	71.8	38.3
YOLOv5s	7.03	15.8	69.2	71.8	72.6	38.9
YOLOv7	37.20	105.2	63.8	68.6	70.7	34.8
YOLOv7-tiny	6.03	13.2	65.9	57.0	66.9	31.2
YOLOv8n	3.01	8.1	60.4	68.4	69.9	36.2
YOLOv8s	11.10	28.4	70.1	70.1	73.5	37.2
YOLOv10n	2.70	8.2	66.0	61.0	67.4	35.3
YOLOv11n	2.58	6.3	64.3	68.8	72.0	38.2
FMG-YOLOv8s ^[23]	2.90	6.7	—	—	75.9	—
YOLOX ^[24]	7.23	20.7	—	—	77.0	—
YOLOv8n-SDEC ^[25]	4.86	—	71.2	71.3	76.7	—
YOLOv8n-CSG	1.95	5.1	71.7	73.0	76.8	40.4

文算法虽然 mAP@ 0.5 高了 0.2%, 但是模型更加复杂, 参数量和计算量达到了 7.23×10^6 和 20.7 GFLOPs, 浪费了过多的资源不利于在设备上部署。文献[25]在精度、召回率和 mAP@ 0.5 都和本文相差无几, 但是本文算法参数量低于文献[25]。相比之下, YOLOv8n-CSG 在较低的参数量的计算量的同时却能保持较高的 mAP@ 0.5 值, 展现了较好的轻量化和性能平衡。

2.7 可视化对比

钢材缺陷同一类中差异较小, YOLOv8n-CSG 网络引入了上下文引导模块, 从而提升了邻近特征之间的相关

性, 提高邻近特征的提取能力。钢材表面缺陷不同类中的大小差异较大且某些缺陷非常微小, 捕获的图像往往较为模糊, 这使得特征的有效融合变得困难, 进而影响了缺陷的边缘、纹理和形状的区分, 而引入星型网络可以将输入映射到高维、非线性的特征空间, 从而有效增强特征融合能力, 而高效的检测头能够更加有效的识别微小缺陷。

为验证本文算法的检测效果, 将其与原始模型算法在 NEU-DET 数据集上进行预测, 并使用不同颜色的锚定框区分不同缺陷类别, 可视化结果如图 10 所示。由图 8(a)~(c)、(e) 可见, YOLOv8n 在对裂纹、杂质和斑块

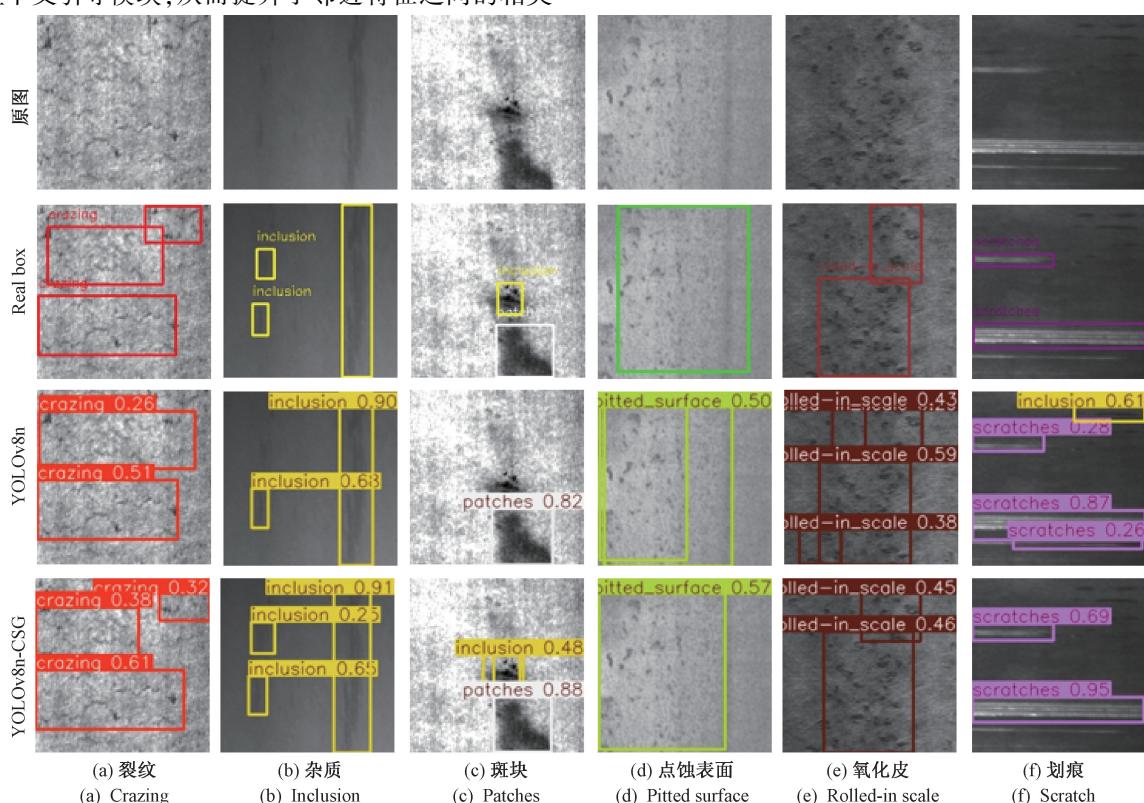


图 10 可视化钢材表面缺陷检测图

Fig. 10 Visualize a steel surface defect detection diagram

的检测时产生了漏检的情况,而在对氧化皮检测时出现了误检,而 YOLOv8n-CSG 不仅解决的这些问题,还增加了检测的准确度。显然本文提出的算法在微小缺陷、伪装缺陷以及相似缺陷的检测上更加优异。

不同模型训练的边界回归损失图如图 11 所示。与基线模型相比,早期训练 YOLOv8n-CSG 损失下降得更快并且最终值低于 YOLOv8n,表明引入的上下文引导模块、星型网络模块,以及高效轻量级检测头优化了特征提取和特征融合能力,证明了改进措施的有效性。

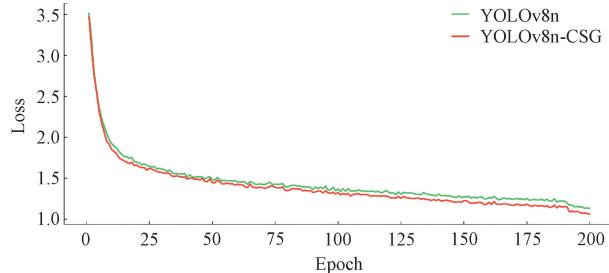


图 11 改进前后模型损失曲线对比

Fig. 11 Comparison of model loss curves
before and after improvement

为验证改进方法的有效性,本文在 NEU-DET 数据集上将 YOLOv8n-CSG 与基线模型进行了详细对比,如图 12 所示。由图 12 可见,该方法在精度、召回率和 mAP@0.5 等关键指标上均有显著提升。同时,计算量和参数量分别减少了 37.0% 和 35.2%,实现了性能与效率的双重优化。性能的提升主要得益于更有效的微小缺陷信息提取与多尺度特征融合。综合来看,YOLOv8n-CSG 在钢材缺陷检测中展现出优越性能,不仅提升了检测准确性与实时性,也为工业应用提供了更可靠的技术支持。

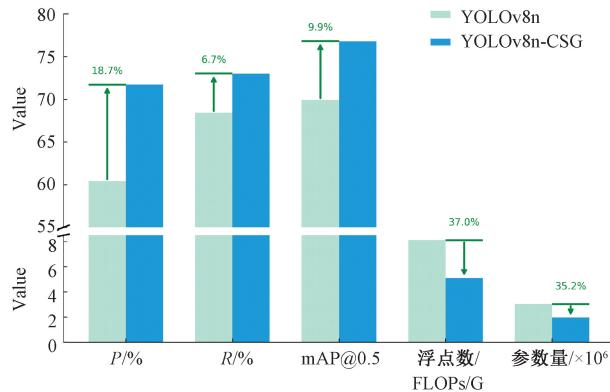


图 12 模型检测性能对比

Fig. 12 Model detection performance comparison chart

3 结 论

针对钢材缺陷检测模型复杂度高、检测精度低等问题,本文提出一种基于 YOLOv8n 改进的检测算法命名为 YOLOv8n-CSG。YOLOv8n-CSG 在检测精度和模型复杂度都表现出了出色的性能。首先,在特征提取网络引入 CG Block 增强对周围特征的捕捉能力,增强关联性信息的提取能力;其次,在特征融合网络加入 Star Block,可以将输入数据映射到一个极高维的非线性特征空间,能够生成丰富的特征表示,使得模型在处理复杂数据时更加有效。最后,在检测头中集成了 GSConv 和 EMA 注意力机制,在保持原检测头高效的同时大大降低了复杂度。在 NEU-DET 数据集上进行多组实验,结果表明,改进后的 YOLOv8n-CSG 网络模型相比 YOLOv8n 模型 mAP@0.5 提升了 6.9%、准确率提升了 11.3%、召回率提升了 4.6%、浮点数下降了 37%、参数量降低了 35.2%。通过对其他模型的对比,表现出本文提出的 YOLOv8n-CSG 模型对钢材表面缺陷更佳的定位与识别能力,且检测速度满足工业应用需求,平衡了模型的性能和复杂度。在未来的研究中,将考虑继续对检测头进行轻量化处理,减少计算开销,使其能部署在小型的嵌入式平台上,并将其应用于钢材缺陷检测领域。

参考文献

- [1] 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 surface defects in steel [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 108-116.
- [2] ZHANG T, PANG H, JIANG C. GDM-YOLO: A Model for Steel Surface Defect Detection Based on YOLOv8s [J]. IEEE Access, 2024, 12: 148817-148825.
- [3] HE K, GKIOXARI G, DOLLAR P, GIRSHICK R. Mask R-CNN [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [4] HUSSAIN M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection [J]. Machines, 2023, 11(7): 677.
- [5] REN S, HE K, 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: 1137-1149.
- [6] 向宽,李松松,栾明慧,等. 基于改进 Faster RCNN 的铝材表面缺陷检测方法 [J]. 仪器仪表学报, 2021, 42(1): 191-198.

针对钢材缺陷检测模型复杂度高、检测精度低等问题

- XIANG K, LI S S, LUAN M H, et al. Aluminum surface defect detection method based on improved FasterRCNN [J]. Chinese Journal of Scientific Instrument, 2021, 42(1): 191-198.
- [7] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection [C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2980-2988.
- [8] TAN M, PANG R, LE Q. Efficientdet: Scalable and efficient object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 10781-10790.
- [9] JIANG P Y, ERGU D, LIU F Y, et al. A review of YOLO algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066-1073.
- [10] HUSSAIN M. YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection [J]. Machines, 2023, 11(7): 677.
- [11] CHU Y, YU X, RONG X. A lightweight strip steel surface defect detection network based on improved YOLOv8 [J]. Sensors 2024, 24: 6495.
- [12] 高春艳, 秦燊, 李满宏, 等. 改进 YOLOv7 算法的钢材表面缺陷检测研究 [J]. 计算机工程与应用, 2024, 60(7): 282-291.
- GAO CH Y, QIN S, LI M H, et al. Research on steel surface defect detection with improved YOLOv7 algorithm [J]. Computer Engineering and Application, 2024, 60(7): 282-291.
- [13] 徐洪俊, 唐自强, 张锦东, 等. 钢材表面缺陷检测的 YOLOv5s 算法优化研究 [J]. 计算机工程与应用, 2024, 60(7): 306-314.
- XU H J, TANG Z Q, ZHANG J D, et al. Research on optimisation of YOLOv5s algorithm for steel surface defect detection [J]. Computer Engineering and Application, 2024, 60(7): 306-314.
- [14] KIM Y. Convolutional neural networks for sentence classification [J]. ArXiv preprint arXiv: 1408.5882, 2014.
- [15] 王素珍, 吕基岳, 邓成禹, 等. 基于改进 YOLOv5 的钢铁表面缺陷检测算法 [J]. 国外电子测量技术, 2023, 42(10): 43-50.
- WANG S ZH, LYU J Y, DENG CH Y, et al. Steel surface defect detection algorithm based on improved YOLOv5 [J]. Foreign Electronic Measurement Technology, 2023, 42(10): 43-50.
- [16] 王浩, 张其猛. 复杂场景下的交通标志小目标检测算法 [J]. 电子测量技术, 2025, 48(2): 158-169.
- WANG H, ZHANG Q M. Small target detection algorithm for traffic signs in complex scenes [J]. Electronic Measurement Technology, 2025, 48(2): 158-169.
- [17] VARGHESE R, S M. 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), 2024: 1-6.
- [18] WU T, TANG S, ZHANG R, et al. CGNet: A light-weight context guided network for semantic segmentation [J]. IEEE Transactions on Image Processing, 2021, 30: 1169-1179.
- [19] MA X, DAI X, BAI Y, et al. Rewrite the stars [C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024: 5694-5703.
- [20] LI H, LI J, WEI H, et al. Slim-neck by GSConv: A better design paradigm of detector architectures for autonomous vehicles [J]. ArXiv preprint arXiv: 2206.02424, 2022.
- [21] OUYANG D L, HE S, ZHAN J, et al. Efficient multi-scale attention module with cross-spatial learning [C]. 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2023: 1-5.
- [22] YAO L, ZHAO B, WANG X, et al. A detection algorithm for surface defects of printed circuit board based on improved YOLOv8 [J]. IEEE Access, 2024, 12: 170227-170242.
- [23] 梁礼明, 龙鹏威, 李俞霖. 改进轻量高效 FMG-YOLOv8s 的钢材表面缺陷检测算法 [J]. 计算机工程与应用, 2025, 61(3): 84-93.
- LIANG L M, LONG P W, LI Y L. Improved steel surface defect detection algorithm of lightweight and efficient FMG-YOLOv8s [J]. Computer Engineering and Application, 2025, 61(3): 84-93.
- [24] LI C, XU A, ZHANG Q, et al. Steel surface defect detection method based on improved YOLOX [J]. IEEE Access, 2024, 12: 37643-37652.
- [25] JIANG X, CUI Y, CUI Y, et al. Optimization algorithm of steel surface defect detection based on YOLOv8n-SDEC [J]. IEEE Access, 2024, 12: 95106-95117.

作者简介



赵佰亭, 2002 年于青岛科技大学获得学士学位, 2005 年于青岛科技大学获得硕士学位, 2010 年于哈尔滨工业大学获得博士学位, 现为安徽理工大学教授, 主要研究方向为深度学习、图像处理。

E-mail: btzhao@ aust. edu. cn

Zhao Baiting received his B. Sc. degree from Qingdao University of Science and Technology in 2002, M. Sc. degree from Qingdao University of Science and Technology in 2005 and Ph. D. degree from Harbin Institute of Technology in 2010. He is

now a professor at Anhui University of Science and Technology. His main research interests include deep learning and image processing.



张敏(通信作者), 现为安徽理工大学硕士研究生, 主要研究方向为图像处理、缺陷检测。

E-mail: 1090918228@ qq. com

Zhang Min (Corresponding author) is now a M. Sc. candidate at Anhui University of Science and Technology. His main research interests include image processing and defect detection.