

DOI: 10.13382/j.jemi.B2508149

PGS-YOLO: 一种轻量高效的带钢表面缺陷检测模型*

马俊杰¹ 张继红¹ 王强² 刘文广³ 吴振奎¹

(1. 内蒙古科技大学自动化与电气工程学院 包头 014010; 2. 包头威丰新材料有限公司设备处 包头 014020;
3. 内蒙古科技大学机械工程学院 包头 014010)

摘要: 钢材是我国的支柱产业,其表面质量问题是影响钢材性能与价格的关键。针对带钢表面缺陷检测存在精度差、效率低、模型复杂度高等诸多问题,提出并改进了一种轻量级的带钢表面缺陷检测模型(PGS-YOLO)。首先,引入更为灵活的 PReLU 激活函数,通过可学习参数自适应调整输入数据负区域斜率,从而提高模型定位缺陷的准确性;其次,将 Re-VGG 融入 C3,构建轻量高效的 Re-C3 模块,降低模型复杂度并提高计算效率,进一步地,基于 GELAN 网络联合设计了全新的 G-GELAN 特征提取-融合模块,通过融合多尺度、多层次的高级语义信息,增强模型对不同类型缺陷及复杂背景的适应能力;最后,采用轻量级的 SCDown 下采样操作,在减少冗余计算的同时提升特征融合的丰富度。在 NEU-DET 数据集上的实验结果表明,该模型相比基准模型平均精度均值(mAP)提高 6.7%,达到 79.9%;参数数量和计算量分别减少 29.7%、27.2%,帧率提升 2.7%,更好地平衡了检测精度、推理速度与轻量化之间的关系。此外,该模型在 WF10-DET 数据集和 PCB_DATASET 数据集上均表现出良好的泛化能力,满足实际工程部署需求,预期在工程应用中具有重要推广应用价值。

关键词: 带钢表面缺陷检测; PReLU; 轻量化; G-GELAN; SCDown

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

PGS-YOLO: A lightweight and efficient strip surface defect detection model

Ma Junjie¹ Zhan Jihong¹ Wang Qiang² Liu Wenguang³ Wu Zhenkui¹

(1. School of Automation and Electrical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China;
2. Equipment Division, Baotou Weifeng New Materials Company Limited, Baotou 014020, China; 3. School of
Mechanical Engineering, Inner Mongolia University of Science and Technology, Baotou 014010, China)

Abstract: Steel is the pillar industry of China, and its surface quality is the key to affecting the performance and price of steel. In order to solve the problems of poor accuracy, low efficiency and high model complexity in strip surface defect detection, a lightweight strip surface defect detection model (PGS-YOLO) was proposed and improved. Firstly, a more flexible PReLU activation function was introduced, and the slope of the negative region of the input data was adaptively adjusted through the learnable parameters, so as to improve the accuracy of the model to locate defects. Secondly, the Re-VGG is integrated into C3 to build a lightweight and efficient Re-C3 module to reduce the complexity of the model and improve the computational efficiency. Finally, the lightweight SCDown downsampling operation is adopted to reduce redundant calculations and improve the richness of feature fusion. Experimental results on the NEU-DET dataset show that the mAP of the model is increased by 6.7% to 79.9% compared with the benchmark model. The number of parameters and the amount of computation are reduced by 29.7% and 27.2%, respectively, and the FPS is increased by 2.7%, which better balances the relationship between detection accuracy, inference speed and lightweight. In addition, the model shows good generalization ability on both the WF10-DET dataset and the PCB_DATASET dataset, which meets the needs of actual engineering deployment and is expected to have important promotion and application value in engineering applications.

Keywords: strip surface defect detection; PReLU; lightweight; G-GELAN; SCDown

收稿日期: 2025-02-06 Received Date: 2025-02-06

* 基金项目: 内蒙古自治区重点研发与成果转化项目(2022YFHH0019)、内蒙古自治区科技攻关大平台建设项目(2023PTXM001)资助

0 引言

随着经济的快速增长,我国的钢铁生产量和需求量持续上升^[1]。然而,在实际钢材的生产及加工过程中,易受原料质量、生产设备、生产环境等因素的影响,导致其表面易产生多种类型缺陷,不仅影响钢材的外观表面,还会造成经济损失,甚至存在疲劳强度和耐磨性等安全隐患^[2-4],尤其生产的带钢产品更为突出。因此,探索一种高效、精准的带钢表面缺陷检测方法意义重大。

早期,带钢表面缺陷检测往往采用人工视觉、涡流探测和红外扫描等方法。这类检测方法存在很大的主观性,检测效率不高,已不能适应现代大规模生产的需要,逐渐被摒弃^[5-6]。随着信息科技的迅猛发展,图像处理技术也取得了巨大的进步。机器视觉检测主要分为传统机器视觉检测方法和基于深度学习的检测方法^[7]。传统方法仍需借助人工分析并提取缺陷纹理特征,这不仅导致了较高的计算复杂度,还影响检测的时效性,同时容易受到复杂背景的干扰,并且展现出较弱的泛化能力^[8]。

近年来,由于卷积神经网络(CNN)技术的普及以及计算机算力的提升,以深度学习为基础的检测算法获得了长足的发展,并逐渐应用到物体表面缺陷检测上^[9-10]。与传统机器视觉检测相比,深度学习检测方法更关注于目标检测的应用,其优点在于能够通过监督学习有效提取目标的特征信息,具有覆盖范围广、适应性强等特点,削弱了特征工程对识别精度的影响^[11],并逐步受到了相关学者和专家的重视。文献[12]设计一种融合HGnetv2和注意力机制的改进YOLOv5算法,满足实时性要求的条件下,提高了检测精度。文献[13]提出快速DCAM-Net(变形卷积和注意力网络)带钢表面缺陷检测方法,在CLAHE(contrast limited adaptive histogram equalization)基础上设计增强型变形特征提取模块,使其在特征提取时获得更多的感受野;并替换空间金字塔池化(SPP)为CA(coordinate attention)模块,提升定位缺陷的能力,在取得较高检测精度的同时,显著提高了带钢表面缺陷检测效率。文献[14]将多分枝结构RepGhost替换网络部分卷积,并设计GG-FP(GPU efficient ghost-feature pyramid networks)进一步增强特征提取能力,实现了精度与速度的双重提升。文献[15]基于YOLO框架设计新的主干网,通过捕捉高阶空间交互和数据特征内相关性来降低检测误检率。文献[16]将OSA-C3(one-shot aggregation-C3)模块代替骨干原C3模块,引入BiC-PAN(bidirectional cascade-path aggregation network)结构增强特征融合能力,使得改进后的YOLOv5钢筋表面缺陷检测算法在检测精度与速度之间达到了较好的平衡。文献[17]将注意力机制和加权双向特征金字塔网

络(BiFPN)集成到YOLOv5架构中,有效解决了图像中背景干扰及缺陷尺寸变化所带来的挑战。尽管上述研究在缺陷检测领域取得了一定的进展,但在实际工业生产和应用中,由于背景复杂、缺陷类型多样等因素,仍存在较高的漏检率、误检率,同时未能较好地平衡检测精度和轻量化,难以满足工业检测要求。鉴于上述研究的局限性,本文在YOLOv5的基础上,提出一种PGS(PReLU+G-GELAN+SCDown)-YOLO检测模型。

为更好地适应带钢表面缺陷数据集图像背景复杂、亮度不高等非线性特性,首先更换激活函数,选用的PReLU(parametric rectified linear unit)可在深度复杂网络中展现更灵活的激活行为,通过学习负区域斜率,更好地适应数据分布,从而精确的捕捉缺陷细微特征,提升模型性能;同时可避免“死神经元”现象,提高网络的训练稳定性。受RepVGG和GELAN启发,将更新后的VGG添加至C3形成Re-C3模块,然后将Re-C3与常规卷积结合形成不同的计算块,并纳入GELAN结构,构建G-GELAN(GhostConv-GELAN)特征提取模块,取代基准模型中的C3;通过多层次、多尺度的信息融合,能够高效地提取缺陷特征,同时,减少了网络的计算量和内存占用,达到轻量化结构的效果以便适应不同设备终端部署。采用轻量级的SCDown(selective channel down-sampling)下采样方法,分开调整通道和空间信息,根据输入特征的实际需求动态调整通道数量和空间尺寸,灵活地学习不同通道特征,减少信息丢失,增强了模型对于不同尺寸和形态的缺陷检测任务的适应能力,减少了传统下采样的计算量和参数量,提高计算效率并避免了不必要的开销。

1 YOLOv5 模型

目前,基于深度学习的机器视觉检测方法可划分为单阶段(One-stage)和两阶段(Two-stage)。其中One-stage代表模型有SSD^[18]和YOLO^[19]系列,这类网络不需要生成候选框便可直接进行定位和类别的检测,具有检测速度快,精度高等特点,广泛应用于工业领域^[20-21]。YOLOv5根据其网络的深度和宽度,提供了多种结构版本,如YOLOv5n、YOLOv5s、YOLOv5m、YOLOv5x。尽管YOLOv11和YOLOv12等新版本相继开源,但由于推出时间较短,社区版本尚未成熟,在复杂工业环境下长期运行的稳定性和可靠性仍需进一步验证。此外,本研究主要面向工业现场的应用场景,硬件资源通常有限。为避免新版本中可能出现的过高计算需求或实现复杂度问题,并确保有效兼容不同硬件环境,综合考虑在带钢表面缺陷数据集上的检测精度、网络大小以及模型检测任务的实现需求,本文选取YOLOv5s进行改进。YOLOv5s结构主要分为输入端(Input)、骨干网(Backbone)、颈部

网(Neck)、检测头(Head)。Input 将接收的待检图统一裁剪为 640×640 的尺寸,并用 Mosaic 数据增强进行优化,同时也可以选择是否根据实际的待检物大小进行自适应锚框或者调整超参数文件。Backbone 由卷积块(CBS)、C3 模块和快速空间金字塔池化(spatial pyramid pooling-fast, SPPF)组成,主要负责特征提取。Neck 由特征金字塔网络(feature pyramid network, FPN)和路径聚合网络(path aggregation network, PAN)两部分组成双塔结构网络,将 Backbone 提取到的特征信息进行多尺度融合以增强目标特征信息多样化。Head 一般由 3 个不同大小的检测头组成,主要用于目标的检测和分类,采用完全交并比作为回归损失函数,再通过 NMS(non-maximum suppression)筛选生成边界框并预测类别及置信度。

2 本文模型

2.1 PGS-YOLO 模型

为了提高现有带钢表面缺陷检测的精度和效率,并兼顾终端检测设备的计算能力与内存资源的局限性,本文提出一种能有效提高检测精度并减少网络参数量与计算量的 PGS-YOLO 检测模型,该模型整体布局与原结构相似,设计的结构如图 1 所示,主要有 3 方面的改进。1) 在激活函数方面,本文将传统卷积的 SiLU(sigmoid linear unit)激活函数替换为更为灵活的 PReLU 激活函数;该激活函数通过引入一个可学习参数,自适应地调整负区域斜率,能有效避免“神经元死亡”问题,更好地适应工业场景中带钢表面缺陷目标复杂的数据分布,提升模型的检测精度,并增强模型的泛化能力;2) 由于输入特征图在逐层特征提取和空间变换等操作过程中会丢失大量信息,受 RepVGG(Re-param VGG)和 GELAN(generalized efficient layer aggregation network)启发,在 C3 的基础上将更新后的 Re-VGG 融入其中形成 Re-C3 模块,该模块能够充分利用各种卷积核组合,更好地应对复杂任务,提高计算效率。为进一步提高参数利用率并增强模型的表达能力,在 Re-C3 之后分别引入幽灵卷积(GhostConv)和 CBP,构建一种全新的 G-GELAN 特征提取网络替换原 C3 模块;该轻量级网络能有效减少输入特征图在特征提取和空间变换等操作中丢失的大量信息,从而提高模型的检测精度与效率;3) 针对卷积神经网络在深度增加时可能出现的有效信息丢失问题,取消了颈部网的部分卷积层。考虑到高维结构语义信息丰富但不易捕获特征信息,本文选取 SCDOWN 进行下采样操作,该模块结合空间维度和通道维度进行下采样,通过高效特征压缩和空间下采样能有效地减少特征图的大小以及降低计算复杂度;同时保留关键缺陷特征,有助于提升模型在复杂背景

下的目标识别能力。

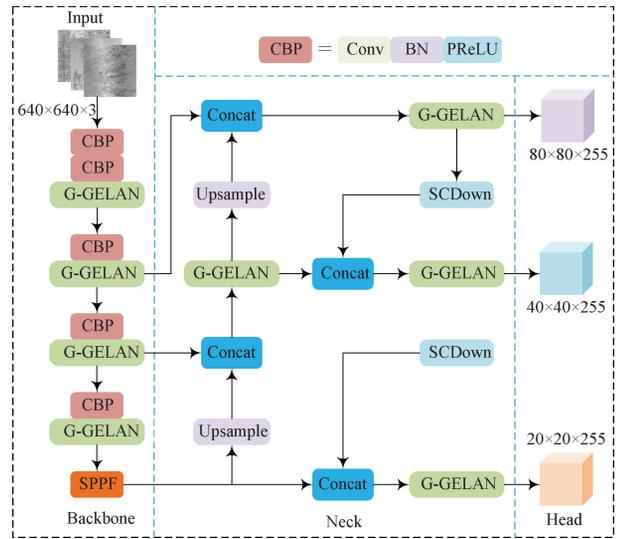


图 1 PGS-YOLO 模型结构

Fig. 1 Structure diagram of the PGS-YOLO model

2.2 PReLU 激活函数

激活函数在缺陷检测任务中起着至关重要的作用,其非线性设计使得网络模型能够学习和表示更加复杂的函数关系或特征,从而增强网络的稳定性。常见的激活函数包括 Sigmoid、ReLU 和 SiLU 等。ReLU 将输入数据分为正负区域,负区域数据被摒弃,仅保留正区域数值并进行非线性转换。相比之下,SiLU 并未直接舍弃负区域数据,它在接近零的区间内拥有更平滑的曲线,然而在负区域的斜率依然会逐渐趋向于 0;由于带钢缺陷数据集较小且具有缺陷种类较多、类间相似高、类别差异大、图像亮度低以及背景信息复杂等特点,模型在训练过程中易产生梯度消失或爆炸问题,难以灵活地学习和捕捉复杂的缺陷特征。因此,为深层网络选择一种更加灵活且表达能力强的激活函数显得尤为重要。

PReLU 激活函数^[22]在 ReLU 的基础上引入一个可学习参数,且该参数的数量等同于通道数。该激活函数可以根据特定任务自适应调整负区域的斜率,在略微增加计算负担的同时,有效避免了 ReLU 激活函数可能会出现“神经元死亡”现象,改善了梯度消失等问题,降低过拟合风险,并提高了模型的灵活性和检测精度,其公式如式(1)所示。

$$f(x_i) = \begin{cases} \alpha_i, & x_i > 0 \\ \alpha_i x_i, & x_i \leq 0 \end{cases} \quad (1)$$

式中: i 为不同的通道; x_i 为第 i 通道上的非线性输入; α_i 为控制负区域斜率的可学习参数。

2.3 G-GELAN 模块

C3 模块在传递残差特征时虽然未增加输出深度,但

信息的部分丢失可能会削弱模型的检测性能,并在一定程度上增加了模型的参数量。因此,设计或选择合适的特征提取模块,以尽可能获取足够的信息,同时减少冗余数据并增强特征表达能力显得尤为重要。受 RepVGG^[23]和 GELAN^[24]思想启发,本文在 GELAN 的基础上设计一

种全新的特征提取模块——G-GELAN,经过该模块处理后,得到的特征图融合了多尺度、多层次的高级语义信息和基本细节,提高模型对不同形态和背景的适应能力,尤其在检测复杂背景下的困难样本时表现较优越,其网络结构如图 2 所示。

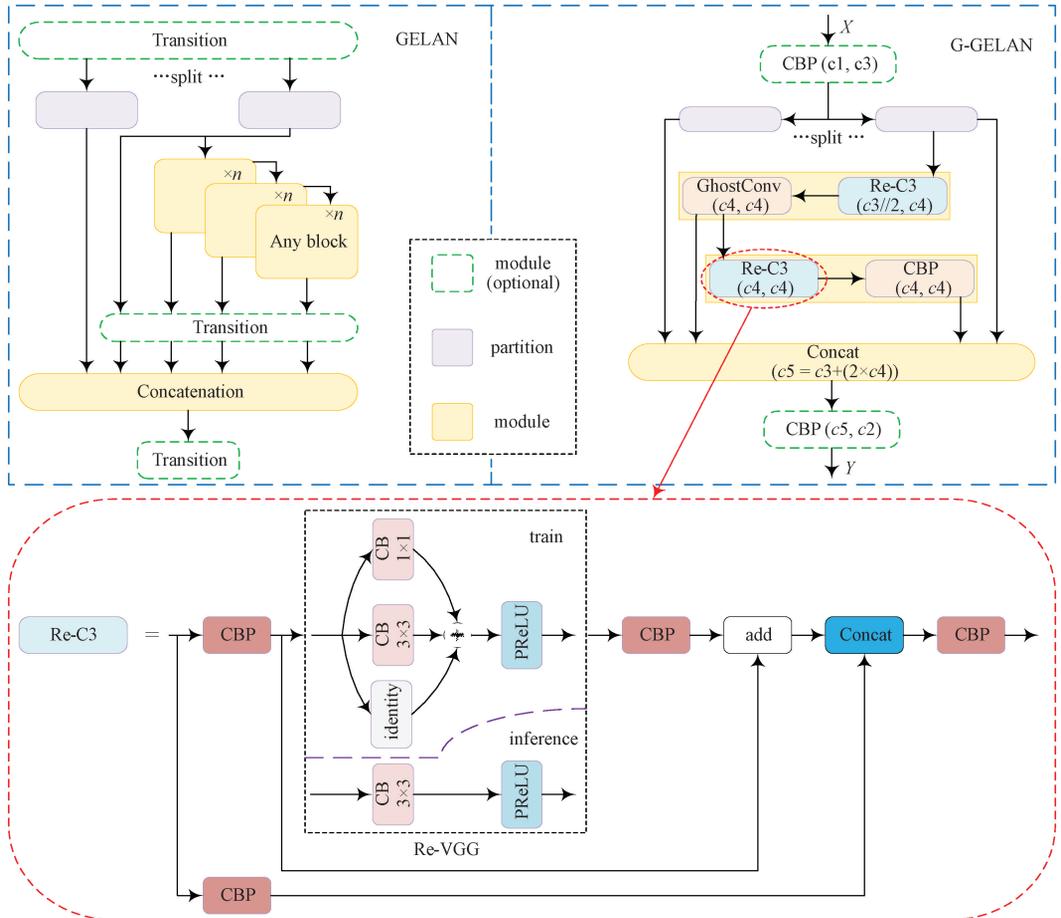


图 2 G-GELAN 结构
Fig. 2 G-GELAN structure diagram

为更好地拟合数据并提高内存利用率,重新设计 C3 模块,采用结构重参数化解耦训练时的多分支与推理时的普通架构,并将更新后的 Re-VGG 融入其中形成 Re-C3 模块。为适应不同任务的需求,在 G-GELAN 中共保留两个 Any block。输入特征图首先经过主干 CBP 卷积层进行特征提取和通道变换,然后使用 split 运算符将 $C \times H \times W$ 的输入张量拆分为两个尺寸相同的张量,以增加梯度源的数量并减小特征图的空间维度。其中一部分张量经过计算块处理,考虑到模型的检测准确性和效率,在两个组合块的首位置分别添加了 Re-C3 模块。由于 GELAN 仅需使用传统的卷积算子,就能实现比基于最先进技术的深度卷积有更好的参数利用率^[25],因此在每个计算块中均添加了卷积操作。不同之处在于,第 1 个块的末尾将 CBP 更换为 GhostConv,该卷积首先通过普通卷积提

取特征,然后利用线性组合以较少的参数生成更多的幽灵特征,在保留重要特征的同时能够有效提高计算效率并降低模型复杂性,从而增强模型的特征表达能力。另一部分张量则直接与经计算块处理后的特征图进行拼接,最后通过主干末端的 CBP 层输出特征图。

2.4 SCDOWN 模块

下采样是提升模型效率及特征提取能力的关键操作,但也伴随着信息丢失和定位精度降低的风险。传统卷积在深层网络中对特征图进行下采样处理时,会对所有通道的特征均匀处理,这可能无法有效保留重要特征信息,同时还会导致参数和计算量的冗余,进而影响模型的检测精度和推理速度。SCDown 是一种轻量级的下采样方法^[26],同时考虑从通道和空间两个维度进行下采

样。通过更加灵活的通道处理,减少冗余计算,其采样过程如图 3 所示。首先,使用点卷积按深度的加权方向生成新的特征图,仅在通道维度上进行操作而不涉及空间维度,从而减少通道数,降低计算复杂度和参数数量;随后,采用深度卷积对每个输入通道分别进行卷积,在维持空间信息完整的同时进行空间下采样处理,有效地捕捉

空间特征。通过这两个卷积操作,SCDown 不仅实现了高效的下采样,还更好地保留了特征图中的关键信息,有效减少信息损失,进而提高模型的检测精度和效率。此外,该模型具有很强的灵活性,可根据具体任务需求调整通道和空间的处理方式,进一步增强网络的适应性。

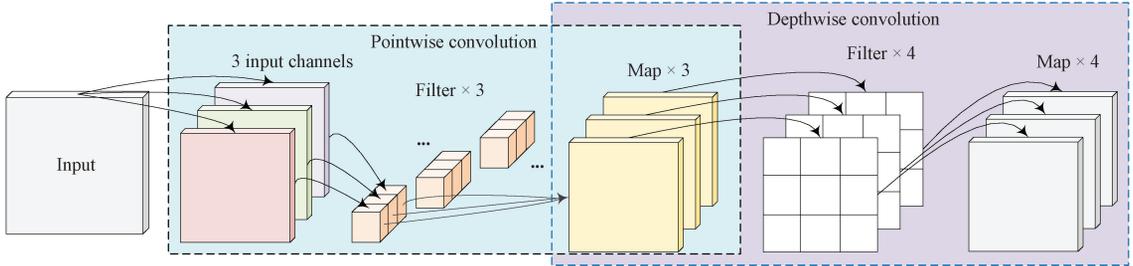


图 3 SCDown 下采样过程

Fig. 3 SCDown downsampling process

3 实验结果与分析

3.1 实验环境及数据集

本文实验环境所需的软硬件配置如表 1 所示。为保证实验的合理与规范性,除非特别说明,本文默认采用表 1 中的配置进行实验。实验训练的参数设置如下:训练 epoch(周期)设置为 300, batch-size(批处理)大小为 16, 交并比(IoU)阈值设置为 0.5, 输入图片尺寸为 640×640, 模型初始学习率为 0.01, 在训练中通过余弦退火算法进行调节, 动量大小为 0.937。

表 1 实验环境的软件和硬件配置

Table 1 Software and hardware configuration of the experimental environment

名称	参数配置
操作系统	Windows10 + 64 GB
CPU	Intel(R) Xeon(R) CPU E5-2686 v4 2.30 GHz
GPU	NVIDIA RTX 3060
深度学习框架	Pytorch2.3.0 + cuda12.1
编译软件	PyCharm

实验采用由东北大学制作的 NEU-DET 数据集^[27]、北京大学制作的 PCB_DATASET 数据集^[28]以及从真实工业环境中收集的带钢表面缺陷数据集 WF10-DET 进行研究。NEU-DET 带钢表面缺陷数据集包含 6 个常见的类别缺陷,分别为裂纹(crazing, Cr)、夹杂(inclusion, In)、斑块(patch, Pa)、麻点(pitted-surface, Ps)、氧化压入(rolled-in-scale, Rs)、划痕(scratches, Sc),每个类别有 300 张;从中按照 8:1:1 的比例随机划分训练集、验证集和测试集,相应图片数量分别为 1 440:180:

180,得到的数据集缺陷信息如图 4 所示。WF10-DET 数据集由底层不良(poor bottom)、色泽不均(uneven color)、接头(connector)、边浪(edge wave)、褶皱(corrugation)、月牙(crescent)、折印(crease)、炉底辊印(stick mark)、压印(coining)、废品(waste)10 种缺陷类型构成,包含 1 000 张分辨率为 4 096×4 096 的灰度图像;PCB_DATASET 数据集共有 693 张瑕疵印刷电路板图片,含有缺孔(missing_hole)、鼠咬(mouse_bite)、断路(open_circuit)、短路(short)、毛刺(spur)和残铜(spurious_copper)6 类缺陷;同样将这两个数据集按照 8:1:1 的比例随机划分该数据集,3 种数据集的典型缺陷如图 5 所示。为减少因 WF10-DET 数据集中图片数量较少导致的过拟合等问题,对其训练集进行数据增强,扩充比例为 1:3,包括垂直翻转、高斯噪声和随机亮度,部分原图及其增强后的图像如图 6 所示。

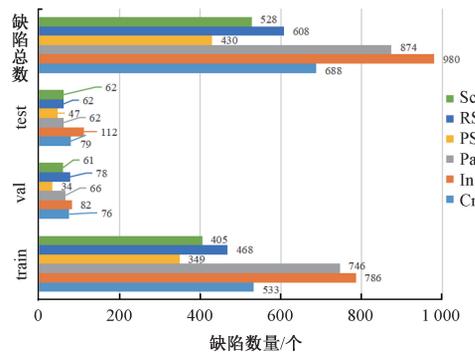
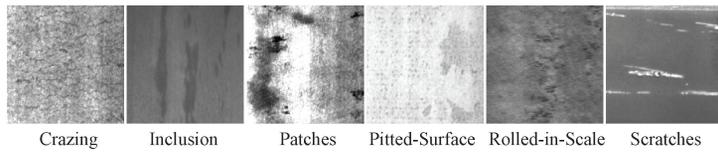


图 4 NEU-DET 数据集缺陷信息

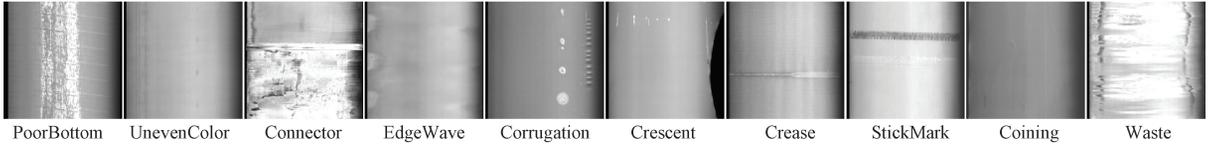
Fig. 4 Defect information of the NEU-DET dataset

3.2 实验评价指标

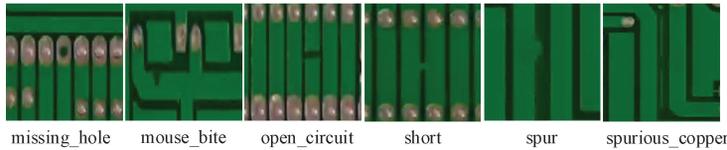
在目标检测算法测试中,主要指标包括精



(a) NEU-DET缺陷类型
(a) NEU-DET defect types



(b) WF10-DET 缺陷类型
(b) WF10-DET defect type



(c) PCB_DATASET缺陷类型
(c) PCB_DATASET defect type

图 5 典型缺陷图

Fig. 5 Typical defect map

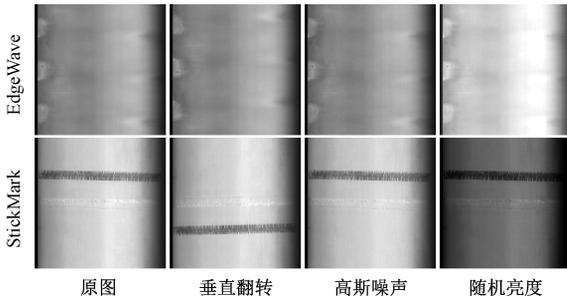


图 6 数据增强前后对比

Fig. 6 Comparison before and after data enhancement

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)、参数量 (parameter)、浮点运算次数 (floating point operations, FLOPs)、每秒帧率 (frame per second, FPS) 和模型权重大小 (weights) 等。本文选取 AP 作为每个类别缺陷的评估指标, 以 mAP 为主要评估指标, 同时参考浮点运算次数、参数量、检测速度等指标来评估模型的整体网络性能。精度和召回率公式如式(2)和(3)所示, 平均精度、平均精度均值和每秒帧率的计算公式如式(4)~(6)所示。

$$P = \frac{TP}{TP + FP} \quad (2)$$

$$R = \frac{TP}{TP + FN} \quad (3)$$

$$AP = \int_0^1 P(R) dR \quad (4)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (5)$$

$$FPS = \frac{1\ 000}{T_{time}} \quad (6)$$

式中: TP (true positive) 为正样本被正确预测的数量; FP (false positive) 为正样本被错误预测的数量; FN (false negative) 为负样本被错误预测的数量; P 代表被检样本中正类占全部正样本的比重, R 代表所有正样本数被正确检测的比重; N 为类别数; T_{Time} 为预处理时间、推理时间以及后处理时间之和。

3.3 实验结果与分析

1) 激活函数实验

为验证 PGS-YOLO 检测模型中 PReLU 激活函数在提升模型学习复杂特征能力方面的有效性, 本文在 YOLOv5 的基础上引入不同的激活函数进行对比, 其中基线 YOLOv5 默认采用的是 SiLU 激活函数, 实验结果如表 2 所示。

表 2 激活函数对比实验

Table 2 Activation function comparison experiments

激活函数	mAP/%	Parameters	计算量/GFLOPs	P/%	R/%
YOLOv5	73.2	7 026 307	15.8	64.6	71.3
MetaAconC	73.7	7 425 459	16.1	65.1	72.4
PReLU	74.2	7 026 364	15.8	65.2	74.3
Hardswish	73.9	7 026 307	15.8	68.3	70.9
SELU	72.1	7 026 307	15.8	64.9	69.5

根据表 2 数据可知,除了 SELU,其他引入的激活函数均在一定程度上提高了模型的 mAP 值。PReLU、Hardswish 和 SELU 这 3 种激活函数在保持与原 YOLOv5 相同计算量的前提下,展现了良好的性能;其中采用 PReLU 激活函数的模型在仅增加少量参数量的情况下,其 mAP 值相较基线提高了 1.0%,达到了最优表现。

2) 改进点消融实验

为了验证改进模型对模型性能的影响以及有效性,本文以 YOLOv5 为基准模型,在 NEU-DET、WF10-DET 和 PCB_DATASET 数据集上对各改进点进行消融实验和分析。一类数据集上设置 5 组实验,其中组别 1、6、11 是基准模型,组 2~4、7~9、12~14 为在基准模型的基础上所做的改进,组 5、10、15 则为 PGS-YOLO 模型。实验结果如表 3 所示。

表 3 消融实验结果

Table 3 Results of ablation experiments

数据集	组别方案	PReLU	G-GELAN	SCDown	mAP/%	Parameters	计算量/GFLOPs	P/%	R/%	Weights/MB
NEU-DET	1				73.2	7 026 307	15.8	64.6	71.3	14.5
	2	✓			74.2	7 026 364	15.8	65.2	74.3	14.5
	3		✓		78.0	5 588 347	12.0	70.4	74.3	11.8
	4			✓	76.4	6 374 787	15.3	67.4	76.1	13.2
	5	✓	✓	✓	79.9	4 936 940	11.5	74.2	72.3	10.6
	6				72.3	7 037 095	15.8	70.2	66.8	14.5
WF10-DET	7	✓			73.4	7 037 152	15.9	72.4	69.2	14.5
	8		✓		78.3	5 599 135	12.0	76.5	73.2	11.9
	9			✓	75.9	6 385 575	15.3	74.6	73.9	13.2
	10	✓	✓	✓	81.4	4 947 615	11.5	75.0	75.5	10.6
	11				92.0	7 026 307	15.8	93.2	86.8	14.5
	12	✓			93.0	7 026 364	15.8	92.8	85	14.5
PCB_DATASET	13		✓		93.5	5 588 347	12.0	94.1	90.3	11.8
	14			✓	93.5	6 374 787	15.3	93.8	89.8	13.2
	15	✓	✓	✓	94.5	4 936 940	11.5	94.9	91.6	10.6

注:“✓”:表示添加,加粗部分为最优值

通过表 3 的消融实验结果可知,3 种改进方法的加入在 3 种不同数据集上的检测精度均有一定的提升。组 2、7、12 仅对基准模型的 SiLU 激活函数进行更改,mAP 值分别增加 1.0%、1.1%和 1.0%,尽管增加了极少量的参数量,但权重大小并未改变,与组 1 保持一致;说明 PReLU 引入一个可学习参数 α 具有很强的灵活性,可以根据特定任务自适应调整负区域的斜率,解决了原激活函数在图像背景信息复杂和较小数据集条件下易出现梯度消失或检测精度低的问题。组 3、8、13 设计了一种全新的特征提取模块替代了 YOLOv5 的 C3 模块,mAP 值较基准模型分别提升 4.8%、6.0%和 1.5%,同时参数量和计算量减少 1 437 960、3.8 GFLOPs;表明设计的 G-GELAN 模块能够更好的保留输入特征图的原始信息,减少冗余信息,优化模型的参数量和计算复杂度,达到了轻量化的效果。组 4、9、14 使用轻量级 SCDown 模块替代基准模型颈部网中的下采样操作,在降低参数量和计算量的同时 mAP 值分别提升 3.2%、3.6%和 1.5%,说明其简单高效的设计不仅节省了计算成本,还能有效地捕获空间特征,保留更多的图像细节信息。组 5、10、15 则为本文所提的 PGS-YOLO 检测模型,综合了所作改进组的优点,mAP 值达 79.9%、81.4%和 94.5%,且参数量和计算量均有所降低;表明本文模型具有一定的合理性,在提

高检测精度的同时能有效降低模型的复杂度,实现了轻量化设计。

3) 不同算法模型对比实验

为验证本文所提出的改进模型在带钢表面缺陷检测任务中的性能,选取了几种目前主流的目标检测模型,并在 NEU-DET 数据集上进行对比,如两阶段的目标检测模型 Faster R-CNN^[29],以及单阶段的 SSD、YOLOv3-tiny、YOLOv5s、YOLOv5s-GG-FPN^[14]、YOLO v8、EHH-YOLOv8s、YOLOv11、Hyper-YOLO^[30]和 YOLOv12^[31]等目标检测模型,对比实验结果如表 4 所示。

由表 4 可知,本文提出的 PGS-YOLO 带钢表面缺陷检测模型在 mAP、参数量、计算量和权重大小方面均表现出色。Faster R-CNN 的 mAP 高于部分单阶段检测模型,尽管检测精度与 YOLOv7 相同,但参数量却高出 YOLOv7 一个数量级,符合两阶段检测特点,不易部署于工业现场。YOLOv3-tiny 虽然检测精度最低,但在原始模型中,其 FPS 表现最佳;文献[14]设计的 GG-FPN 达到了较好的轻量化,但 mAP 低于基准模型 1.5%;YOLOv5s 和 YOLOv5l 在基准模型上增加了网络深度,导致浮点计算量等指标均有所增加,mAP 相应提升 2.1%;在对比的原始模型中,YOLOv8 的检测精度最高,但其参数量较大,计算量和权重大小几乎是基准模型的 2 倍,这不符合

表 4 不同模型对比实验

Table 4 Comparative experiments of different models

模型	AP/%						mAP/%	Parameters	计算量/GFLOPs	FPS/fps	Weights/MB
	Cr	In	Pa	Ps	Rs	Sc					
Faster R-CNN	36.5	78.1	93.2	83.5	60.1	91.6	73.8	136 791 449	200.9	28.3	113.6
SSD	30.0	68.6	86.2	74.9	71.8	62.4	65.6	24 414 218	137.9	52.0	97.7
YOLOv3-tiny	31.4	72.7	88.4	31.2	46.2	54.8	54.1	8 681 426	13.0	120.5	17.5
YOLOv5s	35.3	80.7	83.9	85.9	63.7	89.2	73.2	7 026 307	15.8	68.5	14.5
YOLOv5s-GG-FPN	34.5	83.4	86.8	77.2	65.3	83.1	71.7	4 410 019	10.9	68.9	8.7
YOLOv5l	42.7	84.2	87.8	84.9	59.8	92.5	75.3	46 135 203	107.7	61.1	92.9
YOLOv7	43.1	83.3	87.2	81.4	56.8	91.0	73.8	36 508 742	103.2	59.2	74.5
YOLOv8	44.2	85.2	85.1	91.1	64.2	91.0	76.8	11 127 906	28.4	70.9	22.6
EHH-YOLOv8s	46.3	86.7	90.6	91.4	65.8	92.3	78.9	7 002 683	15.6	71.3	16.7
YOLOv10	37.3	78.2	90.0	86.4	54.5	91.9	73.0	8 039 604	24.5	66.7	16.6
YOLOv11	37.6	86.6	91.4	91.4	61.5	90.7	76.5	9 415 122	21.3	73.6	19.2
Hyper-YOLO	40.6	82.6	90.9	87.4	61.5	90.1	75.5	13 518 370	33.8	63.3	27.5
YOLOv12	35.9	85.8	90.4	88.5	63.1	92.9	76.1	9 233 202	21.2	71.6	19.0
本文	44.7	87.7	92.0	89.0	70.5	95.3	79.9	4 936 940	11.5	70.4	10.6

注:加粗部分为最优值

工业轻量化的要求;Hyper-YOLO 模型,在 NEU-DET 数据集上的整体性能并未表现出色,检测精度只有 75.5%且计算量和参数量较大;在同一数据上,YOLOv12 相比 YOLOv11 拥有较少的参数量和计算量,但 YOLOv11 的 mAP 和检测速度均高于 YOLOv12;相比之下,EHH-YOLOv8s 作为改进后的模型,较好地平衡了检测精度和轻量化,但与本文模型相比不具备优势;虽然 YOLOv7、YOLOv8、YOLOv11 和 YOLOv12 版本模型在 NEU-DET 数据集上的 mAP 值高于 YOLOv5,但它们的计算量、参数量和权重大小也显著高于 YOLOv5;PGS-YOLO 与基准模型相比,mAP 高达 79.9%,在多个类别上的检测精度均有所提升,尤其是在 Cracking 类别上,检测精度由原来的 35.3%提升至 44.7%,这验证了本文模型针对困难样本有较好的检测效果,且检测速度较 YOLOv5s 提升了 1.9 fps。综上所述,本文所改进的模型在与主流检测模型的对比中,表现出更优的性能,以简易模型实现了较高的检测精度和效率,可以满足对带钢表面缺陷检测的更高要求,更易结合现场实际开展部署。

4) 检测效果对比

为更加直观地评估 PGS-YOLO 模型和 YOLOv5s 在带钢表面缺陷检测中的性能差异,本文将改进模型与基准模型在相同测试集上进行定性分析。置信度和 IoU 阈值分别设定为 0.25、0.45,检测结果如图 7 所示。由图 7 可以看出,带钢表面缺陷呈现较大的类别差异、高度的类间相似性以及多类别随机分布等特点,这也是 YOLOv5s 在处理较小或较大目标时容易出现漏检或误检的主要原因。通过比对结果可知,本文模型在多个类别上的检测效果明显优于基准模型,特别是针对困难样本、小尺度以及特征模糊等缺陷,具有较好的定位和分类能力,有效缓解了漏检、误检等问题,进一步验证了本文模型的有

效性。

5) 泛化能力对比实验

为了进一步验证 PGS-YOLO 检测模型的泛化能力,在 WF10-DET(A)和 PCB_DATASET(B)数据集上对基准模型、改进模型和 YOLOv5-TDD^[28]等算法模型进行实验。由于 PCB 缺陷尺寸较小,在真实标注框图中引用了颜色箭头作为辅助标识。不同数据集的泛化实验结果如表 5 和图 8 所示。

表 5 不同数据集的实验结果

Table 5 Experimental results for different datasets

数据集	模型	mAP/%	Parameters	计算量/ GFLOPs	FPS/ fps
A	YOLOv5	72.3	7 037 095	15.8	64.7
	EHH-YOLOv8s	79.2	7 013 471	15.6	66.8
	Hyper-YOLO	75.3	13 529 158	33.8	63.4
	本文	81.4	4 947 615	11.5	67.5
B	YOLOv5	92.0	7 026 307	15.8	66.7
	EHH-YOLOv8s	94.2	7 002 683	15.6	69.2
	YOLOv5-TDD	95.1	8 627 494	17.9	64.9
	本文	94.5	4 936 940	11.5	69.6

注:加粗部分为最优值

由表 5 实验结果可知,本文模型在真实数据集和 PCB 板上均表现出色,取得了较高的 mAP 值;对于复杂纹理、小目标缺陷有一定的检测能力,同时具备较好的实时性,表现出较强的泛化性。与基准模型相比,PGS-YOLO 模型在 A 数据集上的 mAP 获得了 9.1%的增益,达到 81.4%,模型的参数量和计算量分别减少了 2 089 480、4.3 GFLOPs,检测速度提高 2.8 fps。在 B 数据集上,相比基准模型 mAP 提高了 2.5%,在模型的参数量和计算量减少的同时,检测速度提升了 4.3%;对比模型 EHH-YOLOv8s 较好的均衡了检测精度和速度,

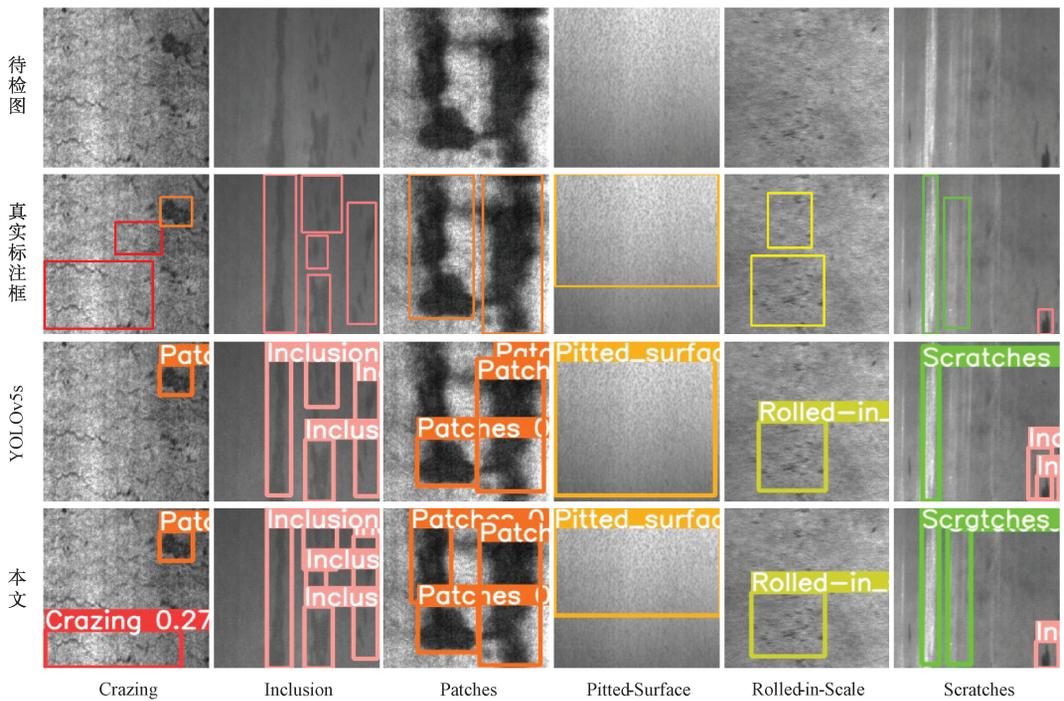


图 7 检测效果

Fig. 7 Detection effect

YOLOv5-TDD 取得了较高的检测精度, mAP 为 95.1%, 甚至高于本文模型, 但其参数量和计算量分别比本文模型高出 3 690 554、6.4 GFLOPs, 占用了较多的计算资源, 不易部署生产实际。

从图 8 可以看出, 针对缺陷种类复杂多样、无固定特征表现、微小缺陷图像占比小且与背景对比度低等问题, 本文改进模型表现出了较好的效果。在真实数据集的多个类别检测中, PGS-YOLO 模型的表现优于基准模型; 尽管在色泽不均类别中检测效果较差, 但整体模型能够更精确地捕捉缺陷信息, 有效减少了漏检和误检情况, 并取得了较高的置信度。在 PCB 板上除缺孔和短路两类缺陷的置信度略低基准模型外, 其余类别均维持较高的精度。对进一步突出了 PGS-YOLO 检测模型的优越性, 更符合实际工业检测需求, 为带钢安全高质量生产提供了可靠保障。

4 结 论

带钢的表面缺陷对其质量影响极大, 甚至危及生产安全。针对带钢表面缺陷类别差异大、类间相似高而导致缺陷检测准确率低且泛化能力不强的问题, 本文提出了一种灵活高效的 PGS-YOLO 带钢表面缺陷检测模型。首先更换 SiLU 激活函数为 PReLU 激活函数, 通过增加可学习参数方式避免“神经元死亡”问题, 更加精确地拟

合特征分布, 提高模型性能; 其次构建一种轻量级的 G-GELAN 特征提取模块, 通过融合多层次、多尺度的信息, 实现缺陷特征的高效提取, 在保持高精度的同时, 也显著降低模型的计算开销; 最后在颈部网络中选择轻量级的 SCDown 下采样, 通过更灵活的通道和空间处理, 使得模型能更好地保留有效特征, 增强对复杂背景和不同尺寸缺陷的适应能力, 进一步提高了计算效率。

在 NEU-DET 数据集上, 本文提出的 PGS-YOLO 检测模型相较于基准模型, mAP 提升了 6.7%, 达到 79.9%, 同时参数量和计算量分别减少 29.7%、27.2%, 权重大小下降至 10.6 MB, 且 FPS 提高了 2.7%; 特别对于裂纹困难样本, 检测效果显著, mAP 提升了 9.4%。实验结果表明, 该模型不仅具备较高的检测精度和简洁的结构设计, 还实现了轻量化, 有效地平衡了检测精度、计算量和速度之间的关系, 且展现出良好的泛化能力。

参考文献

[1] 孙建亮, 孙孟乾, 郭贺松, 等. 热轧带材多缺陷和单缺陷表面质量综合预报[J]. 钢铁, 2021, 56 (1): 51-58.
 SUN J L, SUN M Q, GUO H S, et al. Comprehensive prediction of surface quality of hot-rolled strip with multiple and single defects[J]. Iron & Steel, 2021, 56 (1): 51-58.
 [2] AMERI R, HSU C C, BAND S S. A systematic review of

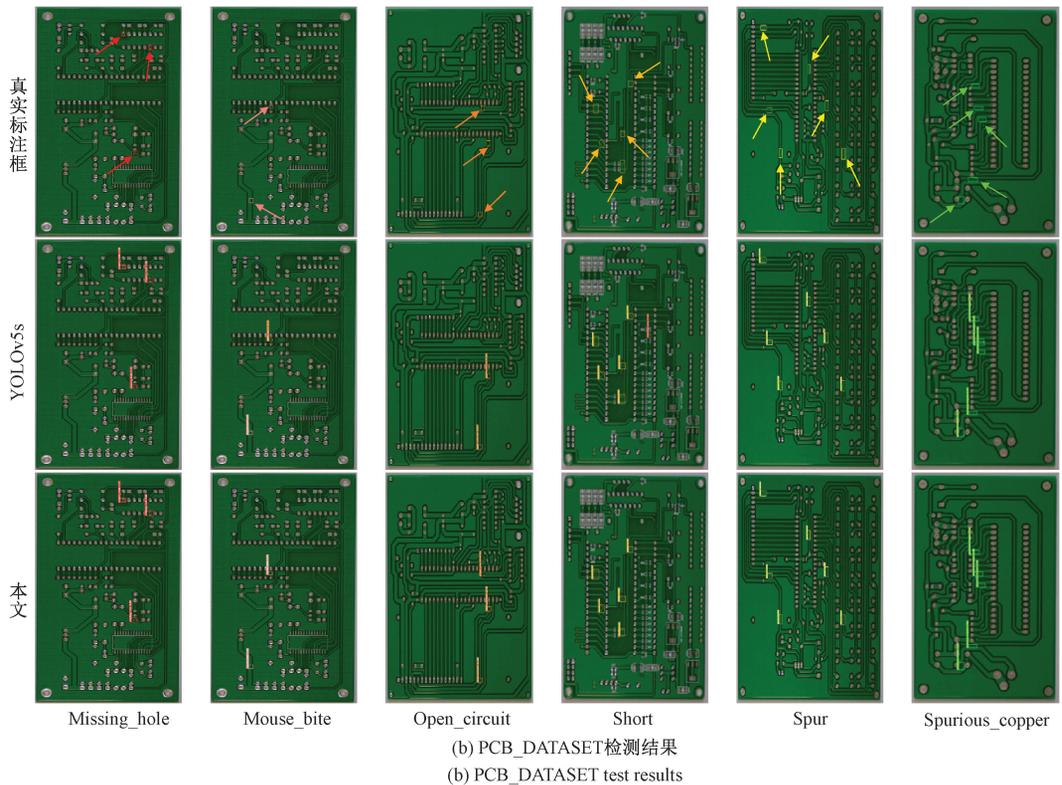
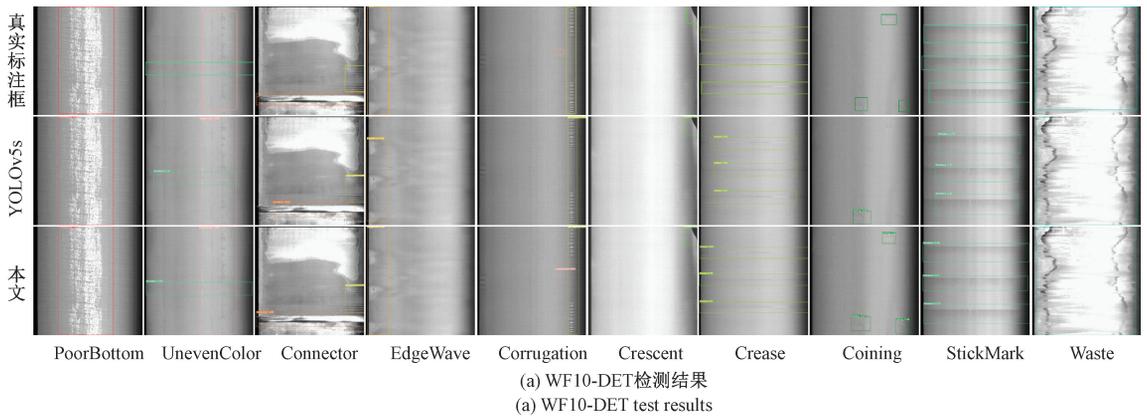


图 8 泛化能力效果

Fig. 8 Effect of generalization ability

deep learning approaches for surface defect detection in industrial applications [J]. *Engineering Applications of Artificial Intelligence*, 2024, 130: 107717.

[3] 伍麟, 郝鸿宇, 宋友. 基于计算机视觉的工业金属表面缺陷检测综述 [J]. *自动化学报*, 2024, 50 (7): 1261-1283.

WU L, HAO H Y, SONG Y. A review of metal surface defect detection based on computer vision [J]. *Acta Automatica Sinica*, 2024, 50 (7): 1261-1283.

[4] LIU G H, CHU M X, GONG R F, et al. DLF-YOLOF: An improved YOLOF-based surface defect detection for steel plate [J]. *Journal of Iron and Steel Research*

International, 2024, 31 (2): 442.

[5] CHEN Y J, DING Y Y, ZHAO F, et al. Surface defect detection methods for industrial products: A review [J]. *Applied Sciences*, 2021, 11 (16): 7657-7657.

[6] 杨春龙, 吕东浩, 张勇, 等. 融合自适应下采样的带钢表面缺陷检测算法 [J]. *钢铁研究学报*, 2024, 36 (6): 806-816.

YANG CH L, LYU D H, ZHANG Y, et al. Fusion of adaptive down-sampling for strip steel surface defect [J]. *Journal of Iron and Steel Research*, 2024, 36 (6): 806-816.

[7] 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 on steel[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38 (4): 108-116.
- [8] 马磊, 李晔, 王宇翔. YOLOv8-FD: YOLOv8 改进的钢板表面缺陷检测方法[J]. 计算机工程与应用, 2024, 60 (24): 211-221.
- MA L, LI Y, WANG Y X. YOLOV8-FD: YOLOv8 improved method for detecting surface defects on steel plates [J]. Computer Engineering and Applications, 2024, 60 (24): 211-221.
- [9] 梁礼明, 龙鹏威, 卢宝贺, 等. EHH-YOLOv8s: 一种轻量级的带钢表面缺陷检测算法[J/OL]. 北京航空航天大学学报, 1-15 [2024-10-29]. <https://doi.org/10.13700/j.bh.1001-5965>. 2024. 0426.
- LIANG L M, LONG P W, LU B H, et al. EHH-YOLOv8s: A lightweight surface defect detection algorithm for strip steel [J/OL]. Journal of Beijing University of Aeronautics and Astronautics, 1-15 [2024-10-29]. <https://doi.org/10.13700/j.bh.1001-5965>. 2024. 0426.
- [10] SINGH S A, DESAI K A. Automated surface defect detection framework using machine vision and convolutional neural networks[J]. Journal of Intelligent Manufacturing, 2023, 34(4): 1995-2011.
- [11] 陶显, 侯伟, 徐德. 基于深度学习的表面缺陷检测方法综述[J]. 自动化学报, 2021, 47 (5): 1017-1034.
- TAO X, HOU W, XU D. A review of surface defect detection methods based on deep learning [J]. Acta Automatica Sinica, 2021, 47 (5): 1017-1034.
- [12] 张航, 周毅, 邱宇峰. 融合 HGnetv2 和注意力机制的钢材表面缺陷检测方法[J]. 电子测量与仪器学报, 2025, 39(1): 36-49.
- ZHANG H, ZHOU Y, QIU Y F. Detection method of steel surface defects with fusion of HGnetv2 and attention mechanism[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(1): 36-49.
- [13] CHEN H, DU Y, FU Y, et al. DCAM-Net: A rapid detection network for strip steel surface defects based on de-formable convolution and attention mechanism [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-12.
- [14] 苏盈盈, 何亚平, 邓圆圆, 等. 基于改进 YOLOv5s 轻量化带钢表面缺陷检测方法[J]. 光电子·激光, 2024, 35 (7): 723-730.
- SU Y Y, HE Y P, DENG Y Y, et al. Surface defect detection method based on improved YOLOv5s for lightweight steel strip [J]. Journal of Optoelectronics Laser, 2024, 35 (7): 723-730.
- [15] WANG L, SONG C, WAN G, et al. A surface defect detection method for steel pipe based on improved YOLO[J]. Mathematical Biosciences and Engineering: MBE, 2024, 21(2): 3016-3036.
- [16] ZHONG H, WU B, ZHANG X, et al. Steel surface defect detection based on an improved YOLOv5 model [C]. Proceedings of the Intelligent Control, Measurement and Signal Processing (ICMSP), 2023: 51-55.
- [17] YANAG G, GANG L, DONG X, et al. Research on steel surface defect classification method based on deep learning [J]. Scientific Reports, 2024, 14 (1): 8254-8254.
- [18] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector [C]. European Conference on Computer Vision. Cham: Springer International Publishing, 2016: 21-37.
- [19] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: Unified, real-time object detection[C]. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016: 779-788.
- [20] ZHENG W, TANG W, JIANG L, et al. SE-SSD: Self-Ensembling single-stage object detector from point cloud[J]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 14489-14498.
- [21] 郎松, 曹选, 张艳微, 等. 融合改进 YOLOv5 算法的图像全站仪全自动测量方法[J]. 仪器仪表学报, 2022, 43 (5): 120-127.
- LANG S, CAO X, ZHANG Y W, et al. Fully automatic measurement method of image total station by incorporating improved YOLOv5 algorithm [J]. Chinese Journal of Scientific Instrument, 2022, 43 (5): 120-127.
- [22] HE K, ZHANG X, REN S, 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.
- [23] DING X H, ZHANG X Y, MA N N, et al. RepVGG: Making VGG-style ConvNets great again [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 13728-13737.
- [24] WANG C Y, YEH I H, LIAO H Y M. YOLOv9:

- Learning what you want to learn using programmable gradient information [J]. ArXiv preprint arXiv: 2402.13616, 2024.
- [25] BALAKRISHNAN T, SENGAR S S. RepVGG-GE LAN; Enhanced GELAN with VGG-STYLE ConvNets for brain tumour detection[J]. ArXiv preprint arXiv:2405.03541, 2024.
- [26] WANG A, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection [J]. ArXiv preprint arXiv. 2405.14458, 2024.
- [27] HE Y, SONG K, MENG Q, et al. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features [J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493-1504.
- [28] 廖鑫婷, 张洁, 吕盛坪. 融合浅层特征和注意力机制的 PCB 缺陷检测方法[J]. 计算机集成制造系统, 2024, 30(3): 1092-1104.
LIAO X T, ZHANG J, LYU SH P. Defect detection for PCB by combining shallow features and attention mechanism [J]. Computer Integrated Manufacturing Systems, 2024, 30(3): 1092-1104.
- [29] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(6): 1137-1149.
- [30] FENG Y, HUANG J, DU S, et al. Hyper-YOLO: When visual object detection meets hypergraph computation [J]. ArXiv preprint arXiv:2408.04804, 2024.
- [31] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors [J]. ArXiv preprint arXiv: 2502.12524, 2025.

作者简介



马俊杰, 2022 年于河南大学民生学院获得学士学位, 现为内蒙古科技大学硕士研究生, 主要研究方向为深度学习、缺陷检测。
E-mail: 206353056@qq.com

Ma Junjie received his B. Sc. degree from Minsheng College of Henan University in 2022. He is currently a M. Sc. candidate at Inner Mongolia University of Science and Technology. His main research interests include deep learning and defect detection.



张继红 (通信作者), 2015 年于内蒙古工业大学获得博士学位, 现为内蒙古科技大学教授, 主要研究方向为电气工程领域、可再生能源发电及智能微电网保护与控制。
E-mail: zjh00318@imust.com.cn

Zhang Jihong (Corresponding author) received his Ph. D. degree from Inner Mongolia University of Technology in 2015. He is currently a professor at Inner Mongolia University of Science and Technology. His main research interests include electrical engineering, renewable energy power generation, and smart microgrid protection and control.