

DOI:10.16410/j.issn1000-8365.2025.5120

面向钢表面缺陷的双模态目标检测方法

李芹芹,王奎越,宋宝宇,宋 君,马晓国
(鞍钢集团北京研究院有限公司 北京 102209)

摘 要:钢表面缺陷检测是工业质量管控核心环节,针对现有基于 RGB 单模态的缺陷检测模型鲁棒性不足,空间形态缺陷误检漏检错检发生率高,提出了并行双模态空间感知融合的 PMSF-YOLOv8(Parallel multi-modal spatial-aware fusion YOLOv8)算法,采用双分支异构网络分别强化 RGB 纹理和深度空间特征学习,在中期融合阶段采用双模态特征融合模块 DFFM(dual-modal feature fusion module)通过动态权重实现多尺度特征自适应融合。使用 NUE-RSDDS-AUG 数据集进行验证,结果表明,PMSF-YOLOv8 网络模型检测准确率 $mAP@0.5$ 达 98.6%,误报率较单模态方法降低 2.1%,实现了“高精度-低误报”平衡。

关键词:双模态;钢表面缺陷检测;YOLOv8;特征融合;注意力机制

中图分类号: TP391

文献标识码: A

文章编号: 1000-8365(2025)10-0973-09

Dual-Modal Target Detection Method for Steel Surface Defects

LI Qinqin, WANG Kuiyue, SONG Baoyu, SONG Jun, MA Xiaoguo
(Ansteel Beijing Research Institute Co., Ltd., Beijing 102209, China)

Abstract: Steel surface defect detection is a core aspect of industrial quality control. Given the insufficient robustness of existing RGB single-modality-based defect detection models, which often suffer from high rates of false positives, false negatives, and incorrect detections of spatial morphological defects, the parallel multi-modal spatial-aware fusion YOLOv8 (PMSF-YOLOv8) algorithm was proposed. This algorithm employs a dual-branch heterogeneous network to enhance the learning of RGB texture and depth spatial features. In the mid-fusion stage, the dual-modal feature fusion module (DFFM) was utilized to achieve adaptive fusion of multiscale features through dynamic weights. The NUE-RSDDS-AUG dataset was used for validation. The results show that the PMSF-YOLOv8 network model achieves a detection accuracy of $mAP@0.5$ of 98.6%, with a false alarm rate reduced by 2.1% compared with that of single-modality methods, striking a balance between "high accuracy and low false alarms".

Key words: dual-modality; steel surface defect detection; YOLOv8; feature fusion; attention mechanism

钢铁作为国民经济的支柱产业,强化钢铁产品质量管控至关重要^[1],高质量的钢产品检测系统能有效地检测出产品的质量。然而,现有工业检测系统面临两大挑战^[2]:传统光学检测易受水渍、油污、锈蚀等环境干扰,导致误检率居高不下;缺陷深度信息如凸起高度、凹坑深度的缺失,难以量化评估缺陷严重程度,进而影响工艺优化决策。

目前,深度学习算法应用于钢产品的质量检测上已经引起国内外学者的广泛关注。Liu 等^[3]率先研究钢表面缺陷的分类问题,但并未解决缺陷定位难题。为攻克定位问题,单阶段与两阶段检测算法^[4]

被相继提出。典型的两阶段检测算法包括 Fast R-CNN^[5]和 Mask R-CNN^[6]。如 Chen 等^[7]采用 Faster R-CNN 对钢轨表面缺陷检测进行研究;Wang 等^[8]针对钢轨缺陷形式的多样性和随机性,提出基于 Mask R-CNN 的新型表面缺陷检测网络来检测钢轨缺陷,有效解决小尺寸缺陷检测难题。与两阶段相比,单阶段检测算法主要有 YOLO^[9-11]系列算法以及 SSD^[12]算法,其检测速度更具优势。Song 等^[13]采用 YOLOv3 算法的深度学习方法来实现钢轨表面缺陷检测。Hu 等^[14]将 Transformer 结构的编码器模块应用于 YOLOv5 模型,以检测钢轨缺陷。杜少聪等^[15]基

收稿日期: 2025-06-23

基金项目: 国家重点研发计划(2022YFB3304800)

作者简介: 李芹芹,1995 年生,硕士,助理研究员。主要从事机器视觉智能感知与检测研究。Email: 18738357832@163.com

引用格式: 李芹芹,王奎越,宋宝宇,宋君,马晓国.面向钢表面缺陷的双模态目标检测方法[J].铸造技术,2025,46(10): 973-981.

LI Q Q, WANG K Y, SONG B Y, SONG J, MA X G. Dual-modal target detection method for steel surface defects[J]. Foundry Technology, 2025, 46(10): 973-981.

于 YOLOv5 设计专用模型用于铁路轨道表面缺陷识别,解决了实时检测小缺陷和集群缺陷。张涛源等^[16]提出 TRSD-YOLO 算法并融合了 Transformer,在带钢表面检测效率以及小目标缺陷检测精度上有了很大的突破。

尽管基于深度学习的钢产品质量检测技术取得一定进展,但工业现场复杂环境导致传统单模态方法依然面临着严峻的挑战。钢表面易存在多种干扰信息,如水渍、氧化锈蚀等,使得钢表面图像信息非常复杂,仅使用二维图像提供的颜色或者纹理难以区分真假缺陷。而深度信息能够衡量钢表面损伤程度,为此考虑结合二维图像和深度信息这两种不同模态的互补优势进行缺陷检测。Wang 等^[17]为了解决无轨钢轨表面缺陷检测存在缺陷边界不规则、前景和背景相似等挑战,使用深度图像为 RGB 图像提供互补的空间信息,提出协作学习注意力网络(CLANet)模型用于特征提取、跨模态信息融合和缺陷定位与分割。赵宏伟等^[18]通过两个骨干网络分别提取二维信息和深度信息,融合阶段采用通道级拼接融合,但该融合方式未考虑这两种模态之间的差异和互相干扰问题,导致融合后的双模特征存在过多的冗余信息,更难捕获小缺陷特征。

根据上述问题,本文提出 PMSF-YOLOv8 模型双模态特征学习与融合的差异化设计,即双分支主干网络采用“模态适配”策略,RGB 分支通过可变形卷积动态调整感受野,针对性解决钢表面划痕、疤痕等不规则形变缺陷的“检不出”问题;深度分支保留标准卷积的平移不变性,有效抑制噪声导致的“检错”。DFFM 模块创新性引入 PPA 并行补丁感知注意力与残差滤波,通过局部-全局多尺度特征建模

与通道-空间自适应增强,解决了传统融合方法中低质量模态主导与特征冗余问题。

1 YOLOv8 算法

YOLOv8 作为典型的目标检测模型,网络主体由 3 部分组成,分别为骨干网络(Backbone)、颈部网络(Neck)以及检测头结构(Head)。Backbone 部分负责特征提取,采用一系列卷积核反卷积,同时使用了残差连接和瓶颈结构来减小网络的大小;Neck 部分通过构建双流特征金字塔网络(feature pyramid networks, FPN)^[20],负责多尺度特征融合,通过将来自 Backbone 不同阶段的特征图进行融合,增强特征表示能力;Head 部分负责最终的目标检测与分类任务。

2 PMSF-YOLOv8 算法

PMSF-YOLOv8 算法构建基于 YOLOv8n 结构,如图 1 所示。其主干网络在改进的 YOLOv8 基础上,设计了一种并行双分支多尺度特征提取与融合策略,在多尺度特征提取之后分别进行多尺度双模态图像特征融合,这种策略既保证骨干网络保留各自模态的独立特征,又可以实现特征之间的互补性。

2.1 双分支骨干网络

双分支骨干网络作为特征提取器,对模型的检测性能起着关键作用。针对钢表面缺陷形态不规则、易受环境干扰的特点,本文提出的基于可变形卷积改进型骨干网络 DeformBackbone(图 1)。与几何结构固定的标准卷积核相比,可变形卷积通过可学习偏移量实现感受野动态调整,如式(1)所示,显著增强对不规则缺陷的几何适配能力:

$$y(p_0) = \sum_{p_k \in R} w(p_k) \cdot x(p_0 + p_k + \Delta p_k) \quad (1)$$

式中, p_0 为输出特征图当前位置; p_k 为预设采样位

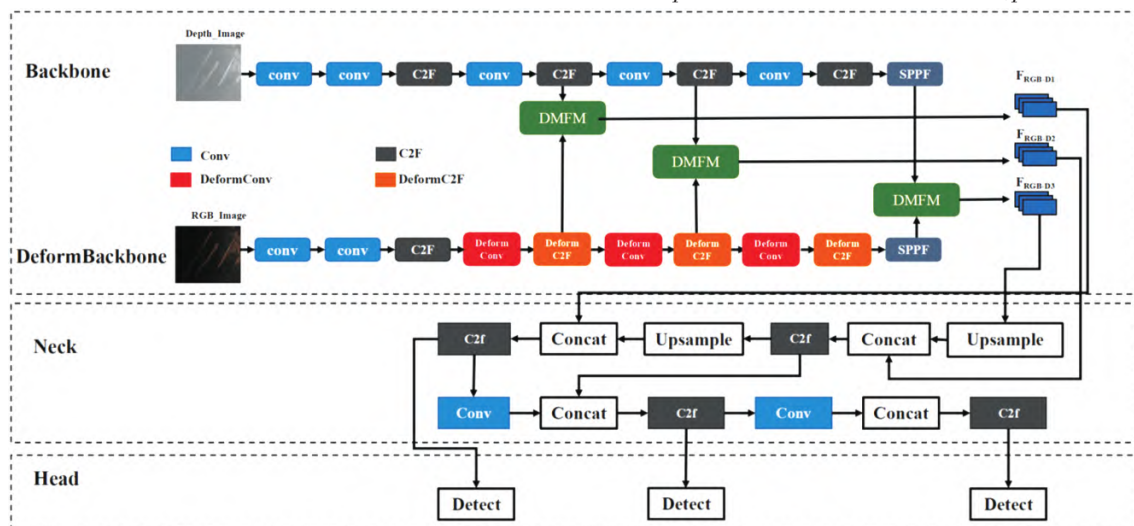


图 1 PMSF-YOLOv8 算法的网络架构

Fig.1 Network structure of the PMSF-YOLOv8 algorithm

置; Δp_k 为网络学习的空间偏移量; $w(p_k)$ 为卷积核的权重。

2.1.1 网络架构设计

如表1所示,DeformBackbone在YOLOv8原始架构基础上进行改进。如图2所示,DeformC2F模块与常规C2F的不同之处在于DeformC2F中的DeformBottleneck单元是将2个 3×3 标准卷积替换为DeformConv,保留跨层短路连接(shortcut)实现多尺度特征融合,通过下式表示:

$$y=x+D[D(x)] \quad (2)$$

式中, D 表示DeformConv操作。

表1 DeformBackbone 参数
Tab.1 DeformBackbone parameters

Type	Filters	Size	Output
Conv	64	3×3	320×320
Conv	128	3×3	160×160
C2F	128	1×1	160×160
		3×3	
DeformConv	256	3×3	80×80
DeformC2F	256	1×1	80×80
		3×3	
DeformConv	512	3×3	40×40
DeformC2F	512	1×1	40×40
		3×3	
DeformConv	512	3×3	20×20
DeformC2F	512	1×1	20×20
		3×3	
SPPF	512	1×1	20×20

2.1.2 双模态差异化策略

模型采用双分支骨干网络并行处理RGB图像和深度图像,基于模态的互补特性实现更鲁棒的缺陷检测。具体策略如下。

(1)对于输入端为RGB图像的分支,由于钢表面疤痕、裂纹等缺陷常呈现局部形变,因此采用可变形卷积构建RGB分支的主干网络。如检测钢表面的划痕时,可变形卷积能聚焦于划痕边缘的突变梯度区域,而传统卷积可能因固定感受野漏检部分片段。这一结论通过消融实验验证。

(2)针对深度图输入分支,深度图通过距离信息表征物体几何结构,但易受环境噪声干扰。为此保留了标准卷积结构,利用其平移不变性抑制噪声。

例如,钢表面水渍在RGB图像中可能被误判为缺陷。为此采用标准卷积通过全权重共享平滑噪声构建深度图分支的主干网络。实验部分的对比实验进一步验证这一论断。

综上,并行双模态特征融合的思想对于RGB模态采用可变形卷积解决“检不出”等形变缺陷;对深度模态采用标准卷积解决“检错”等噪声误报问题。

2.2 双模态特征融合模块 DFFM

在RGB图像中,钢表面复杂的背景和低对比度的缺陷会导致缺陷与背景难以区分。而深度图像在深度较小时,缺陷易被复杂背景掩盖。根据RGB图像和深度图像自身特征,常见的特征融合方法(如简单相加或直接拼接)存在明显局限性。这些方法易忽视低质量图像对整体融合效果的潜在影响,且未能充分考量不同模态特征之间的本质差异。此外,其难以有效利用双模态信息的互补优势,甚至可能出现一种模态特征过度主导而抑制另一种模态特征的现象,或者在融合过程中引入误导性的特征信息,进而产生不利影响。

针对上述问题,为有效避免双模态信息之间的相互干扰并充分发挥其互补特性,构造双模态特征融合模块DFFM,其结构如图3所示。DFFM模块引入的并行化补丁感知注意力模块PPA的网络结构如图4所示。通过引入深度可分离卷积(depthwise separable convolution, DSConv)^[21]来替代传统的常规卷积操作,有效降低了模型的参数量,从而避免了模型过拟合的问题。

DFFM融合模块分为两个阶段。在第一阶段,将RGB特征图 F_{RGB} 与深度特征图 F_{Depth} 进行拼接,从而生成新的特征图 F_{RGB-D} ,具体表达式如式(3)所示。

$$F_{RGB-D}=\text{Concat}(F_{RGB};F_{RGB-D}) \quad (3)$$

通过深度可分离卷积模块对 F_{RGB-D} 执行卷积操作,构建相关性表示。接着,PPA通过其独特的多分支特征提取融合模块和自适应注意力机制模块,捕获不同尺度和级别的特征信息,如图4所示。

多分支特征提取融合包括补丁感知和级联卷积。补丁感知中的参数 p 设置为2和4,分别表示局部和全局分支,不同分支采用不同大小的补丁,建模多尺度特征。具体地,通过将输入特征划分为 $p\times p$ 的

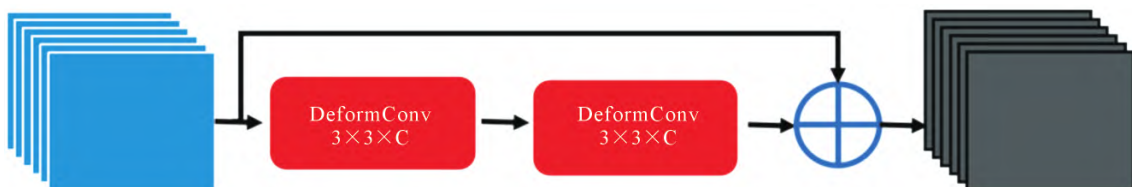


图2 DeformC2f中的DeformBottleneck单元
Fig.2 DeformBottleneck unit in DeformC2f

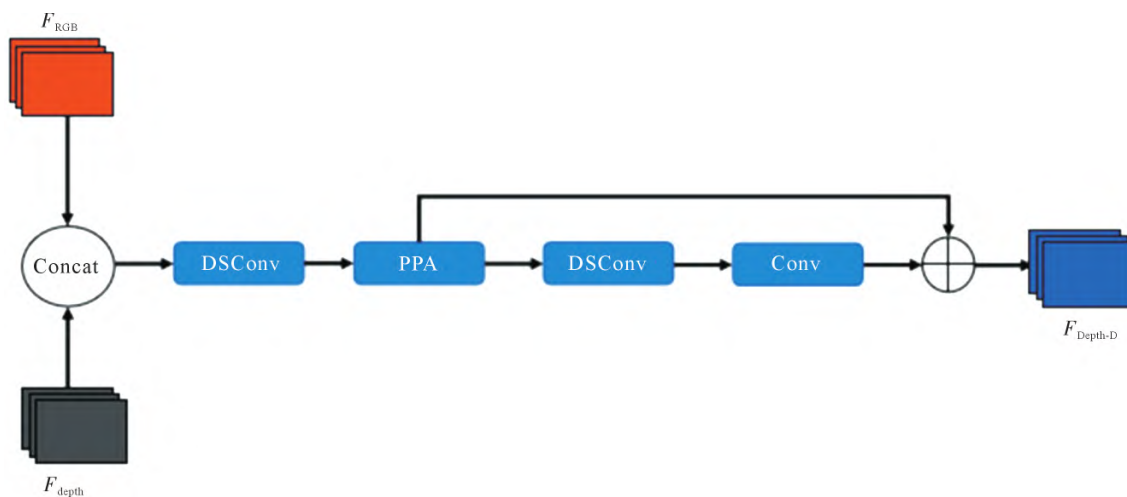


图 3 DFFM 模块结构

Fig.3 The DFFM module structure

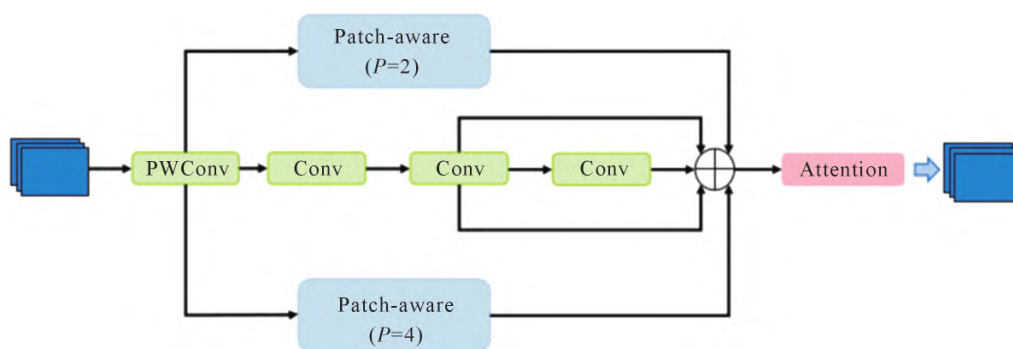


图 4 PPA 模块示意图

Fig.4 Schematic diagram of the PPA module

局部补丁,从而捕获局部上下文信息,并对每个补丁内的特征进行均值处理以生成精简的特征表示。在此基础上,模块通过 MLP 和 Softmax 实现重要特征的筛选,并使用残差加权方式融合局部上下文,最终生成局部特征 F_{local} 和全局特征 F_{global} 。级联卷积包含 3 个 3×3 卷积层串行卷积,逐步融合局部和上下文信息,得到串行卷积输出 F_{conv} 。融合上述 3 个分支的特征,得到综合特征 $\tilde{F}$,如下所示:

$$\tilde{F} = F_{\text{local}} + F_{\text{global}} + F_{\text{conv}} \quad (4)$$

在经过多分支特征提取后,采用自适应注意力机制进行自适应通道和空间维度上的增强, $\tilde{F}$ 在经过一维通道注意力网络模块 M_c 处理后,处理的结果再经过二维空间注意力图 M_s 处理,公式为:

$$F_c = M_c(\tilde{F}) \otimes \tilde{F} \quad (5)$$

$$F_s = M_s(F_c) \otimes F_c \quad (6)$$

$$F = \delta\{B[D(F_s)]\} \quad (7)$$

式中, $\otimes$ 表示逐元素相乘, F_c 和 F_s 分别表示经过通道和空间选择后的特征, δ 、 B 和 D 分别表示修正线性单元、批量归一化和正则化函数(dropout), F' 为 PPA 的最终输出。

在第 II 阶段,通过构建残差连接结构 Res,进一步去除冗余信息,凸显关键特征,从而构建出缺陷在三维空间中的特征表示 F'' ,通过下式计算获得:

$$F'' = F' + \text{Res}(F) \quad (8)$$

3 实验和结果分析

3.1 实验数据集

此次实验使用的数据集为 NEU-RSDDS-AUG^[22]。该数据集包含 4000 对 RGB 图像与深度图像,其图像质量足以支持模型性能的测试以及满足实际需求。数据集涵盖了 4 种典型缺陷,分别是 hole、gutter、scar 和 protrusion。数据集图像大小为 600×600 像素。如图 5 显示了 4 种钢表面缺陷的部分样本图像,这些图像的缺陷位置是随机的,因此满足实际的应用需求。

将数据集按照 0.7:0.15:0.15 的比例划分为训练集、验证集和测试集。每个数据集中包含的各种缺陷数量如表 2 所示。

3.2 实验环境

本实验是在 Pytorch 1.10.0+cu11.1 的框架下搭建的,详细的实验环境如表 3 所示。

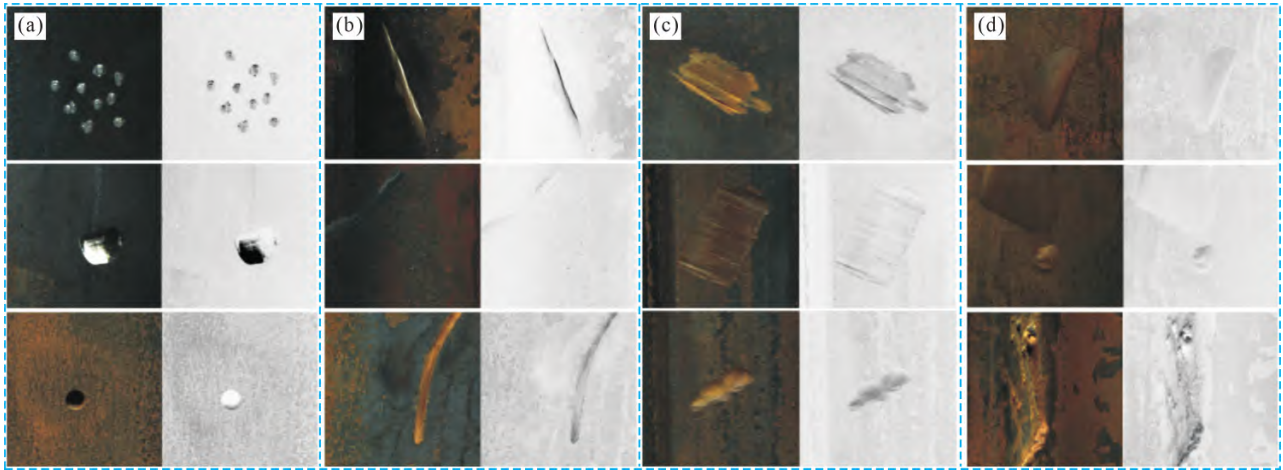


图5 NEU-RSDDS-AUG 数据集:(a) hole; (b) gutter; (c) scar; (d) protrusion

Fig.5 NEU-RSDDS-AUG dataset: (a) hole; (b) gutter; (c) scar; (d) protrusion

表2 钢表面缺陷数据集详细信息

Tab.2 Details of the steel surface defect dataset

Dataset (number of images)	Hole	Gutter	Scar	Protrusion
Train set(2 800)	730	426	952	692
Validation set(600)	157	218	63	162
Test set(600)	132	162	76	230

表3 实验环境

Tab.3 Experimental environment

Software/Hardware	Version
Operating system	Window10
GPU	NVIDIA GeForce RTX 3080Ti(16GB)
CPU	11th Gen Inter(R) Core(TM) i9-11950H@
Programm language	Python 3.8
Deep learning framework	Pytorch 1.10.0
Computing platform	CUDA 11.1
GPU acceleration rate	CUDNN 8005

在训练过程中开启了 Mixup 模式增强数据,避免过拟合。此外将迭代次数 epochs 设置为 200,学习率(learning rate, lr)设置为 0.02, batch size 为 16,其中动量 Momentum 设置为 0.8。训练过程中采用了学习率衰减策略,随着训练轮次 epochs 的增加,学习率逐渐减小。这种方法使得模型训练过程更加稳定,在接近最优解时能够平稳收敛,避免震荡。

3.3 实验结果分析

采用召回率(recall, R)、精确率(precision, P)和平均精度(mean average precision, mAP)作为模型的评价指标, IoU 阈值选取 0.5, 即平均精度为 mAP@0.5。

3.3.1 消融实验

为验证本文算法的有效性,在 NEU-RSDDS-AUG 数据集上进行了消融实验,分别对每个改进点进行钢表面质量检测性能的测试,包括骨干网络中的可变形卷积(DeformConv)、DFFM、以及DFFM 中的 PPA 模块。从实验结果表 4 中得出以下结论。

(1)DeformConv 的影响 通过对比方案 3 和 4,

表4 消融实验

Tab.4 Ablation experiment

Scheme	Add	Concat	DeformConv	+PAA	DFFM	mAP@0.5
1	√					0.921
2		√				0.923
3					√	0.954
4			√		√	0.972
5				√		0.961
6			√	√		0.986

二者的不同之处在于方案 3 中的深度图分支主干网络采用传统卷积,方案 4 引入 DeformConv 替代传统卷积,与方案 3 相比,方案 4 的 mAP@0.5 从 0.954 提升到 0.972,表明 DeformConv 的引入能够更好地适应目标的形变和姿态变化,解决“检不出”等形变缺陷。

(2)DFFM 的影响 通过对比方案 1、2 和 5,方案 1 的 Add 模块表示两个分支多尺度融合采用直接相加的方式,方案 2 的 Concat 模块表示两个分支多尺度融合采用拼接操作,方案 5 采用引入 PAA 模块的 DFFM 模型进行两个分支多尺度融合,与方案 1 和 2 相比,方案 5 的评价指标 mAP@0.5 分别提升了 4.0%和 3.8%。证实使用直接相加和拼接操作进行特征融合会造成特征层内的信息互相干扰,而 DFFM 模型能够充分利用双模态信息的互补特性,捕获不同尺度和级别的特征信息,增强目标的检测能力。

(3)PPA 模块的影响 通过对比方案 4 和 6,加入 PPA 模块后,方案 6 的 mAP@0.5 提升了约 1.4%,表明 PPA 模块在特征融合方面发挥了作用,能够更有效地利用双模态信息,避免低质量模态数据干扰,提升小缺陷检出率。

3.3.2 对比实验

为全面验证 PMSF-YOLOv8 的检测性能,将其与主流单模态、双模态目标检测算法在 NEU-RS-

DDS-AUG 数据集上进行对比,并从整体性能、缺陷类型适应性、尺度敏感性三个维度展开分析。

表 5 不同模型对比实验

Tab.5 Comparison experiments of different models

Method	Input	Precision	Recall	mAP@0.5
Transform+YOLOv5	Rgb	0.937	0.899	0.955
Fast R-CNN	Rgb	0.941	0.909	0.968
YOLOv8n	Rgb	0.946	0.908	0.965
CLANet	Rgb+Depth	0.95	0.913	0.974
PMSF-YOLOv8	Rgb+Depth	0.952	0.925	0.986

从表 5 不同模型对比实验可见,PMSF-YOLOv8 在所有评价指标上均优于对比算法:其精确率达 95.2%、召回率达 92.5%、mAP@0.5 达 98.6%,较单模态最优模型 YOLOv8n(mAP@0.5=96.5%) 提升 2.1%,较双模态方法 CLANet^[17](mAP@0.5=97.4%) 提升 1.2%,验证了 RGB 纹理信息与深度空间信息的互补,显著提升了缺陷表征能力。同时,相较于 Transform+YOLOv5 跨模态模型,PMSF-YOLOv8 对复杂背景下缺陷的判别能力更优,表明 DFFM 模块中深度可分离卷积与注意力机制的协同优势。

表 6 各模型对不同缺陷的检测性能对比

Tab.6 Comparison of the detection performance of various models for different defects

Method	hole	gutter	scar	protrusion
YOLOv8n	0.952	0.95	0.966	0.958
PMSF-YOLOv8	0.97	0.975	0.991	0.988

表 6 各模型对不同缺陷的检测性能对比进一步揭示了 PMSF-YOLOv8 对不同类型缺陷的适配性。对于 scar 类,PMSF-YOLOv8 的 mAP 达 99.1%,较 YOLOv8n 提升 2.5%。这是由于 RGB 分支可变形卷积对不规则边缘的动态捕捉,解决了传统卷积对形变缺陷的“漏检”问题;对于 protrusion 类,mAP 提升 3.0%(从 95.8%至 98.8%),验证了深度分支标准卷积对三维几何特征的有效建模,增强了对立体缺陷的量化识别能力;对于 gutter 类和 hole 类,mAP 分别提升 2.5%和 1.8%,表明 DFFM 模块通过动态权重分配,有效抑制了水渍、反光等干扰对低对比度缺陷的影响。

表 7 不同尺度缺陷性能对比聚焦缺陷尺度差异的影响。在小尺度缺陷(<300 pixel²)上,PMSF-YOLOv8 的 mAP 达 96.2%,较 YOLOv8n 提升 6.5%。这主要源于 PPA 模块的局部-全局多尺度补丁建模,能够聚焦小缺陷的细微特征;在中尺度(300~1 000 pixel²)和大尺度(>1 000 pixel²)缺陷上,mAP 分别达 99.0%和 99.5%,较 YOLOv8n 提升 6.9%和 5.8%,证实了双分支特征融合在全尺度范围内的有效性,尤其在大尺度缺陷上,深度信息的加入减少了 RGB 图像中背景纹理的误判。

表 7 不同尺度缺陷性能对比

Tab.7 Comparison of defect performance at different scales

Method	<300 pixels ²	300~1 000 pixels ²	>1 000 pixels ²
YOLOv8n	0.897	0.921	0.937
PMSF-YOLOv8	0.962	0.990	0.995

图 6~10 可视化结果进一步直观验证了上述结论。图 6 展示了原始 YOLOv8n 模型容易出现误检漏检错检的现象,而模型 PMSF-YOLOv8 检测效果有了明显的改善。图 7 中,YOLOv8n 对 hole 类缺陷因反光导致边缘定位偏移,而 PMSF-YOLOv8 结合深度信息精准框定缺陷边界;图 8 显示,YOLOv8n 对细长 gutter 存在片段漏检,PMSF-YOLOv8 通过深度梯度信息实现完整检出;图 9 中,小尺度 scar 在 YOLOv8n 中易被背景噪声掩盖,PMSF-YOLOv8 通过 PPA 模块增强特征响应,检出率提升显著;图 10 验证了 protrusion 类缺陷的检测优势,PMSF-YOLOv8 利用深度信息消除了 YOLOv8n 对凸起附近复杂背景干扰下的误报。

3.3.3 鲁棒性测试

为验证模型在实际工业环境中的可靠性,除高斯噪声干扰测试外(图 11),图 12 新增对某钢厂现场采集图像的鲁棒性验证,该批图像含氧化层脱落、轧制痕迹等未标注干扰项,更贴近真实生产场景。

在图 11 高斯噪声测试中,当噪声标准差 σ 从 0 增至 0.2 时,PMSF-YOLOv8 的 mAP@0.5 仅下降 9.72%(从 98.6%至 88.88%),而 YOLOv8n 下降 32.2%(从 96.5%至 64.3%)。这表明深度模态的几何特征在 RGB 模态受噪声退化时,仍能提供稳定的缺陷表征,验证了双模态互补的抗干扰能力。

图 12 展示了现场图像的检测结果。在含氧化层脱落和轧制痕迹的复杂场景中,对图像中存在的 hole 类和 protrusion 类,模型既能通过 RGB 分支识别 hole 的纹理特征,又能通过深度分支区分 protrusion 与周围氧化层的高度差异,实现精准分类与定位。证实 PMSF-YOLOv8 不仅在标准数据集上表现优异,在未见过的工业干扰场景中仍能保持低误报、高检出的稳定性能,具备实际应用价值。

4 结论

(1)研究提出的双模态目标检测算法—PMSF-YOLOv8 采用“模态适配”双分支骨干网络与 DFFM 融合模块,实现了 RGB 纹理特征与深度空间特征的互补增强。RGB 分支的可变形卷积动态适配划痕、疤痕等不规则缺陷的几何形变,解决“检不出”问题;深度分支的标准卷积利用平移不变性抑制水

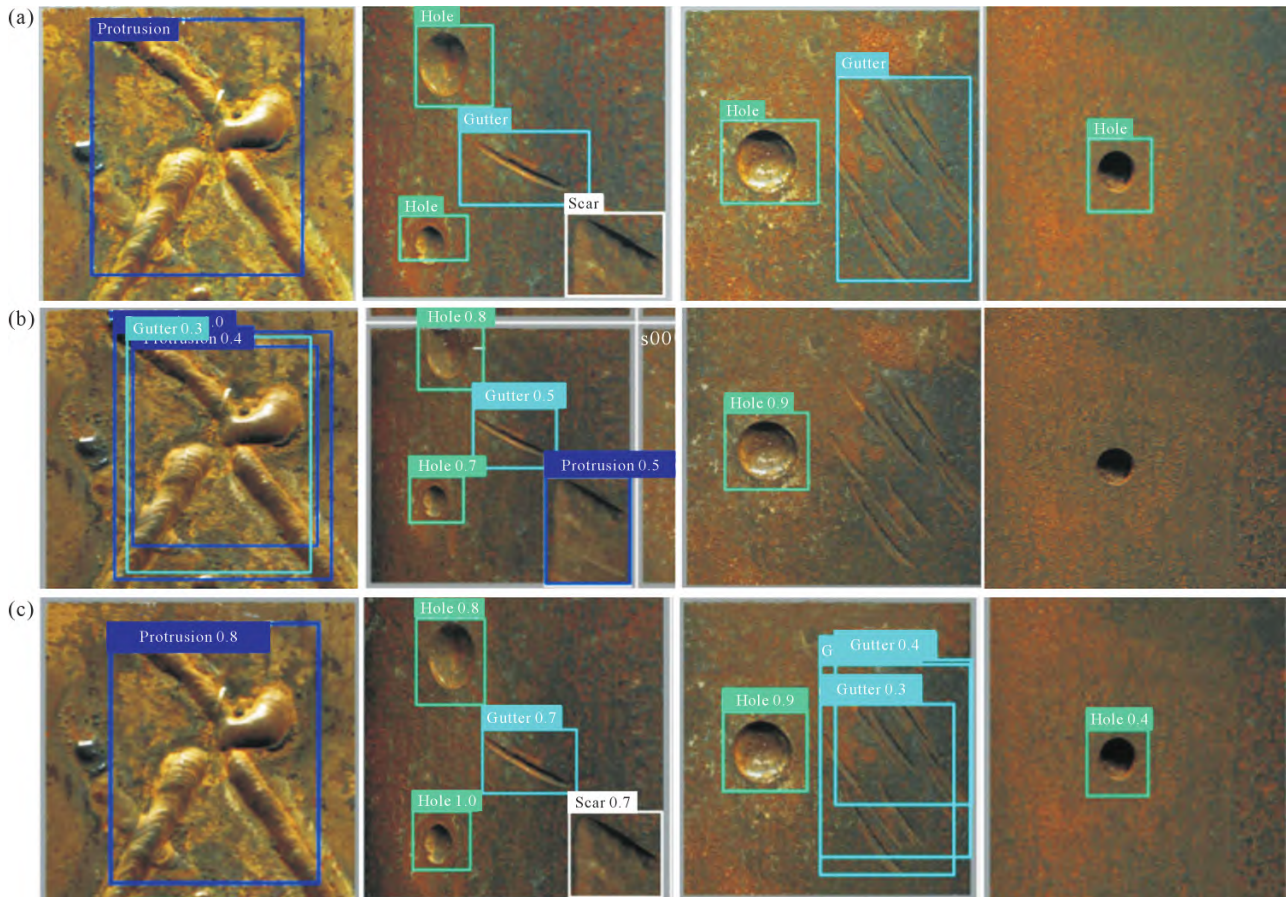


图 6 模型检测结果对比:(a) 真实标签;(b) YOLOv8n 模型检测;(c) PMSF-YOLOv8n 模型检测
Fig.6 Comparison of model detection results: (a) true label; (b) YOLOv8n model detection; (c) PMSF-YOLOv8n model detection

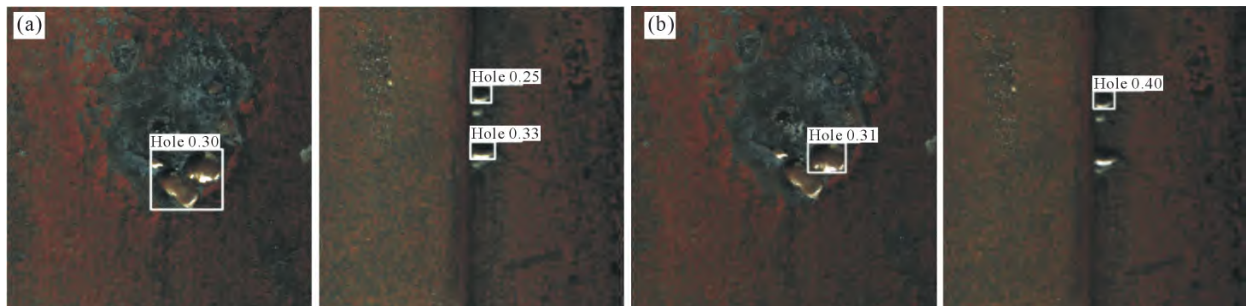


图 7 Hole 类在 YOLOv8n 与 PMSF-YOLOv8 中的检测效果对比:(a) PMSF-YOLOv8n 模型检测;(b) YOLOv8n 模型检测
Fig.7 Comparison of the detection performance of the hole class between YOLOv8n and PMSF-YOLOv8: (a) PMSF-YOLOv8n model detection; (b) YOLOv8n model detection

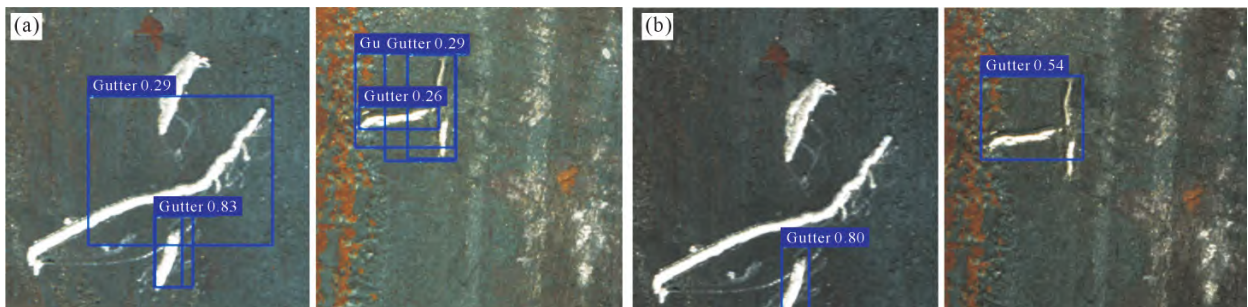


图 8 Gutter 类在 YOLOv8n 与 PMSF-YOLOv8 中的检测效果对比:(a) PMSF-YOLOv8n 模型检测;(b) YOLOv8n 模型检测
Fig.8 Comparison of the detection performance of the gutter class between YOLOv8n and PMSF-YOLOv8: (a) PMSF-YOLOv8n model detection; (b) YOLOv8n model detection

渍、锈蚀等噪声干扰,减少“检错”现象。在 NEU-RS DDS-AUG 数据集上,模型 $mAP@0.5$ 达 98.6%,较单模态 YOLOv8n 提升 2.1%,较双模态方法 CLANet

提升 1.2%,验证了双模态融合在复杂工业场景中的优越性。

(2)虽然 PMSF-YOLOv8 模型通过对比实验与

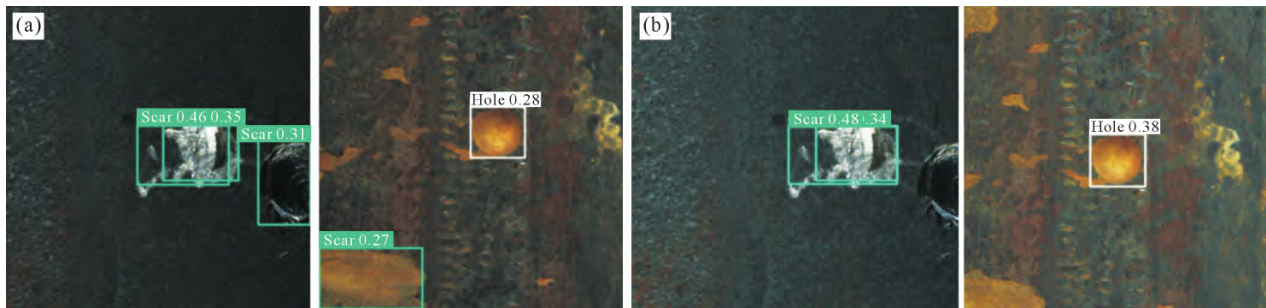


图 9 Scar 类在 YOLOv8n 与 PMSF-YOLOv8 中的检测效果对比:(a) PMSF-YOLOv8n 模型检测;(b) YOLOv8n 模型检测
Fig.9 Comparison of the detection performance of the scar class between YOLOv8n and PMSF-YOLOv8: (a) PMSF-YOLOv8n model detection; (b) YOLOv8n model detection

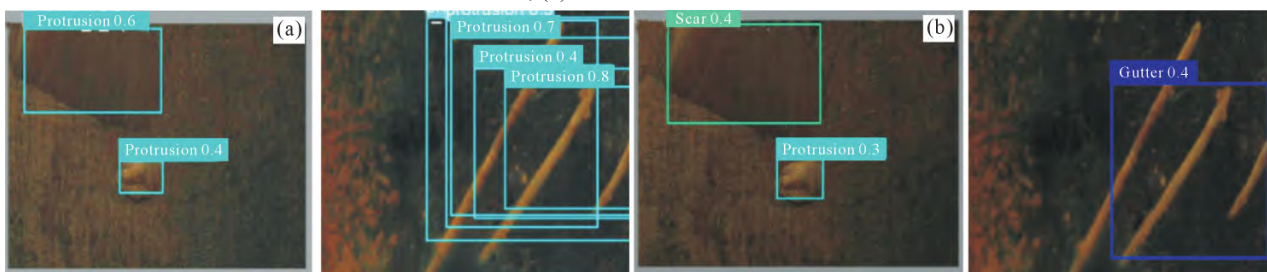


图 10 Protrusion 类在 YOLOv8n 与 PMSF-YOLOv8 中的检测效果对比:(a) PMSF-YOLOv8n 模型检测;(b) YOLOv8n 模型检测
Fig.10 Comparison of the detection performance of the protrusion class between YOLOv8n and PMSF-YOLOv8:
(a) PMSF-YOLOv8n model detection; (b) YOLOv8n model detection

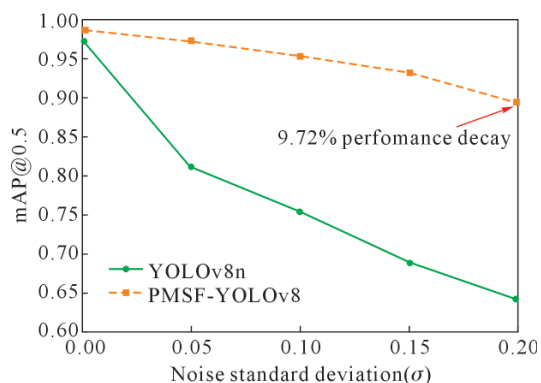


图 11 噪声强度与检测精度关联曲线
Fig.11 Correlation between noise intensity and detection accuracy

细分分析,模型展现出良好的泛化能力与针对性优势,但现有数据集缺陷类型覆盖有限仅 4 类典型缺陷,且深度信息依赖特定采集设备,模型对不同深度传感器的适配性尚未测试。

(3)未来将模型扩展到其他类型的工业缺陷检测任务,利用双模态信息的互补特性,进一步验证 PMSF-YOLOv8 模型在其他工业缺陷检测任务中的适用性和泛化能力。

参考文献:

- [1] HE Y, SONG K C, MENG Q G, YAN Y H. An end-to-end steel surface defect detection approach via fusing multiple hierarchical

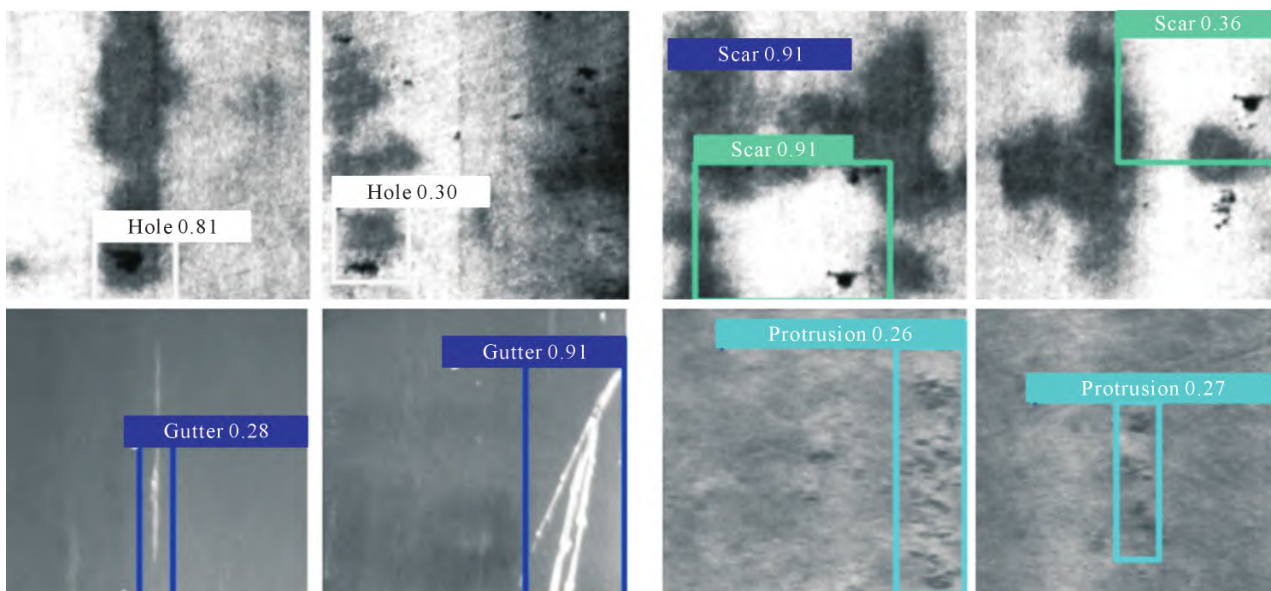


图 12 现场图像的检测结果
Fig.12 Detection results for on-site images

- features[J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.
- [2] 罗东亮,蔡雨萱,杨子豪,章哲彦,周瑜,白翔. 工业缺陷检测深度学习综述[J]. 中国科学: 信息科学, 2022, 52(6): 1002-1039.
- LUO D L, CAI Y X, YANG Z H, ZHANG Z Y, ZHOU Y, BAI X. Survey on industrial defect detection with deep learning[J]. Science in China: Information Sciences, 2022, 52(6): 1002-1039.
- [3] LIU K, LI A M, WEN X, CHEN H Y, YANG P. Steel surface defect detection using GAN and one-class classifier[A]. 2019 25th International Conference on Automation and Computing[C]. New York: IEEE, 2019. 1-6.
- [4] 孙颖,侯志强,杨晨,马素刚,范九伦. 基于双模态融合网络的目标检测算法[J]. 光子学报, 2023, 52(1): 203-215.
- SUN Y, HOU Z Q, YANG C, MA S G, FAN J L. Object detection algorithm based on dual-modal fusion network[J]. Acta Photonica Sinica, 2023, 52(1): 203-215.
- [5] REN S Q, HE K M, GIRSHICK R, SUN J. Faster R-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2015, 39(6): 1137-1149.
- [6] HE K M, GKIOXARI G, DOLLÁR P, GIRSHICK R. Mask R-CNN [A]. 2017 IEEE International Conference on Computer Vision (ICCV)[C]. Venice: IEEE, 2017. 2961-2969.
- [7] CHEN X, ZHANG H. Rail surface defects detection based on Faster R-CNN[A]. 2020 International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)[C]. Tianjin: IEEE, 2020. 819-822.
- [8] WANG H, LI M J, WAN Z B. Rail surface defect detection based on improved Mask R-CNN[J]. Computers and Electrical Engineering, 2022, 102: 108269.
- [9] REDMON J, DIVVALA S, GIRSHICK R, FARHADI A. You only look once: Unified, real-time object detection[A]. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)[C]. New York: IEEE, 2016. 779-788.
- [10] CHOI J, CHUN D, KIM H, LEE H J. Gaussian YOLOv3: An accurate and fast object detector using localization uncertainty for autonomous driving[A]. Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision[C]. New York: IEEE, 2019. 502-511.
- [11] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. Scaled YOLOv4: Scaling cross stage partial network[A]. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE, 2021. 13029-13038.
- [12] LIU W, ANGUELOV D, ERHAN D, SZEGEDY C, REED S, FU C Y, BERG A C. SSD: Single shot multibox detector [A]. Computer Vision- ECCV 2016[C]. Cham: Springer, 2016. 21-37.
- [13] SONG Y N, ZHANG H, LI L, ZHONG H. Rail surface defect detection method based on YOLOv3 deep learning networks [A]. 2018 Chinese Automation Congress(CAC)[C]. Xi'an: IEEE, 2018. 1563-1568.
- [14] HU Q, TANG B, JIANG L, ZHU F X, ZHAO X K. Rail surface defects detection based on Yolov5 integrated with transformer [A]. 2022 IEEE 5th International Conference on Electronics Technology (ICET)[C]. Chengdu: IEEE, 2022. 1131-1135.
- [15] 杜少聪,张红钢,王小敏. 基于改进 YOLOv5 的钢轨表面缺陷检测[J]. 北京交通大学学报, 2023, 47(2): 129-136.
- DU S C, ZHANG H G, WANG X M. Rail surface defect detection based on improved YOLOv5[J]. Journal of Beijing Jiaotong University, 2023, 47(2): 129-136.
- [16] 张涛源,谢新林,谢刚,张琳. 融合 Transformer 的带钢缺陷实时检测算法[J]. 计算机工程与应用, 2023, 59(16): 232-239.
- ZHANG T Y, XIE X L, XIE G, ZHANG L. Real-time strip steel defect detection algorithm fused with transformer[J]. Computer Engineering and Applications, 2023, 59(16): 232-239.
- [17] WANG J P, SONG K C, ZHANG D F, NIU M H. Collaborative learning attention network based on RGB image and depth image for surface defect inspection of no-service rail [J]. IEEE/ASME Transactions on Mechatronics, 2022, 27(6): 4874-4884.
- [18] 赵宏伟,郑嘉俊,赵鑫欣,王胜春,李涓东. 基于双模态深度学习的钢轨表面缺陷检测方法[J]. 计算机工程与应用, 2023, 59(7): 285-293.
- ZHAO H W, ZHENG J J, ZHAO X X, WANG S C, LI Y D. Rail surface defect method based on bimodal-modal deep learning[J]. Computer Engineering and Applications, 2023, 59(7): 285-293.
- [19] DAI J F, QI H Z, XIONG Y W, LI Y, ZHANG G D, HU H. Deformable convolutional networks[A]. Proceedings of 2017 IEEE International Conference on Computer Vision (ICCV)[C]. New York: IEEE, 2017. 764-773.
- [20] LIN T Y, DOLLÁR P, GIRSHICK R, HE K M, HARIHARAN B, BELONGIE S. Feature pyramid networks for object detection[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE, 2017. 2117-2125.
- [21] CHOLLET F. Xception: Deep learning with depthwise separable convolutions[A]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition[C]. New York: IEEE, 2017. 1251-1258.

(责任编辑:宫文婧)