

控制与决策

Control and Decision

DCG-DETR: 基于双分支与上下文引导的钢铁缺陷检测算法

丁方也, 段先华, 张静

引用本文:

丁方也, 段先华, 张静. DCG-DETR: 基于双分支与上下文引导的钢铁缺陷检测算法[J]. 控制与决策, 2026, 41(3): 626-638.

在线阅读 View online: <https://doi.org/10.13195/j.kzyjc.2025.0705>

您可能感兴趣的其他文章

Articles you may be interested in

[基于局部信息增量与MPLS的质量相关故障检测方法](#)

Quality-related fault detection method based on local information increment and MPLS

控制与决策. 2021, 36(7): 1647-1654 <https://doi.org/10.13195/j.kzyjc.2019.1402>

[基于核典型相关性-熵成分分析的工业过程质量监测方法](#)

A quality monitoring method for industrial process based on kernel canonical correlation-entropy component analysis

控制与决策. 2021, 36(12): 2999-3006 <https://doi.org/10.13195/j.kzyjc.2020.0548>

[l_p-范数约束下MKL-OC-ELM的装备故障检测](#)

MKL and OC-ELM fault detection based on l_p-norm constraint

控制与决策. 2021, 36(10): 2379-2388 <https://doi.org/10.13195/j.kzyjc.2020.0443>

[基于2D-OTSU图像边缘检测的回转窑工况识别方法](#)

Condition recognition method of rotary kiln based on 2D-OTSU image edge detection

控制与决策. 2021, 36(10): 2427-2434 <https://doi.org/10.13195/j.kzyjc.2020.0348>

[改进集成深层自编码器在轴承故障诊断中的应用](#)

Application of improved ensemble deep auto-encoder in bearing fault diagnosis

控制与决策. 2021, 36(1): 135-142 <https://doi.org/10.13195/j.kzyjc.2019.0270>

DCG-DETR: 基于双分支与上下文引导的钢铁缺陷检测算法

丁方也, 段先华[†], 张静

(江苏科技大学 计算机学院, 江苏 镇江 212003)

摘要: 针对钢材表面缺陷检测任务中存在的模型参数量大、小目标漏检率高及复杂背景干扰等关键问题, 提出一种基于 RT-DETR (real-time detection transformer) 架构的双分支与上下文引导的协同优化检测方法, 命名为 DCG-DETR. 首先, 设计双分支特征增强模块 DFEM, 通过通道注意力机制与动态感受野卷积的并行融合, 显著提升复杂纹理背景下微小缺陷的特征判别力; 其次, 构建内容-上下文引导聚合特征金字塔模块 CCGAFP, 采用内容感知上采样 CARAFE 与全局-局部双分支特征融合, 解决多尺度特征错位问题, 增强小目标定位精度; 进一步引入轻量化特征融合模块 VoV-GSCSPC, 通过压缩冗余计算与跨阶段梯度传播优化, 在保持精度的同时降低模型复杂度. 在 NEU-DET 数据集上的实验表明, 改进模型 mAP@0.5 指标达 81.5%, 较基准 RT-DETR-L 提升 3.2%, 同时参数量降低 11%、计算量减少 25.9%. 实验结果表明, 改进后的 DCG-DETR 性能整体优于其他同类主流算法, 在进行轻量化的同时提高了检测精度, 为工业质检提供了新方案.

关键词: 钢铁表面缺陷; 缺陷检测; RT-DETR; 双分支特征增强; 特征金字塔; 轻量化

中图分类号: TG142; TP391.41 文献标志码: A

DOI: 10.13195/j.kzyjc.2025.0705

引用格式: 丁方也, 段先华, 张静. DCG-DETR: 基于双分支与上下文引导的钢铁缺陷检测算法 [J]. 控制与决策, 2026, 41(3): 626-638.

DCG-DETR: Steel defect detection via dual-branch and context-guided optimization

DING Fang-ye, DUAN Xian-hua[†], ZHANG Jing

(School of Computer Science, Jiangsu University of Science and Technology, Zhenjiang 212003, China)

Abstract: To address the critical challenges in steel surface defect detection, such as high model complexity, frequent missed detection of small targets, and interference from complex backgrounds, we propose DCG-DETR, a novel detection framework based on the real-time detection transformer (RT-DETR) architecture, which leverages dual-branch feature enhancement and context-guided optimization. First, we design a dual-branch feature enhancement module that integrates an efficient channel attention mechanism with dynamic receptive field convolution in parallel, effectively enhancing the discriminative power for micro-defects against complex textures. Second, we construct a content-context guided aggregation feature pyramid (CCGAFP) module, which adopts content-aware upsampling CARAFE and global-local dual-path fusion to resolve multi-scale feature misalignment and improve the localization accuracy of small defects. Furthermore, we introduce a lightweight feature fusion module VoV-GSCSPC (generalized cross stage partial network with GSConv) that compresses redundant computations and optimizes cross-stage gradient propagation, significantly reducing model complexity while maintaining detection performance. Experiments on the NEU-DET dataset demonstrate that the DCG-DETR achieves 81.5% mAP@0.5, outperforming the baseline RT-DETR-L by 3.2%, while reducing parameters by 11% and computational load by 25.9%. The results validate that the DCG-DETR achieves a superior balance between accuracy and efficiency, offering a practical and deployable solution for industrial quality inspection.

Keywords: steel surface defects; defect detection; RT-DETR; dual-branch feature enhancement; feature pyramid; lightweight design

收稿日期: 2025-07-04; 录用日期: 2025-12-13.

基金项目: 国家自然科学基金项目 (62276118).

责任编辑: 柴利.

[†]通信作者. E-mail: dxh@just.edu.cn.

0 引言

随着工业智能化进程的加速,带钢作为冶金、高速铁路等领域的核心材料,其表面质量直接影响下游产品的安全性和可靠性。在热轧、冷轧等生产过程中,受工艺参数波动、设备磨损和环境因素的影响,带钢表面易出现裂纹、划痕、氧化皮残留、孔洞等缺陷^[1]。据统计,全球每年因带钢表面缺陷导致的直接经济损失超过 50 亿美元,且可能引发重大安全事故(如汽车结构件断裂)。因此,实现带钢表面缺陷的实时、精准检测是保障工业产品质量和安全生产的关键环节^[2]。

当前工业质检领域广泛采用基于卷积神经网络(CNN)的目标检测方法^[3],主要包含双阶段与单阶段两类架构。双阶段检测器如 Faster R-CNN^[4]通过区域建议网络生成候选区域,并执行精细化分类与回归,但其两阶段处理机制导致计算冗余,推理速度受限,难以满足高速产线需求。单阶段算法以 YOLO^[5]系列为代表,采用端到端回归策略^[6]简化检测流程,YOLOv8s^[7]在保持较低参数数量的同时能够实现较高的实时性能,但对微小缺陷(如微裂纹、点蚀)的检测精度显著低于宏观缺陷,主要受限于局部感受野与多尺度特征融合能力不足。基于 Transformer 架构的 DETR(detection transformer)^[8]为缺陷检测带来新突破。该模型摒弃传统锚框设计与 NMS^[9]后处理,通过全局自注意力机制实现端到端检测,在复杂纹理背景下展现出更强的上下文建模能力,特别适合处理带钢表面氧化皮干扰、反光伪影等长程依赖问题。然而,DETR 存在计算复杂度高、需要占用较大现存资源、对微小目标^[10]定位精度较差等显著工程缺陷。对此,学界提出多项改进方案。Deformable-DETR^[11]通过可变形注意力机制降低计算量,并结合多尺度特征金字塔增强微小目标捕获能力。Anchor DETR^[12]创新性地融合空间锚点先验与 Transformer 查询机制,通过位置约束提升对目标的定位精度。尤其是近期,百度飞桨团队提出的 RT-DETR(real-time detection transformer)^[13]通过 3 项关键技术实现成功克服了 DETR 在训练效率、计算复杂度和实时性方面的不足,既保持了高检测精度,又提高了检测速度,并降低了计算资源的需求。

近年来在钢铁表面缺陷检测研究中,崔克彬等^[14]提出了 MCB-FAH-YOLOv8 钢材表面缺陷检测算法,通过加入改进后的卷积注意力机制模块更好地确定密集目标,将 FPN 和 SPPF 模块分别替换为 BiFPN 和 Sim CSPSPF 模块,以便更好地提取上下文信息并提高精度,同时提出四头 ASFF 预测头来改善对小

目标的检测效果,改善了以往钢铁表面缺陷的误检、漏检、检测精度低等问题;张瑞芳等^[15]以 YOLOv5s 为基础,通过在主干网络添加 SE 注意力机制,用 C2f 模块代替 C3 模块,将双向特征金字塔网络 BiFPN 代替颈部网络中的路径聚合网络 PAN,从而改善了因小目标检测效果差而导致的钢铁表面缺陷检测精度差的问题;为了解决钢铁缺陷检测任务中小尺度、形态复杂、结构模糊目标导致现有算法精度低、漏检率高的问题,王素珍等^[16]提出了基于 YOLOv5s 的 SDD-YOLO 算法,通过使用双层路由 Transformer 将局部特征与全局特征结合,提高对结构模糊的缺陷的检测效果,并设计全新的 CSDA 注意力,增强空间和通道的信息交互能力。但上述算法对于模型轻量化的改进不足,模型的参数量和计算量显著增加,需要占用大量内存和计算资源,实际部署起来较为困难。根据以往钢铁表面缺陷检测技术的检测效能较低、准确性低的情况,张世强等^[17]通过在 YOLOv5 模型中加入 CA 注意力机制的 ASPP 扩大模型感受野,并加入改进的 SK 注意力机制,同时将模型的损失函数替换为 SIoU,从而提升了模型检测性能并加快收敛速度;张伟健等^[18]通过在 RT-DETR 主干网络中使用 RepNCSPeLan4 模块更准确地提取数据特征,同时提出 SliderVarfocalLoss 提升模型对难样本的学习能力,提升精度的同时实现了模型的轻量化。但两者依然存在对特定缺陷类型的检测效果不佳的问题,存在漏检、误检的风险。

针对上述模型在进行钢铁表面缺陷检测任务时存在的参数量、计算量较大,对特定类型缺陷检测精度较低,检测速度较慢等问题,本文提出一种基于 RT-DETR 模型改进的钢铁表面缺陷检测模型 DCG-DETR。改进工作如下:

- 1) 设计双分支特征增强模块(DFEM),通过通道注意力与动态感受野协调机制,抑制复杂背景干扰,实现双分支特征协同增强,从而提升钢铁缺陷检测的精度。

- 2) 设计新的内容-上下文引导聚合特征金字塔模块(CCGAFP),通过内容感知上采样与上下文引导融合,实现多尺度特征的自适应聚合,提升小目标特征分辨率,提高检测精度。

- 3) 引入轻量化特征融合模块 VoV-GSCSPC (generalized cross stage partial network with GSCConv),通过压缩冗余计算,结合跨阶段梯度传播优化,增强多尺度表达,在保持精度的同时实现结构精简,从而满足工业质检场景中实时检测与边缘部署的需求。

改进后的模型对钢铁表面缺陷检测的精度较高,

并且模型的体积较小,有效解决了传统方法检测精度不高、运算量大的问题,在工业缺陷检测方面具有广阔前景.

1 RT-DETR 模型

RT-DETR 是一种基于 Transformer 架构的高效实时目标检测模型,其通过融合 CNN 的局部特征提取能力与 Transformer 的全局建模优势,构建了创新的混合编码器 (Hybrid Encoder),显著提升了检测速度与精度. RT-DETR 模型采用可扩展的 HGNetv2 (hierarchical ghost network) 作为主干网络,结合深度可分离卷积与跨阶段连接结构,提取多尺度特征. HGNetv2 通过动态通道剪枝与特征重用机制,在 ResNet 基础上减少 30% 计算量,同时保留丰富的语义信息. 在 Neck 部分设计特征融合模块,包括基于注意力的尺度内特征交互模块 (AIFI) 与跨尺度特征融合模块 (CCFM). 其中, AIFI 模块对高层特征进行全局语义建模,通过自注意力机制增强目标上下文关联; CCFM 模块利用 CNN 的局部感知能力,实现多尺度特征的渐近式融合,解决特征金字塔的语义鸿沟问题. Head 部分创新性地引入 IoU 感知查询选择机制,将交并比引入分类损失函数以动态筛选高置信度候选框,同时采用去噪训练策略对真实框添加随机噪声,强制解码器学习噪声抑制与特征重建,使预测框更逼近真实分布. RT-DETR 的端到端特性

摒弃了传统 NMS 后处理流程,支持动态输入分辨率,针对复杂场景下的目标检测效果好,尤其擅长检测微小目标. 该模型还可扩展应用于自动驾驶、多目标跟踪与遥感图像分析,其混合架构设计为特征融合与边缘设备部署提供了高效解决方案,成为实时检测领域的重要技术突破.

RT-DETR 按照参数可分为 RT-DETR-R18、RT-DETR-R50、RT-DETR-R101、RT-DETR-L、RT-DETR-X. RT-DETR-L 在检测任务中性能相对突出,因此本研究将其作为基线模型. 尽管 RT-DETR-L 模型在目标检测任务中展现出了强大的性能,特别是其端到端设计的优势,但其仍存在若干显著的局限性. 首先,虽然得益于 Transformer 架构和混合编码器设计,RT-DETR-L 的处理速度与检测精度有着显著优势,但对于复杂场景以及极小目标的检测精度和鲁棒性,有时仍不及一些专门优化于此的密集检测器或级联结构模型,小目标漏检和误检问题相对突出. 其次,RT-DETR-L 作为 Large 规模的模型,其参数数量和计算复杂度相对较高,导致模型体积庞大且推理速度相较于更轻量的实时检测器仍有差距,对部署环境的计算资源提出了较高要求,限制了其在资源受限边缘设备上的广泛应用潜力. 因此,RT-DETR-L 模型在小目标检测精度、模型轻量化以及部署成本方面,仍存在改进空间. RT-DETR-L 模型结构如图 1 所示.

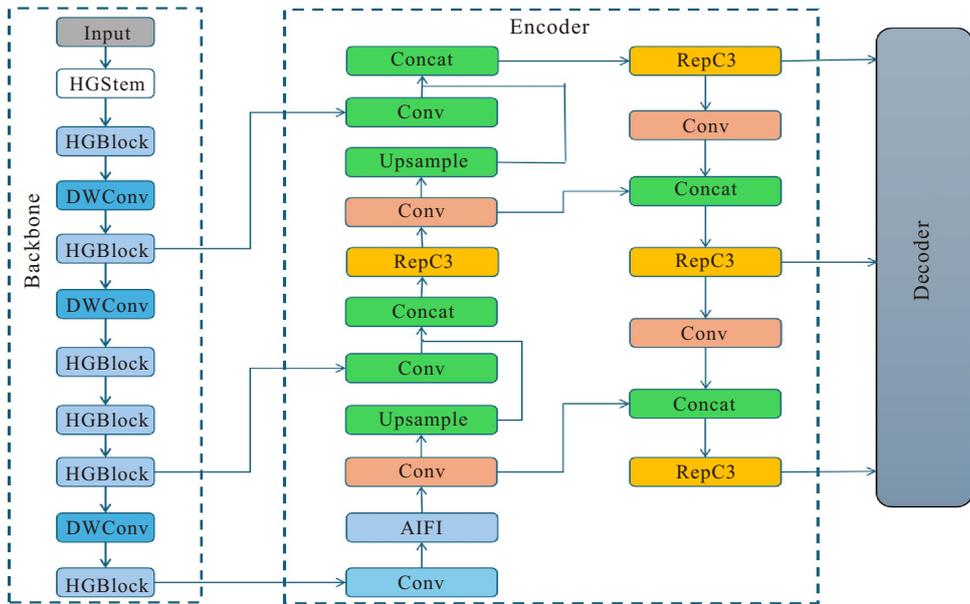


图1 RT-DETR-L 结构

2 DCG-DETR 模型

本文提出的 DCG-DETR 模型从以下方面对 RT-DETR-L 模型进行改进:设计双分支特征增强模块 (DFEM),基于通道注意力与动态感受野协同机制,有效抑制复杂背景噪声干扰,并利用双分支间的特

征交互增强策略提升钢铁缺陷的判别性表征能力;设计新的内容-上下文引导聚合特征金字塔模块 (CCGAFP),通过内容感知上采样与上下文引导融合,实现多尺度特征的自适应聚合,提升小目标特征分辨率,抑制背景干扰,提高检测精度;引入轻量化

特征融合模块 VoV-GSCSPC, 通过压缩冗余计算, 结合跨阶段梯度传播优化, 增强多尺度表达, 在保持精

度的同时实现结构精简. DCG-DETR 模型结构如图 2 所示.

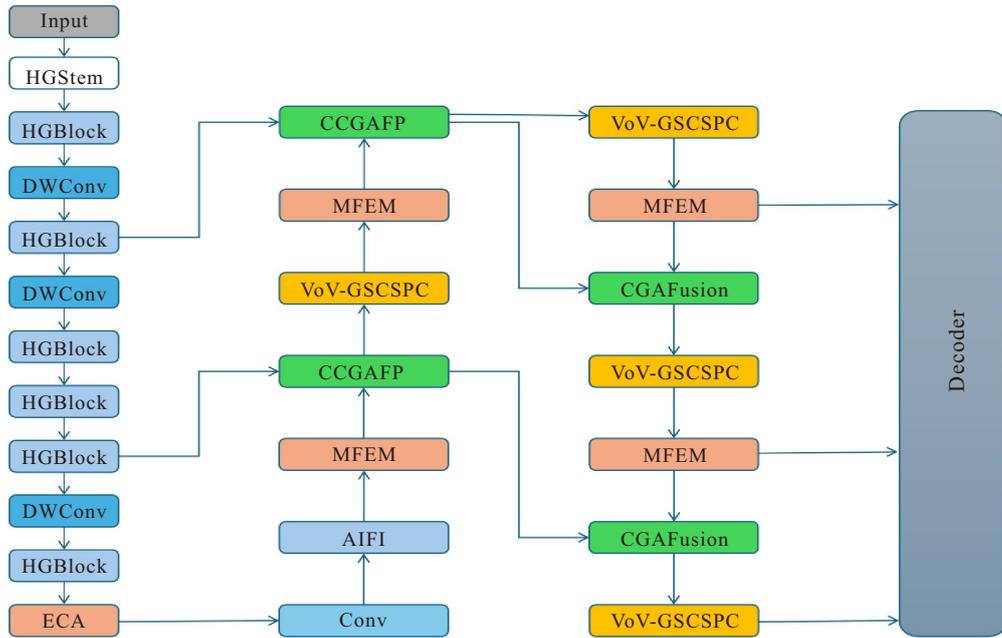


图2 DCG-DETR 结构

2.1 DFEM 模块的设计

在钢铁表面缺陷检测领域中, RT-DETR-L 的混合编码器虽然结合了 CNN 和 Transformer, 但并未充分优化通道间的交互. 因此, 钢铁表面的复杂纹理和噪声可能会导致通道特征冗余或重要信息被掩盖, 从而影响缺陷定位和分类的准确性. 同时, 钢铁表面常受轧制纹理、锈蚀、油污等痕迹干扰, 原始 RT-DETR-L 的特征提取受 CNN 骨干网络的局部感受野限制, 难以捕捉全局上下文信息, 可能导致误检或漏检. 为解决传统通道注意力机制 (如 SENet) 在性能提升与模型复杂度之间矛盾的问题, Wang 等^[19]

提出了动态通道注意力 ECA-Net, 通过低成本的一维卷积操作捕获局部跨通道交互信息, 显著提升了细粒度特征判别力. Zhang 等^[20] 并非单纯依赖减少 FLOPs 来提升卷积神经网络性能, 而是引入一种简单但高效的卷积变体 RFCACConv, 通过结合感受野空间特征和通道注意力机制, 解决传统卷积参数共享问题, 兼顾了高效性与检测性. 此外, 在视频超分辨率重建任务中, 卞鹏程等^[21] 提出的基于注意力融合网络的双分支结构, 通过并行路径实现特征增强, 进一步验证了双分支设计在提升模型表征能力方面的潜力. 基于此, 本文融合通道注意力优化与动态感

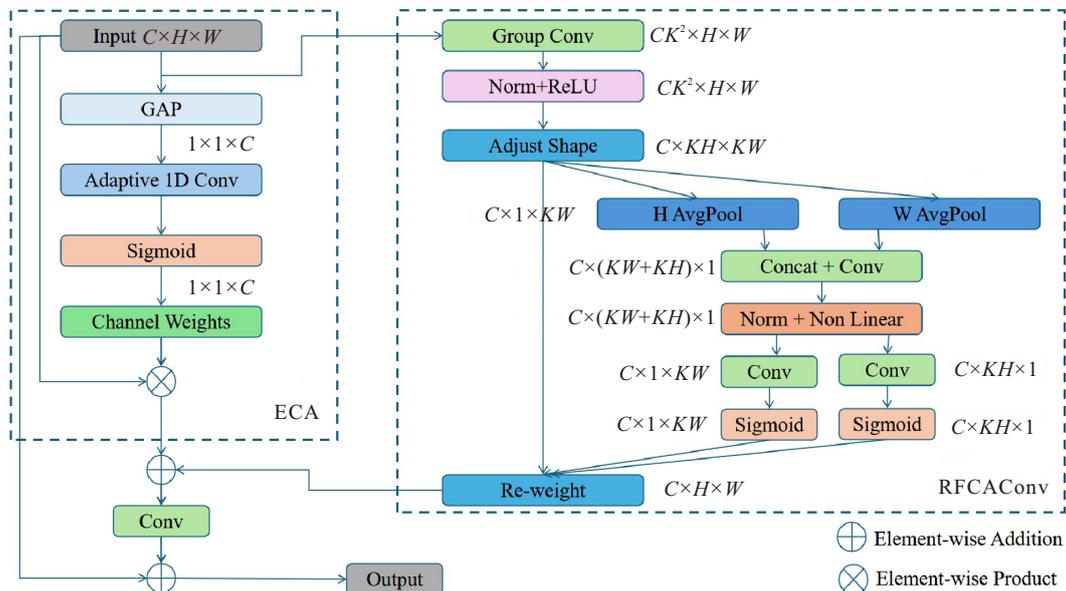


图3 DFEM 模块结构

受野调节思想,创新性地提出双分支特征增强模块(DFEM),其核心架构如图3所示。

DFEM 模块由高效通道注意力(ECA)分支与感受野协调卷积(RFCACConv)分支并行构成。首先,对输入特征图进行全局平均池化(GAP),将每个通道的二维空间特征压缩为一个标量,生成表征各通道全局信息的通道描述符向量。与传统的通道注意力相比,ECA摒弃了使用全连接层进行降维和升维的操作,而是直接在通道描述符向量上应用一个动态一维卷积(adaptive 1D Conv)来生成通道注意力权重。该一维卷积的核大小 k 计算公式如下:

$$k = \left\lceil \frac{\log_2 C}{\gamma} + \frac{b}{\gamma} \right\rceil_{\gamma_{\text{odd}}}, \quad \gamma = 2, b = 1. \quad (1)$$

其中: k 为最终确定的一维卷积核尺寸; C 为通道数; γ 为调节因子,用来控制通道数 C 对卷积核大小的增长速率; b 为偏置项,用来调整卷积核的基础大小,从而能够根据通道数 C 自适应地覆盖有意义的局部跨通道交互邻域。

最后通过 Sigmoid 函数将一维卷积的输出归一化为 0 到 1 之间的值,得到每个通道的注意力权重系数。这些系数随后与原始输入特征图的对应通道相乘,实现通道级的特征重标定,强调重要特征并抑制相对不重要的特征;同时,RFCACConv 分支中, $C \times H \times W$ 的输入特征图首先通过 $K \times K$ 分组卷积提取局部特征,输出特征图尺寸变为 $CK^2 \times H \times W$,经过归一化处理,再通过 ReLU 激活函数,增强模型非线性表达能力;随后通过形状调整将特征图重组为 $C \times KH \times KW$ 的非重叠感受野区块,每个 $K \times K$ 窗口对应一个感受野;之后,对该重组特征分别沿高度与宽度方向进行全局平均池化,得到尺寸为 $C \times 1 \times KW$ 和 $C \times KH \times 1$ 的两个特征图,将二者拼接后通过 1×1 卷积生成通道交互特征,再次进行归一化处理,并通过非线性激活函数,进一步增强特征表达能力。经分割与 Sigmoid 激活输出双向空间权重图,将权重图与重组特征相乘实现感受野级特征重加权,最终通过步长为 K 的 $K \times K$ 卷积将特征还原至原始空间尺寸 $C \times H \times W$ 。双模态输出经 1×1 卷积融合后,通过引入可学习权重参数实现动态特征互补,并与原始输入进行残差连接,这样的设计既保留底层语义完整性,又降低了梯度反向传播的方差,显著提升模型训练稳定性,有效抑制了复杂背景干扰。

为了更直观地展示 DFEM 模块在复杂背景条件下对缺陷检测的有效性,本研究选取了开裂(Cr)与轧制氧化皮(RS)两类缺陷进行单独实验,其可视化

结果如图4所示。在 NEU-DET 数据集中,由于 Cr 与 RS 的形态、纹理以及灰度特性与背景环境存在本质性耦合,导致算法难以区分,极大地提升了对这两类缺陷的检测难度,从而导致检测效果不佳。而对比引入 DFEM 模块前后的可视化结果图,可以观察到基线模型 RT-DETR-L 的特征响应较为弥散,缺陷轮廓与背景轧制纹理存在明显混淆,特征激活区域边界模糊,信噪比较低。相比之下,引入 DFEM 模块后,特征图在缺陷区域呈现出较为集中且尖锐的强激活响应,裂纹形态特征得到完整保留,同时背景纹理干扰得到有效抑制。

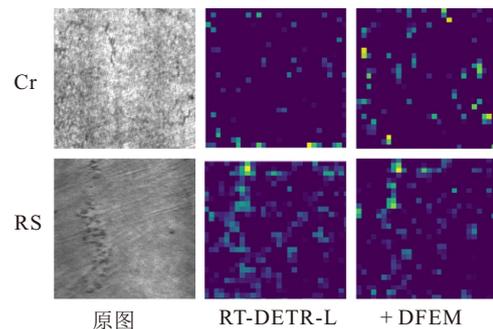


图4 DFEM 模块效果可视化

2.2 CCGAFP 模块的设计

为进一步改善 RT-DETR-L 在钢铁缺陷检测中面临的通道交互不足、全局上下文建模受限所导致的可能发生的误检、漏检等问题,本文提出一种基于内容-上下文双引导的特征金字塔模块(CCGAFP)。通过多尺度内容感知与全局上下文建模的双路径设计,结合动态通道交互机制,能够显著增强模型对微小缺陷的敏感性及复杂背景的鲁棒性,并通过动态特征重组与跨层级语义融合来优化多尺度特征表达。

原始模型中的上采样模块缺乏内容感知能力,上下文建模能力较差,且特征表达僵化,适应性较差。为提升多尺度特征对齐能力,本文用内容感知的特征重组模块 CARAFE^[22]来替代原始模型中的上采样层,其模块结构如图5所示。CARAFE 模块针对钢铁表面复杂背景干扰,可以通过动态核生成策略,实现自适应特征上采样。对于给定的 $C \times H \times W$ 的输入特征图 X 和上采样比例 σ ,CARAFE 通过两个核心步骤生成输出特征图 X' 。在重组核预测阶段,首先通过 1×1 的卷积层将输入通道数从 C 压缩至 C_1 以降低计算量,通过内容编码模块,使用 $k_{\text{encoder}} \times k_{\text{encoder}} \times \sigma^2 k_{\text{up}}^2$ 的卷积核对输入特征内容进行编码,编码后的特征经过展开操作,在空间维展开通道维,得到一个 $\sigma H \times \sigma W \times k_{\text{up}}^2$ 的上采样核,最后对得到的上采样核进行 softmax 归一化。在特征重组模块中,对特征图上的每个点 $l = (i, j)$,取出以之为中心的

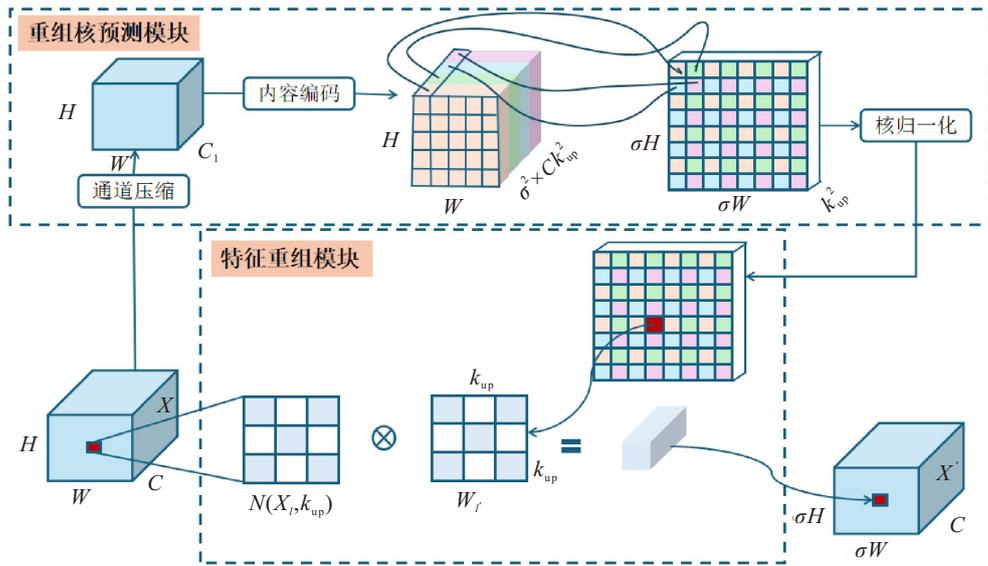


图5 CARAFE 模块结构

$k_{up} \times k_{up}$ 区域, 并与预测的上采样核进行点积, 输出大小为 $\sigma H \times \sigma W \times C$ 的特征图 X' 。

在此基础上, 由于受到 DEA-Net^[23] 启发, 结合 Chen 等^[24] 的上下文引导聚合思想, 进一步优化 Concat 层, 由 CGAFusion 替代, 形成全新的特征金字塔网络 CCGAFP, 从而实现动态特征选择、上下文自适应融合与计算效率优化, 其核心结构如图 6 所示。首先, 将高层特征经过的 1×1 卷积替换成 CARAFE 模块, 高层特征经 CARAFE 算子进行内容感知上采样对齐空间分辨率, 而低层特征依旧通过 1×1 卷积压缩通道维度, 从而提取各自的特征表示; 随后, 这两个初步的特征表示通过一个加法操作进行初步融合; 接着, 双路特征输入通道协调注意力 CGA 模块生成通道专属空间权重 W 及互补权重 $1 - W$, 引导特征通过逐元素乘积进行空间自适应加权融合, 同时保留原始特征的残差路径; 最终, 融合结构经 1×1 卷积整合通道信息, 输出判别性增强的特征 F_{fuse} 。

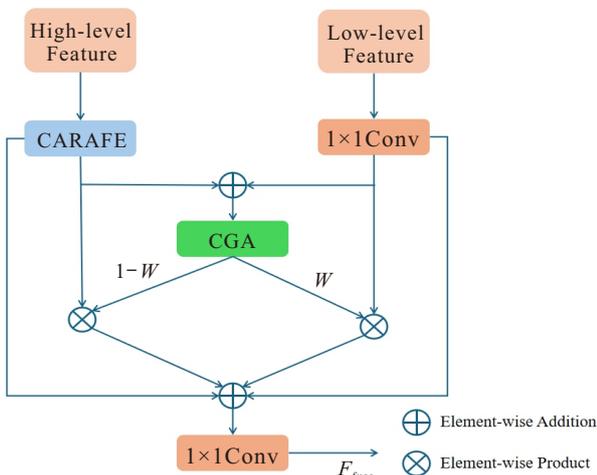


图6 CCGAFP 模块结构

为了更好地评估 CCGAFP 模块在小目标缺陷检测中的性能表现, 选取内含物 (In)、点蚀表面 (PS) 与划痕 (Sc) 3 类典型小目标缺陷进行专项实验, 其可视化结果如图 7 所示。在 NEU-DET 数据集中, 由于 In 直径通常为 5 ~ 15 像素、PS 直径为 3 ~ 8 像素、Sc 宽度仅为 2 ~ 5 像素, 尺寸极其微小, 传统特征金字塔在多次下采样过程中易丢失关键细节, 导致小目标漏检率居高不下。而对比引入 CCGAFP 模块前后的可视化结果图, 可以观察到基线模型 RT-DETR-L 在小目标区域的特征响应微弱且分布稀疏, 导致微小目标与背景难以区分。相比之下, 引入 CCGAFP 模块后, 特征图在微小缺陷区域呈现出显著增强的激活强度与空间集中度, 小目标的完整轮廓和内部纹理细节均得到清晰表征, 同时通过多尺度特征对齐有效提升了小目标的定位精度, 为后续检测头提供了更具判别力的特征输入。

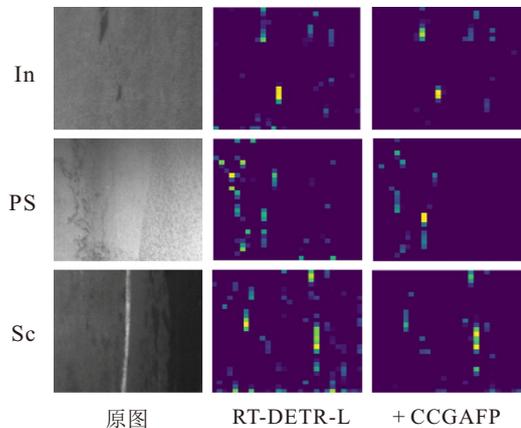


图7 CCGAFP 模块效果可视化

2.3 VoV-GSCSPC 模块的引入

在实时目标检测领域, 轻量化模型设计面临计

算效率与特征表达能力的根本性矛盾. 传统深度可分离卷积 (DSC) 虽可降低计算成本, 但其通道分离特性限制了特征交互能力, 尤其影响高层特征融合精度. 现有改进方案如 MobileNet^[25] 系列通过密集 1×1 卷积补偿通道交互, ShuffleNet^[26] 引入通道混洗机制, 但这些方法或引入额外计算负担, 或未能充分弥合 DSC 与标准卷积 (SC) 间的性能差距. 针对这一挑战, Li 等^[27] 提出 GSConv 技术, 通过混合 SC 与 DSC 的特征输出, 并采用均匀混洗策略实现跨通道信息渗透, 使得轻量化卷积在保留约 50% 计算量的同时, 特征表达能力逼近 SC. 然而, 若将 GSConv 直接应用于检测器全部阶段, 网络深度的增加会加剧数据流动阻力, 显著延长推理延迟. 因此, 如何将 GSConv 的高效特征融合能力与网络结构优化相结合, 成为轻量化检测器设计的关键.

而 VoV-GSCSP 模块, 旨在重构检测器颈部结构, 实现特征复用率与计算效率的协同优化. VoV-GSCSP 模块中, 包含 C_1 个通道的输入特征图首先经过一个标准卷积层, 该卷积层将通道数减半, 输出具

有 $C_1/2$ 个通道的中间特征图, 这个中间特征图进入 GSbottleneck 进行处理; 与此同时, 原始的输入特征图也通过另一个独立的卷积层进行处理, 随后将经过处理的两个特征图经由 Concat 层进行拼接, 输出具有 C_2 个通道的特征图. GSbottleneck 模块如图 8 所示, 该模块由两个 GSConv 层和一个 Conv 层构成. 在此结构中, 输入特征被分别送入两个 GSConv 层和一个 Conv 层进行处理, 并在输出阶段合并三个层的处理结果. 其中, GSConv 模块会将输入特征图经过一个普通卷积层 (Conv) 进行下采样, 然后使用深度可分离卷积层 (DSConv) 对每个通道进行独立卷积. 卷积完成后, 将 Conv 层和 DSConv 层的输出进行经由 Concat 层拼接, 拼接后的特征图经过一个随机排列 (Shuffle) 操作以重新排列特征通道, 从而提高特征间的信息流动, 最后输出特征图. 在模型中引入 VoV-GSCSPC 模块后, 模型在钢铁缺陷检测任务中可以保持高精度来降低模型的参数量与计算量, 从而满足工业质检场景中边缘部署的需求.

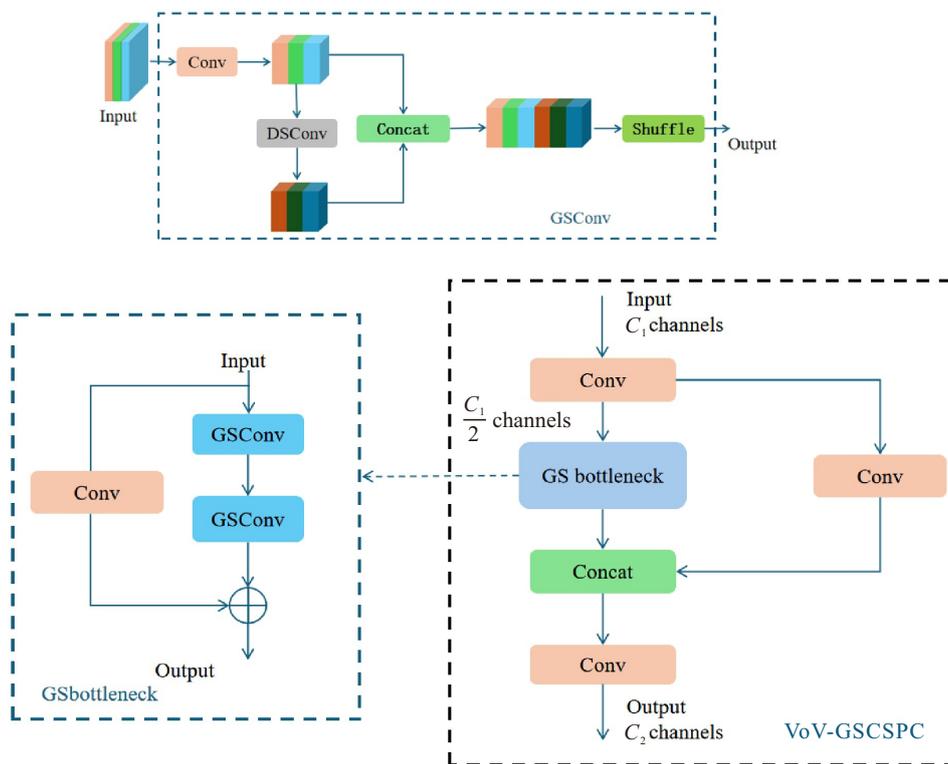


图8 VoV-GSCSPC 模块结构

3 实验与结果分析

3.1 数据集分析

本文采用了公开数据集 (东北大学钢材检测数据集 NEU-DET^[28]) 进行实验. NEU-DET 数据集是专门为钢材表面缺陷检测设计的公开数据集. 该数据集包含 1800 张 200×200 图片, 包含开裂 (Cr)、内含

物 (In)、斑点 (Pa)、点蚀表面 (PS)、轧制氧化皮 (RS) 以及划痕 (Sc) 六个类别. 为了提高图片的质量, 防止训练过拟合, 增强模型的泛化能力, 通过翻转、随机变形缩放以及亮度变化等数据增强操作, 对原始数据集进行扩充, 最终生成 6280 张钢材表面缺陷图像, 并按照 7:1:2 的比例随机划分为训练集、验证

集和测试集.

3.2 实验环境

实验基于 Linux 操作系统下 Python3.10 版本的运行环境, 并在此基础上搭建 Pytorch2.4.0 深度学习框架, 所用的 GPU 为 RTX4090, CUDA 版本为 11.8. 训练轮次设置为 300, 输入图像尺寸为 200×200 像素, 批次大小为 32, 采用 AdamW 优化器并将初始学习率设置为 0.000 2. 本文实验统一在上述环境和参数配置下进行.

3.3 评价指标

为了检验改进算法的有效性, 本文的实验结果采用了目标检测领域常用的评估指标, 包括准确率 P 、召回率 R 、平均精度 mAP、浮点运算量 GFLOPs 以及参数量 Params. 在计算平均精度时, 分别使用了 IoU 阈值为 50 和 $50 \sim 95$ 的情况, 以评估模型在不同 IoU 下的表现. 计算公式如下:

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

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

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

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

式 (2) 和 (3) 中, TP 为真正例, 即模型正确检测到的目标; FP 为假正例, 指模型误检测为目标的负样本; FN 为假反例, 即实际存在目标但未被模型检测到的样本; R 衡量的是在所有实际为正样本的目标中, 被模型正确识别出来的比例. 式 (4) 和 (5) 中, $P(R)$ 是由精确率和召回率构成的曲线, AP 表示某个类别的准确率, N 是类别个数.

3.4 模块效果实验

3.4.1 DFEM 模块效果

为了更好地评估 DFEM 模块在复杂背景条件下的缺陷检测性能, 本研究选取开裂 (Cr) 与轧制氧化皮 (RS) 两类缺陷进行单独实验, 表 1 为实验结果.

表1 DFEM 模块效果 %

类别	Baseline			+ DFEM		
	P	mAP@0.5	mAP@0.5~0.95	P	mAP@0.5	mAP@0.5~0.95
Cr	74.8	65.8	35.7	79.6	70.3	40
RS	73.8	70	37.5	79.4	73	41.3

在 NEU-DET 数据集中, 由于 Cr 与 RS 的形态、纹理以及灰度特性与背景环境存在本质性耦合, 导

致算法难以区分, 极大地提升了对这两类缺陷的检测难度, 从而导致检测效果不佳. 而表 1 实验结果中, 这两类缺陷的各项指标均实现明显提升, Cr 的 mAP@0.5 从 65.8% 跃升至 70.3%, RS 的 mAP@0.5 从 70.0% 提升至 73.0%, 有效解决了复杂背景导致的检测精度低的问题, 验证了 DFEM 模块对于抑制复杂背景干扰的有效性.

3.4.2 CCGAFP 模块效果

为验证 CCGAFP 模块在小目标缺陷检测中的有效性, 本研究针对内含物 (In)、点蚀表面 (PS) 与划痕 (Sc) 三类典型小目标缺陷进行专项实验, 表 2 为实验结果.

表2 CCGAFP 模块效果 %

类别	Baseline			+ CCGAFP		
	P	mAP@0.5	mAP@0.5~0.95	P	mAP@0.5	mAP@0.5~0.95
In	76.2	75.3	39.2	79.4	77.5	44.8
PS	87.8	83.5	60.4	88.8	87.2	66.9
Sc	88.1	90.8	49.1	89.6	91.8	55.9

在 NEU-DET 数据集中, 由于 In、PS、Sc 尺寸微小, 传统特征金字塔在多次下采样过程中易丢失关键细节, 导致小目标漏检率、误检率居高不下. 而实验数据显示, 引入 CCGAFP 模块后 3 类缺陷检测性能全面提升: In 的 mAP@0.5 ~ 0.95 从 39.2% 跃升至 44.8%, PS 的 mAP@0.5 从 83.5% 提升至 87.2%, Sc 的 mAP@0.5 ~ 0.95 从 49.1% 增至 55.9%, 显著改善了小目标的特征可辨识度与空间定位精度, 验证了 CCGAFP 模块在小目标检测任务中的优越性.

3.5 消融实验

3.5.1 参数消融实验

为了进一步探究 DFEM 与 CCGAFP 模块中关键参数对模型检测性能的影响, 本文针对各模块的核心超参数进行了系统的消融实验. 对于 DFEM 模块, 重点评估了控制一维卷积核大小的参数 γ 与 b 的取值, 结果如表 3 所示.

表3 DFEM 模块参数 γ 和 b 的消融实验结果

γ	b	mAP@0.5/%	mAP@0.5 ~ 0.95/%	Params/M	GFLOPs
1	1	81.2	51.5	32.1	103.6
2	1	81.6	52.5	32.1	103.6
3	1	81.3	51.8	32.1	103.6
2	0	81.1	51.6	32.1	103.6
2	2	81.4	52.1	32.1	103.6

由表 3 可知, 当固定 $b = 1$ 时, $\gamma = 1, 2, 3$ 对应的 mAP@0.5 分别为 81.2%、81.6% 和 81.3%, mAP@0.5 ~ 0.95 分别为 51.5%、52.5% 和 51.8%, 表明 $\gamma = 2$ 时模型在保持参数量 32.1 M 与计算量 103.6 GFLOPs

不变的前提下性能最优, 进一步在 $\gamma = 2$ 条件下测试 b 取值, $b = 0, 1, 2$ 对应的 $\text{mAP}@0.5$ 分别为81.1%、81.6%和81.4%, $\text{mAP}@0.5 \sim 0.95$ 分别为51.6%、52.5%和52.1%, 表明 $\gamma = 2, b = 1$ 为DFEM模块的最佳参数组合, 可有效平衡局部跨通道交互的覆盖范围与计算效率.

对于CCGAFP模块, 重点评估了内容感知上采样比例 σ 的影响, 结果由表4所示.

表4 CCGAFP模块参数 σ 的消融实验结果

σ	mAP@0.5/%	mAP@0.5~0.95/%	Params/M	GFLOPs
2	81.4	52.3	32.1	103.8
4	81.1	51.9	31.2	104.1
8	80.8	51.5	32.1	104.5

由表4可知, $\sigma = 2, 4, 8$ 对应的 $\text{mAP}@0.5$ 分别为81.6%, 81.1%和80.8%, $\text{mAP}@0.5 \sim 0.95$ 分别为52.5%, 51.9%和51.5%, 同时计算量随 σ 增大而轻微上升, $\sigma = 2$ 时更有利于保留细微缺陷的纹理特征并抑制冗余计算, 而过大的 σ 易引入背景噪声, 尤其不利于小尺度缺陷的定位与分类. 上述参数消融实验表明, 本文对DFEM与CCGAFP模块的参数选择具有充分的实验依据, 最优参数组合在显著提升模型精度的同时保持较低的计算复杂度, 进一步强化了DCG-DETR在工业质检场景中的实用性与可靠

性.

3.5.2 模块消融实验

为了进一步验证各个模块在钢铁表面缺陷检测任务中的有效性, 本文以RT-DETR-L为基准模型, 在此基础上依次添加改进模块进行消融实验, 实验结果如表5所示. 由实验结果可知, 单独使用DFEM模块时, 实验结果与RT-DETR-L相比, P 和 R 分别提升了5.4%和3.0%, $\text{mAP}@0.5$ 和 $\text{mAP}@0.5 \sim 0.95$ 分别提升了3.3%和4.0%. 单独使用CCGAFP模块时, 与RT-DETR-L相比, P 和 R 分别提升2.2%和1.7%, $\text{mAP}@0.5$ 和 $\text{mAP}@0.5 \sim 0.95$ 分别提升3.1%和3.8%. 单独使用VoV-GSCSPC模块时, 与RT-DETR-L相比, P 和 R 分别提升1.7%和1.0%, $\text{mAP}@0.5$ 和 $\text{mAP}@0.5 \sim 0.95$ 分别提升2.6%和1.5%, 且参数量降低了11%, GFLOPs降低了35.5%. 同时采用3个改进模块之后, 相比于基准模型, P 、 R 、 $\text{mAP}@0.5$ 、 $\text{mAP}@0.5 \sim 0.95$ 分别提升3.4%、3.2%、3.2%、3.3%, 在钢铁表面缺陷检测任务中体现出了较好的性能, 有效解决了因小目标检测效果差而导致的钢铁表面缺陷检测精度差的问题; 同时, 在提升检测精度的基础上, 参数量和GFLOPs分别减少了11%和25.9%, 使得模型能够以更低的成本实现更好的检测效果.

表5 模块消融实验

DFE M	CCGA FP	VoVGSCSPC	P /%	R /%	mAP@0.5/%	mAP@0.5~0.95/%	GFLOPs	Params/M
×	×	×	80.5	74.2	78.3	48.5	103.5	32.0
✓	×	×	85.9	77.2	81.6	52.5	103.6	32.1
×	✓	×	82.7	75.9	81.4	52.3	103.8	32.1
×	×	✓	82.2	75.2	80.9	50.0	67.0	20.0
✓	✓	×	87.1	78.4	82.3	52.6	104.0	32.0
✓	×	✓	87.5	75.0	81.3	50.0	67.8	25.8
×	✓	✓	81.1	77.1	81.2	50.8	69.4	25.8
✓	✓	✓	83.9	77.4	81.5	51.8	76.7	28.5

由表5可知, 同时引入DFEM、CCGAFP与VoV-GSCSPC三个模块后, 综合性能指标相较于DFEM+CCGAFP组合出现小幅下降, 但显著优于基线模型. 该现象主要源于不同模块的设计目标差异. DFEM与CCGAFP模块旨在通过增强特征判别力与多尺度融合以提升检测精度, 但同时会增加模型复杂度, 而VoV-GSCSPC模块的核心目标在于压缩冗余计算、降低参数量与计算成本, 其轻量化过程会不可避免地引入微量的特征信息损失. 当三者协同工作时, 轻量化模块对特征通道的压缩和精简效应与特征增强模块的丰富化效应之间存在一定的权衡, 导致最终性能指标相较于纯精度优化方案DFEM+CCGAFP略有降低. 然而, 此举换来了模型复杂度的

大幅下降. 最终模型在牺牲较小精度的前提下, 实现了更优的综合效能, 显著提升了算法的部署可行性. 此外, 对比DFEM+VoVGSCSPC组合, 引入CCGAFP模块构成三模块方案后, GFLOPs与Params略有增加, 其原因在于CCGAFP模块为增强多尺度特征对齐与上下文感知能力, 其内部的内容感知上采样与双分支融合机制引入了额外的计算开销. 然而, CCGAFP模块以相对较小的计算成本增幅, 换取了对小目标及复杂缺陷特征捕捉能力的大幅提升, 最终的方案实现了模型在轻量化与检测精度之间所采取的有效平衡. 值得注意的是, 三模块组合的准确率 P 相较于DFEM+VoVGSCSPC组合有所下降, 而召回率 R 则显著提升. 这一现象符合目标检测中经

典的“召回率-精确率权衡”规律. CCGAFP 模块通过增强对微小与模糊缺陷的敏感性, 成功地找到了更多真实缺陷, 从而大幅提升了召回率. 然而, 其带来的特征丰富度也使得模型生成了更多候选预测, 其中包含少量将复杂背景误判为缺陷的假正例, 从而导致整体精确率略微降低. VoV-GSCSPC 模块的轻量化设计在高效传递这些特征的同时, 可能略微削弱了对这些负样本的极致甄别能力. 但综合性能指标 mAP, 尤其是 mAP@0.5 ~ 0.95 的提升表明, 模型的平均性能得到了优化. 在实际工业质检场景中, 优先保证高召回率以杜绝漏检至关重要, 轻微的精确率下降可通过后续设置置信度阈值或二次复检等低成本手段进行改善. 因此, 该性能变化趋势符合实际应用需求, 验证了三模块协同设计的有效性.

3.6 模型对比实验

为综合评估改进算法的优越性, 本研究在 NEU-DET 数据集上进行实验, 选取了主流目标检测器、同类改进模型与本文提出的 DCG-DETR 模型进行横向对比实验, 并对结果进行了分析. 表 6 为不同模型对比实验结果. 根据实验结果可知, DCG-DETR 的精度 P 达到 83.9%, 召回率 R 达到 77.4%, 优于多数对比模型. 在 mAP@0.5 指标上, 改进的 DCG-DETR 模型以 81.5% 的准确率与 YOLOv8-L 持平, 但显著高于其他轻量级模型; 在更为严格的 mAP@0.5 ~ 0.95 指标上, DCG-DETR 以 51.8% 的成绩领先于大多数对比算法, 验证了其高 IoU 阈值下的鲁棒性. 同时, DCG-DETR 的参数量达到了 28.5 M, 优于大多数对比模型, 仅次于一些轻量级 YOLO 模型, 但这些模型在精度上低于改进模型. 此外, 改进模型拥有更低的计算量, 表明其在保持高精度的同时, 也实现了较好的计算效率和模型复杂度控制.

表6 模型对比实验

模型	$P/\%$	$R/\%$	mAP@0.5 0.5%	mAP@0.5 ~ 0.95%	GFLOPs	Params/M
YOLOv5-L	73.3	65.9	71.8	41.7	107.7	46.1
YOLOv8-L	81.4	76.2	81.5	49.0	164.8	43.6
YOLOv11-L	71.8	75.8	77.3	48.2	86.6	25.3
YOLOv12-L	75.7	74.4	79.7	49.3	82.1	26.4
Deformable-DETR	79.2	75.3	78.2	48.8	173.2	40.1
RT-DETR-R18	81.3	73.3	78.5	45.5	86.3	28.9
RT-DETR-R50	79.8	74.4	80.2	49.3	125.6	41.9
RT-DETR-R101	83.2	75.2	79.9	49.8	186.3	60.9
RT-DETR-X	84.2	77.8	82.1	51.6	232.4	67.3
RDD-YOLO ^[29]	83.2	78.5	80.7	50.7	16.8	7.1
CSE-YOLOv7 ^[30]	76.9	74.6	78.8	51.2	32.8	22.5
EDH-DETR ^[31]	76.8	73.5	81.4	48.2	187.4	36.3
DCG-DETR	83.9	77.4	81.5	51.8	76.7	28.5

与 RT-DETR 系列模型对比时, DCG-DETR 同样展现出卓越的效率优势. 虽然 RT-DETR-X 的部分精度指标略高, 但其计算量和参数量均远高于 DCG-DETR. 具体而言, DCG-DETR 的参数量仅为 RT-DETR-X 的 42.3%, 计算量降低约 67.0%, 而精度损失控制在 1% 以内. 同样, 与 RT-DETR-R50 和 RT-DETR-R101 相比, DCG-DETR 在保持相当或更优精度的同时, 显著降低了模型复杂度.

综合分析表 6 结果可知, DCG-DETR 在多项关键指标上实现了最佳平衡. 虽然 YOLOv8-L 在 mAP@0.5 指标上与本文模型持平, 但其参数量和计算复杂度分别较 DCG-DETR 高出 52.9% 和 114.8%, 在实际工业部署中将面临显著的资源约束. 相比之下, DCG-DETR 仅以 28.5 M 参数和 76.7 GFLOPs 计算量实现了相当甚至更优的综合性能, 体现出更高的计算效率. 在更严格的 mAP@0.5 ~ 0.95 指标上, DCG-DETR 以 51.8% 的性能领先于大多数对比模型, 表明 DCG-DETR 在高 IoU 阈值和常规 IoU 阈值下均保持稳定的检测性能. 在对比模型中, 专为钢铁缺陷设计的 RDD-YOLO 虽具有极高的计算效率, 但其 mAP@0.5 和 mAP@0.5 ~ 0.95 均低于 DCG-DETR, 表明本文提出的多模块协同优化策略在保持轻量化的同时, 能够实现更优的特征表征能力. 上述对比结果表明, 精度-效率平衡优化的 DCG-DETR 模型较为适合资源受限的工业质检场景.

3.7 缺陷类别对比实验

为了具体分析 DCG-DETR 模型对于各个缺陷类别的检测效果, 本研究对 NEU-DET 数据集中的 6 个类别分别进行实验, 并对不同类别的检测结果进行分析. 表 7 为不同缺陷类别的对比实验结果, 可以看出, 本文提出的 DCG-DETR 模型在各类缺陷检测中均实现了显著性能提升. 在开裂 (Cr) 检测中, 改进后准确率 P 提升 3.9 个百分点至 78.7%, mAP@0.5 提升 4.3 个百分点至 70.1%, 表明模型在细微裂纹与背景纹理交织的复杂场景下, 对缺陷的定位能力显著增强; 内含物 (In) 类缺陷的 mAP@0.5 ~ 0.95 指

表7 缺陷类别对比实验 %

类别	改进前			改进后		
	P	mAP@0.5	mAP@0.5 ~ 0.95	P	mAP@0.5	mAP@0.5 ~ 0.95
Cr	74.8	65.8	35.7	78.7	70.1	40.3
In	76.2	75.3	39.2	81.3	76.3	43.7
Pa	85.5	91.7	63.5	88.5	93.5	65.5
PS	87.8	83.5	60.4	89.8	87.2	65.7
RS	73.8	70	37.5	75.4	70.4	40.6
Sc	88.1	90.8	49.1	89.8	91.6	54.9

标提升最为突出, 达 4.5 个百分点至 43.7%, 反映模型对不规则形状异物的小尺寸特征捕捉能力优化. 对于斑点 (Pa) 和点蚀表面 (PS) 两类高发缺陷, 改进后 P 值分别达 88.5% 和 89.8%, 其中 PS 类的 $mAP@0.5 \sim 0.95$ 提升 5.3 个百分点至 65.7%, 表明模型对点蚀表面边缘纹理的判别力显著增强. 在轧制氧化皮 (RS) 类别中, $mAP@0.5 \sim 0.95$ 提升 3.1 个百分点至 40.6%, 结合其 P 值 1.6% 的提升, 显示模型对线性缺陷的连续性特征提取能力增强; 划痕 (Sc) 类虽原始精度较高, 但改进后 $mAP@0.5 \sim 0.95$ 仍实现 5.8 个百分点的最大幅度提升, 达到 54.9%, 验证了模型对小尺寸点状缺陷的鲁棒性优化. 综合而言, 6 类缺陷的 $mAP@0.5 \sim 0.95$ 平均提升 4.1 个百分点, 表明改进模型在不同类别的缺陷检测中具有良好的泛化能力.

3.8 泛化实验

为了进一步验证 DCG-DETR 模型的有效性, 实验分别选取谢韦尔钢板公司 (Severstal) 数据集^[32]、GC10-DET 钢铁表面缺陷公开数据集^[33]以及 PCB 缺陷数据集 PKU-Market-PCB^[34] 分别进行实验, 从而验证改进模型在不同数据集上的泛化能力. 如表 8 所示, 本文提出的 DCG-DETR 模型在 3 个不同数据集上均实现了检测性能的显著提升. 在 Severstal 数据集上, 改进后 P 提高了 4.2 个百分点, $mAP@0.5$ 和 $mAP@0.5 \sim 0.95$ 分别提升了 3.1 个百分点和 1.5 个百分点. 在 GC10-DET 数据集上, 改进后 P 提高了 2.3 个百分点, $mAP@0.5$ 和 $mAP@0.5 \sim 0.95$ 分别提升了 1.9 个百分点和 2.3 个百分点. 在 PCB 数据集上, 改进后 P 提高了 2.8 个百分点, $mAP@0.5$ 和 $mAP@0.5 \sim 0.95$ 分别提高了 1.5 个百分点和 1.7 个百分点. 这些结果表明, 本文方法在不同类型的数据集上均能有效提升检测性能, 验证了所提出方法的有效性和泛化能力.

表8 泛化实验 %

数据集	改进前			改进后		
	P	$mAP@0.5$	$mAP@0.5 \sim 0.95$	P	$mAP@0.5$	$mAP@0.5 \sim 0.95$
Severstal	73.7	70.9	45.7	77.9	74	47.2
GC10-DET	76.4	73.7	42.6	78.8	75.6	44.9
PCB	84.1	80	50.7	86.9	81.5	52.4

3.9 可视化

3.9.1 特征可视化

为了更直观地表示出 DCG-DETR 模型对特征提取的改进作用, 以及从特征层面深入阐释本文算法的改进机理, 对基准模型 RT-DETR-L 与 DCG-DETR 在典型缺陷样本上的输出特征图进行可视化对比, 结果如图 9 所示.

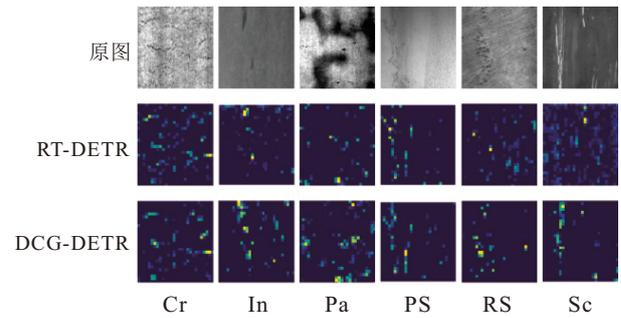


图9 特征可视化

由图 9 可以直观看出, 本文提出的 DCG-DETR 模型所提取的特征具有更强的判别性和语义清晰度. 在开裂 (Cr) 缺陷样本中, RT-DETR-L 的特征图响应较为弥散, 缺陷轮廓与背景纹理存在混淆, 而 DCG-DETR 的特征图则在缺陷区域呈现出更为尖锐和集中的高强度激活, 有效抑制了轧制纹理背景的干扰, 验证了双分支特征增强模块对于复杂背景的强大抑制能力. 对于小尺度内含物 (In) 缺陷, 对比尤为明显. RT-DETR-L 的特征响应微弱且信噪比低, 导致微小目标特征在后续处理中极易丢失, 相比之下, DCG-DETR 的特征图不仅保留了完整的缺陷形态, 其激活强度也显著高于背景区域, 凸显了内容上下文引导聚合特征金字塔在增强小目标特征分辨率和保持细节信息方面的卓越性能. 此外, 在各类缺陷的边缘区域, DCG-DETR 的特征图梯度变化更为剧烈和清晰, 表明其定位能力更加精确. 而 RT-DETR-L 的特征图边缘则相对模糊, 存在不同程度的特征扩散现象. 特征图的直观对比表明, DCG-DETR 通过多模块协同优化, 能够学习到更具判别性的缺陷表征, 其特征图具有高激活强度、高语义清晰度和高定位精度.

3.9.2 结果可视化

为了能够更直观地看出主流算法的检测效果, 选取 NEU-DET 数据集集中的部分图像进行可视化实验, 如图 10 所示.

由图 10 可以看出, 相比于其他检测模型, 改进模型的检测精度更高, 在微小缺陷的检测效果上都有较好的提升. 基准算法 RT-DETR-L 在轧制氧化皮类缺陷检测中检测出 5 个目标, 多于标签数量, YOLOv12-L 在轧制氧化皮类缺陷以及斑点类缺陷中检测出的目标均多于标签数量, 属于误检. 在正常缺陷情况下, 本文模型的精度检测效果明显优于其他主流算法模型, 可视化的结果验证了召回率的提升. 本文提出的算法几乎不存在漏检、误检现象; 同时, 与其他主流目标检测探测器相比, 本文算法在缺陷检测中展现出良好的检测性能.

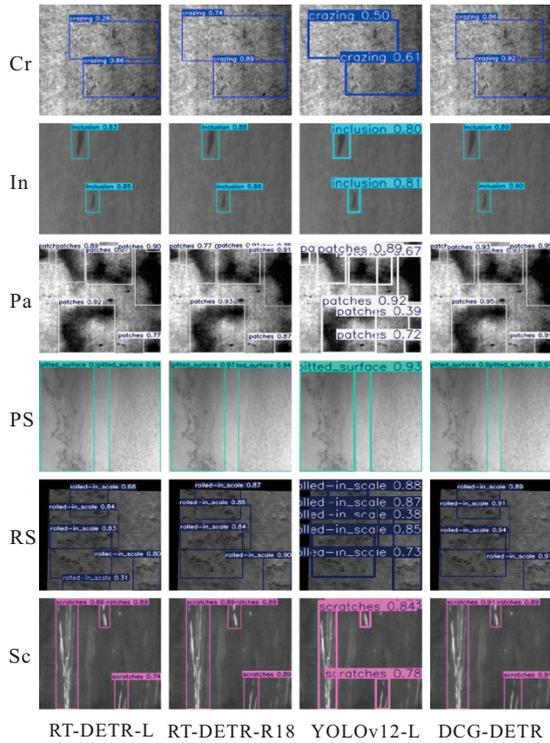


图10 实验效果对比

4 结论

针对钢材表面缺陷检测任务中存在的模型参数量大、小目标误检率高以及复杂背景干扰等关键问题, 本文提出了一种融合双分支与上下文引导协同优化的高效检测框架 DCG-DETR. 首先, 设计多模态特征增强模块 DFEM, 通过通道注意力与动态感受野协调机制, 实现双分支特征协同增强, 有效抑制复杂背景干扰; 其次, 提出内容-上下文引导聚合特征金字塔模块 CCGAFP, 通过内容感知上采样与上下文引导融合, 显著提升微小缺陷分辨率, 缓解小尺度缺陷的漏检与误检问题; 在此基础上, 引入 VoV-GSCSPC 模块, 通过压缩冗余计算并优化跨阶段梯度传播, 实现模型高效精简.

在 NEU-DET 数据集上的实验表明: 相较于原始 RT-DETR 算法, 本文方法在精度、召回率、mAP@0.5 和 mAP@0.5 ~ 0.95 指标上分别提升 3.4%、3.2%、3.2% 和 3.3%, 同时模型参数量减少 11%, 运算量减少 25.9%. 此外, 改进算法还改善了小目标类别中漏检和误检现象, 表现出比其他主流算法更好的性能. 相较于主流的目标检测算法, 本文提出的方法在缩小模型体积的同时提高了检测精度, 为实际工业环境中的钢铁表面缺陷检测带来了新的可能性.

未来, 将继续对该模型算法加以优化和改进, 在保持模型低参数量与较高精度的前提下, 进一步提高模型检测速度, 更好地满足工业场景的实时性需求, 同时, 此项工作也为未来缺陷检测领域提供了一

种新的解决方案.

参考文献 (References)

- [1] Song G R, Song K C, Yan Y H. EDRNet: Encoder-decoder residual network for salient object detection of strip steel surface defects[J]. *IEEE Transactions on Instrumentation and Measurement*, 2020, 69(12): 9709-9719.
- [2] 张德富, 宋克臣, 牛孟辉, 等. 基于一维卷积的生产线冷态重轨表面缺陷快速检测[J]. *东北大学学报: 自然科学版*, 2021, 42(2): 276-281.
(Zhang D F, Song K C, Niu M H, et al. Rapid detection of cold heavy rail surface defects of production line based on one-dimensional convolution network[J]. *Journal of Northeastern University: Natural Science*, 2021, 42(2): 276-281.)
- [3] 卢迪, 赵庆. 空间分组内卷积轻量级目标检测算法[J]. *控制与决策*, 2025, 40(10): 3127-3135.
(Lu D, Zhao Q. Lightweight object detection algorithm based on SGWInvo[J]. *Control and Decision*, 2025, 40(10): 3127-3135.)
- [4] 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.
- [5] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[J/OL]. 2015, arXiv: 1506.02640.
- [6] 丁浩晗, 贺万程, 万俊, 等. DI-YOLO: 一种面向无人机航拍图像的高效小目标检测框架[J]. *控制与决策*, 2025, 40(10): 3106-3116.
(Ding H H, He W C, Wan J, et al. DI-YOLO: An efficient small object detection framework for UAV aerial imagery[J]. *Control and Decision*, 2025, 40(10): 3106-3116.)
- [7] 赵阳, 康飞, 万刚. 基于改进 CycleGAN 与 YOLOv8s 的混凝土坝水下裂缝识别方法[J]. *水电能源科学*, 2025, 43(4): 158-162.
(Zhao Y, Kang F, Wan G. Underwater crack identification for concrete dams based on improved CycleGAN and YOLOv8s[J]. *Water Resources and Power*, 2025, 43(4): 158-162.)
- [8] Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]. *Computer Vision — ECCV 2020*. Cham: Springer, 2020: 213-229.
- [9] Bodla N, Singh B, Chellappa R, et al. Soft-NMS — Improving object detection with one line of code[C]. *2017 IEEE International Conference on Computer Vision*. Venice, 2017: 5562-5570.
- [10] 周威楠, 吴治海, 张正道, 等. 基于弱特征增强的轻量化小目标检测方法[J]. *控制与决策*, 2024, 39(2): 381-390.
(Zhou W N, Wu Z H, Zhang Z D, et al. Lightweight small target detection method based on weak feature enhancement[J]. *Control and Decision*, 2024, 39(2): 381-390.)

- [11] Zhu X Z, Su W J, Lu L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection[J/OL]. 2020, arXiv: 2010.04159.
- [12] Wang Y M, Zhang X Y, Yang T, et al. Anchor DETR: Query design for transformer-based detector[J]. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2022, 36(3): 2567-2575.
- [13] Zhao Y A, Lv W Y, Xu S L, et al. DETRs beat YOLOs on real-time object detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2024: 16965-16974.
- [14] 崔克彬, 焦静颐. 基于 MCB-FAH-YOLOv8 的钢材表面缺陷检测算法[J]. *图学学报*, 2024, 45(1): 112-125.
(Cui K B, Jiao J Y. Steel surface defect detection algorithm based on MCB-FAH-YOLOv8[J]. *Journal of Graphics*, 2024, 45(1): 112-125.)
- [15] 张瑞芳, 伏铭强, 程小辉. 基于 YOLOv5s 的钢铁表面缺陷检测算法[J]. *科学技术与工程*, 2024, 24(23): 9980-9988.
(Zhang R F, Fu M Q, Cheng X H. Steel surface defect detection algorithm based on YOLOv5s[J]. *Science Technology and Engineering*, 2024, 24(23): 9980-9988.)
- [16] 王素珍, 吕基岳, 葛润东, 等. 基于 YOLO 的双层注意力缺陷检测算法[J]. *组合机床与自动化加工技术*, 2024(5): 91-95.
(Wang S Z, Lyu J Y, Ge R D, et al. YOLO-based bi-level attention defect detection algorithm[J]. *Modular Machine Tool & Automatic Manufacturing Technique*, 2024(5): 91-95.)
- [17] 张世强, 史卫亚, 张绍文, 等. 基于改进 YOLOv5 算法的钢铁表面缺陷检测[J]. *科学技术与工程*, 2023, 23(35): 15148-15157.
(Zhang S Q, Shi W Y, Zhang S W, et al. Steel surface defect detection based on improved YOLOv5 algorithm[J]. *Science Technology and Engineering*, 2023, 23(35): 15148-15157.)
- [18] 张伟健, 徐向英, 章永龙. 改进 RT-DETR 的钢材表面缺陷检测算法[J]. *机械科学与技术*, DOI: 10.13433/j.cnki.1003-8728.20250009.
(Zhang W J, Xu X Y, Zhang Y L. Improved RT-DETR algorithm for detecting surface defects in steel[J]. *Mechanical Science and Technology for Aerospace Engineering*, DOI:10.13433/j.cnki.1003-8728.20250009.)
- [19] Wang Q L, Wu B G, Zhu P F, et al. ECA-net: Efficient channel attention for deep convolutional neural networks[C]. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle, 2020: 11531-11539.
- [20] Zhang X, Liu C, Yang D G, et al. RFACConv: Innovating spatial attention and standard convolutional operation[J/OL]. 2023, arXiv: 2304.03198.
- [21] 卞鹏程, 郑忠龙, 李明禄, 等. 基于注意力融合网络的视频超分辨率重建[J]. *计算机应用*, 2021, 41(4): 1012-1019.
(Bian P C, Zheng Z L, Li M L, et al. Attention fusion network based video super-resolution reconstruction[J]. *Journal of Computer Applications*, 2021, 41(4): 1012-1019.)
- [22] Wang J Q, Chen K, Xu R, et al. CARAFE: Content-aware reassembly of features[J/OL]. 2019, arXiv: 1905.02188.
- [23] Chen Z X, He Z W, Lu Z M. DEA-net: Single image dehazing based on detail-enhanced convolution and content-guided attention[J]. *IEEE Transactions on Image Processing*, 2024, 33: 1002-1015.
- [24] Chen L Y, Liu H W, Mo J, et al. Cross channel aggregation similarity network for salient object detection[J]. *International Journal of Machine Learning and Cybernetics*, 2022, 13(8): 2153-2169.
- [25] Howard A, Sandler M, Chu G, et al. Searching for MobileNetV3[J/OL]. 2019, arXiv: 1905.02244.
- [26] Ma N N, Zhang X Y, Zheng H T, et al. ShuffleNet V2: Practical guidelines for efficient CNN architecture design[J/OL]. 2018, arXiv: 1807.11164.
- [27] Li H L, Li J, Wei H B, et al. Slim-neck by GSConv: A lightweight-design for real-time detector architectures[J]. *Journal of Real-Time Image Processing*, 2024, 21(3): 62.
- [28] He Y, Song K C, Meng Q G, 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.
- [29] Zhao C, Shu X, Yan X, et al. RDD-YOLO: A modified YOLO for detection of steel surface defects[J]. *Measurement*, 2023, 214: 112776.
- [30] Shi T, Wu R X, Zhu W X, et al. Steel surface defect detection based on lightweight YOLOv7[J]. *Optoelectronics Letters*, 2025, 21(5): 306-313.
- [31] 韩慧健, 邢怀宇, 张云峰, 等. 基于 Transformer 多元注意力的钢材表面缺陷视觉检测[J]. *郑州大学学报: 工学版*, 2025, 46(5): 69-76.
(Han H J, Xing H Y, Zhang Y F, et al. Visual detection of steel surface defects based on transformer and multi-attention[J]. *Journal of Zhengzhou University: Engineering Science*, 2025, 46(5): 69-76.)
- [32] Demir K, Ay M, Cavas M, et al. Automated steel surface defect detection and classification using a new deep learning-based approach[J]. *Neural Computing and Applications*, 2023, 35(11): 8389-8406.
- [33] Yeung C C, Lam K M. Efficient fused-attention model for steel surface defect detection[J]. *IEEE Transactions on Instrumentation and Measurement*, 2022, 71: 2510011.
- [34] Huang W B, Wei P. A PCB dataset for defects detection and classification[J/OL]. 2019, arXiv: 1901.08204.

作者简介

丁方也 (2000-), 女, 硕士生, 主要研究方向为计算机视觉、目标检测, E-mail: 15370695906@163.com;

段先华 (1965-), 男, 教授, 博士, 主要研究方向为模式识别、图像处理及计算机视觉, E-mail: dxh@just.edu.cn;

张静 (1979-), 女, 副教授, 硕士, 主要研究方向为数字图像处理, E-mail: 20992791@qq.com.