计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
[doi: 10.15888/j.cnki.csa.010082] [CSTR: 32024.14.csa.010082]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

# 改进 RT-DETR 算法的金属杆件表面缺陷检测①

王震洲, 李成哲, 宿景芳, 王建超

(河北科技大学 信息科学与工程学院, 石家庄 050018)
通信作者: 李成哲, E-mail: 359281979@qq.com

**摘　要**: 针对金属杆件表面缺陷检测中存在的微小缺陷漏检、背景干扰及实时性不足等问题, 提出改进 RT-DETR 的高效检测算法 RDGS-DETR. 设计轻量化特征提取模块 RPFN (reparameterized-partial feature network), 融合结构重参数化与稀疏通道计算, 在减少参数量的同时提升微小裂纹特征表达; 构建动态特征精炼融合模块 (dynamic feature refinement fusion module, DFRFM), 集成动态上采样算子 *DySample*, 通过自适应偏移预测提高曲面成像场景下的多尺度特征对齐精度; 引入几何感知归一化损失 (geometric-sensitive normalized loss, *GSNL*) 函数, 解决传统 *IoU* 对非重叠小目标敏感度不足及复杂缺陷回归偏差问题; 设计稀疏全局交互注意力模块 (sparse global interaction attention, SGIA), 采用高效加法注意力机制, 以线性复杂度实现缺陷区域的全局上下文建模. 实验结果表明, 相较原始模型, RDGS-DETR 推理速度提升 8.55 FPS, *mAP*@0.5 提升 2.8%, 并验证了鲁棒性. 该算法兼顾精度与实时性, 为智能制造场景下的金属杆件表面质检提供可靠支撑.

**关键词**: RT-DETR; 缺陷检测; 轻量化; 特征融合; 微小缺陷特征表达

引用格式: 王震洲,李成哲,宿景芳,王建超.改进 RT-DETR 算法的金属杆件表面缺陷检测.计算机系统应用. http://www.c-s-a.org.cn/1003-3254/10082.html

## Surface Defect Detection in Metal Rods Using Enhanced RT-DETR Algorithm

WANG Zhen-Zhou, LI Cheng-Zhe, SU Jing-Fang, WANG Jian-Chao

(College of Information Science and Engineering, Hebei University of Science and Technology, Shijiazhuang 050018, China)

**Abstract**: To address the issues of missed detection of tiny defects, background interference, and insufficient real-time performance in surface defect detection of metal rods, an improved RT-DETR-based efficient detection algorithm, RDGS-DETR, is proposed. A lightweight feature extraction module, called the reparameterized-partial feature network block (RPFN block), is designed. It integrates structural re-parameterization with sparse channel computation to reduce parameter complexity while enhancing the representation of micro-crack features. A dynamic feature refinement fusion module (DFRFM) is also developed, incorporating the *DySample* dynamic upsampling operator, which improves multi-scale feature alignment accuracy in curved surface imaging scenarios by adaptively predicting offsets. Furthermore, a geometric-sensitive normalized loss (*GSNL*) is introduced to address the limited sensitivity of traditional IoU metrics to non-overlapping small targets and to reduce regression bias for complex-shaped defects. In addition, a sparse global interaction attention (SGIA) module is developed, which employs an efficient additive attention mechanism to achieve global context modeling of defect regions with linear complexity. Experimental results demonstrate that, compared with the original model, RDGS-DETR improves inference speed by 8.55 FPS and increases *mAP*@0.5 by 2.8%, while also verifying its robustness. The algorithm achieves a balance between detection accuracy and real-time performance,

① 基金项目: 国家重点研发计划 (2024YFD2402205); 河北省高等学校科学技术研究项目 (QN2025371)
收稿时间: 2025-07-29; 修改时间: 2025-08-19, 2025-09-10; 采用时间: 2025-09-26; csa 在线出版时间: 2025-12-26

providing reliable technical support for surface quality inspection of metal rods in intelligent manufacturing scenarios.
**Key words**: RT-DETR; defect detection; lightweight; feature fusion; small defect feature representation

金属杆件作为工业基础设施与高端装备制造的核心构件, 其表面质量直接关系到产品的可靠性与安全性. 传统人工目视检测和早期图像处理方法因效率低、误检率高、环境适应性差, 已难以满足智能制造对精度、鲁棒性和实时性的要求. 近年来, 深度学习为表面缺陷检测提供了新思路, 但受材质光学特性、缺陷形态多样及曲面成像畸变等因素影响, 金属杆件的缺陷检测在算法与应用层面仍存在技术瓶颈.

早期研究多采用灰度阈值分割[1]、形态学滤波[2]等方法实现缺陷粗定位, 但对光照和噪声极为敏感, 难以适应复杂工业环境. 卷积神经网络 (CNN) 的兴起极大推动了检测技术的变革, YOLO[3,4]、Faster R-CNN[5]等端到端目标检测模型在精度和效率之间实现较好平衡, 广泛应用于金属表面的缺陷检测场景. YOLO 系列凭借其多尺度预测结构, 在实时性和小目标识别方面表现突出; 后续研究者通过引入注意力机制、轻量化网络等策略, 有效提升了模型检测缺陷的精确度. 例如, 2023 年, 周彦等[6]提出改进 YOLOv5s 算法, 通过融合通道注意力机制与优化损失函数, 显著增强钢板表面裂纹与划痕的识别能力; 2024 年, 马淑华等[7]基于轻量化架构 PP-PicoDet-XS 嵌入无参注意力模块 SimAM, 结合量化蒸馏策略压缩模型, 在铝型材表面微小缺陷检测中兼顾精度与实时性.

近年来, Transformer 架构的引入为目标检测开辟了新路径. 2020 年, Carion 等[8]提出的 DETR 模型首次将 Transformer 应用于目标检测任务, 利用二分图匹配机制消除了对非极大值抑制 (NMS) 的依赖, 实现真正意义上的端到端检测, 但其计算复杂度高、收敛慢等问题限制了在工业场景的推广. 为提升其实用性, Zhu 等[9]提出 Deformable-DETR, 通过引入稀疏注意力机制加速训练与推理. 2024 年, Zhao 等[10]进一步提出 RT-DETR, 采用混合编码器与查询选择机制, 在保持 Transformer 优势的同时实现了工业级实时检测性能.

尽管金属表面缺陷检测取得显著进展, 但针对金属杆件这类复杂工件, 现有方法仍面临挑战. 图像模糊、背景干扰和特征弱表达易导致小缺陷漏检; 多尺度融合难以兼顾精度与效率, 复杂模型在工业部署中还存在高能耗、低兼容性的问题. 为此, 本文结合 RT-DETR 的高效结构优势, 提出改进模型, 通过特征提取优化、多尺度感知增强与轻量化设计, 在提升精度与鲁棒性的同时降低计算开销, 实现性能与效率的平衡, 为构建可工程化的金属杆件智能检测系统提供有力支撑.

## 1 RT-DETR 改进方法

### 1.1 改进路线

本研究围绕 RT-DETR 模型在金属杆件表面缺陷检测中的性能瓶颈, 提出“特征提取-融合-优化-推理”全链路改进方案, 从主干网络轻量化、多尺度特征对齐、损失函数适配性以及注意力效率优化 4 个维度进行系统性改进, 旨在实现高精度与实时性的提升, 改进后的 RDGS-DETR 网络模型结构如图 1 所示.

(1) 主干网络动态化重构

构建 RPFN block 轻量化特征提取模块, 通过动态结构重参数化与稀疏通道计算策略重构 Basic block. 在保留局部细节的同时显著降低计算冗余, 解决传统卷积对反光噪声敏感及嵌入式设备部署效率低的问题. 该改进为后续模块提供高效且鲁棒的特征表达基础.

(2) 跨尺度特征融合优化

在混合编码器中构建动态特征精炼融合模块 DFRFM, 集成动态上采样算子 *DySample*, 替代传统双线性插值. *DySample* 通过点采样策略与规则内存访问模式, 以更低的计算成本实现特征图分辨率恢复, 增强微小缺陷的边缘特征对齐能力, 提升跨层级信息交互效率, 对初始特征进行“强化”.

(3) 缺陷感知损失函数设计

设计几何感知归一化损失函数 *GSNL*, 融合 *Shape-IoU* 的边界框形变自适应机制与归一化 Wasserstein 距离 (normalized Wasserstein distance, *NWD*) 的分布相似性度量. 其中, *Shape-IoU* 通过引入形状权重系数动态调整边界框回归梯度, 增强模型对裂纹、划痕等不规则缺陷的定位精度; *NWD* 则将目标匹配问题转化为高斯分布建模, 有效缓解微小缺陷因像素占比低导致的 *IoU* 计算偏差. 二者协同优化显著提升模型对工业场景

尺度与形态多样性的适应能力.

(4) 全局注意力效率升级

设计 SGIA 稀疏全局交互注意力模块, 利用高效加法注意力 (efficient additive attention), 以线性元素乘法替代传统自注意力的矩阵乘操作. 该机制通过动态权重分配保留全局上下文建模能力, 同时降低计算复杂度, 显著降低高分辨率特征图处理延迟, 满足实时推理需求.

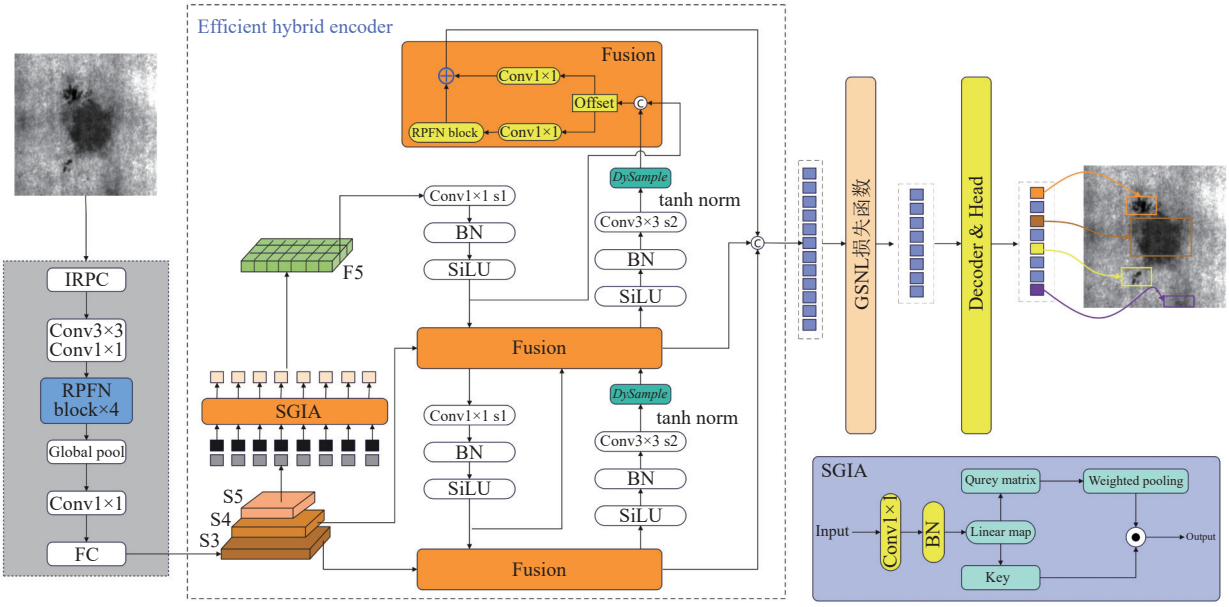



图 1 RDGS-DETR 网络结构图

### 1.2 RPFN 轻量化特征提取模块

传统 ResNet-18 的 Basic block 采用固定 3×3 卷积核堆叠的残差结构, 虽然在通用目标检测任务中表现稳定, 但其固有的结构特性在金属杆件表面缺陷检测场景中存在局限: 刚性卷积模式难以适配金属表面高反光、多噪声干扰下的复杂纹理特征, 导致局部细节表征能力不足. 冗余的密集计算导致嵌入式设备推理效率低下, 难以满足工业实时检测需求.

为此, 本研究融合结构重参数化 RepConv[11]与稀疏卷积 PConv[12]核心思想, 构建轻量化复合卷积单元 IRPC. PConv 通过稀疏通道采样策略大幅降低 *FLOPs*, 同时保留关键空间细节; 而 RepConv 通过多分支训练与参数融合技术, 在提升特征多样性的同时确保推理效率. 二者的协同优化赋予 Basic block 动态特征适应能力, 训练阶段利用多分支拓扑增强模型对金属表面噪点与微小缺陷的鲁棒性, 推理阶段通过参数合并与稀疏计算实现硬件友好的高效部署, 从而在特征表达能力与计算效率之间达成平衡. PConv 卷积单元与改进后的 IRPC 卷积模块结构如图 2 所示, 图中“*”表示卷积.

设输入特征为$X \in R^{c_p \times h \times w}$, $h$、$w$ 分别为特征图的维度, $c_p$表示重参数化分支通道数. 改进后的 IRPC 复合卷积模块通过以下流程完成特征映射.

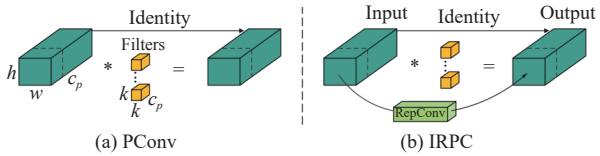



图 2 PConv 与 IRPC 结构图

(1) 动态稀疏训练策略

为强化模型对目标缺陷的敏感性, 首先利用动态稀疏训练策略, 在 PConv 的通道选择阶段集成可学习的掩码向量$\alpha \in [0,1]^{c_p}$, 量化各通道对缺陷特征的判别贡献度, 在训练阶段, $\alpha$ 作为软掩码与输入特征逐通道相乘以渐进抑制低贡献通道; 在推理阶段, 引入可学习的动态阈值参数$\tau$进行二值化决策, 生成硬掩码 $M$, 仅保留重要性高于$\tau$的通道执行 PConv 计算, 其余通道直接跳过以降低冗余.

$$M^{(0)} = \begin{cases} 1, & \alpha_i > \tau \\ 0, & \text{else} \end{cases} \tag{1}$$

为确保$\tau$在训练中稳定收敛, 采用分层初始化与动

态梯度更新机制: $\tau$初始化时根据网络结构中不同模块对特征敏感度的差异进行分配, 训练过程中则通过梯度下降方式动态优化, 并限制其更新步长$\eta_\tau$为主干网络学习率的 1/10, 避免因梯度波动导致阈值过度偏离初始区间[13]. 其更新公式如式 (2):

$$\tau^{(t+1)} = \tau^{(t)} - \eta_\tau \cdot \frac{\partial_L}{\partial_\tau} \tag{2}$$

其中, $\frac{\partial_L}{\partial_\tau}$为损失函数对动态阈值参数的偏导数. 从而实现通道子集的端到端动态筛选, 使模型能自适应聚焦于金属杆件缺陷检测任务中更具判别性的特征.

PConv 的稀疏计算定义为:

$$X_{\text{PConv}} = Concat(Conv_{3\times3}(X_{\text{selected}}), X_{\text{bypass}}) \tag{3}$$

其中, $X_{\text{selected}} = X \cdot M$, $\cdot$为 Hadamard 乘积.

(2) RepConv 参数融合

经 PConv 动态筛选后的稀疏特征由重参数化卷积 RepConv 处理, 其多分支训练模式增强了局部特征的非线性表达能力, 并通过动态参数融合机制提升了特征适应性. 训练阶段采用多分支结构 (3×3 卷积、1×1 分支、Identity 连接), 如式 (4) 所示:

$$Y = BN(W_{3\times3} * X_{\text{PConv}}) + BN(W_{1\times1} * X_{\text{PConv}}) + Iden(X_{\text{PConv}}) \tag{4}$$

其中, “*”表示卷积.

推理时通过参数重映射融合为单一 3×3 卷积核, 能够显著降低部署时的计算复杂度. 该方法通过空间特征强化与通道计算稀疏化的协同设计, 在同等通道容量的约束下实现计算量的可控压缩. 因此, IRPC 的 *FLOPs* 的计算方式为:

$$FLOPs = h \times w \times k^2 \times c_p^2 \tag{5}$$

其中, $k$ 表示卷积核的大小.

动态稀疏 PConv 通过通道重要性评估动态筛选卷积通道, 在浅层保留更多高频纹理通道以适配微小缺陷, 深层压缩语义通道以减少冗余, 整体降低了计算量, 实现了缺陷适应性与计算效率的平衡.

RepConv 在推理阶段通过参数融合进一步减少内存访问次数, 此外, 其多分支训练模式使模型在金属反光、油污噪声等干扰下仍能保持稳定的特征提取能力, 增强了特征鲁棒性. 融合 IRPC 卷积单元的轻量化特征提取模块 RPFN block 的结构如图 3 所示.

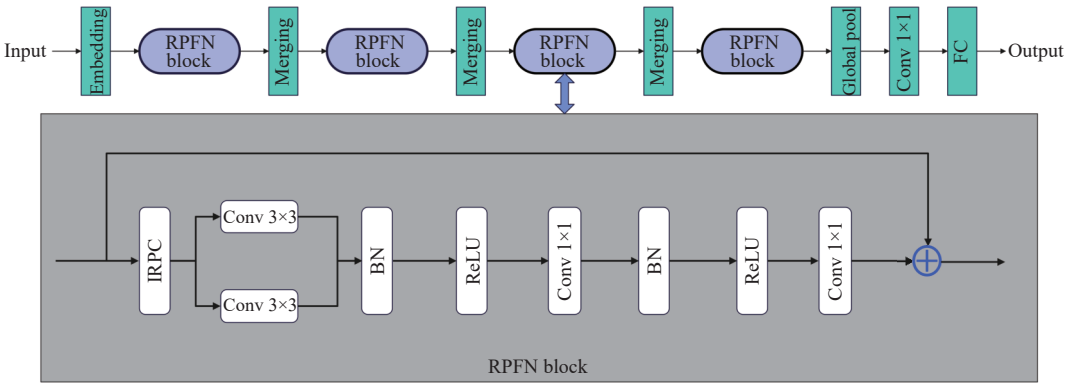



图 3 RPFN block 轻量化特征提取模块

### 1.3 跨尺度特征采样优化

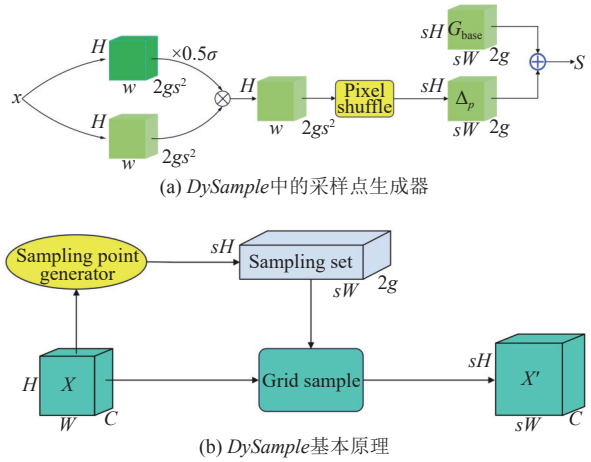
针对 RT-DETR 混合编码器中跨尺度特征融合模块 CCFM 在金属杆件曲面成像场景下存在的特征错位与微小缺陷边缘模糊问题, 本研究提出动态特征精炼融合模块 DFRFM, 利用动态上采样器 *DySample*[14], 取代传统双线性插值算法, 以动态采样策略优化特征图分辨率恢复过程. 传统 CCFM 模块采用固定权重的双线性插值, 该算法虽计算高效但其固定采样网格难以适应金属杆件曲面成像的几何畸变, 导致微小缺陷的边界特征在跨尺度融合中出现偏移. *DySample* 通过动态生成采样偏移量实现自适应插值, 其核心公式定义为:

$$DySample(X') = GridSample(X, S) \tag{6}$$

其中, $S$ 为动态采样网格函数生成的采样集. 具体而言, 对输入特征图$X \in R^{C\times H\times W}$, *DySample* 通过轻量级子网络 (含 1×1 卷积与 Sigmoid 激活) 与像素洗牌预测每个目标位置的偏移向量:

$$\Delta_p \in R^{2\times sH\times sW} \tag{7}$$

并与标准双线性网格$G_{\text{base}}$叠加生成优化后的采样网格$S = G_{\text{base}} + \Delta_p$. 动态偏移量使插值点向缺陷边缘的高梯度区域聚集, 从而增强特征对齐精度. 动态上采样 *DySample* 中的采样点生成器以及 *DySample* 的原理如图 4(a)、(b) 所示.




(a) *DySample*中的采样点生成器

(b) *DySample*基本原理

图 4 *DySample* 采样点生成器及 *DySample* 基本原理

同时, 为了避免偏移量学习过程中因剧烈波动导致的越界风险 (如采样点超出特征图边界), 在 DFRFM 模块中对 *DySample* 的偏移学习施加偏移范围约束机制, 将生成的偏移图数值通过 torch.tanh 函数归一化至 [−1, 1]区间, 再结合特征图尺寸缩放至实际坐标范围 (如对于高为 $H$、宽为 $W$ 的特征图, 偏移量实际作用范围为$\left[-\frac{H}{2}, \frac{H}{2}\right]$和$\left[-\frac{W}{2}, \frac{W}{2}\right]$), 确保采样网格始终落在有效特征区域内.

实现层面, 在 DFRFM 模块中 *DySample* 的部署包含 3 步: 首先, 对低分辨率特征图进行通道压缩 (1×1 卷积) 生成偏移图并进行归一化处理; 其次, 将偏移图与基础网格叠加构建自适应采样网格; 最后, 调用 PyTorch 内置函数 F.grid_sample 完成特征插值. 相较于传统方法, 动态采样策略能够有效缓解曲面畸变导致的特征错位; 规则化的内存访问模式减少数据搬运次数, 推理速度提升, 显著优化硬件适配性. DFRFM 模块结构如图 5 所示.

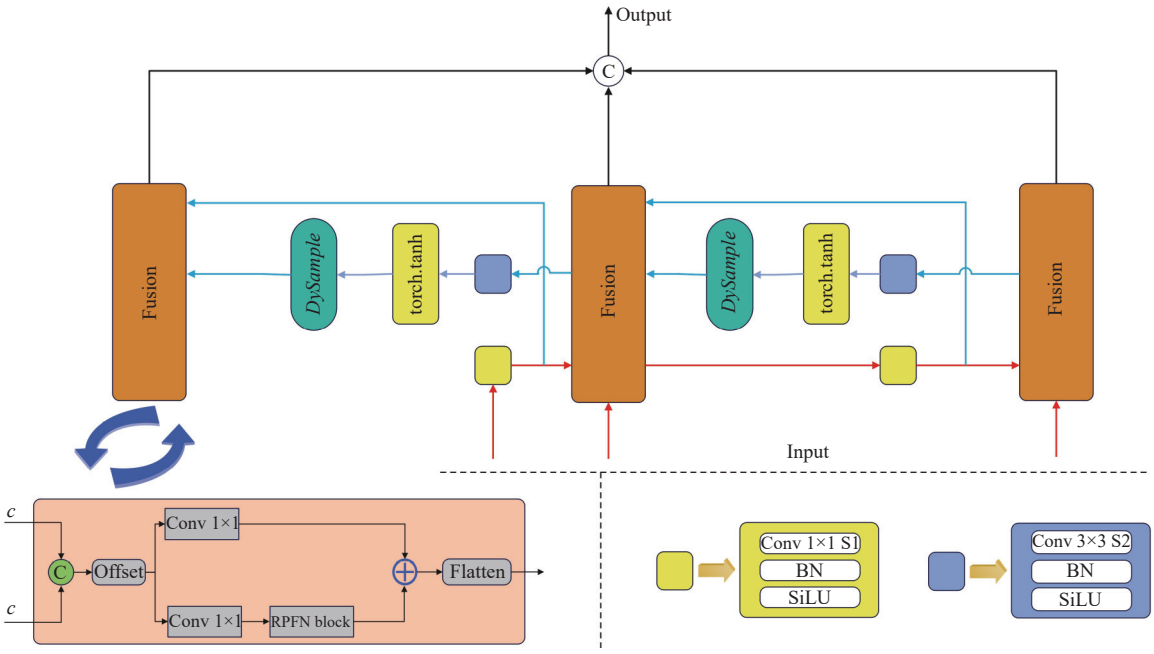



图 5 DFRFM 模块

### 1.4 *GSNL* 损失函数

针对金属杆件表面缺陷检测中面临的小目标漏检与不规则缺陷定位偏差问题, 本研究提出几何感知归一化损失函数 *GSNL*. 该复合损失融合归一化 Wasserstein 距离 (*NWD*)[15]与 *Shape-IoU*[16], 通过分布相似性度量与边界框形变自适应机制, 显著增强模型对微小缺陷的敏感性及不规则形态缺陷的定位精度.

#### 1.4.1 *NWD* 损失函数

*NWD* 将边界框建模为二维高斯分布, 通过计算预测框与真实框的 Wasserstein 距离衡量其分布相似性. Wasserstein 距离的主要优点是, 即使没有重叠或重叠可以忽略不计, 也可以度量分布的相似度. 此外, *NWD*

对不同尺度的物体不敏感, 因此更适合测量微小物体之间的相似性.

*NWD* 的核心分为 3 部分, 首先进行高斯建模, 预测框 $B_p=(C_x^p,C_y^p,W_p,h_p)$ 与真实框 $B_g=(C_x^g,C_y^g,W_g,h_g)$ 分别建模为均值为中心坐标、协方差为宽高平方的二维高斯分布:

$$\Sigma_p=diag\left(\frac{W_p^2}{4},\frac{h_p^2}{4}\right)\quad \Sigma_g=diag\left(\frac{W_g^2}{4},\frac{h_g^2}{4}\right) \tag{8}$$

其次为 Wasserstein 距离计算:

$$W_2^2(B_p,B_g)=(C_x^p-C_x^g)^2+(C_y^p-C_y^g)^2+\frac{(W_p-W_g)^2+(h_p-h_g)^2}{4} \tag{9}$$

最后, 归一化处理转换为相似性分数:

$$NWD=\exp\left(-\frac{\sqrt{W_2^2(B_p,B_g)}}{c}\right) \tag{10}$$

其中, $c$ 为数据集中目标平均尺寸的统计值.

1.4.2 *Shape-IoU* 损失函数

*Shape-IoU* 通过引入形状权重系数优化边界框回归, 聚焦边界框自身形状与尺度对回归的影响, 定义为:

$$Shape\text{-}IoU=(1-s)\cdot IoU(B_p,B_g)+s\cdot Y(B_p,B_g) \tag{11}$$

其中, $s$ 为 *Shape-IoU* 内部尺度因子 (scale), 控制形状相似性的贡献权重, 本研究采取 scale=0.5 的策略, 意味着位置 *IoU* 与形状相似性各占一半, 既保留中大型目标的位置准确性, 又强化形状匹配能力. $IoU(B_p,B_g)$ 为传统交并比, $Y(B_p,B_g)$ 为形状相似性, 采用边界轮廓的 *Hausdorff* 距离归一化:

$$Y=1-\frac{Hausdorff(B_p,B_g)}{Max(W,H)} \tag{12}$$

其中, $W$、$H$ 为图像宽度、高度.

1.4.3 实现策略

*GSNL* 损失函数旨在结合二者优势, 构建对小目标检测与不规则缺陷回归具有双重优化能力的复合损失函数. 该策略采用动态权重因子实现多粒度监督: *NWD* 利用 Wasserstein 距离提升检测小目标缺陷的敏感性, *Shape-IoU* 通过形状权重优化中大型目标的轮廓精度. 在实现层面, 该复合损失函数定义为:

$$GSNL=\alpha\cdot NWD+(1-\alpha)\cdot Shape\text{-}IoU \tag{13}$$

其中, $\alpha$ 为基于目标尺寸分布自适应划分的权重系数, 具体而言, 通过对训练集中目标框的尺寸进行聚类获得尺度中心点, 构造若干连续区间, 并在每一区间内定义分段线性函数, 使得 $\alpha$ 从小尺度段的 1 线性平滑下降至大尺度段的 0. 该策略使得损失函数在不同尺度之间过渡自然, 避免了权重突变带来的梯度不稳定问题, 同时保持训练过程的可微性和尺度自适应性, 从而更有效地兼顾小目标定位精度与大目标形状回归性能.

在工业检测场景中的硬件适配性优化方面, *NWD* 的矩阵加速计算与 *Shape-IoU* 的轻量化算子设计, 有效控制计算开销增长, 保障算法在嵌入式设备中的实时推理能力, 并提升对复杂工业场景的鲁棒性.

### 1.5 SGIA 稀疏全局交互注意力模块

在 RT-DETR 的混合编码器中, AIFI 模块位于混合编码器头部, 通过全局自注意力建模长距离依赖关系, 负责融合主干网络输出的高层特征图 (如 S5 层级), 其原始实现依赖多头自注意力 MHSA[17]进行全局交互, 但计算负载随特征图分辨率呈平方级增长. 在处理金属杆件高分辨率表面图像时, 传统自注意力的二次计算复杂度会导致推理延迟显著增加. 为适配金属杆件表面缺陷检测任务中高分辨率输入以及满足工业实时性的需求, 本研究提出稀疏全局交互注意力模块 SGIA, 在 AIFI 中集成高效加法注意力机制 (efficient additive attention)[18], 该机制通过线性复杂度架构实现全局建模: 以逐元素乘积替代矩阵乘法, 结合动态生成的全局查询向量, 在保持跨像素依赖捕获能力的同时降低计算冗余, 从而优化特征交互效率. 高效加法注意力原理如图 6 所示.

(1) 查询与键的线性投影

输入特征 $X\in R^{N\times d}$ ($N=H\times W$ 为空间位置数, $d$ 为通道维度), 经过两个线性投影生成查询矩阵 $Q$ 和键矩阵 $K$: $Q=XW_q$, $K=XW_k$, 其中 $W_q,W_k\in R^{d\times d}$ 为可学习的投影矩阵.

(2) 全局查询向量生成

通过可学习参数向量 $W_a\in R^{d+1}$ 动态生成注意力权重 $\alpha=Q\cdot\frac{W_a}{\sqrt{d}}$ ($W_a$ 为可学习的注意力权重向量), 随后对查询矩阵 $Q$ 进行加权池化, 生成全局查询向量 $q=\sum_{i=1}^{N}\alpha_iQ_i$, $q\in R^d$, 该向量综合了所有空间位置的语义信息, 表征全局上下文特征.

(3) 键与全局查询的逐元素交互

将全局查询向量 $q$ 与键矩阵 $K$ 进行逐元素乘法, 生成全局上下文特征. 最终输出为原始查询与上下文

特征的残差连接:

$$Output(X') = Q_\rho + T(K \odot q) \tag{14}$$

其中, $Q_\rho$表示归一化查询矩阵, $T$ 表示线性变换, “⊙”表示逐元素乘法.

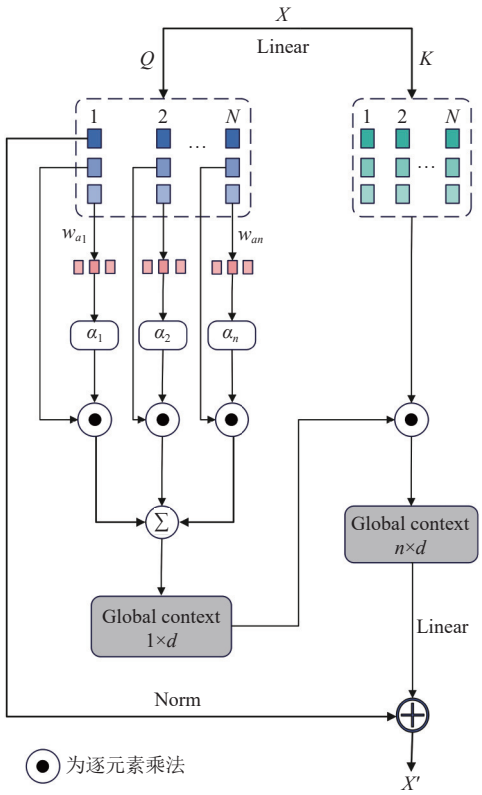



图 6 高效加法注意力机制原理

具体实施过程中, 首先会移除冗余键值交互分支, 舍弃传统自注意力中复杂的键值投影过程, 仅保留查询与键的轻量化交互, 通过单投影路径压缩计算冗余, 降低参数量; 其次, 引入动态权重适配机制, 基于通道注意力门控动态调整注意力权重分布, 抑制反光噪声区域的响应, 同时增强缺陷边缘区域的聚焦能力; 最后, 构建跨层级特征融合通路, 将 SGIA 模块与 DFRFM 模块输出的多尺度特征图深度融合, 通过层级间特征互补与尺度对齐策略, 强化模型对微裂纹、氧化斑等跨尺度缺陷的表征一致性. 此重构策略在降低计算复杂度的同时, 实现了局部细节保留与全局语义建模的平衡. SGIA 模块如图 7 所示.

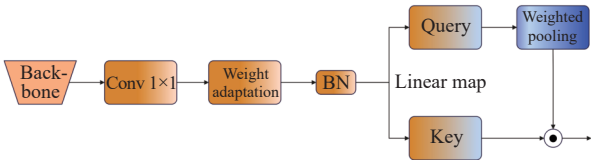



图 7 SGIA 模块

## 2 实验分析

### 2.1 实验环境及参数配置

实验环境使用的是 Linux 5.15.0 操作系统, GPU 为 NVIDIA GeForce RTX 3090, 24 GB 显存, 运行内存 24 GB. 详细设置如表 1 所示.

表 1 环境详细设置

| Type | Version | Type | Value |
|---|---|---|---|
| GPU | RTX 3090 | Epoch | 300 |
| Python | 3.10.14 | Batch | 24 |
| PyTorch | 2.3.0 | Learning rate | $1\times10^{-4}$ |
| CUDA | 12.1.1 | *Shape-IoU* scale | 0.5 |
| 显存 | 24 GB | Warmup_epochs | $2\times10^{3}$ |
| 运行内存 | 24 GB | Weight_decay | $1\times10^{-4}$ |

### 2.2 数据集介绍

本研究采用东北大学发布的公开金属表面缺陷数据集 NEU-DET[19]作为实验基准. 该数据集包含裂纹 (Cra)、轧制氧化皮 (Inc)、夹杂 (Pat)、麻点 (Pis)、斑块 (Ris) 和划痕 (Scr) 这 6 类典型工业缺陷, 每类包含 300 张灰度图像, 总样本量 1 800 张. 为更清晰地展示缺陷特征, 本文在示例图中对原始样本进行了局部放大处理, 并适度增强放大缺陷区域与周围区域的对比度, 以便更直观地突出缺陷细节, 如图 8 所示.

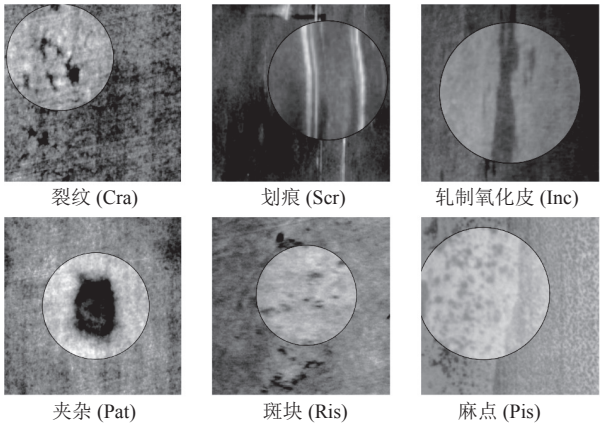



图 8 各类缺陷示例图

数据集按 8:1:1 比例划分为训练集 (1 440 张)、验证集 (180 张) 与测试集 (180 张), 所有图像均采集自真实钢铁产线, 涵盖不同光照强度、金属表面氧化程度及缺陷尺度, 具有强工业场景代表性.

### 2.3 评价指标

为了准确评估本研究的有效性, 本研究采用平均精确率 *AP*、*mAP*@0.5 以及 FPS 作为核心评价指标. 平均精确率即为衡量模型正确检测各类缺陷的能力, 具体公式为:

$$P = \frac{TP}{TP + FP} \tag{15}$$

其中, $TP$ 为正确检测的缺陷数量, $FP$ 为误检的背景区域数量.

$mAP$@0.5 为综合评估类别缺陷的平均检测精度, 计算方式为:

$$AP_i = \int_0^1 P_i(R_i)\mathrm{d}R_i \tag{16}$$

$$mAP = \frac{\sum_{i=1}^{N} AP_i}{N} \tag{17}$$

其中, $P_i$、$R_i$ 分别表示第 $i$ 类缺陷的精确率-召回率曲线积分结果, $N$=6 为缺陷类别总数.

$FPS$ 即衡量模型实时性, 定义为:

$$FPS = \frac{1}{\text{平均单帧推理时间(s)}} \tag{18}$$

精确率与 $mAP$@0.5 侧重检测精度, FPS 则保障工业部署的实时性要求, 三者共同构建算法性能的综合评估体系.

## 2.4 实验结果

为评价改进算法对金属杆件表面缺陷的检测性能, 实验对比了改进模型和 RT-DETR 模型在 NEU-DET 数据集上的多种评价指标, 包括 $mAP$@0.5、$mAP$@0.5:0.95、FPS、GFLOPs、参数量以及 6 种缺陷的 $AP$ 值. 具体实验结果如表 2 所示.

可以看出, 改进算法模型在 6 种缺陷上 $AP$ 值均有提高. 其中, 裂纹、轧制氧化皮、夹杂、麻点、斑块、划痕的 $AP$ 值分别提高了 10.4%、2%、7.7%、2.7%、5.6%、5.9%, mAP@0.5 和 $mAP$@0.5:0.95 分别提升 2.8%、0.6%. 此外, 与原始 RT-DETR 相比, 改进算法模型在推理速度上表现出明显的优越性, FPS 达到 166.51, 比 RT-DETR 高出了 8.55 f/s. 通过算法优化, 计算复杂度从 57 GFLOPs 降至 43.1 GFLOPs (降低 24.4%), 有效减少运算能耗, 并降低对计算单元和内存带宽的压力, 提升数据复用率, 满足工业检测等实时响应场景的严苛需求. 综合实验结果表明, 本文改进模型能够有效识别金属杆件表面缺陷.

表 2 实验结果

| Model | $AP$ (%) | | | | | | $mAP$@0.5 (%) | $mAP$@0.5:0.95 (%) | FPS (f/s) | GFLOPs | Params (MB) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | Cra | Inc | Pat | Pis | Ris | Scr | | | | | |
| RT-DETR | 58.9 | 71.2 | 77.5 | 71.5 | 61.3 | 91.6 | 72.6 | 41.8 | 157.96 | 57.0 | 20.09 |
| RDGS-DETR | 69.3 | 73.2 | 85.2 | 74.2 | 66.9 | 97.5 | 75.4 | 42.4 | 166.51 | 43.1 | 14.02 |

## 2.5 可视化分析

图 9 展示了本文算法网络与原始的 RT-DETR 网络在相同配置下进行 300 轮训练的 $mAP$@0.5 曲线图. 图中红色曲线代表使用本文网络训练时的 $mAP$@0.5 曲线, 黑色代表使用原始 RT-DETR 网络训练时的 mAP@0.5 曲线, 横坐标表示训练的轮数, 纵坐标表示 $mAP$@0.5 的值. 可以明显观察到随着训练次数的增加, 本文改进模型表现得更为稳定, 在整体精度上表现更优, 最终取得更高的 $mAP$@0.5, 体现出更好的检测精度和收敛性. 图 10 展示了在同样配置下训练过程中损失值的变化曲线图, 可以明显看出本文改进算法 RDGS-DETR 在训练过程中的损失值要明显低于原始 RT-DETR 算法, 数值越低, 代表预测框与真实框越重合, 定位越准确. 而数值高代表预测框存在较大位置偏差或与真实框无重叠.

在训练过程中, 2 000 次迭代的长预热机制通过逐步提升学习率至目标值, 使模型在初期均匀学习所有类别特征, 为后续收敛奠定稳定性基础; 同时, 权重衰减 (Weight_decay=0.0001) 作为正则化策略, 配合同等量级的低学习率形成双重约束: 低学习率确保模型在细微缺陷的局部最优解区域精细更新参数, 避免因步长过大跳过关键特征, 而权重衰减则通过温和的正则化抑制复杂背景噪声引发的过拟合, 使训练过程呈现较为平滑的收敛曲线.

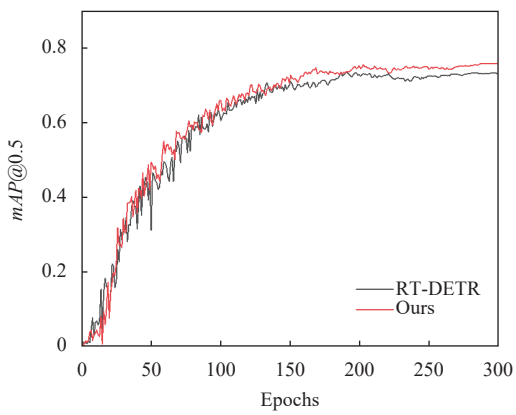



图 9 $mAP$@0.5 曲线对比图

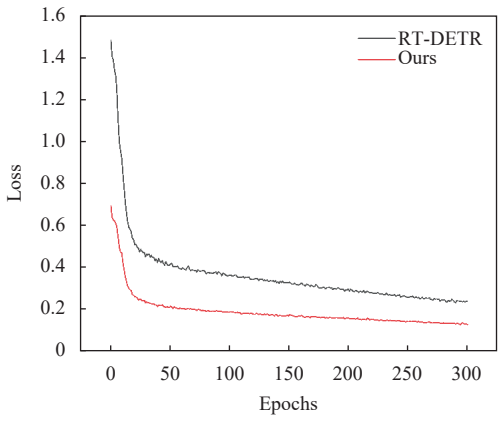



图 10　损失值曲线对比图

此外, 在 RDGS-DETR 模型中, *Shape-IoU* 的内部尺度因子 scale 的取值对平均损失值的影响呈现规律性. 如图 11 所示, 当 $s$=0.5 时, 平均损失值降至所有配置中的最低, 实现了形状特征与位置特征的均衡贡献, 既保留了对缺陷轮廓的识别能力, 又不忽略边界框的定位精度, 因此成为该任务下的最优参数选择. 若 $s$ 过大, 会因过度强调形状特征而削弱对边界框位置准确性的关注, 导致平均损失值较大. 若 $s$ 过小, 形状特征的权重占比不足, 无法有效区分缺陷与背景的形状差异, 损失值同样也会增大.

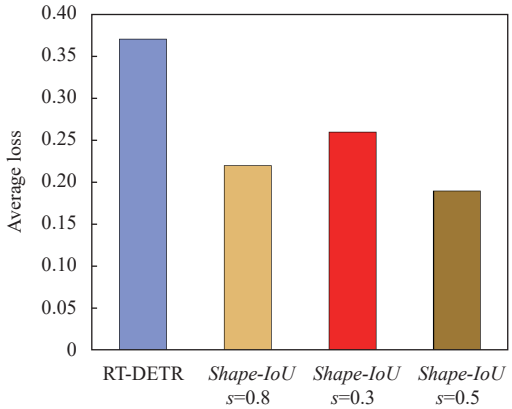



图 11　不同 *Shape-IoU* 尺度因子下平均损失值

为深入探究模型的内在机制与决策依据, 本研究对模型生成的特征响应热力图同样进行了可视化分析. 该热力图通过颜色深浅直观地映射神经网络对输入图像不同区域的响应强度, 其中颜色越深的区域对应模型关注度越高, 通常也意味着该区域具备更优的检测精度. 热力图对比结果如图 12 所示.

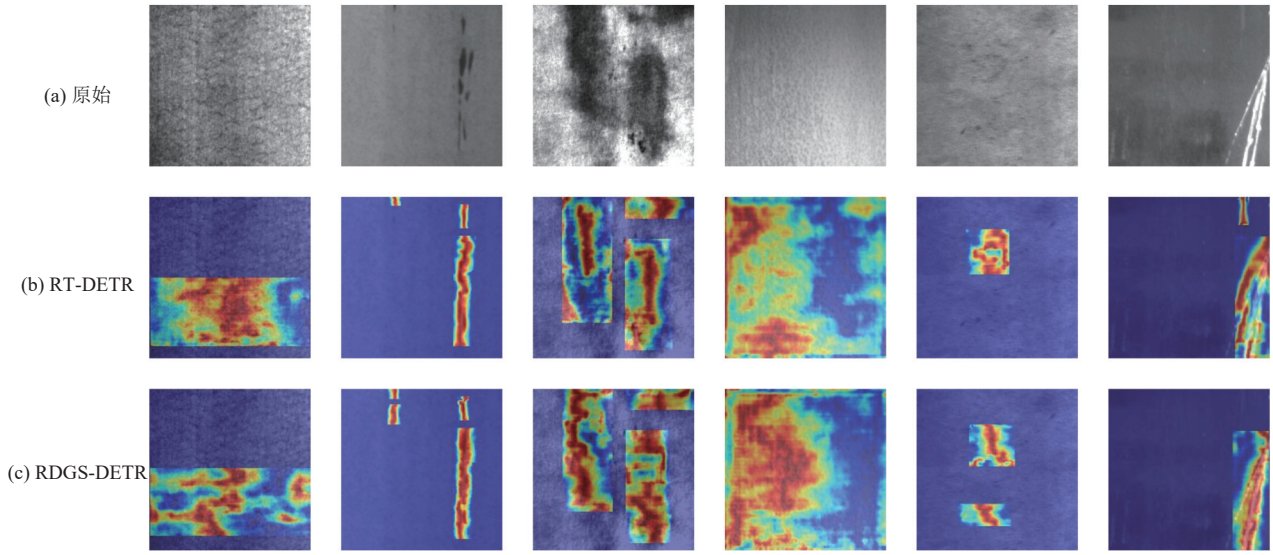
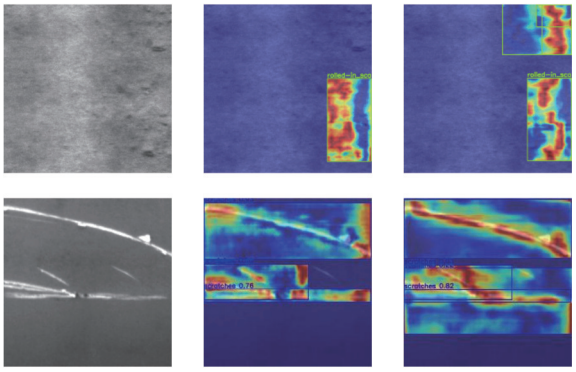



图 12　检测结果对比热力图

通过细致对比热力图分布特征发现: 原始 RT-DETR 虽能较完整地框定潜在缺陷区域, 但在检测精度方面存在局限. 具体表现在其对目标区域关键特征的聚焦程度不足, 且背景干扰抑制能力有待提升. 相比之下, 本研究提出的 RDGS-DETR 在目标区域特征提取方面展现出优势. 其生成的热力图能更精准、清晰地凸显目标缺陷的关键视觉特征, 同时有效抑制非目标背景区域的响应. 这种增强的聚焦能力不仅有效提升模型在复杂背景下识别微小目标缺陷的鲁棒性, 也优化了其注意力分配机制, 降低对无关区域的关注度. 最终, 这些改进共同优化模型性能, 使其整体检测效果获得实质性提升, 尤其在精度和抗干扰性方面.

### 2.6 典型错误检测案例分析

如图 13 所示, (a) 为原始的斑块及划痕缺陷图像, (b) 为初始 RT-DETR 算法的检测结果. 可以观察到, 该算法在处理小目标缺陷时存在一定的漏检现象. 例如, 图 13 第 1 行样例中, 表面右上角的微小斑块类缺陷未能被 RT-DETR 有效识别. 图 13 第 2 行样例中, 金属表面中部偏右位置的细小划痕同样被遗漏, 而改进后的 RDGS-DETR 则能够实现准确且完整的检测.

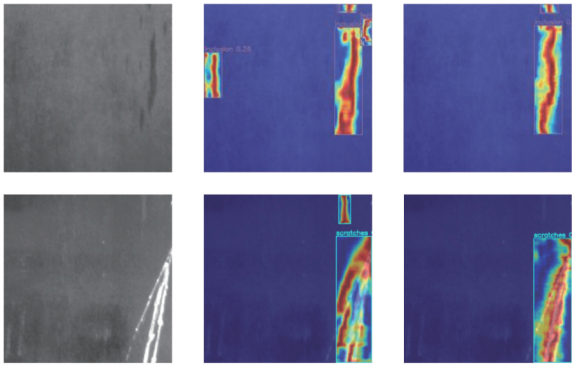


(a) 原始 (b) RT-DETR检测 (c) RDGS-DETR检测

图 13 小目标缺陷漏检案例对比

此外, 从热力图的响应情况来看, RT-DETR 在缺陷区域的注意力明显不足. 在图 13(b) 中, 划痕缺陷区域的响应以低强度蓝色为主, 仅在划痕右端点出现了较弱的红色响应, 而在划痕的中段及细微分支处几乎没有得到有效响应. 这说明模型未能从纹理背景中提取出划痕独特的边缘梯度特征, 导致对缺陷区域的注意力被分散, 难以聚焦真实缺陷.

如图 14 所示, 在存在相似干扰纹理的情况下, RT-DETR 容易出现误判. 例如, 在图 14 第 1 行中, 真实的轧制氧化皮缺陷位于图像右侧, 但由于背景中存在与缺陷形态相似的干扰纹理, RT-DETR 在左侧的非缺陷区域产生错误检测. 进一步地, 从图 14 第 2 行可以观察到, 图像右侧确实存在明显的纵向划痕缺陷, 而右上角则有一处与划痕纹理高度相似的非缺陷区域 (如加工遗留的肌理等). 该区域虽然呈现出与划痕相似的线性纹理特征, 但缺乏划痕所特有的深度感以及形态上的连贯性, 因此仍被错误地识别为缺陷. 相比之下, 改进后的 RDGS-DETR 能够更准确地抑制干扰响应, 仅在真实缺陷区域给出检测结果, 避免将正常纹理误判为缺陷. 这表明, RDGS-DETR 在特征判别能力与抗干扰性方面具有明显优势, 能够在复杂背景条件下实现更可靠的缺陷检测.



(a) 原始 (b) RT-DETR检测 (c) RDGS-DETR检测

图 14 干扰纹理误判案例对比

### 2.7 消融实验

为验证改进模块对性能的提升效果, 以 RT-DETR-r18 为基准模型进行消融实验, 通过逐步引入 RPFN block、DFRFM 模块、*GSNL* 损失函数、SGIA 模块, 从 *mAP*@0.5、推理速度 FPS 与模型参数量这 3 方面验证各模块的作用与协同效应. 消融实验结果如表 3 所示, “√”代表在算法中使用该改进方法, “—”代表未使用该改进方法.

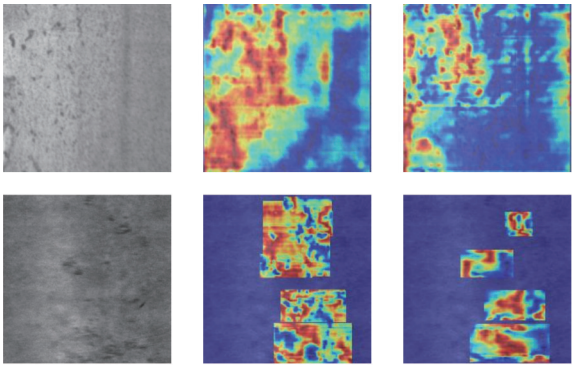
表 3 消融实验结果

| RT-DETR-r18 | RPFN bolck | DFRFM | *GSNL* | SGIA | *mAP*@0.5 (%) | FPS (f/s) | Params (MB) |
|---|---|---|---|---|---|---|---|
| √ | — | — | — | — | 72.6 | 157.96 | 20.09 |
| √ | √ | — | — | — | 73.9 | 151.55 | 14.01 |
| √ | — | √ | — | — | 72.8 | 160.32 | 19.90 |
| √ | — | — | √ | — | 74.1 | 168.03 | 19.88 |
| √ | — | — | — | √ | 72 | 177.92 | 19.88 |
| √ | √ | √ | — | — | 73.3 | 169.52 | 14.02 |
| √ | √ | — | — | √ | 72.7 | 178.68 | 14.01 |
| √ | — | √ | √ | — | 73.8 | 164.19 | 19.9 |
| √ | √ | √ | √ | — | 73.5 | 164.00 | 14.03 |
| √ | √ | √ | √ | √ | 75.4 | 166.51 | 14.02 |

基准模型的 *mAP*@0.5 为 72.6%、FPS 为 157.96 f/s、参数量约 20.09 MB, 为性能对比提供基线. 单独引入 RPFN block 时, *mAP*@0.5 提升至 73.9% (+1.3%), 参数量得到大幅优化, 降至 14.01 MB, FPS 略降至 151.55

f/s, 表明 RPFN block 在增强特征交互、轻量化模型的同时, 对推理速度存在一定权衡取舍; DFRFM 模块单独加入后, *mAP*@0.5 仅微增至 72.8% (+0.2%), FPS 提升至 160.32 f/s, 参数量变化平缓, 说明其对精度增益有限, 更侧重维持速度与轻量化; *GSNL* 损失函数单独引入后, *mAP*@0.5 提升至 74.1% (+1.5%), FPS 跃升至 168.03 f/s, 体现该模块在优化边界框匹配与形状感知的同时, 有效平衡了速度.

SGIA 模块单独作用时, 推理速度相较原始 RT-DETR 有较大提升, 达到 177.92 f/s, 但 *mAP*@0.5 出现了小幅度下降. 通过图 15 麻点与斑块的注意力热力图可以观察到, SGIA 通过线性复杂度的加法注意力实现了更广泛的全局响应覆盖, 但其在峰值抑制与边界聚焦能力不足. 原因在于: 首先, 移除冗余键值交互与采用单投影路径虽然有效降低计算冗余, 但特征交互维度受限, 难以充分建模细粒度空间依赖, 从而削弱了表征能力; 其次, 加法注意力在相似性度量上缺乏动态放大特性, 即使有通道门控动态调节, 也难以在小尺度高对比度缺陷上形成强响应, 导致局部定位精度下降. 对于边界不规则的缺陷, 虽然 SGIA 能够捕获主体区域, 但边界描述相对模糊, 从而影响整体检测精度. 由此可见, SGIA 模块在降低计算复杂度、提升推理速度的同时, 牺牲了一部分对小目标和复杂边缘的表征能力, 体现出一种典型的实时性与检测精度之间的权衡关系.



(a) 原始 (b) RT-DETR-SGIA (c) RDGS-DETR

图 15 SGIA 模块单独应用时热力图对比

模块组合实验中, RPFN block 与 DFRFM 的叠加使 *mAP*@0.5 达到了 73.3%, 该数值低于 RPFN block 单独引入时的 73.9%, 却高于 DFRFM 单独作用下的 72.8%. FPS 提升至 169.52 f/s, 这表明特征融合逻辑与动态采样策略的协同, 并非简单的性能“折中”, 反而在推理速度维度实现了“1+1>2”的增益.

RPFN block 与 SGIA 组合时, *mAP*@0.5 为 72.7%, 虽低于 RPFN 单独贡献的 73.9%, 却高于 SGIA 单独作用下的 72%, FPS 进一步提升至 178.68 f/s, 此结果说明二者组合在“速度优先”场景下的协同优势显著, 并且能够弥补 SGIA 单独作用时的权衡关系. 此外, Params 为 14.01 MB, 保持与 RPFN block 单独引入时相近的轻量化水平. DFRFM 与 *GSNL* 协同时, *mAP*@0.5 达 73.8%, 该数值高于 DFRFM 单独引入时的 72.8%, 略低于 *GSNL* 单独作用下的 74.1%, FPS 为 164.19 f/s, 维持在相对高效的水平, Params 为 19.9 MB, 与 DFRFM 单独引入时的参数量基本一致. 这反映 DFRFM 对数据分布的适配能力辅助 *GSNL* 优化检测逻辑, 二者在缺陷特征增强与样本匹配层面的功能耦合, 可在效率代价可控的范围内推动检测精度进阶.

当 RPFN block、DFRFM、*GSNL* 协同作用时, *mAP*@0.5 达 73.5%, FPS 保持在 164 f/s, 展现多模块在轻量化与精度间的部分协同; 最终全部模块 (RPFN block、DFRFM、*GSNL*、SGIA) 集成后, *mAP*@0.5 显著提升至 75.4%, 较基准提升 2.8%, 参数量稳定在 14.02 MB, FPS 保持 166.51 f/s, 证明各模块在特征提取、干扰抑制与速度优化的协同设计下, 实现检测精度、轻量化与推理效率的综合提升, 验证了改进策略的有效性与模块间的互补性.

### 2.8 鲁棒性测试

为全面评估改进模型的鲁棒性, 本研究进一步在 GC10-DET[20]金属表面缺陷数据集上进行鲁棒性验证. 该数据集涵盖冲孔 (Pu)、油斑 (Os)、轧坑 (Rp)、焊缝 (Wl)、新月形缝隙 (Cg)、水斑 (Ws)、丝斑 (Ss)、夹杂 (Inc)、折痕 (Cre)、腰部折痕 (Wf) 这 10 类复杂工业缺陷, 图像分辨率高达 2048×1000 像素, 其缺陷形态、光照条件及背景噪声与 NEU-DET 存在显著分布差异, 是验证模型泛化能力的理想基准. 实验采用零样本迁移策略, 将基于 NEU-DET 训练的改进模型直接部署于 GC10-DET 测试集, 不进行任何参数微调. 鲁棒性测试实验结果如表 4 所示.

结果显示, 改进模型的 *mAP*@0.5 为 68.4%, 较原始 RT-DETR 提升了 3.1%. *mAP*@0.5:0.95 达到 33.1%, 超越原始模型的 30.1%. 推理速度达到 201.69 f/s, 较原始模型提升 6.7%, 验证了改进模型在不同场景下仍保持高效率优势.

表 4 鲁棒性测试结果

| Model | *AP* (%) | | | | | | | | | | *mAP*@0.5 (%) | *mAP*@0.5:0.95 (%) | FPS (f/s) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Pu | Os | Rp | Wl | Cg | Ws | Ss | Inc | Cre | Wf | | | |
| RT-DETR | 88.1 | 71 | 50.7 | 85.6 | 84.2 | 85.2 | 76.3 | 50.4 | 85.2 | 70.7 | 65.3 | 30.1 | 189.04 |
| RDGS-DETR | 89.3 | 72.1 | 67.5 | 88.5 | 86.5 | 91.7 | 77 | 55.3 | 87.2 | 78.1 | 68.4 | 33.1 | 201.69 |

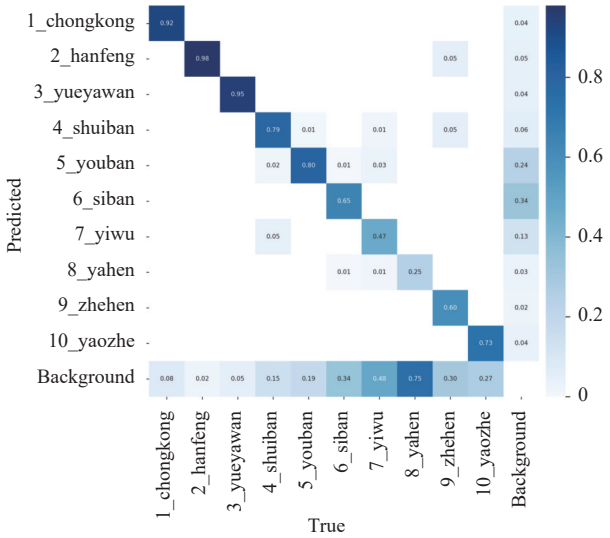
在类别精确度方面, RDGS-DETR 对不同缺陷类型展现出良好的适应性. 轧坑类提升最为显著, *AP* 值由 50.7% 提高至 67.5%, 增幅达 16.8%, 夹杂和水斑类缺陷的 *AP* 值也取得较大提升, 分别增加 4.9% 与 6.5%, 表明模型在形态复杂缺陷的特征建模能力得到增强; 焊缝及腰部折痕类缺陷分别提升 2.9%、7.4%, 体现出方法在边界模糊缺陷检测中的鲁棒性. 对于新月形缝隙、冲孔与折痕等边界清晰、形态稳定的缺陷, 模型 *AP* 值持续保持在 86% 以上, 体现了其在边界框回归精度上的稳健性.

值得注意的是, 油斑以及丝斑类的 *AP* 提升幅度相对有限, 反映出模型在细粒度特征区分和小样本类别泛化方面仍存在一定挑战, 后续可通过针对性数据增强与类别自适应学习策略进一步优化. 总体来看, RDGS-DETR 在性能方面表现优越, 证明了其在工业场景质检中的应用能力.

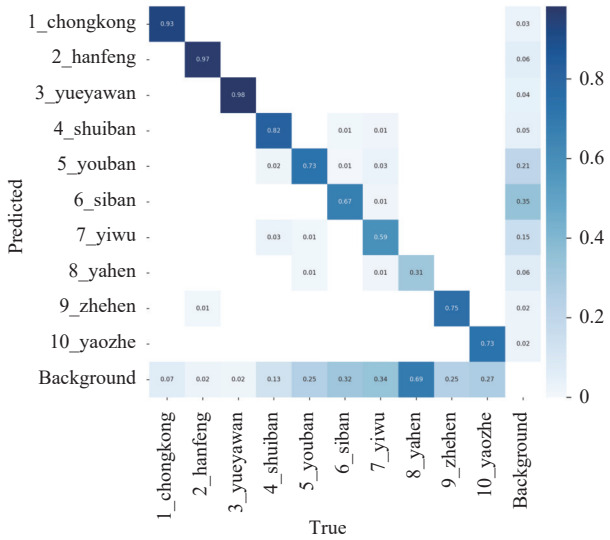
此外, 混淆矩阵作为评估模型性能的核心工具, 直观展示了模型在鲁棒性测试实验中对 GC10-DET 数据集 10 类缺陷的预测结果与真实标签的对应关系, 其对角线元素代表正确分类的样本数, 非对角线元素则代表误分类的样本数. RT-DETR 和 RDGS-DETR 的混淆矩阵如图 16(a)、图 16(b) 所示.

由此可见, RDGS-DETR 的对角线元素数值普遍较高, 非对角线元素数值较低, 说明 RDGS-DETR 在大多数缺陷上的分类准确性明显优于 RT-DETR, 充分验证了改进算法的鲁棒性优势.

性能提升源于 4 项协同改进: 轻量化 RPFN 模块通过优化卷积稀疏性增强微小缺陷特征捕获能力, 保证跨分辨率场景下的稳定表达; 动态上采样算子 *DySample* 自适应调整采样策略, 改善高分辨率图像中多尺度特征的空间错位, 提升形变缺陷几何对齐精度; 复合损失函数结合形状感知与分布相似性约束, 降低复杂背景噪声干扰, 强化缺陷形态与空间关系判别能力; 高效加法注意力机制以线性复杂度建模全局语义, 兼顾上下文关联性与计算效率. 这 4 项改进从特征提取、空间对齐、监督优化到计算架构形成闭环, 使模型在跨数据集场景中实现鲁棒性突破, 为工业质检提供精准且高效的解决方案.



(a) RT-DETR



(b) RDGS-DETR

图 16 RT-DETR 和 RDGS-DETR 的混淆矩阵

## 2.9 对比实验

为了更加直观地体现出本文改进算法的性能提升以及有效性, 在相同的实验环境下, 将改进模型同 YOLO 系列等一些主流检测模型在 NEU-DET 数据集上, 进行性能对比. 对比实验结果如表 5 所示.

表 5 对比实验结果

| 方法 | *mAP*@0.5 (%) | *mAP*@0.5:0.95 (%) | FPS (f/s) |
|---|---|---|---|
| 文献[21] | 75.9 | 39.2 | 121.95 |
| 文献[22] | 74.8 | 38.8 | — |
| 文献[23] | 73.7 | — | 73.46 |
| 文献[24] | 74.4 | — | — |
| YOLOv7 | 71.2 | 37.9 | 115.52 |
| YOLOv8 | 71.7 | 39.2 | **169.20** |
| YOLOv9 | 70.4 | 39.4 | 150.74 |
| YOLOv10 | 66.0 | 37.7 | 162.03 |
| RT-DETR | 72.6 | 41.8 | 157.96 |
| RDGS-DETR | 75.4 | 42.4 | 166.51 |

从表 5 对比实验结果可知, 所提 RDGS-DETR 在检测精度与推理效率上均展现出优势. 在 *mAP*@0.5 维度, RDGS-DETR 达到 75.4%, 不仅高于基础 RT-DETR 的 72.6%, 也超越 YOLOv7、YOLOv8 等主流目标检测算法与多数对比文献; 在更严格的 *mAP*@0.5:0.95 指标中, RDGS-DETR 以 42.4% 的成绩领先于 RT-DETR 的 41.8%, 且优于 YOLO 系列及对比文献方法, 体现模型对不同交并比下缺陷特征的精准识别与分类能力. 同时, RDGS-DETR 的 FPS 达 166.51 f/s, 相比 RT-DETR 的 157.96 f/s 进一步提升, 且高于 YOLOv9 等多数算法, 在保障检测精度的同时强化了实时推理性能. 结合金属杆件表面缺陷检测对精度与效率的双重需求, RDGS-DETR 通过优化特征交互机制实现的性能突破, 为工业场景下复杂缺陷的高效识别提供了更可靠的技术方案.

## 3 结论与展望

本文围绕工业生产中金属杆件表面缺陷检测任务展开研究, 针对工业质检中对检测精度、实时性及复杂场景适应性的要求, 在 RT-DETR 原模型基础上提出了改进算法 RDGS-DETR. 通过引入轻量化的 RPFN block 模块, 提升了对微小裂纹的特征提取能力并减少模型参数; 融合 *DySample* 动态上采样算子优化多尺度特征对齐精度; 设计融合归一化 Wasserstein 距离与 *Shape-IoU* 的复合监督函数, 有效解决传统 *IoU* 在小目标和复杂形态回归中的局限; 采用高效加法注意力机制以降低计算复杂度, 实现实时性的保障. 实验结果表明, RDGS-DETR 在检测精度和推理速度方面均优于原始 RT-DETR 和主流 YOLO 系列模型, 具备良好的综合性能和应用前景. 未来的研究工作将在现有成果的基础上进一步深入, 持续提升模型的实用性与智能化水平. 随着工业检测场景的不断发展, 后续可结合更多实际需求, 不断拓展算法的应用范围, 完善系统性能, 推动智能质检技术在工业领域的深入落地与广泛应用.

## 参考文献


1 Liu JF, Cao XL, Xu J, *et al*. A new method for threshold determination of gray image. Geomechanics and Geophysics for Geo-energy and Geo-resources, 2020, 6(4): 72. [doi: 10.1007/s40948-020-00198-2]

2 Cousty J, Najman L, Dias F, *et al*. Morphological filtering on graphs. Computer Vision and Image Understanding, 2013, 117(4): 370–385. [doi: 10.1016/j.cviu.2012.08.016]

3 Redmon J, Divvala S, Girshick R, *et al*. You only look once: Unified, real-time object detection. Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE, 2016. 779–788.

4 Terven J, Córdova-Esparza DM, Romero-González JA. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680–1716. [doi: 10.3390/make5040083]

5 Zhao QH, Liu YQ. RETRACTED ARTICLE: Design of apple recognition model based on improved deep learning object detection framework Faster-RCNN. Advances in Continuous and Discrete Models, 2024, 2024: 49. [doi: 10.1186/s13662-024-03835-2]

6 周彦, 孟江南, 吴佳, 等. 改进 YOLOv5s 的钢板表面缺陷检测算法. 激光与光电子学进展, 2023, 60(4): 0415009.

7 马淑华, 李立振, 秦汉民, 等. 基于 PP-PicoDet-XS 的改进铝型材表面缺陷检测算法. 东北大学学报 (自然科学版), 2024, 45(11): 1557–1564.

8 Carion N, Massa F, Synnaeve G, *et al*. End-to-end object detection with Transformers. Proceedings of the 16th European Conference on Computer Vision. Glasgow: Springer, 2020. 213–229.

9 Zhu XZ, Su WJ, Lu LW, *et al*. Deformable DETR: Deformable Transformers for end-to-end object detection. Proceedings of the 9th International Conference on Learning Representations. OpenReview.net, 2021.

10 Zhao YA, Lv WY, Xu SL, *et al*. DETRs beat YOLOs on real-time object detection. Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle: IEEE, 2024. 16965–16974.

11 Ding XH, Zhang XY, Ma NN, *et al*. RepVGG: Making VGG-style ConvNets great again. Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern

Recognition. Nashville: IEEE, 2021. 13728–13737.

12 Chen JR, Kao SH, He H, *et al*. Run, don't walk: Chasing higher FLOPS for faster neural networks. Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023. 12021–12031.

13 Luo JH, Wu JX. AutoPruner: An end-to-end trainable filter pruning method for efficient deep model inference. Pattern Recognition, 2020, 107: 107461. [doi: 10.1016/j.patcog.2020.107461]

14 Liu WZ, Lu H, Fu HT, *et al*. Learning to upsample by learning to sample. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 6004–6014.

15 Zhou LL, Ran HM, Xiong RL, *et al*. NWD-YOLOv5: A YOLOv5 model for small target detection based on NWD loss. Proceedings of the 6th International Conference on Robotics, Intelligent Control and Artificial Intelligence (RICAI). Nanjing: IEEE, 2024. 542–546.

16 Zhang H, Zhang SJ. Shape-IoU: More accurate metric considering bounding box shape and scale. arXiv:2312.17663, 2023.

17 Tan HC, Liu XP, Yin BC, *et al*. MHSA-Net: Multihead self-attention network for occluded person re-identification. IEEE Transactions on Neural Networks and Learning Systems, 2023, 34(11): 8210–8224. [doi: 10.1109/TNNLS.2022.3144163]

18 Shaker A, Maaz M, Rasheed H, *et al*. SwiftFormer: Efficient additive attention for Transformer-based real-time mobile vision applications. Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision. Paris: IEEE, 2023. 17379–17390.

19 He Y, Song KC, Meng QG, *et al*. An end-to-end steel surface defect detection approach via fusing multiple hierarchical features. IEEE Transactions on Instrumentation and Measurement, 2020, 69(4): 1493–1504. [doi: 10.1109/TIM.2019.2915404]

20 Lv XM, Duan FJ, Jiang JJ, *et al*. Deep metallic surface defect detection: The new benchmark and detection network. Sensors, 2020, 20(6): 1562. [doi: 10.3390/s20061562]

21 马小林, 邓翔予, 钱亚飞, 等. 改进 YOLOv7 的金属表面缺陷检测方法. 北京邮电大学学报, 2025, 48(2): 119–125. [doi: 10.13190/j.jbupt.2024-026]

22 Ma XL, Deng XY, Kuang HL, *et al*. YOLOv7-BA: A metal surface defect detection model based on dynamic sparse sampling and adaptive spatial feature fusion. Proceedings of the 6th IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC). Chongqing: IEEE, 2024. 292–296. doi: 10.1109/IMCEC59810.2024.10575022.

23 Ye ZM, Terekhov VI, He J. Research of different neural network architectures for real-time metal surface defect detection. Proceedings of the 7th International Youth Conference on Radio Electronics, Electrical and Power Engineering (REEPE). Moscow: IEEE, 2025. 1–5.

24 梁礼明, 龙鹏威, 冯耀, 等. 改进轻量化 VTG-YOLOv7-tiny 的钢材表面缺陷检测. 光学 精密工程, 2024, 32(8): 1227–1240.

(校对责编: 李慧鑫)