

文章编号 1004-924X(2026)15-2422-16

基于双重频域增强与动态核引导的 DFD-DETR 钢材表面缺陷检测

敬天池, 孙晓明*, 郝亮, 徐凯歌, 陈言, 韩永琦, 刘洋

(哈尔滨理工大学 黑龙江省激光光谱技术及应用重点实验室, 黑龙江 哈尔滨 150080)

摘要:针对钢铁表面缺陷检测中复杂背景干扰导致的低对比度缺陷特征难以表征,以及跨尺度特征融合过程中局部细节易丢失的问题,提出一种融合双重频域增强与动态核引导的改进型检测框架 DFD-DETR。首先,针对轧制纹理等低频背景对缺陷边缘特征的干扰,设计了双重频域增强模块(Dual-Frequency Domain Enhancement, DFE)。该模块由前置高通滤波与后置频谱门控两部分组成,前者通过固定高通响应抑制轧制纹理等低频背景信息,后者利用可学习频谱门控对融合后的频域特征进行自适应重校准,从而增强微弱缺陷的边缘响应。随后,为在一定程度上克服传统特征融合过程中浅层细节被深层语义覆盖的困难,提出动态核引导模块(Dynamic Kernel Guidance, DKG)。利用深层语义特征生成空间自适应动态卷积核,对浅层高分辨率特征进行局部引导融合,从而增强多尺度缺陷的结构表征能力。在 NEU-DET 与 GC10-DET 数据集上的实验结果表明,本文方法的 mAP@0.5 分别达到 80.8% 与 73.4%,相比基线 RF-DETR,所提出方法在复杂背景下的低对比度及多尺度缺陷检测任务中取得了更优的检测结果,具有较好的检测性能与一定的泛化能力。

关键词:钢铁表面;缺陷检测;RF-DETR;频域增强;多尺度特征融合

中图分类号:TP391.4 **文献标识码:**A

doi:10.37188/OPE.20263415.2422

CSTR:32169.14.OPE.20263415.2422

Steel surface defect detection based on DFD-DETR with dual-frequency enhancement and dynamic kernel guidance

JING Tianchi, SUN Xiaoming*, HAO Liang, XU Kaige, CHEN Yan, HAN Yongqi, LIU Yang

(Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application,
Harbin University of Science and Technology, Harbin 150080, China)

* Corresponding author, E-mail: sunxiaoming@hrbust.edu.cn

Abstract: Steel surface defect detection is challenged by complex industrial backgrounds, which weaken low-contrast defect representations, as well as by the loss of local structural details during cross-scale feature fusion. To address these issues, an improved detection framework, termed DFD-DETR, is proposed, integrating a dual-frequency-domain enhancement strategy with a dynamic kernel guidance mechanism. Specifically, a Dual-Frequency Domain Enhancement (DFE) module is designed to suppress low-frequency background interference caused by rolling textures while enhancing the high-frequency edge re-

收稿日期:2026-05-26;修订日期:2026-07-07.

基金项目:2025年度黑龙江省自然科学基金联合基金培育项目(No. PL2025F021)

sponses of subtle defects. The DFE module consists of a pre-stage high-pass filtering module and a post-stage spectral gating module. In the pre-stage, low-frequency background components, such as rolling textures, are suppressed by a fixed high-pass frequency response, whereas in the post-stage, the fused frequency-domain features are adaptively recalibrated through a learnable spectral gating mechanism, thereby strengthening the edge responses of subtle defects. To further mitigate the loss of fine-grained information during multi-scale feature interaction, a Dynamic Kernel Guidance (DKG) module is introduced. Spatially adaptive dynamic convolution kernels are generated from high-level semantic features, and locally guided aggregation is performed on high-resolution shallow features, thereby enhancing structural representation across multiple defect scales. Extensive experiments on the NEU-DET and GC10-DET datasets demonstrate that the proposed method achieves mAP@0.5 scores of 80.8% and 73.4%, respectively. Compared with the baseline RF-DETR, the proposed framework exhibits superior performance in low-contrast and multi-scale defect detection under complex backgrounds, indicating improved detection capability and generalization performance.

Key words: steel surface; defect detection; RF-DETR; frequency enhancement; multi-scale feature fusion

1 引言

随着钢铁工业智能制造与质量控制需求的不断提升,钢材表面缺陷检测已成为工业视觉领域的重要研究方向。钢板在轧制、冷却及运输过程中容易产生裂纹、压痕及水斑等表面缺陷,这类缺陷往往会削弱材料的力学性能,并对航空航天、高速铁路及压力容器等关键装备的安全性产生影响。因此,构建高精度、高稳定性的钢铁自动化缺陷检测方法具有重要的工程意义^[1]。

传统机器视觉方法通常依赖人工设计特征进行缺陷识别,在复杂工业背景下容易受到光照变化、纹理干扰及噪声分布差异的影响,泛化能力有限。近年来,深度学习技术凭借其强大的特征表征能力,在工业缺陷检测领域得到广泛应用^[2-3]。其中,以 YOLO^[4] 为代表的卷积神经网络(Convolutional Neural Network, CNN)方法具有较高的推理效率与局部纹理建模能力,在工业场景中应用广泛。然而,卷积操作受限于局部感受野,对长距离上下文关系的建模能力不足;同时,基于锚框与非极大值抑制(Non-Maximum Suppression, NMS)的后处理方式也会增加复杂场景下的误检与漏检风险。

为增强全局特征建模能力,以 DETR^[5] 为代

表的 Transformer 检测架构逐渐被引入工业缺陷检测任务。该类方法利用自注意力机制建立全局依赖关系,能够在复杂背景下获得更丰富的上下文信息,不过存在收敛速度慢的问题。随后提出的 Deformable DETR^[6] 通过可变形注意力机制实现多尺度特征交互,在一定程度上解决了 DETR 收敛速度慢及小目标检测能力不足的问题。尽管 Transformer 架构提升了全局语义建模能力,但在真实钢材缺陷场景中,现有方法仍面临两个关键问题。

一方面,钢材表面通常伴随复杂轧制纹理与随机噪声,低对比度缺陷的边缘响应容易被背景纹理淹没。现有方法大多基于空间域进行特征增强,缺少对频域信息的有效利用,在复杂背景下仍容易出现缺陷特征弱化的问题。另一方面,现有特征金字塔结构(如 FPN^[7]、PAN^[8])通常采用固定拼接或逐元素相加方式进行跨尺度特征融合,浅层高分辨率特征中的局部纹理信息在深层语义传递过程中容易被覆盖,进而影响微小缺陷与非结构化缺陷的检测稳定性。尤其对于裂纹、划痕等尺度跨度较大且边缘不规则的缺陷,固定卷积与静态融合方式难以充分建立局部结构与全局语义之间的关联。

针对上述问题,研究人员开展了大量相关

工作,部分方法通过特征增强来提升模型对复杂缺陷的表征能力。例如,Zhang等^[9]引入可变形卷积增强局部特征建模。郭伟等^[10]通过融合通道移位与倒置残差结构改善多尺度缺陷表达能力。张晓檬等^[11]利用注意力融合机制增强关键缺陷区域响应。这类方法主要集中于空间域的特征增强,在复杂背景下仍容易受到低频纹理干扰。另一方面,也有方法尝试优化跨尺度特征融合过程。彭有有等^[12]通过多尺度膨胀残差结构增强局部感受野。Liu等^[13]利用全局注意力机制改善非结构化缺陷检测能力。Li等^[14]设计了频率感知上采样算法,引入频域信息重构特征来提高密集小目标的检测效果。然而,已有频域增强方法多采用统一频谱变换或通道频率建模策略,缺少针对钢材表面缺陷背景纹理与缺陷边缘频率分布差异的设计,在复杂纹理背景下容易同时增强背景噪声与缺陷响应。另一方面,现有跨尺度特征融合方法主要采用固定融合或内容感知重组方式,对深浅层特征的交互过程缺乏语义引导,难以兼顾全局语义与局部细节信息。因此,如何充分利用频域信息抑制背景干扰,并提高跨尺度特征融合过程中局部细节的保留能力,仍是钢材表面缺陷检测需要解决的问题。

基于上述分析,本文以 RF-DETR^[15]为基础框架,提出一种融合双重频域增强模块与动态核引导机制的钢材表面缺陷检测方法 DFD-DETR。本研究受 Pei 等提出的 FFTFormer^[16]用于增强低光照图像的启发,充分利用频域技术在特征提取过程中保留关键细节的优势,针对复杂背景下低对比度缺陷边缘响应弱的问题,基于傅里叶变换设计了双重频域增强模块(Dual Frequency Enhancement Module,DFE)。该模块在主干网络与颈部特征处理中引入高通滤波模块与频谱门控模块,前置高通滤波侧重于抑制编码阶段的低频背景干扰,后置频谱门控则针对多尺度融合后的特征进行频率自适应重校准,两阶段频域增强分别承担背景抑制与频谱重标定功能。随后,为缓解跨尺度融合过程中浅层细节信息被深层语义覆盖的情况,提出动态核引导模块(Dynamic Kernel Guidance,

DKG)。不同于传统特征金字塔中的固定融合方式,该模块利用深层语义信息生成空间自适应动态卷积核,对浅层高分辨率特征进行局部引导融合,从而增强模型对多尺度及非结构化缺陷的结构表征能力。

2 DFD-DETR 模型

2.1 方法概述

RF-DETR 是一种兼顾检测精度与推理效率的端到端目标检测框架,在 COCO 数据集上能够在较低推理延迟下取得较高检测性能。该模型采用 DINOv2^[17]预训练权重作为主干网络(Backbone),使用大规模视觉预训练策略获得较强的特征表征能力。在特征重组阶段,RF-DETR 通过多尺度 Transformer 编码器(Multi-scale Transformer Encoder)对不同层级特征进行统一建模。具体而言,模型首先对 P3~P6 层特征进行展平与拼接,并引入层级嵌入(Level Embedding)以保留不同尺度特征的空间来源信息。在注意力计算过程中,模型并未采用全局密集注意力,而是基于参考点(Reference Points)在多尺度特征图中预测少量关键采样位置,并对采样区域进行加权聚合。该机制在降低高分辨率特征计算复杂度的同时,能够保留一定的长距离依赖建模能力,从而改善小目标检测性能。

尽管 RF-DETR 在多尺度目标检测任务中表现出较好的检测能力,但在钢铁表面缺陷场景中仍存在一定局限。一方面,ViT 主干网络采用固定 Patch 划分方式,浅层局部纹理信息在特征编码过程中容易受到复杂背景干扰;另一方面,现有特征融合过程仍缺少对浅层细节信息的动态引导机制,在低对比度与尺度变化较大的缺陷场景下,容易出现边缘特征弱化与局部结构信息丢失的问题。

基于上述分析,本文以 RF-DETR 为基线框架,在特征流中引入双重频域增强模块(DFE)与动态核引导模块(DKG),形成“编码特征增强—跨尺度语义引导—频谱重校准”的特征增强流程。其中,前置频域增强用于削弱背景纹理干扰,动态核引导负责利用深层语义增强浅层细节

表达,后置频域增强进一步对融合后的多尺度特征进行频域重校准。整体网络结构如图 1 所示。

2.2 双重频域增强模块

针对 RF-DETR 存在的问题,提出一种双重频域增强模块 (Dual Frequency Enhancement Module, DFE),分别在主干网络的输出端与特征重组阶段引入前置高通滤波模块与后置频谱门控模块。前置高通滤波模块作用于编码器浅层特征,通过固定径向高通掩码抑制低频背景响应,在保持预训练特征分布稳定的前提下突出局部边缘信息;后置频谱门控模块部署于特征投影之后,利用可学习频率掩码、复数频域权重以及频率感知通道注意力,对融合后的频谱进行自适应重校准,从而进一步增强不同尺度缺陷的频率响应,整体结构如图 2 所示。

2.2.1 高通滤波模块

该模块被部署于 Backbone 输出端,作为前置频域滤波结构,通过显式构建一个基于频率中心距离的软性高通掩码,对频域中的低频背景响应进行抑制,同时与保留缺陷相关的高频边缘信息。与传统可学习频域增强不同,该模块仅采用固定高通掩码完成频率筛选,并结合残差结构进行空间域特征精炼,避免在编码器输出端直接引入复杂可学习频域参数,从而降低对预训练特征

分布的扰动。

设编码器最浅层输出特征表示为 $F_e^{(0)} \in \mathbb{R}^{B \times C \times H \times W}$,其中 B 表示批大小, C 表示通道数, H 和 W 分别表示特征图的高度和宽度。首先,对输入特征进行二维实值快速傅里叶变换 (Real Fast Fourier Transform, RFFT),得到复数频谱表示 $X_{\text{freq}} \in \mathbb{C}^{B \times C \times H \times W_f}$,其中 $W_f = W/2 + 1$, C 表示复数域。在频域表示的输出中,低频分量集中在左上角 $(0,0)$,该模块通过计算频域网格中各点到原点的归一化欧式距离,生成了一个距离越远权重越大的软掩码 $M(u, v)$:

$$D(u, v) = \sqrt{u^2 + v^2}, \quad (1)$$

$$M(u, v) = \frac{D(u, v)}{\max(D)}. \quad (2)$$

通过将该掩码与频域特征相乘,可以直接实现对低频背景的滤除和抑制,保留高频边缘特征。提取出的高频信息(如边缘、缺陷轮廓)经过逆变换回到空间域:

$$X_{\text{high_freq}} = \mathcal{F}^{-1}(F \odot M). \quad (3)$$

随后利用卷积注意力机制 (Convolutional Block Attention Module, CBAM) 聚焦缺陷位置,再通过一个轻量级的网络层(由 1×1 卷积、组归一化 Group Norm 及 GELU 激活函数组成)对特征完成精细化处理,最终作为增强特征以残差形

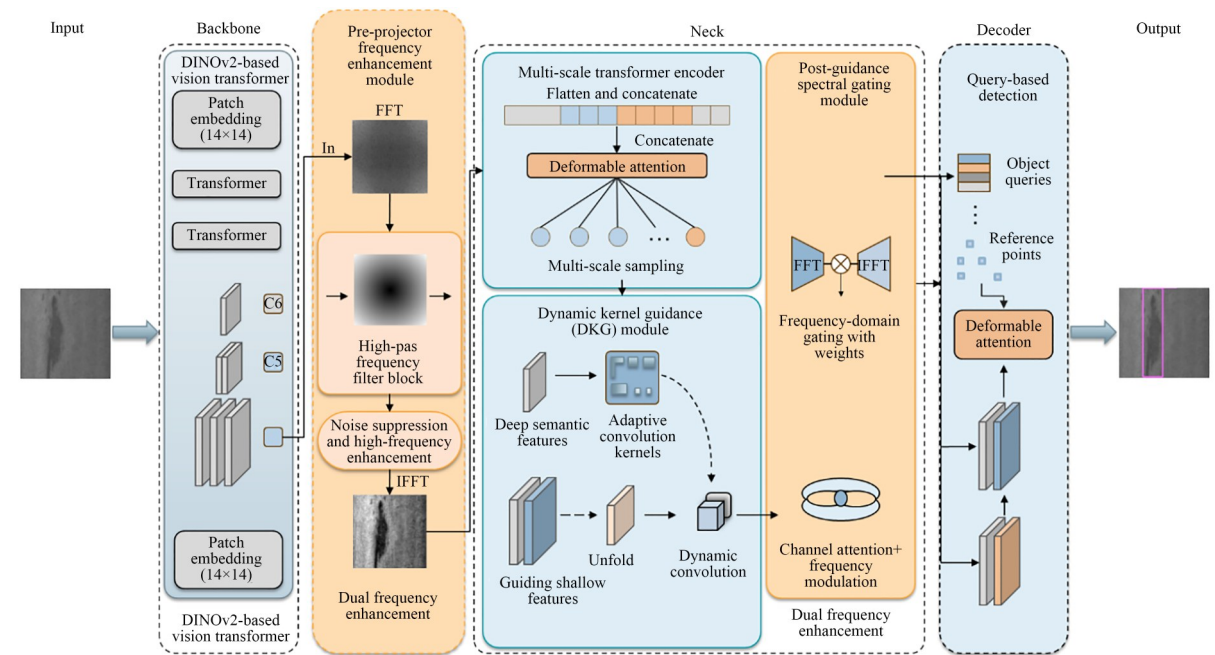


图 1 DFD-DETR 网络架构

Fig. 1 DFD-DETR network architecture

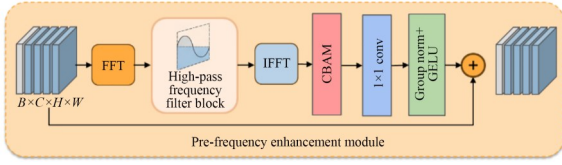


图2 前置频率增强模块

Fig. 2 Pre-frequency enhancement module

式叠加回原始输入：

$$F_e^{(0)'} = F_e^{(0)} + E, \quad (4)$$

其中： E 表示高频增强分支输出。该模块主要用于浅层特征阶段，使高频缺陷边缘在进入后续特征编码前得到保留，从而减弱复杂背景纹理对特征提取过程的干扰。

2.2.2 频谱门控模块

在完成特征重组阶段之后，进一步设计了后置的频谱门控模块，对频域响应进行动态调制。与前置高通滤波不同，该模块更加关注不同频率成分的重要性分配，通过联合通道注意力与可学习的频域权重，对有效频率信息进行增强，模块结构图如图3所示。

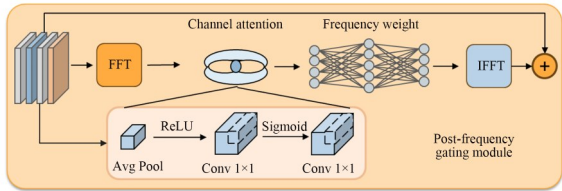


图3 后置频率门控模块

Fig. 3 Post-frequency gating module

设投影器输出的第 i 层特征表示为 $P_i \in \mathbb{R}^{B \times C \times H_i \times W_i}$ ，其中 $C = 256$ ， H_i 和 W_i 分别表示对应尺度的空间分辨率。频谱门控模块对每个尺度独立进行频域增强，输出特征与输入保持相同维度。首先利用全局平均池化和多层感知机生成通道缩放因子 $S \in \mathbb{R}^{B \times C \times 1 \times 1}$ ：

$$S = \sigma(W_2 \cdot \delta(W_1 \cdot \text{AvgPool}(P_i))), \quad (5)$$

其中， δ 为 ReLU 激活函数， σ 为 Sigmoid 激活函数， W_1 ， W_2 为权重矩阵。之后通过二维实数快速傅里叶变换，得到频域表征：

$$F = F(X) \in \mathbb{C}^{B \times C \times H \times \left(\frac{W}{2} + 1\right)}. \quad (6)$$

频域变换之后对频域特征进行调制，引入一组可学习的复数权重为 W_c ， W_r 和 W_i 分别表示

实部和虚部权重，两者均为可学习参数，通过联合调制频率幅值与相位信息，实现不同频率成分的自适应增强：

$$W_{\text{complex}} = W_r + iW_i. \quad (7)$$

A_i 表示复数频域权重， M 表示二维频率掩码，调制后的频域特征为：

$$\tilde{F} = F \odot W_c \odot A_i \odot M, \quad (8)$$

其中： $\odot$ 表示逐元素乘法。

最后，为避免频域调制对原始空间特征造成过度扰动，引入可学习残差系数 α 控制增强特征贡献：

$$Y = X + \alpha F^{-1}(\tilde{F}). \quad (9)$$

相比单一频域增强方式，双重频域增强模块能够分别从背景抑制与频率响应调制两个层面对缺陷特征进行增强。其中，高通滤波模块更关注低频背景过滤，而频谱门控模块则进一步强化微弱缺陷边缘对应的有效频率成分。不同于 FF-TFormer，FcaNet 等频域增强方法，DFE 并非直接利用频域注意力对整体频谱进行增强，而是结合钢材表面缺陷的频率分布特点，将频域增强划分为前置背景抑制与后置频谱调制两个阶段。其中，前置高通滤波主要抑制轧制纹理等低频背景干扰，后置频谱门控在训练后依据缺陷检测任务需求，对不同频率响应进行自适应重分配，在保持低频结构信息的同时增强中高频缺陷响应，并抑制无效极高频噪声，从而实现更符合钢材表面缺陷分布特点的频率表示。

2.3 动态核引导模块

RF-DETR 采用 DINOv2 ViT 作为骨干网络，虽具备较强的全局特征提取能力，但浅层局部纹理特征在跨尺度特征交互过程中仍容易受到深层语义信息的覆盖而弱化。尤其对于裂纹、划痕等边缘不规则且尺度跨度较大的缺陷，传统固定融合方式在不同空间位置采用统一的特征融合策略，难以根据缺陷尺度变化与局部结构差异进行自适应调整。因此，需要一种能够结合空间语义信息进行动态特征重组的融合机制，以增强模型对局部结构信息的保留能力。

针对上述问题，受动态卷积 (Dynamic Convolution)^[18] 以及 CARAFE^[19] 算法中核预测机制与特征重组机制的启发，本研究在 RF-DETR 的特征重组阶段引入了动态核引导模块 (Dynamic

Kernel Guidance, DKG)。该模块利用深层语义特征生成空间自适应动态卷积核,对浅层高分辨率特征进行局部引导重组,以增强跨尺度特征交互过程中对边缘细节与局部结构信息的保留能力。

不同于固定卷积核或静态融合策略,DKG将深层语义信息作为动态引导先验,通过预测空间相关的卷积核权重,对浅层局部邻域特征进行自适应聚合。具体而言,首先利用深层特征生成动态核权重,并通过上采样映射至对应空间分辨率;随后采用 Unfold 操作提取浅层特征的局部窗口信息,并利用动态核对局部区域进行加权重组,从而实现基于局部动态卷积的特征融合过程。经过动态引导后,浅层特征在保留局部纹理与边缘响应的同时,能够进一步引入深层语义信息,从而提升模型对多尺度缺陷的特征表征能力。

此外,考虑到跨层动态融合可能带来的额外计算开销,本文采用局部窗口展开与深度可分离卷积相结合的轻量化设计,并在特征融合后引入轻量级空间注意力机制,对关键区域的响应进行增强。最终,通过残差连接将引导后的特征与原始浅层特征进行融合,在控制参数量与计算复杂度增长的同时,提升模型对复杂背景下微小缺陷及非结构化缺陷的建模能力,模块整体结构如图 4 所示。

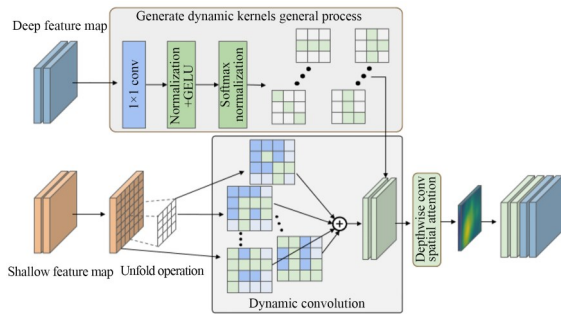


图 4 动态核引导模块

Fig. 4 Dynamic kernel guidance module

设浅层特征为 $F_s \in \mathbb{R}^{B \times C_s \times H \times W}$, 深层特征为 $F_d \in \mathbb{R}^{B \times C_d \times H_d \times W_d}$, 模块输出特征与浅层输入保持一致, 即 $F_{out} \in \mathbb{R}^{B \times C_s \times H \times W}$ 。局部核尺寸为 $k \times k$ 。将深层特征经过 1×1 卷积降维、插值上采样和 Softmax 归一化后, 生成针对每个空间像素的动

态核权重 K , 这一过程为:

$$K_{raw} = \text{Conv}_{1 \times 1} \left(\text{GELU} \left(\text{BN} \left(\text{Conv}_{1 \times 1} (F_d) \right) \right) \right), \quad (10)$$

$$K = \text{Softmax} \left(\text{Upsample} (K_{raw}), \dim = 1 \right). \quad (11)$$

随后, 利用滑动窗口提取浅层特征的邻域信息, 采用 Unfold 操作将浅层特征展开为 $U_s \in \mathbb{R}^{B \times C_s \times k^2 \times H \times W}$, 其中每个空间位置对应一个 $k \times k$ 局部窗口, 为后续动态核加权聚合提供邻域特征。将其与动态核进行加权求和, 得到引导后的特征 F_g , 计算过程为:

$$U_s = \text{Unfold} (F_s, \text{kernel}_{size} = k), \quad (12)$$

$$F_g = \sum_{i=1}^{k^2} U_s[:, :, i, :, :] \odot K[:, :, i, :, :]. \quad (13)$$

为进一步增强关键区域响应, 降低额外参数开销, 引入轻量化的空间注意力机制, 计算空间注意力权重图 A , 通过一个 3×3 深度卷积, 之后经过 BatchNorm、GELU 和 1×1 卷积, 经过 Sigmoid 激活函数最终取得空间注意力权重 A :

$$A = \sigma \left(\text{Conv}_{1 \times 1} \left(\text{GELU} \left(\text{BN} \left(\text{DW}_{3 \times 3} (F_g) \right) \right) \right) \right). \quad (14)$$

最后将引导后的特征 F_{guided} 与空间注意力权重 A 进行逐元素乘法和 F_s 拼接, 最终得到输出 F_{out} :

$$F_{out} = F_s + F_{guided} \odot A. \quad (15)$$

与固定融合方式不同, DKG 通过深层语义预测空间相关动态卷积核, 通过深层语义信息对浅层局部特征进行动态引导, 对浅层局部区域进行自适应重组, 在保留边缘纹理信息的同时增强跨尺度语义关联, 使不同空间位置采用不同的融合权重, 从而改善复杂背景下多尺度缺陷的特征表达能力。相比 CARAFE 和 DySample 等内容感知重组方法, DKG 更加关注深浅层语义交互过程, 在保持计算复杂度基本不变的情况下, 提高了跨尺度特征融合能力。

3 实验

3.1 实验环境与参数设置

本研究实验基于 Visual Studio Code 开发环境实现, 运行于 Windows 11 操作系统。硬件平台包括 Intel Core (TM) i5-13490F 和 NVIDIA

RTX 5060Ti GPU(16 GB 显存)。软件环境包括 Python 3.10.19、PyTorch2.8.0 以及 CUDA12.9。

实验分别在 NEU-DET^[20] 与 GC10-DET^[21] 两个公开钢材表面缺陷数据集上进行,数据集都按照 7:1.5:1.5 的比例划分为训练集、验证集与测试集,为保证实验结果的公平性,所有模块替换实验均采用相同的数据划分随机种子、训练轮数、输入尺寸及优化器参数,训练配置保持一致。所有实验均在相同硬件环境下完成,FPS 测试采用单卡推理结果。初始学习率设置为 1×10^{-4} ,采用余弦退火学习率衰减策略,并结合 Early Stopping 机制防止过拟合。Batch Size 设置为 32,对比试验中针对整个检测器的训练轮数(Epoch)设置为 150,考虑到训练成本问题,针对模块有效性验证的对比实验中 Epoch 均设置为 50。

3.2 数据集

NEU-DET 数据集由东北大学发布,包含六类典型的表面缺陷:裂痕(Crazing, Cr)、夹杂物(Inclusion, In)、斑块(Patches, Pa)、麻点(pitted surface, Ps)、氧化皮压入(rolled-in scale, Rs)和划痕(Scratches, Sc)。每种缺陷类型包含 300 幅图像,图像大小为 200×200 像素。6 种缺陷在 NEU-DET 数据集中分布情况如图 5 所示,数据集中夹杂和斑块缺陷较多,点蚀表面缺陷较少。从目标框的大小分布图可知绝大多数缺陷为中小型缺陷。缺陷位置分布在图片中轴线位置的缺陷较多,小型缺陷较多。

GC10-DET 数据集收集自真实的工业产线,客观反映了复杂多变的实际生产环境。该数据集包含 3 563 幅灰度图像,分辨率为 $2\,048 \times 1\,000$ 像素共涵盖了十类细粒度的表面缺陷类型,包括:冲孔(PH)、焊缝(WL)、月牙弯(CG)、水斑(WS)、油斑(OS)、丝斑(SS)、夹杂(In)、轧辊印(RP)、折痕(Cr)和腰折(WF),包含了钢板加工中易发的各类典型质控难题,存在长尾效应,在实际检测工程中充满挑战性。10 种缺陷在 GC10-DET 数据集中分布情况如图 6 所示,从目标框的大小可知绝大多数缺陷为中小型缺陷,尺度差异大。从缺陷位置的分布可以看出处在图片中轴线位置的缺陷较多,从缺陷的宽高来看,小型缺陷最多,长条形尺寸的缺陷较多。

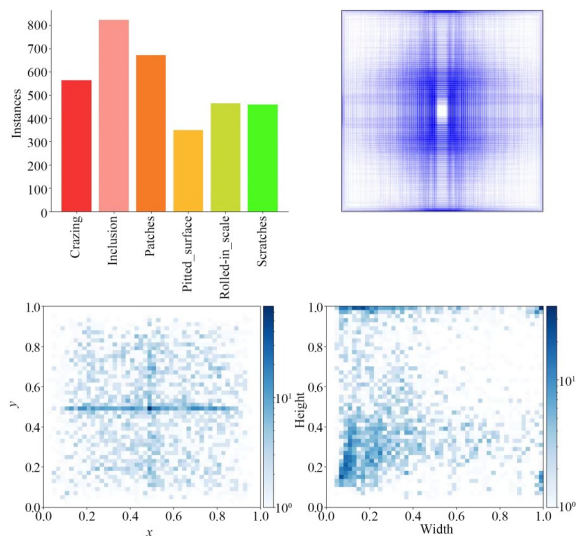


图 5 NEU 数据分布

Fig. 5 Data distribution of NEU

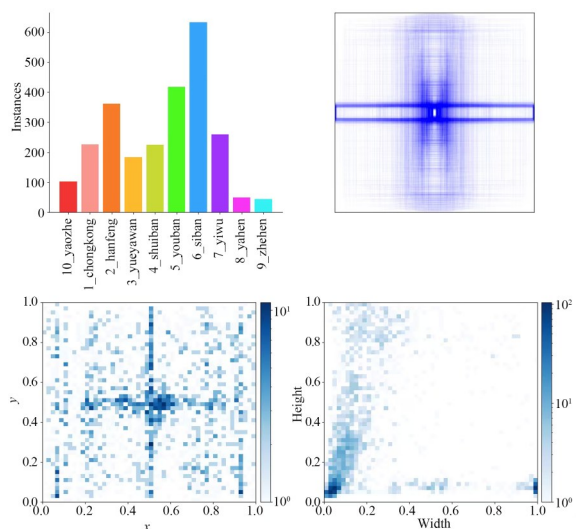


图 6 GC10-DET 数据分布

Fig. 6 Data distribution of GC10-DET

3.3 评价指标

本文采用准确度(Precision, P)、召回率(Recall, R)、平均精度均值(mean Average Precision, mAP)、每秒千兆浮点运算数(GFLOPs)、参数量(Params)和每秒帧数(FPS)等指标来评估模型效果。Precision 与 Recall 分别表示模型预测结果的准确性与完整性, TP, FP, TN 和 FN 的含义如表 1 所示,计算公式如下:

$$P = \frac{TP}{TP + FP}, \quad (16)$$

$$R = \frac{TP}{TP + FN}. \quad (17)$$

表 1 混淆矩阵
Tab. 1 Confusion matrix

		Prediction	
		Positive	Negative
Actual	Positive	TP	FN
	Negative	FP	TN

mAP 衡量了模型在不同目标类别上的平均精度。平均精度 (Average Precision, AP) 是 Precision-Recall 曲线下的面积。mAP 通过对所有类别的 AP 值进行平均得到。mAP@0.5 是在 IoU 阈值设为 0.5 情况下计算的 mAP。计算公式为:

$$AP = \int_0^1 P(R) dR, \quad (18)$$

$$mAP = \frac{\sum_{i=0}^N AP_i}{N}, \quad (19)$$

$$mAP@0.5 = \frac{\sum_{i=0}^N AP@0.5_i}{N}. \quad (20)$$

IoU 通过比较预测框 (Predicted Bounding Box) 与真值框 (Ground-Truth Bounding Box) 的重叠区域、间距能够评估目标检测模型的效果, 是两个框重叠区域的面积除以两个框并集区域的面积, 当 IoU 的值高于 0.5 时认为该模型的表现较为理想。

$$IoU = \frac{GT \cap Pred}{GT \cup Pred}. \quad (21)$$

GFLOPs 表示模型在前向传播过程中所需执行的浮点运算次数反映模型的计算开销; Params 表示模型训练和存储所需的参数总量, 是用来衡量模型复杂度与存储需求的重要依据; FPS 是衡量模型检测速度的指标, 表示模型每秒可处理的图像帧数。

3.4 模块先进性验证实验

为验证 DFE 模块在训练过程中是否能够学习到与缺陷检测任务相关的频率响应, 本文统计了模型初始化、训练 10 epoch 及训练 50 epoch 三个阶段的频谱可视化结果如图 7 所示, 该图展示了不同训练阶段的频谱变化情况。后续数据分析表明模型初始化阶段, 由于通道注意力和复数权重尚未学习, 所有频率分量均受到较强抑制, 表现为对频谱的无差别压制。随着训练进行, 高频能量下降幅度逐渐减小, 至第 50 个 epoch 时,

HFR 仅下降 12.4%, 与此同时, 高低频能量比 (High/Low Ratio) 由初始阶段下降 51% 逐渐提升至增加 83%。

该结果说明, DFE 并非简单提高整体高频能量, 而是在训练过程中逐步学习输入特征对应的频率分布规律, 通过频率感知注意力和复数频谱调制, 对不同频率成分进行选择性地加权。其中, 与缺陷边缘相关的中高频信息得到保留和增强, 而随机噪声对应的极高频响应受到抑制, 实现符合钢材表面缺陷检测需求的频率表达。后续进一步统计 20 个测试样本发现, 所有样本均表现出一致的频谱演化趋势, 说明该频率调制机制具有较好的稳定性和泛化能力。

为了验证本文提出的双重频域增强策略 (DFE) 的有效性, 本文分别选取高通滤波 (HPF)、FFT Enhancement 以及频域通道建模方法 FcaNet 作为对比模块。由表 2 可知, 仅采用固定高通滤波时, 模型性能最差。这说明简单抑制低频信息可能同时削弱部分对目标检测有贡献的结构特征, 难以兼顾背景抑制与缺陷信息保留。FFT Enhancement 能够对整体频域特征进行增强, 但 mAP 仅提升 0.1%, 说明通用频域增强方式对钢材缺陷检测的改善有限。FcaNet 利用频域信息对通道响应进行建模, 使 mAP 提升至 76.0%, 验证了频域建模对于检测任务具有一定作用。

相比上述方法, 本文提出的 DFE 取得 76.4% 的最高 mAP, 同时 Recall 达到 75.1%, 均优于其他频域增强策略。分析认为, DFE 采用前置高通滤波与后置频谱门控相结合的双阶段设计, 在抑制轧制纹理等背景干扰的同时, 对缺陷边缘频率响应进行自适应调制, 使频域信息更加符合钢材表面缺陷检测任务的特点。

为了进一步验证 DKG 模块的有效性, 本文分别采用 Add、Concat、CARAFE 以及 DySample 等具有代表性的特征融合方式进行对比。由表 3 可以看出, 不同融合方式均能够完成跨层特征交互, 但检测性能存在明显差异。传统 Add 与 Concat 属于固定融合方式, 缺乏针对空间位置的自适应建模能力, 因此性能提升有限。CARAFE 通过内容感知核实现特征重组, 但在当前 Transformer 检测框架下并未取得更优结果, 说明仅依

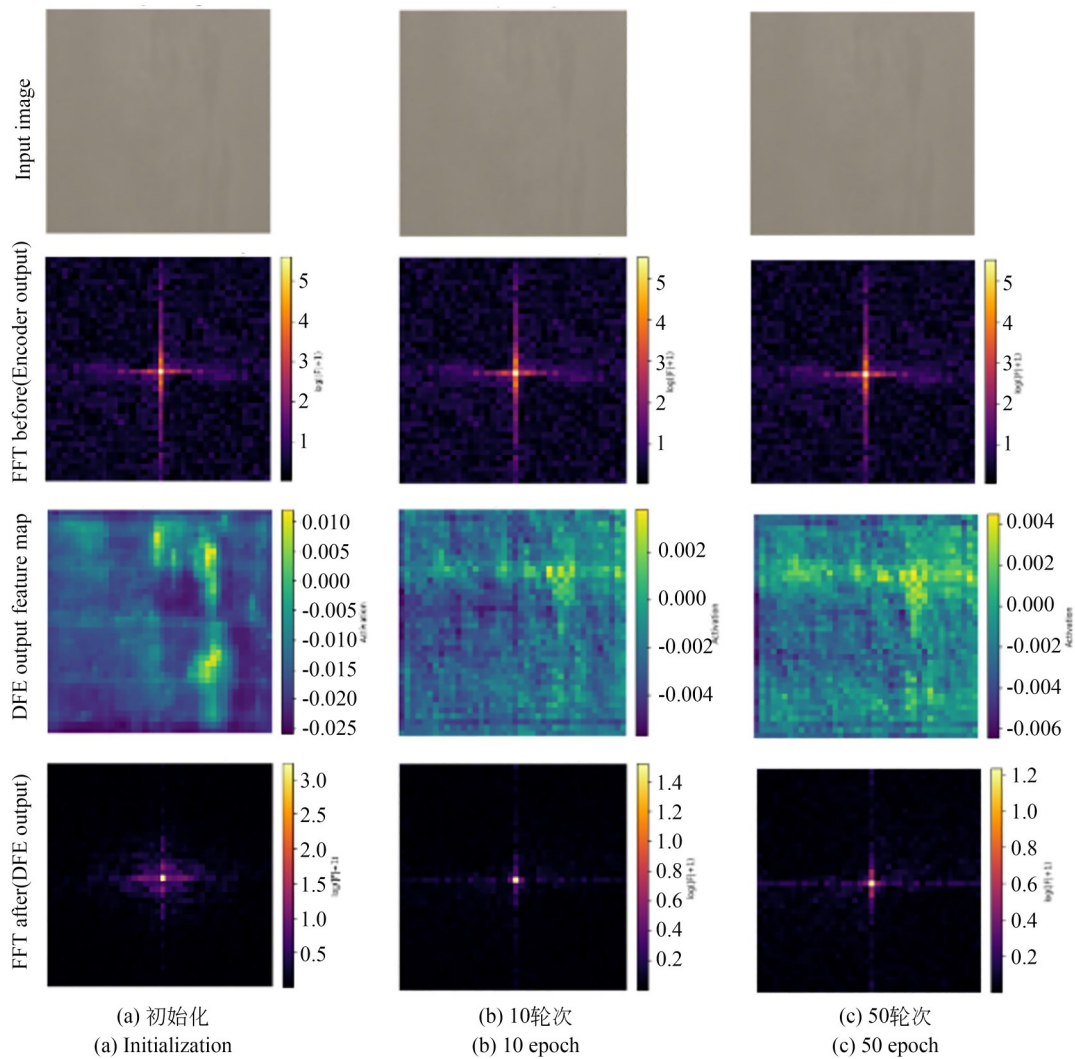


图 7 频谱结果可视化对比

Fig. 7 Visual comparison of spectrum results

表 2 NEU 数据集频域增强方法对比

Tab. 2 Comparison of frequency-domain enhancement methods on the NEU dataset

Method	<i>P</i>	<i>R</i>	mAP@ 0.5/%	FLOPs	Param s/M	FPS
HPF	74.3	72.0	73.6	36.45	32.24	24
FFT	75.1	73.5	75.0	36.34	32.24	22.5
FcaNet	79.3	71.7	76.0	36.45	32.27	20.9
Ours	75.8	75.1	76.4	36.34	32.31	25.4

赖局部重组难以充分利用深层语义信息。DySample 采用动态采样完成特征对齐,但对于尺度变化较大的钢材缺陷仍存在一定局限。

DFE 模块和 DKG 模块在 GC10-DET 数据集上的实验结果分别如表 4~表 5 所示,实验结果

表 3 在 NEU 数据集跨尺度特征融合方法对比

Tab. 3 Comparison of cross-scale feature fusion methods on the NEU dataset

Method	<i>P</i>	<i>R</i>	mAP@ 0.5/%	Params/ M	FPS
Add	79.8	73.3	76.0	30.82	35.0
Concat	79.9	73.3	75.9	31.87	41.7
CARAFE	80.0	71.7	74.6	32.66	38.1
DySample	75.9	71.5	73.6	31.90	33.9
Ours	80.2	73.8	76.8	31.90	40.7

进一步验证了上述结论。相比其他频域增强方法均出现不同程度性能下降,DFE 仍取得 74.4% 的最高检测精度,DKG 模块在 NEU-DET 与 GC10-DET 数据集上均取得最高检测精

度,同时模型参数量和推理速度变化较小,说明该模块能够在保持计算效率的基础上提升跨尺度特征融合效果,充分说明本文所提出的方法在复杂工业背景下具有较好的稳定性和跨数据集泛化能力。

表 4 在 GC10-DET 数据集频域增强方法对比
Tab. 4 Comparison of frequency-domain enhancement methods on the GC10-DET dataset

Method	<i>P</i>	<i>R</i>	mAP@ 0.5/%	FLOPs	Params /M	FPS
Baseline	80.9	69.7	73.4	36.34	32.17	45.7
HPF	78.9	72.0	73.1	36.45	32.24	51.4
FFT	81.5	68.7	72.5	36.35	32.24	47.2
FcaNet	79.5	70.6	72.3	36.45	32.27	44.3
Ours	85.7	70.8	74.4	36.34	32.31	46.8

表 5 在 GC10-DET 数据集跨尺度特征融合方法对比
Tab. 5 Comparison of cross-scale feature fusion methods on the NEU dataset

Method	<i>P</i>	<i>R</i>	mAP@ 0.5/%	Params/ M	FPS
Add	81.1	70.2	72.7	30.84	54.8
Concat	82.4	68.6	73.0	31.89	54.5
CARAFE	77.9	69.6	70.8	32.67	51.4
DySample	80.9	68.7	72.8	31.91	49.5
Ours	79.6	71.4	73.9	31.92	54.5

3.5 消融实验

为验证本文提出的动态核引导模块(DKG)与双重频域增强模块(DFE,包括前置高通滤波模块与后置频谱门控模块)的有效性,本文在 NEU-DET 数据集上进行了消融实验。所有实验均采用相同的训练参数与数据划分方式,实验结果如表 6 所示。

从实验结果可以看出,单独引入 DKG 模块后,模型的 mAP@0.5 由基线模型的 72.4% 提升至 75.8%。说明基于动态卷积核的跨尺度特征引导方式能够在一定程度上缓解传统特征融合过程中浅层细节信息被深层语义覆盖的情况。与此同时,模型 Precision 提升至 76.7%,表明该模块对于复杂背景下缺陷区域的定位具有一定

表 6 消融实验结果 Tab. 6 Ablation experiments results					
Method			mAP@	<i>P</i>	<i>R</i>
DKG	Pre-Freq	Post-Freq	0.5/%		
			72.4	73.6	68.8
✓			75.8	76.7	68.7
	✓		72.7	68.3	68.9
		✓	71.1	68.6	70.5
✓	✓		75.5	68.7	71.5
✓		✓	75.1	69.5	73.2
	✓	✓	76.2	73.1	69.4
✓	✓	✓	80.8	76.3	72.9

的积极作用。
单独引入前置高通滤波模块后,模型 mAP@0.5 提升至 72.7%。该结果说明频域高通滤波能够减少部分低频背景纹理对缺陷特征提取的干扰,使边缘与纹理信息得到增强。相比之下,仅使用后置频谱门控模块时,模型 Recall 指标有所提升,一定程度上说明了动态频谱调制对于低对比度缺陷特征保留具有一定帮助。
当同时引入前置与后置频域模块后,模型 mAP@0.5 可以达到 76.2%,进一步说明双重频域增强模块在复杂工业背景下具备较好的互补性。其中,前置频率模块主要用于抑制背景噪声干扰,而后置频谱门控则进一步强化了特征重组后的有效频率响应。

在同时引入 DKG 与 DFE 后,模型最终取得 80.8% 的 mAP@0.5,相较于基线模型提升了 8.4%。结果表明,DFE 模块通过频域增强削弱了复杂背景中的低频干扰,强化了缺陷边缘的响应,从而改善了浅层特征中的局部结构质量,DKG 模块进一步利用深层语义信息对局部区域进行动态引导融合,使增强后的边缘信息能够在跨尺度特征交互过程中得到充分保留。两类模块分别作用于特征增强与特征融合阶段,具有一定的互补性。

3.6 对比实验

消融实验已经证实本文所提出模块的有效性,为进一步验证本方法在钢材表面缺陷检测任务中的性能,本文分别在 NEU-DET 与 GC10-DET 数据集上进行了对比实验。

为了全面评估本文提出的算法在复杂工业场景下的性能优势,首先在 NEU-DET 数据集上将本模型与多种主流和前沿的目标检测算法进行了定量对比实验。对比模型涵盖了经典的双阶段网络 (Faster R-CNN)^[22]、单阶段网络 (SSD^[23], YOLOv8n^[24], YOLOv11^[25]), 以及近期针对复杂缺陷检测提出的多种先进算法 (SDF-YOLO^[26], VTG-YOLOv7-tiny^[27], UAL-YOLO^[28], GC-NET^[13]), RF-DETR 作为基准模型。详细对比结果如表 7 所示,表中标注来源数据为引用结果。

表 7 在 NEU 数据集对比实验结果

Tab. 7 Results of comparative experiments on NEU datasets

Models	<i>P</i>	<i>R</i>	mAP@ 0.5/%	FPS
Faster R-CNN ^[22]	56.9	72.8	69.3	19.9
SSD ^[23]	67.9	60.1	68.2	56.8
YOLOv8n ^[24]	73.3	68.1	67.6	63.2
YOLOv11 ^[25]	76.1	67.3	74.1	66.3
SDF-YOLO ^[26]	76.8	72.6	76.9	—
VTG-YOLOv7 ^[27]	68.5	74.5	74.4	87
UAL-YOLO ^[28]	—	—	79.5	84
GC-NET ^[13]	—	—	77.1	40
RF-DETR	73.6	68.8	72.4	64.2
Ours	76.3	73.0	80.8	66.8

实验结果表明,本文方法在 NEU-DET 数据集上取得 80.8% 的 mAP@0.5,高于参与对比的其他检测模型,同时保持了较高的推理速度。

与基准模型 RF-DETR 相比,本文方法 mAP@0.5 提升了 8.4%,精确率与召回率分别达到 76.3% 与 73.0%。结果表明,引入 DFE 与 DKG 后,模型在复杂背景下的特征提取与跨尺度特征表达能力得到改善。相比 Faster R-CNN、SSD 以及 YOLOv8n 等传统检测模型,本文方法在低对比度与多尺度缺陷场景下具有更好的检测性能。与 YOLOv11 相比,本文方法在保持相近精确率的同时取得了更高的召回率,说明其对复杂缺陷场景中的漏检问题具有一定改善作用。

与近年来针对工业缺陷检测提出的先进模

型相比,本文方法同样取得了较优结果。其中,UAL-YOLO 在 NEU-DET 数据集上的 mAP@0.5 达到 79.5%,而本文方法进一步提升至 80.8%。

从机制角度分析,DFE 模块通过频域增强削弱复杂轧制纹理对缺陷边缘响应的干扰,有助于提升低对比度缺陷的特征可分性;DKG 模块则利用深层语义信息对浅层局部区域进行动态引导融合,在一定程度上改善了多尺度特征交互过程中细节信息易弱化的问题。两类模块的协同作用提升了模型对复杂工业场景中多尺度缺陷的检测能力。

双重频域增强模块 (DFE) 通过频域响应调制减弱了轧制纹理等背景信息对钢材缺陷边缘特征的干扰,能够在一定程度上增强裂纹与划痕等弱边缘缺陷的可分性。从实验结果来看,该策略能在提升检测精度的同时,还能够保持较稳定的 Precision 指标。面对 NEU-DET 数据集中存在大量低对比度缺陷与尺度差异较大的目标,复杂轧制纹理背景容易导致微弱边缘信息在特征融合过程中被削弱。当前 YOLO 系列及级联检测方法多采用固定的跨尺度融合策略,对局部结构信息的自适应建模能力有限,在复杂背景下容易出现细节特征丢失的问题。针对上述问题,本文引入动态核引导模块 (DKG),利用深层语义特征生成空间自适应核权重,对浅层高分辨率特征进行局部动态重组,从而增强了跨尺度特征交互过程中对边缘纹理与局部结构信息的保留能力。结合实验结果可以发现,该策略在复杂背景场景下对多尺度缺陷检测具有一定改善作用。

表 8 进一步给出了 NEU 数据集上不同类别缺陷的检测结果。相比基线模型 RF-DETR,除夹杂物 (Inclusion, In) 类别出现小幅度下降外,其余类别的检测精度均有所提升。其中,裂痕 (Cracking, Cr) 与氧化皮压入 (Rolled-in Scale, Rs) 两类低对比度缺陷的 mAP@0.5 分别提高 11.1% 与 10.8%。这一结果说明,频域增强策略能够在一定程度上改善弱边缘缺陷的特征表达,而动态核引导机制对于局部结构信息的保留具有一定积极作用。

为进一步验证模型在复杂工业场景中的泛化性能,本文还在 GC10-DET 数据集上开展对比

表 8 NEU 数据集上各类缺陷对比实验结果
Tab. 8 Comparative experimental results of various types of defects on NEU datasets

Models	mAP@0.5/%					
	Cr	In	Pa	Ps	Rs	Sc
YOLOv11	36.0	79.5	92.1	89.2	53.7	96.0
RF-DETR	37.8	81	90	82	58.2	97
Ours	48.9	79.6	94.2	85.1	69.0	99.0

实验,结果如表 9 所示。

GC10-DET 数据集具有较强背景干扰、明显尺度变化以及类别不均衡等特点,对模型特征建模能力具有较高要求。从实验结果可以看出,传统双阶段与单阶段检测模型在该数据集上的性能均出现不同程度下降。这表明常规的卷积网络在面对极其复杂的非结构化工业背景时,可能导致复杂背景下的特征区分能力下降。采用全局注意力机制的 GC-NET 也仅取得 63.3% 的结果。

RF-DETR 基线模型在该数据集上取得 71.3% 的 mAP@0.5,体现出 Transformer 主干网络较强的全局特征建模能力。在此基础上,本文方法进一步将 mAP@0.5 提升至 73.4%,同时精确率与召回率分别提升至 75.3% 与 69.5%。结果表明,所提出方法在复杂工业场景下仍能够保持较好的检测性能,具备较强的稳定性与特征解耦能力。

表 9 在 GC10-DET 数据集上的对比实验结果
Tab. 9 Results of comparative experiments on GC10-DET datasets

Models	P	R	mAP@0.5
Faster R-CNN	55.3	66.9	61.5
SSD	61.0	55.1	60.9
YOLOv8n	73.3	68.1	68.4
YOLOv11	68.5	63.8	65.9
文献 ^[8]	76.8	70.4	70.9
GC-NET ^[29]	—	—	63.3
RF-DETR	69.9	65.0	71.3
Ours	75.3	69.5	73.4

表 10 给出了不同类别缺陷的检测结果。对于“水斑”等低对比度缺陷,本文方法相较 RF-DETR 提升较为明显,其检测精度由 72.7% 提升至 78.0%,说明 DFE 模块在抑制复杂背景低频干扰方面具有一定作用。

对于“异物”等小尺度缺陷,本文方法同样取得一定性能提升。这表明 DKG 模块通过深层语义信息引导浅层局部特征重组,有助于增强模型对小目标与局部细节信息的表征能力。

针对 GC10-DET 数据集中折痕(Creases)类别检测性能下降的问题,本文进一步对 DFE 模块的频域响应进行了分析。随机选取 20 幅折痕样本,统计 Encoder 输入特征与 DFE 输出特征在不同频率区间的能量分布变化。结果表明,DFE 模块对 DC 分量及极低频成分(归一化半径小于 10%)进行了较明显的抑制,其中 DC 能量下降 67.7%,极低频能量下降 36.4%,说明该模块能够有效抑制由整体亮度基线和缓慢光照变化带来的背景干扰。与此同时,折痕结构所对应的低频区域(10%~20% 归一化频率)能量仅下降 3.2%,表明其主体结构信息得到了较好保留;而中低频至高频区域的能量分别提升 23.7%~36.9%,说明 DFE 进一步增强了缺陷边缘及局部纹理特征,提高了特征表达能力。原始特征通过恒等映射直接保留,频域增强分支仅对补充信息进行调制。因此,即使频谱门控对 DC 及极低频成分进行了选择性抑制,原始空间域结构信息仍能够完整保留,从而避免了传统高通滤波容易造成的结构信息丢失问题。频谱分析结果表明,DFE 模块在训练过程中逐渐由初始的全局频率抑制学习为针对不同频率成分的选择性调制,这说明所提出的频谱门控机制能够根据检测任务自适应学习频率响应,而非采用固定的高通滤波策略,进一步验证了本文频域增强设计的合理性。折痕类别检测性能出现一定下降,说明当前频率重分配策略对于依赖低频形状信息的缺陷类别仍存在一定高频偏置,后续工作将进一步研究基于缺陷类型的自适应频率调制策略,以兼顾低频结构缺陷与高频纹理缺陷的检测性能。

图 8 展示了部分典型困难样本的可视化检测结果。相比原始 RF-DETR,本文方法在复杂背

表 10 GC10-DET数据集各类别 mAP 结果

Tab. 10 Results of mAP of various types on GC10-DET datasets

Models	mAP@0.5/%									
	冲孔	焊缝	月牙弯	水斑	油斑	丝斑	异物	压痕	折痕	腰折
YOLOv11	93.4	94.3	96.5	71	70.7	63.6	49.2	40.6	53.2	70.6
RF-DETR	94.9	94.8	98.0	72.7	71.8	62.7	53.5	50.2	51.8	83.4
Ours	95.3	95.3	98.3	78.0	73.4	63.6	54.9	55.2	41.0	83.7

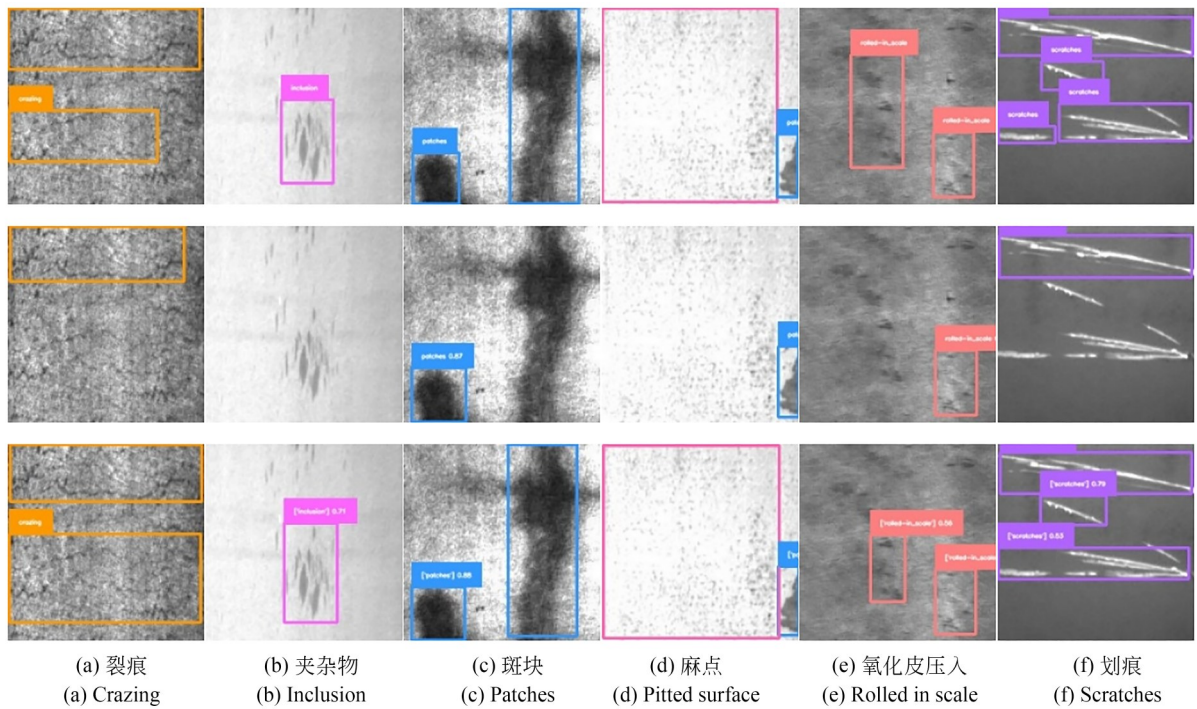


图 8 检测结果可视化对比

Fig. 8 Visual comparison of test result

景干扰与尺度变化较大的场景下,对部分弱边缘缺陷与小尺度缺陷具有更完整的检测结果,漏检现象有所减少。第一行是数据集各类别的真实标注框(Ground truth)作为基准对照,第二行则是 RF-DETR 原始模型的检测效果,可以看到在面对复杂场景(背景噪声干扰强)和缺陷尺度差异大的情况下,均存在漏检现象,第三行则是本文改进后的检测效果,在所有类别上都可以直观观察到明显的改进效果。

4 结 论

针对复杂工业场景下钢材表面缺陷检测中存在的低对比度缺陷响应弱、跨尺度特征融合过

程中局部细节易丢失等问题,本文提出了一种基于 RF-DETR 的改进型检测框架 DFD-DETR。

本文构建了双重频域增强模块(DFE),通过前置高通滤波与后置频谱门控模块对特征频域响应进行调制,在一定程度上抑制复杂背景中的低频干扰,并增强微弱缺陷边缘信息。实验结果表明,该策略对裂纹、水斑等低对比度缺陷具有较好的检测增益。

同时,针对传统跨尺度融合过程中浅层局部结构信息易弱化的问题,设计了动态核引导模块(DKG)。该模块利用深层语义信息生成空间自适应动态卷积核,对浅层高分辨率特征进行局部引导重组,从而增强模型对多尺度缺陷的结构表征能力。

在NEU-DET与GC10-DET两个数据集上的实验结果表明,本文方法分别取得80.8%与73.4%的mAP@0.5,相较基线模型均取得一定提升。此外,针对GC10-DET数据集中“折痕”类别检测性能下降的问题,进一步分析了频域高通滤波对低频结构信息可能带来的影响,为后续频域增强策略的优化提供了参考。

总体来看,本文提出的频域增强与动态核引导协同机制在复杂背景钢材表面缺陷检测任务中表现出较好的适应性与检测性能,同时兼顾了一定的推理效率。尽管本文方法在两个公开数据集上取得了较好的检测效果,但对于具有明显低频结构特征的大尺度缺陷仍存在一定

性能权衡。后续工作将进一步研究自适应频率选择机制及更加轻量化的动态特征交互策略,以提升模型在不同类型工业缺陷检测任务中的适应能力。

作者贡献声明:

敬天池:测量方法的提出,论文构思和撰写;

孙晓明:论文审核与编辑写作;

郝亮:实验的设计;

徐凯歌:实验数据整理和分析;

陈言:实验数据整理和分析;

韩永琦:论文审核与编辑写作;

刘洋:论文审核与编辑写作。

参考文献:

- [1] 唐东林,杨洲,程衡,等. 浅层卷积神经网络融合Transformer的金属缺陷图像识别方法[J]. 中国机械工程, 2022, 33(19): 2298-2305, 2316.
TANG D L, YANG ZH, CHENG H, *et al.* Metal defect image recognition method based on shallow CNN fusion transformer[J]. *China Mechanical Engineering*, 2022, 33 (19): 2298-2305, 2316. (in Chinese)
- [2] 邓志鹏,何施茗,杨根,等. 基于深度学习的纹理表面缺陷检测方法综述[J]. 计算机集成制造系统, 2025, 31(3): 721-745.
DENG ZH P, HE SH M, YANG G, *et al.* Survey of texture surface defect detection method based on deep learning[J]. *Computer Integrated Manufacturing Systems*, 2025, 31(3): 721-745. (in Chinese)
- [3] 李少波,杨静,王铮,等. 缺陷检测技术的发展与应用研究综述[J]. 自动化学报, 2020, 46(11): 2319-2336.
LI SH B, YANG J, WANG ZH, *et al.* Review of development and application of defect detection technology [J]. *Acta Automatica Sinica*, 2020, 46 (11): 2319-2336. (in Chinese)
- [4] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection[C]. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016: 779-788.
- [5] CARION N, MASSA F, SYNNAEVE G, *et al.* End-to-end object detection with transformers[EB/OL]. 2020: *arXiv*: 2005.12872. <https://arxiv.org/abs/2005.12872>.
- [6] ZHU X Z, SU W J, LU L W, *et al.* Deformable DETR: deformable transformers for end-to-end object detection [EB/OL]. 2020: *arXiv*: 2010.04159. <https://arxiv.org/abs/2010.04159>.
- [7] LIN T Y, DOLLÁR P, GIRSHICK R, *et al.* Feature pyramid networks for object detection [C]. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. July 21-26, 2017, Honolulu, HI, USA. IEEE, 2017: 936-944.
- [8] LIU S, QI L, QIN H F, *et al.* Path aggregation network for instance segmentation[C]. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*. June 18-23, 2018, Salt Lake City, UT, USA. IEEE, 2018: 8759-8768.
- [9] XIAO H Y, ZHANG W Y, ZUO L, *et al.* WFF-Net: Trainable weight feature fusion convolutional neural networks for surface defect detection[J]. *Advanced Engineering Informatics*, 2025, 64: 103073.
- [10] 郭伟,孟巧巧,金海波,等. 多尺度交互与动态协同的钢表面缺陷检测[J]. 计算机工程, 2025: 1-14.
GUO W, MENG Q Q, JIN H B, *et al.* Multi-scale Interaction and dynamic collaboration for steel surface defect detection [J]. *Computer Engineering*, 2025: 1-14. (in Chinese)
- [11] 张晓榕,黄卫阳,吕文丹,等. 基于GAB-YOLO

- 的钢材表面缺陷检测方法[J]. 计算机测量与控制, 2026: 1-10.
- ZHANG X M, HUANG W Y, LYU W D, *et al.* Steel surface defect detection method based on GAB-YOLO[J]. *Computer Measurement & Control*, 2026: 1-10. (in Chinese)
- [12] 彭有有, 李维刚, 胡佳琪. 基于DMI-YOLO算法的带钢表面缺陷检测[J]. 钢铁研究学报, 2026, 38(3): 373-383.
- PENG Y Y, LI W G, HU J Q. Strip steel surface defect detection based on DMI-YOLO algorithm[J]. *Journal of Iron and Steel Research*, 2026, 38(3): 373-383. (in Chinese)
- [13] LIU G H, CHU M X, GONG R F, *et al.* Global attention module and cascade fusion network for steel surface defect detection[J]. *Pattern Recognition*, 2025, 158: 110979.
- [14] LI J, LUO W, *et al.* PCBA pin solder defect detection method based on improved RT-DETR[J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(10): 1532-1544.
- [15] ROBINSON I, ROBICHEAUX P, POPOV M, *et al.* RF-DETR: neural architecture search for real-time detection transformers [EB/OL]. 2025: *arXiv*: 2511.09554. <https://arxiv.org/abs/2511.09554>.
- [16] PEI X, HUANG Y D, SU W J, *et al.* FFTFormer: a spatial-frequency noise aware CNN-Transformer for low light image enhancement [J]. *Knowledge-Based Systems*, 2025, 314: 113055.
- [17] LIU S L, ZENG Z Y, REN T H, *et al.* Grounding DINO: marrying DINO with grounded pre-training for open-set object detection[C]. *Computer Vision-ECCV 2024*. Cham: Springer, 2025: 38-55.
- [18] CHEN Y P, DAI X Y, LIU M C, *et al.* Dynamic convolution: attention over convolution kernels [C]. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. June 13-19, 2020. Seattle, WA, USA. IEEE, 2020: 11027-11036.
- [19] WANG J Q, CHEN K, XU R, *et al.* CARAFE: content-aware ReAssembly of FEatures[C]. 2019 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 27-November 2, 2019. Seoul, Korea. IEEE, 2019: 3007-3016.
- [20] 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.
- [21] LV X M, DUAN F J, JIANG J J, *et al.* Deep metallic surface defect detection: the new benchmark and detection network[J]. *Sensors*, 2020, 20(6): 1562.
- [22] REN Q R, GENG J H, LI J Y. Slighter Faster R-CNN for real-time detection of steel strip surface defects [C]. 2018 *Chinese Automation Congress (CAC)*. November 30-December 2, 2018, Xi'an, China. IEEE, 2019: 2173-2178.
- [23] LIU W, ANGUELOV D, ERHAN D, *et al.* SSD: Single Shot MultiBox Detector[M]. Computer Vision-ECCV 2016. Cham: Springer International Publishing, 2016: 21-37.
- [24] VARGHESE R, M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness [C]. 2024 *International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS)*. April 18-19, 2024, Chennai, India. IEEE, 2024: 1-6.
- [25] KHANAM R, HUSSAIN M. YOLOv11: an overview of the key architectural enhancements [EB/OL]. 2024: *arXiv*: 2410.17725. <https://arxiv.org/abs/2410.17725>.
- [26] WANG C L, XU K. SDF-YOLO: an efficient neural network for detecting surface defects in steel strips [J]. *Journal of Physics: Conference Series*, 2025, 2990(1): 012024.
- [27] 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[J]. 光学精密工程, 2024, 32(8): 1227-1240.
- LIANG L M, LONG P W, FENG Y, *et al.* Improving the lightweight VTG-YOLOv7-tiny for steel surface defect detection[J]. *Optics and Precision Engineering*, 2024, 32(8): 1227-1240. (in Chinese)
- [28] 张新君, 赵志闯. 基于UAL-YOLO的钢材表面缺陷检测算法[J]. 电子测量与仪器学报, 2026: 1-12.
- ZHANG X J, ZHAO ZH CH. Steel surface defect detection algorithm based on UAL-YOLO [J].

Journal of Electronic Measurement and Instrumentation, 2026: 1-12. (in Chinese)

[29] XIA Y K, LU Y, JIANG X H, *et al.* Enhanced

multiscale attentional feature fusion model for defect detection on steel surfaces[J]. *Pattern Recognition Letters*, 2025, 188(C): 15-21.

作者简介:



敬天池(2002-),男,黑龙江大庆人,硕士研究生,主要从事图像处理与机器视觉方面的研究。E-mail: 15663740563@163.com

通讯作者:



孙晓明(1983-),女,黑龙江海林人,博士,教授,博士生导师,2007年、2012年于哈尔滨工业大学分别获得硕士、博士学位,主要从事机器视觉、图像处理和信号特征提取等方面的研究。E-mail: sunxiaoming@hrbust.edu.cn