

# 基于DCD-YOLO的钢材表面缺陷检测算法

唐世龙, 徐正豪, 候华毅\*

武汉工程大学光电信息与能源工程学院、数理学院, 湖北 武汉

收稿日期: 2025年12月25日; 录用日期: 2026年1月22日; 发布日期: 2026年1月29日

## 摘要

本文针对钢材表面缺陷检测中多尺度特征适应性不足、上下文信息融合不充分、变形缺陷捕捉能力有限的问题, 提出一种基于YOLOv11n的改进模型DCD-YOLO。首先, 设计动态混合卷积模块(DCMB), 替换主干网络C3k2模块的Bottleneck部分, 通过动态卷积核权重机制自适应调整卷积核权重大小, 增强对多尺度缺陷的特征提取能力; 其次, 设计了上下文引导的空间特征重构金字塔网络(CGRFPN), 通过矩形自校准模块(RCM)与金字塔上下文提取模块(PCE), 加强模型对缺陷前背景的建模能力, 提升缺陷与复杂背景的区分度; 最后, 通过引入可变形注意力机制(DAttention)替换PSA模块中的固定注意力机制, 完成注意力的动态采样, 强化了对变形缺陷的适应性。实验结果表明, 改进后的模型在GC10-DET数据集上的mAP@0.5达到66.9%, 较原YOLOv11n提升3.3%。同时, 模型检测精度与召回率分别提升1.7%和2.8%, 有效解决了多尺度、背景抑制等检测难题, 满足了工业场景对准确性与召回率的要求。

## 关键词

钢材表面缺陷, YOLOv11n, 动态混合卷积, 上下文引导, 可变形注意力

# Steel Surface Defect Detection Algorithm Based on DCD-YOLO

Shilong Tang, Zhenghao Xu, Huayi Hou\*

School of Mathematics and Physics, School of Optical Information and Energy Engineering, Wuhan Institute of Technology, Wuhan Hubei

Received: December 25, 2025; accepted: January 22, 2026; published: January 29, 2026

## Abstract

This paper addresses the issues of insufficient multi-scale feature adaptability, inadequate contextual information fusion, and limited capability in capturing deformed defects in steel surface defect

\*通讯作者。

文章引用: 唐世龙, 徐正豪, 候华毅. 基于 DCD-YOLO 的钢材表面缺陷检测算法[J]. 计算机科学与应用, 2026, 16(2): 15-28. DOI: 10.12677/csa.2026.162035

detection by proposing an improved model, DCD-YOLO, based on YOLOv11n. First, a Dynamic Convolutional Mixed Block (DCMB) is designed to replace the Bottleneck part of the C3k2 module in the backbone network. Through a dynamic convolution kernel weight mechanism, it adaptively adjusts the convolution kernel weights, enhancing feature extraction capabilities for multi-scale defects. Second, a Context-Guided Spatial Feature Reconstruction Pyramid Network (CGRFPN) is designed. By using the Rectangular Self-calibration Module (RCM) and Pyramid Context Extraction Module (PCE), the model's ability to model defect foreground and background is strengthened, improving the distinction between defects and a complex background. Finally, by introducing a Deformable Attention mechanism (DAttention) to replace the fixed attention mechanism in the PSA module, dynamic sampling of attention is achieved, enhancing adaptability to deformed defects. Experimental results show that the improved model achieves a mAP@0.5 of 66.9% on the GC10-DET dataset, an increase of 3.3% compared to the original YOLOv11n. Meanwhile, the model's detection precision and recall increased by 1.7% and 2.8%, respectively, effectively addressing detection challenges such as multi-scale defects and background suppression, meeting industrial requirements for accuracy and recall.

## Keywords

Steel Surface Defects, YOLOv11n, Dynamic Convolutional Mixing, Context Guidance, Deformable Attention

Copyright © 2026 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

## 1. 引言

钢材是现代工业体系的核心基础材料，广泛应用于机械制造、船舶工程、航空航天、建筑结构等关键[1]领域，其表面质量直接决定了终端产品的安全性与可靠性。然而，在热轧、冷轧、焊接等生产加工环节中，钢表面易产生裂纹、斑点、划痕、氧化皮等缺陷[2] [3]。这些缺陷不仅会降低钢材的力学性能，还可能在后续使用中引发断裂、泄漏等严重安全事故，给企业造成巨大经济损失。因此，高效、准确的钢材表面缺陷检测技术是保障钢铁产业高质量发展的关键环节。

传统钢表面缺陷检测主要依赖人工目视和物理检测两类方法，但均存在明显不足。人工目视检测：依赖检测人员的经验判断，易受疲劳、情绪、光线等主观因素影响，漏检率与误检率高，且检测效率低，难以适应现代钢铁生产的高速流水线需求；物理检测方法：涡流探伤仅能检测表面或近表面的浅缺陷，无法识别深埋于氧化皮下方的裂纹；超声波检测对环境湿度敏感，需使用耦合剂，增加了清洁成本；磁粉检测则仅适用于铁磁性材料，且对微小缺陷检测能力有限。总体而言，传统方法易受人为因素与环境干扰，难以有效检测复杂背景中的微小缺陷，无法满足工业场景的要求。

近年来，基于深度学习的目标检测方法凭借其强大的自动特征学习能力，在钢表面缺陷检测中展现出显著优势。当前主流方法分为两类：二阶检测方法以 Faster R-CNN [4]、Mask R-CNN [5]为主，通过“候选区域生成 + 分类回归”两阶段流程，可有效提升小目标[6]检测精度，但因多阶段计算导致推理速度慢，难以满足工业实时性要求；一阶检测方法以 SSD [7]、YOLO [8]为主，将目标分类与边界框回归整合到单一网络中，直接输出目标类别与位置，推理速度快，且并行处理能力强，更适合钢铁生产的高速流水线场景。为进一步提升一阶模型的检测精度，现有研究主要围绕特征提取能力、特征融合机制、注意力机制三大方向展开优化：文献[9]提出了一种改进的缺陷检测模型 CFE-YOLOv8s，该模型有效地解决了钢材表面缺陷检测的问题，但在应用于其他缺陷检测任务时仍然存在局限性，文献[10]提出了一种

YOLOv8n-MDC 的钢材表面缺陷检测算法,对特征提取、目标定位和多尺度检测方面进行了优化,提高了钢材缺陷检测的精度。Zhang [11]等提出了一种 ESI-YOLOv8 的钢材表面缺陷检测方法,该方法融合了大分离卷积注意力机制、空间金字塔池化策略,增加了多尺度特征融合以增强感受野和特征表达能力。

上述研究在特征提取与特征融合方面取得一定进展,但在复杂场景目标检测中仍存在多尺度特征捕捉不充分、多特征融合低效冗余,自适应变形差等问题。为此,本文设计了三个改进策略,提出检测模型 DCD-YOLO,具体工作包括:

- (1) 针对传统卷积核形状固定尺寸单一的局限性提出一种动态混合卷积结构(DIMB),通过多尺度特征提取和自适应权重调整捕获更高质量的缺陷特征,以此提高检测的准确率和灵活性。
- (2) 针对传统特征金字塔网络抗背景抑制差的问题,设计上下文引导空间特征重构金字塔(CGRFPN)。该结构通过上下文空间引导和特征重建,提升了模型在复杂背景下目标的识别能力。
- (3) 针对 C2PSA 中固定注意力导致场景适应性差的问题,引入可变形注意力机制(DAttention)替换形成 C2DA 结构,加强了对缺陷的动态适应性。
- (4) 在开源 GC10-DET 数据集上验证检测算法的有效性。

上述改进后的优势主要在于传统 Inception 模块通过固定卷积核组合提取多尺度特征,但未考虑特征的自适应性同时参数量较大。本文提出的 DIMB 模块,融合了动态核权重机制与多尺度深度卷积,在轻量化基础上实现自适应特征提取。对于特征融合,现有 FPN/PAN 结构通过简单拼接融合多尺度特征,缺乏上下文引导。本文设计的 CGRFPN 引入矩形自校准模块(RCM)和金字塔上下文提取模块(PCE),通过空间上下文建模提升复杂背景下的缺陷识别能力。在注意力机制方面,可变形注意力(Deformable Attention)通过动态采样增强对变形目标的适应性,本文将其融入 C2PSA 模块,形成 C2DA 结构,进一步提升对不规则缺陷的感知能力。

## 2. YOLOv11 算法介绍

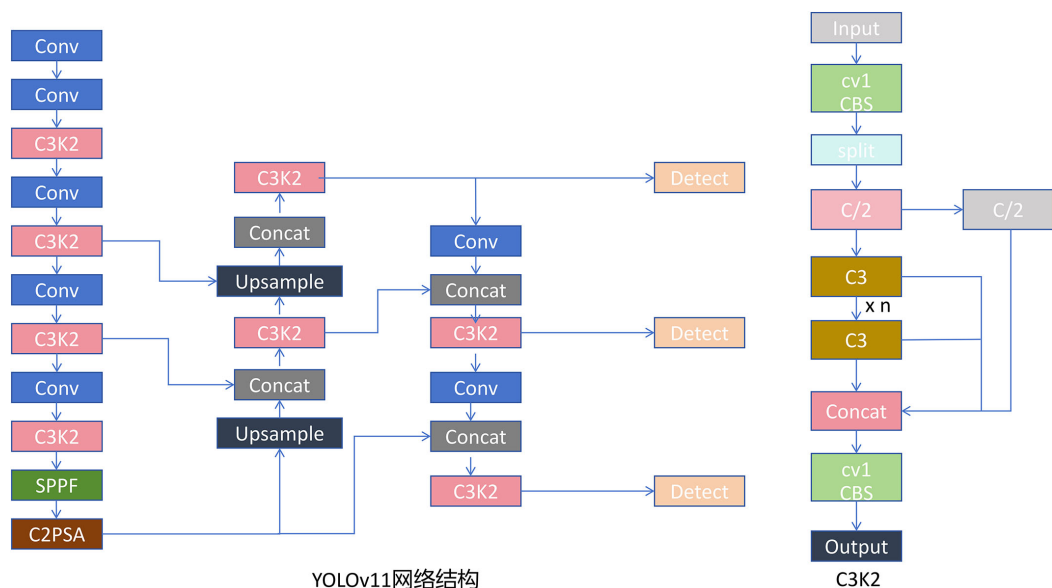


Figure 1. YOLOv11 network architecture diagram  
图 1. YOLOv11 网络结构图

YOLOv11 继承并发展了 YOLO 系列的优点,具有检测速度快和精度高的特点。其结构包括主干网络、颈部特征融合网络和检测头。主干网络负责特征提取,包含 C3k2 模块、卷积模块、SPPF [12]模块以及 C2PSA

模块。C3K2 模块是 YOLOv8 中 C2F 模块的演变，具有可调的 C3k 参数，优化模型处理特征的能力。模型颈部采用 FPN [13]和 PANet [14]结构，对主干网络提取的特征图进行融合生成特征金字塔，提升了模型对不同尺度缺陷的检测能力。检测头采用解耦结构和无锚框设计，将分类头与检测头分离，简化模型设计。通过这些步骤，YOLOv11 可以有效检测多种场景下的目标。YOLOv11 网络结构图如图 1 所示。

### 3. DCD-YOLO 钢材表面缺陷检测模型

尽管 YOLOv11 在通用目标检测任务中表现出色，但在钢材表面缺陷检测中，其面临复杂背景与缺陷尺度变化巨大的问题，这将大大降低钢材缺陷检测的效果，例如，细长弯曲的裂纹易被误判为背景纹理，从而限制了通用检测模型在该任务中的适用性。因此，本研究基于 YOLOv11 架构进行改进，提出了 DCD-YOLO 检测模型，旨在实现更精准的钢材表面缺陷检测。改进后的检测模型提出了动态混合卷积模块 (DIMB)，用于替换 C3k2 模块中的 Bottleneck 结构中的卷积模块，有助于动态提取多尺度特征；面对 YOLO11 的 FPN/PAN 结构通过简单的特征拼接进行多尺度融合，但缺乏有效的上下文信息提取的问题，提出了上下文引导空间特征重构金字塔网络(CGRFPN)，面对 C2PSA [15]模块采用固定注意力采样，无法适应钢材表面缺陷多变的特性，引入了可变形注意力机制形成(C2DA)结构，以适应钢材表面缺陷的变形特性。

#### 3.1. 动态混合卷积模块

钢材表面缺陷通常种类繁多，尺度差异巨大。C3k2 模块在特征提取过程中采用了固定的卷积核尺度和权重，未能考虑到缺陷多尺度和自适应问题，因此导致漏检和误检现象频发。为此，本文提出了 C3k2-DIMB 动态多尺度混合残差模块，该模块由动态多尺度融合器和卷积线性门控单元两部分组成。DIMB 模块首先将特征图输入到动态多尺度融合器中，以生成具有多尺度信息的特征图；随后将该特征图送入卷积线性门控单元中以提高特征非线性表达的能力，进而提高缺陷检测的能力与精度。其 DIMB 模块分解结构如图 2 所示。

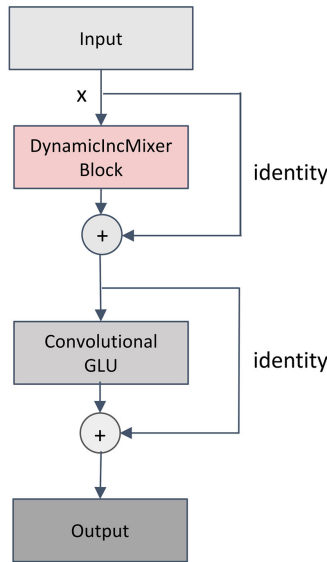


Figure 2. C3K2-DIMB module decomposition structure  
图 2. C3K2-DIMB 模块分解结构

##### 3.1.1. 动态多尺度融合器

动态多尺度融合器首先将输入特征均分成两份，让每份特征图各经过一个尺度不同的动态多尺度深

度卷积模块，然后将新生成的两份特征图通过拼接方式送入  $1 \times 1$  卷积核中，最终得到带有多尺度信息的自适应缺陷特征图。

这里的动态多尺度深度卷积模块是一种自研的创新卷积模块。它采用了尺度卷积核组合：包含方形核(square kernel，如  $3 \times 3$ )、水平条带核(band kernel，如  $1 \times 11$ )、垂直条带核(如  $11 \times 1$ )，以覆盖不同尺度的特征。此外，它还引入了引入动态核权重机制：通过特征自适应池化得到通道注意力特征，再通过  $1 \times 1$  卷积学习不同核的权重大小，以实现自适应调整卷积核权重。此机制将强化网络对多尺度缺陷的自适应调整，以实现更高的检测准确率与召回率。其网络结构如图 3 所示。

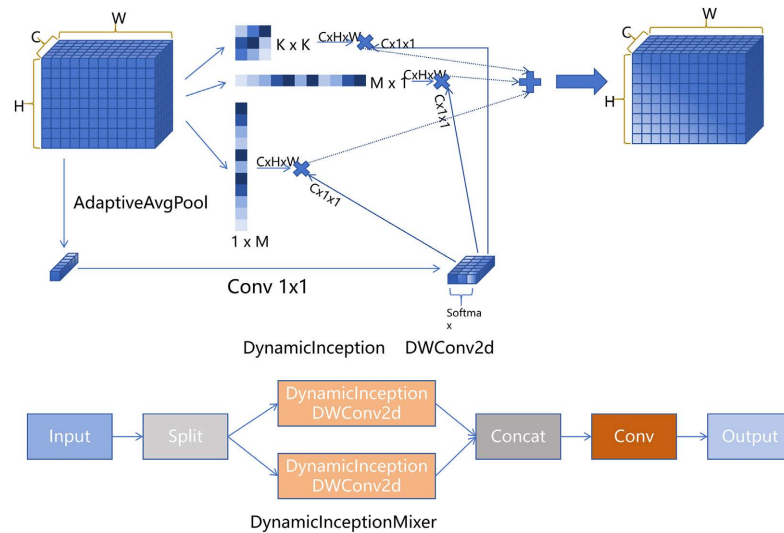


Figure 3. Dynamic multi-scale fusion structure  
图 3. 动态多尺度融合结构

### 3.1.2. 卷积线性门控单元

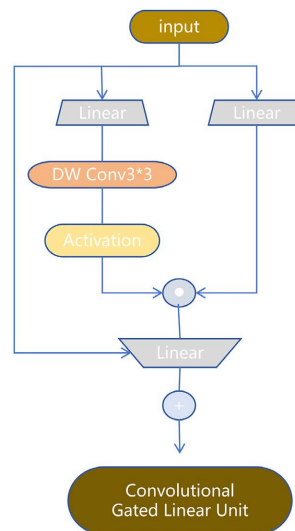


Figure 4. Convolutional linear gated unit  
图 4. 卷积线性门控单元

卷积线性门控单元(CGLU) [16]是一个用卷积层替代全连接层的轻量级 MLP [17]结构，核心是通过 1

$\times 1$  卷积保持特征图的空间维度, 同时实现通道维度的变换。它通过深度卷积引入局部空间信息, 增强了模型对空间特征的感知能力, 此外门控机制允许网络动态地选择传递信息, 提高了特征表达的灵活性。CGLU 的模块结构如图 4 所示。

卷积线性门控单元通过“卷积 + 归一化 + 激活 + 卷积”的结构, 增强了特征的非线性表达能力。相比单一卷积层, ConvMlp [18] 能更好地捕捉特征之间的高阶关联, 提升分类和检测的精度。卷积线性门控单元的实现可描述为:

$$x_1 = \sigma \left( N \left( \text{Conv2d}(x; W_1, b_1) \right) \right) \odot d \quad (1)$$

$$y = \text{Conv2d}(x_1; W_2, b_2) \quad (2)$$

$$N(z) = \gamma \odot \frac{z - \mu_z}{\sqrt{\sigma_z^2 + \varepsilon}} + \beta \quad (3)$$

式中  $\sigma$  表示激活函数, 用于引入非线性特性;  $N$  表示归一化层, 其详细公式为式(3)所示,  $\odot$  为点乘操作,  $d$  为注意缩放因子,  $\text{Conv2d}(x; W_1, b_1)$  和  $\text{Conv2d}(x_1; W_2, b_2)$  表示  $1 \times 1$  卷积操作, 其中  $x$  和  $x_1$  为输入,  $W_1$  和  $W_2$  为权重参数,  $b_1$  和  $b_2$  为偏置项。

### 3.2. 上下文引导空间特征重构金字塔网络

PAN 是 YOLOv11 的特征聚合核心模块, 通过 Top-Down (自顶向下) 和 Bottom-Up (自底向上) 路径融合 backbone 输出的多尺度特征, 其核心逻辑是特征传递与拼接。但 PAN 存在以下局限性: 空间上下文利用不足, 多尺度融合动态性差, 特征重构效率低, 通道不匹配等问题。为此本文创新的提出了一种 CGRFPN 网络结构。它主要由三部分组成, 分别是金字塔上下文提取, 动态插值融合和多特征融合三个子模块。其网络融合结构如图 5 所示。图中, RCM: Rectangular Self-Calibration Module 为矩形自校准模块; FBM: Feature Block Module 为多特征融合模块; DIF: Dynamic Interpolation Fusion 为动态插值融合模块。

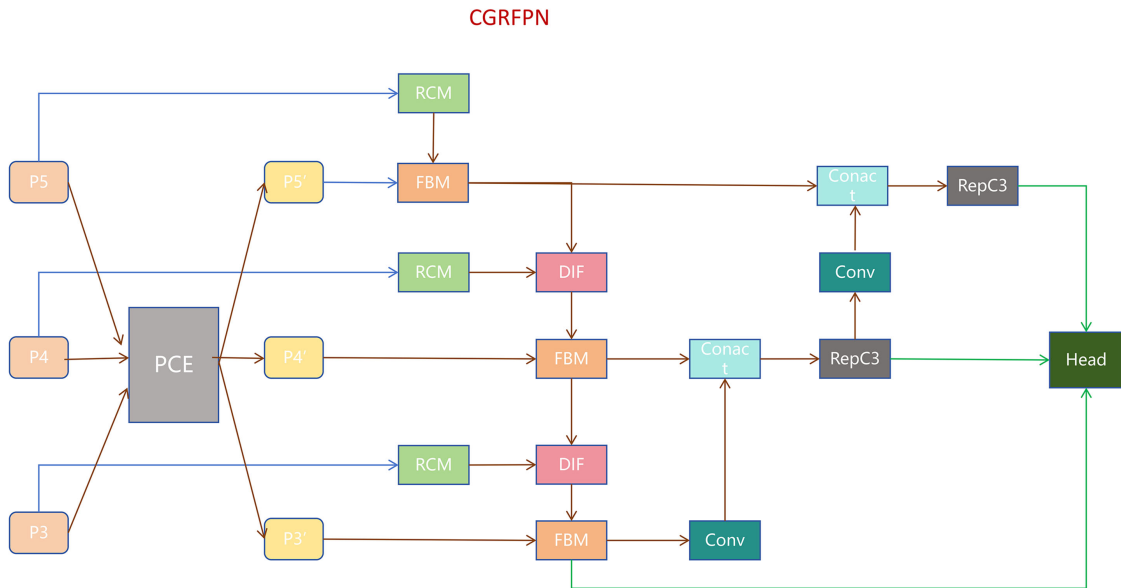


Figure 5. Contextual reconstruction guided structure

图 5. 上下文重构引导结构



CGRFPN 网络结构设计的核心依据为传统特征金字塔网络(FPN)仅通过简单的特征拼接进行多尺度融合,缺乏有效的上下文信息提取,导致模型难以区分缺陷与复杂背景。为解决这一问题,CGRFPN 引入了金字塔上下文提取模块(PCE),通过金字塔池化聚合(PPA)和矩形自校准模块(RCM)来捕获全局上下文信息,从而增强模型对缺陷前景的聚焦能力。在复杂背景下,缺陷往往与背景纹理具有相似性,但缺陷的局部特征与整体上下文关系具有独特性。通过金字塔池化聚合,可以捕获不同尺度的上下文信息,而矩形自校准模块则通过水平/垂直自适应池化,捕捉缺陷的长宽比特征,使模型能够更精确地定位缺陷区域。其次,CGRFPN 设计了动态插值融合模块,解决了不同尺度特征直接融合时,通道数差异导致信息传递不畅的问题。最后,多特征融合模块通过高尺度特征生成动态权重,自适应调整低尺度特征的融合强度。其理论依据是:在钢材表面缺陷检测中,不同尺度的特征对缺陷识别的重要性不同。高尺度特征包含更丰富的语义信息,可以指导低尺度特征的融合,使模型能够根据缺陷的类型和大小动态调整特征融合策略。特征的空间分辨率与语义信息的平衡,通过动态调整通道数,可以实现语义信息与空间细节的最优融合。

### 3.2.1. 金字塔上下文提取模块

金字塔上下文提取模块主要由金字塔池化聚合(PPA)和上下文处理模块组成。金字塔池化聚合通过自适应平均池化将特征堆叠到同一尺度,再通过矩形自校准注意力机制(RCA) [19]生成矩形注意力区域,加强对缺陷前景的聚焦能力。其 PPA 模块主要是为了融合多尺度的特征图,而矩形自校准模块(RCM),则先是通过 RCA 水平/垂直自适应池化捕捉全局上下文,再通过水平/垂直卷积优化注意力区域,最后经多层感知机(MLP)结构精炼特征,以增强特征的非线性表达能力。其各模块结构图如图 6 所示。

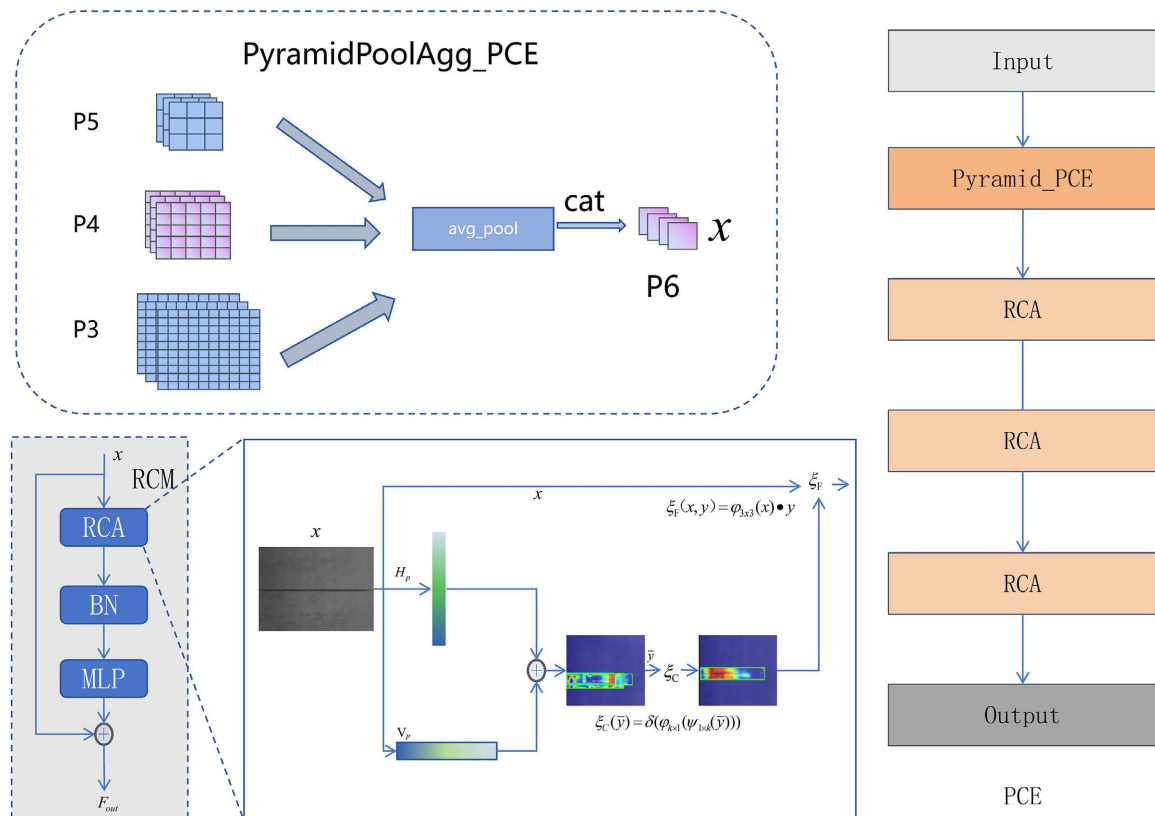


Figure 6. Pyramid context extraction module  
图 6. 金字塔上下文提取模块

### 3.2.2. 动态插值融合模块

动态插值融合主要是解决了不同尺度特征(如 P3 的  $C=256$ , P4 的  $C=512$ )直接融合时, 通道数差异导致信息传递不畅的问题。设计该模块, 先通过插值上采样将高层语义与低层特征做尺寸对齐, 然后通过  $1 \times 1$  卷积将高尺度特征的通道数调整为低尺度特征的通道数, 再通过加法进行融合, 进而解决通道不匹配和信息冗余的问题。其模块结构图如图 7 所示。

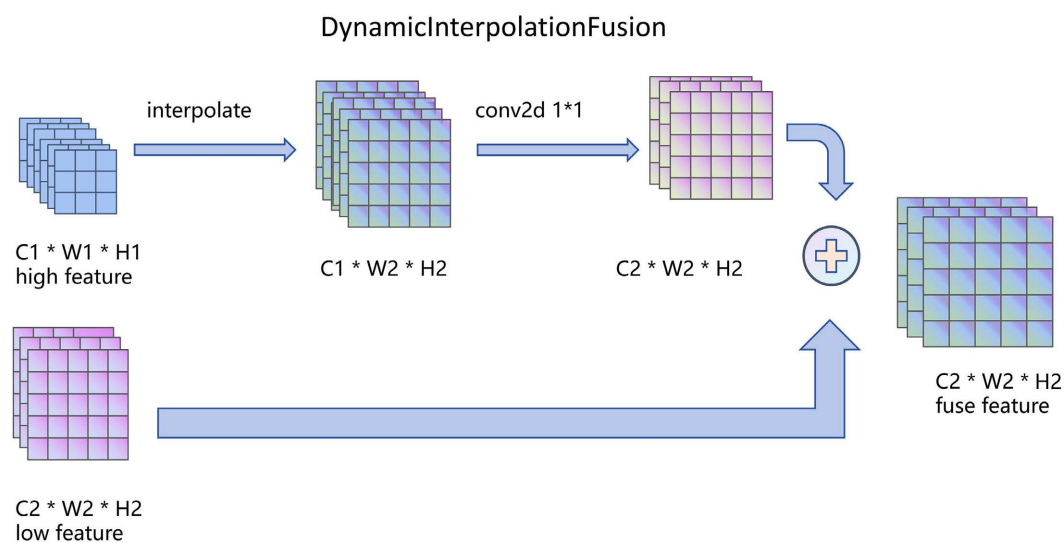


Figure 7. Dynamic interpolation fusion

图 7. 动态插值融合

### 3.2.3. 多特征融合模块

基线模型采用固定权重融合不同尺度特征, 无法根据高尺度特征调整低尺度特征的融合比例, 导致特征冗余。设计多特征融合模块, 通过高尺度特征生成动态权重, 自适应调整低尺度特征的融合强度。例如, 当高尺度特征包含明显的目标语义(如缺陷类别)时, 动态权重会增强低尺度特征中的空间细节, 提升融合效果。其模块结构图如图 8 所示。

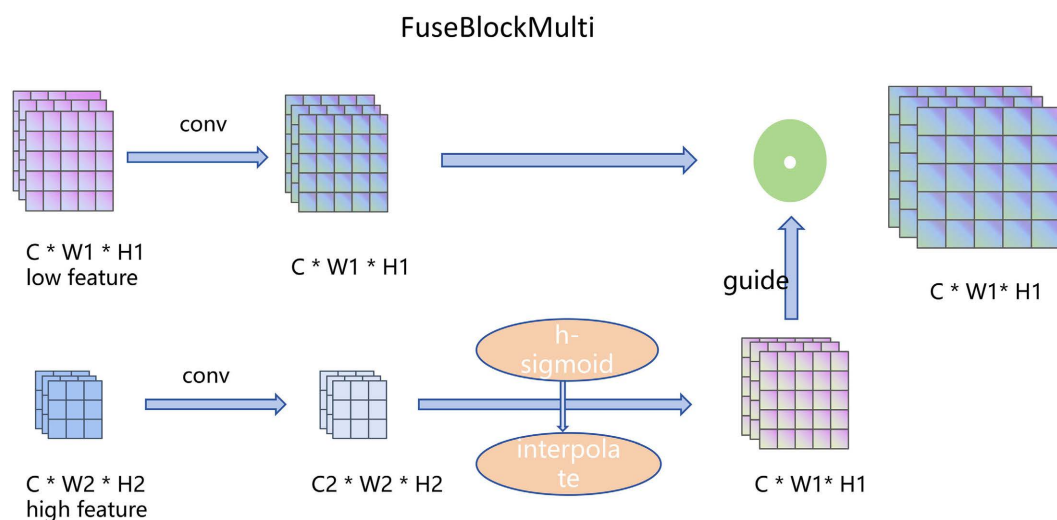


Figure 8. Multi-feature fusion structure

图 8. 多特征融合结构



3.3. 可变形注意力增强 C2PSA 模块(C2DA)

YOLO11 的 C2PSA 模块采用固定注意力采样(均匀分布的采样点),无法适应表面缺陷的变形特性.为此本文引入可变形注意力机制(DAttention) [20]替换原有的固定注意力采样,提升了对变形缺陷的适应性.该改进根据输入特征的变形信息,动态调整注意力采样点的位置,此外还引入位置编码信息,提升注意力的位置感知能力,最后还保持了 C2PSA 的轻量化,不增加过多参数.其改进后的结构如图 9 所示.

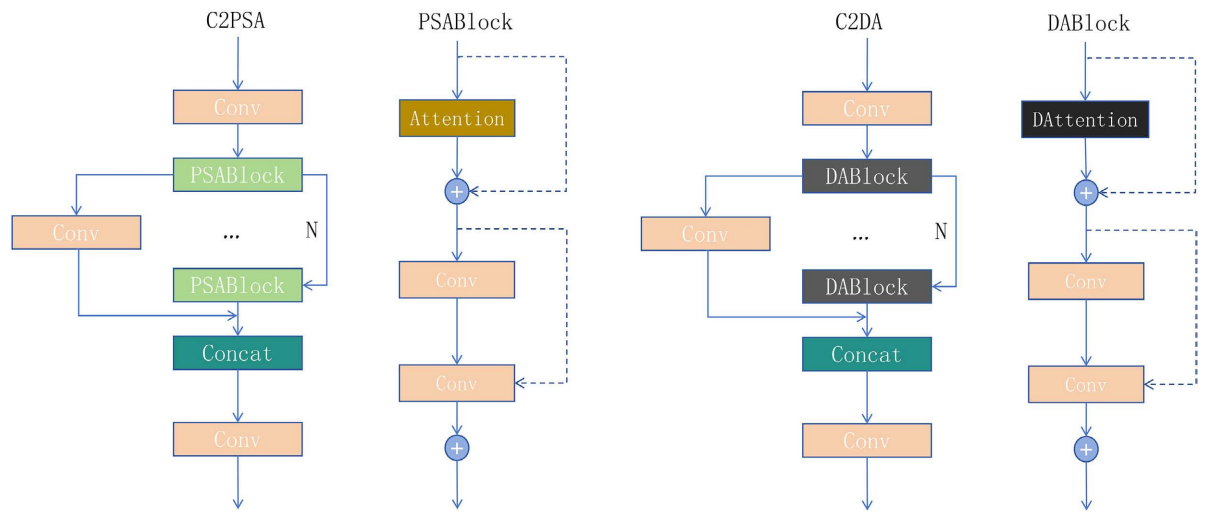


Figure 9. Deformable attention enhancement module  
图 9. 可变形注意力增强模块

4. 实验结果及分析

4.1. 实验环境

实验平台使用 NVIDIA GeForce RTX 3090 的 GPU, 操作系统为 Windows。深度学习框架为 Pytorch 2.6.0, Python 3.12, Cuda 12.6。实验环境参数如表 1 所示。

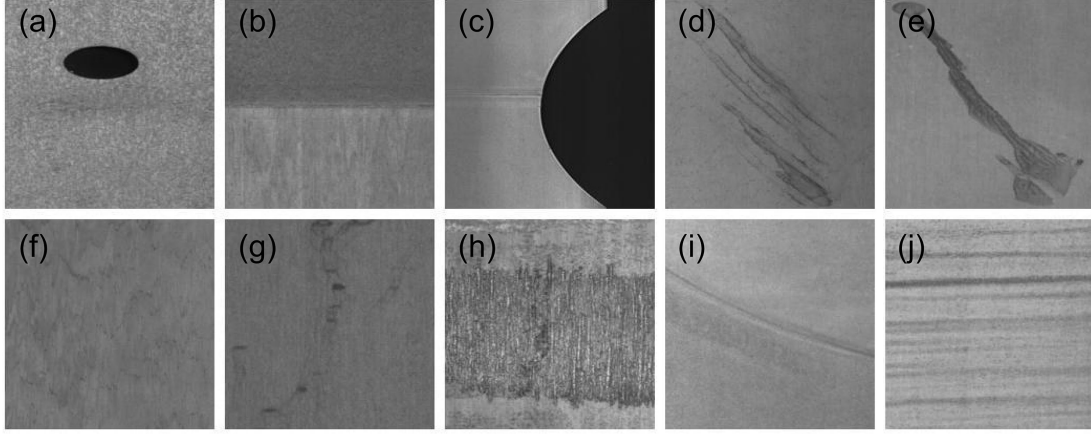
Table 1. Parameters of the experimental environment  
表 1. 实验环境参数

Experimental parameter	Value
Epoch	300
Batch size	32
Optimize	SGD
Image Size	640
Learn Rate	0.01
Momentum	0.937

4.2. 数据集及预处理

实验使用 GC10-DET [21]数据集来训练和测试模型, 其包含 10 种常见的金属表面缺陷类型, 如冲孔、焊缝、新月形缝隙、水斑、油斑、丝斑、夹杂物、轧坑、折痕和腰部折痕。该数据集包括 2294 张

灰度图像，所有缺陷都在钢板的表面上。实验中按照 7:2:1 的比例随机划分训练集、验证集和测试集，通过旋转、裁切、增加高斯噪声、相机传感器噪声以图像虚焦增加数据集样本多样性，更好地应对实际生产中的复杂场景并提升模型的泛化能力，扩充后的数据集共 3570 张图像，缺陷图像示例如图 10 所示。



**Figure 10.** Examples of image defects. (a) Punching; (b) Weld; (c) Crescent-shaped fine crack; (d) Water stain; (e) Oil stain; (f) Silk stain; (g) Inclusion; (h) Rolling pit; (i) Crease; (j) Waist crease

**图 10.** 图像缺陷示例。(a) 冲孔；(b) 焊缝；(c) 新月形细缝；(d) 水斑；(e) 油斑；(f) 丝斑；(g) 夹杂物；(h) 轧坑；(i) 折痕；(j) 腰部折痕

### 4.3. 评价指标

为了评估本文提出的检测模型在钢表面缺陷检测任务中的性能，实验采用精确率  $P$  (precision)、召回率  $R$  (recall)、平均精度  $AP$  (average precision)、平均精度均值  $mAP$  (mean average precision) 与参数量  $Params$  (parameters) 作为评价指标。各指标计算公式为：

$$P = \frac{TP}{TP + FP} \times 100\%, \quad (4)$$

$$R = \frac{TP}{TP + FN} \times 100\%, \quad (5)$$

$$AP = \int_0^1 P dR, \quad (6)$$

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i, \quad (7)$$

式中： $TP$  为正确预测为正类的正样本数量； $FP$  为错误预测为正类的负样本数量； $FN$  为错误预测为负类的正样本数量； $AP$  为模型平均准确度； $n$  为类别数； $AP_i$  为第  $i$  种缺陷类别的  $AP$  值； $mAP@0.5$  为检测到的边界框与真实框之间的交集大于或等于 0.5 时的均值平均精度。

### 4.4. 对比实验

为了进一步验证 DCD-YOLO 模型的先进性，将改进模型与 SSD、Faster R-CNN、YOLOv5s、YOLOv8n、YOLOv8 以及 YOLO11n 和 YOLO11s 的模型进行对比，在 GC10-det 数据集上评估各个模型的准确率、召回率、均值平均精度和参数量，不同检测网络对比实验如表 2 所示。

**Table 2.** Comparison experiments of different detection networks  
**表 2.** 不同检测网络对比实验

Model	P/%	R/%	mAP@0.5/%	Params/M
SSD	62.8	62.4	64.1	24.4
Faster RCNN	59.8	63.6	62.7	136.8
YOLOv5s	61.5	63.4	64.2	46.2
YOLOv8n	68.2	59.4	63.3	3.01
YOLOv8s	73.4	60.0	64.5	11.1
YOLO11n	64.7	61.0	63.6	2.58
YOLO11s	63.2	63.3	63.6	9.42
ours	66.4	63.8	66.9	3.07

从表 2 可知,DCD-YOLO 在各项关键指标上均优于其他算法,其中 mAP@0.5 指标较 SSD、YOLOv5s、YOLOv8n、YOLOv8s、YOLOv11n 和 YOLOv11s 分别提高 2.8%、2.7%、3.6%、2.4%、3.3% 和 3.3%。YOLOv8n 的模型的准确率虽然略高于 DCD-YOLO,但其召回率却远远低于我们的模型。通过上述检测网络对比实验分析可知,DCD-YOLO 网络模型综合性能最优,模型兼具高精度以及轻量化的优势,充分体现了本文模型能更好的完成钢表面缺陷检测任务。

#### 4.5. 消融实验及分析

为了进一步验证每个改进模块的有效性,以 YOLOv11n 为基线在 GC10-DET 数据集上设计 7 组消融实验,使用 P、R、AP 和 mAP@0.5 作为衡量模型性能的指标,实验均采用同一参数和网络环境测试,消融实验结果如表 3 所示。

**Table 3.** Ablation experiment results  
**表 3.** 消融实验结果

Model	A	B	C	P	R	mAP@50	AP%									
				%	%	%	Pu	WI	Cr	WS	Os	Ss	In	Rp	Cr	Wf
M1	×	×	×	64.7	61.0	63.6	93.8	78.4	93.0	80.5	66.6	57.5	25.8	20.4	58.3	61.3
M2	√	×	×	62.4	63.9	64.1	94.6	88.2	94.3	<b>82.2</b>	64.1	<b>62.6</b>	24.4	23.2	45.8	61.5
M3	×	√	×	68.2	60.3	65.1	<b>95.9</b>	87.8	94.3	79.2	69.1	59.7	26	30.4	44.8	63.6
M4	×	×	√	<b>70.2</b>	62.0	64.5	93.1	86.4	93.3	78.1	70.0	53.4	26.9	24.9	51.2	<b>67.0</b>
M5	√	√	×	65.2	61.3	63.5	94.4	86.6	95.7	79.8	60.6	56.1	22.1	19.4	<b>60.0</b>	58.6
M6	√	×	√	68.5	59.0	64.5	93.7	84.9	95.2	80.6	65.2	60.6	<b>28.9</b>	17.4	55.7	62.1
M7	×	√	√	61.9	<b>66.0</b>	64.1	94.7	<b>89.7</b>	<b>96.1</b>	78.4	67.0	53.7	26.9	7.3	59.9	65.1
M8	√	√	√	66.4	63.8	<b>66.9</b>	95.4	88.4	95.9	75.3	<b>70.9</b>	53.5	22.6	<b>41.1</b>	59.9	65.3

注: × 为不添加模块; √ 为添加模块; 加粗数据为最优值。

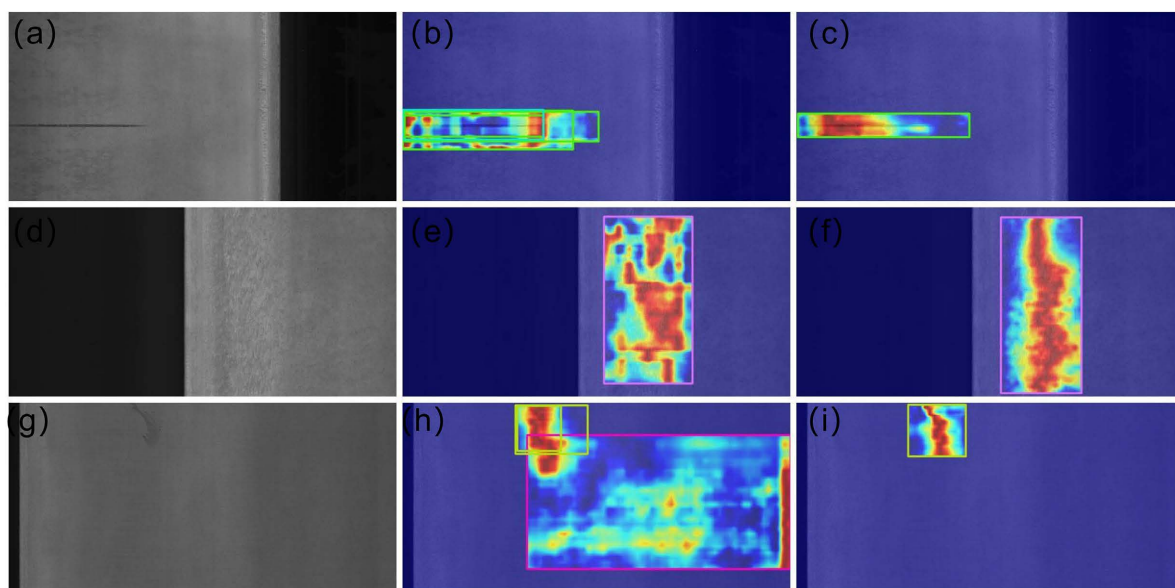
表 3 中 M1 为未经任何改进的基线模型 YOLOv11n,该模型下 Rp 的 AP 最低, Pu 的 AP 最高,表明基线模型对细长、低对比度缺陷的识别能力较弱,难以满足工业场景“零遗漏”的核心要求; M2 中单独使用 C3k2-DIMB,与基线模型相比, mAP@0.5 提升 0.5%,召回率显著提升 2.9%,但精准率略有下降。

表明该改进增强了模型对细长裂纹的边缘感知能力,能有效提高了召回率。M3 为在基线模型上替换颈部结构为 CGRFPN 其  $mAP@0.5$  提升 1.5%, 精准率提升 3.5%, 但召回率微降 0.7%, 表明该模块有效地抑制了背景干扰, 提高了精准率。M4 通过将原有 C2PSA 模块替换为改进后的 C2DA 模块与基线 M1 相比,  $mAP@0.5$  提升 0.9%, 精准率提升 5.5%, 召回率(R)微升 1.0%。表明该改进通过可变形注意力优化了复杂缺陷的分类精度, 且不影响模型效率, 是单一模块中性价比最高的改进。

单一改进均能使均值平均精度有 0.6% 以上的提升。两两组合的实验中, M5 在 M2 的基础上增加 CGRFPN, 其  $mAP@0.5$  下降 0.6%, 召回率下降 2.6%, 表明 C3k2-DIMB 模块的“边缘感知”与 CGRFPN 模块的“局部小缺陷感知”存在特征冲突, 导致裂纹检测性能下降; M6 在 M2 基础上增加 C2DA 模块,  $AP@0.5$  提升 0.4%, 但召回率下降 4.9%, 表明: C3k2-DIMB 模块的“召回率提升”与 C2DA 模块的“精准率提升”未形成协同, 导致划痕检测性能退化。M7 在 M3 的基础上增加 C2DA 模块,  $mAP@0.5$  下降 1.0%, 精准率下降 6.3%, 表明 CGRFPN 模块的“局部小缺陷感知”与 C2DA 模块的“复杂缺陷分类”过度强化了裂纹、腐蚀的检测, 忽略了划痕的识别, 导致划痕 AP 大幅下降。

在 M8 中使用了全组合改进方式, 其性能较基线 M1 大幅提升, 其中  $mAP@0.5$ , 较基线提升 3.3%; 精准率较基线提升 1.7%; 召回率较基线提升 2.8%, 参数量仅较基线增加 0.49 M, FLOPs = 8.2 G 仅较基线增加 1.9G。在缺陷类型表现方面: Rp 划痕较基线提升 20.6%; Os 氧化的较基线提升 4.2%; Wf 焊接缺陷较基线提升 4.0%。综上结果表明: C3K2-DIMB、CGRFPN、C2DA 三个模块协同增强了模型的特征提取能力, 全面提升了钢表面缺陷检测的性能, 实现了“高精度 + 高召回 + 轻量化”的平衡, 为钢表面缺陷检测提供了最优解决方案。

在钢铁制造的质量管控体系中, 召回率的战略地位远超其他指标。高召回率意味着模型能够最大限度捕捉产线上的潜在缺陷, 从源头杜绝质量问题流向终端客户。本研究中改进模型达成了 63.8% 召回率, 意味着对缺陷特征的深层语义理解得到了实质性增强。为直观呈现模型在特征提取过程中的这种理解程度的加深, 我们特别绘制了基线模型和改进后的模型的特征热力图, 以直观展示改进后的模型的优势。如图 11 所示。



**Figure 11.** Comparison of detection heatmaps before and after improvement. (a) (d) (g) Original images; (b) (e) (h) Baseline heatmaps; (c) (f) (i) Ours heatmaps

**图 11.** 改进前后检测的热力图对比。(a) (d) (g) 原图; (b) (e) (h) Baseline 热力图; (c) (f) (i) Ours 热力图

热力图[22]的核心价值是可视化模型对图像区域的“注意力分配”，颜色越深表示模型认为该区域与缺陷特征的相关性越高，颜色越浅表示相关性越低。上图直观展示了全组合改进模型相比基线模型在特征提取能力上的提升：从模糊的背景误激活到精准的目标聚焦，从块化的特征识别到细节的清晰勾勒，模型实现了对缺陷特征的深层语义理解。

4.6. 缺陷检测结果可视化对比

为了直观展示 DCD-YOLO 检测模型对钢表面缺陷检测的效果，从 GC10-DET 数据集中随机抽取 3 张不同缺陷的样本，分别使用 YOLOv11n 和改进后的模型进行实验。检测结果可视化如图 12 所示。

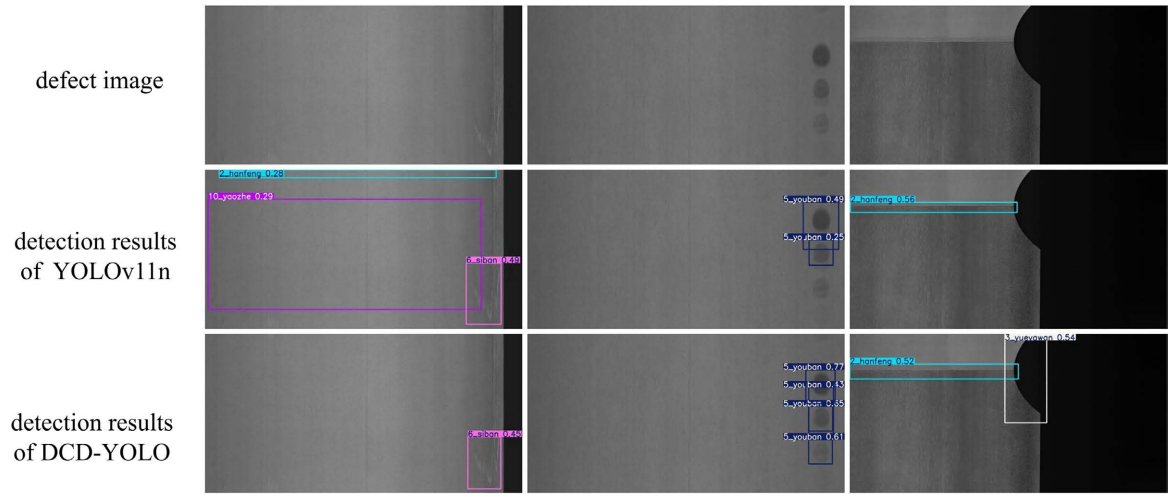


Figure 12. Visualization of detection results  
图 12. 检测结果可视化图

通过对比检测结果可知，YOLOv11n 在检测钢材表面缺陷过程中存在着大量漏检误检问题。相比之下，DCD-YOLO 模型因为加强了对全局和细节纹理特征的理解能力进而减少冗余背景信息的干扰，从而有效的改善漏检误检的问题，因此 DCD-YOLO 在检测复杂背景中的缺陷时效果也更好。

5. 结论

针对钢材表面缺陷检测任务中，缺陷多样、形状复杂和背景纹理干扰巨大等因素导致的检测不准确问题，提出了 DCD-YOLO 检测模型。首先，设计 DIMB 模块对 C3k2 结构进行改进，提升模型对多尺度特征的捕获能力。其次，更换 PANet 为 CGRFPN，在实现多尺度特征的灵活组合的同时，加强了高层语义对低层细节的信息引导作用，增强了特征的融合能力。最后，提出 C2DA 模块，通过引入可变形注意力，加强了对不规则缺陷的认识程度。实验结果表明，本文提出的模型在 GC10-DET 数据集上的 mAP@0.5 达到 66.9%，提升 3.3%，同时模型在准确率上提升 1.7%，召回率提升 2.8%，在单一缺陷精度表现上，轧坑的精度提升高达 20.6%，因此相较于其他主流算法，改进后的模型更适应于工业钢材表面缺陷检测任务。然而，改进模型在检测划痕缺陷方面的精度仍有提升的潜力，未来将扩充划痕类缺陷样本以平衡各类缺陷的数量，提升对划痕缺陷的识别能力，同时优化算法结构，在保持检测精度的前提下进一步提升速度保持轻量化。

参考文献

[1] Luo, Q., Fang, X., Liu, L., Yang, C. and Sun, Y. (2020) Automated Visual Defect Detection for Flat Steel Surface: A



- Survey. *IEEE Transactions on Instrumentation and Measurement*, **69**, 626-644. <https://doi.org/10.1109/tim.2019.2963555>
- [2] Tang, B., Chen, L., Sun, W. and Lin, Z. (2023) Review of Surface Defect Detection of Steel Products Based on Machine Vision. *IET Image Processing*, **17**, 303-322. <https://doi.org/10.1049/ipr2.12647>
  - [3] Fang, X.X., Luo, Q.W., Zhou, B.X., et al. (2020) Research Progress of Automated Visual Surface Defect Detection for Industrial Metal Planar Materials. *Sensors*, **20**, 5136. <https://doi.org/10.3390/s20185136>
  - [4] Ren, S., He, K., Girshick, R. and Sun, J. (2017) Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **39**, 1137-1149. <https://doi.org/10.1109/tpami.2016.2577031>
  - [5] He, K., Gkioxari, G., Dollar, P. and Girshick, R. (2017) Mask R-CNN. 2017 *IEEE International Conference on Computer Vision (ICCV)*, Venice, 22-29 October 2017, 2961-2969. <https://doi.org/10.1109/iccv.2017.322>
  - [6] Lu, J., Zhu, M., Ma, X. and Wu, K. (2024) Steel Strip Surface Defect Detection Method Based on Improved YOLOV5s. *Biomimetics*, **9**, Article 28. <https://doi.org/10.3390/biomimetics9010028>
  - [7] Wei, L., Dragomir, A., Dumitru, E., et al. (2016) SSD: Single Shot MultiBox Detector. Springer.
  - [8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
  - [9] Yang, S., Xie, Y., Wu, J., Huang, W., Yan, H., Wang, J., et al. (2024) CFE-YOLOV8s: Improved YOLOV8s for Steel Surface Defect Detection. *Electronics*, **13**, Article 2771. <https://doi.org/10.3390/electronics13142771>
  - [10] Zhang, X., Wang, Y. and Fang, H. (2024) Steel Surface Defect Detection Algorithm Based on ESI-YOLOV8. *Materials Research Express*, **11**, Article 056509. <https://doi.org/10.1088/2053-1591/ad46ec>
  - [11] Huang, Y., Tan, W., Li, L. and Wu, L. (2023) WFRE-YOLOV8s: A New Type of Defect Detector for Steel Surfaces. *Coatings*, **13**, Article 2011. <https://doi.org/10.3390/coatings13122011>
  - [12] He, K., Zhang, X., Ren, S. and Sun, J. (2015) Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**, 1904-1916. <https://doi.org/10.1109/tpami.2015.2389824>
  - [13] Lin, T., Dollar, P., Girshick, R., He, K., Hariharan, B. and Belongie, S. (2017) Feature Pyramid Networks for Object Detection. 2017 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, 21-26 July 2017, 936-944. <https://doi.org/10.1109/cvpr.2017.106>
  - [14] Liu, S., Qi, L., Qin, H., Shi, J. and Jia, J. (2018) Path Aggregation Network for Instance Segmentation. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake, 18-23 June 2018, 8759-8768. <https://doi.org/10.1109/cvpr.2018.00913>
  - [15] Wang, K., Liu, J. and Cai, X. (2025) C2PSA-Enhanced YOLOv11 Architecture: A Novel Approach for Small Target Detection in Cotton Disease Diagnosis. ArXiv, abs/2508.12219.
  - [16] Shi, D. (2023) Transnext: Robust Foveal Visual Perception for Vision Transformers. 2024 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 16-22 June 2024, 17773-17783. <https://doi.org/10.1109/cvpr52733.2024.01683>
  - [17] Tolstikhin, I.O., Housby, N., Kolesnikov, A., et al. (2021) MLP-Mixer: An All-MLP Architecture for Vision. *Advances in Neural Information Processing Systems*, **34**, 24261-24272.
  - [18] Li, J., Hassani, A., Walton, S. and Shi, H. (2023) ConvMLP: Hierarchical Convolutional Mlps for Vision. 2023 *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, Vancouver, 17-24 June 2023, 6307-6316. <https://doi.org/10.1109/cvprw59228.2023.00671>
  - [19] Ni, Z., Chen, X., Zhai, Y., Tang, Y. and Wang, Y. (2024) Context-Guided Spatial Feature Reconstruction for Efficient Semantic Segmentation. In: Leonardis, A., Ricci, E., Roth, S., Russakovsky, O., Sattler, T. and Varol, G., Eds., *Lecture Notes in Computer Science*, Springer, 239-255. [https://doi.org/10.1007/978-3-031-72943-0\\_14](https://doi.org/10.1007/978-3-031-72943-0_14)
  - [20] Xia, Z., Pan, X., Song, S., Li, L.E. and Huang, G. (2022) Vision Transformer with Deformable Attention. 2022 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, New Orleans, 18-24 June 2022, 4784-4793. <https://doi.org/10.1109/cvpr52688.2022.00475>
  - [21] Lv, X., Duan, F., Jiang, J., Fu, X. and Gan, L. (2020) Deep Metallic Surface Defect Detection: The New Benchmark and Detection Network. *Sensors*, **20**, Article 1562. <https://doi.org/10.3390/s20061562>
  - [22] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. and Batra, D. (2020) Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *International Journal of Computer Vision*, **128**, 336-359. <https://doi.org/10.1007/s11263-019-01228-7>