

基于改进YOLO11的复杂场景带钢表面缺陷检测算法

王 盼, 李 莉

天津职业技术师范大学电子工程学院, 天津

收稿日期: 2026年1月2日; 录用日期: 2026年2月2日; 发布日期: 2026年2月10日

摘要

针对钢铁表面缺陷检测中存在的形态不规则、特征细微及背景干扰复杂等难题, 本文提出一种改进YOLO11的钢铁表面缺陷检测算法。该算法在主干网络的C3k2模块中引入可变形注意力机制, 增强模型对不规则缺陷如裂纹(Cr)的空间特征捕捉能力。采用风车形卷积替代部分标准卷积, 提升模型对夹杂(Pa)、斑点(Ps)等低对比度细微目标的特征提取性能。同时在C2PSA模块中集成基于变分率削减的线性复杂度注意力机制(TSSA), 有效实现全局上下文信息交互。在NEU-DET数据集上的实验表明, 改进模型mAP@0.5达到80.76%, 相比原始YOLO11提升3.16%, 对比其他的目标检测主流方法, 改进算法精度上都有了明显的提升, 满足钢铁表面缺陷检测的精度要求。

关键词

钢铁表面缺陷检测, 目标检测, YOLO11, 可变形注意力, TSSA

Complex Scene Strip Steel Surface Defect Detection Algorithm Based on Improved YOLO11

Pan Wang, Li Li

School of Electronic Engineering, Tianjin University of Technology and Education, Tianjin

Received: January 2, 2026; accepted: February 2, 2026; published: February 10, 2026

Abstract

To address challenges in steel surface defect detection, such as irregular shapes, subtle features, and complex background interference, this paper proposes an improved YOLO11 algorithm for

steel surface defect detection. The algorithm introduces a deformable attention mechanism in the C3k2 module of the backbone network, enhancing the model's ability to capture spatial features of irregular defects such as cracks (Cr). Part of the standard convolution is replaced with pinwheel-shaped convolutions to improve the model's feature extraction performance for low-contrast, fine targets such as inclusions (Pa) and spots (Ps). Additionally, the C2PSA module integrates a linear-complexity attention mechanism (TSSA) based on variance reduction, effectively enabling global context information interaction. Experiments on the NEU-DET dataset show that the improved model achieves an mAP@0.5 of 80.76%, an increase of 3.16% compared to the original YOLO11. Compared with other mainstream object detection methods, the improved algorithm demonstrates a significant improvement in accuracy, meeting the precision requirements for steel surface defect detection.

Keywords

Steel Surface Defects Detection, Object Detection, YOLO11, Deformable Attention, TSSA

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

钢铁工业作为现代工业社会的重要基石之一，支撑并推动着交通、机械、电子等行业的协同发展。然而，受生产设备与工艺水平的限制，钢材表面常会产生斑点、裂纹、划痕等各类缺陷与瑕疵，这些问题不仅会削弱钢材的力学性能与结构强度，甚至会危及生产安全。因此，对这类缺陷开展高效、自动化的检测工作，对于保障产品质量、维护人员安全具有至关重要的意义。此类表面缺陷绝非单纯的外观质量问题，更有可能成为应力集中的源头，进而导致产品使用寿命缩短，甚至引发威胁人身安全的生产事故[1]。

传统人工目视检测存在效率低、标准不一致、易受主观因素干扰等问题[2]。随着智能制造发展，深度学习技术为缺陷自动检测提供了新方案。马冬梅等[3]利用 YOLOv5 对热轧带钢进行缺陷检测，樊嵘等[4]改进 YOLOv7 提升带钢表面小目标识别，马燕婷等[5]通过空间金字塔池化增强多尺度感知能力来提升对带钢表面缺陷精度；周亚罗等[6]提出 STCS-YOLO 降低带钢表面纹理干扰，王春梅等[7]设计 VSC 结构增强带钢表面不规则缺陷识别，戴林华等[8]开发 HSED-YOLO 降低计算负担，刘凤春等[9]利用可变形卷积提高复杂背景识别率；吕宝展等[10]通过部分卷积和通道注意力改进 YOLOv7 提升精度，周世阳等[11]的 SKS-YOLO 结合 EfficientNetv2 和 ASPP 模块实现高精度检测，黄帆等[12]证实注意力机制对 YOLOv9 和 YOLOv10 均有提升。

尽管这些方法取得了一定的进展，在小目标检测能力和整体检测精度上仍有提升空间。因此，本文提出一种深度集成前沿视觉机制的改进 YOLO11 模型，集成可变形注意力(DAttention)至 C3k2 模块，引入风车形卷积(PSCConv)替换标准卷积，构建基于变分率削减原理(MCR²)的线性复杂度注意力机制(TSSA)。

2. YOLOV11 目标检测算法改进

2.1. 改进算法总体设计

随着深度学习技术的迅猛发展以及卷积神经网络(CNNs)在目标检测领域的广泛应用，检测速度与检

测精度均得到了显著提升。卷积神经网络通过对大规模数据集内的目标特征进行学习, 实现了检测性能的优化。在单阶段目标检测算法中, YOLO 系列无疑是应用最为广泛的一类算法。该算法将目标检测问题建模为回归问题, 通过单个神经网络直接预测目标的类别与位置, 因此在各领域均有着十分广泛的应用, 在学术界与工业界均备受青睐。其中, YOLOv11 的基础网络架构在实时处理与目标检测精度的平衡上展现出卓越性能。

然而, 直接将其应用于钢铁表面缺陷检测时, 标准网络往往难以应对高度形变的裂纹以及低对比度的纹理干扰。鉴于此, 本文对网络架构进行了深度重构, 其总体改进布局如图 1 所示。

考虑到钢材表面缺陷(如裂纹)的高度复杂性, 本文在骨干网络(Backbone)的 C3k2 模块中嵌入可变形注意力机制(DAttention), 使特征采样过程能够根据目标的实际形态进行动态调整。针对颈部网络(Neck)对微小夹杂物识别困难的问题, 本文摒弃传统卷积操作, 引入风车形卷积(PSConv), 借助其非对称感受野特性增强弱特征的鲁棒性。同时, 本文采用线性复杂度注意力机制对 C2PSA 模块进行改进, 在实现全局信息聚合的同时, 解决了算力冗余问题。上述一系列改进措施, 确保模型在保持高效运行的同时, 进一步提升了检测精度。

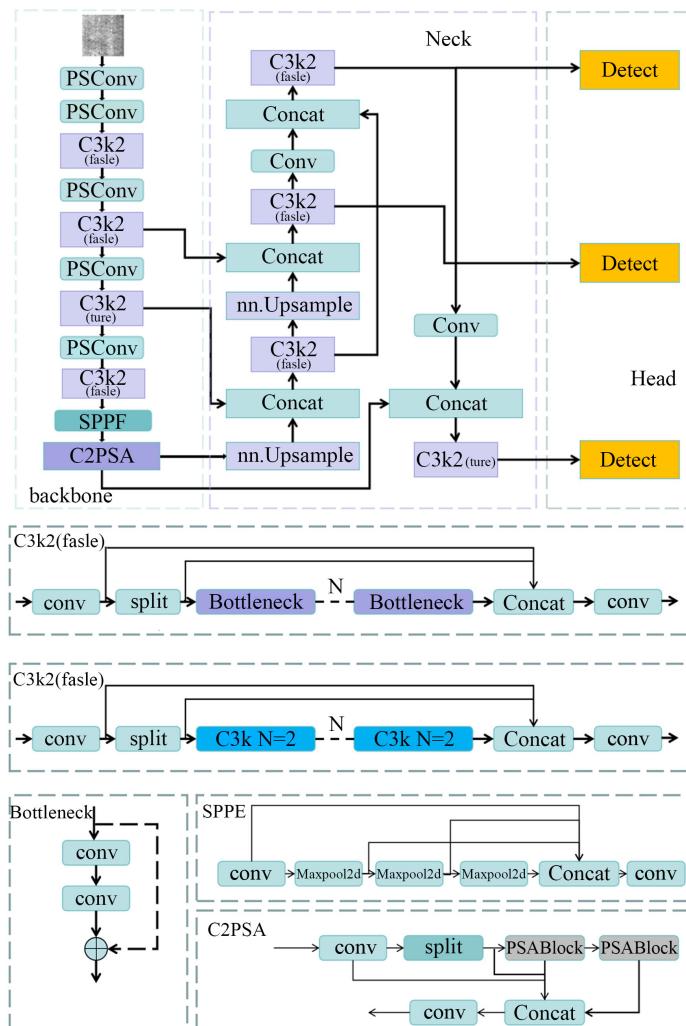


Figure 1. Improved YOLOv11 architecture diagram

图 1. 改进 YOLOV11 结构图

2.2. 集成可变形注意力的空间增强模块(C3k2-DAT)

在带钢表面缺陷检测任务中, 裂纹等缺陷目标的空间分布往往具有极强的随机性, 而原始 C3k2 模块的固定采样方式难以很好地适配其不规则边缘。为此, 本文对原始模块的算子结构进行重构, 在特征提取分支中引入可变形注意力机制(DAttention) [13], 最终设计出具备空间自适应感知能力的 C3k2-DAT 模块。

2.2.1. C3k2-DAT 架构设计与工作原理

原始自注意力机制仅能在固定网格上对特征进行采样; 为实现利用自注意力机制对特定区域对应的目标特征进行映射, 本文提出 C3k2-DAT 模块, 使模型能够动态捕捉关键特征。该模块的演化逻辑与工作流程可分为以下三个阶段(见图 2):

偏移量生成:首先将输入特征送入一个轻量级二维网络(由深度可分离卷积中的深度卷积(DWConv)搭配高斯误差线性单元(GELU)激活函数构成), 基于查询特征的局部上下文信息, 为锚定在每个参考点上的二维空间网格生成偏移量。

可变形采样:借助生成的偏移量对原始参考点网格进行形变, 得到自适应采样点; 随后在这些不规则位置上, 对特征图执行双线性插值采样, 提取对应特征。

注意力聚合:将采样得到的特征映射为注意力计算所需的键(Keys)与值(Values), 并与原始查询特征(Queries)共同参与注意力运算; 同时通过对偏置表进行双线性采样得到相对位置偏置, 使模型能够有效学习形变后特征间的空间拓扑关系。

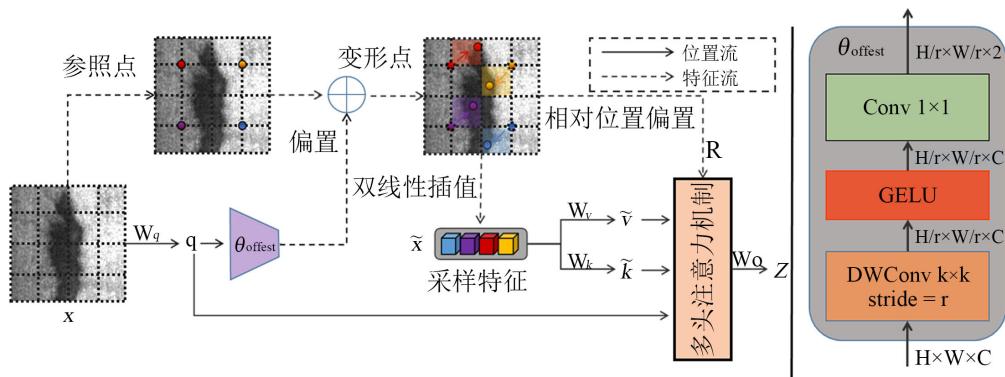


Figure 2. Schematic diagram of deformable attention mechanism

图 2. 可变形注意力机制示意图

2.2.2. 针对缺陷识别的物理动机

本文在 C3k2 模块中引入的 DAT 机制, 仅聚焦于钢材缺陷本身的相关几何特征, 以及工业场景带来的技术干扰复杂性。由于带钢表面的裂纹或其他易混淆缺陷大多呈现细长且随机分布的形态, 标准卷积内固定尺寸的采样矩阵受限于规则感受野, 导致模型在提取特征的同时, 不可避免地引入了大量噪声背景信息。通过融入 DAT 机制, 模型能够利用可学习的空间偏移量, 引导采样点精准定位在缺陷区域而非背景区域, 从而缓解了感受野与不规则目标之间的空间不匹配问题。针对工业现场存在的反光、尺度差异、金属纹理等干扰因素, DAT 的可变形采样机制可使模型突破常规区域的限制, 将计算权重集中于感兴趣区域(ROI)内的少量像素点, 形成一种核心特征捕捉方法, 有效提升了复杂背景下低对比度边缘的提取精度与识别鲁棒性。

2.3. 基于风车形卷积的钢铁缺陷特征强化模块(PSCConv)

针对钢材表面缺陷(如夹杂物 Pa 与斑点 Ps)在成像时呈现出的中心能量集中、边缘呈类高斯分布平缓衰减的特征, 本文采用风车卷积(PSCConv) [14]对 YOLO11 网络中的通用卷积单元进行了替换。

2.3.1. PSCConv 架构设计与非对称填充机制

如图 3 所示, 本文所提方法(风车形卷积, PSCConv)的核心设计思路为: 将标准方形卷积核分解为四个方向的条形卷积核, 形成类似“风车”的采样模式。通过采用非对称填充策略, PSCConv 可在图像的不同区域生成水平与垂直方向的卷积核, 其权重分布从中心向四周延展, 近似于高斯分布的权重衰减规律。

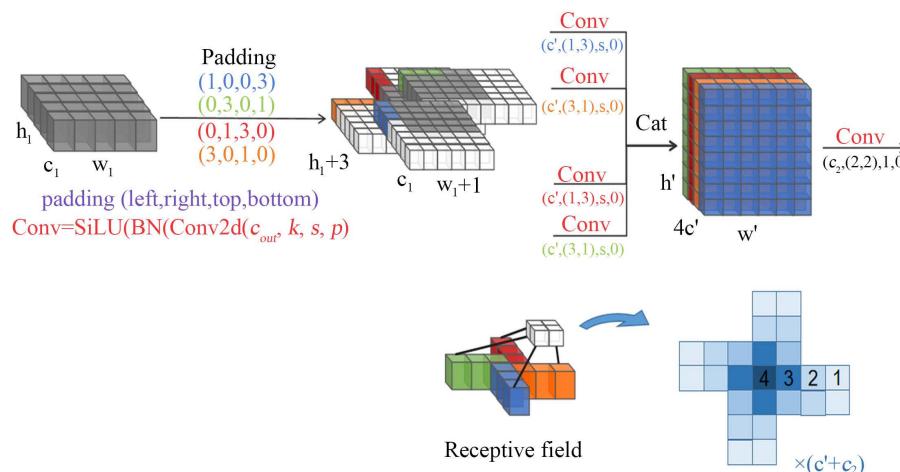


Figure 3. Schematic diagram of the pinwheel-shaped convolution module architecture

图 3. 风车状卷积模块的架构示意图

非对称填充策略(Asymmetric Padding): 给定输入特征图 $X^{(h_1 \times w_1 \times c_1)}$, 通过定义四个方向的互补填充模式: $P = \{(k, 0, 1, 0), (0, k, 0, 1), (0, 1, k, 0), (1, 0, 0, k)\}$ 打破标准卷积采样位置的对称性。这种设计使模型能够生成覆盖不同空间维度的水平和垂直条形卷积核, 从而有效捕捉钢铁表面的各向异性纹理(如具有明显方向性的划痕 Sc)。

分组条形卷积(Directional Group Convolution): 系统执行四个并行的条形卷积分支, 每个分支分别负责不同象限的特征提取。水平卷积($1 \times k$)组与垂直卷积组($k \times 1$)协同工作, 既显著扩大了有效感受野(在 $k = 3$ 时感受野由 9 提升至 25), 又通过减少参数量维持了计算效率。

特征拼接与归一化(Contact & Normalize): 四个方向的特征输出在通道维度进行拼接, 随后通过一个 2×2 的标准卷积进行跨通道信息交互和空间归一化, 最终输出具备深度语义信息的特征图 Y_{out} 。

2.3.2. 物理动机: 类高斯采样与缺陷匹配

本文采用 PSCConv 的核心逻辑在于其采样权重分布与缺陷物理特性的高度一致性:

中心密集采样: 四个条形卷积核在中心位置重叠, 形成了极高密度的采样点分布, 这天然适配夹杂、斑点类缺陷的中心聚集特性。

权重向外衰减: 采样密度由中心向四周递减, 形成隐式的类高斯权重分布。其数学表达可近似为:

$$D(x, y) \approx A \cdot e^{-\frac{(x-x_0)^2 + (y-y_0)^2}{2\sigma^2}} \quad (1)$$

该机制使模型能够更精准地拟合缺陷灰度的空间衰减规律, 从而在复杂的工业背景噪声中有效提取

缺陷核心特征。

2.4. 基于变分率削减原理的线性注意力重构(C2TSSA 模块)

针对工业现场检测对实时性与高维特征融合深度的双重需求, 本文引入基于变分率削减原理 (Maximal Coding Rate Reduction, MCR²) 的线性复杂度注意力机制[15] (TSSA), 并对 YOLO11 中的 C2PSA 模块进行了重构, 设计了具备线性计算复杂度的 C2TSSA 模块。

2.4.1. Token 统计自注意力(TSSA)设计机理

如图 4 图例所述, TSSA 算子可将传统自注意力机制中两两相似度计算的二次复杂度降低至线性复杂度, 其内部工作逻辑如下。首先, 将输入特征映射至多个低维空间, 并从映射后特征的能量分布中提取隶属度矩阵, 以此精细调控特征中每个标记(token)在不同特征子空间的分配方式。在每个特征子空间内, 计算标记的加权二阶矩统计量, 并通过变分优化准则得到一个对角注意力权重矩阵; 该矩阵可保留具备最大编码率增益的主成分方向, 同时抑制所有背景噪声, 从而实现数据驱动的低秩特征映射。随后, 利用这些预估的隶属度概率对更新后的特征进行相应加权, 并对各特征子空间进行聚合, 完成标记级的全局上下文建模。

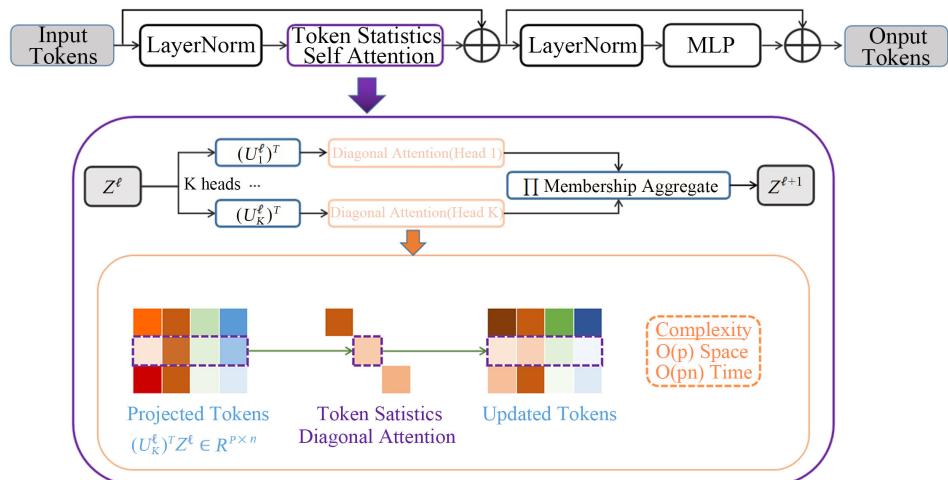


Figure 4. Architecture of the C2TSSA module
图 4. C2TSSA 模块架构

2.4.2. C2TSSA 模块架构与 CSP 设计

为保证感受野具备足够的尺寸, 且能够为空间特征提取与网络内的特征流转提供丰富的信息, 本文将 TSSA 算子嵌入至 CSP 模块中。从网络架构图可知, 输入特征沿两条路径分支传递: 其一为恒等映射路径, 该路径可直接保留空间特征信息, 同时维持梯度流无偏传递; 其二为 TSSA 增强路径, 该路径通过堆叠 TSSA 模块单元(由层归一化、TSSA 算子、多层感知机构成), 捕捉复杂场景下的长距离依赖关系, 随后在通道维度上进行特征拼接, 再经由卷积操作完成特征的压缩与注意力加权, 最终实现细节信息与统计特征的良好平衡。

2.4.3. C2TSSA 模块设计优势

TSSA 模块准则适配于钢材缺陷检测任务, 具有丰富的物理内涵:

- 1) 更强的类别区分能力: TSSA 旨在最大化不同子空间特征之间的编码率增益。例如, 在处理划痕

(Sc)与裂纹(Cr)等易混淆缺陷时, TSSA 可自动拉大类间特征在流形空间中的距离。

2) 高分辨率图像实时处理能力: 钢材表面图像通常包含大量工业纹理噪声, 而 TSSA 的线性复杂度特性, 可确保模型即便在输入高分辨率图像、执行标记更新操作时, 依旧能够维持实时检测速度。

3. 实验结果分析

3.1. 数据集

本文主要通过 NEU-DET 钢材表面缺陷数据集来验证模型的性能。如图六所示, NEU-DET 数据集包括裂纹(crazing)、内含物(inclusion)、斑块(patches)、麻点(pitted_surface)、压入氧化铁皮(rolled_in_scale)和划痕(scratches)共六种钢材表面缺陷。其为每个类别 300 张图像, 共 1800 张图像, 每张图像大小为 200×200 的灰度图。按照 8:1:1 的比例将处理好的数据集随机划分成训练集、测试集和验证集。其中训练集有图片样本 1440 张, 测试集有图片样本 180 张, 验证集有图片 180 张。

3.2. 实验环境

该实验模型在训练过程中, 选择 YOLOv11 作为基本模型, epochs 设置为 300, batchsize 为 24, imagesize 为 640×640 , works 为 8, 初始学习率为 0.01, 学习动量为 0.937, 衰减权重系数为 0.0005, 表 1 是训练使用的硬件和软件配置环境。

Table 1. Experimental environment

表 1. 实验环境

配置对象	环境参数
操作系统	Windows10 64 位
内存	16 G
GPU	NVIDIA GeForce RTX3090
CUDA	11.6
Python	3.8
Pytorch	1.12.1

3.3. 评价指标

文中采用 mAP@0.5 (IoU 阈值为 0.5 时各个类别平均 AP 值)评价模型的检测精度, 其他评估指标有: 精确率(Precision, P)、召回率(Recall, R)、平均精度(Average Precision, AP), 表达式如公式(2)~(5)。

$$P = P_T / (P_T + P_F) \times 100\% \quad (2)$$

$$R = P_T / (P_T + N_F) \times 100\% \quad (3)$$

$$\delta_{AP} = \sum_{i=1}^N p(i) \Delta R(i) = \int_0^1 p(R) dR \quad (4)$$

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

式中: P 表示精确率; R 表示召回率; N 表示总类别个数; P_T 表示正样本预测出正样本数量; P_F 表示负样本预测出正样本数量; N_F 表示正样本预测出负样本数量。

3.4. 实验结果

3.4.1. 消融实验

为验证所提改进模块的有效性, 本文在东北大学钢铁表面缺陷数据集上进行了系统的消融实验。实验分别评估了 C3K2-DAT、C2TSSA 和 PSConv 三个改进模块对模型性能的贡献, 具体方案如表 2 所示, 实验结果如表 3 所示。

Table 2. Ablation experiment plan

表 2. 消融实验方案

实验组别	C3K2-DAT	C2TSSA	PSConv
1	×	×	×
2	√	×	×
3	×	√	×
4	×	×	√
5	√	√	×
6	√	√	√

注: √表示有, ×表示无。

Table 3. Ablation experiment results

表 3. 消融实验结果

实验组别	Cr (%)	In (%)	Pa (%)	Ps (%)	Rs (%)	Sc (%)	mAP50 (%)
1	51.58	94.3	84.94	81.64	81.36	71.76	77.6
2	55.12	89.96	85.66	76.39	93.63	73.78	79.09
3	52.69	92.61	88.75	85	92.15	69.92	80.19
4	49.4	95.1	82.2	83.4	89.3	71	78.2
5	51.31	91.67	88.5	88.08	92.84	69.37	80.29
6	53.63	92.23	87.17	88.27	91.21	72.07	80.76

从表 3 的实验结果可知: 实验组 1 为基准模型, 即未添加任何改进模块的原始模型, 其 mAP50 达到 77.6%, 为后续实验提供了性能参照基准。实验组 2、3、4 用于验证各单个模块的有效性: 其中, C3K2-DAT 模块将模型的 mAP50 从 77.6% 提升至 79.09%, 性能提升 1.49 个百分点; C2TSSA 模块将 mAP50 提升至 80.19%, 相较于基准模型提升 2.59 个百分点, 是对模型整体性能提升最为显著的模块; 而 PSConv 模块则将 mAP50 提升至 78.2%, 较基准模型提升 0.6 个百分点。实验组 5 组合了 C3K2-DAT 与 C2TSSA 两个模块, 其 mAP50 达到 80.29%, 较基准模型提升 2.69 个百分点。实验组 6 为所有组合中的最优结果, 该组融合了全部三个改进模块, 最终 mAP50 达到 80.76%, 相较于原始模型提升 3.16 个百分点, 且所有缺陷类别的检测精度指标均得到提升。

为直观对比检测性能, 图 5 展示了改进算法与 YOLOv11 基线模型的预测效果。在复杂几何形态及多目标干扰场景下, 对于裂纹(Crazing)与杂质(Inclusion)、轧制氧化皮(Rolled-in Scale)等目标, YOLOv11 基线模型出现了部分漏检, 而改进模型能够更有效地捕捉到更多细微特征, 降低了漏检率; 针对斑块(Patches)与划痕(Scratches)之类密集型缺陷, 改进模型生成的预测框与目标实际形状契合度更高。

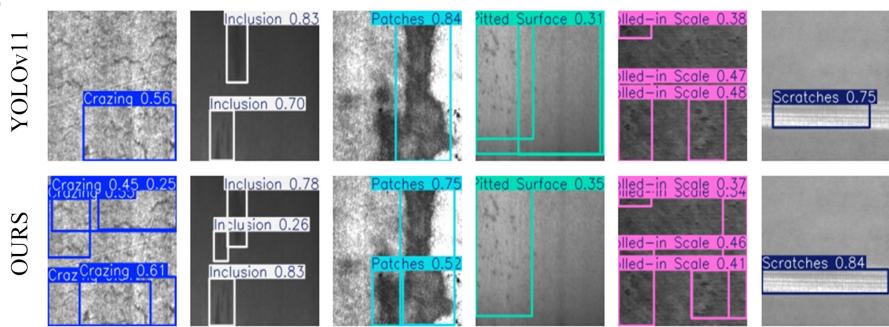
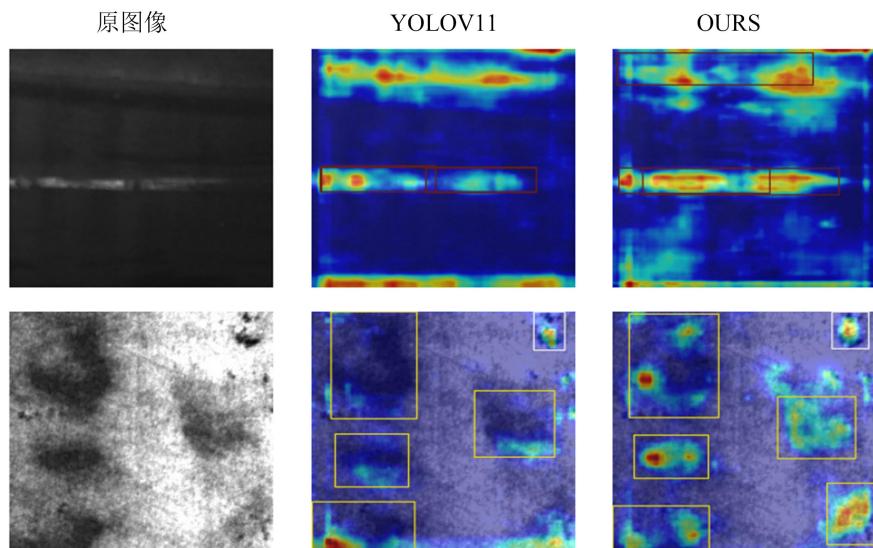
**Figure 5.** Comparison of effects between improved and original algorithms**图 5.** 改进算法与原始算法检测效果对比**Figure 6.** Comparison of heatmaps between the improved algorithm and the original algorithm**图 6.** 改进算法与原始算法热力图对比

图 6 展示了改进模型(OURS)与 YOLOv11 基线模型的 Grad-CAM 热力图对比。在划痕(Scratches)场景中, 相比基线模型散乱的响应, 改进模型的高亮区域呈现出紧贴缺陷路径的长条状分布, 验证了 DAttention 引导采样点匹配非规则几何形状的能力; 而在斑块(Patches)场景中, 改进模型在目标区域产生了更集中且强度更高的响应中心, 证明了 PSConv 凭借非对称感受野设计, 显著增强了对背景干扰下微弱特征信号的捕获灵敏度。

3.4.2. 对比实验

为进一步验证改进算法在检测精度方面的优势, 在相同数据集和硬件环境下, 与部分主流算法进行了对比实验, 结果见表 4。

从实验结果可以看出, 本文提出的改进算法(ours)在 mAP50 指标上达到 80.76%, 相比基线模型 YOLOV11 的 77.6% 提升了 3.16 个百分点, 表现出明显的性能优势。与其他主流检测算法均低于本文算法。在各类别缺陷检测精度方面, 本文算法在 Cr、In、Pa、Ps、Rs 和 Sc 六类缺陷上分别达到 53.63%、92.23%、87.17%、88.27%、91.21% 和 72.07%, 整体表现优于对比算法。特别是在 Cr、Pa 和 Sc 类缺陷的检测上, 本文算法相比 YOLOV11 分别提升了 2.05、2.23 和 0.31 个百分点, 显示出改进模块对复杂缺陷特征提取的有效性。本改进模型在计算效率与实时性方面展现出显著优势, 其参数量仅为 2.52 M, 计算

量为 6.4 GFLOPs, 均优于基线模型 YOLOv11n, 其检测速度高达 155.5 FPS, 仅次于 hyper-yolo。本文提出的改进算法在钢铁表面缺陷检测任务中具有更高的检测精度和检测速度。

Table 4. Average accuracy of different algorithms in detecting various defects
表 4. 不同算法检测各类缺陷的平均精度

算法	Cr (%)	In (%)	Pa (%)	Ps (%)	Rs (%)	Sc (%)	MAP50 (%)	FPS	GFLOPs	参数量(M)
SSD [16]	27.52	65.57	85.92	61.71	56.19	67.51	60.74	37.7	88.2	25.1
YOLOV3 [17]	39.23	78.30	82.41	72.85	63.26	61.63	66.28	68	155.3	61.53
YOLOV5	45.11	81.25	84.92	83.27	71.84	65.72	72.02	83	15.8	7.05
YOLOV8	48.2	82.41	86.32	84.65	85.11	67.92	75.77	165	8.2	3.01
YOLOV11n	51.58	94.3	84.94	81.64	81.36	71.76	77.6	118.9	6.3	2.58
YOLOV12n [18]	41.85	90.07	84.27	82.65	89.65	64.49	75.37	148.9	5.8	2.51
hyper-yolo [19]	46.43	88.48	85.33	86.28	89.72	74.12	73.57	162	9.5	3.6
rt-detr-18 [20]	41.15	94.13	85.10	79.85	92.13	66.37	73.57	113	57	19.9
ours	53.63	92.23	87.17	88.27	91.21	72.07	80.76	155.5	6.4	2.52

4. 结论

钢材表面缺陷目标往往形态不规则、相对尺度差异显著, 且目标局部密度变化频繁, 这导致各类算法的检测精度普遍偏低。针对这一问题, 本文将可变形注意力机制、风车形卷积以及线性复杂度注意力机制(TSSA)进行融合, 提出一种基于 YOLO11 的高性能钢材表面缺陷检测方法。

在 NEU-DET 数据集上的实验结果表明: 可变形注意力机制(DAttention)通过自适应采样捕捉长距离依赖关系, 显著增强了模型对不规则形状缺陷的感知能力; 风车形卷积(PSCConv)凭借独特的非对称结构特征与近似正态的权重分布特性, 适配于常见及稀有缺陷类别的检测任务, 有效提升了模型对低对比度小目标的特征提取能力; TSSA 机制则实现了模块不同阶段深层特征的融合与更新。该方法在钢材表面缺陷检测任务中取得了良好效果, 在 NEU-DET 数据集上的测试显示, 改进算法的 mAP50 达到 80.76%, 相较于基准模型精度提升 3.16%; 同时, 与基准算法相比, 模型对小目标、裂纹、夹杂物等难检测类别具备更强的鲁棒性, 整体检测性能达到预期标准。

参考文献

- [1] 米春风, 卢琨, 汪文艳, 等. 基于机器视觉的带钢表面缺陷检测研究进展[J]. 钢铁研究学报, 2022, 34(5): 401-414.
- [2] 李跃, 李铁柱, 张宝, 等. 带钢表面缺陷检测方法研究进展[J]. 钢铁研究学报, 2023, 35(8): 950-962.
- [3] 马冬梅, 朱佳浩. 面向热轧带钢表面缺陷检测的 YOLOv5 算法优化分析[J]. 计算机应用与软件, 2023, 40(2): 178-184.
- [4] 樊嵘, 张雪峰, 罗鑫, 等. 面向带钢表面小目标缺陷检测的改进 YOLOv7 算法[J]. 合肥工业大学学报(自然科学版), 2024, 47(3): 367-376.
- [5] 马燕婷, 郭晓峰, 高毅, 等. 改进 YOLOv5 网络的带钢表面缺陷检测方法[J]. 电子测量与仪器学报, 2023, 36(8): 150-159.
- [6] 周亚罗, 李志强, 刘凤春, 等. 基于 STCS-YOLO 的带钢表面缺陷检测算法[J]. 中国冶金, 2023, 33(12): 128-138.
- [7] 王春梅, 刘欢. YOLOv8-VSC: 一种轻量级的带钢表面缺陷检测算法[J]. 计算机科学与探索, 2024, 18(1): 151-160.
- [8] 戴林华, 黎远松, 石睿. HSED-YOLO: 一种轻量化的带钢表面缺陷检测模型[J]. 广西师范大学学报(自然科学版), 2025, 43(2): 65-77.
- [9] 刘凤春, 李志强, 周亚罗, 等. YOLO-VDCW: 一种新的轻量化带钢表面缺陷检测算法[J]. 中国冶金, 2024, 34(6):

- 125-135.
- [10] Lv, B., Duan, B., Zhang, Y., *et al.* (2024) Research on Surface Defect Detection of Strip Steel Based on Improved YOLOv7. *Sensors*, **24**, Article 2667. <https://doi.org/10.3390/s24092667>
 - [11] Zhou, S., Ao, S., Yang, Z., *et al.* (2024) Surface Defect Detection of Steel Plate Based on SKS-YOLO. *IEEE Access*, **12**, 76723-76737. <https://doi.org/10.1109/ACCESS.2024.3422244>
 - [12] Huang, F., Wen, H., Zhang, Z., *et al.* (2024) A Comparative Study of YOLOv9 and YOLOv10 for Steel Surfaced Effect Detection. 2024 30th International Conference on Mechatronics and Machine Vision in Practice (M2VIP), Leeds, 3-5 October 2024, 1-6. <https://doi.org/10.1109/M2VIP62491.2024.10746104>
 - [13] 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>
 - [14] Yang, J., Liu, S., Wu, J., Su, X., Hai, N. and Huang, X. (2025) Pinwheel-Shaped Convolution and Scale-Based Dynamic Loss for Infrared Small Target Detection. *Proceedings of the AAAI Conference on Artificial Intelligence*, **39**, 9202-9210. <https://doi.org/10.1609/aaai.v39i9.32996>
 - [15] Wu, Z., Ding, T., Lu, Y., Pai, D., Zhang, J., Wang, W., *et al.* (2024) Token Statistics Transformer: Linear-Time Attention via Variational Rate Reduction. arXiv: 2412.17810.
 - [16] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C., *et al.* (2016) SSD: Single Shot MultiBox Detector. In: Leibe, B., Matas, J., Sebe, N. and Welling, M., Eds., *Computer Vision—ECCV 2016*, Springer, 21-37. https://doi.org/10.1007/978-3-319-46448-0_2
 - [17] Farhadi, A. and Redmon, J. (2018) YOLOv3: An Incremental Improvement. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-22 June 2018, 7794-7803.
 - [18] Tian, Y., Ye, Q. and Doermann, D. (2025) YOLOv12: Attention-Centric Real-Time Object Detectors. arXiv: 2502.12524.
 - [19] Feng, Y., Huang, J., Du, S., Ying, S., Yong, J.H., Li, Y., *et al.* (2024) Hyper-YOLO: When Visual Object Detection Meets Hypergraph Computation. arXiv: 2408.04804.
 - [20] Zhao, Y., Lv, W., Xu, S., Wei, J., Wang, G., Dang, Q., *et al.* (2024) DETRs Beat YOLOs on Real-Time Object Detection. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, 16-22 June 2024, 16965-16974. <https://doi.org/10.1109/cvpr52733.2024.01605>