

# 基于YOLOv8s的钢材表面缺陷检测方法研究

芦志远

辽宁科技大学电子与信息工程学院, 辽宁 鞍山

收稿日期: 2026年3月23日; 录用日期: 2026年4月21日; 发布日期: 2026年4月29日

## 摘要

钢材表面缺陷的准确检测是保障钢铁产品质量的关键环节。针对热轧带钢表面缺陷种类多样、部分类别与背景纹理相似导致漏检率高等问题, 本文针对NEU-DET数据集对YOLOv8s进行系统的优化与评估, 构建强有力的基准检测模型。以东北大学热轧带钢表面缺陷数据集(NEU-DET)为实验基础, 该数据集涵盖裂纹、夹杂物、斑块、麻面、轧入氧化皮和划痕6类缺陷, 共1800张图像。通过高分辨率输入(800×800像素)、AdamW优化器、Mosaic/MixUp/CopyPaste多策略数据增强及余弦学习率调度等组合优化策略, 相比YOLOv8s默认配置(640×640, 默认数据增强)显著提升了模型性能。实验结果表明, 优化后模型在验证集上mAP@0.5达到0.739 (基线0.701), 较基线提升了3.8%。其中斑块类AP最高为0.923, 划痕类为0.868, 最优F1分数为0.69 (置信度阈值0.292)。混淆矩阵及可视化分析揭示了各类缺陷的检测特性, 为后续改进指明了方向。

## 关键词

目标检测, YOLOv8, 钢材表面缺陷, NEU-DET, 数据增强, 深度学习

# Steel Surface Defect Detection Based on YOLOv8s

Zhiyuan Lu

School of Electronic and Information Engineering, University of Science and Technology Liaoning, Anshan Liaoning

Received: March 23, 2026; accepted: April 21, 2026; published: April 29, 2026

## Abstract

Accurate detection of steel surface defects is essential for quality assurance in steel manufacturing. Aiming at the problems of diverse defect types and a high miss-detection rate caused by similarity between certain defect categories and background textures in hot-rolled strip steel, this paper

systematically optimizes and evaluates YOLOv8s on the NEU-DET dataset to construct a strong baseline detection model. Experiments are conducted on the NEU-DET dataset released by Northeastern University, which covers six types of defects—crazing, inclusion, patches, pitted surface, rolled-in scale, and scratches—totaling 1800 images. Through a combination of high-resolution input (800 × 800 pixels), AdamW optimizer, multi-strategy data augmentation (Mosaic/MixUp/CopyPaste), and cosine learning rate scheduling, the optimized model achieves  $mAP@0.5 = 0.739$  on the validation set, representing a +3.8% improvement over the YOLOv8s default configuration baseline (640 × 640,  $mAP@0.5 = 0.701$ ). Among individual categories, patches achieve the highest AP of 0.923 and scratches reach 0.868, while the optimal F1 score is 0.69 at a confidence threshold of 0.292. Confusion matrix and visualization analyses reveal the detection characteristics of each defect type and indicate directions for future improvement.

## Keywords

Object Detection, YOLOv8, Steel Surface Defect, NEU-DET, Data Augmentation, Deep Learning

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

钢材表面缺陷的快速准确检测是钢铁制造业质量控制的核心环节。热轧过程中产生的裂纹、划痕、夹杂物等表面缺陷若未被及时发现，将直接影响钢材的力学性能和使用寿命，严重时可引发安全事故。传统人工目检方式效率低下、受主观因素影响大，难以满足高速生产线的稳定检测需求[1]。

近年来，基于深度学习的目标检测方法以其强大的特征自动提取能力，已成为工业表面缺陷检测领域的主流技术路线。YOLO (You Only Look Once) 系列单阶段检测器凭借端到端检测框架在实时性和精度之间取得了良好平衡，被广泛应用于工业质检场景[2]。Ultralytics 于 2023 年发布的 YOLOv8 引入 C2f 骨干模块、解耦检测头与 Anchor-Free 策略，在多个标准基准上取得了优异成绩，是目前工业实时检测的优选模型之一[3]。

本文选用轻量级的 YOLOv8s 版本作为基础网络，以东北大学热轧带钢表面缺陷公开数据集(NEU-DET)为实验平台，针对 NEU-DET 数据集对 YOLOv8s 进行系统的优化与评估，通过高分辨率输入、多策略数据增强等组合优化策略构建强有力的基准检测模型，量化组合优化策略对 6 类缺陷检测性能的实际增益，并与相关工作中的 SOTA 方法进行客观比较，为后续研究提供可复现的基准参考。

## 2. 相关工作

### 2.1. 钢材表面缺陷检测方法研究

赵朗月和吴一全[1]进一步综述了基于机器视觉的表面缺陷检测方法研究进展，总结了 NEU-DET 等公开数据集在不同方法上的性能对比，为后续研究提供了全面参考。

针对带钢表面缺陷检测的专项研究持续深入。王延舒和余建波[4]提出基于自适应全局定位网络 (AGLNet) 的带钢表面缺陷检测方法，引入自适应树型候选框提取网络和全局定位回归算法，在 NEU-DET 上  $mAP$  达 79.90%，检测速度 11.8 帧/秒。卢俊哲等[5]提出面向复杂背景带钢缺陷检测的轻量级 DCN-YOLO，在保持高精度的同时显著减少模型参数量。梁礼明等[6]设计改进轻量化 VTG-YOLOv7-tiny 模型，

在 NEU-DET 上取得了良好的精度与速度平衡。

## 2.2. YOLO 系列检测器在缺陷检测中的应用

YOLO 系列检测器因其实时性强、部署便捷的特点被广泛应用于钢材缺陷检测。徐彦威等[2]对 YOLO 系列目标检测算法进行了系统综述,梳理了从 YOLOv1 到 YOLOv8 的演进路线,分析了各版本的核心创新及在工业检测中的适用性。王春梅和刘欢[7]提出 YOLOv8-VSC 轻量化带钢缺陷检测算法,通过引入可分离卷积和通道注意力机制,在 NEU-DET 数据集上取得了良好的精度 - 效率平衡。朱成杰等[8]提出基于 YOLOv8-NFMC 的带钢表面缺陷检测算法,改进了特征融合模块并引入多尺度上下文感知机制,进一步提升了小目标缺陷的检测能力。刘毅和蒋三新[9]基于改进 YOLOX 在 NEU-DET 数据集上的钢材缺陷检测研究表明,结合注意力机制的改进方法能有效降低复杂纹理背景下的误检率。

## 3. NEU-DET 数据集

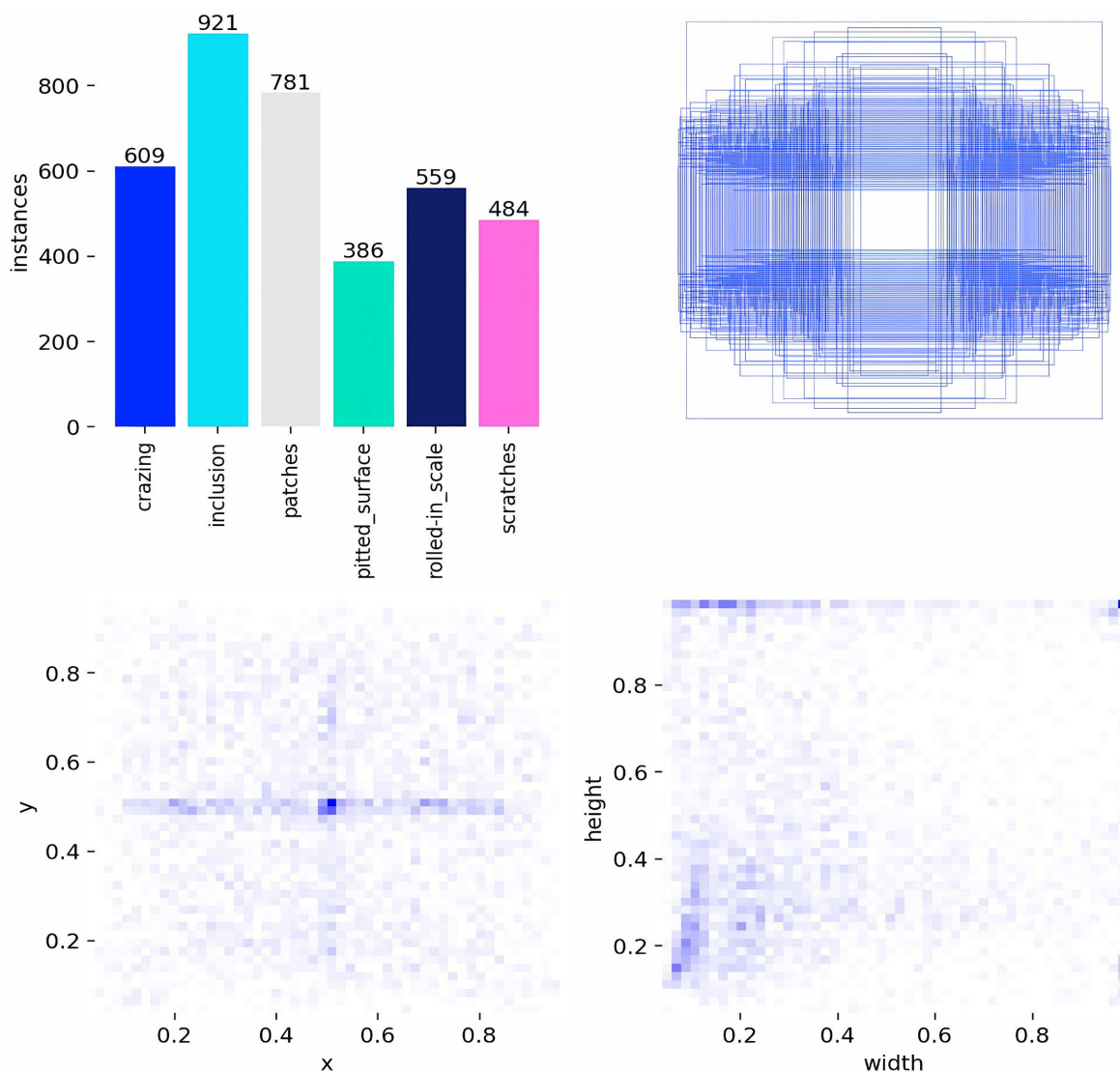


Figure 1. Instance count statistics and bounding box distribution of each category in NEU-DET dataset

图 1. NEU-DET 数据集各类别实例数量统计及边界框分布

NEU-DET (Northeastern University Surface Defect Database)是钢材表面缺陷检测领域最具代表性的公开基准数据集之一,由东北大学宋克臣课题组建立并公开发布。数据集收录热轧带钢表面 6 类典型缺陷,每类 300 张灰度图像,原始分辨率  $200 \times 200$  像素,共计 1800 张,所有图像均配有矩形边界框(bounding box)标注,适用于目标检测任务。

6 类缺陷的物理成因与视觉特征各异:① 裂纹(crazing)因热轧受力不均产生,呈细密网状分布,与正常粗糙表面纹理极为相似,检测难度最大(训练集 609 例);② 夹杂物(inclusion)为轧制混入的非金属杂质,呈点状或条状深色区域,样本量最多(921 例);③ 斑块(patches)为大面积明暗不均匀区域,面积大、对比强,检测难度最低(781 例);④ 麻面(pitted\_surface)为密集分布的微小孔坑,样本量最少(386 例);⑤ 轧入氧化皮(rolled-in\_scale)为氧化皮压入钢板的印痕,纹理平滑,与正常表面相似(559 例);⑥ 划痕(scratches)为细长线形刮痕,形态独特(484 例)。

按 8:2 比例划分训练集(1440 张)与验证集(360 张)。图 1 展示了各类缺陷实例数量统计及边界框尺寸/位置分布,可见存在一定程度的类别不平衡,训练中需关注少数类的检测性能。

## 4. 方法

### 4.1. YOLOv8s 网络架构

本文选用 YOLOv8s (小型版)作为基础检测器。相较于 YOLOv8m/l, YOLOv8s 参数量更少、推理延迟更低,更适合工业边缘端实时部署,同时在充分训练后仍能保持竞争性的检测精度[3]。YOLOv8s 由三大模块构成:

(1) 骨干网络(Backbone): 采用改进 CSPDarknet, 核心为 C2f (Cross Stage Partial with 2 Convolutions) 模块。C2f 将输入特征图分为主路和短接路并行处理,主路经多个 Bottleneck 层提取深层语义特征,短接路直接连接,最终 Concat 融合,在减少冗余计算的同时保留完整梯度信息流。

(2) 特征聚合颈部(Neck): 采用 PA-FPN (Path Aggregation Feature Pyramid Network)双向特征融合结构,自顶向下传递深层语义信息,自底向上传递浅层位置细节,两次双向融合充分整合多尺度特征,对大小各异的缺陷目标检测具有重要作用[6]。

(3) 检测头(Head): 采用解耦检测头将分类与定位任务独立优化,配合 Anchor-Free 策略直接预测目标中心坐标与宽高,无需预设锚框,有效降低超参数调优难度。

### 4.2. 训练策略设计

针对 NEU-DET 数据集小样本、类别不平衡的特点,本文从四个维度优化训练策略:

(1) 高分辨率输入: 将输入图像从默认  $640 \times 640$  提升至  $800 \times 800$  像素,保留更丰富的缺陷细节纹理信息,改善对裂纹、麻面等细小特征的感知能力。

(2) AdamW 优化器: 采用 AdamW 优化器,初始学习率  $lr_0 = 0.001$ ,最终衰减系数  $lrf = 0.01$ ,动量 0.937,权重衰减 0.0005,配合余弦退火学习率调度策略实现平滑收敛。

(3) 多策略数据增强: 综合使用 Mosaic 增强(4 图拼接,概率 1.0)、MixUp 图像混合(概率 0.1)、CopyPaste 目标复制粘贴(概率 0.1),以及 HSV 颜色扰动( $h: 0.015, s: 0.3, v: 0.4$ )、随机旋转( $\pm 10$  度)、水平/垂直翻转等增强操作,有效扩充训练样本多样性,缓解过拟合与类别不平衡问题。

(4) 训练调度: 前 3 个 epoch 执行线性学习率预热;设置  $patience = 50$  的早停机制防止无效过拟合;开启混合精度训练(AMP)加速计算并降低显存占用。

主要超参数配置见表 1。

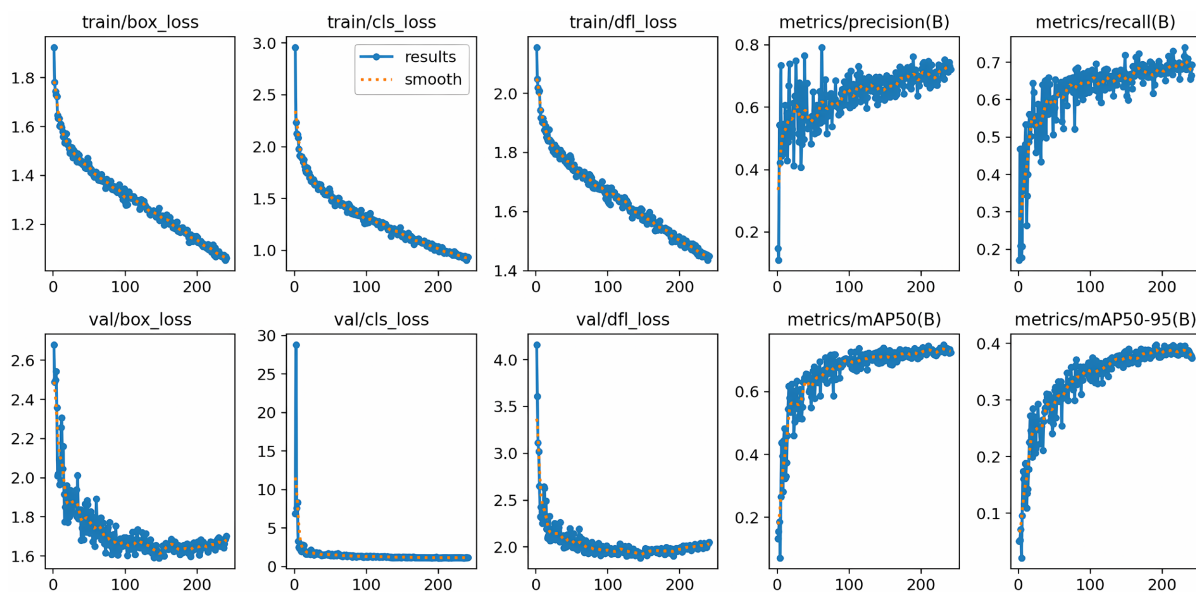
**Table 1.** Main training hyperparameter configuration  
**表 1.** 主要训练超参数配置

参数名称	参数值	说明
基础模型	YOLOv8s	轻量级小型版本
输入分辨率	800 × 800	提升细节感知能力
最大训练轮数	300	含早停 patience = 50
批次大小	32	单 GPU
优化器	AdamW	解耦权重衰减
初始学习率	0.001	余弦退火衰减
权重衰减	0.0005	L2 正则化
Mosaic 增强	1.0	全程开启
MixUp 增强	0.1	随机图像混合
CopyPaste 增强	0.1	目标复制粘贴
NMS IoU 阈值	0.7	后处理去重叠框

## 5. 实验结果与分析

### 5.1. 实验环境

实验在 AutoDL 云计算平台进行，GPU 为 NVIDIA RTX 系列，深度学习框架为 PyTorch，检测框架基于 Ultralytics YOLOv8 开源库，操作系统 Ubuntu 22.04。开启混合精度训练(AMP)以降低显存占用并加速训练。



**Figure 2.** Training curves of loss terms and key evaluation metrics

**图 2.** 训练过程各损失项及关键评估指标变化曲线

## 5.2. 评估指标

采用目标检测标准评估指标：精确率  $P = TP/(TP + FP)$ ；召回率  $R = TP/(TP + FN)$ ；F1 为精确率与召回率的调和均值；AP@0.5 为 IoU 阈值 0.5 时精确率 - 召回率曲线下面积；mAP@0.5 为各类 AP@0.5 均值；mAP@0.5:0.95 为 IoU 从 0.5 到 0.95 (步长 0.05) 的平均 AP 均值。

## 5.3. 训练过程分析

图 2 展示了训练集与验证集各损失项(box\_loss、cls\_loss、dfl\_loss)及关键评估指标随训练轮次的变化曲线。三类损失均平稳下降，训练集与验证集损失变化趋势一致，未出现明显过拟合现象，说明多策略数据增强有效缓解了小样本条件下的过拟合问题。验证集 mAP@0.5 在约 100 至 150 轮后进入平台期，最终趋于稳定；mAP@0.5:0.95 收敛于 0.39 至 0.40，表明模型具备较好的目标定位精度。

## 5.4. 整体检测性能

表 2 汇总了本文 YOLOv8s 优化方法与默认配置基线在 NEU-DET 验证集上的整体检测性能对比(见表 2)。相比 YOLOv8s 默认配置(640 × 640)，组合优化策略在 mAP@0.5 上取得 +3.8% 的增益(0.701 → 0.739)，验证了高分辨率输入与多策略数据增强组合的有效性。

**Table 2.** Overall detection performance on NEU-DET validation set  
**表 2.** NEU-DET 验证集整体检测性能

方法	Precision	Recall	mAP@0.5	mAP@0.5:0.95
YOLOv8s 基线(640 × 640, 默认配置)	≈0.72	≈0.65	0.701	≈0.36
本文(YOLOv8s 优化, 800 × 800)	≈0.75	≈0.69	0.739	≈0.39

为进一步客观评估本文方法的研究水平，表 3 将本文方法与相关工作中在 NEU-DET 数据集上报告结果的代表性 SOTA 方法进行了性能对比。

**Table 3.** Performance comparison with SOTA methods on NEU-DET dataset  
**表 3.** NEU-DET 数据集上 SOTA 方法性能对比

方法	mAP@0.5	骨干网络	来源
AGLNet [4]	0.799	AGLNet	自动化学报 2024
VTG-YOLOv7-tiny [6]	0.821	YOLOv7-tiny	光学精密工程 2024
YOLOv8-VSC [7]	0.764	YOLOv8	计算机科学与探索 2024
YOLOv8-NFMC [8]	0.782	YOLOv8	国外电子测量技术 2024
本文(YOLOv8s 优化)	<b>0.739</b>	<b>YOLOv8s</b>	<b>本文</b>

由表 3 可见，本文方法(mAP@0.5 = 0.739)与采用结构改进的 SOTA 方法相比仍存在一定差距(AGLNet: 0.799, VTG-YOLOv7-tiny: 0.821)，主要原因在于本文以轻量级 YOLOv8s 为基础，未引入注意力机制或特征融合改进，侧重于训练策略优化。但相比同样基于 YOLOv8 的方法(YOLOv8-VSC: 0.764)，本文以更简洁的配置验证了组合优化策略的有效性，为后续网络结构改进提供了可靠的基准参考。

图 3 展示了各类别精确率 - 召回率(PR)曲线。斑块(patchs)类曲线面积最大(AP = 0.923)，在整个召回率区间均保持较高精确率；划痕(scratches)类 AP 为 0.868，曲线形态良好；裂纹(crazing)类曲线在高召回率段迅速下降，AP 仅 0.477，反映该类缺陷与正常表面纹理高度相似导致的检测困难。

图 4 为 F1-置信度曲线。全类 F1 在置信度阈值 0.292 时达最优值 0.69，该阈值可作为实际部署的推荐置信度设定值，实际应用中可根据业务对误检与漏检的容忍度灵活调整。

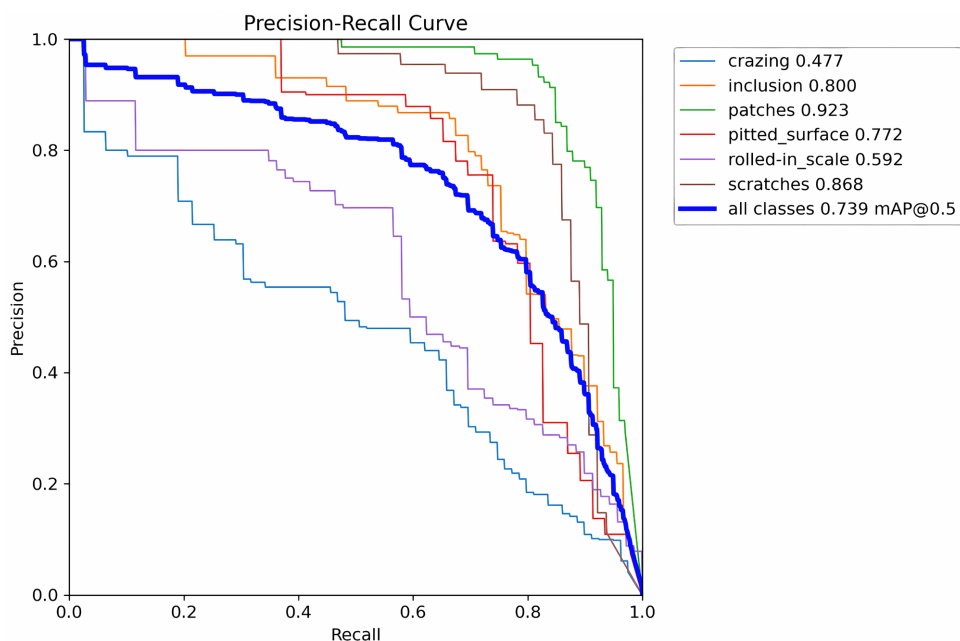


Figure 3. Precision-Recall (PR) curves for each defect category

图 3. 各类别精确率 - 召回率(PR)曲线

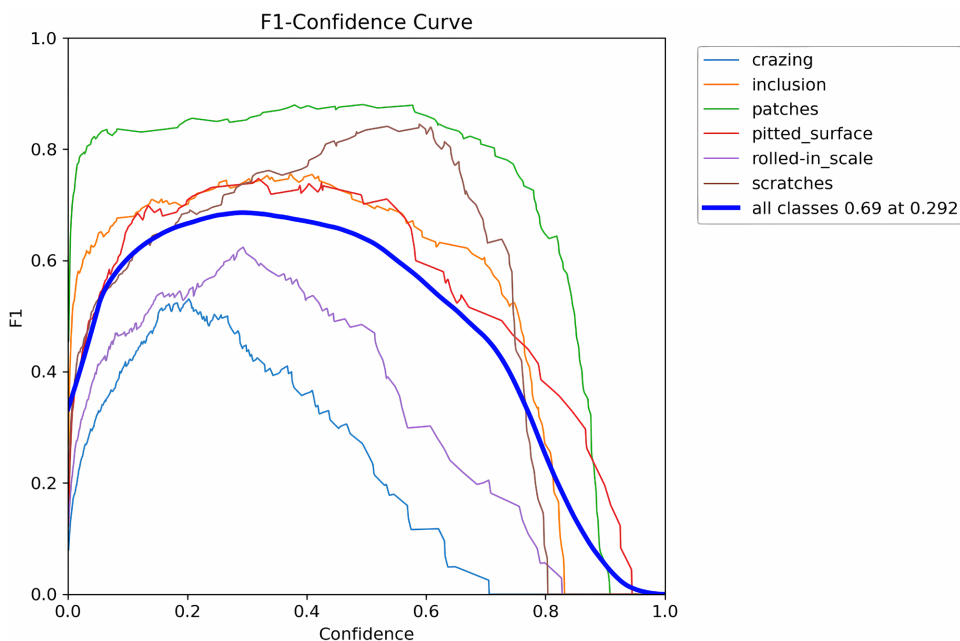


Figure 4. F1-confidence curves for each defect category

图 4. 各类别 F1-置信度曲线

### 5.5. 各类别检测性能详细分析

表 4 列出 6 类缺陷的 AP@0.5 及归一化混淆矩阵对角线正确检测率。

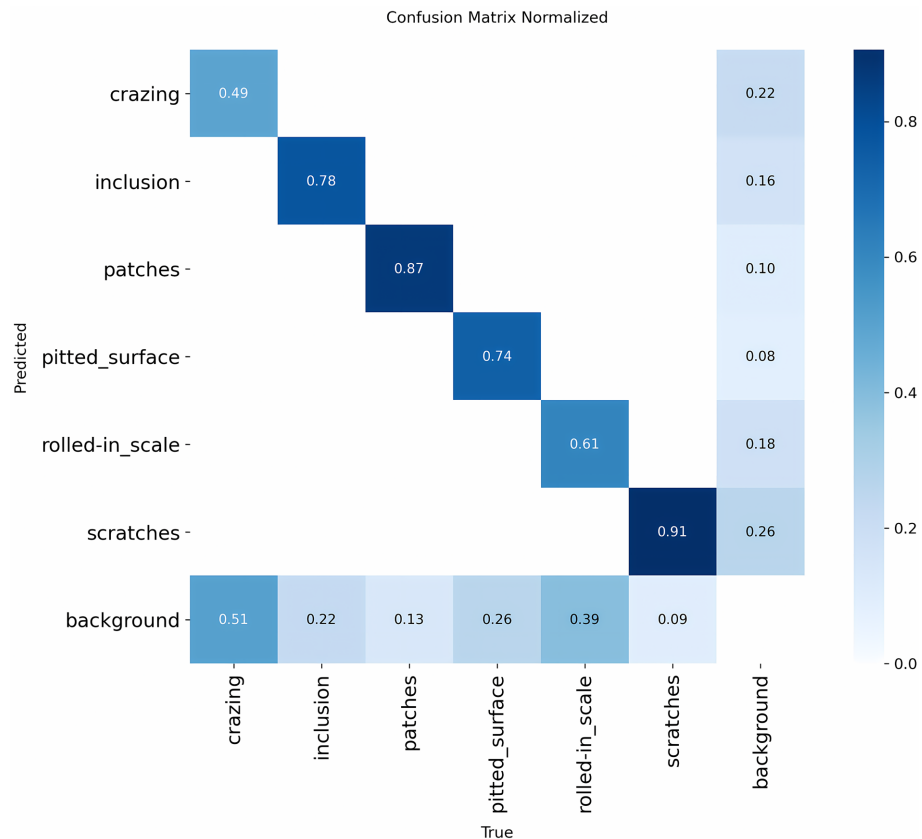
**Table 4.** Detailed detection performance results for each defect category

**表 4.** 各类别缺陷检测性能详细结果

缺陷类别	AP@0.5	正确率	漏检率	主要特征难点
裂纹 crazing	0.477	0.49	0.51	低对比度网状纹理
夹杂物 inclusion	0.800	0.78	0.22	点状特征相对明显
斑块 patches	0.923	0.87	0.13	面积大对比强
麻面 pitted surface	0.772	0.74	0.26	小坑点背景易混淆
轧入氧化皮	0.592	0.61	0.39	纹理与正常表面相近
划痕 scratches	0.868	0.91	0.09	线形形态独特

### 5.6. 混淆矩阵分析

图 5 展示了归一化混淆矩阵。对角线元素为各类别正确检测比例，最后一列表示真实样本被误判为背景的漏检比例。



**Figure 5.** Normalized confusion matrix

**图 5.** 归一化混淆矩阵

主要发现如下：(1) 裂纹(crazing)类正确率仅 0.49，漏检率高达 0.51，且有 22%的背景被误检为裂纹(高假阳性)，根本原因在于裂纹网状纹理与钢板正常粗糙表面极为相似，低对比度使卷积特征提取器难以建立稳定的语义表征；(2) 轧入氧化皮(rolled-in\_scale)类漏检率 0.39，该类缺陷区域纹理过渡平滑、边界模糊，且训练样本量相对有限，模型对其特征表征不充分；(3) 斑块和划痕类正确率最高(分别为 0.87 和 0.91)，与其鲜明的视觉特征——大面积明暗对比和独特线形形态直接相关；(4) 误检模式以漏检为主，类别间互相混淆较少，表明模型分类能力基本可靠，主要瓶颈在于低置信度导致的漏检。

## 6. 结论

本文针对 NEU-DET 数据集对 YOLOv8s 进行了系统的优化与评估，构建了强有力的基准检测模型。通过高分辨率输入(800 × 800)、AdamW 优化器与 Mosaic/MixUp/CopyPaste 多策略数据增强的组合优化，相比默认配置基线(mAP@0.5 = 0.701)，模型在验证集上取得 mAP@0.5 = 0.739 (+3.8%)的整体性能提升，量化验证了组合优化策略的实际增益。斑块(patches)和划痕(scratches)类缺陷检测效果最佳(AP 分别达 0.923 和 0.868)，视觉特征突出是其优良检测性能的根本原因；裂纹(crazing)和轧入氧化皮(rolled-in\_scale)类因低对比度纹理特征仍是主要检测瓶颈(AP 分别为 0.477 和 0.592)。与 NEU-DET 上的 SOTA 方法对比，本文以轻量级骨干和纯训练策略优化实现了有竞争力的基准性能，为后续结构改进研究提供可复现的参考基线。混淆矩阵分析表明漏检为主要失误模式，推荐部署置信度阈值 0.292 (F1 = 0.69)。

后续工作将以本文基准模型为起点，围绕以下方向展开：① 引入通道-空间注意力机制(如 CBAM)或 Transformer 模块，增强模型对低对比度缺陷特征的感知，以期超越当前 SOTA 方法；② 针对裂纹、轧入氧化皮类设计专项数据增强或合成样本生成策略，缓解漏检问题；③ 探索知识蒸馏压缩方案以进一步降低模型部署开销；④ 利用实际工业生产数据进行领域自适应微调，提升模型在真实场景中的泛化能力。

## 参考文献

- [1] 赵朗月, 吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. 仪器仪表学报, 2022, 43(1): 198-219.
- [2] 徐彦威, 李军, 董元方, 等. YOLO 系列目标检测算法综述[J]. 计算机科学与探索, 2024, 18(9): 2221-2238.
- [3] Jocher, G., Chaurasia, A. and Qiu, J. (2023) Ultralytics YOLO (Version 8.0.0). <https://github.com/ultralytics/ultralytics>
- [4] 王延舒, 余建波. 基于自适应全局定位算法的带钢表面缺陷检测[J]. 自动化学报, 2024, 50(8): 1550-1564.
- [5] 卢俊哲, 张铖怡, 刘世鹏, 等. 面向复杂环境中带钢表面缺陷检测的轻量级 DCN-YOLO [J]. 计算机工程与应用, 2023, 59(15): 318-328.
- [6] 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.
- [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] 朱成杰, 刘乐乐, 朱洪波. 基于 YOLOv8-NFMC 的带钢表面缺陷检测算法[J]. 国外电子测量技术, 2024, 43(7): 97-104.
- [9] 刘毅, 蒋三新. 基于改进 YOLOX 的钢材表面缺陷检测研究[J]. 现代电子技术, 2024, 47(9): 131-138.