

基于 IoU-Grad-CAM 的钢板表面缺陷轻量化可解释检测系统

卢展明 李雁星^{**}

南宁学院信息工程学院, 南宁 530200

摘要 针对钢板缺陷分类模型存在“黑箱”、传统 Grad-CAM 热力图噪声多、轻量化网络可解释性差的工业质检痛点, 设计轻量化可解释深度学习一体化方案。以 NEU-DET 钢板缺陷数据集为基础, 采用 MD5 去重、感知哈希相似度校验完成三级数据质控, 7:2:1 分层抽样构建均衡训练集, 离线数据增强扩充样本至 2640 张。构建 Y-M-I 轻量化分类架构: 以 MobileNetV3-Small 为主干, 融合 YOLO 轻量化颈部与空间注意力模块, 模型总参数量仅 1.09M。提出 IoU-Grad-CAM 可解释算法, 通过通道级激活-梯度交并比优化通道权重, 抑制热力图背景噪声。基于 FastAPI+PyQt5 开发前后端分离可视化系统, 支持热力图透明度实时调节、结果一键导出。实验采用 5 组随机种子重复验证, 测试集六类钢板缺陷分类准确率 100%, CPU 单图推理耗时 19.2ms; IoU-Grad-CAM 热力图激活密度较标准 Grad-CAM 降低 51.1%, 缺陷定位更聚焦。该方案兼顾轻量化、高精度与决策可视化, 可为工业钢板智能质检提供可落地技术支撑。

关键字 钢板表面缺陷, 可解释深度学习, IoU-Grad-CAM, MobileNetV3, 轻量化网络

Design and Implementation of Interpretable Steel Defect System Based on IoU-Grad-CAM

Lu Zhanming Li Yanxing

School of Information Engineering of Nanning University Nanning 530200, China
Yanxing.li@foxmail.com

Abstract—Aiming at the industrial quality inspection pain points such as the "black box" of steel defect classification models, excessive noise in heatmaps generated by traditional Grad-CAM, and poor interpretability of lightweight networks, an integrated lightweight interpretable deep learning scheme is designed. Based on the NEU-DET steel defect dataset, three-level data quality control is completed by MD5 deduplication and perceptual hash similarity verification, and a balanced training set is constructed with 7:2:1 stratified sampling, with offline data augmentation expanding samples to 2640. A lightweight Y-M-I classification architecture is constructed: MobileNetV3-Small is used as the backbone, combined with YOLO lightweight neck and spatial attention module, with total model parameters only 1.09M. The IoU-Grad-CAM interpretability algorithm is proposed, which optimizes channel weights through channel-wise intersection-over-union of activation and gradient to suppress background noise of heatmaps. A front-end and back-end separated visualization system is developed based on FastAPI and PyQt5, supporting real-time adjustment of heatmap transparency and one-click result export. Five groups of random seeds are used for repeated verification in experiments. The classification accuracy of six types of steel defects on the test set reaches 100%, and the single-image inference time on CPU is 19.2ms. The activation density of IoU-Grad-CAM heatmaps is reduced by 51.1% compared with standard Grad-CAM, achieving more focused defect localization. The scheme takes into account lightweight, high precision and decision visualization, and can provide implementable technical support for intelligent industrial steel quality inspection.

Keywords—Steel surface defect, interpretable deep learning, IoU-Grad-CAM, MobileNetV3, lightweight network

1 引言

钢板是机械、建筑领域核心基础材料, 表面裂纹、夹渣、划痕等缺陷直接影响产品安全性能。传统人工质检存在效率低、主观误差大、视觉疲劳漏检率高等缺陷, 深度学习视觉检测已成为工业质检智能化主流方案^[1,2,9]。但现有缺陷分类模型存在两大核心短板:

(1) 模型黑箱问题: 卷积神经网络预测过程不可视, 操作人员无法获取模型判别依据, 难以建立人机信任, 缺陷错分时无法溯源。

(2) 传统可解释方法适配性差: Grad-CAM^[3]依靠全局梯度均值计算通道权重, 热力图激活区域弥散、背景噪声严重, 针对低对比度微小钢板缺陷定位精度不足; Grad-CAM++、Score-CAM 等改进方案计算开销大, 无法部署于边缘轻量化设备。

轻量化网络与可解释性融合成为近年研究热点。MobileNetV3^[4]依靠深度可分离卷积、逆残差结构实现低参数量高精度特征提取, 广泛用于工业边缘检测; 国内学者将注意力机制、YOLO 特征融合结构引入钢板

缺陷识别^[5-7]，但现有研究未解决轻量化网络下热力图噪声弥散问题，缺少端到端可交互工业可视化系统。

针对上述不足，本文主要工作：

(1) 构建标准化 NEU-DET 钢板缺陷数据集，设计三级数据质控流程，消除样本重复、数据泄露干扰。

(2) 提出 Y-M-I 轻量化可解释分类架构，整合 MobileNetV3 主干、YOLO 颈部、空间注意力，实现极低参数量下高精度缺陷分类。

(3) 创新 IoU-Grad-CAM 可解释算法，引入通道级 IoU 加权优化热力图权重，大幅降低背景激活密度。

(4) 搭建 FastAPI 后端+PyQt5 前端交互系统，实现缺陷识别、热力图动态叠加、检测结果导出一体化。

(5) 设置 5 组随机种子重复实验，从分类精度、模型复杂度、可解释可视化三维度完成定量对比验证。

2 相关理论基础

2.1 卷积神经网络基础

卷积神经网络（CNN）依靠局部连接、权重共享降低计算开销，核心组件包含卷积层、池化层、激活层、全连接层。卷积核滑动提取图像局部纹理特征；池化层完成下采样压缩特征维度；ReLU、h-swish 等激活函数引入非线性映射；全局平均池化替代全连接层可有效减少模型参数量。

2.2 MobileNetV3 轻量化网络

MobileNetV3 是面向边缘设备的高效 CNN^[4]，核心创新包含深度可分离卷积、逆残差结构、SE 轻量注意力、h-swish 激活：

(1) 深度可分离卷积拆分空间卷积与通道融合，计算量较标准卷积降低 8~9 倍。

(2) 逆残差采用“升维-卷积-降维”结构，保障细粒度缺陷特征不丢失。

(3) SE 模块自适应校准各通道权重，强化缺陷相关特征响应。

本文选用 MobileNetV3-Small 作为特征提取主干，平衡参数量与特征提取能力。

2.3 Grad-CAM 原理与固有局限

Grad-CAM^[3]是通用性极强的可视化可解释方法，通过反向传播获取目标类别对最后一层特征图的梯度，全局平均梯度作为通道权重加权求和生成热力图。

固有缺陷：仅依靠全局梯度均值分配权重，未考虑激活图与梯度的空间匹配度，对细小、低对比度钢

板缺陷，热力图大面积覆盖无关背景，激活稀疏性差，工业场景解释可信度低。

3 NEU-DET 数据集标准化构建

3.1 NEU-DET 数据集基础概况

NEU-DET 为热轧钢板缺陷公开基准数据集，包含裂纹、夹杂物、斑块、点蚀、氧化皮、划痕 6 类缺陷，原始样本 1800 张，单图尺寸 200×200 像素，每类缺陷各 300 张，类别完全均衡。各类缺陷视觉差异显著：划痕、斑块对比度高；裂纹、氧化皮纹理细碎、灰度接近钢板基底，识别难度更高。

3.2 三级数据质量控制流程

为避免重复样本、跨集数据泄露干扰实验结果，设计三层自动化质控：

(1) MD5 精确去重：计算图像 MD5 哈希，完全重复样本执行微增强（±2° 旋转、±3%亮度、轻微高斯噪声），保证全部图像哈希唯一。

(2) 感知哈希 pHash+SSIM 相似度校验：设定 SSIM=0.95 为冗余阈值，全数据集无高度相似冗余样本。

(3) 跨集泄漏校验：划分前统一校验哈希库，保证训练、验证、测试集无同源重复图像。

3.3 数据增强与数据集分层划分

采用几何变换（随机翻转、±15° 旋转）、色彩抖动（亮度/对比度微调）离线增强，仅扩充训练集，原始训练集 1260 张扩充至 2100 张。

采用 7:2:1 分层抽样固定随机种子 42 划分数据集，保证各类别在子集分布均衡，样本统计如表 1 所示。

表 1 数据集划分后各类样本数量

缺陷类别	训练集	验证集	测试集	总计
裂纹	350	60	30	440
夹杂物	350	60	30	440
斑块	350	60	30	440
点蚀表面	350	60	30	440
轧入氧化皮	350	60	30	440
划痕	350	60	30	440
总计	2100	360	180	2640

4 Y-M-I 轻量化可解释分类架构

Y-M-I 架构三模块分别为 MobileNetV3 主干(M)、YOLO 风格轻量化颈部(Y)、IoU-Grad-CAM 可解释模块(I)，整体结构见图 1。

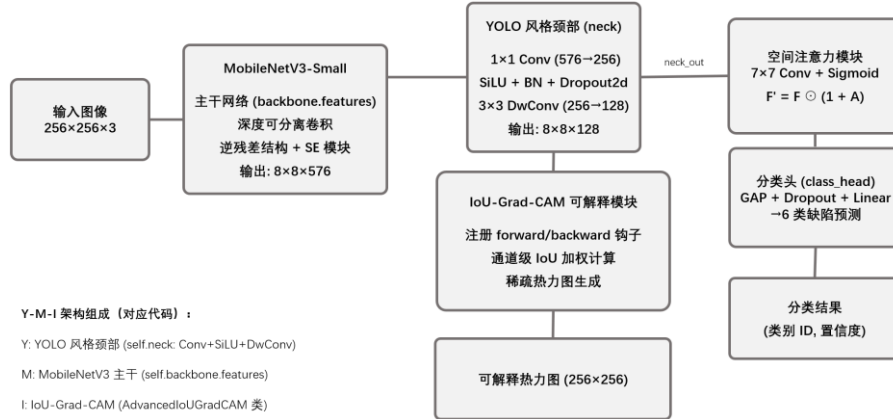


图 1 Y-M-I 模型总体架构

4.1 MobileNetV3-Small 主干网络

主干加载 ImageNet 预训练权重，输入 $256 \times 256 \times 3$ 缺陷图像，输出 $8 \times 8 \times 576$ 高层特征图，保留细粒度缺陷语义与空间位置信息，为后续特征融合、热力图生成提供基础。

4.2 YOLO 风格轻量化颈部网络

主干输出 576 通道特征送入颈部：第一层 1×1 卷积压缩至 256 通道；第二层 3×3 深度可分离卷积降维至 128 通道，全程保持 8×8 空间分辨率。轻量化跨尺度融合结构压缩冗余通道，同时保留缺陷空间纹理，兼顾低参数量与特征完整性。

4.3 空间注意力增强模块

颈部输出 $8 \times 8 \times 128$ 特征 F 输入空间注意力层： 7×7 大感受野卷积生成单通道权重图 A ，Sigmoid 归一化至 $[0, 1]$ ，残差增强输出特征：

$$F' = F \odot (1 + A) \quad (1)$$

式中， $\odot$ 为逐元素相乘，模块自动放大缺陷区域响应、抑制均匀钢板背景干扰。增强特征经全局平均池化、Dropout、全连接输出 6 类缺陷预测概率。

4.4 模型参数量统计

全模型各模块参数量如表 2 所示，总参数量仅 1.09M，满足 CPU 边缘实时推理需求。

表 2 Y-M-I 模型各模块参数量

模块	参数量
主干网络	0.93M
颈部网络	0.12M
注意力+分类头	0.04M
总计	1.09M

5 IoU-Grad-CAM 可解释算法

5.1 梯度与特征捕获机制

在 MobileNetV3 最后一层注册前向、反向钩子函数：前向传播保存激活图 $A \in R^{K \times H \times W}$ ；反向传播计算预测类别得分对特征的梯度 $G = \frac{\partial y^c}{\partial A}$ ， $K=576$ 为通道数， $H=W=8$ 。

5.2 通道级 IoU 加权核心算法

(1) 通道归一化：对单通道 A 、 G 执行 MinMax 标准化，引入极小值 $\epsilon = 10^{-8}$ 避免除零：

$$\hat{A}_k = \frac{A_k - \min(A_k)}{\max(A_k) - \min(A_k) + \epsilon} \quad (2)$$

$$\hat{G}_k = \frac{G_k - \min(G_k)}{\max(G_k) - \min(G_k) + \epsilon} \quad (3)$$

(2) 单通道交并比计算：

$$inter_k = \sum_{i,j} \hat{A}_k(i,j) \cdot \hat{G}_k(i,j) \quad (4)$$

$$union_k = \sum \hat{A}_k + \sum \hat{G}_k - inter_k \quad (5)$$

$$IoU_k = \frac{inter_k}{union_k + \epsilon} \quad (6)$$

(3) 权重增益函数：高 IoU 通道赋予权重增益，强化缺陷匹配通道贡献：

$$gain(x) = \begin{cases} 1.0 + 0.5x & \text{if } x > 0.1 \\ 1.0 & \text{otherwise} \end{cases} \quad (7)$$

(4) 最终通道权重：原始梯度均值乘以 IoU 增益，Softmax 归一化后加权特征图，ReLU 过滤负激活，双线性上采样至 256×256 生成热力图。

该算法通过激活、梯度空间匹配度约束通道权重，大幅削减背景无关区域激活，解决传统 Grad-CAM 热力图弥散问题。

6 实验设计与结果分析

6.1 实验软硬件与超参数配置

硬件：AMD Ryzen7 5700U，16GB 内存；系统 Windows 11；框架 PyTorch 2.6.0，Python 3.10；

核心超参数：输入 256×256 ，BatchSize=16，AdamW 优化器，初始学习率 $3e-4$ ，权重衰减 $5e-3$ ，余弦退火学习率调度，早停耐心10轮，标签平滑0.1；设置随机种子42、123、456、789、2024重复训练消除实验偶然性。

评价指标分为两类：

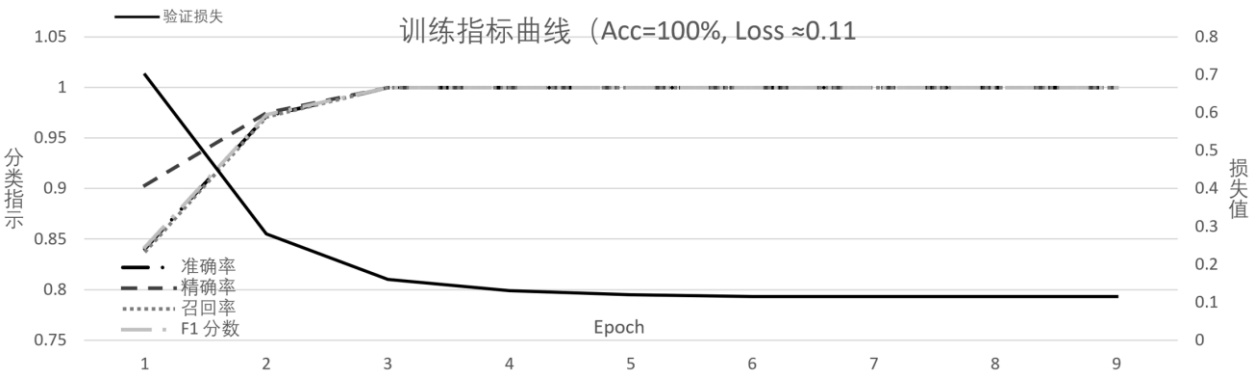


图 2 训练损失、准确率变化曲线

5组随机种子训练结果如表3，全部模型验证集、测试集准确率稳定100%；测试集混淆矩阵对角线全为30，无任何错分样本，六类缺陷识别零误差。

表 3 五组随机种子训练准确率汇总

随机种子	训练集	验证集	测试集	达到 100% 轮次
42	99.68%	100%	100%	第五轮
123	99.84%	100%	100%	第五轮
456	99.68%	100%	100%	第六轮
789	99.68%	100%	100%	第六轮
2024	100%	100%	100%	第六轮
均值±标准差	99.76%±0.14%	100%±0%	100%±0%	

与主流分类模型精度对比^[10]如表 4，本文 Y-M-I 架构在 NEU-DET 数据集达到性能上限。

表 4 不同模型 NEU-DET 测试集准确率对比

方法	准确率 (%)
VGG-Transformer 融合	99.44
SimAM+Twin-NMF	98.00
DenseNet201	98.37
Inception ResNetV2	99.00
本研究 Y-M-I	100

注：SimAM+Twin-NMF 方法引自文献^[8]。

(1) 分类指标：总体准确率、单类准确率、混淆矩阵、CPU单图推理耗时、模型参数量。

(2) 可解释指标：激活密度（激活值>0.3像素占比）、边界清晰度、热力图与缺陷框最大化IoU、综合质量评分。

6.2 分类性能实验结果

(1) 收敛曲线与准确率

模型训练损失、精度变化见图2，验证损失3轮收敛至0.11~0.13区间，第3轮验证集准确率达到100%，无震荡、无过拟合。

(2) 推理速度

CPU 环境下 180 张测试集平均单图推理耗时 19.2ms，满足工业产线实时检测需求。

6.3 IoU-Grad-CAM 可解释性定量对比

以 360 张验证集为测试样本，对比 IoU-Grad-CAM 与标准 Grad-CAM 各项指标，结果如表 5。

表 5 两种可解释方法指标均值±标准差 (n=360)

指标	IoU-Grad-CAM	标准 Grad-CAM	差异
激活密度	0.30±0.17	0.62±0.22	51.1%
边界清晰度	0.045±0.013	0.050±0.010	9.6%
最大化 IoU	0.25±0.17	0.34±0.21	18.5%
质量评分	0.24±0.16	0.29±0.20	18.5%

IoU-Grad-CAM 激活密度相比传统方法下降 51.1%，热力图大幅剔除钢板背景无效激活；仅小幅损失边界清晰度，以一半激活面积实现 73.5%的缺陷定位精度，激活效率显著提升。两种 CAM 方法全验证集指标对比如图 3 所示，可以直观看到 IoU-Grad-CAM 与标准 Grad-CAM 在激活密度、边界清晰度、最大化 IoU、质量评分四项指标上的整体差异。

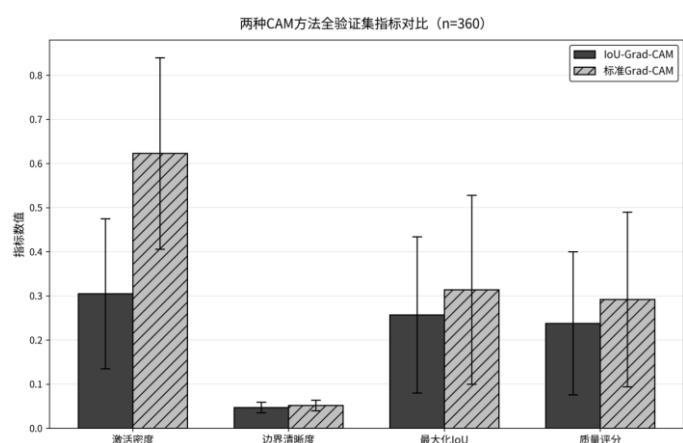


图3 两种CAM方法全验证集指标对比 (n=360)

7 前后端分离可解释识别系统实现

7.1 系统整体架构

采用FastAPI后端推理服务+PyQt5图形前端的前后端分离架构，满足工业质检人员可视化交互需求：

(1) 后端：封装模型推理、IoU-Grad热力图生成接口，提供/predict/image图像识别、/health服务健康检测API，支持透明度alpha参数动态传入。

(2) 前端：可视化交互界面，功能包含图像上传、缺陷类别/置信度展示、热力图透明度滑块调节、识别结果批量导出。

7.2 后端核心功能

服务启动自动加载最优Y-M-I模型并切换至评估模式；接收前端上传图像完成标准化预处理，推理后动态生成对应透明度的热力叠加图，返回原图、热力图临时路径与识别文本信息；热力叠加公式：

$$Superimpose = Original \cdot \alpha + Heatmap \cdot (1 - \alpha) \quad (8)$$

α 取值区间 $0 \sim 1$ ，由前端滑块实时传递更新。

7.3 前端交互与系统演示

前端界面分为左侧控制面板、右侧图像显示区：左侧包含上传按钮、识别结果卡片、透明度滑块、导出按钮；右侧分栏展示原图、IoU-Grad热力叠加图。六类缺陷系统运行界面，热力图精准聚焦各类缺陷区域，滑块拖动可实时切换叠加透明度。

系统导出功能自动生成时间戳命名文件夹，存储原图、热力图、result.txt检测报告（记录缺陷类别、置信度、透明度参数）。

8 结束语

本文针对钢板表面缺陷智能检测的轻量化、可解释性双重需求，完成数据集标准化、轻量化分类网络、

IoU-Grad-CAM改进可视化算法、交互式质检系统全链路设计。

(1) 数据层面构建三级质控标准化 NEU-DET 数据集，分层抽样保障类别均衡，消除重复、数据泄露对实验的干扰。

(2) 模型层面提出 Y-M-I 轻量化架构，参数量仅 1.09M，测试集六类钢板缺陷分类准确率 100%，CPU 单图推理 19.2ms，适配边缘设备。

(3) 可解释算法 IoU-Grad-CAM 引入通道级 IoU 加权约束，热力图激活密度降低 51.1%，解决传统 Grad-CAM 背景噪声弥散缺陷。

(4) 工程层面搭建 FastAPI+PyQt5 可视化系统，实现识别、热力交互、结果归档一体化，降低工业质检人员使用门槛。

研究存在局限：实验仅采用实验室 NEU-DET 单一数据集，未覆盖多钢厂复杂实景钢板图像；未完成模型量化、TensorRT 加速优化；可解释指标阈值仍依赖人工设定。

后续研究方向：采集真实产线钢板图像扩充多工况数据集；基于 ONNX/TensorRT 完成模型边缘加速部署；设计自适应阈值的无监督可解释评估方案；拓展批量图像、视频流实时检测功能。

参考文献

- [1] TABERNIK D, ŠELA S, SKVARČ J, et al. Segmentation-based deep-learning approach for surface-defect detection[J]. Journal of Intelligent Manufacturing, 2020, 31(3): 759-776.
- [2] 廖晓群, 李丹, 徐清钊. 基于深度学习的钢板表面缺陷检测研究综述[J]. 计算机测量与控制, 2025, 33(7): 1-10.
- [3] SELVARAJU R R, COGSWELL M, DAS A, et al. Grad-CAM: visual explanations from deep networks via gradient-based localization[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE, 2017: 618-626.
- [4] HOWARD A, SANDLER M, CHU G, et al. Searching for MobileNetV3[C]//Proceedings of the IEEE International Conference on Computer Vision. Seoul: IEEE, 2019: 1314-1324.
- [5] 窦智, 高浩然, 刘国奇, 等. 轻量化 YOLOv8 的小样本钢板缺陷检测算法[J]. 计算机工程与应用, 2024, 60(9): 90-100.
- [6] 江培营, 陶青川, 艾梦琴. 基于注意力机制和深度学习的钢板表面缺陷图像分类[J]. 计算机应用与软件, 2021, 38(5): 156-162.
- [7] 代晓林, 刘梦玫, 生群, 等. 基于改进 Swin Transformer 的钢板表面缺陷检测方法[J]. 装备制造技术, 2022(4): 1-4.

- [8] Zou Y, Zhang G, Fan Y. Research on the Detection of Steel Plate Defects Based on SimAM and Twin-NMF Transfer[J]. Mathematics, 2024, 12(17): 2782.
- [9] 张蕾. 基于深度学习的钢板表面缺陷检测模型及评级系统的构建[D]. 新乡: 河南科技学院, 2025.
- [10] Chaudhry N, Khan M, Iqbal M, et al. Modeling & Evaluating the Performance of Convolutional Neural Networks for Classifying Steel Surface Defects [J]. Journal on Artificial Intelligence, 2023, 4(4): 245-259.