

基于 YOLO-AMM 的轻量化带钢表面缺陷检测算法

孙艳龙 陈广锋

(东华大学机械工程学院 上海 201620)

摘 要: 针对现有带钢表面缺陷检测算法参数量大、检测精度不足及在嵌入式设备部署难等问题,提出了一种基于改进 YOLOv11n 的轻量化检测算法 YOLO-AMM。首先,引入 ADown 模块,解决传统跨步卷积下采样导致的微小缺陷特征丢失问题以提升精度,减少参数量以提升实时性;其次,构建 C3k2_MDSB 模块,增强多尺度特征提取能力和计算效率;最后,在特征融合网络中引入混合局部通道注意力(MLCA)机制,提升对不同尺度缺陷的特征提取能力。实验表明, YOLO-AMM 算法的 mAP50 较 YOLOv11n 提升 3.6%,召回率提升 2.3%,精确率提升 2.9%,同时参数减少 18.0%,计算量减少 14.1%,模型大小下降 16.5%,优于多种经典轻量级与同类改进算法,适合部署于资源受限的嵌入式设备中。

关键词: 缺陷检测;轻量化;ADown;YOLOv11n;MLCA

中图分类号: TN911;TP391.4 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Lightweight strip steel surface defect detection algorithm based on YOLO-AMM

Sun Yanlong Chen Guangfeng

(College of Mechanical Engineering, Donghua University, Shanghai 201620, China)

Abstract: To address the issues of large parameter counts, insufficient detection accuracy, and difficulty in deployment on embedded devices in existing strip steel surface defect detection algorithms, a lightweight detection algorithm named YOLO-AMM based on improved YOLOv11n is proposed. Firstly, the ADown module is introduced to solve the problem of tiny defect feature loss caused by traditional strided convolution downsampling, thereby improving accuracy and to reduce the number of parameters for enhanced real-time performance. Secondly, the C3k2_MDSB module is constructed to strengthen multi-scale feature extraction capability and computational efficiency. Finally, the mixed local channel attention (MLCA) attention mechanism is incorporated into the feature fusion network to improve the feature extraction ability for defects of different scales. Experimental results show that compared with YOLOv11n, the mAP50 of the YOLO-AMM algorithm is increased by 3.6%, the recall rate by 2.3%, and the precision by 2.9%. Meanwhile, the number of parameters is reduced by 18.0%, the computational load by 14.1%, and the model size by 16.5%. The proposed algorithm outperforms a variety of classic lightweight algorithms and similar improved algorithms and is suitable for deployment on resource-constrained embedded devices.

Keywords: defect detection; lightweight; ADown; YOLOv11n; MLCA

0 引 言

钢铁行业产品应用广泛,带钢作为典型钢铁制品,凭借高强度、良好可塑性及易加工性,成为航空航天、机械制造等领域关键原材料^[1]。但受环境、原材料质量、生产工艺等影响,带钢表面易产生裂纹、夹杂、斑块、麻点、压入氧化皮、划痕等缺陷^[2],这些缺陷不仅会影响外观,还会降低钢材性能、缩短产品使用寿命,给下游关联产品带来质量隐患,甚至造成重大经济损失^[3]。目前多数厂商依赖人工筛查,存

在干扰大、速度慢、误检率高的问题^[4]。因此,研发兼具高时效特性与精准缺陷识别能力的检测算法,已成为把控钢材表面质量的核心技术路径,对带钢表面缺陷检测意义重大^[5]。

近年来,深度学习技术的崛起为带钢表面缺陷检测带来革命性突破,该技术凭借高精度、快速检测的优势,减少了人力资源消耗,提升了生产效率和质量控制精度,已成为该领域的主导研究算法。借助深度学习构建的目标检测算法,可划分为单阶段与双阶段两类^[6]。双阶段算法以区域

卷积神经网络 (region-convolutional neural network, R-CNN)^[7] 为典型, 衍生的快速区域卷积神经网络 (fast region-convolutional neural network, Faster R-CNN)^[8]、更快区域卷积神经网络 (faster region-convolutional neural network, Faster R-CNN)^[9] 等虽通过优化提升检测精度, 但速度偏慢。单阶段算法如 YOLO(you only look once)系列^[10]、单发多框检测器 (single shot multibox detector, SSD)系列^[11]等, 将目标检测转化为回归问题, 具备端到端处理、速度快、部署灵活等特性。任帅等^[12]改进 YOLOv11, 集成双向特征金字塔网络 (bidirectional feature pyramid network, BiFPN) 与全局注意力机制 (global attention mechanism, GAM) 优化小目标识别, 引入 SlimNeck 降低参数量与复杂度。张泽辰等^[13]通过动态通道空间卷积模块改进 YOLOv8 网络、设计含多尺度 Transformer 模块的小目标感知增强网络、采用双线性插值与上下文引导注意力融合机制及引入自适应分布交并比损失函数, 实现叶片小缺陷的高精度定位。严赫等^[14]提出 GMS2-YOLOv11 模型, 以多尺度特征提取、特征融合增强与分离归一化检测头相结合, 提高金属表面缺陷检测精度。方忠志等^[15]提出用于表面异常检测的判别式训练重建嵌入 (discriminatively trained reconstruction embedding for surface anomaly detection, DRAEM) 算法, 通过加入 GrabCut 分割、跳跃连接、通道和空间双重注意力模块、Transformer 模块及空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 模块, 提高纤维绳索的缺陷检测的识别精度与鲁棒性。Chen 等^[16]基于 YOLOv5, 改进特征提取网络、采用自适应特征融合网络、使用 Ghost 卷积等优化, 提升冷轧带钢缺陷检测质量。Xu 等^[17]提出 YOLOv8n-GAM, 以 Con5v 卷积、注意力机制及多检测头提升缺陷检测性能。包广清等^[18]改进 k-means++ 优化先验框, 利用 FocalSIoU 损失函数优化回归, 设计多尺度特征融合模块 (multi-scale feature fusion module, MFFM) 增强特征提取, 使用空到深 (space to depth, SPD) 卷积减少漏检。Ni 等^[19]提出 EGC-YOLO 算法, 设计并嵌入边缘细节增强模块 (edge detail enhancement module, EDEM), 捕捉缺陷纹理特征, 引入广义动态特征金字塔网络 (generalized dynamic feature pyramid network, GDFPN) 适配多尺度缺陷检测, 采用 CGAFusion 模块优化浅层和深层特征的融合。戴林华等^[20]将 YOLOv8 主干网络替换为 HGNetV2 并采用 Slim-Neck 设计降低算法复杂度, 引入高效多尺度注意力机制 (efficient multi-scale attention, EMA) 和搭配 DIoU 损失函数提高精度。Zhang 等^[21]设计 ESI-YOLOv8 算法, 通过引入高效多尺度卷积与逐点卷积 (efficient multi-scale convolution and pointwise convolution, EP) 模块、空间金字塔池化-大可分离核注意力 (spatial pyramid pooling with filter-large separable kernel attention, SPPF-LSKA) 模块并替换为内部交并比损失 (inner intersection

over union, INNER-CIOU) 损失函数, 实现精度提升和模型参数降低。计忠平等^[22]提出集成全局局部感知与分层特征融合的检测算法, 通过特征增强、分层融合及全局局部感知模块提升检测精度。但上述方法仍难兼顾工业场景的高精度与快速度需求, 部分轻量化设计还会显著降低精度。

针对上述问题, 本文提出改进 YOLOv11n 的轻量级带钢表面缺陷检测算法 YOLO-AMM, 具体工作为: 引入自适应下采样 (adaptive downsampling, ADown)^[23], 将输入特征图分两路处理, 减少参数与 GFLOPs, 保留微小缺陷特征, 提升精度和实时性。设计多尺度深度可分离瓶颈 (multi-scale depthwise separable bottleneck, MDSB) 模块对 C3k2 模块进行改进, 优化跨尺度特征融合策略, 增强多尺度特征提取能力与效率。引入混合局部通道注意力 (mixed local channel attention, MLCA)^[24], 通过双分支处理融合信息, 生成注意力图, 提升对不同尺度缺陷的提取能力, 减少漏检, 提升检测精度。

1 YOLO-AMM 算法

YOLOv11 作为 Ultralytics 团队研发的新一代 YOLO 系列算法, 其核心目标在于提升检测精度与速度、精简模型参数量, 并增强复杂场景下的多任务处理能力。针对 YOLOv11n 应用于带钢表面缺陷检测时存在的微小缺陷特征丢失、多尺度特征提取能力不足及轻量化程度有限等问题, 同时结合现有带钢表面缺陷检测算法中普遍存在的参数量庞大、检测精度欠佳及在资源受限的嵌入式设备上部署难度较大等现状, 本文以 YOLOv11n 为基础进行改进, 提出 YOLO-AMM 改进算法, 整体结构如图 1 所示。首先, 采用 ADown, 替代传统跨步卷积, 将输入特征图分两支处理, 既减少微小缺陷特征丢失, 又降低参数量与计算量, 提升精度和实时性; 其次, 将原 C3k2 模块融入了 MDSB 模块, 通过多尺度并行卷积增强多尺度特征捕捉能力, 同时利用深度可分离卷积实现轻量化; 最后, 在特征融合网络中引入 MLCA 注意力机制, 结合全局与局部信息生成注意力图, 强化对不同尺度缺陷的特征提取, 减少漏检误检。通过上述改进, YOLO-AMM 算法在保持轻量化优势的同时, 显著提升了带钢表面缺陷检测的精度与效率。

1.1 引入 ADown

在带钢表面缺陷检测中, YOLOv11n 虽能满足实时检测基础需求, 但传统跨步卷积下采样易丢失微小缺陷特征, 且受计算方式限制, 难以在轻量化基础上同步提升实时性与精度。结合带钢生产线对检测算法的严苛要求, 引入 ADown 替代 YOLOv11n 中部分传统下采样模块, 其结构如图 2 所示。ADown 通过多路径处理实现轻量化与高精度协同优化, 其输入特征图 F_i 在通道维度被平均分为 F_1 和 F_2 后, F_1 通过步长为 2 的 3×3 跨步卷积下采样提取高维特征, F_2 经 3×3 最大池化强化局部特征再通过 1×1 非跨步卷积映射, 表达式为:

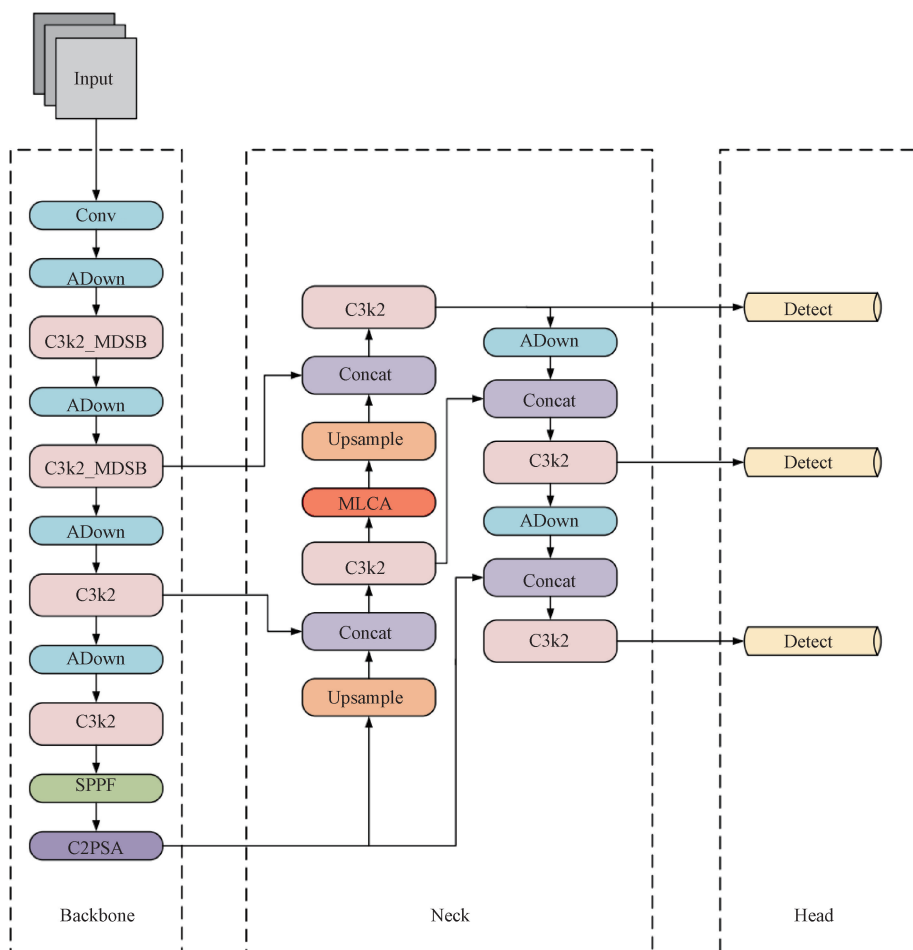


图1 YOLO-AMM 算法结构

Fig. 1 Structure of the YOLO-AMM algorithm

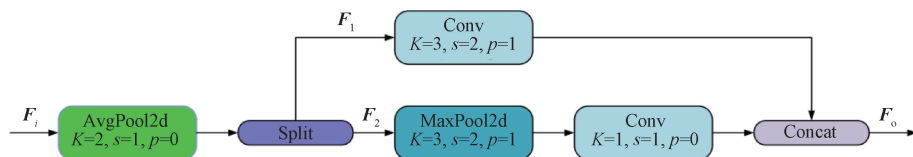


图2 ADown 结构

Fig. 2 Structure of ADown

$$\mathcal{F}_{\text{path1}} = \text{Conv}_{3 \times 3, s=2, p=1}(\mathbf{F}_1) \quad (1)$$

$$\mathcal{F}_{\text{path2}} = \text{Conv}_{1 \times 1}(\text{MaxPool}_{3 \times 3, s=2, p=0}(\mathbf{F}_2)) \quad (2)$$

$$\mathbf{F}_o = \text{Concat}(\mathcal{F}_{\text{path1}}, \mathcal{F}_{\text{path2}}) \quad (3)$$

式中:“ $\times$ ”符号用于表示卷积核; $\mathcal{F}_{\text{path1}}(\cdot)$ 表示对 $\mathbf{F}_1$ 路径的卷积操作; $\mathcal{F}_{\text{path2}}(\cdot)$ 表示对 $\mathbf{F}_2$ 路径的池化和卷积操作, $\text{Concat}(\cdot)$ 表示通道拼接。

ADown 一方面通过分路处理优化卷积参数,相较传统跨步卷积显著减少参数量与计算量,在保证特征提取效果、缩减模型规模的同时,还能通过降低特征图分辨率减轻带钢生产线连续图像的处理负担,加快图像处理速度以适配高速轧制场景的实时检测需求,实现轻量化与实时性优化。另一方面,其双路设计可减少微小缺陷信息丢失,

传统跨步卷积单次大跨步压缩易割裂微小缺陷局部特征关联,而 ADown 中 $\mathbf{F}_1$ 的 3×3 卷积能捕捉微小缺陷全局关联特征, $\mathbf{F}_2$ 的最大池化可强化局部纹理,两路特征拼接为后续特征提取提供更全面初始特征以降低微小缺陷漏检率,且 ADown 在降参的同时能避免因计算资源不足导致的特征提取简化,保障检测稳定性。

1.2 构建 C3k2_MDSB

在 YOLOv11n 框架中, C3k2 模块是特征提取核心组件,其密集卷积与特征融合性能直接影响检测精度与效率。但带钢表面缺陷存在类型多、尺度差异大、背景干扰等问题,对特征提取模块的多尺度感知与计算效率要求严苛,而传统 C3k2 模块,依赖常规卷积与残差连接,卷积核

尺度固定、特征融合简单,在复杂缺陷检测中灵敏度不足且易受背景冗余信息干扰。

针对上述挑战,本文提出 C3k2_MDSB 模块,该模块是将 YOLOv11n 中 C3k2 模块的原始瓶颈模块替换为 MDSB 模块(如图 3(a)所示),构建的改进型特征提取模块(如图 3(b)所示)。该改进主要体现在 3 个方面。首先是设计 MDSB 模块,利用 3×3 、 5×5 、 7×7 的多尺度并行深度可分离卷积分支,通过不同感受野的卷积核针对性捕捉带钢表面不同尺度的缺陷特征,尤其增强对小、大尺度的缺陷特征响应。其次,采用深度可分离卷积替代传统标准卷积,在保持特征提取能力的同时避免参数量与计算量大幅增加,实现模型轻量化与效率优化。最后,优化特征融合策略,通过通道维度的分割与拼接操作保留各尺度缺陷

的原始特征,避免融合信息丢失,同时利用 1×1 卷积实现跨尺度特征的自适应整合,抑制带钢背景中的冗余特征,提升夹杂物、裂缝等特征复杂缺陷的辨识度。表达式为:

$$\mathbf{X}_3 = \text{DWConv}_{3\times 3}(\mathbf{X}) \tag{4}$$

$$\mathbf{X}_5 = \text{DWConv}_{5\times 5}(\mathbf{X}) \tag{5}$$

$$\mathbf{X}_7 = \text{DWConv}_{7\times 7}(\mathbf{X}) \tag{6}$$

$$\mathbf{X}_{\text{concat}} = \text{Concat}(\mathbf{X}_3, \mathbf{X}_5, \mathbf{X}_7) \tag{7}$$

$$\mathcal{F}_{\text{MDSB}} = \mathbf{X} + \text{Conv}_{1\times 1}(\mathbf{X}_{\text{concat}}) \tag{8}$$

式中:“ $\times$ ”符号用于表示卷积核; $\mathbf{X}$ 为输入特征图; $\mathbf{X}_3$ 代表经过 3×3 尺度深度可分离卷积分支的结果; $\mathbf{X}_5$ 代表经过 5×5 尺度深度可分离卷积分支的结果; $\mathbf{X}_7$ 代表经过 7×7 尺度深度可分离卷积分支的结果; $\text{Concat}(\cdot)$ 表示通道拼接; $\mathbf{X}_{\text{concat}}$ 为拼接后的结果; $\mathcal{F}_{\text{MDSB}}$ 为 MDSB 输出结果。

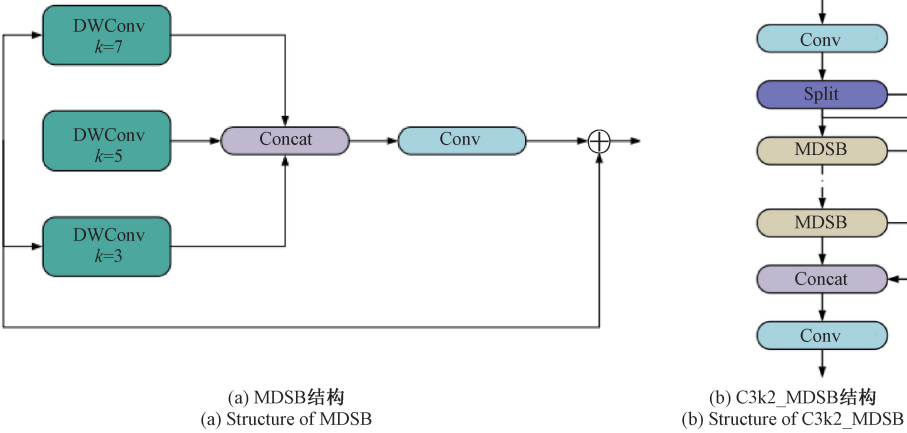


图 3 MDSB 结构与 C3k2_MDSB 结构

Fig. 3 Structures of MDSB and C3k2_MDSB

1.3 引入 MLCA 注意力机制

在带钢表面缺陷检测中,存在缺陷类型繁杂、部分缺陷尺度细微且易受背景噪声干扰的问题,导致检测常出现漏检、误检。YOLOv11n 作为轻量级模型,其颈部网络在

特征压缩过程中易丢失细节信息,难以精准捕捉细微及不同尺度缺陷特征,影响检测精度。为此,在 YOLOv11n 颈部网络引入 MLCA 注意力机制以提升性能,其结构如图 4 所示,图中“ $\times$ ”符号用于表示尺寸规格。

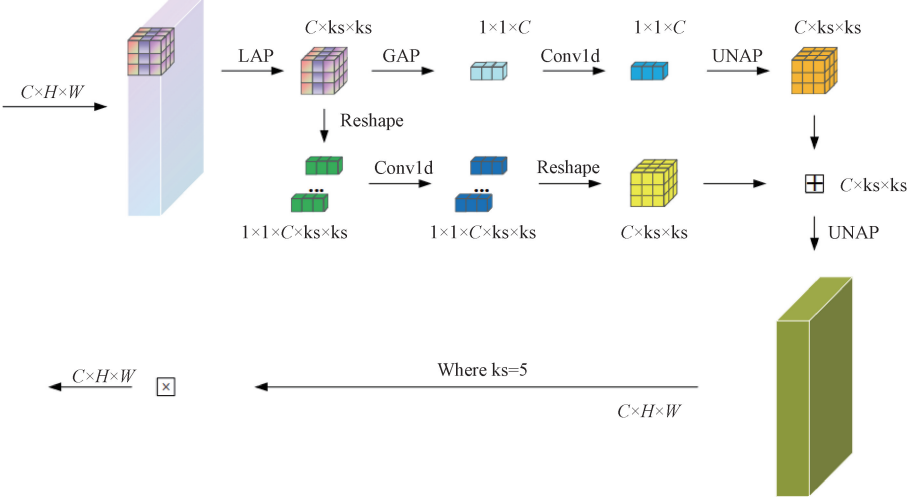


图 4 MLCA 结构

Fig. 4 Structure of MLCA

MLCA 注意力机制核心优势在于可同时融合通道与空间信息、局部与全局信息。对输入特征图,先通过局部平均池化(local average pooling, LAP)将其转换为 $C \times ks \times ks$ 的局部特征张量,捕获细微缺陷细粒度空间细节,基于全局平均池化(global average pooling, GAP)提取 $1 \times 1 \times C$ 的全局特征向量捕获上下文信息,以理解缺陷与背景关系。全局分支计算特征图通道间关系获取全局信息,引导模型关注高语义通道区域,局部分支将特征张量重排为 $1 \times 1 \times C \times ks \times ks$ 的扩展向量,捕获局部细节并计算空间关系生成注意力权重,提升对细微缺陷纹理、边缘的感知能力。随后两分支注意力图经逐元素加法融合,再通过反池化(un-average pooling, UNAP)恢复至原始空间维度,最终与输入特征图逐元素乘法运算,输出细化特征图。

MLCA 为轻量化模块,通过一维卷积减少计算复杂度与参数量,在增加少量参数的情况下,可显著增强 YOLOv11n 对不同尺度带钢缺陷的特征提取能力,有效减少因背景干扰与细微缺陷导致的漏检,提升检测精度,更适配轻量级带钢表面缺陷检测任务。

2 实验结果与讨论

2.1 实验数据集

实验基于东北大学 NEU-DET 数据集开展模型训练,包含 6 类缺陷:裂纹(crazing)、夹杂(inclusion)、斑块(patches)、麻点(pitted surface)、压入氧化皮(rolled in scale)、划痕(scratches)。每类缺陷含 300 张图片,总计 1 800 张。数据集划分采用 8:1:1 比例,训练集 1 440 张、测试集 180 张、验证集 180 张。

2.2 实验环境与评价指标

本实验的环境设定于 Windows11 操作系统上, GPU 为 NVIDIA GeForce RTX 3060 型号, Python 版本是 3.11,深度学习框架使用的是 Pytorch2.5.1。实验训练参数设置如下:训练迭代次数设置为 200,批次大小设置为 16,初始学习率设置为 0.01,学习动量为 0.937,优化器采用 SGD,其权重衰减为 0.000 5,图像大小统一调整为 640×640 pixels。

为全方位评估模型的性能表现,实验选取精确率(precision, P)、召回率(recall, R)、平均精度均值(mean average precision, mAP)、参数量(parameters)、计算量(giga floating point operations per second, GFLOPs)以及模型大小(model size)作为评估指标。 P 用以衡量预测的准确性, R 反映模型对正样本的捕捉能力,mAP 用于综合评估模型在多类别任务中的表现,参数量反映模型参数规模,直接关联存储需求与运行效率,GFLOPs 体现模型运算复杂度,影响实时处理能力。模型大小是衡量算法实用性和部署适应性的重要指标。公式为:

$$P = \frac{TP}{TP + FP} \quad (9)$$

$$R = \frac{TP}{TP + FN} \quad (10)$$

$$AP = \int_0^1 P(R) dR \quad (11)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP(i) \quad (12)$$

式中:TP 指的是正样本里被准确分类的总数量;FP 是正样本中被错误分类的总数量;FN 代表负样本中被错误分类的总数量。 P 反映的是正样本被预测正确的占比, R 体现的是所有正例中被正确预测出来的比例。AP 代表 P - R 曲线下方的面积,mAP 是所有类别 AP 值的平均值。

图 5 所示为损失函数变化的曲线。从图 5 中可以看出,在目标检测模型训练中,对比 YOLO11n, YOLO-AMM 展现显著优势。训练初期,其损失值下降迅猛,能快速适配数据、捕捉特征,为训练筑牢基础;进入中后程,损失始终更低,反映出更强拟合能力,对目标定位与分类或更精准,且曲线波动小,训练稳定性佳,受噪声等干扰弱,可持续优化。更低的损失函数表明 YOLO-AMM 在训练中能更好地拟合带钢表面缺陷数据,对不同类型、不同尺度缺陷的检测精度更高,检测结果更可靠。

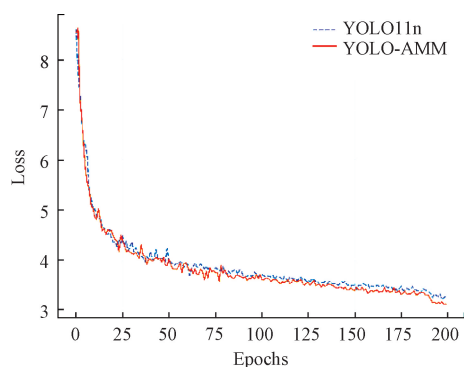


图 5 损失函数变化曲线

Fig. 5 Variation curve of loss function

2.3 消融实验

为验证 YOLO-AMM 算法在带钢表面缺陷检测任务的实际作用,本文对改进的模块展开消融实验,实验结果如表 1 所示。

从表 1 可以看出,将 Conv 模块替换为 ADown, P 实现 2.4% 的提升, R 降低 1.2%, 同时使模型参数量减少 482 560, mAP50 略微提升 0.3%, GFLOPs 降低 1.1, 模型大小缩减 0.9 MB, 这一数据充分印证其在轻量化与准确率的双重价值,尤其适合带钢生产线对实时性的严苛要求。将 C3k2 模块替换为 C3k2_MDSB 模块,虽然 P 降低 2.5%, 参数量增加 15 648, GFLOPs 上升 0.3, 但 R 提升 3.7%, mAP50 提升 0.7%, 说明其多尺度卷积设计可以增强对多尺度缺陷的捕捉能力。MLCA 模块的表现更为均衡,在 P 提升 1.0%、 R 提升 2.1% 的同时,参数和模型大

表 1 消融实验结果

Table 1 Results of ablation experiments

方法	ADown	C3k2_MDSB	MLCA	$P/\%$	$R/\%$	mAP50/ $\%$	parameters	GFLOPs	model size/MB
YOLOV11n				71.1	73.6	77.6	259 101 0	6.4	5.238
A	✓			73.5	72.4	77.9	210 845 0	5.3	4.331
B		✓		68.6	77.3	78.3	260 665 8	6.7	5.281
C			✓	72.1	75.7	78.3	259 102 0	6.5	5.240
D	✓	✓		76.3	72.6	79.6	212 409 8	5.5	4.375
E	✓	✓	✓	74.0	75.9	81.2	212 410 8	5.5	4.376

小基本不变,凸显了注意力机制在不增加模型负担前提下的特征优化能力。单独使用 C3k2_MDSB 模块时 P 下降,其核心原因是数据集中压入氧化皮等缺陷存在集中分布特征,大卷积核会捕捉缺陷周围冗余区域,导致核心特征权重被稀释。该现象是多尺度特征提取对特定缺陷分布场景的适应性表现,通过 ADown 的高效优化与 MLCA 的特征聚焦机制协同作用,已有效缓解此问题,实现 P 的平衡。整体来看,YOLO-AMM 算法 mAP50 较基础模型提升 3.6%,为所有方法中最高, R 提升 2.3%, P 提升 2.9%。同时,3 个模块融合后参数降低 18.0%,GFLOPs 减少 14.1%,模型大小下降 16.5%,验证改进策略在性能和效率平衡上的合理性,充分验证改进策略在带钢表面缺陷检测任务中的实用性与有效性,为工业场景下的高精度、高效率缺陷检测提供技术支撑。

图 6 为 mAP50 指标变化折线。从图 6 中可以看出,训练初期,YOLO-AMM 算法的 mAP50 提升更快,随着训练进行,两者 mAP50 都逐渐上升并趋于稳定,且后期 YOLO-AMM 算法整体精度更高,说明 YOLO-AMM 改进可能有助于模型在训练过程中更高效地提升检测精度。

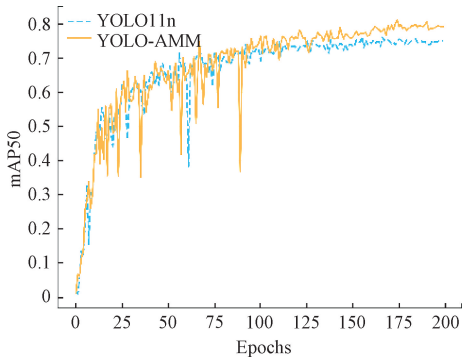


图 6 mAP50 指标变化折线

Fig. 6 Line chart of mAP50 metric changes

2.4 对比实验

为验证 YOLO-AMM 算法在带钢表面缺陷检测任务中的性能与潜力,基于 NEU-DET 数据集,本节选取多种目标检测算法和常用的轻量化算法开展横向对比实验,相关结果如表 2 所示。

表 2 对比实验

Table 2 Results of comparative experiments

方法	mAP50/ $\%$	parameters/ M	GFLOPs
YOLOv3-tiny	74.6	12.10	19.0
YOLOv5n	76.0	2.51	7.1
YOLOv6n	77.3	4.20	11.8
YOLOv8n	79.0	3.00	8.2
YOLOv10n	73.2	2.70	8.4
YOLOv11n-MobileNetv4	72.9	1.80	4.3
YOLOv11n-EfficientNetv2	73.0	2.10	5.4
文献[18]	78.3	7.50	—
文献[19]	80.2	3.30	8.3
文献[20]	79.3	2.10	6.1
文献[21]	78.0	2.90	7.8
文献[22]	80.7	4.50	—
YOLO-AMM	81.2	2.10	5.5

通过对比分析可以发现,YOLO-AMM 的 mAP50 达到 81.2%,不仅比 YOLOv3-tiny、YOLOv5n、YOLOv6n、YOLOv8n 经典 YOLO 系列轻量级模型分别提升 6.6%、5.2%、3.9%、2.2%,且远高于 MobileNetv4、EfficientNetv2 工业中常用的轻量化模型,也分别超出文献[18-21]多数同类改进算法 2.9%、1.0%、1.9%、3.2%,略高于文献[22]的 80.7%,充分体现对带钢表面多种缺陷的精准识别能力。在轻量化与效率上,YOLO-AMM 算法参数量为 2.1 M,远低于 YOLOv3-tiny、YOLOv6n 等,与文献[20]和 EfficientNetv2 持平,仅略高于 MobileNetv4,但检测精度更高,GFLOPs 为 5.5,显著低于 YOLOv3-tiny、YOLOv5n 等,略高于 MobileNetv4、EfficientNetv2 通用轻量化网络,在保证检测精度的同时仍能满足实时检测需求。这种优势源于 ADown、C3k2_MDSB 和 MLCA 模块的协同作用,ADown 轻量化与精度优化,C3k2_MDSB 增强多尺度缺陷特征捕捉,MLCA 模块聚焦关键细节,三者共同实现检测性能与效率的平衡,适配带钢高速轧制场景。

2.5 机制分析

YOLO-AMM 之所以能打破传统轻量化算法降参数

必降精度的固有困境,实现精度提升与参数量下降的双重突破,核心在于 ADown、C3k2_MDSB、MLCA 三模块在特征处理全链路中形成的深度协同互补机制。三者并非孤立叠加,而是围绕轻量化特征提取、高效特征表达、精准特征筛选的核心逻辑层层递进、深度联动。ADown 大幅降低参数量,又避免微小缺陷特征割裂,为后续处理提供完整、高质量的初始输入并减轻计算负担。C3k2_MDSB 模块基于该特征进行多尺度和高效率的特征表达,挖掘不同缺陷的核心信息,同时维持轻量化特性。MLCA 模块对表达后的特征进行精准化与抗干扰的筛选,突出有效特征、剔除冗余噪声。在前向推理层面,三级模块层层递进,输出高纯度、高辨识度的缺陷特征,为检测任务提供精准支撑。在反向训练层面,三者的损失信号相互引导,MLCA 的注意力权重反馈优化 C3k2_MDSB 的多尺度卷积参数, C3k2_MDSB 的特征表达效果反向调整 ADown 的分路特征融合比例,ADown 的轻量化特性则为整体模型的参数优化提供更大空间,形成动态迭代的优化闭环,最终实现参数量下降 18.0%与 mAP50 提升 3.6%,为带钢表面缺陷检测提供高精度和轻量化的最优解。

2.6 部署测试

将本文所提出的 YOLO-AMM 算法,部署于 NXP 公司所生产的 i.MX 8 M Plus 边缘计算的嵌入式设备,验证改进算法的部署适配性与综合性能,如图 7 所示。随机抽取 NEU-DET 测试集 180 张图片,开展部署测试与性能对比。如表 3 所示,YOLO-AMM 在该平台上展现出优异的综合优势。在检测精度方面,该模型 mAP50 达 81.2%,超越 YOLOv11n 的 77.6%,实现 3.6%的性能提升,模型轻量化后体积仅 4.376 MB,较 YOLOv11 减小 16.5%,资源

消耗更低,适配资源受限的嵌入式部署场景,推理速度达 29.9 fps,通过精度与速度的平衡设计,满足带钢生产线现场检测需求。

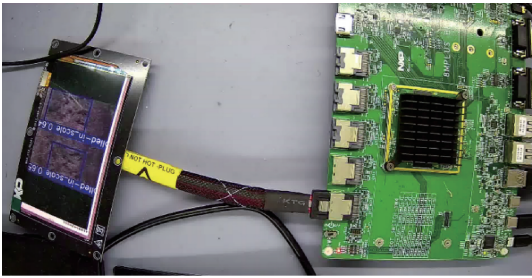


图 7 嵌入式部署效果

Fig. 7 Embedded deployment effect

表 3 嵌入式设备对比结果

Table 3 Comparison results of embedded devices

方法	mAP50/%	model size/MB	FPS/fps
YOLOv11n	77.6	5.238	28.6
YOLO-AMM	81.2	4.376	29.9

2.7 可视化分析

针对 YOLO-AMM 模型与 YOLOv11n 模型的检测性能对比,从测试集上各类缺陷图片的检测效果,如图 8 所示。可以看出,YOLOv11n 模型在检测过程中不仅平均精度较低,还存在对部分缺陷的漏检、误检问题。相比之下,YOLO-AMM 模型在相同位置的检测表现有显著改善,这表明 YOLO-AMM 模型在检测精度、缺陷捕捉能力和抗干扰能力上均优于 YOLOv11n 模型,更能满足带钢表面缺陷检测对准确性和完整性的严苛要求,适配高速生产线的质量管控需求。

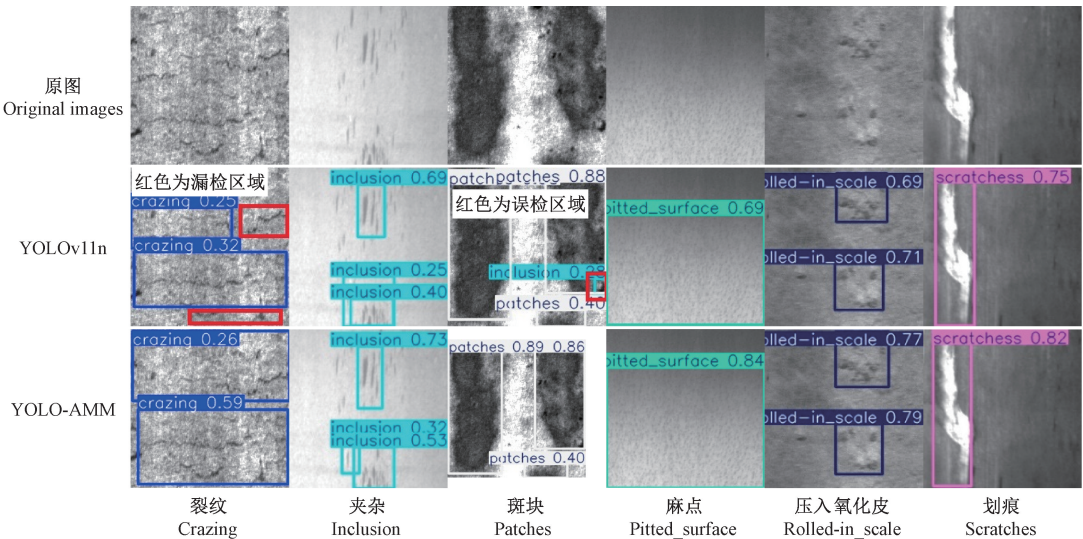


图 8 可视化结果

Fig. 8 Visualization results

3 结 论

为解决带钢表面缺陷检测领域中,现有算法存在参数量大、检测精度不足以及在资源受限的嵌入式设备上部署难度较大等问题,本文提出改进 YOLOv11n 的 YOLO-AMM 算法,引入 ADown 解决微小缺陷特征丢失问题以提升精度,同时减少参数量和 GFLOPs;将 C3k2 模块的原始瓶颈块替换为 MDSB 模块,构建 C3k2_MDSB 模块,增强多尺度特征提取能力和计算效率;在特征融合网络中引入 MLCA 注意力机制,提升对缺陷特征的精准筛选。实验表明, YOLO-AMM 算法的 mAP₅₀ 较 YOLOv11n 提升 3.6%,*R* 提升 2.3%,*P* 提升 2.9%,同时参数减少 18.0%,GFLOPs 减少 14.1%,模型大小下降 16.5%,优于多种经典轻量级模型和同类改进算法,适合部署于资源受限的嵌入式设备。未来研究可结合网络剪枝、知识蒸馏等技术,进一步缩减模型参数量,提升推理速度,推动轻量化缺陷检测技术在钢铁及相关制造业的落地应用。

参考文献

- [1] LI H L, LIU M, YIN Y F, et al. Steel surface defect detection based on multi-layer fusion networks [J]. Scientific Reports, 2025, 15(1): 10371.
- [2] 孙克鑫. 基于 YOLOv8 的钢材表面缺陷识别算法[D]. 大连:大连交通大学, 2025.
SUN K X. Research on steel surface defect recognition algorithm on YOLOv8 [D]. Dalian: Dalian Jiaotong University, 2025.
- [3] 薛文亮, 靳伍银, 王全. 改进关系网络的小样本带钢表面缺陷分类方法[J]. 电子测量技术, 2021, 44(19): 167-172.
XUE W L, JIN W Y, WANG Q. The surface defects classification method of strip steel with small samples based on improved relation network [J]. Electronic Measurement Technology, 2021, 44(19): 167-172.
- [4] 杨森泉, 丁凡, 文昊翔, 等. 基于 CA-YOLOv5 的热轧带钢表面缺陷检测方法[J]. 光电子·激光, 2024, 35(1): 21-28.
YANG S Q, DING F, WEN H X, et al. Hot-rolled strip steel surface defect detection method based on CA-YOLOv5 [J]. Journal of Optoelectronics·Laser, 2024, 35(1): 21-28.
- [5] 王永浩. 基于加权结构 SVM 的钢板表面缺陷分类[J]. 电子测量技术, 2020, 43(11): 69-73.
WANG Y H. Steel plate surface defect classification based on weighted structured SVM [J]. Electronic Measurement Technology, 2020, 43(11): 69-73.
- [6] 周迅, 李帆, 张岩. LiteSteel-YOLO: 小目标低对比度轻量化级钢材缺陷检测网络[J]. 电子测量技术, 2026, 49(6): 202-210.
- [7] ZHOU X, LI F, ZHANG Y. LiteSteel-YOLO: A lightweight steel defect detection network for small targets with low contrast [J]. Electronic Measurement Technology, 2026, 49(6): 202-210.
- [8] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]//2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580-587.
- [9] GIRSHICK R. Fast R-CNN [C]//2015 IEEE International Conference on Computer Vision, 2015: 1440-1448.
- [10] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [11] JIANG P Y, ERGU D J, LIU F Y, et al. A review of yolo algorithm developments [J]. Procedia Computer Science, 2022, 199: 1066-1073.
- [12] QIAN H M, WANG H L, FENG S F, et al. FESSD: SSD target detection based on feature fusion and feature enhancement [J]. Journal of Real-Time Image Processing, 2023, 20(1): 2.
- [13] 任帅, 杨思念, 曹立佳, 等. 面向边缘计算设备的轻量化带钢缺陷检测算法[J]. 电子测量技术, 2026, 49(5): 85-94.
REN S, YANG S N, CAO L J, et al. Lightweight strip steel defect detection algorithm for edge computing devices [J]. Electronic Measurement Technology, 2026, 49(5): 85-94.
- [14] 张泽辰, 姚顺春, 梁骁俊, 等. 基于小目标感知增强的风机叶片缺陷智能检测[J]. 仪器仪表学报, 2025, 46(9): 159-172.
ZHANG Z C, YAO S C, LIANG X J, et al. Wind turbine blade intelligent defect detection based on small object perception enhancement [J]. Chinese Journal of Scientific Instrument, 2025, 46(9): 159-172.
- [15] 严赫, 史晓亮, 吴超华, 等. GMS2-YOLO: 融合改进 MANet 和检测头的金属表面缺陷检测方法[J]. 电子测量与仪器学报, 2026, 40(3): 176-188.
YAN H, SHI X L, WU C H, et al. GMS2-YOLO: A metal surface defect detection method fusing improved MANet and detection head [J]. Journal of Electronic Measurement and Instrumentation, 2026, 40(3): 176-188.

- [15] 方忠志,朱华波,陶友瑞,等.改进DRAEM的非监督纤维绳索表面缺陷检测算法[J].电子测量与仪器学报,2026,40(2):44-54.
FANG Z Z, ZHU H B, TAO Y R, et al. Improved unsupervised DRAEM algorithm for surface defect detection in fiber ropes [J]. Journal of Electronic Measurement and Instrumentation, 2026, 40(2): 44-54.
- [16] CHEN S Z, JIANG S Q, WANG X Y, et al. An efficient detector for detecting surface defects on cold-rolled steel strips [J]. Engineering Applications of Artificial Intelligence, 2024, 138: 109325.
- [17] XU H W, XIAO X Y, ZHAO Z W, et al. YOLOv8n-GAM: An improved surface defect detection network for hot-rolled strip steel [J]. Engineering Research Express, 2024, 6(2): 025557.
- [18] 包广清,周芷意,孟庆成.基于融合注意力和多尺度特征的热轧带钢表面缺陷检测方法[J].北京工业大学学报,2025,51(8):944-956.
BAO G Q, ZHOU Z Y, MENG Q C. Hot-rolled strip steel surface defect detection method based on fused attention and multi-scale features [J]. Journal of Beijing University of Technology, 2025, 51(8): 944-956.
- [19] NI Y F, ZI D X, CHEN W, et al. EGC-YOLO: Strip steel surface defect detection method based on edge detail enhancement and multiscale feature fusion[J]. Journal of Real-Time Image Processing, 2025, 22(2): 65.
- [20] 戴林华,黎远松,石睿,等. HSED-YOLO: 一种轻量化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): 95-106.
DAI L H, LI Y S, SHI R, et al. HSED-YOLO: A lightweight detection model for strip steel surface defects [J]. Journal of Guangxi Normal University (Natural Science Edition), 2025, 43(2): 95-106.
- [21] ZHANG X R, WANG Y L, FANG H S. Steel surface defect detection algorithm based on ESI-YOLOv8[J]. Materials Research Express, 2024, 11(5): 056509.
- [22] 计忠平,林晨伟,何志伟,等.集成全局局部感知与分层特征融合的带钢表面缺陷检测[J].计算机工程与应用,2026,62(4):373-381.
JI Z P, LIN C W, HE Z W, et al. Strip steel surface defect detection integrating global-local perception and hierarchical feature fusion[J]. Computer Engineering and Applications, 2026, 62(4): 373-381.
- [23] WANG C Y, YE H I, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]//European Conference on Computer Vision, 2024: 1-21.
- [24] WAN D H, LU R S, SHEN S Y, et al. Mixed local channel attention for object detection[J]. Engineering Applications of Artificial Intelligence, 2023, 123: 106442.

作者简介

孙艳龙,硕士研究生,主要研究方向为机器视觉。

E-mail:sylpro@foxmail.com

陈广锋(通信作者),博士,副教授,硕士研究生导师,主要研究方向为机器视觉。

E-mail:chengf@dhu.edu.cn