

基于改进YOLOv8n的钢材缺陷检测算法

李梓豪

五邑大学电子与信息工程学院, 广东 江门

收稿日期: 2025年3月16日; 录用日期: 2025年4月16日; 发布日期: 2025年4月24日

摘要

针对目前钢材表面缺陷检测方法中存在的算法复杂度高、模型计算量大等问题, 本文选择YOLOv8n为基准模型, 提出了一种基于YOLOv8n改进的跨尺度特征融合的轻量化的钢材表面缺陷检测方法, 并命名为YOLOv8n-FMR。首先, 新模型利用多尺度卷积模型MSC (Multi-Scale Conv)对模型进行改进, 有效减少因堆叠不同尺寸的特征提取层而产生的精度损失问题; 其次, 为了降低融合MSC模块所带来的召回率和准确率减少的影响, 引入了Focal Modulation焦点调制模块对Neck网络中的SPPF模块进行改进。最后, 为了提高模型整体的检测性能, 引入了应用于遥感领域的RFD模块, 有效增加了模型对小物体的识别能力。实验结果表明, 改进后的YOLOv8n-FMR模型在NEU-DET数据集上的mAP达到了72.8%, 比原模型提高了2.7%, 并且模型的参数量也比原模型下降了0.1 M。改进后的模型计算量仅为9.5 GFLOPs, 提出的模型不仅实现了轻量化的设计, 而且提升了检测效果。

关键词

YOLOv8n, 目标检测, 特征融合, 多尺度卷积, 焦点调制器, 鲁棒特征降采样

Steel Defect Detection Algorithm Based on Improved YOLOv8n

Zihao Li

School of Electronics and Information Engineering, Wuyi University, Jiangmen Guangdong

Received: Mar. 16th, 2025; accepted: Apr. 16th, 2025; published: Apr. 24th, 2025

Abstract

In response to the high algorithm complexity and large model computation of current steel surface defect detection methods, this paper selected YOLOv8n as the benchmark model and proposed a lightweight steel surface defect detection method based on YOLOv8n improved cross scale feature fusion, named YOLOv8n FMR. First, the new model utilized the multi-scale convolution model (MSC)

to improve the model, effectively reducing the accuracy loss caused by stacking different size feature extraction layers. Second, to reduce the impact of the recall rate and accuracy reduction caused by the fusion of the MSC module, the Focal Modulation (FoM) module was introduced to improve the SPPF module in the Neck network. Finally, to improve the overall detection performance of the model, the RFD module, which is commonly used in remote sensing, was introduced to effectively increase the model's ability to recognize small objects. The experimental results show that the mAP of the improved YOLOv8n-FMR model on the NEU-DET dataset reached 72.8%, an improvement of 2.7% over the original model, and the model's parameter count was reduced by 0.1 M. At the same time, the model's computation volume was only 9.5 GFLOPs. The proposed model not only achieves a lightweight design, but also improves detection performance.

Keywords

YOLOv8n, Object Detection, Feature Fusion, MSC, Focal Modulation, RFD

Copyright © 2025 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

随着制造及加工工艺水平的不断发展,由铁矿精炼而成的各种金属产品(如不锈钢)具有耐腐蚀、高硬度,高韧性和高抗压强度等特点,已然成为了人们日常生活中不可或缺的物品。因此,如何快速、批量筛选合格的、具有较高质量的金属产品自然成为了工业制造不可忽视的重要话题。然而,由于钢铁在加工时受到机械设备、加工方式和材料质量等多种因素的影响,在工业的实际生产加工以及保存运输的过程中,金属产品表面容易产生各种各样的缺陷,如:划痕(scratch)、烧损(burn)、生锈(dense_rust)以及孔洞(pitting)等。这些缺陷不仅降低了产品的外观质量,还可能会对机器加工设备造成损坏,甚至还会降低衍生产品的安全性能及使用寿命[1]。

过去依靠人工检测的方法不仅增加了企业的人力成本,而且检测速度比较慢[2]。随着计算机硬件的不断发展,出现了机器视觉检测方法。机器视觉检测方法分为机器学习与深度学习。机器学习方法的图像特征提取过程不仅依赖专家的设计策略,而且仅适用于单一场景的检测,不同场景下的物体会弱化机器学习方法的识别能力。而基于深度学习算法能够自动捕捉物体的高级特征,在学习特征的同时还能通过模型的验证优化学习过程,使其能够减少人工设计特征提取方法所带来的不确定性的影响,并且深度学习拥有很强的泛化能力,对于处理不同场景与带有噪声的图像具有较高的鲁棒性。

目前,应用在目标检测领域的深度学习方法有 Faster R-CNN [3]、SSD [4]以及 YOLO [5]等算法。其中, Faster R-CNN 是当前主流的双阶段检测算法。双阶段检测算法[6]会先生成多个候选区域,然后将候选区域输入到神经网络中,并对其进行目标检测。这种算法具有较高的检测准确性,但由于需要先筛选出候选框,因此检测速度较慢。而 SSD 和 YOLO 为单阶段端到端检测算法,与双阶段检测算法相比,单阶段检测算法[7]能够直接将输入的图片转化为检测结果,跳过候选区域的生成,其计算量更小,且检测速度较快,能够运行在低成本的嵌入式设备上。与 SSD 相比, YOLO 具备更高的检测速度和准确性,并且在处理小目标时检测精度更高,因此广泛用于表面缺陷检测领域。例如, Cui 等人[8]针对钢材表面缺陷检测中误检和漏检率高等问题提出了 MCB-FAH-YOLOv8,该方法引入了改进后的 CBAM 和预测头 FAH,并将 FPN 替换为 BiFPN,这种方法虽然降低了网络的运行速率,但模型对于微小物体和密集目标

的检测能力却得到了增强。Huang 等人[9]在 YOLOv8s 的基础上提出了 WFRE-YOLOv8s 模型,该方法使用了 CFN 模块来代替 C2f 模块,并引入了 RFN 模块,同时,将 EMA 注意力模块融入主干网络。该改进方法不仅使特征提取网络能更好地融合不同尺度的特征,提高了模型的检测精度,而且还提高了网络的检测速度。Gao 等人[10]注意到钢材表面微小缺陷的识别精度不高等问题,通过改进 YOLOv7 模型,提出了 SRN-YOLO 模型,该改进方法采用分割残差卷积网络捕捉梯度特征信息,构建特征融合金字塔网络以最小化特征损失,通过增大模型参数量牺牲了检测速度,极大地提高了检测准确率。Liu 等人[11]提出一种改进 YOLOX 的钢材表面缺陷检测算法,通过在 Backbone 部分引入改进的 SE 注意力机制,并增加最大池化层分支以增强比较重要的特征通道;此外,该作者还在 Neck 部分融合了 ASFF 模块,使得模型能够更好地融合不同尺度的特征;最后,该作者还通过将 IOU 损失函数替换为 EIOU 损失函数,来改善模型的定位精度。有的作者还改进了 YOLOv8n,结合引入 CA 机制减少冗余特征提取,并使用幽灵卷积减少模型参数。

而为了更好地解决原有目标检测算法难以精确识别小对象、少特征缺陷的问题,减少原有算法因大量堆叠模块而产生检测精度损失的影响,本研究针对 YOLOv8n 网络的模块进行了创新性的改进,提出了一种基于多尺度融合的新模型 YOLOv8n-FMR,在保证实时性的情况下,充分提高了模型的检测精度。模块的改进部分分布在 YOLOv8n 的主干网络和颈部网络。其中,在主干网络引入了 Focal Modulation [12] 焦点调制模块,不仅有效地整合了细节特征与上下文信息,还能降低了模型的参数量和复杂度,这使其能够替代 SPPF 模块进行更好的特征提取,降低 SPPF 中因大量堆叠的最大池化层[13]所造成的面对多细节图像时处理能力下降的影响。而在 YOLOv8n 的主干网络与颈部网络中,则引入了轻量化的多尺度特征融合模块 MSC 以及能够更好检测小目标的模块 RFD [14]。在 MSC 模块中,我们创新性地将 MSC 模块与 7×7 大核卷积进行融合,这使得 MSC 模块能够更好地融合不同尺度的缺陷图像的特征,并且能够显著提升模型对尺度变化的适应性,这不仅提高了模型提取不同大小特征的能力,还能够增加模型的计算效率。最后,为了进一步提升模型在小缺陷图像的特征提取方面的性能,本文还将用于遥感领域的 RFD 模块与 YOLOv8n 进行结合。

2. 相关方法

2.1. 基准模型

通过对 YOLO 系列的模型进行测试分析,本研究最终选定 YOLOv8n 作为改进的基准模型。YOLOv8n 为 YOLO 系列的第 8 个版本,不仅检测速度快、检测精度高,而且可用于多种检测任务,如:图像分类任务和实例分割任务。目前 YOLOv8 中使用最多的模型为 YOLOv8n,它是 YOLOv8 的最小版本,由于其轻量化的特性,更容易加载到嵌入式设备上面,目前常应用于工业目标检测领域的检测。然而,面对表面缺陷检测,YOLOv8n 模型目前依然存在许多显著的问题:在 SPPF 模块中,存在池化层大量堆叠的情况,虽然池化层计算速度快,但是在面对多细节的物体时,池化层的特征提取效果仍会受到一定的影响。而在面对小目标检测时,普通的下采样卷积模块对于小目标的特征提取并不十分有效,容易导致其特征丢失。因此针对上述 YOLOv8n 模块的不足之处,我们做了如下的改进。

2.2. 引入 Focal Modulation 焦点调制模块

SPPF 作为 SPP 的改进版本,能够处理多尺度物体的特征,确保模型在面对各种大小的目标时,能够获得丰富的信息。这种特性使得 SPPF 模块在小目标或大目标频繁出现的情况下,模型的准确性大幅提升。然而,SPPF 中堆叠的多个最大池化操作虽然加快了模块的处理速度,但却降低了网络在感知复杂背景以及面对多细节物体时的处理能力,这使得我们将目光转向自注意力机制。自注意力机制能够让图片的全局信

息得到交互，并使网络提取特征的能力得到加强。然而，自注意力计算复杂度较高，不能很好地在嵌入式设备上运行。这时，焦点调制模块(Focal Modulation Module)成为了能够取代自注意力机制的一种新方法。

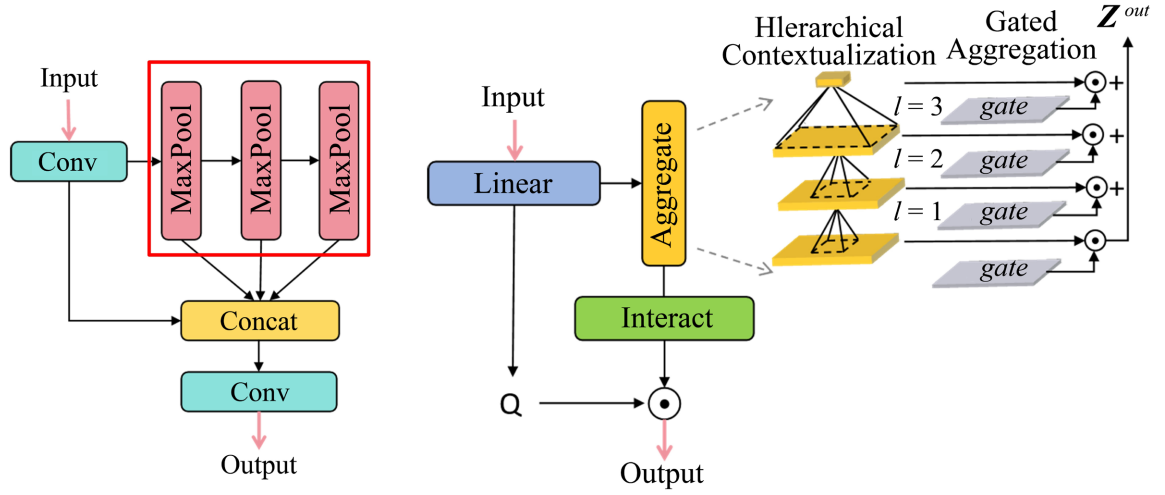


Figure 1. SPPF module and focal modulation module
图 1. SPPF 模块和焦点调制模块

Focal Net 使用一组深度可分离卷积，通过对短程到长程的视觉语境进行编码，将不同粒度的空间上下文编码为总结标记(Summarized Tokens)，如图 1 所示，然后根据查询内容选择性地融合到查询中。如式(1)所示，

$$y_i = q(x_i) \odot M(x_i, X) \quad (1)$$

其中 M 表示聚合过程， $q(\cdot)$ 是一个查询映射函数， $\odot$ 是按元素的乘法运算符。

与自注意力 SA 相比，这种先聚合再交互的做法使 Focal Net 运算效率得到明显提升。这种使用分层语境化感知输入的图像的方式，能够将输入的特征映射投射到新的特征空间中。相比 SPPF 中大量堆叠的最大池化层，Focal Net 使用的 L 个 DW 卷积是可学习的，并且具有结构感知能力，能够使 Focal Net 获得更大的感受野以捕获全局上下文的信息，其中，第 l 线性卷积层如式(2)、(3)所示。

$$\mathbf{Z}^0 = f_z(\mathbf{X}) \in \mathbb{R}^{H \times W \times C} \quad (2)$$

$$\mathbf{Z}^l = f_a^l(\mathbf{Z}^{l-1}) \triangleq \text{GeLU}(\text{Conv}_{dw}(\mathbf{Z}^{l-1})) \quad (3)$$

此外，该作者还使用门控聚合机制来聚合不同层级的线性卷积层，其公式如下所示，

$$\mathbf{Z}^{out} = \sum_{l=1}^{L+1} \mathbf{G}^l \odot \mathbf{Z}^l \quad (4)$$

其中， $\mathbf{G}^l \in \mathbb{R}^{H \times W \times 1}$ 是第 l 层权重。最后，作者还使用额外计算层 $h(\cdot)$ 获得最终的输出调制器 $M = h(\mathbf{Z}^{out})$ 。

2.3. 使用 MSC 模块代替 C2f 进行特征提取

在深度学习中，为了提高模型的学习特征的能力，往往会选择叠加网络的层数以及增加卷积核的大小、个数等，这些做法虽然能在一定程度上稍微提高模型的测试精度，但是在面对指数级增长的参数量时，微小的精度提升将无法抵消这种巨大的代价。这使得我们将目光转向了扁平化的多尺度卷积操作，通过结合不同尺寸大小的卷积核进行并行计算，能够在减少参数的条件下提取丰富的特征。

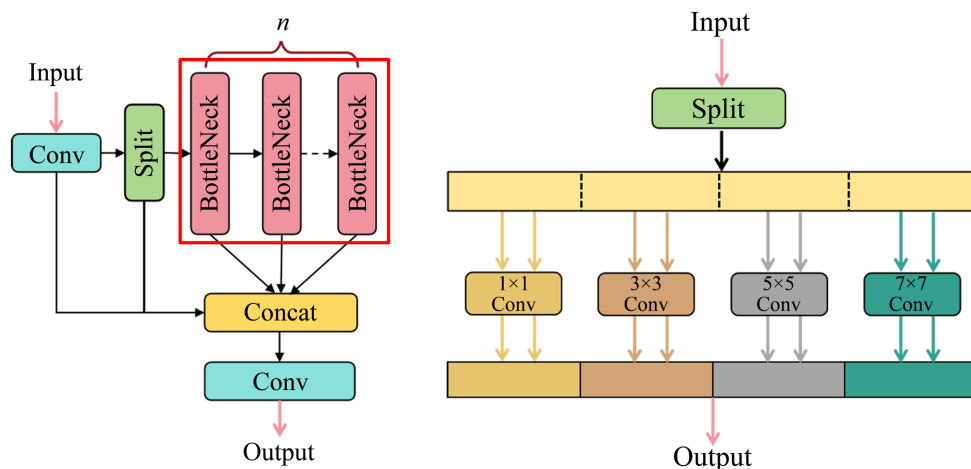


Figure 2. C2f module and MSC module
图 2. C2f 模块和 MSC 模块

此外，本次实验还进一步改进了 MSC 模块，增加了 7×7 的大核卷积，如图 2 所示，该模块将在一定程度上降低模型的计算复杂度。同时实验也表明，该模块能够增加了网络的特征提取能力，并且能够在一定程度上减少 C2f 中因大量堆叠的 Bottleneck 而造成的精度损失问题。模块参数数量的减少不仅有助于降低模型的复杂度，还能进一步降低小数据集上的过拟合，提升模型在小数据集上的性能，使得模型在实际应用中更加具有鲁棒性。

2.4. 代替卷积模块的 RFD

当网络接收原始图像时，原始图像通常包含大量像素级的噪声和冗余信息，这些噪声会对特征提取造成影响。因此，尽可能地保留图像特征，减少噪声对图像的影响，是主干网络中下采样模块需要解决的事情。虽然 YOLOv8 中 3×3 的传统卷积模块计算速度较快，然而，这种下采样方法对于小目标的特征描述并不十分有效。该操作容易导致小目标的特征丢失，并增加了原始噪声对特征图产生干扰的可能性，从而影响了最终性能。

因此，我们引入了鲁棒特征降采样(RFD)模块，该模块分为两种形式：分别是 shallow RFD (SRFD)和 Deep RFD (DRFD)，用来取代 YOLOv8 网络的普通卷积模块，其中 shallow RFD (SRFD)取代网络中第一个卷积模块，而 Deep RFD (DRFD)则用来取代其他卷积模块。

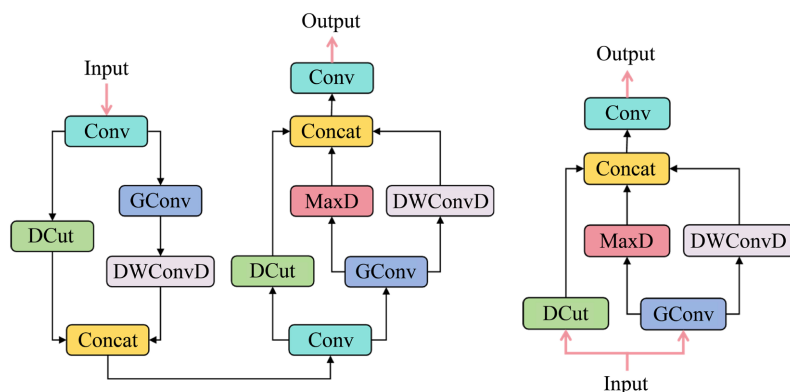


Figure 3. SRFD module and DRFD module
图 3. SRFD 模块和 DRFD 模块

在 DRFD 中, 输入的特征图输入特征图 Y 通过 DConv、DCut 和 DMax 进行下采样, 分别得到输出特征图 X1、X2 和 X3。随后, X1、X2 和 X3 被连接在一起, 形成一个通道数为 6C 的特征映射。最后, 使用步长为 1 的 1×1 卷积层将通道数减少到 2C。这种设计确保了在下采样过程中, 减少了语义信息的丢失, 如图 3 左所示。

在 SRFD 模块中, 输入特征图 X 经过步长为 1 的 7×7 卷积处理, 将通道数从 3 扩展到 1/4C, 该卷积操作能够融合局部特征和过滤冗余信息, 并增强原始图像中重要特征。接着, 特征图 x 被复制为 y1 和 y2, 其中 y1 通过切片下采样(DCut)保留原始特征信息, 同时生成一个大小减半的矩阵。y2 则分别经过步长为 1 和步长为 2 的 3×3 降采样卷积块, 该卷积块将输出通道数提升为之前的两倍, 从而整合局部特征并进一步减少特征图的尺寸。接着, y1 与 y2 合并成通道数为 C 的总特征图, 并经过步长为 1 的 1×1 卷积将通道数减少到 1/2C。最后, 再对上一个输出进行卷积操作后送入 SRFD 的后半部分, 该后半部分的模块结构与 DRFD 模块一致。总而言之, SRFD 模块在下采样过程中尽量避免最大池化的操作, 确保了在不同尺度上的特征表达能力, 同时在保留关键特征的基础上, 降低了原始图像中的噪声的数量, 如图 3 右所示。

3. 实验

3.1. 实验环境

本文实验环境的操作系统为 Windows 10, CPU 型号为 Intel(R) Core(TM) i5-12400F, 运行内存 32GB, GPU 型号为 NVIDIA Geforce RTX-4060ti 16G。实验采用 SGD 优化器进行训练, 并且在训练时, 对训练数据使用 Mosaic 增强, 增加模型的泛化能力。

同时, 将学习率设置为 0.01; 动量默认设置 0.937; 权重衰减系数默认设置 0.0005; 迭代次数为 300。

3.2. 实验数据集

本研究主要选用 NEU-DET 钢带表面缺陷数据集, 该数据集收集了热轧带钢表面的六种典型缺陷, 分别为裂纹(crazing)、内含物(inclusion)、斑块(patchs)、点蚀表面(pitted_surface)、轧制氧化皮(rolled-in_scale)和划痕(scratches)缺陷。该数据集包括 1800 张钢材缺陷图像, 每张图像的分辨率为 200×200 像素, 涵盖了不同类型和不同程度的钢带表面缺陷, 能够充分评估和比较不同缺陷检测算法在实际应用中的性能。NEU-DET 数据集已成为金属表面缺陷检测领域中广泛使用的公共数据集之一。在本次实验中, 将整个 NEU-DET 数据集分为三份, 分别为训练集, 验证集和测试集, 其比例为 8:1:1, 即训练集 1440 张、验证集 180 张、测试集 180 张。同时, 为了保证数据集的识别准确率、兼顾网络的推理速度, 我们将输入网络的图片大小提升为 384×384 像素, 使该网络更适合在嵌入式设备上运行。

3.3. 评价指标

本文使用准确率(Precision)、召回率(Recall)、平均准确率均值(mAP)、参数量(Params)、计算量(GFLOPs)这五个指标来评估模型的性能。其中, 准确率(Precision)为预测正样本的数据中, 真实值为正样本的比例, 该值反应了模型预测正样本的准确性, 如式(5)如所示。

$$P = \frac{T_p}{(T_p + F_p)} \quad (5)$$

召回率(Recall)为正样本能够预测正确的概率, 该值反映了模型检测的查全率, 如式(6)如所示。而平均准确率均值(mAP)衡量了模型在多个类别上的平均检测精度, 对于每个类别, 可计算其对应的 AP 值, 再取均值可得到平均准确率均值(mAP)。平均准确率均值反映的是模型整体的检测精度, 该值的大小能够

直接作为评估模型优劣的重要指标，该式(7)如所示。

在式(5)~(7)中，TP 表示模型正确识别到的目标数量、FP 表示模型未正确识别到的目标数量、FN 表示模型错误检测到的目标数量、k 表示数据集中类别数量。

$$R = \frac{T_p}{(T_p + F_N)} \quad (6)$$

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

参数量(Params)反应的是模型的大小，即模型占用内存空间的大小。而计算量(GFLOPs)反应的是模型的训练速度和推理速度。计算量越小，表示模型推理速度越快。较小的计算量能够使模型更好运行在工业嵌入式设备上。

3.4. 消融实验

为了探究改进的各个模块对模型整体性能的影响，本文以 YOLOv8n 为基准模型，设计了如下的消融实验，以充分验证各个融合模块的有效性。本次消融实验由 4 组实验构成，为了保证实验的准确性，4 组实验所采用的环境及训练参数均保持一致，实验结果如表 1 所示。

Table 1. Ablation experiments

表 1. 消融实验

No.	YOLOv8n	MSC	Focal	RFD	mAP/%	Recall/%	Precision/%	Params/M	GFLOPs
1	✓	×	×	×	70.1	68.5	70.5	3.1	8.2
2	✓	✓	×	×	70.7	64.1	70.1	2.8	7.9
3	✓	✓	✓	×	71.2	68.9	68.8	2.9	8.0
4	✓	✓	✓	✓	72.8	69.8	70.4	3	9.5

第一个实验未融合模型，只采用 YOLOv8n 基准模型进行实验，第二个实验在基准模型的基础上融合了 EMSCP 模块，从表中可看出，该模块不仅降低了基准模型的计算量和参数值，使其分别降低了 0.3 M 和 0.3 GFLOPs，而且还使基准模型的 mAP 检测精度提高了 0.5%，这也意味着 EMSCP 可以有效结合来自不同尺度的特征，减少因堆叠不同尺寸的特征提取层而产生的精度损失问题，并有效降低模型的复杂程度，使模型更加易于优化。

然而，EMSCP 模块带来精度提升的同时却使基准模型的召回率和准确率分别降低了 0.4%和 4.4%。为了降低融合 EMSCP 模块所带来的召回率和准确率减少的影响，第三个实验在实验 2 的基础上融合了 Focal Modulation 模块，从实验结果中可看出，相比实验 2，Focal Modulation 在牺牲了少量的计算速度、增加少量参数量的情况下，不仅提高了 0.5%的 mAP 检测精度，而且还弥补了 EMSCP 所带来的召回率减小问题。这意味着引入 Focal Modulation 模块能够保证模型在计算量和参数量不变的情况下，提升模型对局部图像的特征提取能力，从而提高网络的整体检测精度。

最后，为了提高模型整体的检测性能，实验 3 引入了 RFD 模块，虽然融合后模型的参数量没有发生较大变化，但是计算量却出现明显的增加，即该模块牺牲了一定的检测速度，换取了检测精度。相比 YOLOv8n 基准模型，最终模型的计算量增加了 1.3 GFlops，而 mAP 检测精度、Recall 召回率分别提升了 2.7%，1.3%。从实验 3 结果可看出 RFD 模型能够增加特征提取能力，从而提高模型的检测精度。

综上所述，相比 YOLOv8n，改进的最终模型在缺陷图像检测的准确性上实现了显著提升，同时，其轻量化的改进特性也适合在工业嵌入式设备上运行。

同时，为了明确改进方案在各个缺陷类型上的检测精度变化，本文还将最终改进模型与原始 YOLOv8n 模型进行 mAP 检测精度对比，结果如表 2 所示。crazing 缺陷类型的 mAP 提升了 4.1%，inclusion 缺陷类型的 mAP 提升了 4.3%，而 patches 缺陷类型的 mAP 保持稳定，此外，pitted_surface 缺陷和 rolled-in_scale 缺陷的 mAP 分别提升了 2.0%和 0.4%。我们注意到，在最终改进方案下，crazing 缺陷类型和 inclusion 缺陷类型的检测精度得到了明显的提升。

Table 2. Comparison of mAP detection accuracy (%) of the model in improvement on different defect categories
表 2. 改进前后模型在不同缺陷类别上的 mAP 检测精度(%)对比

No.	Model	Crazing	Inclusion	Pitted Surface	Patches	Scratch	Rolled-in Scale	All
1	YOLOv8n	27.8	66.5	91.5	97.5	50.4	86.8	70.1
4	YOLOv8n-FMR	31.9	70.8	93.5	96.1	57.4	87.2	72.8

3.5. 对比实验

为了进一步验证改进模型的优越性，本文在相同的实验环境和参数设置下，引入 YOLOv3-tiny、YOLOv5n、YOLOv6n、YOLOv10n 等经典模型进行实验对比，并通过实验结果对上述模型的性能及检测精度进行比较分析，实验结果如表 3 所示。

Table 3. Performance comparison between mainstream algorithms
表 3. 主流算法之间性能对比

Model	mAP/%	Params/M	GFLOPs
YOLOv3-tiny	69.3	12.1	18.9
yolov5	69.4	2.5	7.1
yolov6	71.2	4.2	11.8
yolov8	70.1	3.0	8.1
yolov10n	69.6	2.2	6.7
Ours	72.8	3	9.5

由表 3 可知，YOLOv3-tiny 在上述对比模型中表现最差，其 mAP 检测精度为 69.3%；同时其参数量、计算量在上述对比模型中也是最大的，分别为 12.1 M 和 18.9 GFLOPs，这意味着在性能和计算效率上都处于较差水平。而在上述对比模型中，YOLOv6 模型检测精度最高，其 mAP 检测精度为 71.3%，然而其参数量和计算量却同样相对偏大，分别为 4.2 M 和 11.8 GFLOPs。而本文的改进模型则取得了较高的提升，相对 YOLOv6 模型，其 mAP 检测精度提升了 1.6%，并且参数量、计算量也同时更低，分别少了 1.2 M 和 2.3 GFLOPs，相当于降低 28.5%参数量和 19.4%的计算量。这意味着改进后的模型显示出了更明显的优越性。综上所述，本研究针对 YOLOv8n 基准模型进行了多次改进实验，改进实验融合了 MSC 模块、Focal Modulation 模块和 RFD 模块，实验后的结果显示改进后模型不仅使缺陷图像的检测性能得到了提升，而且其轻量化的特性还使该模型能够部署在计算资源受限的嵌入式设备上。

4. 结论

本文根据钢材表面缺陷检测中存在的实际问题、详细分析了 YOLOv8n 模型中各个模块的不足之处，

提出了 YOLOv8n-FMR 模型, 试图通过融合其他模块来降低 YOLOv8n 模型中原有模块的缺陷所带来的影响, 从而增加原模型的特征提取能力, 提高模型对图像的缺陷检测精度。首先, 通过融合 Focal Modulation 模块, 在一定程度上解决了 SPPF 模块中多个堆叠的最大池化层会降低网络在感知复杂背景或者多细节物体的处理能力的问题, 同时也能够增加原模型的特征提取能力。其次, 为了降低 C2f 中因 BottleNeck 大量堆叠而造成的精度损失问题, 本文引入了具有 7×7 大核卷积的 MSC 模块, 该模块将在一定程度上降低模型的计算复杂度。同时消融实验也表明, 相比 C2f 模块, MSC 模块能够更进一步增加网络的特征提取能力。最后, 为了能更好地处理对象较小、特征较少的缺陷目标, 进一步提高网络模型对微小缺陷的识别精度, 本文引入了运用在遥感领域的下采样方法 RFD。实验结果表明, 改进后的 YOLOv8n-FMR 模型在 NEU-DET 数据集上的 mAP 达到了 72.8%, 比原模型提高了 2.7%, 并且模型的参数量也比原模型下降了 0.1 M。同时, 改进后的模型计算量仅为 9.5 GFLOPs。与当前先进模型相比, 本文提出的模型不仅实现了轻量化的设计, 而且在检测效果上也更为出色, 这使得它更适合在工业终端设备上部署。

但本文仍有许多不足之处需要进一步研究与改进, 尤其是使用 RFD 模块后导致模型计算量大幅增加的问题, 后续需要考虑使用数据增强技术和更高效的注意力算法提升模型对这两类缺陷的检测精度, 进一步提升模型的整体精确度, 并兼顾模型的计算效率。

参考文献

- [1] Gui, Z. and Geng, J. (2024) YOLO-ADS: An Improved YOLOv8 Algorithm for Metal Surface Defect Detection. *Electronics*, **13**, Article 3129. <https://doi.org/10.3390/electronics13163129>
- [2] 宋育斌, 孔维宾, 陈希, 等. 钢材表面缺陷检测研究综述[J]. 软件导刊, 2024, 23(3): 203-211.
- [3] 任钰, 刘全金, 黄忠, 等. 基于 Faster R-CNN 与迁移学习的口罩佩戴检测算法[J]. 安庆师范大学学报(自然科学版), 2021, 27(4): 25-30.
- [4] 卢云聪. 基于 SSD 算法的口罩佩戴检测模型[J]. 科技与创新, 2022(18): 101-103, 109.
- [5] 王婉婷, 姜国龙, 褚云飞, 等. 从 RCNN 到 YOLO 系列的物体检测系统综述[J]. 齐鲁工业大学学报, 2021, 35(5): 9-16.
- [6] 仲家栋. 基于双阶段网络的小目标检测算法研究[D]: [硕士学位论文]. 成都: 电子科技大学, 2024.
- [7] 董甲东, 郭庆虎, 陈琳, 等. 深度学习中单阶段金属表面缺陷检测算法优化综述[J]. 计算机工程与应用, 2025, 61(4): 72-89.
- [8] 崔克彬, 焦静颐. 基于 MCB-FAH-YOLOv8 的钢材表面缺陷检测算法[J]. 图学学报, 2024, 45(1): 112-125.
- [9] 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>
- [10] Gao, S., Chu, M. and Zhang, L. (2024) A Detection Network for Small Defects of Steel Surface Based on YOLOv7. *Digital Signal Processing*, **149**, Article ID: 104484. <https://doi.org/10.1016/j.dsp.2024.104484>
- [11] 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.
- [12] Yang, J.W., et al. (2022) Focal Modulation Networks. arXiv: 2203.11926.
- [13] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet Classification with Deep Convolutional Neural Networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, Lake Tahoe, 3-6 December 2012, 1106-1114.
- [14] Lu, W., Chen, S., Tang, J., Ding, C.H.Q. and Luo, B. (2023) A Robust Feature Downsampling Module for Remote-Sensing Visual Tasks. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, 1-12. <https://doi.org/10.1109/tgrs.2023.3282048>