

基于 YOLOX 改进模型的 金属表面缺陷检测

车国霖, 傅家辉

(昆明理工大学 信息工程与自动化学院, 昆明 650504)

摘要: 针对金属表面缺陷检测中模型精度与推理速度难以兼顾的问题, 提出一种基于 YOLOX 模型改进的 SWE-YOLOX 检测算法。首先, 为解决复杂背景干扰大及缺陷尺度变化显著的问题, 引入混洗通道注意力模块, 增强特征表达能力并抑制无关信息; 其次, 针对缺陷边缘模糊、纹理不清晰的问题, 融合小波卷积以提升频域特征提取能力, 从而强化细节信息表达; 最后, 将原有交并比(IoU)损失函数替换为 EIou 损失函数, 以优化预测框与真实框的回归精度。实验结果表明, 该方法在数据集 NEU-DET 上平均表面精度(mAP)达 76.3%, 较 YOLOX 模型提升 3.86 百分点, 且在参数量与计算复杂度基本不增加的前提下保持了较快的推理速度。

关键词: YOLOX 模型; 注意力模块; 小波卷积; 损失函数; 金属表面缺陷

中图分类号: TP183 **文献标志码:** A **文章编号:** 1671-5489(2026)03-0603-14

Metal Surface Defect Detection Based on Improved YOLOX Model

CHE Guolin, FU Jiahui

(Faculty of Information Engineering and Automation,
Kunming University of Science and Technology, Kunming 650504, China)

Abstract: Aiming at the challenge of balancing model accuracy and inference speed in metal surface defect detection, we proposed an improved SWE-YOLOX detection algorithm based on the YOLOX model. Firstly, in order to solve the problem of large interference of complex backgrounds and significant variation in defect scales, we introduced a channel shuffle attention module to enhance feature expression ability and suppress irrelevant information. Secondly, aiming at the problem of unclear defect edges and weak texture features, we incorporated wavelet convolution to improve the extraction ability of frequency-domain features, thereby enhancing the expression of detailed information. Finally, the original intersection over union (IoU) loss function was replaced with an enhanced intersection over union (EIou) loss function to optimize the regression accuracy between predicted boxes and ground truth boxes. Experimental results show that the proposed method achieves a mean average precision (mAP) of 76.3% on the NEU-DET dataset, which is 3.86 percentage point higher than that of YOLOX model. It also maintains a fast inference speed without increasing the number of

收稿日期: 2025-01-03.

第一作者简介: 车国霖(1975—), 男, 汉族, 硕士, 副教授, 从事智能控制及应用的研究, E-mail: 864111506@qq.com. **通信作者简介:** 傅家辉(2000—), 男, 汉族, 硕士研究生, 从事目标检测的研究, E-mail: 1064084760@qq.com.

基金项目: 国家重点研发计划子课题项目(批准号: 2017YFB0306405).

parameters and computational complexity.

Keywords: YOLOX model; attention module; wavelet convolution; loss function; metal surface defect

0 引言

金属表面缺陷检测是工业制造中至关重要的一环，直接关系到产品的质量、安全性和使用寿命。随着现代制造业对精度、质量和可靠性要求的不断提升，传统人工检测方法已无法满足高效、准确的需求。金属材料广泛应用于航空航天、汽车、船舶等关键领域，表面缺陷如裂纹、划痕、凹坑等，虽然在早期肉眼难以察觉，但在实际使用中可能会导致结构疲劳、断裂甚至灾难性事故。因此，提高对金属表面的缺陷检测精度有一定的应用价值。

目前，针对不同种类的金属表面缺陷，常用的缺陷检测方式主要包括：磁粉检测法、渗透检测法、涡流检测法、超声波检测法、X 射线检测和机器视觉检测法^[1]。除机器视觉外，其他检测方式都较依赖材料的物理、化学特性，且需要专业人员操作和判断，结果受人为因素影响较大，同时成本也较高。而在机器学习中，传统的机器视觉依赖手工设计特征^[2-8]，效率低且难以适应复杂场景。而深度学习中的卷积网络由于其特性是自动学习特征，擅长处理高维度、非线性和复杂的视觉任务^[9]，以及持续的深度学习发展^[10-12]和硬件进步，使深度学习在解决机器视觉中的复杂任务和大规模数据处理需求方面展现了显著优势，被广泛应用于缺陷检测任务中^[13-15]。文献^[16]中的 MaskR-CNN 检测算法解决了传统卷积神经网络(convolutional neural network, CNN)难以准确检测金属表面存在大量纹理复杂、光照不均和对比度低的缺陷。Kou 等^[17]将 YOLOv3 网络中基于锚框的结构修改为无锚框特征选择机制，并在模块中引入专门设计的密集卷积块，有效提高了钢带表面缺陷检测的精度。Shi 等^[18]通过改进 SSD(single shot multiBox detector)算法，引入注意力机制、多层特征融合提高 PCB(printed circuit board)缺陷检测的精度。刘培勇等^[19]在传统 CNN 的基础上构建了多支路神经网络，引入注意力模块和特征融合的策略，模型在磁瓦图像上做到了高速、高精度的检测性能。刘兰兰等^[20]在主干网络中嵌入卷积块注意力机制，并通过使用 CIoU 优化边界框损失函数提升模型的收敛能力，提高了输电线路缺陷检测的精度。耿志强等^[21]提出了构建连续小波卷积层取代卷积神经网络中的初始卷积层，相比于传统的基于 CNN 在检测精度上有所提升。然而，上述方法在提升检测精度的同时，通常引入了额外的计算开销和模型复杂度，导致推理速度下降，从而限制了其在实时工业场景中的应用。另一方面，也有研究尝试通过改进网络结构提升特征表达能力。Finder 等^[22]提出 WTConv 方法通过引入小波卷积增强模型对纹理和细节特征的提取能力，在一定程度上提升了检测精度。但该类方法同样面临计算复杂度增加的问题，在实际部署中需在精度与效率之间进行权衡。

现有金属表面缺陷检测方法虽然在检测精度方面取得了显著进展，但普遍存在精度提升依赖复杂结构设计、计算开销较大、实时性不足的问题。在实际工业检测场景中，检测系统不仅需具备高精度，还需满足高效率 and 低延迟的要求，因此如何在保证检测精度的同时降低模型复杂度，实现精度与速度的有效平衡，仍是亟待解决的关键问题。

尽管现有方法在一定程度上提升了检测精度，但在复杂工业场景下仍面临多种挑战。图 1 为金属表面缺陷检测中难以检测的类型。由图 1 可见，金属表面缺陷检测主要存在以下三方面难点：

- 1) 图 1(A)和(B)的轧制氧化皮(Rs)缺陷区域与背景区域纹理相似，难以区分，导致模型易误检；
- 2) 图 1(C)和(D)的点蚀面(Ps)缺陷包含的缺陷尺度变化较大，网络难以检测到微小缺陷；
- 3) 图 1(E)和(F)的裂纹(Cr)缺陷纹理不明显，导致网络难以捕捉到缺陷的特征。

基于上述分析，本文提出一种改进模型 SWE-YOLOX。首先，选取 YOLOX 作为基础网络，引入混洗通道的注意力机制，增强多尺度处理的能力，聚焦于缺陷区域，减低背景噪声的影响；其次，引入小波卷积，增强图像的对比度和细节纹理，突出缺陷区域，使检测更敏感；再次，修改网络的损失函数，引入额外的惩罚项，优化网络预测框与真实框的位置关系；最后，结合三者提出一种新的用于钢

材表面缺陷的检测网络 SWE-YOLOX, 并在公开钢材表面缺陷数据集 NEU-DET 上验证模型的有效性, 在不明显增加模型参数量和复杂度的情况下, 有效提高了模型的检测精度。

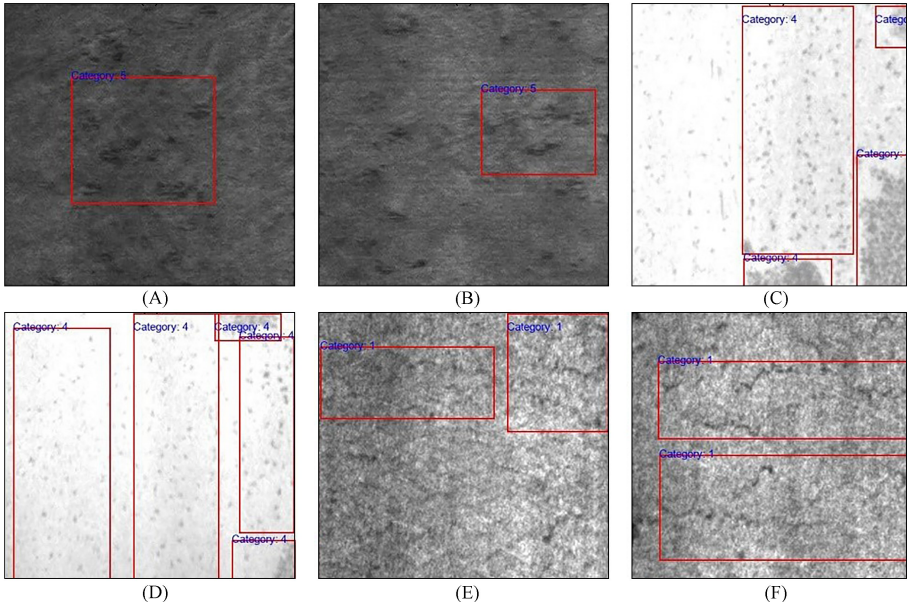


图 1 金属表面缺陷检测中难以检测的类型

Fig. 1 Types that are difficult to detect in metal surface defect detection

1 金属表面缺陷检测模型 SWE-YOLOX

1.1 SWE-YOLOX 网络结构

SWE-YOLOX 网络结构如图 2 所示. SWE-YOLOX 网络主要分为 3 个部分: Backbone, Neck 和 Head. 主干网络 Backbone 主要负责特征的提取, 包括 Focus, Conv, CspLayer, SPPBottleneck 和

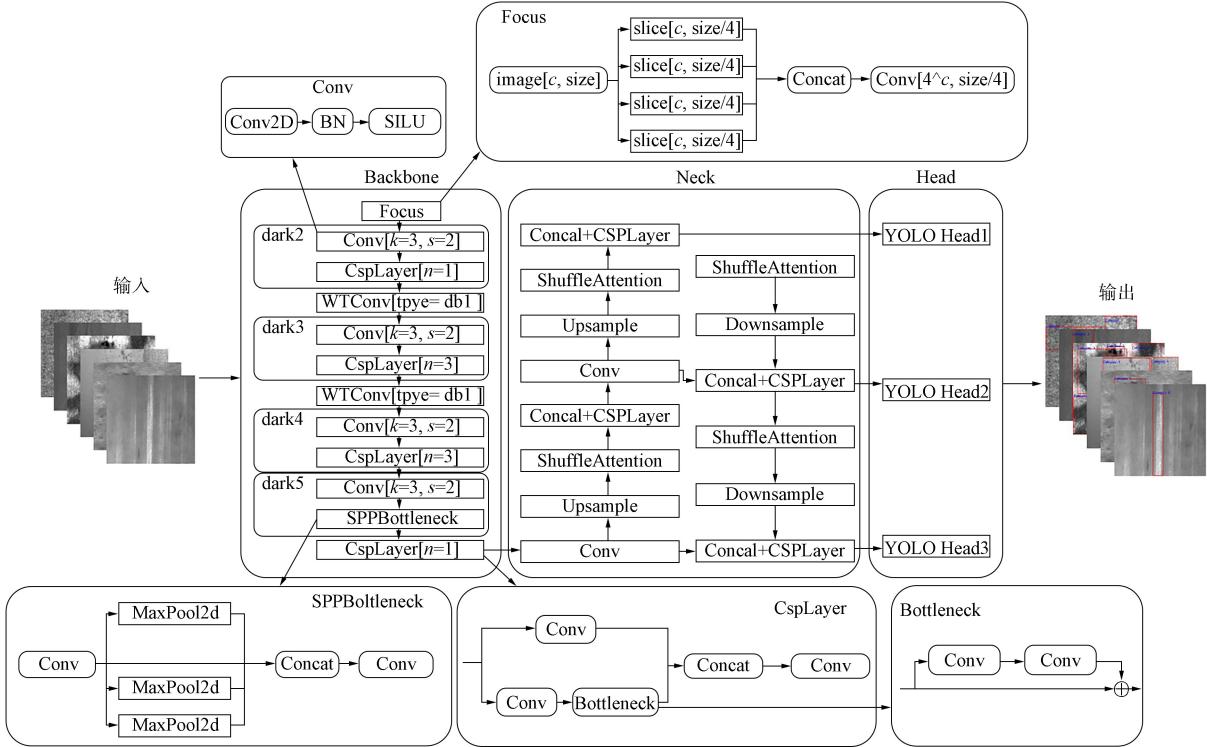


图 2 SWE-YOLOX 网络结构

Fig. 2 SWE-YOLOX network structure

WTConv 模块. Focus 操作将输入的图片分为 4 个切片,然后在通道上进行拼接,将空间信息转化为通道信息,再接一个卷积操作 Conv. Conv 是基本的卷积单元,其操作顺序为二维卷积、归一化和激活值运算. CspLayer 模块的结构为两条分支,一条经过 Conv 卷积操作,另一条经过 Conv 操作后再经过 Bottleneck 操作,此处可多次使用 Bottleneck,这取决于可定义的参数,将两条分支 Concat 拼接,最后再做 Conv 卷积操作. 其中 Bottleneck 主要的结构为残差结构,同样也分为两条分支,一条不做任何操作,另一条做两次卷积操作,最后由网络中的参数决定是否将两个分支相加. SPPBottleneck 的结构为先将输入经过一次卷积操作,然后将结果平行的分别进行多次最大池化操作,再将多个最大池化操作结果与卷积后的结果 Concat 拼接,最后再进行一次卷积操作. WTConv 是本文添加的模块,通过捕捉频域的信息,能增强多尺度特征提取,并保留边缘纹理等细节,且本身所需的参数量较小,可在不增加网络复杂度的情况下提升特征提取能力,理论上对纹理不明显、边缘模糊的特征更有效,Neck 网络主要用来实现不同尺度的特征融合,其主要结构是特征金字塔(FPN)和 ShuffleAttention. FPN 将深层特征图和浅层特征图通过自顶向下的路径进行融合,生成多层特征图,用于处理不同大小的目标;ShuffleAttention 为本文添加的模块,其结构主要为通道混洗、空间注意力和通道注意力三者的结合,通道混洗增强了通道之间的相互依赖性,改进不同层次特征的融合,注意力机制通过对特征图施加注意力权重,使网络更关注有重要信息的通道和空间区域,减少背景的干扰,提升特征表达能力. Head 部分共有 3 个检测头,使用 3 种不同分辨率的特征图进行预测,分别用于检测大、中、小 3 个尺寸的目标.

1.2 ShuffleAttention 模块

ShuffleAttention 模块是一种融合通道注意力、空间注意力和通道混洗机制的改进型注意力结构,其核心目标是在较低计算开销下提升卷积神经网络的特征表达能力. 该模块通过多维度建模特征之间的依赖关系,使网络能在复杂场景中更有效地提取关键信息.

ShuffleAttention 模块首先将输入特征图在通道维度上划分为 G 组,每组分别进行通道注意力和空间注意力操作,从而在分组层面上实现对特征的局部建模. 其次,将各组特征在通道维度上重新拼接,并通过通道混洗操作打乱原有通道顺序,使不同组之间的信息能充分交互,从而缓解分组操作导致的信息隔离问题. 在保证计算效率的同时增强特征表达的全局一致性. 该模块整体结构如图 3 所示. 其输出特征尺寸与输入保持一致,其中 b, c, h, w 分别表示 batch size、通道数、特征图高度和宽度,整体表达式为

$$Y_{SA} = \text{shuffle}[\text{view}(A_s + A_c, b, -1, h, w), 2], \tag{1}$$

其中: view 函数的作用为调整特征图的形状,将得到的 $b * G, c/G, h, w$ 变换为 b, c, h, w ; shuffle 表示对通道排序进行随机混洗排序; Y_{SA} 为模块输出. 在具体实现上, ShuffleAttention 模块对每组特征进一步在通道维度进行分裂,形成上下两条分支,如图 4 所示.

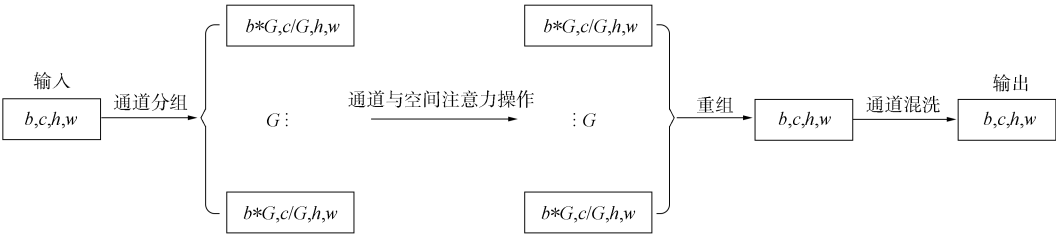


图 3 ShuffleAttention 模块的整体结构

Fig. 3 Overall structure of ShuffleAttention module

图 4 中的上分支为空间注意力机制,其主要过程为:先对输入特征进行组归一化(group normalization, GN),以稳定特征分布;然后通过多层感知机(MLP)进行线性变换和非线性映射;最后将得到的权重与原特征进行逐元素相乘,从而在空间维度上突出关键区域. 该过程可表示为

$$A_s = X_1 \cdot \{\text{MLP}(\text{GN}(X_1))\}, \tag{2}$$

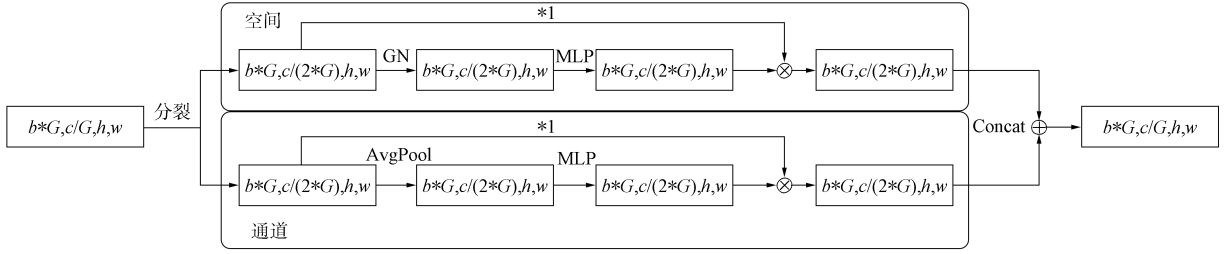


图 4 ShuffleAttention 的通道和空间操作

Fig. 4 Channel and spatial operations of ShuffleAttention

其中 MLP 为简化的多层感知机, MLP 本质为线性变换与非线性激活操作的结合, 也将此操作视为神经元, 此处的操作实际上是多个通道上单神经元的实现. 单个神经元^[23]的通用表达式为

$$y = \sigma \left(\sum_{i=1}^n w_i x_i + b \right). \quad (3)$$

图 4 中的下分支为通道注意力机制, 其主要通过全局平均池化(global average pooling, AvgPool)对特征图进行压缩, 将空间信息聚合为通道描述, 从而获得每个通道的全局响应; 并通过 MLP 学习通道间的依赖关系, 与原特征进行逐元素相乘, 实现对重要通道的增强与冗余通道的抑制, 其过程可表示为

$$A_c = X_0 \cdot \{ \text{MLP}[\text{AvgPool}(X_0)] \}. \quad (4)$$

从机制上看, 通道注意力主要用于建模哪些特征更重要, 空间注意力用于刻画哪些位置更关键, 而通道混洗操作则通过重排通道顺序打破分组导致的信息割裂, 使不同通道间的特征能充分融合. 因此, 该模块在特征选择性与信息交互能力之间取得了有效平衡.

针对金属表面缺陷检测任务的特点, ShuffleAttention 模块的多维度注意力机制有良好的适配性. 一方面, 金属表面通常有较强的重复纹理和较低的对比度, 通道注意力能有效强化与缺陷相关的异常特征; 另一方面, 缺陷区域通常尺寸较小且分布稀疏, 空间注意力有助于模型聚焦潜在的缺陷位置, 抑制背景干扰. 此外, 通过通道混洗增强跨通道的信息流动, 有助于提升特征表达的整体一致性, 从而提高模型对微小缺陷的感知能力和定位精度.

本文在传统 ShuffleAttention 注意力机制的基础上参考刘毅等^[24]的改进方法进行一定改进, 如图 5 中红色部分所示, 在保持空间注意力分支不变的前提下, 对通道注意力机制进行增强. 在原有全局平均池化的基础上引入全局最大池化(global max pooling, MaxPool), 并分别对两种池化结果引入 \$1 \times 1\$ 卷积进行特征映射, 以自适应调整不同统计特征的重要性; 然后将两者结果进行融合, 得到最终的通道权重, 其过程可表示为

$$A_c = X_0 \cdot \{ \text{MLP} \{ \text{Conv}[\text{AvgPool}(X_0)] + \text{Conv}[\text{MaxPool}(X_0)] \} \}. \quad (5)$$

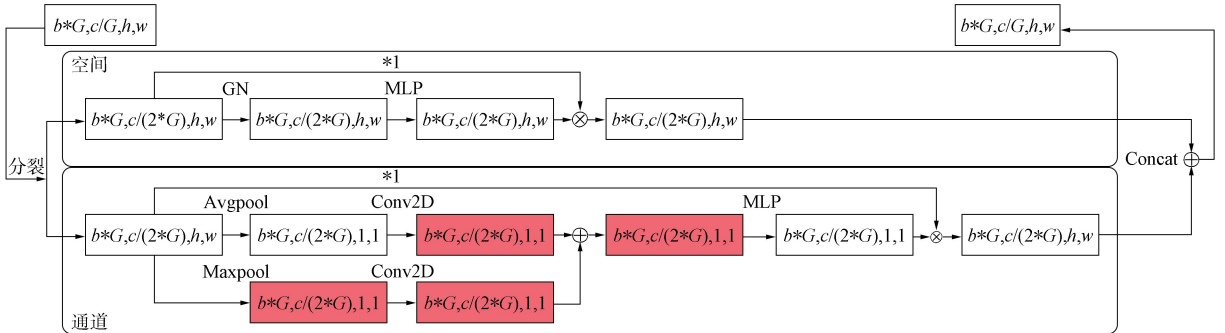


图 5 ShuffleAttention 改进后的通道和空间操作

Fig. 5 Improved channel and spatial operations of ShuffleAttention

该改进通过融合平均池化与最大池化的互补信息, 使模型不仅能捕获全局统计特征, 还能关注局部显著响应, 从而进一步提升通道注意力对复杂特征的建模能力, 增强模型在复杂缺陷场景下的表达

能力.

1.3 小波卷积

小波变换是一种信号处理工具,用于分析信号的频率和时域特征.与 Fourier 变换不同,其可提供信号在时域和频域上的多尺度表示,特别适用于处理非平稳信号和局部特征.在图像处理领域使用离散小波变换.小波变换主要得到了低频分量(LL):包含信号的全局特征或大致轮廓;高频分量(LH,HL,HH):捕捉局部特征,如边缘、纹理、细节等.而小波卷积是将小波变换引入卷积操作的结果,结合了卷积操作的局部性和小波的多尺度分析能力.在 CNN 中,小波卷积用于替代传统的卷积操作,以增强对多尺度特征的提取能力.小波卷积的关键步骤为小波基选择、卷积操作、尺度分解和重构或融合.本文使用小波卷积的主要目的是通过在主干网络中添加小波卷积以达到提高主干网络的增强局部特征能力,增强图像的对比度,突出缺陷区域,使检测更敏感.本文验证了小波卷积的尺度分解对图像细节特征提取能力的影响.尺度分解即为小波卷积中小波卷积的深度.小波卷积的结构如图 6 所示,其中输入图的维度为 B,C,H,W 分别表示 batchsize、通道数、特征图高和特征图宽.在经过小波变换后获得了额外维度,此时共有 5 个维度,如图中的 $[LL],[LH],[HL],[HH]$ 为小波变换后的频域量,包含了高频信息 $[LH],[HL],[HH]$ 与低频信息 $[LL]$;小波变换后进行通道扩展,将多出的维度扩展到通道上,于是得到原来 4 倍的通道数,该操作是为方便后续的卷积操作;在卷积操作后需要恢复到 5 维特征图;将特征图中的高频量和低频量分开存放.本文将上述过程称为小波卷积正变换.图 6 的下半部分称为小波卷积逆变换.将高频分量与低频分量从缓存中取出;高频分量与低频分量 Concat 结合为完整的 5 维特征图;经过小波逆变换后还原得到 4 维特征图;再经过卷积与缩放,得到与输入相同规格的特征图;将得到的结果加上输入特征图即为小波卷积逆变换的全过程.小波卷积深度的含义为多次进行小波卷积正变换与小波卷积逆变换,如深度为 3,则共进行 3 次频域分解,且下次小波变换的输入为上一层小波变换的低频量;将每次小波变换得到的高频与低频量放入缓存中;而小波逆变换则是多次从缓存中取出高频与低频分量,将得到的多次小波变换频域信息全部还原;最后经过卷积与缩放,保证图像尺寸与输入相同,再与输入相加作为最后的输出.

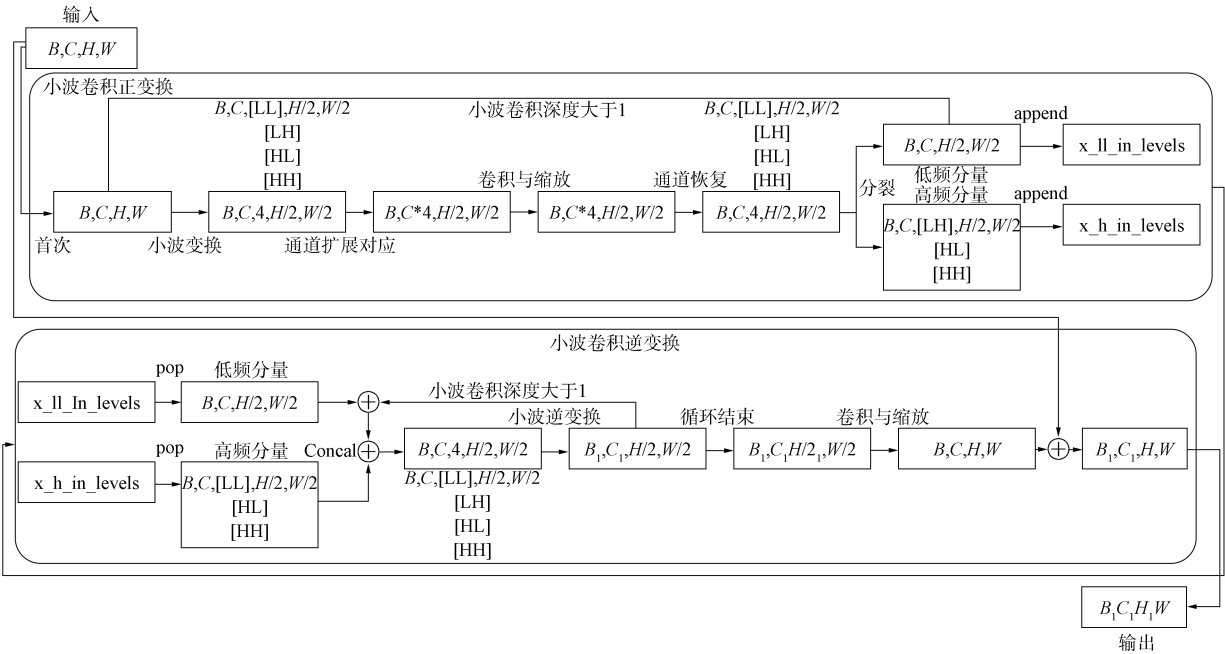


图 6 小波卷积结构

Fig. 6 Wavelet convolution structure

1.4 损失函数

YOLOX 网络中包含 3 种损失:回归损失、种类损失和目标损失.本文讨论的损失函数主要为 YOLOHead 预测的回归损失.YOLOX 网络中回归损失的计算方式为传统的交并比(IoU)损失.这里

研究其他类型的 IoU 回归损失对网络学习特征能力的影响, 主要讨论 EIoU 损失、CIoU 损失和 WIoU 损失带来的影响. 相比传统的 IoU 损失只计算预测框与真实框的交并比, EIoU 损失函数在 IoU 的基础上添加了对框的中心距离、宽高差的惩罚, 旨在解决 IoU 只关心重叠区域而忽略了框中心的位置差异及宽度和高度差异的缺点. EIoU 的计算公式为

$$\text{EIoU} = 1 - \text{IoU} + \frac{d_{\text{center}}^2}{c_{\text{diag}}^2} + \frac{(\Delta w)^2}{c_w^2} + \frac{(\Delta h)^2}{c_h^2}, \tag{6}$$

其中 d_{center} 为预测框和真实框中心点之间的欧氏距离, c_{diag} 为包含预测框和真实框最小闭包矩形的对角线长度, Δw 为预测框和真实框的宽度差异, Δh 为预测框和真实框的高度差异, c_w 为最小闭包矩形的宽度, c_h 为最小闭包矩形的高度. CIoU 和 EIoU 类似, 两者都是在 IoU 的基础上添加了框中心点的距离和预测框宽高差异的惩罚. 但 CIoU 更倾向于宽高比的差异, 即预测框的形状. 其计算公式为

$$\text{CIoU} = 1 - \text{IoU} + \frac{d_{\text{center}}^2}{c_{\text{diag}}^2} + \alpha \times v, \tag{7}$$

其中 d_{center} 为预测框和真实框中心点之间的欧氏距离, c_{diag} 为包含预测框和真实框最小闭包矩形的对角线长度, α 为宽高比调整系数, v 为宽高比差异度量. v 的计算公式为

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w_{\text{gt}}}{h_{\text{gt}}} - \arctan \frac{w_{\text{pred}}}{h_{\text{pred}}} \right)^2, \tag{8}$$

其中 w_{gt} 和 h_{gt} 分别为真实框的宽度和高度, w_{pred} 和 h_{pred} 分别为预测框的宽度和高度. WIoU 与 CIoU 和 EIoU 不相同. WIoU 是 IoU 的一种加权版本, 适用于特定任务中, 如对于不同目标类型、尺寸或位置的目标需要不同的权重进行优化. WIoU 通过在计算 IoU 时引入权重 (如基于目标的位置、尺寸等) 调整对不同目标的惩罚程度. 本文使用基于预测框和目标框的宽高比差异调整权重, 其计算公式为

$$\text{WIoU} = \text{IoU} \times \frac{1}{1 + \left| w_{\text{pred}}/h_{\text{pred}} - w_{\text{target}}/h_{\text{target}} \right|}, \tag{9}$$

其中 w_{pred} 和 h_{pred} 分别为预测框的宽度和高度, w_{target} 和 h_{target} 分别为目标框的宽度和高度. 本文使用 3 种不同回归损失函数验证哪种计算方式能在本文实验环境中提高模型学习图像特征的能力.

2 实验结果与分析

2.1 数据集

实验使用东北大学建立的公开金属表面缺陷数据集 NEU-DET^[25]. 该数据集共有 6 类钢材表面缺陷, 包括裂纹(crazing, Cr)、夹杂(inclusion, In)、斑块(patches, Pa)、点蚀面(pitte-surface, Ps)、轧制氧化皮(rolled-inscale, Rs)和划痕(scratches, Sc)共 1 800 张图像, 图像分辨率为 640×640 , 以 7 : 1.5 : 1.5 随机抽取划分为训练集、验证集和测试集. 图 7 为数据集缺陷类型示例.

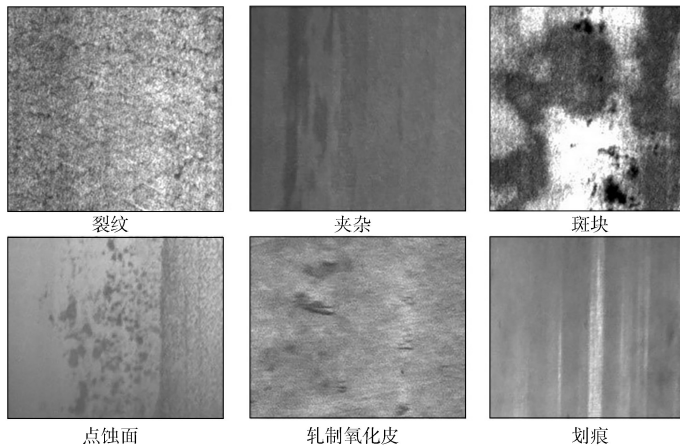


图 7 缺陷类型和缺陷统计

Fig. 7 Defect types and defect statistics

2.2 实验参数设置

实验环境配置如下: 18v CPU AMDEPYC9754 128-Core 处理器, Nvidia RTX 4090D(24 GiB) 显卡, 60 GiB 内存, 操作系统为 Ubuntu20.04, 深度学习框架为 PyTorch 2.0.0, Cuda 11.8, Python 3.8, Cudnn 8.7. 优化器选择 SGD, 动量为 0.937, 权重衰减因子为 0.005, 初始学习率为 0.01, 采用冻结训练 50 个 epoch. 冻结训练时的 Batchsize 为 16, 解冻后的训练为 350 个 epoch, Batchsize 为 8, 共 400 个 epoch. 经过 400 个 epoch 训练后, 验证集上的损失集基本收敛, 如图 8(A)所示; 验证集的平均检测精度(mean average precision, mAP)也基本收敛, 如图 8(B)所示.

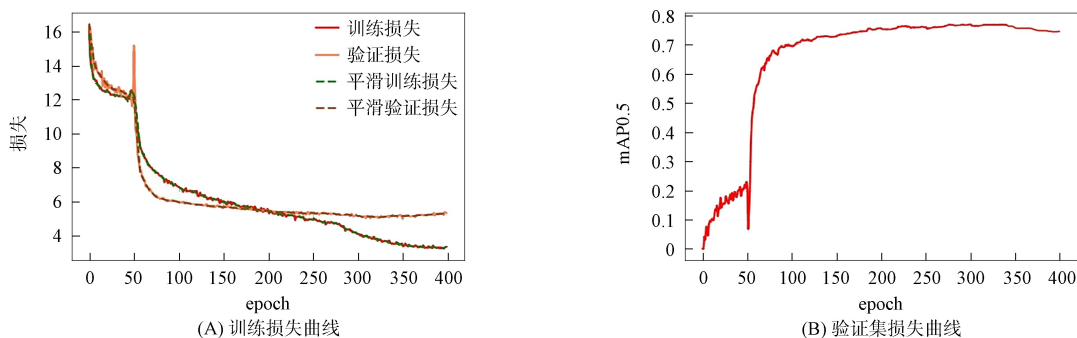


图 8 训练过程收敛曲线

Fig. 8 Convergence curves of training process

2.3 评价指标

本文实验采用的评价指标为 mAP、精确度 AP(average precision)、浮点运算次数 FLOPs(floating point operations)和参数量 Parameters. 精确度 AP 的计算过程依赖 Precision-Recall 曲线. Precision 的计算公式为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (10)$$

Recall 的计算公式为

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad (11)$$

其中: TP 表示记录预测框与真实框的 IoU 值超过一定阈值的预测框数量, 本文使用的阈值为 0.5; IoU 未超过 0.5 的预测框则标记为 FP; FN 表示未被检测到的真实框数量.

AP 的计算方式为计算 Precision-Recall 曲线下的面积:

$$\text{AP} = \int_0^1 P(r) dr, \quad (12)$$

其中 $P(r)$ 表示在给定 Recall 下的 Precision. mAP 的计算公式为

$$\text{mAP} = \frac{1}{N} \sum_{i=1}^N \text{AP}_i, \quad (13)$$

其中 N 表示缺陷类别总数, AP_i 表示第 i 个类别的 AP 值. FLOPs 表示浮点运算次数, 用来衡量网络运行算法的时间复杂度. 本文主要计算了卷积层的 FLOPs, 其计算公式为

$$\text{FLOPs}_{\text{conv}} = 2 \times H_{\text{out}} \times W_{\text{out}} \times C_{\text{in}} \times K_h \times K_w \times C_{\text{out}}, \quad (14)$$

其中 H_{out} 表示输出特征图的高度, W_{out} 表示输出特征图的宽度, C_{in} 表示输入通道数, C_{out} 表示输出通道数, K_h 表示卷积核的高度, K_w 表示卷积核的宽度, 2 表示乘法和加法的计算次数. Parameters 表示模型含有的参数量, 用来衡量算法空间复杂度. 针对卷积层, 其计算公式为

$$\text{Parans}_{\text{conv}} = (K_h \times K_w \times C_{\text{in}} + 1) \times C_{\text{out}}, \quad (15)$$

其中 K_h 表示卷积核的高度, K_w 表示卷积核的宽度, C_{in} 表示输入通道数, C_{out} 表示输出通道数, +1 表示偏置项(如果存在).

2.4 消融实验

为验证本改进方法的有效性, 使用数据集 NEU-DET 进行多组消融实验, 共包括 3 个阶段, 分别

为注意力机制验证、小波卷积深度验证和损失函数验证.

第一阶段是注意力机制验证, 在数据集 NEU-DET 上对注意力机制的消融实验结果列于表 1, 其中 YOLOX-CBMA 模型为在相同位置加入 CBMA 注意力机制, YOLOX-SA(n)模型表示在相同位置加入不带最大池化操作的 ShuffleAttention 注意力机制, YOLOX-SA 模型表示在相同位置加入带最大池化操作的 ShuffleAttention 注意力机制.

表 1 不同注意力模块效果对比

Table 1 Comparison of effects of different attention modules

%

模型	mAP	Cr	In	Pa	Rs	Ps	Sc
YOLOX	72.44	41.66	78.79	89.56	64.01	66.69	93.92
YOLOX-CBMA	72.96	39.63	77.32	88.27	64.22	75.27	93.04
YOLOX-SA(n)	73.54	41.22	79.84	89.21	69.55	66.86	94.59
YOLOX-SA	73.80	32.82	77.26	87.49	74.82	77.85	92.57

由表 1 可见, YOLOX 原型未加入注意力机制前, 缺少一定的注意力聚焦能力, mAP 最低. 在加入 CBMA 后, mAP 有提升. 注意力替换为 SA(n)后提升更明显, 表明此处的通道混洗对通道间的信息流动有一定作用, 提升了模型的特征表达能力. 注意力替换为 SA 后, mAP 提升最大, 表明在原来的 ShuffleAttention 注意力机制中加入最大池化操作对模型的性能有提升作用. 且在 Ps 和 Rs 类型的缺陷中相对前面的 3 种模型有较明显的提升. 在之前的分析中, Ps 和 Rs 类型缺陷的特点为缺陷与背景难以区分, 缺陷的尺度变化大等. 说明此 SA 注意力机制增强了模型多尺度缺陷处理能力并修正了缺陷关注区域, 降低了背景对模型推理的干扰. 虽然某些缺陷类型的 AP 值下降了, 如 Cr 类型缺陷的 AP 值下降较明显. 原因可能是: 在 SA(n)注意力机制中添加了最大池化操作后使模型获得了额外的最大显著特征, 但 Cr 类型缺陷的纹理与边缘不明显, 该操作使模型不仅获得了缺陷位置的特征, 也可能获得了更多背景位置的特征, 模型难关注到正确的位置, 反而使 Cr 的检测精度下降了, 但总的 mAP 提升. 最终选择 YOLOX-SA 模型进行后续阶段的消融实验, 此时最优模型的 mAP 值为 73.8%.

第二阶段是小波卷积深度验证, 在数据集 NEU-DET 上对小波卷积深度的消融实验结果列于表 2, 其中 Level1, Level2 和 Level3 分别表示加入的小波卷积深度为 1, 2, 3.

表 2 小波卷积不同深度效果对比

Table 2 Comparison of effects of wavelet convolution at different depths

%

模型	mAP	Cr	In	Pa	Rs	Ps	Sc
YOLOX-SA	73.80	32.82	77.26	87.49	77.85	74.82	92.57
Level1	74.29	38.48	82.16	89.58	71.56	70.07	93.88
Level2	75.02	39.39	79.63	87.24	77.63	70.95	95.28
Level3	73.91	37.59	77.73	89.26	75.72	68.05	95.12

由表 2 可见, 未加入小波卷积的模型 YOLOX-SA, 此时的 mAP 最低, 且模型对 Cr 类型的缺陷检测效果明显较低. 加入小波卷积后, 3 个层次各有不同程度提升, 其中小波卷积深度为 2 时, 模型的 mAP 提升最大. 深度为 1 时, 小波卷积对某些特征的频域分析以及特征提取能力不足, 未达到理想的最优解. 深度为 3 时, 因为复杂度略增加, 模型可能在测试集上出现过拟合现象, 导致提升效果不太理想. 而小波卷积深度为 2 时, 模型兼顾了特征提取能力与泛化性能, 性能达到最优. 在该消融实验中, 不同小波卷积深度都较大提升了 Cr 缺陷类型的 AP 值. 在之前的分析中, Cr 类型的缺陷特点为纹理边缘不明显. 从而说明该小波卷积提升了模型细节纹理的特征提取能力, 使纹理不清晰的 Cr 类型缺陷的检测精度有较大提升, 同时保持其他类型缺陷的检测精度无明显浮动. 最后选择 Level2 模型进行后续的消融实验, 此时最优模型的 mAP 值为 75.02%.

第三阶段为损失函数的验证, 在数据集 NEU-DET 上对损失函数的消融实验结果列于表 3, 其中 WIoU, CIoU, SWE-YOLOX 分别表示将原来的 IoU 损失修改为 WIoU, CIoU, EIoU 3 种不同的损失函数. 未修改的 IoU 损失函数的 Level2 模型, IoU 只考虑了最基本的预测框与真实框的交并比, 其

mAP 最低. 做不同的修改后各有不同程度的 mAP 值提升, 其中提升最大的是 EIoU 类型损失函数的模型. EIoU 损失函数考虑到了预测框与真实框的位置关系和尺寸, 模型学习到了更好的特征.

表 3 不同损失函数效果对比

Table 3 Comparison of effects of different loss functions							%
模型	mAP	Cr	In	Pa	Rs	Ps	Sc
Level2	75.02	39.39	79.63	87.24	77.63	70.95	95.28
WIoU	75.27	41.15	80.38	89.40	78.23	68.23	94.21
CIoU	75.99	42.14	81.24	89.64	81.43	67.17	94.32
SWE-YOLOX	76.30	42.23	81.02	88.70	83.68	68.89	93.26

虽然 EIoU 的考虑因素与 CIoU 损失大致相同, 但不同的是 CIoU 对于预测框有额外形状的惩罚, 过多的惩罚因素反而影响了模型性能. 同理, 对于 WIoU, 添加基于预测框尺寸和位置的权重对模型的学习确实有帮助, 但其整体考虑的因素仍不如 EIoU, 虽然 EIoU 损失函数在 Rs 和 Sc 两种缺陷上的检测精度略有下降, 但在其他类型的缺陷上都有可观提升. 实验结果表明, EIoU 类型的损失函数更适合本文改进方案. 最后最优模型的 mAP 值达 76.3%, 即为本文提出的 SWE-YOLOX 模型.

整体在数据集 NEU-DET 上的消融实验结果列于表 4, 其中实验 1 表示原 YOLOX 模型, 实验 2 表示加入 SA 注意力模块的模型, 实验 3 表示加入小波卷积的模型. SWE-YOLOX 模型为最后加上修改损失函数的模型.

表 4 整体消融实验结果

Table 4 Overall ablation experiment results								%
模型	mAP	Cr	In	Pa	Ps	Rs	Sc	GFLOPs
实验 1	72.44	41.66	78.79	89.56	66.69	64.01	93.92	26.7
实验 2	73.80	32.82	77.26	87.49	77.85	74.82	92.57	26.7
实验 3	75.02	39.39	79.63	87.24	77.63	70.95	95.28	27.0
SWE-YOLOX	76.30	42.23	81.02	88.70	83.68	68.89	93.26	27.0

由表 4 可见, SWE-YOLOX 模型相比 YOLOX 原型在 mAP 上提升了 3.86 百分点, 其中 AP 值提升较明显的缺陷类型为 Cr, In, Ps, Rs 四种缺陷. 可见, SWE-YOLOX 模型不仅提升了整体的检测性能, 且模型的复杂度(GFLOPs)基本未增加. SWE-YOLOX 模型兼顾了检测的速度与性能.

图 9 为 6 种缺陷类型的检测效果对比, 其中第一行图片为原始图片, 第二行图片为 YOLOX 原型检测的效果, 第三行为 SWE-YOLOX 检测效果, 最后一行为真实的标签图片. 由图 9 可见, 相比于 YOLOX 原型, 检测中对于存在小目标, 如 In 缺陷, 对于 YOLOX 模型漏检的目标, SWE-YOLOX 模型通过引入了通道混洗的注意力机制, 关注正确的缺陷位置从而做到了更低漏检. 对于背景干扰大, 如 Rs 缺陷, YOLOX 模型检测结果中存在误检目标, 而 SWE-YOLOX 模型同样依靠注意力机制, 修正了模型关注的位置, 从而减少了误检. 对纹理不明显, 如 Cr 缺陷, YOLOX 模型同样未正确预测到纹理不明显的目标, 而 SWE-YOLOX 模型通过加入小波卷积的特性, 提取了更多缺陷区域的纹理和边缘, 从而达到了更高的正检. 可见, SWE-YOLOX 模型在这 3 种典型缺陷类型中的检测性能明显更好.

2.5 与主流目标检测算法性能对比

为验证 SWE-YOLOX 算法的金属表面缺陷检测性能, 将其与主流的单阶段目标检测算法 YOLOv3, YOLOv4, YOLOv5, YOLOv7, YOLOv7-tiny 和 YOLOv8 等进行横向对比. 图 10 为不同模型整体的 mAP 与 F_1 曲线. 由图 10 可见, 本文改进模型 SWE-YOLOX 的曲线基本覆盖了其他大部分模型的曲线. SWE-YOLOX 模型在数据集 NEU-DET 上检测精度上明显超过其他模型. 图 11 和图 12 分别为不同算法在数据集 NEU-DET 上每个缺陷类型的 AP 曲线与 F_1 曲线. 由图 11 和图 12 可见, 在所有 6 类缺陷中, 蓝色曲线基本都覆盖了其他大部分曲线, 表明本文改进模型在检测性能上优于其他模型.

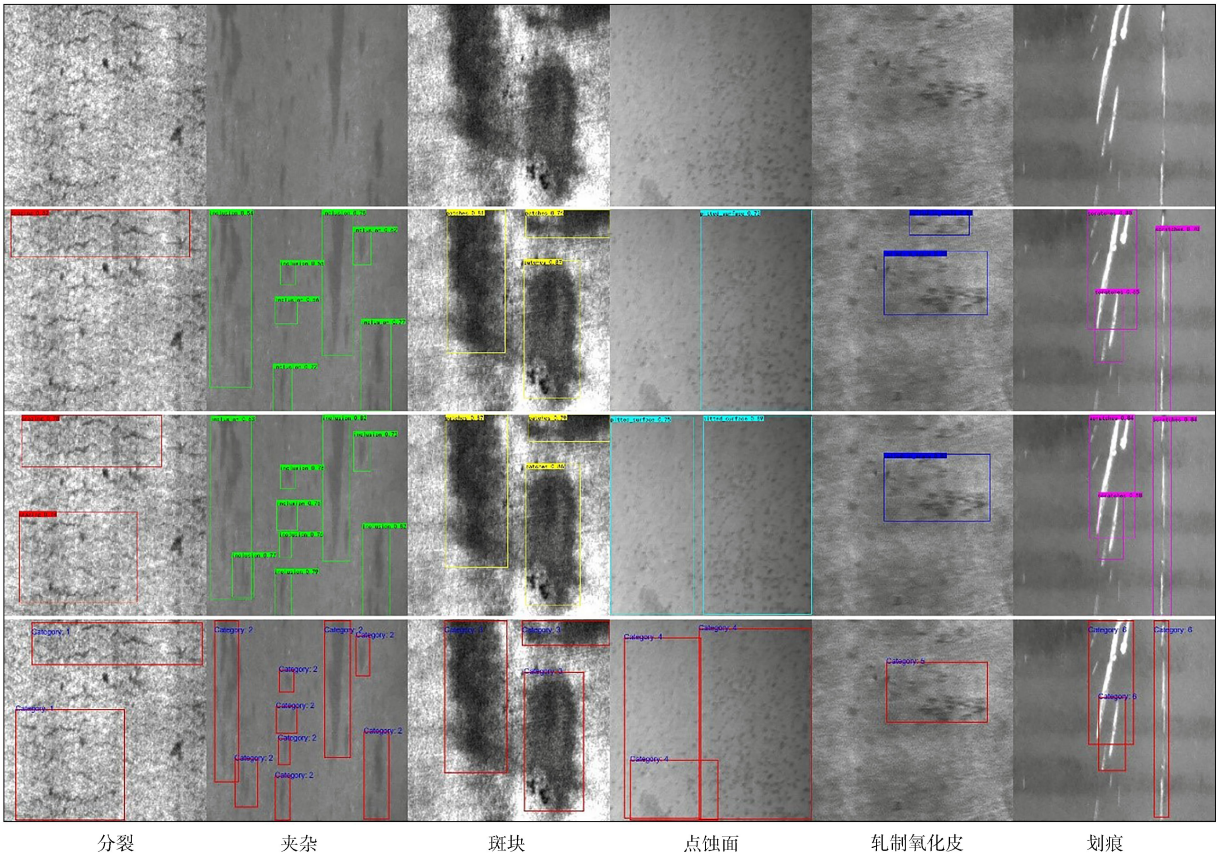


图 9 6 种缺陷检测效果对比

Fig. 9 Comparison of effects of 6 types of defect detection

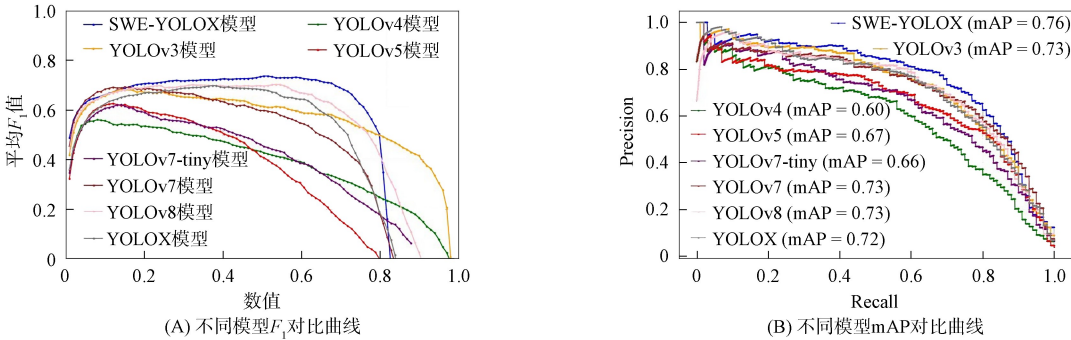


图 10 不同模型的性能对比曲线

Fig. 10 Performance comparison curves of different models

表 5 列出了不同网络模型在数据集 NEU-DET 上的检测性能、模型参数和复杂度. 实验结果表明, SWE-YOLOX 算法的 mAP 值达 76.3%, 比原始的 YOLOX 算法提升了 3.86 百分点. 同时模型的参数量与复杂度基本未增加. 相比于其他目标检测模型, SWE-YOLOX 模型不仅做到了检测性能的领先, 同时在检测速度上也保持了前列, 很好地兼顾了检测性能与速度的平衡.

综上所述, 针对当前工业检测中金属表面检测算法检测性能与速度不平衡的问题, 本文提出了一种 SWE-YOLOX 网络算法. 首先, 在网络中引入混洗通道的注意力模块, 修正网络的关注位置, 减少背景干扰, 增加了不同通道之间的信息交互和特征提取能力, 且在此注意力模块中添加了全局最大池化操作, 结合全局平均池化和全局最大池化的优点, 使模型能在通道注意力中获得更全面的信息; 其次, 在主干网络中添加小波卷积, 小波卷积通过捕捉频域特征, 增强了网络的特征提取能力并减少了信息丢失的情况, 且探究了小波卷积深度对该网络特征提取能力的影响; 最后, 在损失函数处做修

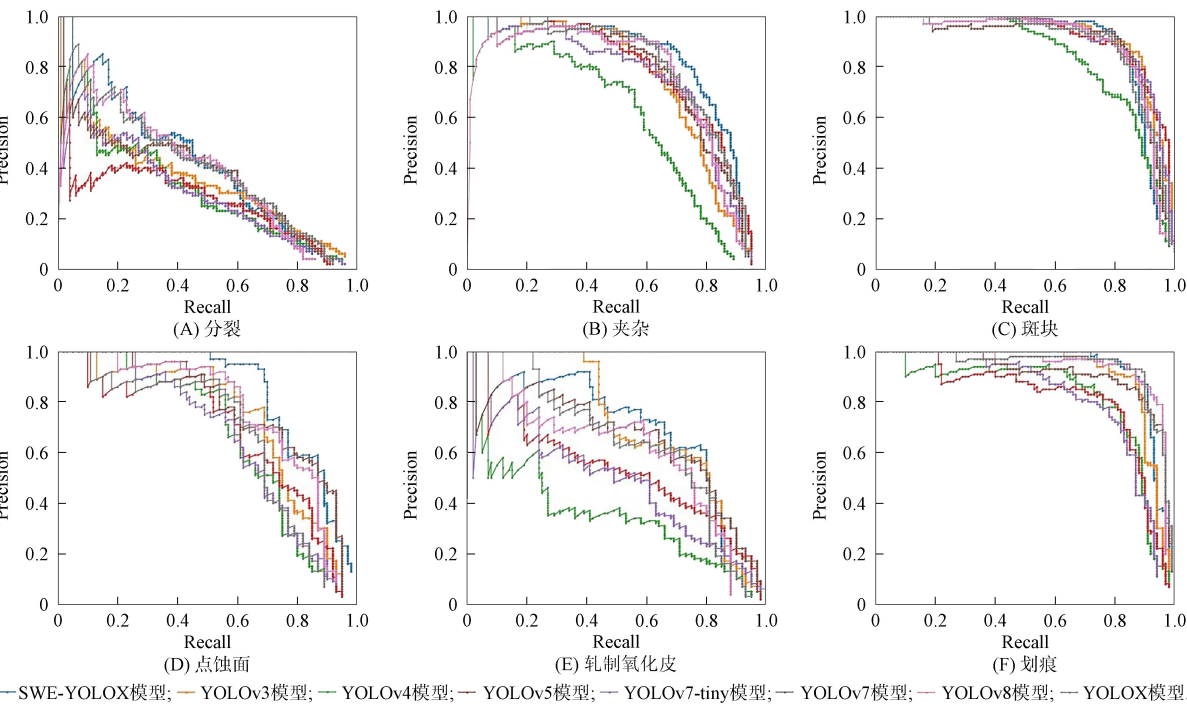


图 11 不同模型的所有缺陷种类 AP 对比曲线

Fig. 11 Comparison curves of AP for all defect types in different models

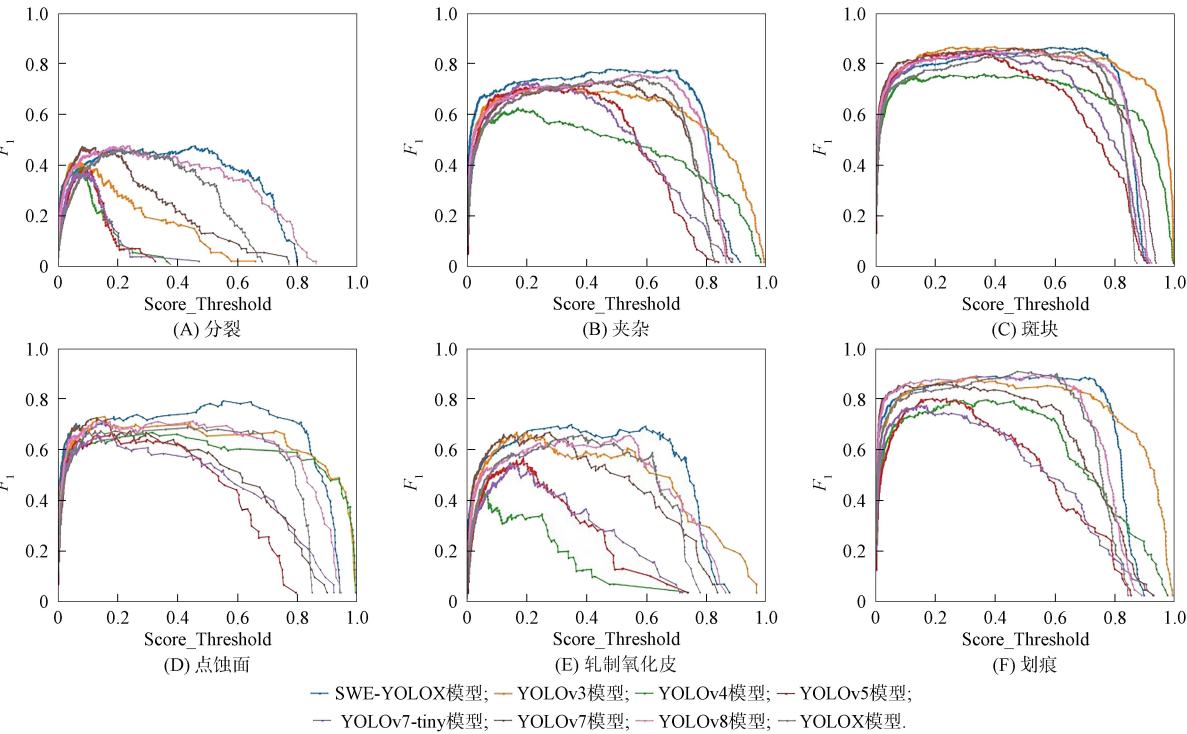


图 12 不同模型的所有缺陷种类 F_1 曲线

Fig. 12 F_1 curves for all defect types in different models

改，将一般的 IoU 损失函数修改为 EIou 损失函数，EIou 引入了额外的惩罚项，进一步提升了对预测框与真实框之间位置关系的优化效果，同时也分析了常用的几种 IoU 损失函数变体对网络学习的影响。实验结果表明，SWE-YOLOX 模型在公开的数据集 NEU-DET 的检测效果相比 YOLOX 模型有显著提升，也优于其他多种主流的 YOLO 检测模型，且改进模型的参数量和复杂度与 YOLOX 模型

相当, SWE-YOLOX 模型的推理速度与 YOLOX 模型的推测速度相比基本没有损失, 实现了对金属表面缺陷的高效检测.

表 5 不同目标检测网络模型性能对比

Table 5 Performance comparison of different object detection network models

模型	mAP/%	Cr/%	In/%	Pa/%	Ps/%	Rs/%	Sc/%	参数量/MiB	GFLOPs
YOLOv3	73.01	35.83	73.67	92.41	73.91	70.86	91.39	61.55	155.3
YOLOv4	60.12	32.59	59.44	83.57	65.77	36.69	82.68	63.90	141.9
YOLOv5	67.02	28.63	77.25	91.99	68.98	53.75	81.49	7.07	16.5
YOLOv7	72.93	36.45	75.83	89.69	77.99	66.50	91.09	37.22	105.1
YOLOv7-tiny	65.85	31.62	74.04	92.10	65.80	49.69	81.89	6.02	13.2
YOLOv8	72.61	40.38	75.74	88.86	76.16	61.15	93.41	11.10	28.6
YOLOX	72.44	41.66	78.79	89.56	66.69	64.01	93.92	8.94	26.7
SWE-YOLOX	76.30	42.23	81.02	88.70	83.68	68.89	93.26	8.98	27.0

参 考 文 献

[1] 李少波, 杨静, 王铮, 等. 缺陷检测技术的发展与应用研究综述 [J]. 自动化学报, 2020, 46(11): 2319-2336. (LI S B, YANG J, WANG Z, et al. Review of the Development and Application of Defect Detection Technology [J]. Acta Automatica Sinica, 2020, 46(11): 2319-2336.)

[2] LIU H W, LAN Y Y, LEE H W, et al. Steel Surface in-Line Inspection Using Machine Vision [C]//First International Workshop on Pattern Recognition. [S.l.]: SPIE, 2016: 187-191.

[3] LIANG Y, XU K, ZHOU P. Mask Gradient Response-Based Threshold Segmentation for Surface Defect Detection of Milled Aluminumingot [J]. Sensors, 2020, 20(16): 4519-1-4519-22.

[4] 刘金海, 付明芮, 唐建华. 基于漏磁内检测的缺陷识别方法 [J]. 仪器仪表学报, 2016, 37(11): 2572-2581. (LIU J H, FU M R, TANG J H. Defect Recognition Method Based on Magnetic Flux Leakage Detection [J]. Journal of Instrumentation and Measurement, 2016, 37(11): 2572-2581.)

[5] TSAI D M, CHEN M C, LI W C, et al. A Fast Regularity Measure for Surface Defectdetection [J]. Machine Vision and Applications, 2012, 23: 869-886.

[6] TIMM F, BARTH E. Non-parametric Texture Defect Detection Using Weibull Features [C]//Image Processing; Machine Vision Applications IV. [S.l.]: SPIE, 2011: 150-161.

[7] 周恺, 张睿哲, 叶宽, 等. 基于同步压缩小波变换的接地扁钢缺陷电磁超声 SH 导波检测方法 [J]. 清华大学学报 (自然科学版), 2022, 62(12): 2013-2020. (ZHOU K, ZHANG R Z, YE K, et al. Grounding Flat Steel Defect Detection Method Based on Synchronous Compression Wavelet Transform [J]. Journal of Tsinghua University (Natural Science Edition), 2022, 62(12): 2013-2020.)

[8] 孙力, 刘晨, 姚红兵. 基于机器视觉的树脂镜片水印疵病检测 [J]. 江苏大学学报 (自然科学版), 2018, 39(4): 425-430. (SUN L, LIU C, YAO H B. Detection of Watermark Defects on Resin Lenses Based on Machine Vision [J]. Journal of Jiangsu University (Natural Science Edition), 2018, 39(4): 425-430.)

[9] 王红梅, 王晓鸽, 王晓燕. 基于深度学习的复杂背景下目标检测 [J]. 控制与决策, 2022, 37(12): 3115-3121. (WANG H M, WANG X G, WANG X Y. Target Detection in Complex Background Based on Deep Learning [J]. Control and Decision, 2022, 37(12): 3115-3121.)

[10] YAN Y H, JIA X J, SONG K C, et al. Specificity Autocorrelation Integration Network for Surface Defect Detection of No-Service Rail [J]. Optics and Lasers in Engineering, 2024, 172: 107862-1-107862-10.

[11] ZHANG N X, ZHONG Y Z, DIAN S Y. Rethinking Unsupervised Texture Defect Detection Using PCA [J]. Optics and Lasers in Engineering, 2023, 163: 107470-1-107470-11.

[12] CAO Y L, ZHU W B, YANG J X, et al. An Effective Industrial Defect Classification Method under the Few-Shot Setting via Two-Stream Training [J]. Optics and Lasers in Engineering, 2023, 161: 107294-1-107294-12.

[13] 张慧, 王坤峰, 王飞跃. 深度学习在目标视觉检测中的应用进展与展望 [J]. 自动化学报, 2017, 43(8):

- 1289-1305. (ZHANG H, WANG K F, WANG F Y. Progress and Prospect of Deep Learning in Target Visual Detection [J]. Acta Automatica Sinica, 2017, 43(8): 1289-1305.)
- [14] 郑明明, 刘胜全, 马前. 基于可变形卷积融合双注意力机制的缺陷检测方法 [J]. 东北师大学报(自然科学版), 2023, 55(2): 52-61. (ZHENG M M, LIU S Q, MA Q. Defect Detection Method Based on Deformable Convolution Fused with Dual Attention Mechanism [J]. Journal of Northeast Normal University (Natural Science Edition), 2023, 55(2): 52-61.)
- [15] 李宗祐, 高春艳, 吕晓玲, 等. 基于深度学习的金属材料表面缺陷检测综述 [J]. 制造技术与机床, 2023(6): 61-67. (LI Z Y, GAO C Y, LV X L, et al. Review of Surface Defect Detection Methods for Metallic Materials Based on Deep Learning [J]. Manufacturing Technology & Machine Tool, 2023(6): 61-67.)
- [16] 郭龙源, 段厚裕, 周武威, 等. 基于Mask R-CNN的磁瓦表面缺陷检测算法 [J]. 计算机集成制造系统, 2022, 28(5): 1393-1400. (GUO L Y, DUAN H Y, ZHOU W W, et al. Surface Defect Detection Algorithm of Magnetic Tile Based on Mask R-CNN [J]. Computer Integrated Manufacturing Systems, 2022, 28(5): 1393-1400.)
- [17] KOU X P, LIU S J, CHENG K Q, et al. Development of a YOLO-V3-Based Model for Detecting Defects on Steel Strip Surface [J]. Measurement, 2021, 182: 109454-1-109454-9.
- [18] SHI W, LU Z S, WU W, et al. Single-Shot Detector with Enriched Semantics for PCB Tiny Defect Detection [J]. The Journal of Engineering, 2020, 13: 366-372.
- [19] 刘培勇, 董洁, 谢罗峰, 等. 基于多支路卷积神经网络的磁瓦表面缺陷检测算法 [J]. 吉林大学学报(工学版), 2023, 53(5): 1449-1457. (LIU P Y, DONG J, XIE L F, et al. Surface Defect Detection Algorithm for Magnetic Tiles Based on Multi-branch Convolutional Neural Network [J]. Journal of Jilin University (Engineering and Technology Edition), 2023, 53(5): 1449-1457.)
- [20] 刘兰兰, 万旭东, 汪志刚, 等. 基于超分辨率重建与多尺度特征融合的输电线路缺陷检测方法 [J]. 电子测量与仪器学报, 2023, 37(1): 130-139. (LIU L L, WAN X D, WANG Z G, et al. Transmission Line Defect Detection Method Based on Super-resolution Reconstruction and Multi-scale Feature Fusion [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(1): 130-139.)
- [21] 耿志强, 陈威, 马波, 等. 基于连续小波卷积神经网络的轴承智能故障诊断方法 [J]. 浙江大学学报(工学版), 2024, 58(10): 2069-2075. (GENG Z Q, CHEN W, MA B, et al. Intelligent Fault Diagnosis Method for Bearings Based on Continuous Wavelet Convolutional Neural Networks [J]. Journal of Zhejiang University (Engineering Science), 2024, 58(10): 2069-2075.)
- [22] FINDER S E, AMOYAL R, TREISTER E, et al. Wavelet Convolutions for Large Receptive Fields [EB/OL]. (2024-07-08)[2025-03-12]. <https://arxiv.org/abs/2407.05848>.
- [23] GOODFELLOW I, BENGIO Y, COURVILLE A. Deep Learning [M]. Cambridge: MIT Press, 2016: 1-800.
- [24] 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. (LIU Y, JIANG S X. Research on Steel Surface Defect Detection Based on Improved YOLOX [J]. Modern Electronics Technology, 2024, 47(9): 131-138.)
- [25] HE Y, SONG K C, MENG Q G, et al. An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features [J]. IEEE Transactions on Instrumentation and Measurement, 2019, 69(4): 1493-1504.

(责任编辑: 韩 啸)