

文章编号:1671-6833(2025)05-0018-08

改进 YOLOv5 的工业产品表面缺陷检测方法

刘兆英¹, 陈志远¹, 张婷¹, 时亚南², 陈迎春³

(1. 北京工业大学 计算机学院, 北京 100124; 2. 新疆维吾尔自治区特种设备检验研究院 新疆特种设备检测技术研究重点实验室, 新疆 乌鲁木齐 830011; 3. 北京工业大学 建筑工程学院, 北京 100124)

摘要: 针对工业场景下资源受限且表面缺陷图像对比度低的问题, 提出了一种改进 YOLOv5 的工业产品表面缺陷检测方法。首先, 在骨干网络中引入感受野增强模块, 用于从不同层次的感受野提取更丰富的视觉特征; 其次, 在特征融合网络中添加混洗注意力模块, 更有效地对不同维度的特征图进行融合; 最后, 采取了任务解耦检测头, 使分类和回归两个任务采用相互独立的网络进行预测, 降低彼此的干扰, 提升检测精度。实验结果表明: 该网络的参数量和计算量均低于 YOLOX、YOLOv7、deformable DETR 等模型, 且在管道数字射线 (DR) 缺陷图像数据集 PDD 和 NEU-DET 数据集上, $mAP@0.5$ 分别提高 2.23 个百分点和 2.99 百分点, 兼顾了工业场景下对缺陷检测实时性和精确性的要求。

关键词: 表面缺陷检测; 计算机视觉; 多尺度特征提取; 注意力机制; 解耦检测头

中图分类号: TP391.4 **文献标志码:** A **doi:** 10.13705/j.issn.1671-6833.2025.02.020

工业产品表面缺陷检测旨在识别并定位工业制品的外观瑕疵, 是保障产品质量、维持生产稳定的重要技术之一。目前, 工业产品表面缺陷检测主要包括人工质检、基于机器学习的方法和基于深度学习的方法^[1]。早期的检测方法主要采取人工质检的方式, 人工筛查的速度和质量过度依赖质检员的水平。企业一方面需要花费额外成本培训质检员; 另一方面, 由于人的主观性和视觉疲劳等问题, 误检和漏检也在所难免。基于机器学习的检测方法主要根据缺陷位置像素块的颜色、形状、纹理等特征, 通过图像处理描述缺陷的特征, 再利用支持向量机、随机森林等传统机器学习方法进行分类^[2]。与人工质检相比, 基于机器学习的检测方法实时性好, 还能提供安全、简便的操作环境, 但不同的缺陷都需要设计对应的缺陷模板, 因此, 检测水平主要依靠缺陷模板的精心设计, 遇到未知缺陷类型时, 没有模板与之匹配就会导致漏检。此外, 对于需要识别多种缺陷的场景, 缺陷模板也容易混淆缺陷类型。

基于深度学习的方法依靠复杂的神经网络自动提取缺陷特征, 比基于机器学习的方法更容易识别并区分缺陷目标, 使得检测精度大幅提高。目前基于深度学习的目标检测模型可分为两阶段和单阶段类型。两阶段检测器在第一阶段中, 生成多个潜在候选区域; 在第二阶段中, 提取这些区域的特征进行分类和回归, 得到最终的目标检测结果。此类模型在两个阶段之间进行显式的区分, 所以需要更多的计算资源和时间。两阶段检测代表模型包括 Faster R-CNN^[3]、Cascade R-CNN^[4] 和 Mask R-CNN^[5] 等。Liu 等^[6] 改进了 Faster R-CNN 用于检测航空发动机叶片表面缺陷, 改进了非极大值抑制算法 NMS, 从而保证了缺陷的连续性; Wei 等^[7] 构建了 Faster VG-RCNN 模型, 引入注意力机制提升模型的视觉感知的敏锐程度, 在检测纺织品上的微小缺陷方面效果显著, 但速度较慢。

单阶段检测器从输入图像中提取特征, 经过一次前向传播过程直接预测目标的类别和位置, 此类

收稿日期: 2025-01-24; 修订日期: 2025-03-28

基金项目: 新疆维吾尔自治区自然科学基金资助项目 (2023D01A22)

作者简介: 刘兆英 (1986—), 女, 山东枣庄人, 北京工业大学副教授, 博士, 主要从事模式识别、图像处理和深度学习方面的研究, E-mail: zhaoying.liu@bjut.edu.cn。

通信作者: 张婷 (1986—), 女, 河南郑州人, 北京工业大学副教授, 博士, 主要从事计算机视觉和图像处理方面的研究, E-mail: zhangting@bjut.edu.cn。

引用本文: 刘兆英, 陈志远, 张婷, 等. 改进 YOLOv5 的工业产品表面缺陷检测方法[J]. 郑州大学学报(工学版), 2025, 46(5): 18-25. (LIU Z Y, CHEN Z Y, ZHANG T, et al. Industrial product surface defect detection of improved YOLOv5[J]. Journal of Zhengzhou University (Engineering Science), 2025, 46(5): 18-25.)

模型的检测速度快,但精度上略低于两阶段方法。单阶段检测代表模型包括 SSD^[8]、RetinaNet^[9]、YOLO 系列^[10-15]等。Zhang 等^[16]为了提高汽车油漆缺陷检测的效率和准确性,提出了一种改进的 MobileNet-SSD 算法,针对汽车油漆的特征设计了特征提取网络,并优化了检测框的匹配策略,实现了实时检测。Pan 等^[17]设计了基于特征增强 YOLO 的表面缺陷检测算法,首先,提出了一种改进的特征金字塔网络,提高了多尺度特征层的空间位置相关性;其次,将深度可分离卷积和密集连接层相结合,降低了模型的计算量;最后,建立了一种新的预测框回归损失,在高 IoU 阈值下提升了检测精度,加速了模型收敛。Liu 等^[18]提出了 LF-YOLO 模型,在 YOLOv3 中引入了高效的特征提取模块和多尺度特征模块,在焊缝缺陷检测方面具有高效、低复杂度的检测性能。

上述方法主要从提升检测的精度和速度的角度进行改进,但没有考虑到在成像条件较差的工业场景中,缺陷图像往往存在弱特征、强干扰、低对比度等缺点,易导致缺陷和背景混淆的问题,这种情况可以通过改进特征提取方式和采用注意力机制改善。张震等^[19]将 CBAM 注意力模块嵌入到 YOLOv5 的特征提取块中,提升了对目标特征的提取能力;魏明军等^[20]将 CBAM 与多头注意力机制相结合,设计了一种混合注意力机制,增强了模型在杂乱背景下的抗干扰能力;Xie 等^[21]针对焊缝 X 射线图像的对比度低、缺陷的形状和大小变化大的问题,提出了 WD-YOLO 模型,该模型引入灰度值曲线增强模型,用于提高模型的对比度,并添加了双重注意力机制,使模型能够更好地区分前景和背景区域;Guo 等^[22]针对钢表面缺陷图像背景干扰大、缺陷类型易混淆的问题,提出了 MSFT-YOLO 模型,该模型将 Transformer 模型的 Trans 模块添加到 YOLO 模型的主干

和检测图中,将特征和全局信息进行融合,还组合了多尺度特征融合结构来融合不同尺度的特征,增强了检测器对不同尺度目标的检测效果。上述方法均使用注意力机制来提升识别和定位能力,但复杂的注意力机制也增加了模型的复杂度,降低了检测速度。

虽然目前工业缺陷检测的性能得到显著提升,但依然存在许多不足。针对缺陷图像的对比度较低、骨干网络提取特征能力不足等问题,本文引入感受野增强模块(RFEM)提升特征提取能力;针对缺陷与背景的差异不明显,难以对缺陷准确定位等问题,引入混洗注意力(SAM)模块,消除噪声背景的干扰;针对分类和回归任务侧重点不同,但却由同一特征图预测等问题,引入任务解耦检测头(TDDH),分别对分类和回归任务设计独立的特征图。

1 改进 YOLOv5 的工业产品表面缺陷检测方法

1.1 网络结构

本文以轻量网络 YOLOv5-n 为基础,提出了一种改进 YOLOv5 的工业产品表面缺陷检测网络结构,如图 1 所示。

该网络由骨干网络、特征融合网络和任务解耦检测头 3 个部分组成。骨干网络采用跨阶段局部网络(cross stage partial networks, CSPNet)^[23],每层通过下采样卷积 DConv 和 C3 模块提取特征,并在快速空间金字塔池化层(spatial pyramid pooling faster, SPPF)后添加感受野增强模块,该模块通过膨胀卷积获取不同尺度的感受野,然后将局部特征和区域特征相融合,提升模型的特征表达能力。特征融合网络的结构采取特征金字塔(feature pyramid, FP)和路径聚合网络(path aggregation net, PANet)^[24]相

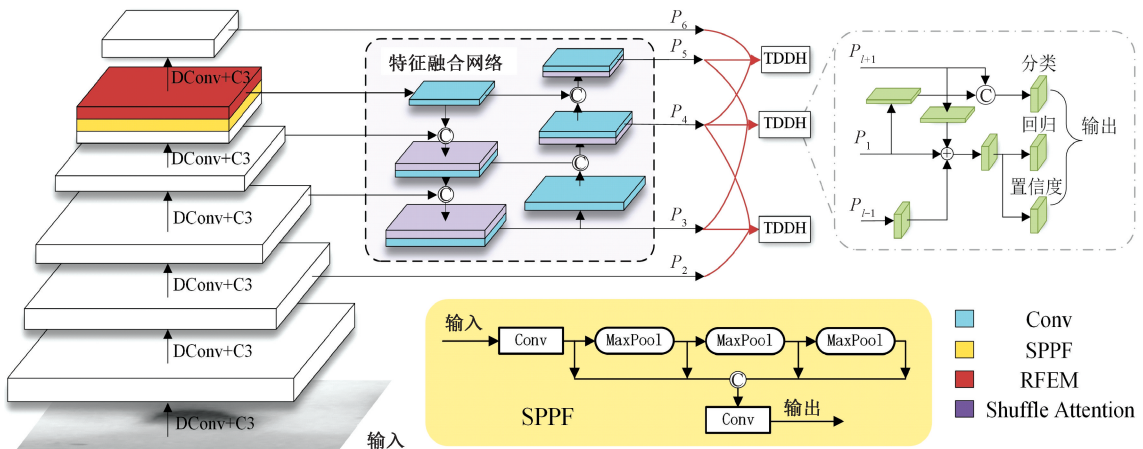


图 1 改进 YOLOv5 的工业产品表面缺陷检测网络结构

Figure 1 Structure of industrial product surface defect detection network for improved YOLOv5

结合的形式,对骨干网络的 P_3 、 P_4 、 P_5 这3层特征图进行特征融合。在上下特征图按通道对齐拼接后,采用混洗注意力模块提取有效信息,该模块通过空间注意力和通道注意力相结合的方式,聚焦特征图中缺陷区域的位置信息,提升模型对缺陷区域的判别能力。最后将融合特征图送入任务解耦检测头中进行预测,检测头同时获取当前尺度及其高低两层的特征图,用以辅助分类和回归任务。其中高层特征图的感受野更广,适合从整幅图像上观察目标的轮廓,有助于判定目标的类别和形状信息;低层特征图拥有更细致的边缘纹理信息,用于辅助精细回归定位。由于任务解耦检测头需要对应3层特征图输入,因此在骨干网络中额外引入 P_2 和 P_6 特征图。任务解耦检测头结合不同尺度特征图的特点,将其分别用于构建分类和回归任务,从而提升了检测精度。

1.2 感受野增强模块

为了增强骨干网络的特征提取能力,引入了感受野增强模块(receptive field enhancement module, RFEM),其结构如图2所示。

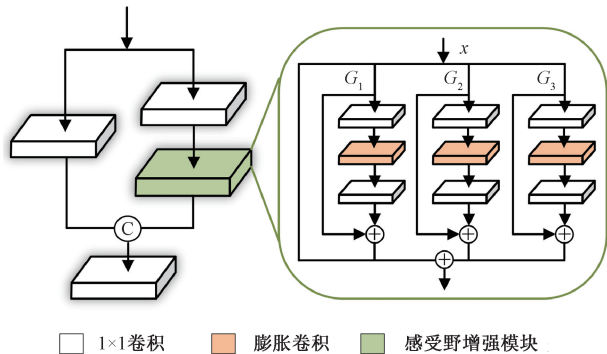


图2 感受野增强模块结构

Figure 2 Structure of receptive field enhancement module

为了从骨干网络中获取不同尺度的特征,将

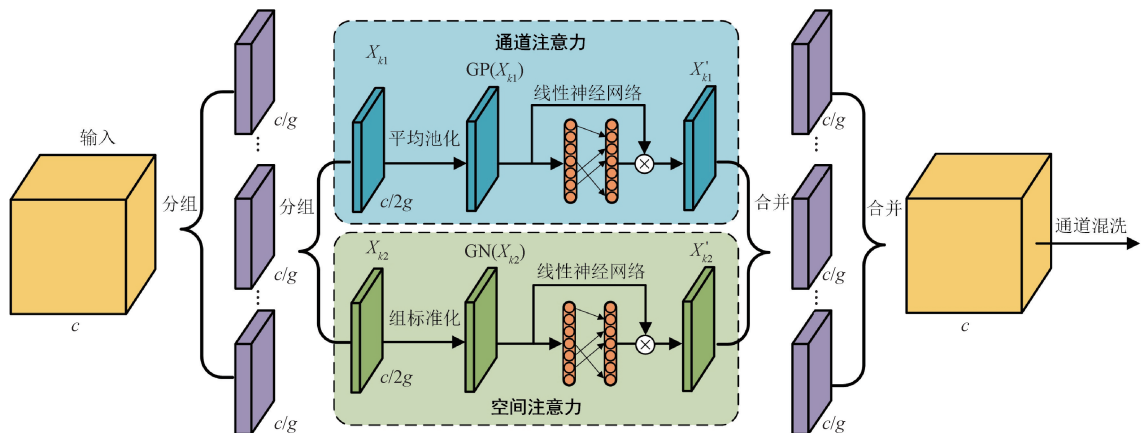


图3 混洗注意力模块结构

Figure 3 Structure of shuffle attention module

YOLOv5网络中C3模块的瓶颈结构替换为3组卷积,每组膨胀卷积首先使用 1×1 卷积收缩通道数,然后分别采用不同的扩张速率的膨胀卷积获取不同尺度的感受野,扩张速率依次设置为1,2和3。最后使用 1×1 卷积扩张至原通道数,类似瓶颈结构的设计降低了膨胀卷积的参数量和计算量。为防止训练过程中梯度爆炸和梯度消失问题,每组膨胀卷积都采取了残差结构,其计算公式为

$$G_i(x) = x + \text{Conv}(\text{DiConv}(\text{Conv}(x), i)). \quad (1)$$

式中: x 表示输入; i 表示膨胀卷积的扩张速率,分别为1,2,3;Conv表示 1×1 卷积;DiConv表示膨胀卷积。每组膨胀卷积参数共享,既降低了参数量,又降低了模块过拟合的风险。最后,每组分支和残差分支对应通道相拼接得到输出:

$$\text{RFEM}(x) = x + G_1(x) + G_2(x) + G_3(x). \quad (2)$$

1.3 混洗注意力模块

特征融合网络在融合上下两层特征图时,采取通道拼接的方式,这种方法虽然能够保留完整的特征信息,但也不可避免地引入了冗余甚至干扰信息。由于注意力机制能够获取偏向信息最丰富的特征分支,而抑制信息较弱的特征分支,因此,在通道拼接之后利用注意力机制获取有效特征信息。由于拼接特征通道数较大,直接采用注意力机制会引入大量的参数,对此,采用了混洗注意力模块(shuffle attention module, SAM),其结构如图3所示。该模块首先将特征的通道数 c 等分为 g 组,得到 g 组子特征,每组子特征再根据通道数等分成两部分,分别用于计算通道注意力和空间注意力。在该模块中,通道注意力由全局平均池化和线性神经网络实现。

$$X'_{k1} = \sigma(W_1 s + b_1) \cdot X_{k1}; \quad (3)$$

$$s = \text{GP}(X_{k1}) = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_{k1}(i, j). \quad (4)$$

式中: X_{k1} 表示子特征对应通道注意力的输入分支; X'_{k1} 表示通道注意力的输出分支; σ 表示 SiLU 激活函数^[25]; W_1 和 b_1 分别表示线性神经网络的权重和偏置; GP 表示平均操作。

空间注意力采取组归一化和线性神经网络实现:

$$X'_{k2} = \sigma(W_2 \cdot \text{GN}(X_{k2}) + b_2) \cdot X_{k2} \quad (5)$$

式中: X_{k2} 表示子特征对应空间注意力的输入分支; X'_{k2} 表示空间注意力的输出分支; σ 表示 SiLU 激活函数; W_2 和 b_2 分别表示线性神经网络的权重和偏置; GN 表示组归一化操作。最后, 将所有子特征拼接起来, 然后使用通道混洗操作打乱不同组特征的顺序, 增强不同组特征之间的信息通信。

1.4 任务解耦检测头

时亚南等^[26]设计了一种轻量的解耦检测头 TIDE, 完全解耦了分类和回归任务, 提升了检测精度。基于此, 本文设计了任务解耦检测头 (task decoupling detection head, TDDH), 其结构如图 4 所示。对分类任务而言, 为了判断目标的类别, 特征图需要提供目标的细节信息。更深层的特征图其感受野更广, 能够从图像全局的角度给出图像的显著区域。同时, 更深层的特征图还拥有更多的通道数, 这就意味着其语义信息更丰富。因此, 对于分类任务, 对 P_l 进行下采样卷积后, 与通道数更多的 P_{l+1} 相连, 二者拼接成 G_l^{cls} 作为分类特征图。 G_l^{cls} 可以表示为

$$G_l^{\text{cls}} = \text{Concat}(\text{DConv}(P_l), P_{l+1}) \quad (6)$$

式中: Concat 表示按照通道对齐的方式拼接特征; DConv 表示下采样卷积。最终得到的 G_l^{cls} 用于预测分类结果。由于对 P_l 进行了下采样卷积, 使得 G_l^{cls} 的通道数扩展为 P_l 的 3 倍, 而分辨率只有 P_l 的一半, 不仅语义信息得到了非常丰富, 还节约了大量参数和计算量。

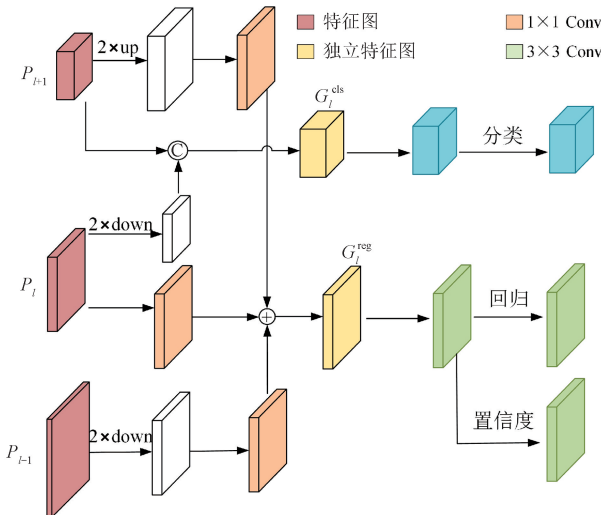


图 4 任务解耦检测头结构

Figure 4 Structure of task decoupling detection head

对回归任务而言, 不但需要从整体上观察到目标轮廓, 还需要充足的边缘细节来辅助检测框回归。因此, 在特征图 P_l 的基础上, 同时引入了 P_l 对应的上下两层特征图 P_{l+1} 和 P_{l-1} 。其中, P_{l+1} 用于提供全局性的边界特征, 有助于模型推断出目标的大致边界定位。 P_{l-1} 则尺寸更大, 拥有更细致的边缘纹理信息, 有助于模型准确定位目标的边界位置。由于回归任务对语义信息不敏感, 过多的通道数不仅容易误导回归边界, 还引入了额外的计算量, 因此, 首先将 3 层特征图分别通过 1×1 卷积统一通道数; 然后将 P_{l+1} 上采样至 P_l 大小, 将 P_{l-1} 下采样至 P_l 大小; 最后 3 层特征图对位相加得到为回归任务设计的特征图 G_l^{reg} :

$$G_l^{\text{reg}} = \mu(\text{Conv}(P_{l+1})) + \text{Conv}(P_l) + \text{DConv}(\text{Conv}(P_{l-1})) \quad (7)$$

式中: Conv 表示 1×1 卷积; μ 表示上采样; DConv 表示下采样卷积。

1.5 损失函数

本文提出的总损失函数包括分类损失、置信度损失和回归损失 3 部分, 其计算公式为

$$\text{Loss} = \alpha_{\text{cls}} L_{\text{cls}} + \alpha_{\text{obj}} L_{\text{obj}} + \alpha_{\text{box}} L_{\text{box}} \quad (8)$$

式中: L_{cls} 、 L_{obj} 、 L_{box} 分别表示分类损失、置信度损失和回归损失; α_{cls} 、 α_{obj} 、 α_{box} 分别表示 3 项损失的权重占比。分类损失和置信度损失均采用二值交叉熵损失 BCELoss 计算, 其计算公式为

$$\text{BCELoss} = -\frac{1}{N} \sum_{n=1}^N [y_n \cdot \log(\gamma(x_n)) + (1 - y_n) \cdot \log(1 - \gamma(x_n))] \quad (9)$$

式中: N 表示样本总数; y_n 表示真实值; x_n 表示预测值; γ 表示 sigmoid 激活函数。

回归损失采用 CIoU 损失, 其计算公式为

$$L_{\text{CIoU}} = 1 - \text{IoU}(A, B) + \frac{\rho^2(A_{\text{ctr}}, B_{\text{ctr}})}{d^2} + \alpha v \quad (10)$$

$$v = \frac{4}{\pi^2} \left(\arctan \frac{w^{\text{gt}}}{h^{\text{gt}}} - \arctan \frac{w}{h} \right)^2 \quad (11)$$

$$\alpha = \frac{v}{(1 - \text{IoU}(A, B)) + v} \quad (12)$$

式中: A 、 B 分别表示预测框和真实框; A_{ctr} 和 B_{ctr} 分别表示预测框和真实框的中心点坐标; $\text{IoU}(A, B)$ 表示 A 框和 B 框的交并比; $\rho(A_{\text{ctr}}, B_{\text{ctr}})$ 表示 A 框和 B 框中心点坐标的欧氏距离; d 表示包围 A 、 B 两框的最小矩形框的对角线距离; α 和 v 是对检测框长宽比的惩罚系数, 使得回归任务在训练过程中更稳定; w^{gt} 和 h^{gt} 分别表示真实框的宽和高; w 和 h 分别表示预测框的宽和高。

2 实验设置

2.1 数据集

为了验证本文方法,采用了管道 DR 缺陷图像数据集 PDD 和钢材表面缺陷检测数据集 NEU-DET^[27] 进行实验。PDD 数据集由本文制作标注,一共包括 6 010 幅图像,共 20 类管道缺陷,每类缺陷样本数量如表 1 所示,管道 DR 缺陷图像样例如图 5 所示,本文选择了 4 866 幅图像作为训练集,542 幅图像作为验证集,602 幅图像作为测试集。

表 1 PDD 数据集中各类缺陷样本数量

Table 1 Number of various defect samples in the PDD dataset

缺陷类型	样本数	缺陷类型	样本数
未熔合	587	根部内凹	417
气孔	355	内部咬边	134
未焊透	426	外部咬边	94
裂纹	372	密集气孔	51
夹钨	401	根部烧穿	385
夹渣	457	横向裂纹	326
咬边	281	热裂纹	153
烧穿	314	焊瘤	417
深孔	417	焊肉不足	134
内凹	204	错边	85

NEU-DET 数据集是东北大学发布的公开缺陷检测数据集^[27]。涉及裂纹、斑块、麻点表面、轧入氧化皮和划痕 6 种热轧带钢的典型表面缺陷,共有 1 800 幅图像,每类缺陷包含 300 张样本。本文选择 1 458 幅图像作为训练集,162 幅图像作为验证集,180 幅图像作为测试集。

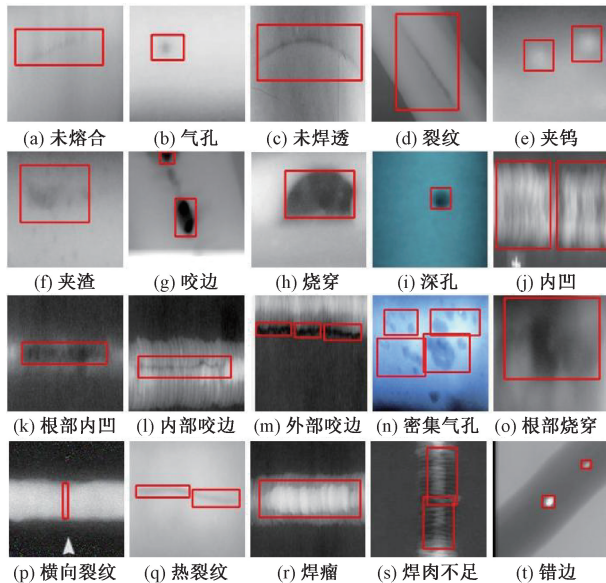


图 5 管道 DR 缺陷图像样例

Figure 5 Sample images of pipeline DR defect

2.2 参数设置

本文实验平台在 windows11 操作系统下进行,处理器型号为 13th Gen Intel (R) Core (TM) i5-13490F, GPU 型号为 NVIDIA GeForce RTX 4060Ti 16 G, 使用可建立深度神经网络和执行张量计算的深度学习框架 PyTorch 1.11, python 版本为 Python3.8, CUDA 版本为 CUDA11.3。超参数设置:训练图片大小为 320×320,训练前对样本进行 Mosaic 数据增强,所有模型均训练 200 个训练轮次,学习率采取线性调整策略,从 0.010 0 逐步下降到 0.000 1,优化器使用 SGD 算法,动量系数设置为 0.937,权重衰减系数为 0.000 5,批次设置为 32,3 项损失函数的权重 α_{cls} 、 α_{obj} 、 α_{box} 均与官方 YOLOv5 保持一致。

3 实验结果与分析

3.1 对比实验

将本文方法与现存的多种目标检测算法进行对比,分别在 PDD 数据集和 NEU-DET 数据集上开展了对比实验,对比实验结果如表 2 所示。从表 2 可以看出,在参数量上,改进方法相比 YOLOv5-n 有所增加,除 EfficientDet-d0 外均低于其他检测方法^[27],但 EfficientDet-d0 的检测精度不高。在计算量上,改进方法比 YOLOv5-n 稍高,但远低于其他方法,满足了工业场景中的硬件和算力需求。在检测精度上,本文方法在 PDD 数据集上的 $mAP@0.5$ 达到 93.31%,比 YOLOv5-n 提高了 2.23 百分点;在 NEU-DET 数据集上的 $mAP@0.5$ 上达到 75.66%,比 YOLOv5-n 提高了 2.99 百分点,均高于其他方法。

为了更直观地展示本文方法的检测效果,图 6 给出了本文方法与其他模型的检测结果。从图 6 可以看出,YOLOv5-n 和 YOLOv7 在图 6(a)、6(e)、6(f) 和 6(g) 中都出现了类似的误检,在图 6(b) 中出现了漏检。在图 6(c) 中,YOLOv5-n 和 YOLOv7 都不能完整地检测到缺陷,并漏检了微小缺陷。在图 6(d) 中,YOLOv5-n 存在多个检测框重叠的问题。

3.2 消融实验

为探究每项改进方法对检测性能的影响,本文在 PDD 数据集和 NEU-DET 数据集上同时评估每个模块对模型性能的影响,实验结果如表 3 所示。从表 3 可以看出,在增加了感受野增强模块后,两个数据集的 $mAP@0.5$ 分别提高了 1.44 百分点和 0.89 百分点。继续增加混洗注意力模块后, $mAP@0.5$ 分别提高了 0.31 百分点和 1.01 百分点,值得注意的是,由于混洗注意力模块采用通道分组以及简单

表 2 本文方法与其他模型在两个数据集上对比实验结果

Table 2 Comparative experiments results of this method and other models on two datasets

模型	PDD 数据集			NEU-DET 数据集		
	$mAP@0.5/\%$	参数量/ 10^6	计算量/GFLOPs	$mAP@0.5/\%$	参数量/ 10^6	计算量/GFLOPs
YOLOv5-n	91.08	1.79	1.0	72.67	1.77	1.0
YOLOX ^[28]	89.53	8.95	26.8	74.35	8.94	26.8
YOLOv7	89.70	36.58	25.9	75.10	36.51	25.8
deformable DETR ^[29]	81.05	40.10	173.0	59.78	40.10	173.0
YOLOv9 ^[30]	92.60	50.74	236.9	73.10	50.71	236.7
EfficientDet ^[31]	83.03	3.84	4.9	62.56	3.83	4.8
YOLOv5-TIDE ^[26]	93.20	4.73	1.7	74.10	4.64	1.7
LF-YOLO ^[18]	90.50	7.29	16.4	67.10	7.26	16.4
YOLOv5-n+RFEM+ SAM+TDDH(本文方法)	93.31	4.33	2.0	75.66	4.14	1.9

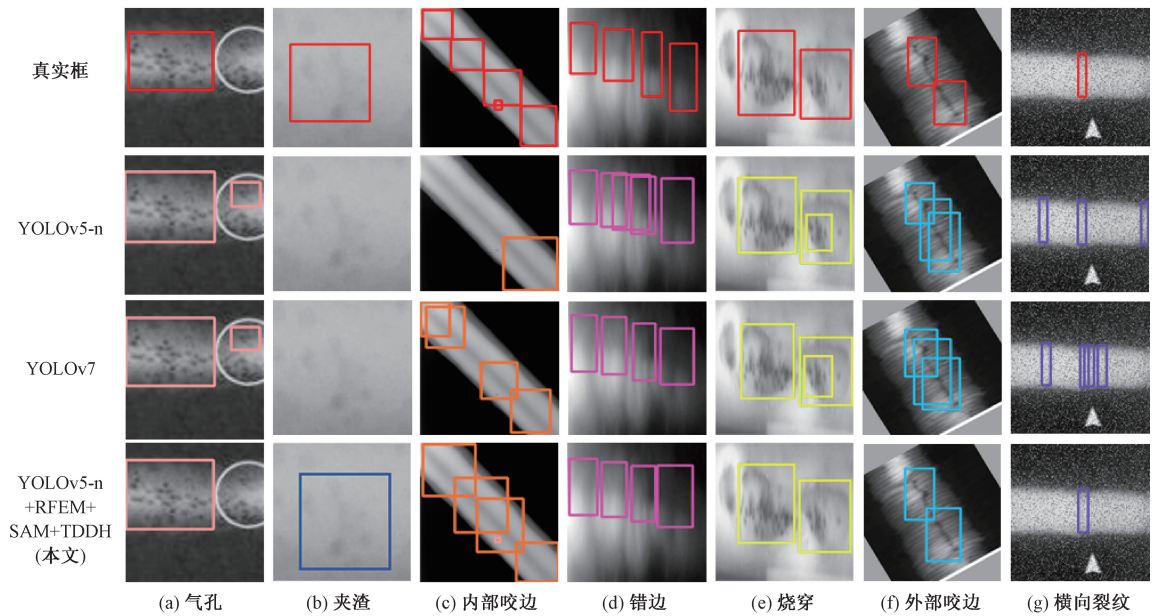


图 6 本文方法与其他模型的检测结果对比

Figure 6 Comparison of detection results between this method and other models

表 3 本文方法与其他模型在两个数据集上消融实验

Table 3 Ablation experiments of this method and other models on two datasets

模型	PDD 数据集			NEU-DET 数据集		
	$mAP@0.5/\%$	参数量/ 10^6	计算量/GFLOPs	$mAP@0.5/\%$	参数量/ 10^6	计算量/GFLOPs
YOLOv5-n	91.08	1.79	1.0	72.67	1.77	1.0
YOLOv5-n+RFEM	92.52	2.00	1.1	73.56	1.98	1.1
YOLOv5-n+RFEM+SAM	92.83	2.00	1.1	74.57	1.98	1.1
YOLOv5-n+RFEM+SAM+TDDH(本文方法)	93.31	4.33	2.0	75.66	4.14	1.9

的线性神经网络,所以几乎不增加参数量和计算量。最后增加任务解耦检测头后, $mAP@0.5$ 分别提高了 0.48 个百分点和 1.09 个百分点。这充分证明了本文中改进模块的有效性。

4 结论

针对工业缺陷图像弱特征、强干扰的特点,以及工业场景中对模型规模和算力的限制,本文提出了

一种快速精确的工业产品表面缺陷检测方法。该模型以 YOLOv5-n 为基础,首先,在骨干网络尾部添加感受野增强模块,提升模型的特征表达能力;其次,在特征融合网络中引入混洗注意力模块,从融合特征图中获取有效信息,抑制无用信息,提升模型对缺陷区域的判别能力;最后,提出了任务解耦检测头,结合不同尺度特征图的特点,将其分别用于构建分类和回归任务,从而提升了检测精度。

(本文受国家市场监督管理总局科技计划项目(2021MK119)的支持)

参考文献:

- [1] CHEN Y J, DING Y Y, ZHAO F, et al. Surface defect detection methods for industrial products: a review[J]. *Applied Sciences*, 2021, 11(16): 7657.
- [2] 罗东亮,蔡雨萱,杨子豪,等. 工业缺陷检测深度学习方法综述[J]. *中国科学:信息科学*, 2022, 52(6): 1002-1039.
- [3] LUO D L, CAI Y X, YANG Z H, et al. Survey on industrial defect detection with deep learning[J]. *Scientia Sinica (Informationis)*, 2022, 52(6): 1002-1039.
- [4] 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*, 2017, 39(6): 1137-1149.
- [5] CAI Z W, VASCONCELOS N. Cascade R-CNN: delving into high quality object detection[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2018: 6154-6162.
- [6] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]//2017 IEEE International Conference on Computer Vision. Piscataway:IEEE, 2017: 2980-2988.
- [7] LIU Y X, WU D B, LIANG J W, et al. Aeroengine blade surface defect detection system based on improved Faster RCNN[J]. *International Journal of Intelligent Systems*, 2023, 2023: 1992415.
- [8] WEI B, HAO K R, GAO L, et al. Detecting textile micro-defects: a novel and efficient method based on visual gain mechanism[J]. *Information Sciences*, 2020, 541: 60-74.
- [9] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector[EB/OL]. (2015-12-08)[2024-11-06]. <https://doi.org/10.48550/arXiv.1512.02325>.
- [10] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for dense object detection[C]//2017 IEEE International Conference on Computer Vision. Piscataway:IEEE, 2017: 2999-3007.
- [11] REDMON J, DIVVALA S, GIRSHICK R, et al. You only look once: unified, real-time object detection[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE, 2016: 779-788.
- [12] REDMON J, FARHADI A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2017: 6517-6525.
- [13] REDMON J, FARHADI A. YOLOv3: an incremental improvement[EB/OL]. (2018-04-08)[2024-11-06]. <https://doi.org/10.48550/arXiv.1804.02767>.
- [14] BOCHKOVSKIY A, WANG C Y, LIAO H Y M. YOLOv4: optimal speed and accuracy of object detection[EB/OL]. (2020-04-23)[2024-11-06]. <https://doi.org/10.48550/arXiv.2004.10934>.
- [15] LI C, LI L, JIANG H, et al. Yolov6: a single-stage object detection framework for industrial applications[EB/OL]. (2022-09-01)[2024-11-06]. <https://doi.org/10.48550/arXiv.2209.02976>.
- [16] WANG C Y, BOCHKOVSKIY A, LIAO H Y M. YOLOv7: trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway:IEEE, 2023: 7464-7475.
- [17] ZHANG J D, XU J B, ZHU L Y, et al. An improved MobileNet-SSD algorithm for automatic defect detection on vehicle body paint[J]. *Multimedia Tools and Applications*, 2020, 79(31): 23367-23385.
- [18] PAN K L, HU H Y, GU P. WD-YOLO: a more accurate YOLO for defect detection in weld X-ray images[J]. *Sensors*, 2023, 23(21): 8677.
- [19] LIU M Y, CHEN Y P, XIE J M, et al. LF-YOLO: a lighter and faster YOLO for weld defect detection of X-ray image[J]. *IEEE Sensors Journal*, 2023, 23(7): 7430-7439.
- [20] 张震,陈可鑫,陈云飞. 优化聚类 and 引入 CBAM 的 YOLOv5 管制刀具检测[J]. *郑州大学学报(工学版)*, 2023, 44(5): 40-45, 61.
- [21] ZHANG Z, CHEN K X, CHEN Y F. YOLOv5 with optimized clustering and CBAM for controlled knife detection[J]. *Journal of Zhengzhou University (Engineering Science)*, 2023, 44(5): 40-45, 61.
- [22] 魏明军,王镜涵,刘亚志,等. 基于特征融合和混合注意力的小目标检测[J]. *郑州大学学报(工学版)*, 2024, 45(3): 72-79.
- [23] WEI M J, WANG M H, LIU Y Z, et al. Small object detection based on feature fusion and mixed attention[J]. *Journal of Zhengzhou University (Engineering Science)*, 2024, 45(3): 72-79.

- [21] XIE Y F, HU W T, XIE S W, et al. Surface defect detection algorithm based on feature-enhanced YOLO[J]. *Cognitive Computation*, 2023, 15(2): 565–579.
- [22] GUO Z X, WANG C S, YANG G, et al. MSFT-YOLO: improved YOLOv5 based on transformer for detecting defects of steel surface[J]. *Sensors*, 2022, 22(9): 3467.
- [23] WANG C Y, MARK LIAO H Y, WU Y H, et al. CSP-Net: a new backbone that can enhance learning capability of CNN[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. Piscataway: IEEE, 2020: 1571–1580.
- [24] LIU S, QI L, QIN H F, et al. Path aggregation network for instance segmentation[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2018: 8759–8768.
- [25] ELFWING S, UCHIBE E, DOYA K. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning[J]. *Neural Networks*, 2018, 107: 3–11.
- [26] 时亚南, 陈志远, 刘兆英, 等. 改进YOLOv5n的管道DR缺陷图像检测方法[J]. *计算机工程与应用*, 2024, 60(12): 366–372.
- SHI Y N, CHEN Z Y, LIU Z Y, et al. Improved YOLOv5n pipeline DR defect image detection method[J]. *Computer Engineering and Applications*, 2024, 60(12): 366–372.
- [27] 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*, 2020, 69(4): 1493–1504.
- [28] GE Z, LIU S T, WANG F, et al. YOLOX: exceeding YOLO series in 2021[EB/OL]. (2021–07–18) [2024–11–06]. <https://arxiv.org/abs/2107.08430>.
- [29] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: deformable transformers for end-to-end object detection[EB/OL]. (2020–10–08) [2024–11–06]. <https://doi.org/10.48550/arXiv.2010.04159>.
- [30] WANG C Y, YE H I, LIAO H Y M. YOLOv9: learning what you want to learn using programmable gradient information[EB/OL]. (2024–02–21) [2024–11–06]. <https://doi.org/10.48550/arXiv.2402.13616>.
- [31] TAN M X, PANG R M, LE Q V. EfficientDet: scalable and efficient object detection[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2020: 10778–10787.

Industrial Product Surface Defect Detection of Improved YOLOv5

LIU Zhaoying¹, CHEN Zhiyuan¹, ZHANG Ting¹, SHI Yanan², CHEN Yingchun³

(1. College of Computer Science, Beijing University of Technology, Beijing 100124, China; 2. Xinjiang Key Laboratory of Special Equipment Testing Technology, Xinjiang Uygur Autonomous Region Inspection Institute of Special Equipment, Urumqi 830011, China; 3. College of Architecture and Civil Engineering, Beijing University of Technology, Beijing 100124, China)

Abstract: Aiming at the problem of limited resources and low contrast of surface defect images in industrial scenarios, an improved YOLOv5 industrial product surface defect detection method was proposed. This method first introduced a receptive field enhancement module in the backbone network to extract richer visual features from different levels of receptive fields. Secondly, a shuffle attention module was added to the feature fusion network to more effectively fuse feature maps of different dimensions. Finally, a task decoupling detection head was adopted, allowing the classification and regression tasks to use independent networks for prediction, reducing mutual interference and improving detection accuracy. The experimental results showed that the parameter and computational complexity of this network were lower than models such as YOLOX, YOLOv7, and deformable DETR. On the pipeline Digital Ray (DR) defect image dataset and NEU-DET dataset, the $mAP@0.5$ were increased by 2.23 percentage points and 2.99 percentage points respectively, balancing the requirements for real-time and accurate defect detection in industrial scenarios.

Keywords: surface defect detection; computer vision; multi-scale feature extraction; attention mechanism; decoupling detection head