

基于改进 YOLOv9 的瓷砖表面缺陷检测^{*}尚梦佳¹ 刘利平^{2,3}(1. 华北理工大学人工智能学院 唐山 063210; 2. 上海电子信息职业技术学院机械与能源工程学院 上海 201411;
3. 华北理工大学矿业工程学院 唐山 063210)

摘要: 针对瓷砖表面缺陷检测中存在的小目标检测精度低、参数量大,以及误检和漏检问题,提出了一种改进的瓷砖表面缺陷检测算法 YOLOv9s-SEFN。首先,本研究设计了 SPNet 多尺度特征融合模块通过增强网络对多尺度特征的捕捉与融合能力,有效提升模型对瓷砖表面小缺陷检测的特征表达;其次,设计 ECG 轻量融合模块减少计算量和参数量以实现轻量化;然后,引入频率自适应扩张卷积(FADC)通过自适应调整膨胀率和频率选择,提升瓷砖小缺陷检测精度;最后,设计新的损失函数 NWD-EIOU 通过结合 EIOU 和 NWD,提高小目标定位的精度。实验结果表明,与原 YOLOv9s 检测算法相比,改进后的 YOLOv9s-SEFN 算法在自建实验数据集上表现更佳,模型 mAP@0.5 提升至 93.2%,提高了 3.5%;召回率提升了 4.96%;参数量减少了 2.3%;浮点运算量降低了 4.0%,能够满足瓷砖表面缺陷检测的需求。

关键词: 瓷砖表面缺陷检测;目标检测;YOLOv9;深度学习;特征融合

中图分类号: TQ174.76;TP391.4;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Detection of surface defects on ceramic tiles based on improved YOLOv9

Shang Mengjia¹ Liu Liping^{2,3}(1. College of Artificial Intelligence, North China University of Science and Technology, Tangshan 063210, China; 2. College of Mechanical and Energy Engineering, Shanghai Technical Institute of Electronics Information, Shanghai 201411, China;
3. College of Mining Engineering, North China University of Science and Technology, Tangshan 063210, China)

Abstract: Aiming at the problems of low accuracy of small target detection, large number of parameters, as well as misdetection and leakage in tile surface defect detection, an improved tile surface defect detection algorithm, YOLOv9s-SEFN, is proposed. Firstly, the SPNet multi-scale feature fusion module is designed in this study to effectively improve the model's detection of small defects on the tile surface by enhancing the network's capability of capturing and fusion of multi-scale feature expression; second, the ECG lightweight fusion module is designed to reduce the computational and parametric quantities to achieve lightweighting; then, the frequency adaptive dilation convolution (FADC) is introduced to improve the accuracy of small defects detection on tiles by adaptively adjusting the dilation rate and frequency selection; and lastly, a new loss function, NWD-EIOU, is designed to improve the accuracy of small target localization by combining EIOU and NWD. The experimental results show that compared with the original YOLOv9s detection algorithm, the improved YOLOv9s-SEFN algorithm performs better on the self-built experimental dataset, with the mAP@0.5 raised to 93.2%, an improvement of 3.5%; the recall rate is raised by 4.96%; the amount of parameters is reduced by 2.3%; and the amount of floating-point arithmetic is reduced by 4.0%, which is able to satisfy the needs of tile surface defect detection.

Keywords: tile surface defect detection; target detection; YOLOv9; deep learning; feature fusion

0 引言

瓷砖作为重要的建筑装饰材料,其生产工艺复杂,各个

工序生产过程中常常会出现滴墨、污渍、光斑、边裂、角裂、缺边、缺角等表面缺陷问题^[1],影响外观与质量,并增加成本与资源浪费。在现代制造业中,智能化生产线的应用已

收稿日期:2024-11-29

^{*} 基金项目:河北省省级科技计划项目(20327218D)、“唐山市卫生陶瓷质量智能监控技术基础创新团队”唐山市科技局项目(21130211D)、2023年上海电子信息职业技术学院高层次与紧缺人才科研启动经费项目(GCC2023006)资助

成为提升生产效率、降低运营成本和提高产品质量的关键。然而,瓷砖表面缺陷的检测仍主要依赖人工操作,存在效率低、易受主观因素影响的问题,难以满足现代工业生产的需求^[2]。因此,研究智能化的瓷砖表面缺陷检测方法显得尤为重要,具有显著的现实意义。

传统检测方法难以满足现代企业的检测需求,因而提出了基于深度学习的检测方法。目前,表面缺陷检测的方法主要分为两类:双阶段检测方法和单阶段检测方法。双阶段方法则包括 R-CNN^[3]、Mask-RCNN^[4]和 Faster-RCNN^[5],这是典型的两阶段方法,首先生成区域建议,再由网络对其进行分类并获取预测信息。例如 Gao 等^[6]提出了一种结合 FCN 和 Faster R-CNN 的深度学习网络模型,并引入自适应边界和感兴趣区域边界层,以实现对城市地铁建设中盾构隧道多种缺陷的快速、准确检测。翁玉尚等^[7]提出了一种改进的 Mask R-CNN 算法,通过引入 K-Means II 聚类算法优化区域建议网络,并移除了掩码分支,该方法在提升检测精度方面取得了一定成效,但检测速度有所欠缺。上述方法虽然通过改进网络算法提升了性能,但仍难以满足实时目标检测的要求。相比之下,单阶段算法无需生成候选框,能够在一个阶段内完成检测,显著缩短了检测时间,因此更适合缺陷检测任务。单阶段方法包括 YOLO 系列^[8-10]、SSD^[11]和 EfficientDet^[12]等,它们直接生成目标的概率和位置坐标,并通过一次检测获得结果。王淑青等^[13]通过使用自制数据集并进行数据增强,采用 YOLOv5 算法,实现了对生产线瓷砖表面缺陷的高效检测,显著提高了检测效率。余松森等^[14]提出了一种改进 YOLOv8 的瓷砖表面缺陷检测方法,通过数据预处理、引入 SPD-Conv、C2f_ECA 模块及微小目标检测头,提升了小目标缺陷检测能力,同时降低了模型复杂度,符合工业应用需求。Cao 等^[15]提出了一种改进的轻量化 YOLOv8-GD 模型,将其应用于光伏缺陷检测中。该方法在 YOLOv8 骨干网络中引入 DW-Conv,使用 GSConv 替代普通卷积,并添加 BiFPN 结构,验证了方法的有效性。李忠科等^[16]提出一种基于轻量级 YOLOv8n 的 PCB 缺陷检测算法,通过新增小目标检测层、设计 C2f-GhostD 与 POne-Detect 模块,并融入 SimAM 注意力机制,优化结构实现模型轻量化与精度提升。虽然上述文献中采用的方法在表面缺陷检测方面取得了一定成效,但普遍存在检测精度较低的问题。

最新提出的 YOLOv9 在表面缺陷检测精度方面得到了提升,但在实际应用当中存在参数量过大的问题。为此, Wang 等^[17]提出一种 YOLOv9s-mobilenetv3-adowncbam 网络模型,采用轻量级 MobileNetv3 网络架构替换主干网络,在颈部采用新颖的 Adown 下采样和 CBAM 注意力机制,以高效实现模型轻量化。Shi 等^[18]提出了一种轻量级 YOLOv9s-Pear 网络模型,将 C2f 和 MobileNetV4 的通用倒瓶颈块与 YOLOv9s 模型的 RepNCSPELAN4 模块集成在一起,形成新的 C2FUIBELAN 模块,SCDown 和

v10detect 模块取代了 YOLOv9s 模型原有的 AConv 和检测头结构,降低了计算成本和参数。在现有的 YOLOv9s 改进方案中,为了实现轻量化通常会限制精度的提升。然而,瓷砖小目标缺陷检测任务,要求在降低参数数量的同时,保持较高的检测精度。为了满足这一需求,必须在参数优化与保持精度之间实现有效的平衡,确保模型在减少计算复杂度的同时,依然能够达到高精度的缺陷检测标准。

针对上述问题,本文提出了一种改进的 YOLOv9s-SEFN 瓷砖表面缺陷检测算法。本文的主要贡献如下: 1)设计多尺度特征融合模块(similarity-based attention module and poly kernel inception network, SPNet),通过增强网络对多尺度特征的捕捉与融合能力,有效提升模型对瓷砖表面小缺陷检测的特征表达;2)设计了轻量融合模块(enhanced convolution ghost, ECG),通过减少模型参数量和计算复杂度,在提升瓷砖表面小缺陷检测效率和精度的同时,实现了模型的轻量化与更快的运行速度;3)引入频率自适应扩张卷积(frequency-adaptive dilated convolution, FADC),通过自适应调整膨胀率和频率选择,提升了瓷砖表面小缺陷检测的精度;4)设计 NWD-EIOU 损失函数,通过结合 EIOU (efficient intersection over union) 和 NWD (normalized wasserstein distance),提升了瓷砖表面小缺陷检测中对目标重叠和尺度不变性的处理能力与精度。

1 YOLOv9 算法

自 2015 年首次推出以来, YOLO 系列算法经历了多次迭代与更新,性能得到了显著提升。YOLOv9 算法引入了可编程梯度信息(programmable gradient information, PGD)机制,该机制由主分支、辅助可逆分支和多级辅助信息构成,有效增强了目标检测能力^[19]。YOLOv9 模型由 3 大核心部分组成:主干网络(Backbone)、颈部网络(Neck)和头部网络(Head)^[20]。主干网络(Backbone)作为特征提取器,负责从图像中提取信息,输入图像的大小为 $640 \times 640 \times 3$,使用的 CBL 模块由卷积层、批量归一化(batch normalization, BN)层和 SiLU 激活函数组成。颈部网络(Neck)位于主干网络和头部网络(Head)之间,进一步优化特征融合。头部网络(Head)则利用这些特征进行最终的目标识别。YOLOv9 中提供了多个版本的选择,分为 t、s、m、c、e, 5 个不同的版本,模型权重大小依次增加。考虑到缺陷检测的需求,因此本文选择模型较小且准确度较高的 YOLOv9s 进行改进。其主体结构如图 1 所示。

2 改进的算法

虽然 YOLOv9s 算法具有较强的目标识别能力,但是在瓷砖小目标缺陷检测方面,仍存在检测精度较低、模型参数量较大以及收敛速度较慢等问题。为此,本文提出了一种改进的 YOLOv9s-SEFN 算法,以提升小目标缺陷的检测性能。在瓷砖表面缺陷检测任务中表现出更高的检测精

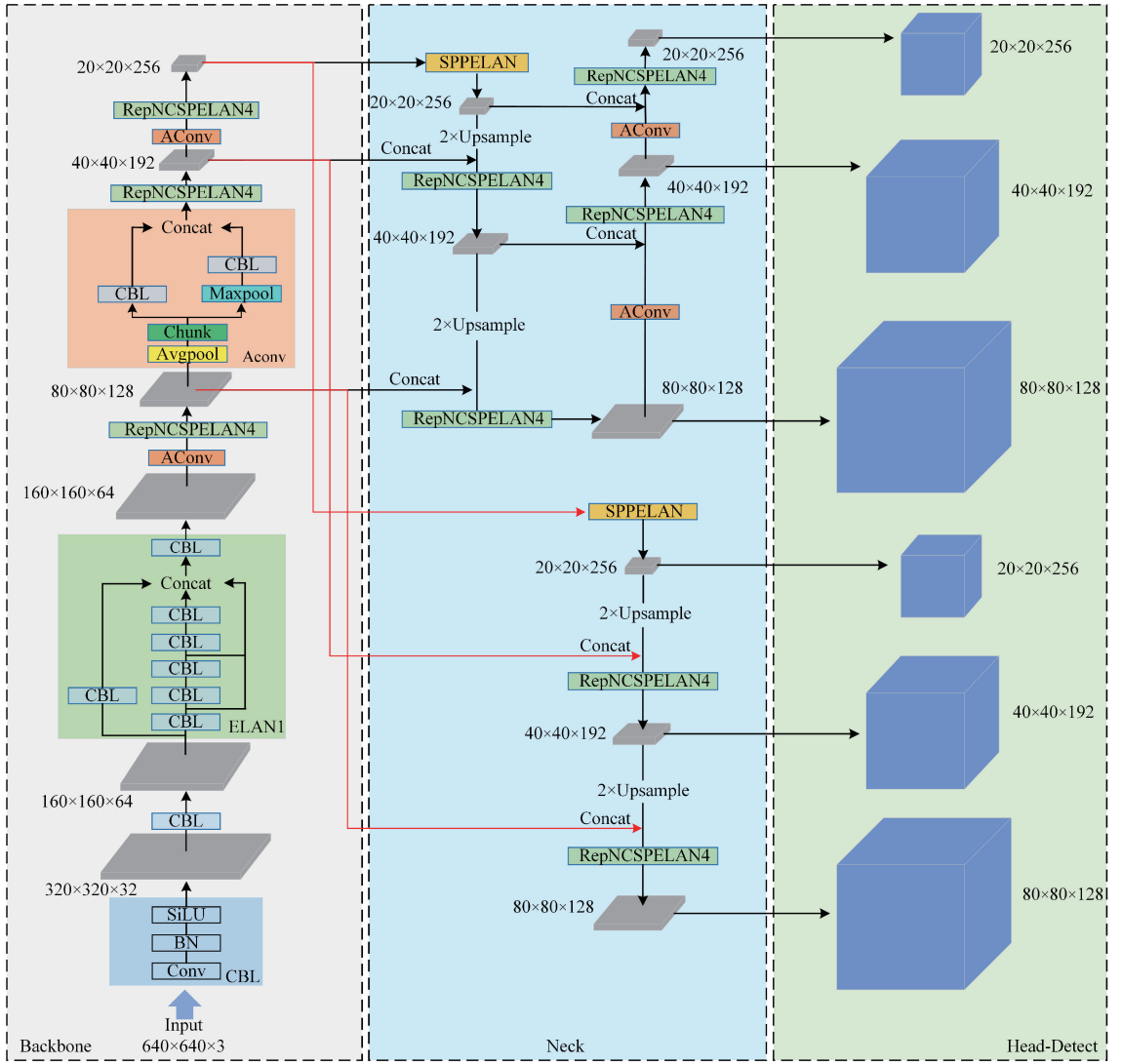


图 1 YOLOv9s 结构框架

Fig. 1 YOLOv9s structural framework

度和效率。本文算法网络结构如图 2 所示。

2.1 SPNet 多尺度特征融合模块

由于瓷砖表面缺陷数据集存在缺陷目标体积较小、形状不规则等问题,导致模型难以精准地识别和定位缺陷目标。为解决这一问题,设计了 SPNet 多尺度特征融合模块,旨在增强网络对多尺度特征的捕捉与融合能力,提升模型对小目标检测的特征表达。其网络结构如图 3 所示。

该模块的实现分为两部分,第 1 部分采用局部自相似性注意力机制^[21],通过三维自适应权重调整,动态优化每个通道的权重分配,突出关键信息,抑制无关特征,使网络更精准地分辨重要区域。具体而言,首先对输入的特征图 $\mathbf{X} \in \mathbf{R}^{B \times C \times W \times H}$ 进行标准化处理, B 表示批次大小 (batchsize), C 表示通道数, H 和 W 分别表示特征图的高度和宽度。为进行标准化,需要在 H 和 W 维度上计算特征图 \mathbf{X} 的均值 $u \in \mathbf{R}^{B \times C \times 1 \times 1}$ 和方差 $\sigma^2 \in \mathbf{R}^{B \times C \times 1 \times 1}$ 。随后,模

块计算每个像素点 $X_{i,j}$ 与其他像素点 $Y_{i,j}$ 的相似度,以进一步突出特征图中的关键信息。然后,计算得到的相似度 \mathbf{Y} 与原始特征图 \mathbf{X} 相乘,生成加权后的特征图 $\mathbf{Z} \in \mathbf{R}^{B \times C \times W \times H}$ 。如式(1)所示。

$$\begin{cases} X_{i,j} = \frac{\lambda - u}{\sqrt{\alpha^2 + \epsilon}} \\ Y_{i,j} = \frac{x_{i,j}^2}{4 \left(\frac{1}{n-1} \sum_{k \neq i,j} x_k^2 + \epsilon \right)} + 0.5 \\ Z_{i,j} = X_{i,j} Y_{i,j} \end{cases} \quad (1)$$

式中: ϵ 是一个极小的常数, $n = HW$ 是特征图中的像素点数, 0.5 是一个偏置项。

第 2 部分由多个 Stage 组成,每个阶段的输入特征图在通道维度上被分割并输入到两条路径中。一条路径通过前向传播的简单前馈网络 (feed forward neural network,

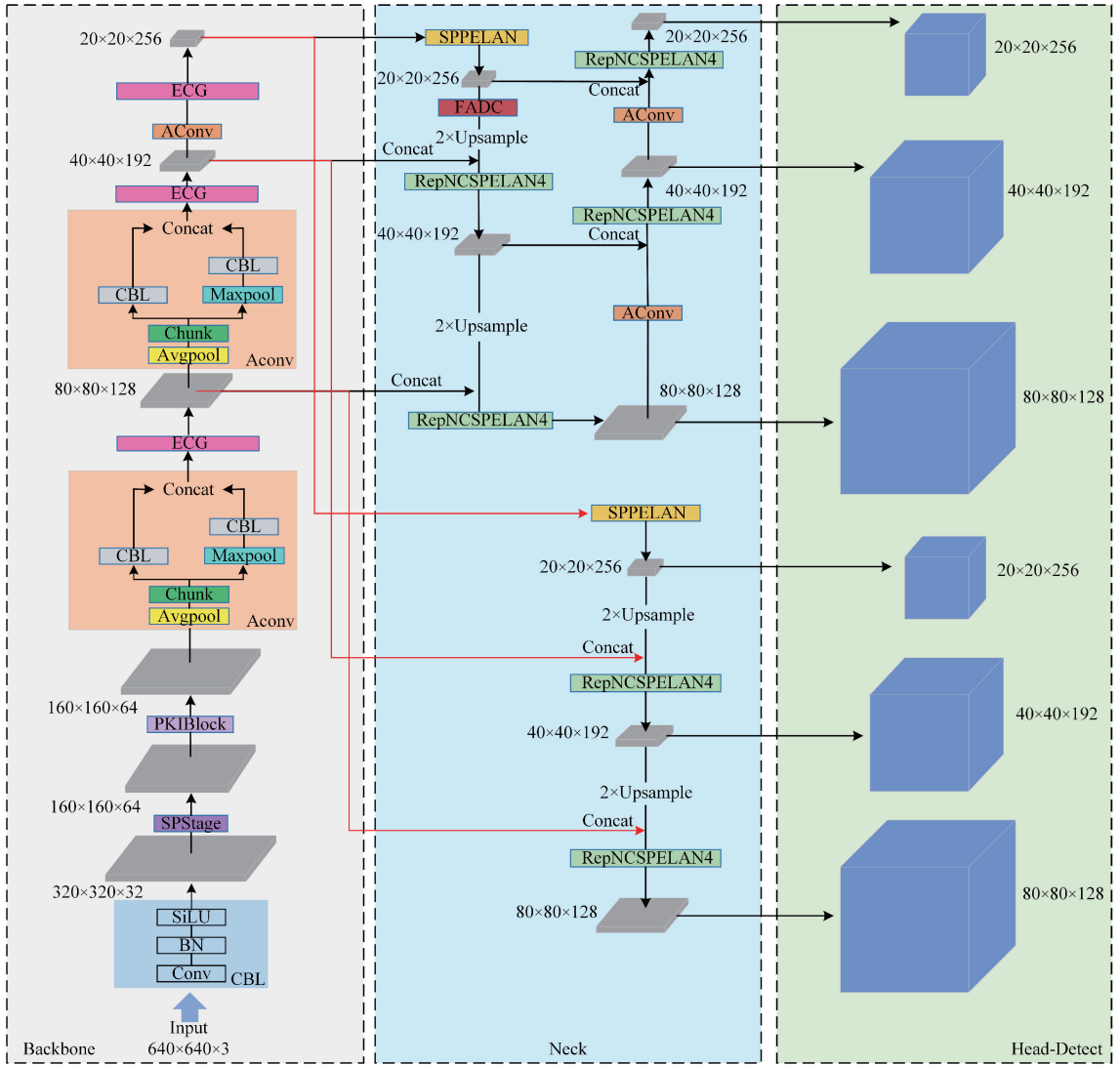


图 2 YOLOv9s-SEFN 网络结构

Fig. 2 YOLOv9s-SEFN network structure

FFN) 处理特征 $Z_{l-1}^{(1)} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$, 输出 $Z_l^{(1)} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$ 。另一条路径包含 N_l 个 PKI Block, 这些块负责处理特征 $Z_{l-1}^{(2)} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$, 输出为 $Z_l^{(2)} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$ 。如式(2)所示。

$$\begin{cases} Z_{l-1} = \text{Conv}_{3 \times 3}(\text{DS}(Z_{l,j})) \in \mathbf{R}^{C_l \times H_l \times W_l} \\ Z_{l-1}^{(1)} = Z_{l-1}[\frac{1}{2}C_l, \dots] \\ Z_{l-1}^{(2)} = Z_{l-1}[\frac{1}{2}C_l, \dots] \end{cases} \quad (2)$$

每个 PKI Block 由 PKI 模块和 CAA 模块组成, PKI 模块首先通过 $k_s \times k_s$ 卷积提取局部特征 $L_{l-1,n} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$, 然后由第 m 个 $k^{(m)} \times k^{(m)}$ 深度可分离卷积 (depth-wise convolution, DWConv) 来跨多个尺度捕获上下文信息

$B_{l-1,n}^{(m)} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$, 设置 $k_s = 3, k^{(m)} = (m+1) \times 2 + 1$ 。最后通过大小为 1×1 的卷积来融合局部和上下文特征, 得到输出特征 $P_{l-1,n} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$ 。如式(3)所示。

$$\begin{cases} L_{l-1,n} = \text{Conv}_{k_s \times k_s}(Z_{l-1,n}^{(2)}), n = 0, \dots, N_l - 1 \\ B_{l-1,n}^{(m)} = \text{DWConv}_{k^{(m)} \times k^{(m)}}(L_{l-1,n}), m = 1, \dots, 4 \\ P_{l-1,n} = \text{Conv}_{1 \times 1}(L_{l-1,n} + \sum_{m=1}^4 B_{l-1,n}^{(m)}) \end{cases} \quad (3)$$

CAA 模块首先对输入特征 $Z_{l-1,n}^{(2)}$ 进行平均池化, 之后通过连接一个 1×1 的卷积来获取局部区域特征 $F_{l-1,n}^{\text{pool}}$, 接着使用两个深度可分离卷积 $\text{DWConv}_{1 \times k_b}$ 和 $\text{DWConv}_{k_b \times 1}$ 用于把握像素之间的上下文信息, 增强中心特征, 最后使用 Sigmoid 函数生成一个注意力权重 $A_{l-1,n} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$, 用于进一步增强 PKI 模块的输出。

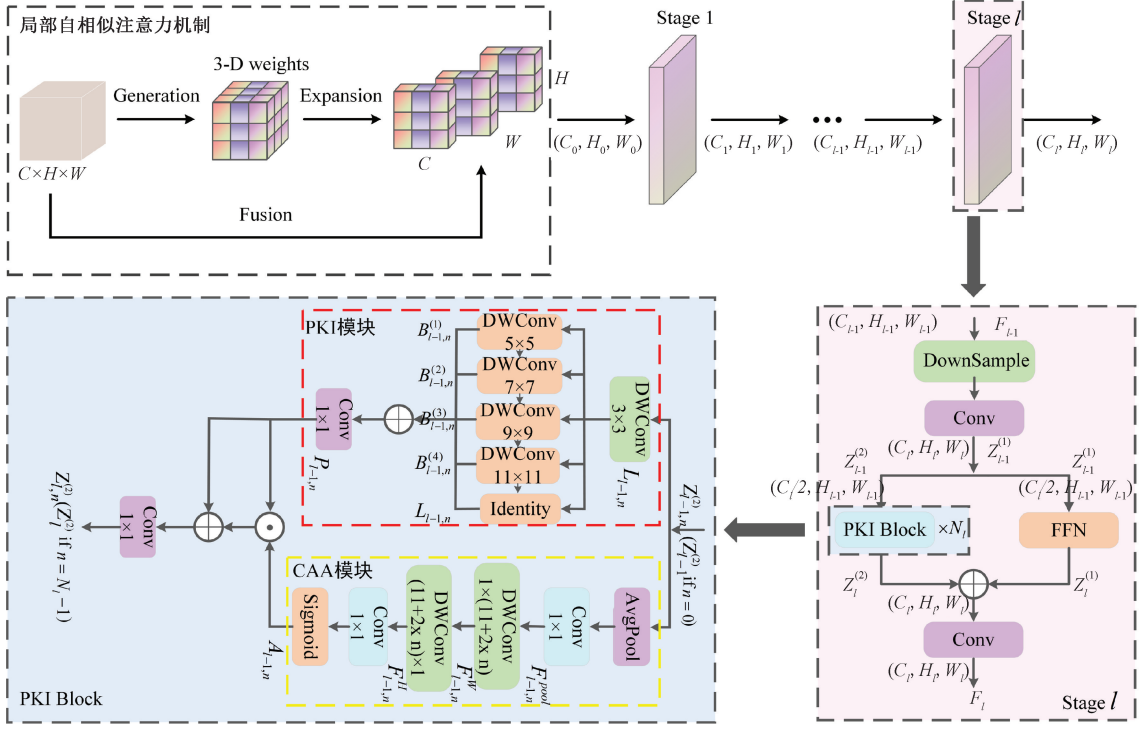


图 3 SPNet 网络结构

Fig. 3 Network structure of SPNet

然后,将 PKI 模块的输出特征 $P_{l-1,n}$ 和 CAA 模块的注意力权重 $A_{l-1,n}$ 进行逐元素相乘再相加,得到增强后的特征 $F_{l-1,n}^{attn} \in \mathbf{R}^{\frac{1}{2}C_l \times H_l \times W_l}$ 。之后,通过一个大小为 1×1 的卷积得到 PKI Block 的最终输出 $Z_{l,n}^{(2)}$ 。如式(4)所示。

$$\begin{cases} F_{l-1,n}^{pool} = \text{Conv}_{1 \times 1}(\rho_{avg}(Z_{l-1,n}^{(2)})), n = 0, \dots, N_l - 1 \\ F_{l-1,n}^w = \text{DWConv}_{1 \times k_b}(F_{l-1,n}^{pool}) \\ F_{l-1,n}^h = \text{DWConv}_{k_b \times 1}(F_{l-1,n}^w) \\ A_{l-1,n} = \text{Sigmoid}(\text{Conv}_{1 \times 1}(F_{l-1,n}^h)) \\ F_{l-1,n}^{attn} = (A_{l-1,n} \odot P_{l-1,n}) \oplus P_{l-1,n} \\ Z_{l,n}^{(2)} = \text{Conv}_{1 \times 1}(F_{l-1,n}^{attn}) \end{cases} \quad (4)$$

式中: ρ_{avg} 表示平均池化操作, $\text{DWConv}_{1 \times k_b}$ 和 $\text{DWConv}_{k_b \times 1}$ 分别表示宽度为 $1 \times k_b$ 和高度为 $k_b \times 1$ 的深度可分离卷积, 其中 $k_b = 11 + 2 \times l$ 。

最后,通过连接操作组合 PKI Block 和 FFN 两条路径的输出,形成该阶段的最终输出,如式(5)所示。

$$F_l = \text{Conv}_{1 \times 1}(\text{Concat}(Z_l^{(1)}, Z_l^{(2)})) \in \mathbf{R}^{C_l \times H_l \times W_l} \quad (5)$$

式中: \odot 表示逐元素乘法, \oplus 表示逐元素加法。

2.2 ECG 轻量融合模块

YOLOv9s 的主干网络使用大量卷积进行特征提取,增加参数和计算量,同时容易引入特征图冗余,从而导致模型运行速度下降。为此,本文设计了 ECG 模块,有效减少了模型参数数量和计算复杂度,优化了网络结构,在保持检测

精度的同时,实现了模型的轻量化与运行速度的提升。ECG 模块的结构如图 4 所示。

ECG 模块首先将输入特征图送入两个并行分支进行处理。第 1 条分支通过标准卷积(Conv)对特征进行提取,紧接着应用批归一化(BN)和 ReLU 激活函数完成特征的初步增强。第 2 条分支通过标准卷积在不同角度提取特征后,将提取的特征输入到 EGG BottleNeck 模块,该模块通过压缩通道维度来减少计算量,并同时保留关键信息。随后,两条分支输出的特征图在通道维度上进行拼接(Concat),提取多尺度特征。最后,通过标准卷积进一步提取关键特征,得到最终的输出特征图。

ECG 模块首先将输入特征图送入两个并行分支进行处理。第 1 条分支通过标准卷积(Conv)对特征进行提取,紧接着应用批归一化(BN)和 ReLU 激活函数完成特征的初步增强。第 2 条分支通过标准卷积在不同角度提取特征后,将提取的特征输入到 EGG BottleNeck 模块,该模块通过压缩通道维度来减少计算量,并同时保留关键信息。随后,两条分支输出的特征图在通道维度上进行拼接(Concat),提取多尺度特征。最后,通过标准卷积进一步提取关键特征,得到最终的输出特征图。

ECG 模块中的 EGG BottleNeck 由 AConv 和 Shortcut 两个分支组成。为提高特征提取效率并减少计算量,在 AConv 和 Shortcut 中设计了增强卷积(Enhanced Conv)。Enhanced Conv 由轻量化卷积和 SE 通道注意力机制组成,其网络结构如图 5 所示。AConv 分支通过 Enhanced Conv

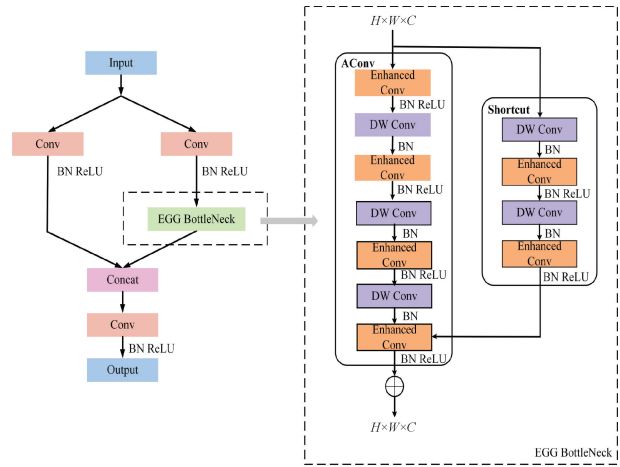


图 4 ECG 网络结构

Fig. 4 Network structure of ECG

和 DWConv 对特征进行细化。Shortcut 分支通过跳跃连接与多层深度可分离卷积来复用特征并优化梯度流动。最

终,将两个分支的输出通过残差连接进行整合,从而生成压缩后的特征输出。

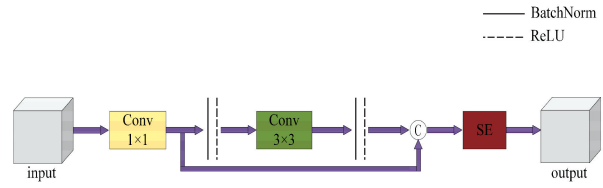


图 5 Enhanced Conv 网络结构

Fig. 5 Network structure of Enhanced Conv

2.3 频率自适应扩张卷积 FADC

瓷砖表面存在多种类型的缺陷,且其差异较大,传统的固定膨胀率卷积方法难以同时兼顾高频细节(如裂纹)与低频全局特征(如光圈),通常会导致高频信息丢失或产生网格伪影。为了解决这一问题,本研究在颈部网络中引入了频率自适应扩张卷积(FADC),其网络结构如图 6 所示。FADC 主要包括 3 个核心策略:频率选择(FreqSelect)、自适应内核(AdaKern)和自适应扩张率(AdaDR)。通过这 3 个策略的协同作用,FADC 增强了网络对不同尺度和频率特征的感知能力,从而有效提升了对小目标的检测精度^[22]。

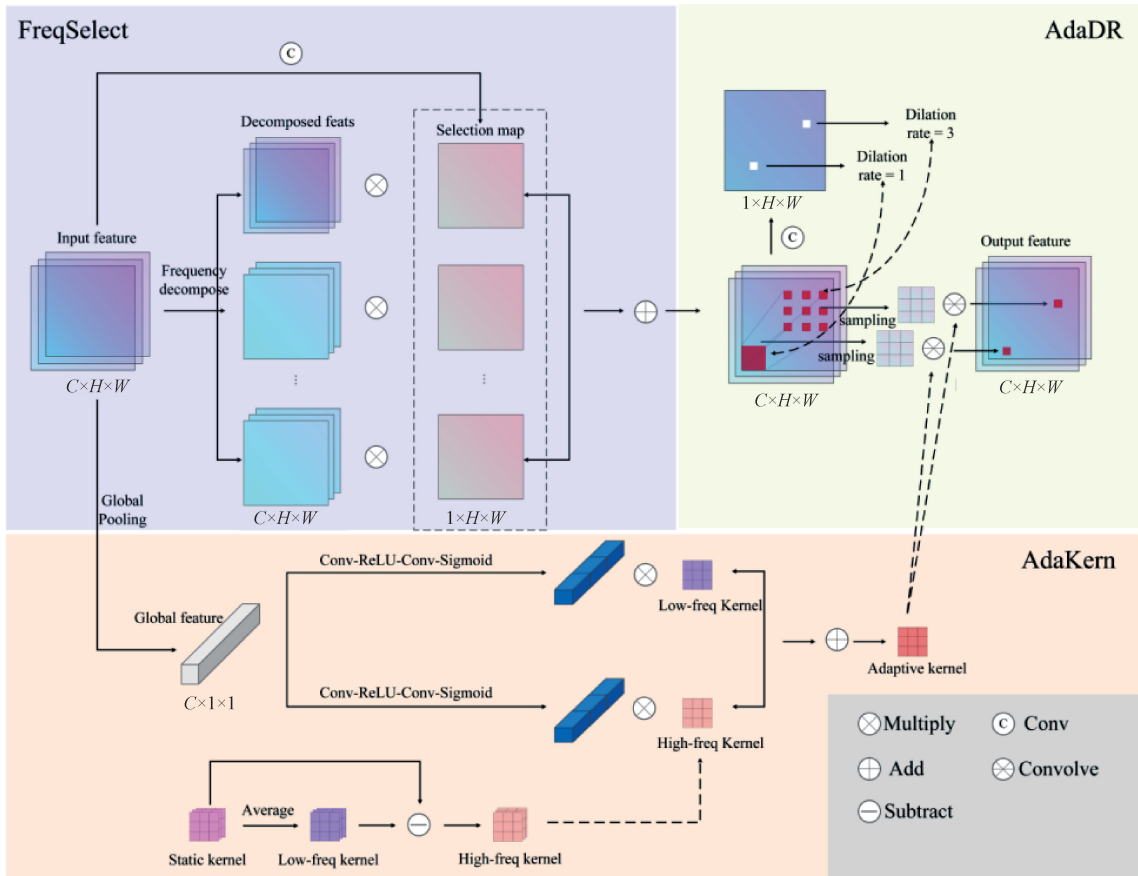


图 6 FADC 网络结构

Fig. 6 Network structure of FADC

1) 频率选择: 首先将输入特征图分解为多个频段, 然后通过选择性掩码对不同频段进行加权处理, 从而动态调整各频率分量的权重 X_b 。最后在空间上动态地对不同频段的频率分量进行加权。计算公式如下:

$$X_b = f^{-1}(M_b X_F) \quad (6)$$

$$\hat{X}(i, j) = \sum_{b=0}^{B-1} A_b(i, j) X_b(i, j) \quad (7)$$

式中: f^{-1} 表示快速傅里叶反变换, M_b 是一个二进制掩码, (i, j) 为 FreqSelect 模块后学习到的频率平衡特征, $A_b \in \mathbf{R}^{H \times W}$ 为第 B 频段的选择映射。

2) 自适应内核: 将卷积核参数分解为低频和高频分量, 然后引入动态加权来调整频率响应, 计算公式如下:

$$W' = \lambda_l W_l + \lambda_h W_h \quad (8)$$

式中: W_l 和 W_h 分别代表低频和高频分量, λ_l, λ_h 为每个通道的动态权值, 通过简单单量级的全局池化+卷积层来预测。

3) 自适应扩张率: 通过动态调整膨胀率优化了特征感受野的大小, 确保可以有效捕捉瓷砖缺陷图像在高频和低频区域的特征, 可以将较低的膨胀率分配给高频区域, 如瓷砖缺陷图像的边界, 将较高的膨胀率分配给低频区域, 如瓷砖缺陷图像的中心和背景。它为每个像素分配一个不同的膨胀率, 计算公式如下:

$$Y(p) = \sum_{i=1}^{K \times K} W_i X(p + \Delta p_i \times \hat{D}(p)) \quad (9)$$

式中: Y_p 为输出特征图中 p 位置的像素值, K 为核大小, W_i 为核的权值参数, Δp_i 为预定义网格采样的第 i 个位置。 $\hat{D}(p)$ 可以通过参数为 θ 的卷积层来预测, θ 的计算公式如下:

$$\theta = \max(\sum_{p \in HP^-} \hat{D}(p) - (\sum_{p \in HP^+} \hat{D}(p))) \quad (10)$$

式中: HP^+ 和 HP^- 分别为最高/最低高频功率的像素。

2.4 损失函数 NWD-EIOU

在目标检测任务中, 损失函数起着至关重要的作用, 主要用于衡量模型预测结果与真实结果之间差异的函数, 并通过引导模型参数的优化来提升检测性能^[23]。选择合适的损失函数不仅能够有效提高检测准确率, 还能够促进模型对特征的高效学习。YOLOv9 模型在边界框回归中采用了损失函数 (complete intersection over union, CIOU)^[24], CIOU 使算法收敛速度更快, 同时考虑了重叠面积、质心距离、纵横比, 也更有助于检测帧与真实帧的匹配。但是, 纵横比是一个相对值, 存在一定的模糊性, 不能很好地平衡正负样本, 特别是对于小物体, 最轻微的位置偏差都会导致 IOU 明显下降。因此, 本文将 EIOU^[25] 和 NWD^[26] 损失函数结合作为小目标检测的损失函数。

EIOU 根据 CIOU 的罚项拆分纵横比的影响因子, 计算边界框的长度和宽度以及地面真值。直接将边界框的高度和宽度与地面真实值之间的差异最小化, 从而实现更

快的收敛和更好的定位结果。损失函数表示为:

$$L_{EIOU} = 1 - IOU + \frac{\rho^2(b, b^{gt})}{(\omega^c)^2 + (h^c)^2} + \frac{\rho^2(\omega, \omega^{gt})}{(\omega^c)^2} + \frac{\rho^2(h, h^{gt})}{(h^c)^2} \quad (11)$$

式中: ω^c 和 h^c 分别为预测帧和实际帧最小外矩形的宽度和高度, ω, h 和 ω^{gt}, h^{gt} 分别为预测帧和实际帧的宽度和高度, 符号 ρ 为 b 和 b^{gt} 、 ω 和 ω^{gt} 、 h 和 h^{gt} 之间的欧氏距离。

NWD 的主要优点在于其对物体尺寸的不敏感性以及对位置偏差引起的代理值变化的平滑性。这一特性使得即便在没有重叠的情况下, 也能够有效计算 Ground truth 与 Bounding Box 之间的相似性。NWD 的表达式函数为:

$$L_{NWD} = 1 - \exp(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}) \quad (12)$$

$$W_2^2(N_a, N_b) = \left\| \begin{pmatrix} cx_a, cy_a, \frac{\omega_a}{2}, \frac{h_a}{2} \\ cx_b, cy_b, \frac{\omega_b}{2}, \frac{h_b}{2} \end{pmatrix}^T \right\|_F \quad (13)$$

其中, N_a, N_b 分别为实际帧和预测帧, C 是数据集中检测类别的数量, $W_2^2(N_a, N_b)$ 是用于衡量边界框 $A = (cx_a, cy_a, \omega_a, h_a)$ 和边界框 $B = (cx_b, cy_b, \omega_b, h_b)$ 高斯分布的二阶 Wasserstein 距离, cx_a, cy_a, ω, h 分别为锚框架的中心点、宽度和高度的坐标, $\| \cdot \|_F$ 为 Frobenius 范数。

本文将 EIOU 和 NWD 损失函数相结合, 提出了 NWD-EIOU 损失函数, 并应用于瓷砖缺陷检测的边界框定位任务。表达式如下:

$$L_{NWD-EIOU} = \alpha \times L_{EIOU} + (1 - \alpha) \times L_{NWD} \quad (14)$$

在瓷砖缺陷检测任务中, NWD-EIOU 损失函数不仅提高了回归精度, 还能有效保持尺度不变性, 尤其在小目标检测方面表现突出。该损失函数的引入使得模型能够更准确地识别和定位瓷砖表面的缺陷, 从而提高了检测结果的精确性, 进一步提升了整体检测性能。

3 实验与结果分析

3.1 实验数据集

本文的研究对象为瓷砖样本, 实验数据均从网络收集, 瓷砖表面缺陷主要以黑点 (dark_spot)、光圈 (aperture)、边异常 (edge anomaly)、角异常 (corners anomaly)、裂纹 (crack) 为主。图 7 为不同类型缺陷的典型样本。

为了增强缺陷检测模型的鲁棒性和泛化能力, 使模型能够更好的检测出缺陷。本实验通过裁剪、平移、旋转等方式对数据集进行增强, 将基础数据集扩充到 3 532 张。使用 labeling 工具对图像数据进行缺陷位置和类别的标注, 并将这些信息保存为 .txt 格式文件。为了能够使模型有效训练, 将数据集以 8:1:1 的比例, 即训练集 2 816 张图片、验证集 358 张图片和测试集 358 张图片。

3.2 实验环境及参数设置

本实验在 GPU 环境下进行模型训练, 操作系统为

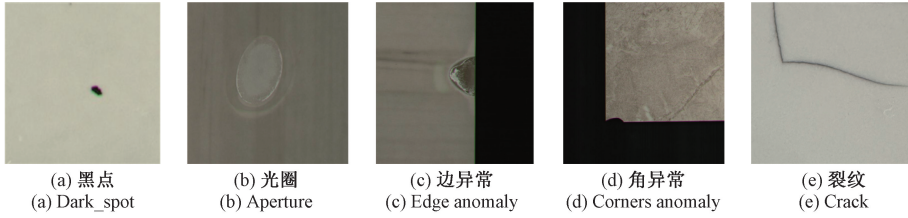


图 7 缺陷示意图

Fig. 7 Schematic diagram of the defect

Linux,显卡为 RTX 3090 E5,CPU 为 Intel (R) Xeon(R) E5-2680,CUDA 版本为 11.3,采用 PyTorch 1.11.0 深度学习框架,使用 Python 3.8 编程语言。训练过程中,输入图像分辨率设置为 640×640 ,批量大小(Batch size)为 8,迭代次数为 200,初始学习率设定为 0.01,动量参数为 0.937,权重衰减系数为 0.000 5。优化器采用随机梯度下降(stochastic gradient descent,SGD)算法,以提升训练的稳定性 and 收敛速度。

3.3 评价指标

为了客观地评估目标检测算法,需要一些评估指标来量化模型的性能。本文实验选取准确率(precision,P)、召回率(recall,R)、平均精度均值(mean average precision,mAP)、参数量和浮点运算作为评价指标来综合评估模型的检测性能。

P 是分类正确的样本在所有样本中的数量比例,计算公式为:

$$P = \frac{T_p}{T_p + F_p} \quad (15)$$

R 指真正为正的样本中被正确地预测为正的样本占比,计算公式为:

$$R = \frac{T_p}{T_p + F_N} \quad (16)$$

式中: T_p 表示正确识别的对象样本的数量, F_p 表示错误识别的对象样本的数量,以及 F_N 表示未识别对象样本的数量。

平均精度(average precision,AP)就是 PR 曲线与坐标轴围成的面积,计算公式为:

$$AP = \int_0^1 P(R) dx \quad (17)$$

mAP 是多个类别 AP 的平均值, $mAP@0.5$ 表示 $IoU=0.5$,计算每一个分类所有图片 AP 值的,然后所有分类求平均 mAP 值,计算公式为:

$$mAP = \frac{\sum_{i=1}^c AP_i}{N} \quad (18)$$

式中: N 表示总类别数目, AP_i 表示第 i 类的 AP 值。

在目标检测系统中,使用平均精度 $mAP@0.5$ 衡量模型检测精度,数值越高表明检测精度越高;GFLOPs 表示每秒千亿次浮点运算,Params 表示模型

的参数量。GFLOPs 数值越低,表明模型或算法的复杂性越低。参数数量 Params 衡量模型大小,数值越低模型越轻量化。

3.4 消融实验

为了分析模型的每项改进点对模型性能提升的作用,设计进行了 8 组消融实验,如表 1 所示。具体来说,模型 YOLOv9s_1 表示 YOLOv9s 基准算法实验。模型 YOLOv9s_2 表示在主干网络中加入设计的 SPNet 多尺度特征融合模块。模型 YOLOv9s_3 表示在主干网络中设计 ECG 轻量融合模块,用其替换原来主干网络中的 RepNCSPELAN4 模块。模型 YOLOv9s_4 表示在颈部网络引入频率自适应扩张卷积 FADC。模型 YOLOv9s_5 表示使用 NWD-EIOU 损失函数。模型 YOLOv9s_6 表示在模型 YOLOv9s_2 的基础上加入 ECG 轻量融合模块。模型 YOLOv9s_7 表示在 YOLOv9s_6 基础上引入频率自适应扩张卷积 FADC。最后的 YOLOv9s_8 模型,即 YOLOv9s-SEFN 模型引入所有改进。

由表 1 可得,仅在主干网络单独设计 SPNet 多尺度特征融合模块,模型的 $mAP@0.5$ 提高了 2.6%,对模型的检测精度提升最为显著。另外,在主干网络单独使用 ECG 轻量融合模块,参数量和计算量分别下降了 3.08% 和 7.3%,但精度略降,其原因是模型的特征提取信息不佳。使用 ECG 轻量融合模块和 SPNet 多尺度特征融合模块配合能够得到更好的性能, $mAP@0.5$ 提高了 2.9%。因此,ECG 轻量融合模块并不适合单独使用。

3.5 对比实验

为了进一步验证改进后的模型的性能,将其与目前主流模型: SSD、Faste-RCNN、YOLOv5、YOLOv7-tiny 和 YOLOX、YOLOv10 等算法进行对比实验,实验结果如表 2 所示。

由表 2 可看出,与基准 YOLOv9s 相比,本文算法在精确度、召回率和平均精度 $mAP@0.5$ 方面分别提升了 1.9%、3.3% 和 3.5%,进一步验证了改进算法的有效性。与其他算法的对比结果显示,相比于 YOLOv5 的精确度、召回率和均值平均精度分别提升了 16.2%、12.5% 和 16.1%,相比于 YOLOv7-tiny 和 YOLOv10 s 的精确度分别提升了 7.5% 和 3%,与 Faste-RCNN 相比 $mAP@0.5$ 提升了 12.7%,参数量减少了 44.09 M。

表 1 消融实验

Table 1 Ablation experiments

模型	SPNet	ECG	FADC	NWD-EIOU	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
YOLOv9s_1	—	—	—	—	85.3	89.0	89.7	9.74	39.6
YOLOv9s_2	✓	—	—	—	86.6	91.5	92.3	9.72	45.0
YOLOv9s_3	—	✓	—	—	85.6	95.3	89.5	6.43	35.2
YOLOv9s_4	—	—	✓	—	85.1	89.4	91.0	9.77	39.5
YOLOv9s_5	—	—	—	✓	86.6	90.2	91.2	9.74	39.6
YOLOv9s_6	✓	✓	—	—	87.4	90.4	92.6	9.29	38.7
YOLOv9s_7	✓	✓	✓	—	87.2	91.2	92.7	9.51	38.0
YOLOv9s_8	✓	✓	✓	✓	87.2	92.3	93.2	9.51	38.0

表 2 算法对比

Table 2 Algorithm comparison

模型	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
Faste-RCNN	70.4	76.9	80.5	53.6	69.4
SSD	75.3	51.2	81.2	26.4	68.2
Mask-RCNN	71.3	77.2	80.9	54.0	69.9
EfficientDet	82.6	80.4	85.9	18.2	12.3
YOLOv5s	71.0	79.8	77.1	2.33	35.7
YOLOv7-tiny	79.3	84.5	85.7	11.9	42.5
YOLOX	89.6	86.4	88.2	9.87	36.9
YOLOv8s	86.4	87.1	88.9	10.8	37.8
YOLOv10s	87.1	89.6	90.2	9.21	36.1
YOLOv9s	85.3	89.0	89.7	9.74	39.6
文献[17]	86.2	90.3	90.3	9.67	33.2
文献[18]	87.1	91.7	90.9	9.63	38.3
YOLOv9s-SEFN	87.2	92.3	93.2	9.51	38.0

在与改进算法文献[17]、文献[18]的比较中,本文的改进模型在 mAP@0.5 上分别提升了 2.9%、2.3%,同时计算复杂度和参数量均有所下降。综上所述,本文算法在当前主流模型中表现出更优的检测性能。

本文改进算法与初始算法在训练过程中的损失函数变化如图 8 所示。从图 8 中可以看出,原始算法的损失曲线(绿色实线)下降较慢,在收敛性和稳定性方面存在一定不足。而改进算法的损失曲线(红色虚线)在整个训练过程中下降显著更快,损失值显著低于原始算法。

为更直观地展示本文算法的有效性,以图像的形式对比基准模型与 YOLOv9s-SEFN 模型的检测结果,用矩形框表示,并标注了识别出的标签和相应的置信度。如图 9 所示,输入 5 种不同的缺陷图像,与原模型相比,本文提出的改进模型在检测黑点、光圈、边异常和裂纹 4 种缺陷时,置信度明显提升,识别精度更高。在角异常缺陷的检测中,改进后的算法不仅保持了更高的置信度,还有效避免了误检,目标定位更加准确。所有实验结果表明,本文提出的 YOLOv9s-SEFN 模型在检测精度上表现出显著优

势,有效地减少了漏检误检,在小缺陷检测中取得了更高的准确率。

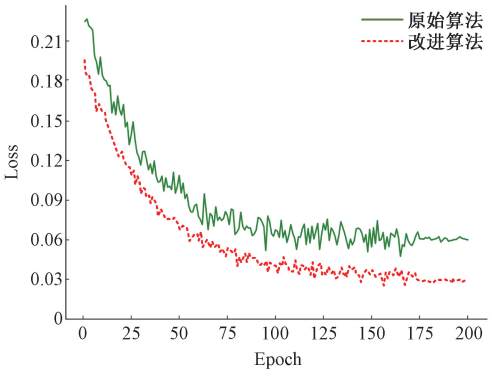


图 8 改进前后算法损失函数对比

Fig. 8 Comparison of loss functions before and after algorithm improvement

3.6 泛化实验

为了验证本文改进算法的泛化能力,本文采用了北京

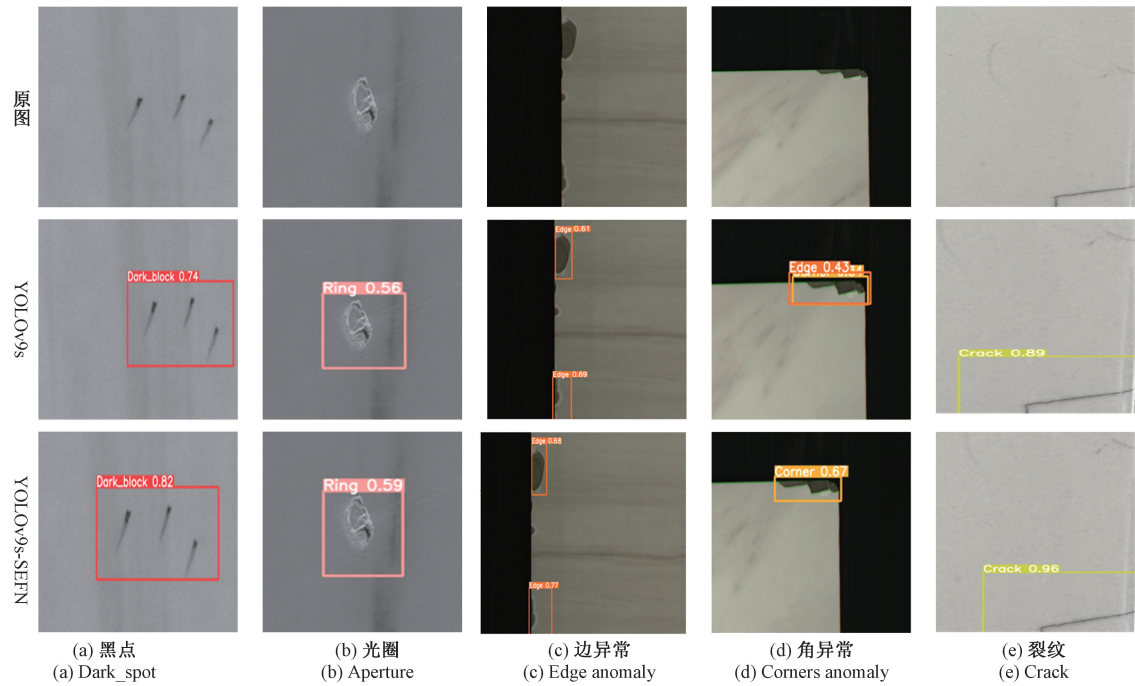


图 9 检测结果可视化图
Fig. 9 The comparison chart of detection result

大学开源的 PCB 缺陷检测数据集进行实验。该数据集包含 693 张高质量的 PCB 图像,每张图像均已标注相应的缺陷类别标签,涵盖了 6 种常见的缺陷类型:缺失孔(missing_hole)、老鼠咬伤(mouse_bite)、开路(open_circuit)、短路

(short)、毛刺(spur)和假铜(spurious_copper)。在相同的实验条件下,本文对 YOLOv9s 及其改进算法进行了性能对比,实验结果如表 3 所示。

表 3 算法对比
Table 3 Algorithm comparison

模型	P/%	R/%	mAP@0.5/%	Params/M	GFLOPs/G
YOLOv9s	94.3	83.8	89.0	9.60	40.1
YOLOv9s-SEFN	89.8	90.1	91.6	8.19	37.7

根据表 3 所呈现的实验数据分析,相较于 YOLOv9s 模型,经过优化的 YOLOv9s-SEFN 模型在 PCB 数据集上依然展现出了很好的性能。改进模型的平均精度 mAP@0.5 提高了 2.6%,在参数量和计算复杂度方面分别减少了 14.7%和 6.0%。实验结果证明,YOLOv9s-SEFN 模型在性能上取得了显著提升,并展现出了良好的泛化能力和鲁棒性。

为了更全面地验证 YOLOv9s-SEFN 模型在北京大学开源的 PCB 缺陷检测数据集上的实际应用效果,本研究选取了 6 种具有代表性的不同缺陷类别的图像进行了检测实验。实验结果如图 10 所示。从图中可以看出,相较于原始模型,本文提出的改进模型在检测缺失孔、鼠咬咬伤、开路与短路四类缺陷时,识别精度显著提升。在毛刺和假铜缺陷两类的检测中,改进后的算法不仅保持了较高的识别精度,还有效避免了漏检,目标定位更加精准。综上所述,

所有实验结果均表明,本文提出的 YOLOv9s-SEFN 模型在小缺陷检测中表现出了更高的准确性,减少了漏检误检,展现出了良好的泛化能力。

4 实验与结果分析

本文提出了一种基于改进 YOLOv9s 框架的瓷砖表面缺陷检测算法,在主干网络中设计 SPNet 模块增强模型对瓷砖表面小缺陷的特征表达能力,设计 ECG 模块减少计算量和参数量,在颈部加入 FADC 卷积,最后采用 NWD-EIOU 损失函数,提高检测精度。通过消融实验与对比实验,验证了所提改进算法的可行性和优越性。改进后的算法显著提升了瓷砖表面缺陷检测的准确度,实现了模型的轻量化,有效减少了漏检和误检现象,满足实际检测需求。未来工作将继续优化模型结构,进一步提升轻量化水平,在保证检测精度的前提下提高检测速度。

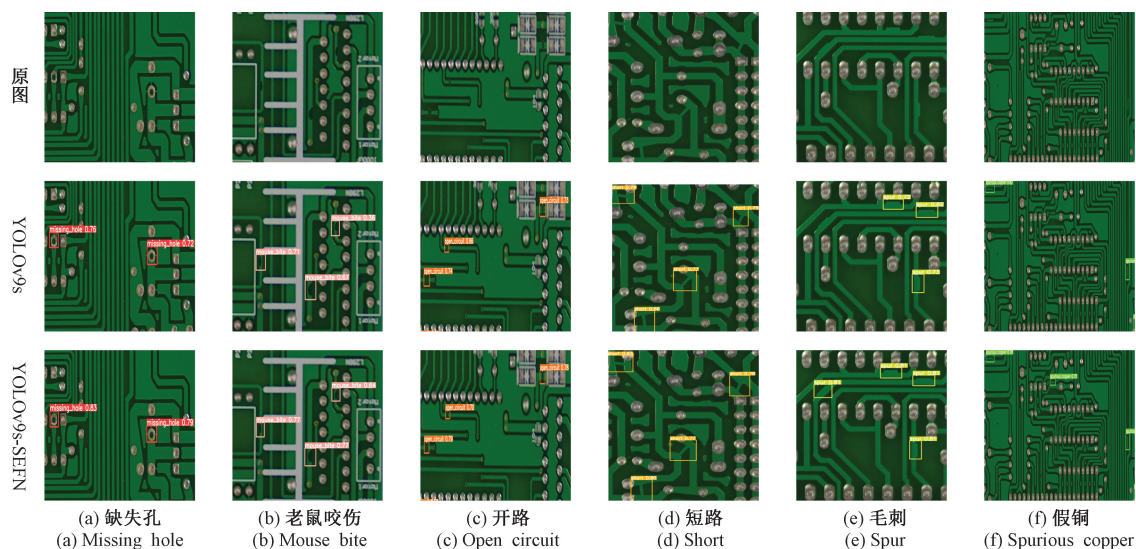


图 10 检测结果可视化图

Fig. 10 The comparison chart of detection results

参考文献

- [1] KARIMI M H, ASEMANI D. Surface defect detection in tiling industries using digital image processing methods: Analysis and evaluation[J]. ISA Transactions, 2014, 53(3): 834-844.
- [2] OZKAN F, ULUTAS B. Use of an eye-tracker to assess workers in ceramic tile surface defect detection[C]. 2016 International Conference on Control, Decision and Information Technologies (CoDIT). IEEE, 2016: 88-91.
- [3] GIRSHICK R, DONAHUE J, DARRELL T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition. New York: IEEE Press, 2014: 580-587.
- [4] AFZAAL U, BHATTARAI B, PANDEYA Y R, et al. An instance segmentation model for strawberry diseases based on mask R-CNN[J]. Sensors, 2021, 21(19): 6565.
- [5] 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.
- [6] GAO X, JIAN M, HU M, et al. Faster multi-defect detection system in shield tunnel using combination of FCN and faster RCNN[J]. Advances in Structural Engineering, 2019, 22(13): 2907-2921.
- [7] 翁玉尚, 肖金球, 夏禹. 改进 Mask R-CNN 算法的带钢表面缺陷检测[J]. 计算机工程与应用, 2021, 57(19): 235-242.
- [8] WENG Y SH, XIAO J Q, XIA Y. Strip surface defect detection based on improved mask R-CNN algorithm [J]. Computer Engineering and Applications, 2021, 57(19): 235-242.
- [9] 朱成杰, 刘乐乐, 朱洪波. 基于 YOLOv8-NFMC 的带钢表面缺陷检测算法[J]. 国外电子测量技术, 2024, 43(7): 97-104.
- [10] ZHU CH J, LIU L L, ZHU H B. Surface defect detection algorithm for strip steel based on YOLOv8-NFMC [J]. Foreign Electronic Measurement Technology, 2024, 43(7): 97-104.
- [11] 吴航星, 张宏怡, 谭湘琼, 等. 改进型 YOLOv5s 的瓷砖表面缺陷检测方法[J]. 厦门理工学院学报, 2023, 31(5): 66-74.
- [12] WU H X, ZHANG H Y, TAN X Q, et al. An algorithm for enhanced tile surface defect detection based on YOLOv5s[J]. Journal of Xiamen University of Technology, 2023, 31(5): 66-74.
- [13] 余松森, 林智凡, 薛国鹏, 等. 基于改进 YOLOv8 的轻量级大幅面瓷砖缺陷检测算法[J]. 计算机应用, 2024, 44(1): 1-7.
- [14] YU S S, LIN ZH F, XUE G P, et al. Lightweight large format tile defect detection algorithm based on improved YOLOv8 [J]. Journal of Computer Applications, 2024, 44(1): 1-7.
- [15] LIU W, ANGUELOV D, ERHAN D, et al. SSD: Single shot multibox detector[C]. Computer Vision-ECCV 2016: 14th European Conference, Springer International Publishing, 2016: 21-37.

- [12] ARUN R A, UMAMAHESWARI S. Effective and efficient multi-crop pest detection based on deep learning object detection models [J]. Journal of Intelligent & Fuzzy Systems, 2022, 43 (4): 5185-5203.
- [13] 王淑青, 顿伟超, 黄剑锋, 等. 基于 YOLOv5 的瓷砖表面缺陷检测[J]. 包装工程, 2022, 43(9): 217-224.
WANG SH Q, DUN W CH, HUANG J F, et al. Ceramic tile surface defect detection based on YOLOv5 [J]. Packaging Engineering, 2022, 43(9): 217-224.
- [14] 余松森, 薛国鹏, 何皇, 等. 改进 YOLOv8 的轻量级瓷砖表面缺陷检测[J]. 计算机工程与应用, 2024, 60(18): 88-102.
YU S S, XUE G P, HE H, et al. Lightweight detection of ceramic tile surface defects on improved YOLOv8[J]. Computer Engineering and Applications, 2024, 60(18): 88-102.
- [15] CAO Y K, PANG D D, ZHAO Q CH, et al. Improved YOLOv8-GD deep learning model for defect detection in electroluminescence images of solar photovoltaic modules[J]. Engineering Applications of Artificial Intelligence, 2024, 131: 107866.
- [16] 李忠科, 刘小芳. 基于轻量级 YOLOv8n 网络的 PCB 缺陷检测算法[J]. 电子测量技术, 2024, 47(4): 120-126.
LI ZH K, LIU X F. PCB defect detection algorithm based on lightweight YOLOv8n network [J]. Electronic Measurement Technology, 2024, 47(4): 120-126.
- [17] WANG G, ZHANG Z, TANG R, et al. A defect recognition method for switchgear infrared images based on improved YOLOv9s [C]. 2024 6th International Conference on Energy, Power and Grid (ICEPG). IEEE, 2024: 1371-1377.
- [18] SHI Y, DUAN Z, QING S, et al. YOLOV9S-Pear: A lightweight YOLOV9S-Based improved model for young red pear small-target recognition [J]. Agronomy, 2024, 14(9): 2086.
- [19] 周建新, 李忠泽, 郝英杰. 基于改进 YOLOv9 的钢板表面缺陷检测的方法[J]. 电子测量技术, 2024, 47(22): 181-188.
ZHOU J X, LI ZH Z, HAO Y J. Method for defect detection on steel plate surface based on improved YOLOv9 [J]. Electronic Measurement Technology, 2024, 47(22): 181-188.
- [20] WANG C Y, YEH I H, MARK LIAO H Y. YOLOv9: Learning what you want to learn using programmable gradient information [C]. European Conference on Computer Vision. Springer, Cham, 2024.
- [21] YANG L, ZHANG R Y, LI L, et al. SimAM: A simple, parameter-free attention module for convolutional neural networks [C]. International conference on machine learning. PMLR, 2021: 11863-11874.
- [22] CHEN L W, GU L, ZHENG D ZH, et al. Frequency-adaptive dilated convolution for semantic segmentation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 3414-3425.
- [23] 谢国波, 林松泽, 林志毅, 等. 基于改进 YOLOv7-tiny 的道路病害检测算法[J]. 图学学报, 2024, 45(5): 987-997.
XIE G B, LIN S Z, LIN ZH Y, et al. Road defect detection algorithm based on improved YOLOv7-tiny [J]. Journal of Graphics, 2024, 45(5): 987-997.
- [24] ZHENG ZH H, WANG P, REN D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 8574-8586.
- [25] ZHANG Y F, REN W, ZHANG Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. Neurocomputing, 2022, 506: 146-157.
- [26] WANG J W, XU CH, YANG W, et al. A normalized Gaussian Wasserstein distance for tiny object detection[J]. ArXiv preprint arXiv: 2110.13389, 2021.

作者简介

尚梦佳, 硕士研究生, 主要研究方向为图像处理、计算机视觉、目标检测。

E-mail: 2574530539@qq.com

刘利平(通信作者), 博士, 教授, 主要研究方向为基于模式识别技术的无损检测方法及应用。

E-mail: 11745430@qq.com