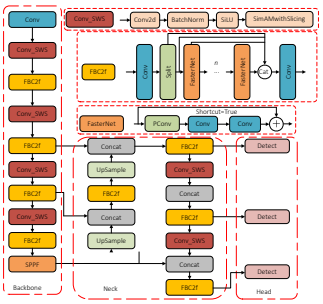
OEE | 光电工程 Opto-Electronic Engineering

Article

2026 年 2 月, 第 53 卷, 第 2 期

PCB-YOLO：改进 YOLOv8n 的 PCB 微小缺陷轻量化检测方法



邓文斌[1]，郭怡希[1]，林源强[2]，鹿业波[1]，杨　俊[2*]

[1]嘉兴大学机械工程学院，浙江 嘉兴 314001；

[2]嘉兴大学人工智能学院，浙江 嘉兴 314001

摘要：针对不同底纹背景颜色的印制电路板 (printed circuit board, PCB) 微小缺陷检测样本少、效率低、检测算法参数量大和精度低的问题，本文基于 YOLOv8n 框架提出了一种轻量化检测模型 PCB-YOLO。首先，对原始的多种颜色背景的大幅面 PCB 数据集进行数据预处理，通过灰度变换和切片操作得到适合 YOLOv8 输入尺寸的 PCB 数据和降低颜色背景干扰影响，防止 PCB 的微小缺陷在缩放的过程中丢失；其次，考虑到 PCB 存在小目标缺陷的占比较大问题，通过结合 SimAM 注意力机制和合并切片操作，设计 Conv_SWS (SimAM with slicing) 模块，替代传统的下采样方式，保留小目标的关键信息，从而增强 PCB 中微小目标缺陷的特征提取并提高检测准确性；然后，对 YOLOv8 中原有的 C2f 模块进行改造，引入部分卷积 (PConv) 对关键特征通道进行综合计算，设计 FBC2f 模块，用于增强特征提取能力并减少模型参数量；最后，采用基于 Wasserstein 距离的度量标准 NWD (normalized Wasserstein distance) 与标准 IoU (intersection over union) 结合改进损失函数来加速模型收敛，进一步提高 YOLOv8 对微小目标的检测能力。实验结果表明，与原始的 YOLOv8n 相比，在 PCB 缺陷检测数据集上的 PCB-YOLO 模型的参数减少了 23.6%，mAP_{50} 提高了 2.9%，达到 95.6%。在工业铝片表面缺陷、PKU-Market-PCB 缺陷数据集上的泛化性实验也验证了 PCB-YOLO 模型的有效性。

关键词：PCB 检测；微小目标检测；YOLOv8n；度量标准；注意力机制

The English title and abstract appear at the end of the paper.

DOI: 10.12086/oee.2026.250252 | CSTR: 32245.14.oee.2026.250252 | 中图分类号：TP391 | 文献标识码：A

邓文斌，郭怡希，林源强等. PCB-YOLO：改进 YOLOv8n 的 PCB 微小缺陷轻量化检测方法 [J]. 光电工程，2026，**53**(2): 250252

Deng W B, Guo Y X, Lin Y Q et al. PCB-YOLO: an enhanced YOLOv8n for lightweight detection of tiny defects in PCB[J]. *Opto-Electron Eng*, 2026, **53**(2): 250252

1　引　言

为了适应市场对复杂电子电路日益增长的需求，印制电路板 (printed circuit board, PCB) 正不断向高精度、多层化与微型化演进，在工业化大规模生产过程中难免导致产品缺陷[1]。PCB 上的微小缺陷可能导致整个电子设备的失效，因此对 PCB 进行高精度的缺陷检测至关重要。由于 PCB 制造工艺复杂，缺陷尺寸越来越小，这给检测带来了新的挑战。如何准确检测 PCB 上的微小缺陷，成为当前 PCB 制造过程研究的热点问题。

在 PCB 生产过程中，常见的缺陷主要有鼠咬、开路、短路、毛刺和伪铜。这些缺陷的特点如下：1) PCB

收稿日期：2025-08-25；修回日期：2025-10-24；录用日期：2025-10-28；网络出版日期：2026-02-13

基金项目：浙江省“尖兵领雁”研发攻关计划 (2024C04028)；浙江省教育厅一般科研项目 (Y202455539)；平湖市科技计划项目 (产业发展攻关专项 202505)；嘉兴大学“智汇平湖•嘉人驻企”科研课题 (00324264AL)

*通信作者：杨俊，juneryoung@zjxu.edu.cn

版权所有©2026 中国科学院光电技术研究所

电路设计多种多样，导致 PCB 缺陷的背景差异很大；2) 缺陷面积小，颜色与背景相似，增大了检测难度；3) 虽然存在多种缺陷，但它们的特征是相似的，这使得检测更加困难，容易出现误检和漏检。这些特性可能会导致 PCB 表面缺陷检测存在以下问题：1) 复杂的纹理导致前景和背景区分困难；2) 微小的尺寸导致检测困难。由于缺陷目标小、种类多、区分难等因素，传统检测技术已经不能满足实际生产需求[2]。

为应对上述挑战，PCB 缺陷检测技术已从传统人工检测演进至智能算法驱动的自动检测。早期 PCB 缺陷检测依赖人工目视，易受视觉疲劳影响，误检率超 10%、效率较低，难以适配大规模生产；X 射线检测虽能穿透多层 PCB，但设备成本超百万元，且难以完成表面缺陷检测。自动光学检测 (Automated optical inspection, AOI)[3] 是当前工业主流检测技术，其中基于传统图像处理的 AOI 通过手工设计特征实现缺陷识别，Tsai 等人[4] 提出了一种全局傅里叶图像重建方法来检测和定位 PCB 表面缺陷，该方法可以适应平移变化和光照变换环境。Hassanin 等人[5] 提出了一种基于 SURF 特征和形态学运算的实时 PCB 自动缺陷检测方法，可以准确地确定缺陷位置和缺陷类型。这类方法的核心局限在于特征依赖人工设计、鲁棒性差，无法适应 PCB 多样化底纹与微小缺陷的检测需求。

为突破人工设计特征的限制，研究者引入机器学习算法实现特征自动提取，Tsai 等人[6] 将颜色与结构特征输入 SVM 分类器，提升了针孔、露铜缺陷的检测精度，但 SVM 对高维特征的处理能力有限，当 PCB 背景存在噪声时准确率下降 15% 以上；Lu 等人[7] 提出贝叶斯特征融合框架，改善了光照不均场景的检测效果，但模型仍需依赖边缘、纹理等手工设计的基础特征，对微小缺陷的特征表征能力不足，这类方法虽比传统图像处理更具适应性，但未摆脱对人工特征的依赖，且模型泛化能力受限于训练样本分布，面对 PCB 工业场景中“小样本、多缺陷类型”的特点时，难以满足检测精度要求。

深度学习通过端到端特征学习，显著提升了 PCB 缺陷检测的鲁棒性，按网络结构可分为两阶段模型与单阶段模型。两阶段模型 (如 Faster R-CNN) 通过“区域提议-分类回归”实现高精度检测，在 PCB 小缺陷检测中得到广泛探索，Hu 等人[8] 使用 GA-RPN (region proposal network by guided anchoring) 改进 Faster RCNN 算法，从而提高了 PCB 中小缺陷的检测能力。同时，Zeng 等人[9] 提出了一种基于非对称平衡特征金字塔网络 (asymmetric balanced feature pyramid network, ABFPN) 的增强多尺度特征融合算法，实现了有效的小目标检测。孙志超等人[10] 通过引入可变形残差卷积替换 VGG16 网络改进 Faster RCNN，并结合 IoU Loss 等组合优化检测网络，实现了对 PCB 缺陷图像数据的检测精度和速度。两阶段模型的核心瓶颈在于参数量大、推理慢，难以适配边缘设备的轻量化需求，因此单阶段模型 (尤其是 YOLO 系列) 成为近年的研究热点。

YOLO 系列模型[11-14] 将检测任务整合为单阶段回归，兼顾速度与精度，成为 PCB 缺陷检测的主流选择，但现有改进仍存在三大待解决问题：1) 小目标特征提取不足。何国忠和梁宇[15] 在 YOLOv4 中引入长距离全局注意力，增强浅层特征信息，但未针对 PCB 微小缺陷的“特征稀疏性”优化，当缺陷尺寸小于 8 pixel×8 pixel 时，注意力权重易被背景噪声干扰。Xia 等人[16] 提出了结合全局上下文注意力和双向加权特征金字塔网络 (bi-directional weighted feature pyramid network, BiFPN) 特征融合的 YOLO 模型，提升了多尺度特征融合效果，但未改进下采样过程，小目标特征在池化操作中仍会丢失。2) 轻量化与精度平衡不足。李振华等人[17] 采用轻量化 GhostNet 网络作为算法模型的特征提取网络，减少 YOLOv5 算法参数量，但 mAP50 下降。Bhattacharya 等人[18] 用 Transformer 替换 YOLOv5 骨干网络，虽然保持了精度，但参数量反增超 20%。Lei 等人[19] 通过张量分解简化卷积，轻量化效果显著，但特征交互能力减弱，对复杂底纹 PCB 的鲁棒性不足。3) 损失函数适配性差。现有 YOLO 模型多采用 CIoU 损失，但 PCB 微小缺陷的边界框标注误差会导致 CIoU 损失值剧烈波动，影响收敛稳定性。Wise-IoU、Shape-IoU 虽优化了回归策略，但未考虑小目标“面积占比低、重叠率低”的特点，当预测框与真实框无重叠时，模型仍无法提供有效梯度。综上，现有 PCB 缺陷检测技术仍存在三大核心问题：微小目标特征易丢失、模型轻量化与精度难以平衡、损失函数对小目标适配性差。这些问题直接导致检测效率低、部署成本高，无法满足工业需求。

针对以上问题，本文以 YOLOv8n 算法为基线模型，提出了 PCB-YOLO 的轻量化 PCB 缺陷检测算法，解决 PCB 底纹颜色差异大、缺陷尺寸小导致误检和漏检及模型复杂不易部署到边缘处理器的问题。首先对原始

PCB 数据集进行预处理，获得符合模型需要的数据集；其次，结合 SimAM 注意力机制和合并切片操作设计了 Conv_SWS 模块，将其用于网络的卷积模块中，代替传统的下采样方式，增强了 PCB 中微小目标缺陷的特征提取并提高了检测准确性；然后通过引入部分卷积来对关键特征通道进行综合计算，设计了 FBC2f 模块，用于替换 YOLOv8 中原有的 C2f 模块，从而增强特征提取能力并降低模型参数量和计算复杂度；最后通过改进损失函数来加速模型收敛，进一步提高 PCB-YOLO 模型对 PCB 微小缺陷目标的检测能力。

2 数据集建立及数据预处理

2.1 数据集建立

本文使用的原始数据集来源于华为云竞赛平台(https://competition.huaweicloud.com/information/1000042047/circumstance)。原始数据集包含 68 张原始 PCB 表面缺陷图像，图片尺寸为 3072 pixel×2048 pixel，每张图片包含一个及以上的缺陷目标。这些目标囊括了 PCB 生产流程中可能出现的多种缺陷类型，包含鼠咬(mouse_bite)、开路 (open_circuit)、短路 (short)、毛刺 (spur) 以及伪铜 (spurious_copper) 5 类缺陷，其缺陷类型如图 1 所示。

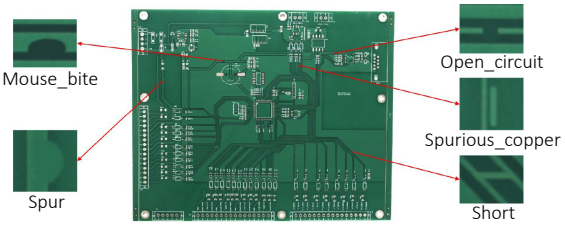


图 1 缺陷类型

Fig. 1 Defect types

2.2 数据预处理

在 PCB 生产过程中，存在多种颜色的底纹，而数据集给出的每种样本数量又少，为了增强模型识别的可靠性，降低不同底纹颜色对缺陷识别的干扰，首先对样本数据进行了灰度化处理，将不同颜色底纹的 PCB 板 RGB 图像使用式 (1) 统一转化为灰度图[20]，其中：R、G、B 分别表示红色、绿色、蓝色。图 2 展示了不同底纹背景的 PCB 转换为灰度图以后的效果，从结果可以看到这样可以减少不同背景颜色对缺陷识别的干扰。

$$Gray = R\times 0.299 + G\times 0.587 + B\times 0.114\,. \tag{1}$$

由于采集的 PCB 图像分辨率较高，且待检测的缺陷目标往往尺寸较小，而 YOLOv8 模型默认的输入图像尺寸为 640 pixel×640 pixel，若直接将原始图像强制缩放到该尺寸，不仅会在训练阶段占用大量显存资源，对计算设备提出较高要求。更严重的是，这种处理方式可能导致小尺度缺陷的关键特征在尺寸变换过程中被过度压缩而消失，最终影响模型的识别效果。因此，为缓解高分辨率 PCB 图像在输入 YOLOv8 模型时因尺寸压缩而可能造成的微小缺陷漏检问题，本文引入图像分块预处理方法，对原始图像进行切分处理。将原图像连续分割为多个尺寸为 640 pixel×480 pixel 的子图像，每个子图像拥有 100 pixel 的重叠区域，避免因缺陷位于滑动窗口边缘而造成漏检，本方法在图像分割处理过程中，依据原始标签文件所提供的缺陷区域坐标，判断当前滑动窗口所覆盖范围内是否包含缺陷目标。通过该机制，可有效筛选出含有缺陷的 PCB 图像切片，并剔除仅包含背景信息的无效切片，从而提升检测效率与准确性。

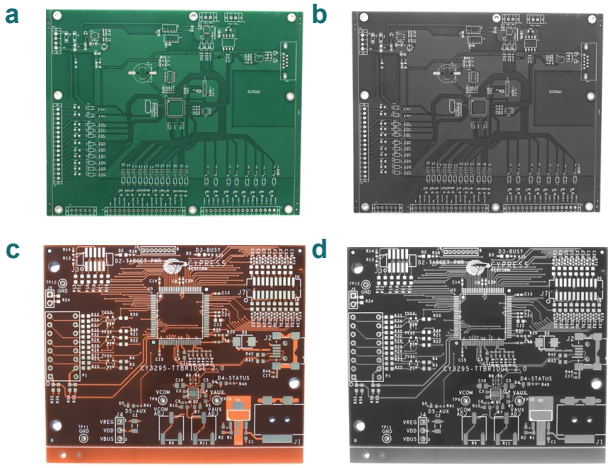


图 2 转换前后的 PCB 图。(a) 绿色底纹 PCB 图；(b) 绿色底纹灰度化后图像；(c) 棕色底纹 PCB 图；(d) 棕色底纹灰度化后图像

Fig. 2 PCB images before and after conversion. (a) PCB diagram with green shading; (b) Image after grayscale processing with green shading; (c) PCB diagram with brown shading; (d) Image after grayscale processing with brown shading

2.3 数据分析及划分

经过图像预处理后，针对 PCB 缺陷数据集中存在的样本分布不平衡及数量有限的问题，本文采用水平镜

像翻转、竖直镜像翻转、添加均值为 0、标准差为 0.1 的高斯噪声等多种数据增强方法对原始数据集进行扩充。经增强后，数据集规模增至 5172 张图像，每幅图像均包含至少一个缺陷，缺陷总数达到 6282 个。图 3 展示了数据集中缺陷的高宽比分布情况，从分析结果可以看出，缺陷的高宽比数值覆盖范围较广，属于多尺度特征目标，给检测带来了较高的难度。

图 4 统计了数据集中所有缺陷的类别与尺寸分布及数量。由图可知，缺陷尺寸主要集中在 32 pixel×32 pixel 以内，所占比例高达 98.3%，符合小目标检测的典型范畴[21]。此外，还存在 546 个尺寸小于 8 pixel×8 pixel 像素的极微小缺陷，占比约 8.7%，图 5 展示了多个小于 8 pixel×8 pixel 的微小缺陷。这类缺陷在多次下采样操作中极易丢失细节信息，导致其在 YOLOv8 模型中的检测难度显著增加。由于较低的区域覆盖率和较少的特征点，微小缺陷和背景混合在一起形成低对比度，这使得微小缺陷的检测具有极大的挑战性。将扩增后的 PCB 数据集采用 8∶1∶1 的比例随机划分为训练集、验证集和测试集，确保各数据集均囊括了多种类型及尺度的缺陷样本，以保证数据的多样性。

3 PCB-YOLO 检测算法

3.1 YOLOv8 检测模型

YOLOv8 模型是 Ultralytics 公司基于 YOLOv5 进行

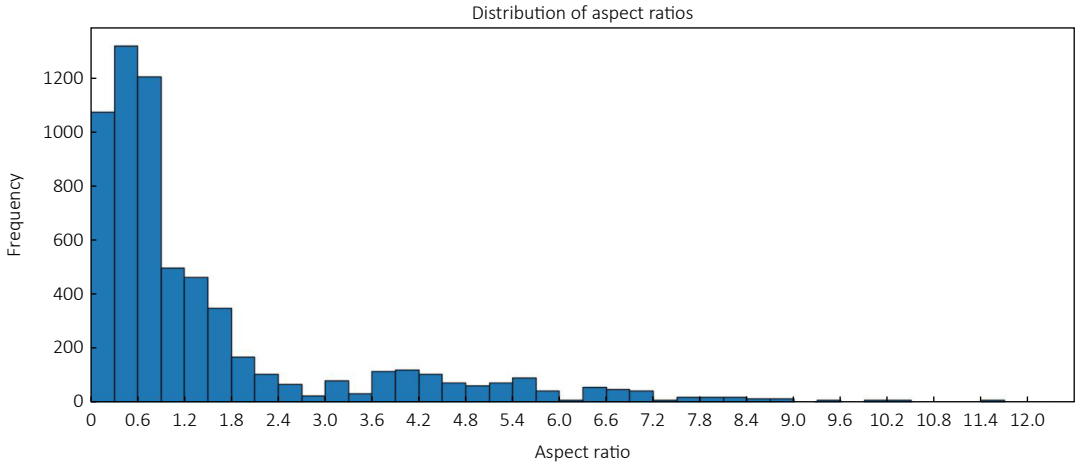


图 3 数据集中缺陷的宽高比

Fig. 3 Aspect ratio of defects in the dataset

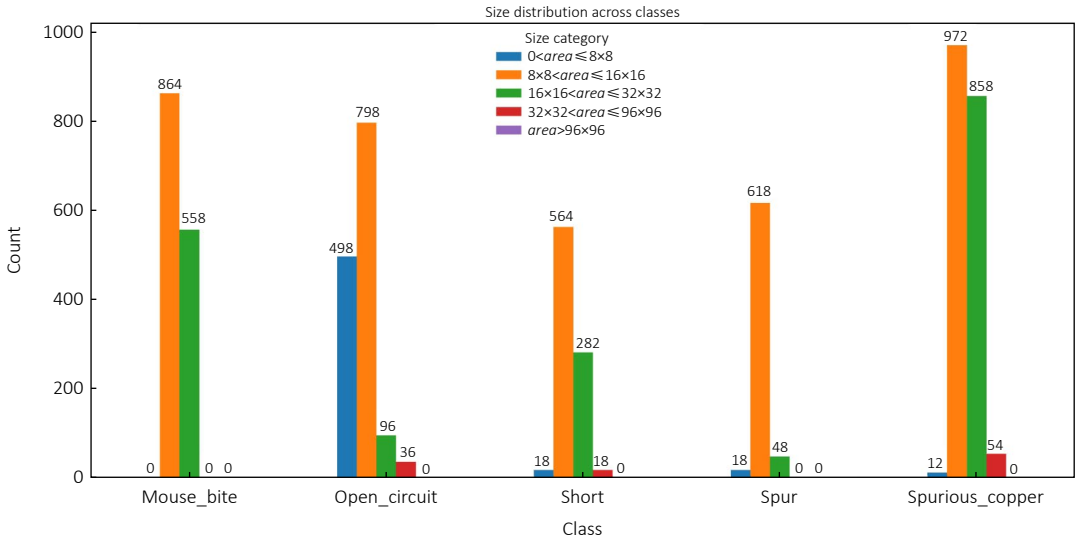


图 4 数据集中缺陷类别与尺寸的分布情况

Fig. 4 The distribution of defect types and sizes in the dataset

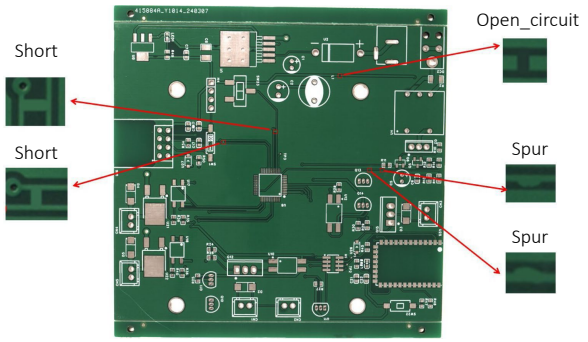


图 5 微小缺陷样例

Fig. 5 Example of minor defects

优化改进的，在准确率和速度方面都得到了较大提升。YOLOv8 的网络结构主要由三个部分组成：提取特征的主干网络、执行特征融合的颈部网络和目标定位及分类的预测头网络，结构细节如图 6 所示。YOLOv8 的骨干网络采用 C2f 模块[22] 替换 C3 模块，并通过 Bottleneck 模块 (*m* 表示 BottleNeck 模块的数量) 和 SPPF 模块提升特征提取能力，同时降低计算复杂度和延迟[23]。颈部网络部分使用 PAN-FAN 结构进行多尺度特征融合，删除上采样层前的卷积层，对模型进行压缩。预测头使用目前主流的解耦头结构，将分类和检测头分离，并使用 Anchor Free 检测方法[24]，不再依赖先验锚框，使模型能够针对不同任务进行优化。

3.2 改进的 PCB-YOLO 微小目标缺陷检测模型

为解决 PCB 缺陷检测中微小目标识别精度低、漏检/误检率高的问题，以及完成对实时性要求的双重挑战，需要保证模型的轻量化和低参数量，本文提出了一种基于 YOLOv8n 改进的 PCB-YOLO 模型，所提出的模型结构如图 7 所示。

PCB-YOLO 的主干网络包含普通卷积 (Conv)、Conv_SWS 卷积模块、FBC2f 和 SPPF 模块。Conv 模块通过 Conv2d、BatchNorm2d 和 SiLU 激活函数对输入图像进行通道数调整与预处理；Conv_SWS 模块则融合卷积与切片操作，有效捕获不同尺度目标的关键特征。FBC2f 模块利用 PConv 有效降低网络的计算量和参数量，

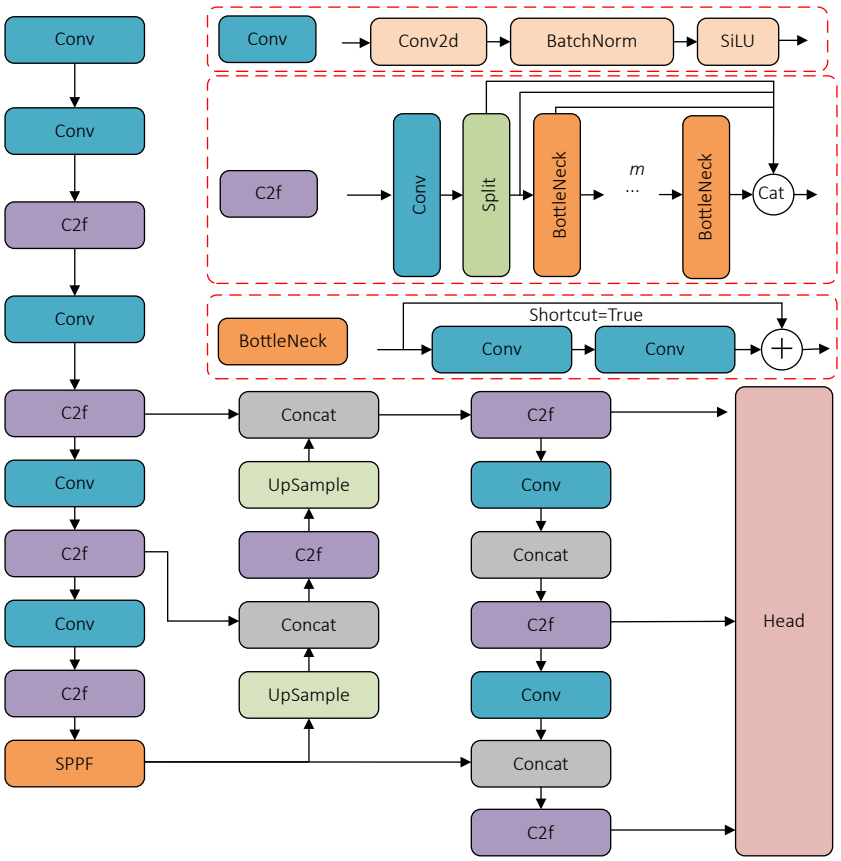


图 6 YOLOv8 模型结构

Fig. 6 YOLOv8 model structure

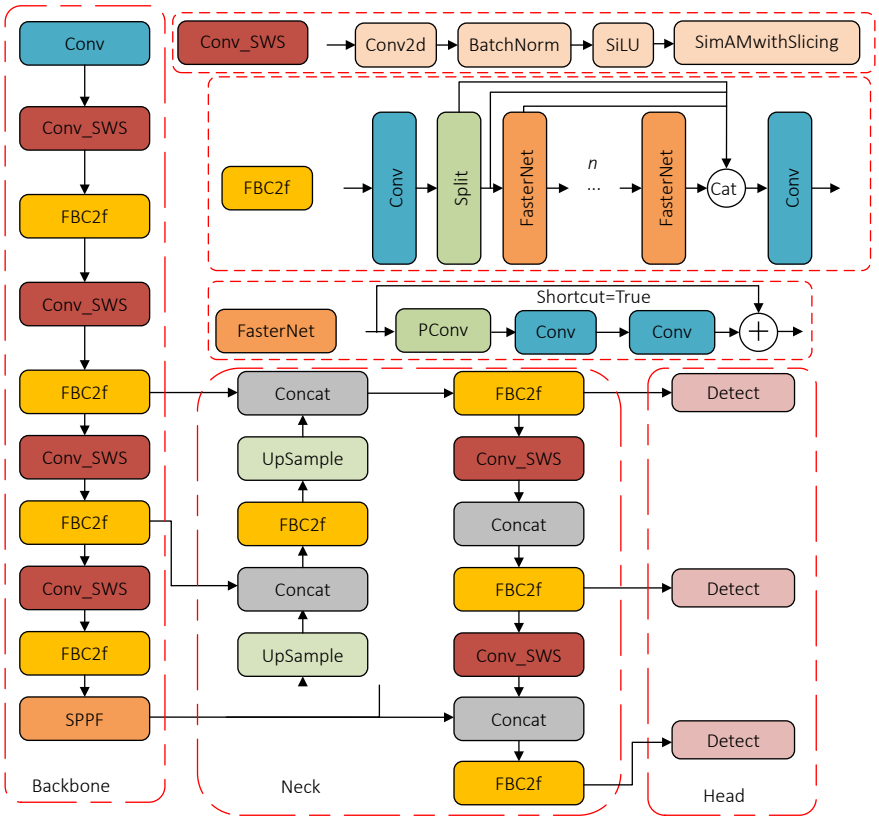


图 7 PCB-YOLO 结构

Fig. 7 PCB-YOLO structure

并通过 FasterNet 增加特征交互；SPPF 模块进一步提升特征提取能力，同时降低计算复杂度，有助于 PCB-YOLO 模型的小目标检测能力。

PCB-YOLO 的颈部网络包含 Conv_SWS 和 FBC2f 模块，Conv_SWS 模块通过融合不同大小特征图尺寸，实现多尺度特征融合；FBC2f 模块采用部分卷积结构构建特征提取单元，在提升局部特征捕获效率的基础上有效降低了计算复杂度，从而进一步强化了 PCB-YOLO 模型的轻量化性能。

3.2.1 Conv_SWS 模块

SimAM (Simple attention mechanism)[25] 是一种三维自注意力模块，能够同时整合空间与通道信息。该模块以轻量化著称，可在不引入额外参数的情况下，有效引导网络聚焦于输入特征中的关键区域，从而增强模型的特征提取与表示能力。与现有的 1D 或 2D 注意力模块不同，SimAM 直接在特征图中的每个神经元上运行。它使用能量函数计算 3D 特征权重，其中权重不仅考虑通道维度，还考虑空间维度。这使得能够更精确地关注特征图中的重要信息。SimAM 注意力机制的集成过程如图 8 所示。首先，从特征 $\boldsymbol{X}$ 生成 3D 注意力权重，然后将该注意力权重集成到特征图中，最后应用扩展操作来增强特征图的表示能力。

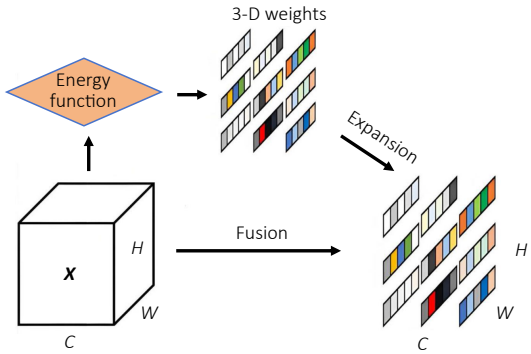


图 8 SimAM 注意力机制

Fig. 8 SimAM attention mechanism

对于特征图 $\boldsymbol{X} \in \mathbf{R}^{C\times H\times W}$(其中 H、W 为特征图高度和宽度，C 为通道数)，SimAM 的最小能量函数 e_t 如式 (2) 所示，用来衡量目标神经元 t 的重要性。最小能量值 e_t 的大小反映了缺陷特征中神经元 t 在对应维度上的

独特性。该值越小，表明该神经元与同维度其他神经元之间的差异性越显著，其重要性也随之提升。因此，神经元的重要性亦可通过 $1/e_t$ 来量化。

$$e_t = \frac{4(\sigma^2+\lambda)}{(t-u)^2+2\sigma^2+2\lambda}, \tag{2}$$

式中：t 表示输入特征图在单一通道内的目标神经元；λ 表示正则化系数；u 和 σ^2 分别表示通道中所有神经元取值的均值和方差，其表达式分别如式 (3) 和式 (4) 所示。

$$u = \frac{1}{M}\sum_{i=1}^{M} x_i, \tag{3}$$

$$\sigma^2 = \frac{1}{M}\sum_{i=1}^{M}(x_i-u)^2, \tag{4}$$

式中：x_i 表示目标神经元在特定维度上的全部神经元集合；$M=H\times W$ 表示该层神经元总数。

最后，依据注意力机制的基本原理，对以上特征信息进行特征增强，获得增强后的输出特征 $\tilde{\boldsymbol{X}}$，如式 (5) 所示。

$$\tilde{\boldsymbol{X}} = sigmoid\left(\frac{1}{E}\right)\odot \boldsymbol{X}, \tag{5}$$

式中：⊙表示按元素相乘；E 表示所有通道和空间维度中 e_t 的集合，通过 sigmoid 对其数值范围进行约束；$\boldsymbol{X}$ 表示图像的输入特征。

然而，当 SimAM 计算整个特征图上像素差异的加权平均值时，它可能会忽略小目标，从而减弱其增强小目标的能力。PCB 图像中的缺陷区域通常占比较小，其在整体图像中的特征与背景均值相近，使得 SimAM 注意力机制在计算权重时难以对其产生显著的增强效果，因此该机制在处理此类小目标缺陷时表现相对有限。为解决这个问题，本文通过引入切片操作来优化 SimAM，从而开发 SWS (SimAM with slicing) 模块。当特征图被划分为多个局部区块时，大目标可能会因分割而损失部分关键特征，进而引起其权重下降；但在后续的特征整合过程中，这些目标仍能够保持较高的可辨识度，同时其显著特征也得到进一步强化；另一方面，经过切片操作的小目标会获得更多的权重，从而增强其特征。改进的 SWS 模块确保了大、小目标都获得增强，从而提高模型的识别能力。切片操作过程如图 9 所示。

Backbone 作为 YOLO 算法中特征提取的关键部分，为提升模型对 PCB 中小目标缺陷的检测效果，设计 Conv_SWS 模块用于捕捉多尺度特征。Conv_SWS 在标准卷积中引入 SWS 模块来增强普通卷积的特征提取能力，且未引入额外的参数量。图 10 为 Conv_SWS 模块的结构图，其中：s 表示步长，p 表示填充。

Conv_SWS 模块的工作流程如下:

1) 对输入特征图 $\boldsymbol{X}\in\mathbf{R}^{H\times W\times C}$ 进行批归一化和 SiLU 激活函数处理，如式 (6) 所示。

$$\boldsymbol{X}_{\text{bn}} = SiLU(BatchNorm2d(\boldsymbol{X})). \tag{6}$$

2) 将标准化后的特征图 $\boldsymbol{X}_{\text{bn}}$ 划分为 K^2 个局部区块，记为 $\boldsymbol{F}(i,j)$，则有

$$\boldsymbol{F}(i,j) = \boldsymbol{X}_{\text{bn}}[i:H:K, j:W:K], \tag{7}$$

式中：$i,j\in[0,K-1]$。

3) 对每个局部区块 $\boldsymbol{X}_{\text{block},i,j}$ 独立计算 SimAM 注意力权重，强化区块内缺陷特征的响应，具体计算过程基于式 (2)~(5)。

4) 将 K^2 个注意力增强后的局部区块 $\boldsymbol{F}(i,j)$ 按原分块位置在通道维度上重新拼接为完整特征图 $\boldsymbol{X}_{\text{merge}}\in\mathbf{R}^{H\times W\times C}$，拼接顺序与步骤 2) 的分块顺序一致，确保特征空间位置的连续性。

5) 对合并后的特征图卷积进行下采样，实现特征图尺寸减半，同时保持通道数不变。

在 PCB 缺陷检测中，存在大量尺寸小于 32 pixel×32 pixel 的微小缺陷。针对此类小目标，传统的 3×3 卷积核难以充分捕获其完整特征，而采用更大尺度的卷积核则容易丢失细节信息，从而削弱网络对小缺陷特征的判别能力。本文所引入的 Conv_SWS 卷积模块，在网络中承担下采样功能。相比原有结构，该模块能够显著缓解下采样过程中的信息损耗，同时增强特征之间的交互关系，并有效扩大感受野范围，进而提升网络对多尺度缺陷的检测性能。

3.2.2 FBC2f 模块

YOLOv8 主干网络采用的 C2f 模块，该结构由多个 Darknet Bottleneck 堆叠而成，在一定程度上增大了模型的复杂性与计算冗余。为缓解该问题，Chen 等[26]引入了部分卷积 (partial convolution, PConv)，该方法通过仅对输入特征图的部分通道进行卷积处理，显著降低了模型的参数总量与计算复杂度。PConv 通过仅对输入通道的一个子集应用标准卷积，同时保持其余通道不变，从而利用特征图的冗余性。图 11 展示了 PConv 的工作原理。

假设 h、w 分别为输入数据的高度和宽度，c 为输入通道数，c_{p} 为部分输入通道数，k 为卷积核大小，标准卷积和部分卷积的计算量如式 (8)~(10) 所示。

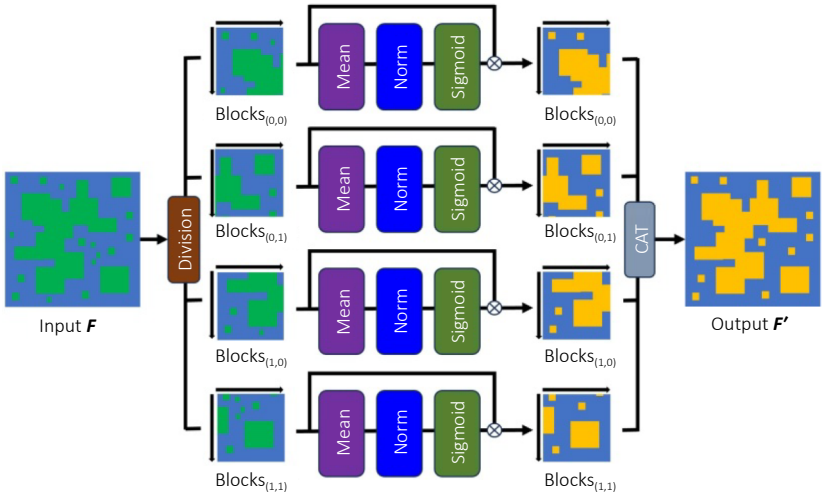


图 9 切分操作原理

Fig. 9 Principle of slicing operation

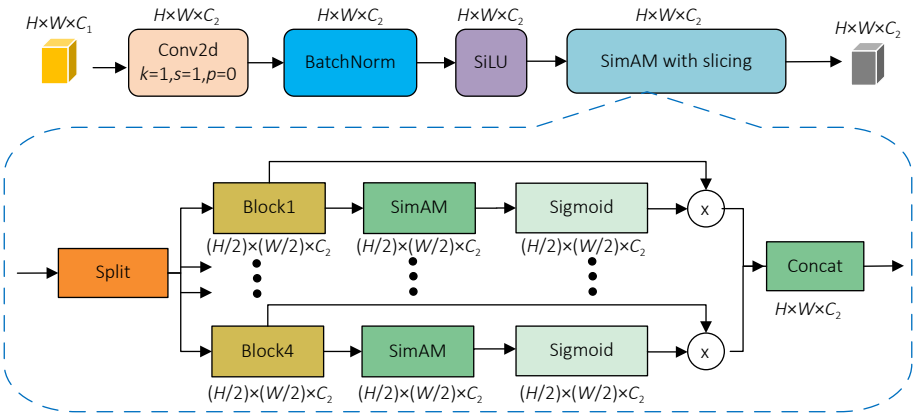


图 10 Conv_SWS 模块结构

Fig. 10 Conv_SWS structure

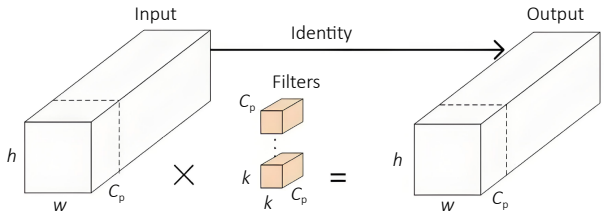


图 11 PConv 结构

Fig. 11 PConv structure

$$FLOP1 = h \times w \times k^2 \times c^2 , \tag{8}$$

$$FLOP2 = h \times w \times k^2 \times c_{\mathrm{p}}^2 , \tag{9}$$

$$\frac{FLOP1}{FLOP2} = \frac{c^2}{c_{\mathrm{p}}^2} . \tag{10}$$

当 c_{p} 取为 c 的数量的通道时，PConv 的浮点计算量为标准卷积的 1/16。因此，PConv 能够有效地提升计算效率和降低算法所需的计算资源。

为进一步提高模型提取特征能力，将 PConv 组合卷积网络，创建 FasterNet 模块，如图 12 所示。为提升模型效率，本文采用 FasterNet Block 替换 C2f 模块中原有的 BottleNeck 模块，同时保持 C2f 的整体结构不变，由此提出一种快速卷积交叉融合模块 (FasterNet block cross stage partial fusion，FBC2f)，其结构如图 13 所示。具体来说，输入特征张量首先通过 1×1 卷积进行处理，使输出特征张量通道的数量翻倍，从而增强模型的特征表示能力。在 FBC2f 模块中，输入特征张量分为两个部分：其中一部分输入至 FasterNet block 进行特征变换，另一部分则直接保留并与处理后的特征进行拼接。经过分割的特征将在多个级联的 FasterNet block 中逐步处理，从

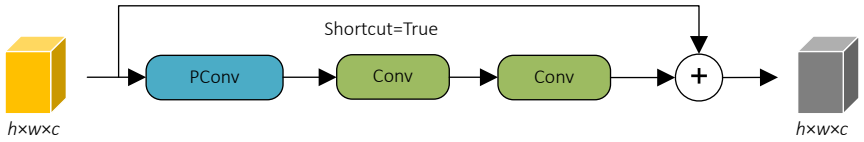


图 12 FasterNet Block 结构

Fig. 12 FasterNet block structure

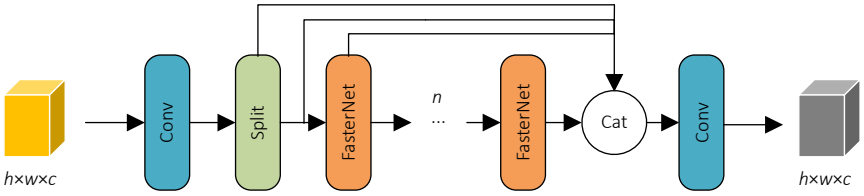


图 13 FBC2f 结构

Fig. 13 FBC2f structure

而有效捕获更深层的特征。集成的 PConv 减少了冗余计算和内存访问，同时保持了有效的空间特征提取。随后，所有 FasterNet Block 的输出与先前分割得到的特征张量进行拼接，以提升特征表达的多样性。最后，通过一个卷积层对拼接后的特征进行通道压缩，使其匹配后续处理所需的输出维度。这种改进不仅保持了高 FLOPs，还显著降低了 FLOPs，有效减少了不必要的计算和内存使用[27]。因此，它减小了模型尺寸并加速了推理时间，更适合轻量化部署。

3.2.3 损失函数的改进

在 PCB 缺陷数据集中，缺陷目标通常尺寸较小、形态多样，且标注质量存在不一致现象，例如部分缺陷定位不精确或边界框标注可靠性较低。原方法采用 CIoU 损失作为边界框回归的损失函数，该函数在 IoU 基础上引入中心点距离惩罚项，以促进预测框与真实框的中心对齐，从而加快收敛速度。然而，该方法对所有样本无论其标注质量高低，均采用统一的损失计算方式。而 IoU 对小目标的位置偏差非常敏感，在基于 Anchor 的检测器中使用时，轻微的位置偏移就会导致标签分配不准，影响检测精度。为了解决这一问题，本文引入了一种基于 Wasserstein 距离的度量标准 NWD (normalized Wasserstein distance)[28] 与标准 IoU 结合，按照一定比例进行加权得到最终的损失函数。由于 NWD 对目标尺度不敏感，并且在目标没有重叠的情况下也能度量相似性，因此更适合测量小目标。NWD 采用二维高斯分布对边界框进行建模，并利用两个高斯分布之间的 Wasserstein 距离来评估边界框之间的相似度，Wasserstein 距离的计算公式为

$$W_2^2(N_a,N_b)=\left\|\left(\left[cx_a,cy_a,\frac{w_a}{2},\frac{h_a}{2}\right]^T,\left[cx_b,cy_b,\frac{w_b}{2},\frac{h_b}{2}\right]^T\right)\right\|_2^2, \tag{11}$$

式中：N_a、N_b 分别代表预测框 A 和真实框 B 的二维高斯分布；(cx_a, cy_a) 和 (cx_b, cy_b) 分别代表预测框 A 和真实框 B 的中心点坐标；w_a 和 h_a 分别代表预测框 A 的宽度和高度；w_b 和 h_b 分别代表真实框 B 的宽度和高度。

对$W_2^2(N_a,N_b)$进行归一化处理，得到 NWD 为

$$NWD(N_a,N_b)=\exp\left(-\frac{\sqrt{W_2^2(N_a,N_b)}}{C}\right), \tag{12}$$

式中：C 是与数据集相关的常数，根据文献 [28] 中的研究结果，实验中将 C 设成数据集中目标的平均绝对尺寸，且在一定范围内，结果都是稳定的。根据图 4 统计的数据集结果，可设定 C 的值为 16。由于$NWD(N_a,N_b)$表示预测框 A 和真实框 B 之间的归一化 Wasserstein 距离，因此损失函数可设为

$$L_{NWD}=1-NWD(N_a,N_b). \tag{13}$$

将 NWD 和 IoU 结合，得到改进后的损失函数计算公式为

$$L_{NWDIoU}=\alpha\times L_{NWD}+(1-\alpha)\times(1-IoU), \tag{14}$$

式中：α为权重系数。调节 NWD 对小目标的优化优先级与 IoU 对大目标的稳定性。经实验验证，本文最终将α设为 0.5。

4 实验结果及分析

4.1 实验环境与评价指标

本文实验使用 Win 10 系统，CPU 是 Intel Core i7-12700KF，GPU 为 NVIDIA GeForce RTX 4060Ti。编程语言是 Python 3.9，采用的深度学习框架是 PyTorch 1.12.0 框架、CUDA 版本是 CUDA 11.6。在模型训练过程中，优化器选用随机梯度下降 (SGD)，初始学习率设为 0.001，动量参数设置为 0.9，权重衰减系数为 0.0005，其余参数保持系统默认配置。输入图像设置为 640 pixel×640 pixel，批处理大小 (batch size) 设置为 16，训练轮数 (epochs) 共计 200 轮。

为了评估算法的检测性能，实验中使用平均精度 (mean average precision，mAP) 作为主要评价指标，由精度 (precision, P)、召回率 (recall, R) 和平均精度 (AP) 计算所得，其计算公式如式 (15)~(18) 所示。此外，为了评估模型的复杂度和计算成本，本文还采用了模型参数量 (Params)、浮点运算量 (FLOPs) 和模型大小 (Weights) 等参数来评价模型的轻量化程度。

$$P=\frac{TP}{TP+FP}, \tag{15}$$

$$R=\frac{TP}{TP+FN}, \tag{16}$$

$$AP=\int_0^1 P(R)\mathrm{d}R, \tag{17}$$

$$mAP=\frac{1}{n}\sum_{i=1}^{n}AP_i, \tag{18}$$

式中：TP 表示正确检测出对应缺陷的预测框的数量；FP 表示检测错误的预测框的数量；FN 表示未能检测出的预测框的数量；AP 表示是 P-R 曲线和坐标轴包围的面积；mAP 表示所有检测类别的 AP 的平均值；AP_i 表示数据集中特定类别的平均精度；n 表示缺陷种类的数量，在本文中，n=5。

4.2 实验结果与分析

4.2.1 数据增强对比实验

采用 YOLOv8n 与 PCB-YOLO 方法对增强前后的 PCB 缺陷数据集进行实验，结果如图 14 所示 (“+”代表增强后的数据集)。实验数据显示，YOLOv8n 在原始数据集上的 mAP_{50} 和 $mAP_{50\text{-}95}$ 分别为 87.0% 与 44.4%；而在数据增强后，这两项指标分别提升至 92.7% 和 59.6%，增幅达到 5.7% 和 15.2%。PCB-YOLO 在增强前实现了 89.3% 的 mAP_{50} 与 44.6% 的 $mAP_{50\text{-}95}$，增强后则进一步提高到 95.6% 和 62.3%，分别上升了 6.3% 和 17.7%。由此可见，数据增强显著提升了两种模型的检测性能，尤其对于本文提出的 PCB-YOLO 方法，在图像数量较多的数据集上展现出更优异的学习能力。

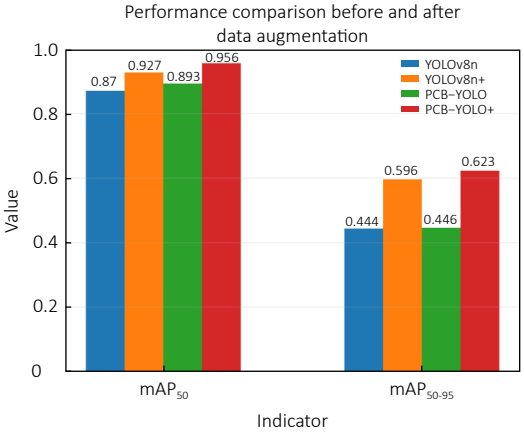


图 14 数据集增强前后不同算法性能对比

Fig. 14 Comparison of algorithm performance before and after dataset augmentation

4.2.2 轻量化对比实验

本实验基于 YOLOv8n 模型，采用了两种方法来减少模型的参数量和计算量。首先，采用主流轻量级特征提取网络，如 GhostNetv2[29]、VanillaNet[30] 用以代替原始模型的标准卷积模块，实现模型的轻量化。其次，采用本文设计的 FBC2f 来代替原始模型的 C2f 模块，实验结果如表 1 所示。

实验结果表明，当引入 VanillaNet 作为特征提取网络时，模型实现了最低的参数量、浮点运算量和模型权重，分别是原始模型的 60.8 %、69.1 % 和 62.7 %，但 mAP_{50} 和 $mAP_{50\text{-}95}$ 显著下降。当使用 FBC2f 作为轻量化方法时，与 GhostNet 和 VanillaNet 相比，mAP_{50} 分别增加了 0.2% 和 1.4%；与原始模型相比，其参数量、浮点运算量和模型权重分别减少了 23.5%、22.2% 和 22.0%。因此，FBC2f 模块更好地平衡了模型轻量化程度和检测精度。

4.2.3 不同注意力机制对比实验

为了验证改进的轻量级注意力机制 (SimAM) 的有效性，对改进 YOLOv8n 中使用的 SimAM 注意力模块进行了替换实验。将 SimAM 分别替换为 ECA (efficient-

表 1 轻量化对比实验结果

Table 1 Light weight comparative experimental results

Model	mAP_{50}/%	$mAP_{50\text{-}95}$/%	Param/M	FLOPs/G	Weight/M
YOLOv8n (baseline)	**92.7**	**59.6**	3.006	8.1	5.9
Baseline + GhostNet	90.3	50.5	2.748	7.5	5.4
Baseline + VanillaNet	89.1	46.7	**1.829**	**5.6**	**3.7**
Baseline + FBC2f	90.5	48.3	2.301	6.3	4.6

channel-attention)[31]、DAT (deformable attention)[32] 注意力机制模块及最初的 SimAM 进行了对比分析，结果详见表 2。

实验结果表明，SimAM 注意力机制和 ECA 注意力机制作比较时，在相似的参数量、浮点运算量和模型权重下，SimAM 的 mAP_{50} 提高了 0.6%。此外，与 DAT 注意力机制作比较，SimAM 将参数量和模型权重分别减少 1.149 M 和 1.800 M。与原始 SimAM 模块相比，改进的 SWS 模块将 mAP_{50} 和 $mAP_{50\text{-}95}$ 分别提高了 1.3% 和 1.8%，达到了 95.1% 和 62.1%。这表明在模型中融合 SimAM 注意力机制和切分方法，更适合多尺度缺陷检测任务。

4.2.4 不同损失函数对比实验

为了验证 Wasserstein 距离损失函数的有效性，在 Wasserstein 距离损失函数 (NWD-IoU) 和其他损失函数之间进行了比较实验，包括 CIoU (complete intersection over union)、EIoU (efficient intersection over union)[33]、Wise-IoU[34] 和 Shape-IoU[35]，实验结果如表 3 所示。

实验结果表明，与原始 CIoU 损失函数相比，NWD-IoU 损失的精度、召回率和 mAP50 分别提高了 3.9%、3.8%、3.4%。与其他损失函数相比，NWD-IoU 实现了最高的精度、召回率和 mAP_{50} 和 $mAP_{50\text{-}95}$。此外在 mAP_{50} 检测性能方面，相比 EIoU、Wise-IoU 和 Shape-IoU 分别提高了 4.8%、1.7% 和 1.1%，这说明 NWD-IoU 损失有利于增强微小目标的检测性能。

4.2.5 损失函数中权重系数灵敏度分析

为验证损失函数中权重系数 α 的最优性，在 PCB 缺陷数据集上设计灵敏度分析实验，将 α 分别设置为 0.1、0.3、0.5、0.7、0.9，实验结果如表 4 所示。

由表 4 可知，当 α<0.5 时，NWD 的权重不足，IoU 的主导作用导致微小缺陷的回归梯度不足，召回率、mAP_{50} 和 $mAP_{50\text{-}95}$ 随 α 的增大而上升，表明 NWD 的加入显著改善了小目标的漏检问题。当 α>0.5 时，NWD 的权重过高，虽对小目标的梯度引导增强，但 IoU 对大目标的约束减弱，导致中等尺寸缺陷的边界框定位精度下降，精度、mAP_{50} 和 $mAP_{50\text{-}95}$ 随 α 的增大而下降，表明过度依赖 NWD 会牺牲大目标的检测稳定性。当 α=0.5 时，模型取得最优综合性能：精度达到 96.3%，

表 2 不同注意力机制对比结果

Table 2 Comparison results of attention mechanisms

Model	mAP_{50}/%	$mAP_{50\text{-}95}$/%	Param/M	FLOPs/G	Weight/M
Baseline + ECA	93.2	61.6	**2.597**	**7.9**	**5.1**
Baseline + DAT	93.5	61.2	3.746	**7.9**	6.9
Baseline + SimAM	93.8	60.3	**2.597**	**7.9**	**5.1**
Baseline + SWS	**95.1**	**62.1**	3.006	8.1	5.9

表 3 不同损失函数对比结果

Table 3 Comparison results of loss function

Model	*P*/%	*R*/%	mAP_{50}/%	$mAP_{50\text{-}95}$/%	Param/M
CIoU	92.4	87.8	92.2	59.3	**2.302**
EIoU	91.5	85.6	90.8	55.6	**2.302**
Wise-IoU	95.8	88.5	93.9	58.7	**2.302**
Shape-IoU	95.4	90.8	94.5	60.5	**2.302**
NWD-IoU	**96.3**	**91.6**	**95.6**	**62.3**	**2.302**

表 4 不同 α 值对比结果

Table 4 Comparison results of α value

α	P/%	R/%	mAP_{50}/%	$mAP_{50\text{-}95}$/%
0.1	94.2	88.9	93.1	55.8
0.3	95.5	90.2	94.5	58.6
0.5	**96.3**	**91.6**	**95.6**	**62.3**
0.7	95.8	90.7	94.8	58.9
0.9	94.5	89.6	93.5	55.4

召回率达到 91.6%，mAP_{50} 达到 95.6%，以及 $mAP_{50\text{-}95}$ 达到 62.3%，证明该取值能较好平衡小目标与大目标的回归需求。

4.2.6 消融实验结果

为了验证每个改进模块在 PCB 缺陷检测中的有效性，以 YOLOv8n 为基线模型，在构建的 PCB 缺陷数据集上进行 6 组消融实验。具体地，M1 实验将主干网络和颈部网络中的 C2f 模块统一替换为 FBC2f 模块；M2 实验则将主干和颈部网络中的下采样模块调整为 Conv_SWS 模块；M3 实验将损失函数替换为 NWD-IoU。通过测试改进模块的不同组合来进行消融实验，验证每项改进的有效性，实验结果见表 5。

在主干和颈部网络融入 FBC2f 模块后，模型在 mAP_{50} 指标上较基线模型 YOLOv8n 降低了 2.2%，但参数量减少了 0.705 M，浮点运算量减少了 1.8 G。使用 Conv_SWS 模块后的方法与基线相比，召回率提高了 5%，达到 92.8%，指标排名第一，mAP_{50} 和 $mAP_{50\text{-}95}$ 分别提高了 2.4% 和 2.5%，参数量、浮点运算量和模型权重保持不变。采用 NWD-IoU 损失函数的方法与基线相比，在参数量、浮点运算量和模型权重保持不变的情况下，mAP_{50} 和 $mAP_{50\text{-}95}$ 分别提高了 2.3% 和 2.8%。结合 FBC2f 和 Conv_SWS 的方法与使用 FBC2f 的方法相比，mAP_{50} 和 $mAP_{50\text{-}95}$ 分别提升 1.7%和 2.8%，与基线模型相比，参数量、浮点运算量和模型权重都有较大幅度下降。融合 FBC2f、Conv_SWS和 NWD_IoU 的方法与基线相比 mAP_{50} 提升 2.9%，指标排名第一，浮点运算量减少 1.8 G，模型权重减少 1.3 M，指标排名第一。

引入 FBC2f 模块后的模型与基线相比，mAP_{50} 下降至 90.5%，但参数量和浮点运算量分别下降了 23.5% 和 22.2%。这表明在网络中使用 FasterNet 模块能显著降低模型复杂度，增加模型轻量化程度。在此基础上进一步引入 Conv_SWS 模块，mAP_{50} 提高至 92.2%，$mAP_{50\text{-}95}$ 提高至 51.1%，并保持了模型的轻量化程度，说明引入的多特征融合注意力机制能有效提取丰富的上下文信息，进而提升模型的多尺度缺陷识别性能。在此基础上进一步引入 NWD_IoU 损失函数后，mAP_{50} 和 $mAP_{50\text{-}95}$ 分别提升至 95.6% 和 62.3%，表明它对于微小目标的检测有较好的提升作用，同时又不影响计算速度。消融实验印证了 PCB-YOLO 模型的有效性，其在检测精度、推理效率与模型复杂度之间取得了良好的均衡性，这一优势为其实际部署与应用奠定了坚实基础。

4.2.7 与其它检测算法性能比较实验

为了进一步验证所提算法的优越性和有效性，使用 Faster R-CNN[36]、YOLOv5n、YOLOv8n、YOLOv10n[37]、YOLOv11n、YOLOv12[38]、最近发表的轻量化模型 PIC2f-YOLO[39] 和所提出的模型进行了比较实验，实验结果如表 6 所示。基于 CPU 和 GPU 的帧率 (frame per second, FPS) 均为将模型转化为.onnx 格式后在实验过程

表 5 消融实验结果

Table 5 Ablation results

Model	P/%	R/%	mAP_{50}/%	$mAP_{50\text{-}95}$/%	Param/M	FLOPs/G	Weight/M
Baseline	93.4	87.8	92.7	59.6	3.006	8.1	5.9
+M1	89.3	89.5	90.5	48.3	**2.301**	**6.3**	**4.6**
+M2	92.8	**92.8**	95.1	62.1	3.006	8.1	5.9
+M3	95.1	90.4	95.0	**62.4**	3.006	8.1	5.9
+M1M2	92.4	87.8	92.2	51.1	2.302	**6.3**	**4.6**
+M1M2M3	**96.3**	91.6	**95.6**	62.3	2.302	**6.3**	**4.6**

中用计算机进行部署后测量所得。

实验结果表明，Faster R-CNN 作为经典的二阶段检测算法，其检测精度最低，推理速度最慢，模型权重最大，不适合实时应用。在 YOLO 系列模型中，PCB-YOLO 方法与 YOLOv5n、YOLOv10n、YOLOv11n 和 YOLOv8n 等模型相比，其在检测精度、召回率、mAP_{50} 和 $mAP_{50\text{-}95}$ 等指标均排名第一，综合表现优于其它模型。与原始的 YOLOv8n 算法相比，PCB-YOLO 算法在精确率、召回率、mAP_{50} 和 $mAP_{50\text{-}95}$ 方面分别提升了 2.9%、3.8%、2.9% 和 2.7%，在参数量、浮点运算量和模型权重指标分别下降了 0.71 M、1.8 G和 1.3 M。对于 FPS 指标，虽然 YOLOv10 和 YOLOv8n分别在 CPU 和 GPU 上获得了最快的推理速度，但其检测精度比改进后的模型下降较多，而改进后的 PCB-YOLO 模型在 GPU 上的推理速度达到了 160 f/s，在 CPU 上也达到了 50 f/s，完全满足实际工业化应用需求，因此，该模型符合轻量化且高精度要求，使其更适合实时检测需求。

4.2.8 可视化对比实验

对 PCB 缺陷数据集的检测效果进行比较，展示了 YOLOv8n 与本文提出的方法在性能上的差异。如图 15(b) 所示，YOLOv8n 在短路缺陷中存在漏检现象，其在工厂对 PCB 高质量生产的实际应用需求中可能难以胜任。相比之下，改进后的 PCB-YOLO 方法能够更精确地识别出有效缺陷区域，显著减少了漏检情况。具体而言，在鼠咬、开路、短路、毛刺和伪铜等缺陷类型上，PCB-YOLO 相比原始 YOLOv8n 表现出更高的检测置信度。因此本文提出的检测方法能更好避免漏检和误检，具有更好的缺陷检测性能。

4.2.9 泛化性验证

为验证 PCB-YOLO 模型的泛化性能，将基线模型、轻量化模型 PIC2f-YOLO 和本文算法在百度飞桨平台上公开的铝片表面工业缺陷检测数据集[40] 和北京大学智

表 6 不同模型对比实验结果

Table 6 Comparison results of different models

Model	*P*/%	*R*/%	mAP_{50}/%	$mAP_{50\text{-}95}$/%	Param/M	FLOPs/G	Weight/M	FPS (CPU)	FPS (GPU)
Faster-RCNN	82.6	83.4	82.3	44.3	41.40	93.6	317.6	8	23
YOLOv5n	92.2	86.3	92.2	51.1	2.50	7.1	5.0	55	227
YOLOv10n	93.7	90.1	93.1	59.9	2.71	8.4	5.5	**58**	225
YOLOv11n	88.6	83.9	87.7	46.8	2.58	**6.3**	5.2	56	216
YOLOv12n	89.1	83.4	88.4	48.0	2.56	**6.3**	5.3	44	177
PIC2f-YOLO	92.0	87.4	92.4	57.0	2.60	10.6	5.2	36	152
YOLOv8n	93.4	87.8	92.7	59.6	3.01	8.1	5.9	55	**229**
PCB-YOLO	**96.3**	**91.6**	**95.6**	**62.3**	**2.30**	**6.3**	**4.6**	50	160

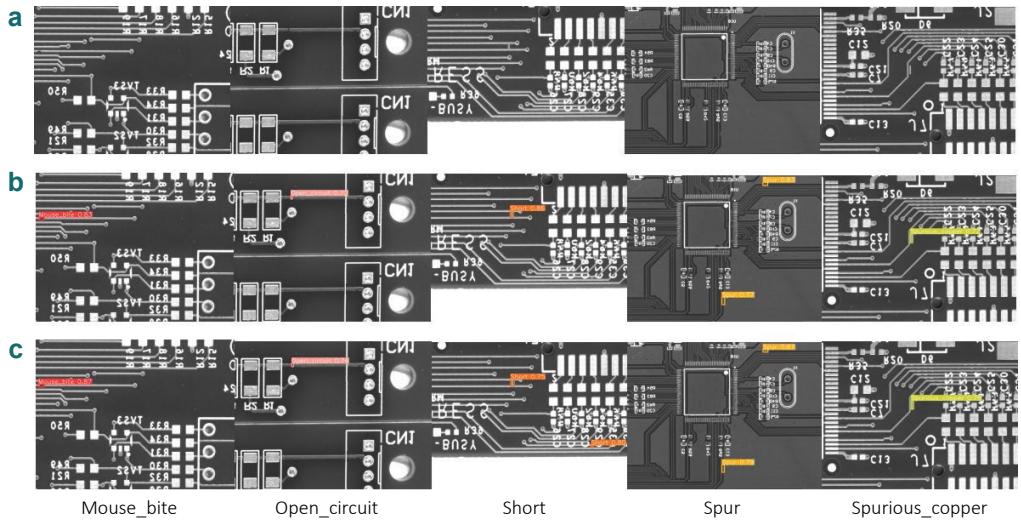


图 15 不同方法在 PCB 缺陷数据集下的检测实例。(a) 原始图像；(b) YOLOv8n；(c) PCB-YOLO

Fig. 15 Detection instance of different methods under the PCB datasets. (a) Original images; (b) YOLOv8n; (c) PCB-YOLO

能机器人开放实验室的开源 PKU-Market-PCB 缺陷数据集[41]上进行实验对比。其中铝片表面工业缺陷检测数据集共有 412 张图片，有针孔、擦伤、脏污、褶皱 4 种缺陷类型，包含小中大型不同尺寸的缺陷类型。PKU-Market-PCB 缺陷数据集包含 693 张图片，共 6 种缺陷类型，且以中小型缺陷为主，具有很大挑战性。实验结果如表 7 和表 8 所示，在铝片表面工业缺陷数据集上，PCB-YOLO 模型的 mAP50达到 98.1%, 相比基线模型和 PIC2f-YOLO 模型分别提高了 0.8% 和 3.1%，同时模型参数量分别减少了 0.71 M 和 0.30 M。在 PKU-Market-PCB 数据集上，其 mAP50 也提升至 93.6%，较基线算法和 PIC2f-YOLO 方法分别提高了 2.7% 和 1.2%，参数量同样分别减少 0.71 M 和 0.30 M。

表 7　铝片表面工业缺陷数据集泛化实验
Table 7　Generalization experiment results on the aluminum sheet surface industrial defect dataset

Model	mAP_{50}/%	Param/M	Weight/M
YOLOv8n	97.3	3.01	6.1
PIC2f-YOLO	95.0	2.60	5.2
PCB-YOLO	98.1	2.30	4.6

表 8　PKU-Market-PCB 数据集泛化实验
Table 8　Generalization experiment results on the PKU-Market-PCB dataset

Model	mAP_{50}/%	Param/M	Weight/M
YOLOv8n	90.9	3.01	4.6
PIC2f-YOLO	92.4	2.60	5.2
PCB-YOLO	93.6	2.30	4.6

上述结果说明，PCB-YOLO 模型在不同类型和缺陷特点的数据集上均表现出良好的检测性能与稳定的泛化能力。

5　结　论

现有 PCB 缺陷检测中因数据样本少、底纹差异大、相机拍摄图像尺寸大，导致识别准确率不高、误检与漏检发生频次较高，以及处理速度慢等局限性。为此，本文设计了 PCB-YOLO 模型，能显著提升检测性能，有效应对以上挑战。通过图像灰度化和切片操作，显著提高微小目标缺陷的检测精度；将改进的 SimAM 注意力机制集成到主干卷积模块中，优化特征间的交互和整合，进一步提高模型对不同尺寸大小的目标感知能力；设计的快速卷积交叉融合模块能在有效提取目标缺陷特征的同时，大幅降低模型的计算量，达到轻量化的目的；最后通过引入 NWD 作为回归损失函数替换 CIoU 损失函数，提升模型的效率和准确性。实验结果表明，PCB-YOLO 模型有效地应对了平衡模型检测精度和轻量级设计的挑战，并且具有较好的泛化性能。

作者贡献

邓文斌：数据集的收集、分析，论文写作；郭怡希：程序编写、实验；林源强：数据预处理；鹿业波：论文修改、提供实验条件；杨俊：指导论文选题、论文修改。

利益冲突

所有作者声明无利益冲突

参考文献

1. Wu Y Q, Zhao L Y, Yuan Y B, et al. Research status and the prospect of PCB defect detection algorithm based on machine vision[J]. *Chin J Sci Instrum*, 2022, **43**(8): 1−17.
吴一全, 赵朗月, 苑玉彬, 等. 基于机器视觉的 PCB 缺陷检测算法研究现状及展望[J]. 仪器仪表学报, 2022, **43**(8): 1−17.
2. Dong J Y, Lü W T, Bao X M, et al. Research progress of the PCB surface defect detection method based on machine vision[J]. *J Zhejiang Sci-Tech Univ*, 2021, **45**(3): 379−389.
董静毅, 吕文涛, 包晓敏, 等. 基于机器视觉的 PCB 板表面缺陷检测方法研究进展[J]. 浙江理工大学学报, 2021, **45**(3): 379−389.
3. Ebayyeh A A R M A, Mousavi A. A review and analysis of automatic optical inspection and quality monitoring methods in electronics industry[J]. *IEEE Access*, 2020, **8**: 183192−183271.
4. Tsai D M, Huang C K. Defect detection in electronic surfaces using template-based Fourier image reconstruction[J]. *IEEE Trans Compon Package Manuf Technol*, 2019, **9**(1): 163−172.
5. Hassanin A A I M, Abd El-Samie F E, El Banby G M. A real-time approach for automatic defect detection from PCBs based on SURF features and morphological operations[J]. *Multimed Tools Appl*, 2019, **78**(24): 34437−34457.
6. Tsai D M, Lin B T. Defect detection of gold-plated surfaces on PCBs using entropy measures[J]. *Int J Adv Manuf Technol*, 2002, **20**(6): 420−428.
7. Lu Z S, He Q Q, Xiang X G, et al. Defect detection of PCB based on Bayes feature fusion[J]. *J Eng*, 2018, **2018**(16): 1741−1745.
8. Hu B, Wang J H. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network[J]. *IEEE Access*, 2020, **8**: 108335−108345.
9. Zeng N Y, Wu P S, Wang Z D, et al. A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection[J]. *IEEE Trans Instrum Meas*, 2022, **71**: 3507014.
10. Sun Z C, Wang B, Zhang X L. PCB defect detection based on deformable residual convolution and scalable feature pyramid algorithm[J]. *Telecommun Eng*, 2023, **63**(6): 798−805.
孙志超, 王博, 张晓玲. 基于可变形残差卷积与伸缩式特征金字塔算法的 PCB 缺陷检测[J]. 电讯技术, 2023, **63**(6): 798−805.
11. Redmon J, Farhadi A. YOLOv3: an incremental improvement[Z]. arXiv: 1804.02767, 2018.
https://doi.org/10.48550/arXiv.1804.02767.
12. Ultralytics. YOLOv5[EB/OL]. (2025-08-25)
https://github.com/ultralytics/YOLOv5.
13. 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]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 7464–7475. https://doi.org/10.1109/CVPR52729.2023.00721.
14. Ultralytics[EB/OL]. (2025-08-25) GitHub, 2023. https://github.com/ultralytics/ultralytics.
15. He G Z, Liang Y. PCB defect detection based on convolutional neural network[J]. *J Graphics*, 2022, **43**(1): 21−27.
何国忠, 梁宇. 基于卷积神经网络的 PCB 缺陷检测[J]. 图学学报, 2022, **43**(1): 21−27.
16. Xia K W, Lv Z L, Liu K. et al. Global contextual attention augmented YOLO with ConvMixer prediction heads for PCB surface defect detection[J]. *Sci Rep*, 2023, **13**(1): 9805.
17. Li Z H, Zhang L. Lightweight PCB defect detection algorithm of Improved YOLOv5[J]. *Radio Eng*, 2023, **53**(6): 1342−1350.
李振华, 张雷. 改进 YOLOv5 的轻量级 PCB 缺陷检测算法[J]. 无线电工程, 2023, **53**(6): 1342−1350.
18. Bhattacharya A, Cloutier S G. End-to-end deep learning framework for printed circuit board manufacturing defect classification[J]. *Sci Rep*, 2022, **12**(1): 12559.
19. Lei L, Li H X, Yang H D. Reliable and lightweight adaptive convolution network for PCB surface defect detection[J]. *IEEE Trans Instrum Meas*, 2024, **73**: 2003208.
20. Gonzalez R C, Wood R E. *Digital Image Processing*[M]. Ruan Q Q, trans. 4th ed. Beijing: Electronics Industry Press, 2020.
Gonzalez R C, Wood R E. 数字图像处理[M]. 阮秋琦, 译. 4 版. 北京: 电子工业出版社, 2020.
21. Tong K, Wu Y Q, Zhou F. Recent advances in small object detection based on deep learning: a review[J]. *Image Vision Comput*, 2020, **97**: 103910.
22. Qiu L L, Zhu W G, Li Y G, et al. YOLO-based multi-class target detection for SAR images[C]//*Proceedings of 2023 IEEE 7th Information Technology and Mechatronics Engineering Conference (ITOEC)*, 2023: 1671–1675. https://doi.org/10.1109/ITOEC57671.2023.10292083.
23. Jooshin H K, Nangir M, Seyedarabi H. Inception-YOLO: computational cost and accuracy improvement of the YOLOv5 model based on employing modified CSP, SPPF, and inception modules[J]. *IET Image Process*, 2024, **18**(8): 1985−1999.
24. Zhang S F, Chi C, Yao Y Q, et al. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection[C]//*Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020: 9756–9765. https://doi.org/10.1109/CVPR42600.2020.00978.
25. Yang L X, Zhang R Y, Li L D, et al. SimAM: a simple, parameter-free attention module for convolutional neural networks[C]//*Proceedings of the 38th International Conference on Machine Learning (ICML)*, 2021: 11863–11874.
26. Chen J R, Kao S H, He H, et al. Run, don't walk: chasing higher FLOPS for faster neural networks[C]//*Proceedings of 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023: 12021–12031. https://doi.org/10.1109/CVPR52729.2023.01157.
27. Wang H R, Cheng H, Zhang J T. Faster-PGYOLO: an efficient framework for floating debris detection in inland waters[J]. *Vis Comput*, 2025, **41**(7): 5087−5104.
28. Wang J W, Xu C, Yang W, et al. A normalized gaussian Wasserstein distance for tiny object detection[Z]. arXiv: 2110.13389, 2021. https://doi.org/10.48550/arXiv.2110.13389.
29. Tang Y H, Han K, Guo J Y, et al. GhostNetV2: enhance cheap operation with long-range attention[C]//*Proceedings of the 36th Conference on Neural Information Processing Systems (NeurIPS)*, 2022: 724.
30. Chen H T, Wang Y H, Guo J Y, et al. VanillaNet: the power of minimalism in deep learning[C]//*Proceedings of the 37th International Conference on Neural Information Processing Systems*, 2023: 308.
31. Wang Q L, Wu B G, Zhu P F, et al. ECA-Net: efficient channel attention for deep convolutional neural networks[C]//*Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020: 11531–11539. https://doi.org/10.1109/CVPR42600.2020.01155.
32. Xia Z F, Pan X R, Song S J, et al. Vision transformer with deformable attention[C]//*Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022: 4784–4793. https://doi.org/10.1109/CVPR52688.2022.00475.
33. Zhang Y F, Ren W Q, Zhang Z, et al. Focal and efficient IOU loss for accurate bounding box regression[J]. *Neurocomputing*, 2022, **506**: 146−157.
34. Tong Z J, Chen Y H, Xu Z W, et al. Wise-IoU: bounding box regression loss with dynamic focusing mechanism[Z]. arXiv: 2301.10051, 2023. https://doi.org/10.48550/arXiv.2301.10051.
35. Zhang H, Zhang S J. Shape-IoU: more accurate metric considering bounding box shape and scale[Z]. arXiv: 2312.17663, 2023. https://doi.org/10.48550/arXiv.2312.17663.
36. Ren S Q, He K M, Girshick R, et al. Faster R-CNN: towards real-time object detection with region proposal networks[J]. *IEEE Trans Pattern Anal Mach Intell*, 2017, **39**(6): 1137−1149.
37. Wang A, Chen H, Liu L H, et al. YOLOv10: real-time end-to-end object detection[C]//*Proceedings of the 38th International Conference on Neural Information Processing Systems*, 2024: 3429.
38. Tian Y J, Ye Q X, Doermann D. YOLOv12: attention-centric real-time object detectors[Z]. arXiv: 2502.12524, 2025. https://doi.org/10.48550/arXiv.2502.12524.
39. Hu Y L, Yang J, Xu C Y, et al. PIC2f-YOLO: a lightweight method for the detection of metal surface defects[J]. *Opto-Electron Eng*, 2025, **52**(1): 240250.
胡依伦, 杨俊, 许聪源, 等. PIC2f-YOLO: 金属表面缺陷检测轻量化方法[J]. 光电工程, 2025, **52**(1): 240250.
40. AI Studio. Surface defect detection for aluminum sheets in industrial applications[EB/OL]. (2025-08-25) https://aistudio.baidu.com/datasetdetail/135640.
AI Studio. 铝片表面工业缺陷目标检测[EB/OL]. https://aistudio.baidu.com/datasetdetail/135640.
41. Ding R W, Dai L H, Li G P, et al. TDD-net: a tiny defect detection network for printed circuit boards[J]. *CAAI Trans Intell Technol*, 2019, **4**(2): 110−116.

作者简介



邓文斌 (1985-)，男，工学博士，嘉兴大学机械工程学院讲师，研究方向为智能装备开发、机器视觉。

E-mail：wbdeng@zjxu.edu.cn



【通信作者】杨俊 (1978-)，男，工学博士，嘉兴大学人工智能学院副教授，研究方向为智能多媒体信息分析与处理、机器学习、计算机视觉。

E-mail：juneryoung@zjxu.edu.cn

PCB-YOLO: an enhanced YOLOv8n for lightweight detection of tiny defects in PCB

Deng Wenbin[1], Guo Yixi[1], Lin Yuanqiang[2], Lu Yebo[1], Yang Jun[2*]

[1] College of Mechanical Engineering, Jiaxing University, Jiaxing, Zhejiang 314001, China;
[2] College of Artificial Intelligence, Jiaxing University, Jiaxing, Zhejiang 314001, China

Abstract

Objective In the production process of printed circuit boards (PCBs), common defects such as mouse_bite, open_circuit, short, spur and spurious_copper are frequently encountered. These defects are characterized by their small size, color similarity to the background, and similar features across different defect types, which make detection challenging and prone to false positives and missed detections. Traditional inspection methods, including manual visual checks and conventional automated optical inspection (AOI), suffer from high error rates, limited adaptability, and dependency on manually selected features. Although deep learning approaches, particularly YOLO-series models, have improved detection robustness, they still face three core limitations in PCB defect inspection: insufficient feature extraction for tiny objects, poor balance between model lightweight and accuracy, and inadequate loss function design for small targets. To address the issues of limited samples, low efficiency, large parameter size, and low accuracy in detecting tiny-defects on PCBs, this paper proposes a lightweight detection model named PCB-YOLO based on the YOLOv8n.

Methods First, the original high-resolution PCB images with diverse background colors are converted to grayscale to minimize color interference. To avoid losing tiny defects during resizing, a sliding-window slicing strategy is applied, generating 640 pixel×480 pixel sub-images with 100 pixels overlaps. Only slices containing defects are retained, effectively increasing the usable dataset while preserving small-object details. Second, considering the significant proportion of small target defects in PCBs, a Conv_SWS (SimAM with slicing) module is designed by integrating the SimAM attention mechanism and merging slicing operations. The input feature map is divided into local patches, each enhanced by SimAM to emphasize small-defect regions, before being recombined. This module replaces standard convolution in downsampling stages, enhancing multi-scale feature extraction without adding parameters. Third, the original C2f module in YOLOv8 is modified by introducing partial convolution (PConv) to comprehensively compute key feature channels, resulting in the design of the FBC2f module. This enhancement improves feature extraction capability while reducing the number of model parameters. Finally, the standard IoU loss is sensitive to minor localization errors and fails when boxes do not overlap. Thus, a combined loss using normalized Wasserstein distance (NWD) and IoU is proposed. NWD models bounding boxes as 2D Gaussian distributions and measures similarity via Wasserstein distance, providing stable gradients even for non-overlapping tiny objects. The final loss is a weighted sum of NWD and IoU, optimized with a balancing factor α=0.5. The model is trained and evaluated on an augmented PCB defect dataset containing 5172 images with 6282 defect instances. Performance is measured using mAP50, mAP50-95, parameters, FLOPs, and frame per second (FPS). Comparative experiments include ablation studies, attention mechanism comparisons, loss function analysis, and benchmarking against state-of-the-art detectors.

Results and Discussions Experimental results demonstrate the effectiveness of PCB-YOLO: it achieves 95.6% mAP50 on the PCB defect dataset, outperforming YOLOv8n by 2.9%. It also reduces parameters by 23.6% (from 3.006 M to 2.302 M), FLOPs by 22.2% (from 8.1 G to 6.3 G), and model weight by 1.3 M. Moreover, ablation studies are finished to evaluate each proposed component contributes positively: the Conv_SWS module boosts recall by 5% and mAP50 by 2.4%; the FBC2f module reduces parameters by 23.5% with a moderate mAP50 trade-off. the NWD-IoU loss increases mAP50 by 2.3% and mAP50-95 by 2.8%. The full integration of all three modules achieves the best balance of accuracy and efficiency. Comparative evaluations demonstrate that PCB-YOLO outperforms Faster R-CNN, YOLOv5n, YOLOv10n, YOLOv11n, YOLOv12n, and the recently introduced lightweight PIC2f-YOLO in terms of both accuracy and model size. Additionally, it achieves a real-time inference speed of 160 f/s on GPU and 50 f/s on CPU, meeting the throughput demands of industrial applications. Moreover, PCB-YOLO also demonstrates strong generalization on external datasets, attaining mAP50 scores of 98.1% on the industrial aluminum surface defects dataset and 93.6% on the PKU-Market-PCB dataset, thereby verifying its robustness and transferability. Visual comparisons reveal that PCB-YOLO substantially decreases the rates of missed detections and false alarms relative to YOLOv8n, especially in the case of minor defects like shorts and mouse bites. In summary, the results validate that the

Foundation item: "Pioneer" and "Leading Goose" R&D Program of Zhejiang (2024C04028), Scientific Research Fund of Zhejiang Provincial Education Department (Y202455539), Pinghu Municipal Science and Technology Plan Project (Industrial Development Research Special Project 202505), and Jiaxing University "Intelligence Convergence in Pinghu: Jiaxing Talents in Enterprises" Research Project (00324264AL)

*Correspondence: Yang J, E-mail: juneryoung@zjxu.edu.cn

Conv_SWS module effectively preserves small-object features during downsampling, the FBC2f module maintains representational capacity while reducing computation, and the NWD-IoU loss stabilizes training and improves localization for miniature targets.

Conclusions This paper proposes PCB-YOLO, a lightweight and accurate detection model designed for tiny PCB defects. The approach integrates grayscale transformation and adaptive slicing during preprocessing to enhance input data quality. It introduces the Conv_SWS module to improve small-object feature retention and designs the parameter-efficient FBC2f module. Additionally, the NWD-IoU loss is adopted to refine the regression of minor defects. Together, these contributions address key challenges in PCB inspection, including limited sample availability, background variation, small defect size, and the accuracy–complexity trade-off. Experimental results show that PCB-YOLO achieves higher detection accuracy with fewer parameters and faster inference than current leading models. It also demonstrates strong generalization across diverse industrial defect datasets. The framework offers a practical, deployable solution for real-time, high-precision PCB quality inspection in manufacturing. Future work may explore further architectural optimizations and extend the method to other micro-defect detection domains.

Keywords: PCB detection; tiny object detection; YOLOv8n; evaluation metrics; attention mechanism

Deng W B, Guo Y X, Lin Y Q et al. PCB-YOLO: an enhanced YOLOv8n for lightweight detection of tiny defects in PCB[J]. *Opto-Electron Eng*, 2026, **53**(2): 250252; DOI: 10.12086/oee.2026.250252

Opto-Electronic Journals Group
www.oejournal.org

扫描二维码，获取 PDF 全文