

# 基于 YOLOv8 的 PCB 缺陷检测改进算法

刘爽<sup>1</sup>, 吕俊良<sup>1</sup>, 秦宇航<sup>1</sup>, 秦丹丹<sup>2</sup>, 孙佳慧<sup>2</sup>

(1. 吉林大学 数学学院, 长春 130012; 2. 空军航空大学, 长春 130022)

**摘要:** 针对工业印刷电路板缺陷检测任务中, 小目标特征不明显且检测精度不足的问题, 提出一种基于 YOLOv8 算法的改进算法。首先, 通过增删特征图尺寸以适应印刷电路板缺陷检测, 并借鉴加权双向特征金字塔网络结构保留原始图像的特征; 其次, 利用分组卷积在颈部设计一个轻量化模块进行特征提取, 提高检测精度的同时降低了模型复杂度; 最后, 在小目标检测头前引入可增强特征表现能力的坐标注意力模块, 进一步提高检验精度。实验结果表明, 改进后的算法能将检测精度 mAP@0.5 提升至 95.4%, 并使检测速度 FPS(帧每秒) 达到 105.4, 可以更好地满足工业检测对精度和实时性的要求。

**关键词:** 印刷电路板缺陷检测; 神经网络; 注意力机制; 分组卷积

**中图分类号:** TP391.41 **文献标志码:** A **文章编号:** 1671-5489(2026)02-0344-07

## Improved Algorithm of PCB Defect Detection Based on YOLOv8

LIU Shuang<sup>1</sup>, LÜ Junliang<sup>1</sup>, QIN Yuhang<sup>1</sup>, QIN Dandan<sup>2</sup>, SUN Jiahui<sup>2</sup>

(1. College of Mathematics, Jilin University, Changchun 130012, China;

2. Aviation University of Air Force, Changchun 130022, China)

**Abstract:** Aiming at the problem of unclear small-target features and insufficient detection accuracy in the defect detection task of industrial printed circuit boards, we proposed an improved algorithm based on YOLOv8 algorithm. Firstly, we adapted the defect detection of printed circuit boards by adding or deleting the feature map sizes, and drew on the experience of the weighted bi-directional feature pyramid network structure to retain the features of the original image. Secondly, we designed a lightweight module at the neck for feature extraction by using group convolution, which improved the detection accuracy while reducing the complexity of the model. Finally, before the small-object detection head, we introduced coordinate attention module that could enhance feature representation capabilities, further improving the inspection accuracy. The experimental results show that the improved algorithm can improve the detection accuracy mAP@0.5 to 95.4% and achieve a detection speed FPS (frames per second) of 105.4, which can better meet the requirements of industrial detection for accuracy and real-time performance.

**Keywords:** defect detection of printed circuit board; neural network; attention mechanism; group convolution

收稿日期: 2024-12-02.

**第一作者简介:** 刘爽(1999—), 女, 汉族, 硕士研究生, 从事机器学习图像处理的研究, E-mail: lshuang1230@126.com. **通信作者简介:** 吕俊良(1979—), 男, 汉族, 博士, 教授, 博士生导师, 从事数学物理反问题和机器学习的研究, E-mail: lvjl@jlu.edu.cn.

**基金项目:** 国家自然科学基金(批准号: 12271209).

印刷电路板 (printed circuit board, PCB) 是电子元器件的基本组成部分, 在各种电子产品中应用广泛。作为电子产品的核心组成部分, PCB 的质量直接影响电子产品的性能和可靠性。特别是在对精度和稳定性要求极高的精密科技领域, PCB 的质量标准更提升到一个新高度, 任何微小的缺陷都可能导致电子设备的性能下降甚至使整个系统的功能失效。因此, 不断探索和优化 PCB 缺陷检测方法, 对推动电子制造业的持续发展, 促进技术创新与产业升级具重大意义。目前已有许多针对 PCB 缺陷检测的方法<sup>[1-3]</sup>, 主要分为两类, 第一类是采用传统图像处理 (有时也结合传统的机器学习方法) 方法进行检测, 传统图像处理方法涵盖了诸如特征提取与模板匹配等多种方法。如文献[1]先将电路板划分为子区域并对其进行特征提取, 然后利用 AdaBoost 算法对特征进行选择, 最后用 CART 分类的回归树确定最终类型。文献[2]通过引入多方向 Sobel 算子改进模板匹配方法, 更精确地实现待测图像与模板图像之间的匹配, 并在此基础上设计了一种基于图像差分法的多级二值化检测流程实现 PCB 板缺陷检测。第二类是基于深度学习的检测方法<sup>[4-8]</sup>, 主要包括单阶段检测算法和两阶段检测算法两种。单阶段检测算法是一种直接在单个网络阶段中完成缺陷检测任务的方法, 其特点是检测速度快, 能实时进行缺陷检测, 同时保持较高的检测精度。两阶段检测算法则将目标检测任务分为两个独立阶段进行: 第一阶段主要负责生成可能包含目标的候选区域; 第二阶段则对这些候选区域进行精细的分类和位置调整。两阶段检测算法的特点是能准确地检测目标, 但相对于单阶段检测算法, 其推理速度可能较慢, 而许多实际问题常对实时性要求较高。

YOLO (you only look once) 目标检测算法是单阶段检测算法的典型代表, 它对目标检测的实时性效果较好。YOLO 算法的主要思想是将目标检测任务转化为一个简单的回归问题, 通过一次前向传播过程即能高效输出所有目标的检测结果。具体步骤是: 先将输入图像划分为网格单元, 每个网格专门负责预测中心点落在该网格区域内的目标; 然后令每个网格预测多个边界框及其对应的置信度, 以及这些边界框所属类别的概率; 最后通过非极大值抑制策略筛选并剔除重叠的边界框, 得到最终的检测结果。目前, 已有许多针对 YOLO 系列算法在不同场景下应用的改进研究。文献[4]通过 YOLOv3 引入 SeNet (squeeze-and-excitation network) 注意力结构将特征层的特征权重校准为特征权值以强化特征, 提高了果实识别的精确度。文献[9]在 YOLOv7 的基础上设计了一种轻量级且具有自适应性的主干网络架构, 并在颈部网络中引入了非参数注意力模块, 强化了关键特征信息的融合与利用。

尽管目前已有许多用 YOLO 系列算法对 PCB 进行缺陷检测的方法, 但由于工业上对缺陷检测的精度、复杂度、实时性等方面的要求越来越高, 因此这类算法仍需进一步改进。本文基于 YOLOv8, 首先, 通过引入双向特征金字塔结构更好地融合原始图像特征; 其次, 在颈部网络引入分组卷积, 更改原有的 C2f 结构, 在提高原有精度的同时降低了模型的参数量及计算开销; 最后, 为进一步提高检验精度, 在检测头前引入了坐标注意力 (CA) 模块, 增强小目标的检验效果。

## 1 YOLOv8 结构

本文基于 YOLOv8 结构实现对 PCB 缺陷的检测。YOLOv8 网络结构主要由三部分组成: 骨干网络、颈部网络和检测头模块, 其网络结构如图 1 所示。主干网络负责提取图像特征, YOLOv8 结构是 YOLOv5 结构的改进版本, 其参考 CSPDarkNet 的主干网络结构设计了 C2f 结构, 通过优化梯度流动增强了模型性能。YOLOv8 结构的空金字塔池化 (SPPF) 模块将不同尺度的特征图混合在一起, 进一步增强了结构的特征提取能力。颈部网络采用 PAN-FPN (金字塔网络-路径聚合网络) 的结构, 通过自底向上的路径和自顶向下的路径将不同尺度的特征图进行融合, 实现了信息的跨尺度传递, 增强了模型的多尺度检测能力, 能更有效地检测不同尺度的目标。最后, 检测头模块输出最终的缺陷分类结果, 每个检测层都包含回归分支和分类分支, 分别用于预测目标的边界框和类别。

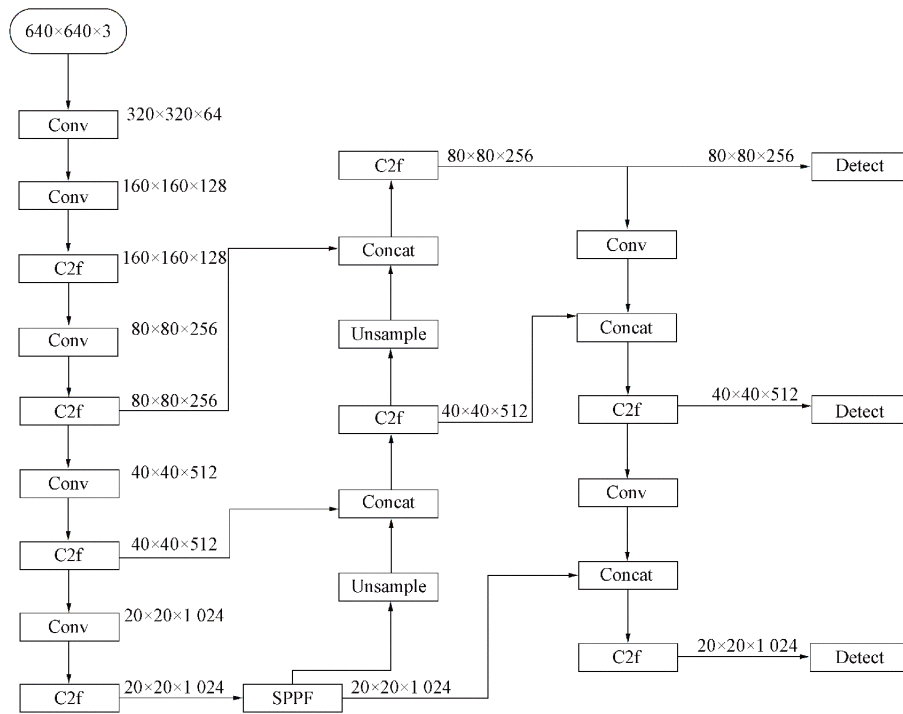


图 1 YOLOv8 网络结构

Fig. 1 YOLOv8 network structure

## 2 加权双向特征金字塔网络

当使用标准的 YOLOv8 结构进行特征提取和特征融合时, 常见的问题是原始图像的细节信息易丢失, 这对最终的检测结果会产生不利影响. 此外, 在 PCB 缺陷检测任务中, 常包含一些较小目标, 而原有的 YOLOv8 结构对这些小目标的检测精度通常较低. 为有效解决这些难题, 本文采用双向特征金字塔网络结构的设计理念, 在原始图像的特征提取阶段, 增加一个横向连接, 将原始特征直接引入到特征融合过程中.

双向特征金字塔网络 (bi-directional feature pyramid network, BiFPN)<sup>[10]</sup> 结构如图 2 所示. 由图 2 可见, 它在原有的自顶向下和自底向上结构基础上引入了一个横向连接, 在不同层级的特征图之间进行信息传递, 有助于确保不同尺度特征之间的信息共享, 提高目标检测的鲁棒性. BiFPN 结构尤其适用于检测小目标或模糊目标、多尺度目标以及对目标进行高精度定位的场景.

BiFPN 加权特征融合时有 3 种方案, 第一种方案是原始的加权特征融合, 计算公式为

$$O = \sum_i w_i I_i,$$

其中  $w_i$  为权重,  $I_i$  为特征图矩阵. 该方法的缺点是对权重没有约束, 会导致训练过程中不稳定, 难以收敛. 第二种方案是基于 Softmax 的特征融合, 计算公式为

$$O = \sum_i \frac{e^{w_i}}{\sum_j e^{w_j}} I_i.$$

该方法的缺点是指数运算会导致训练速度较慢. 第三种方案是快速归一化特征融合, 计算公式为

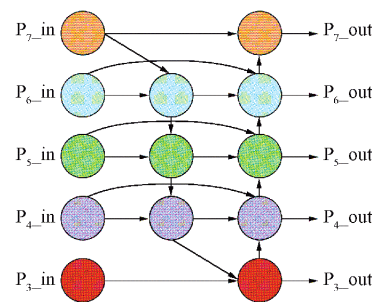


图 2 BiFPN 结构

Fig. 2 BiFPN structure

$$O = \sum_i \frac{\omega_i}{\epsilon + \sum_j \omega_j} I_i,$$

其中  $\epsilon=0.001$ , 以防止分母为 0. 该方法收敛速度较快, 泛化能力较强, 但该计算方式对权重的数值变化较敏感, 一旦某个权重的数值发生明显变化, 即会导致所有权重的占比重新调整, 从而会对网络的收敛速度和结果产生不利影响. 因此, 本文采用第三种方案并对其进行改进, 构造一种由二范数定义的加权融合方式, 计算公式为

$$O = \sum_i \frac{\omega_i}{\epsilon + \sqrt{\sum_j \omega_j^2}} I_i.$$

原始的 YOLOv8 模型结构输出的特征图尺寸分别为  $80 \times 80, 40 \times 40$  和  $20 \times 20$ . 这些特征图中, 尺寸较小的特征图有更大的感受野, 表明它们所对应的原始图像中小目标信息更易被忽略, 从而导致对小缺陷的检测效果不佳.

为提升对小目标的检测能力, 本文在原 YOLOv8 模型结构上引入一个  $160 \times 160$  的小目标检测头, 并剔除原有的  $20 \times 20$  尺寸的检测头, 以捕捉更多关于小目标的细节信息. 从而通过结合不同尺度的特征图, 可更全面地捕捉图像中的信息, 进而提高对小缺陷的检测精度.

基于上述分析, 本文设计了改进的加权金字塔结构, 如图 3 所示. 首先, 调整特征提取过程中保留的特征图尺寸, 以适应 PCB 缺陷检测的小目标任务; 其次, 在得到  $80 \times 80$  尺寸的特征图过程中, 为更好地保留原始图像的特征, 特别加入了直接从原始图像中提取的特征(图 3 中红色线)进行融合, 使最后进行检测的特征图中包含更多的信息, 更有利于缺陷的检测识别.

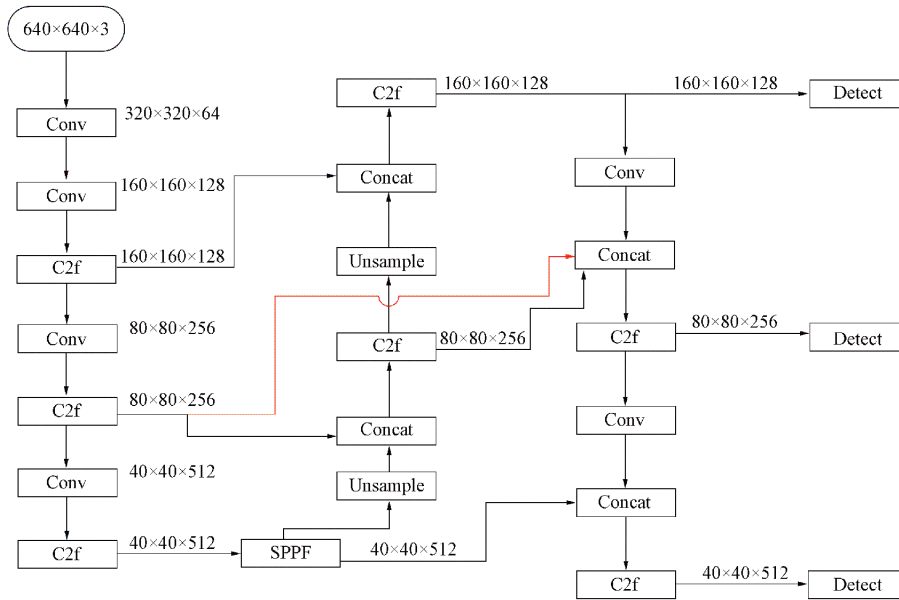


图 3 改进后的加权金字塔结构  
Fig. 3 Improved weighted pyramid structure

### 3 分组卷积结构

C2f 模块在 YOLOv8 中负责特征提取和融合, 它的结构较复杂, 包含多个卷积层和 Bottleneck 模块, 导致 C2f 模块的计算量较大. 此外, 在一些小目标检测任务中, C2f 模块无法充分提取和融合有效的特征信息, 导致模型对小目标的检测性能不佳.

为提高检测精度, 并将模型轻量化, 本文基于 C2f 结构利用分组卷积(group convolution)<sup>[11]</sup>设计一种结构, 并将其命名为 C2fGC, 如图 4 所示. C2fGC 的核心组件 GCBottleneck 如图 5 所示. 该组件的构成主要是通过传统 Bottleneck 中的标准卷积替换为分组卷积. 分组卷积本质上是先将输入通道

和输出通道分成多个组,然后在每组内分别进行卷积操作.分组卷积的优点是不同组可学习到不同的特征表示,从而在一定程度上增强模型的表达能力,使网络能捕获并整合更丰富的特征信息.同时,分组卷积也可以使每组内使用的卷积核维度变小,从而降低了模型的参数量,因此对模型的轻量化效果较好.

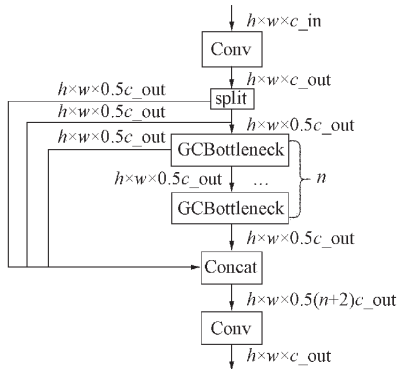


图 4 C2fGC 结构

Fig. 4 C2fGC structure

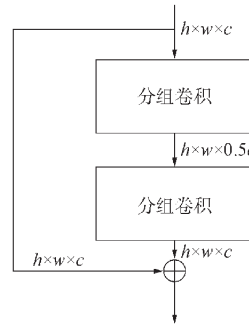


图 5 GC Bottleneck 结构

Fig. 5 GC Bottleneck structure

### 4 坐标注意力机制

标准 YOLOv8 算法识别小目标效果不佳的另一个原因是模型对数据的信息提取不充分.为进一步提高小目标识别的准确率,本文引入坐标注意力(coordinate attention, CA)<sup>[12]</sup>结构,如图 6 所示. CA 是一种为轻量级网络设计的注意力机制,其主要思想是将通道注意力和位置信息相结合,在不增加计算成本的前提下增强特征表达能力. CA 结构主要包括坐标信息嵌入和坐标注意力生成.

进行坐标信息嵌入时,首先,对每个通道分别执行沿宽度和高度的全局平均池化操作,得到两个分别与宽度和高度有关的特征图;其次,将特征图在通道维度上进行拼接,并通过二维卷积层提取融合特征;最后,对得到结果进行批量归一化,并通过非线性激活函数引入非线性特征,增加模型的表达能力.

进行坐标注意力生成时,先将坐标信息嵌入得到的结果分裂为与宽度和高度对应的特征图,再通过卷积及激活函数处理,生成两个注意力分数图.

坐标注意力机制的优点是考虑了位置信息的编码,从而能捕获跨通道和位置感知的信息,使模型可以更精准地定位并识别目标.坐标注意力机制通过使用两个互补的一维全局池化操作得到视觉任务中至关重要的空间位置间的长距离依赖性.此外,它计算开销较小,适合对轻量化的模型进行改进.

在 PCB 缺陷检测任务中,主要是一些细小的目标缺陷检测效果不佳,因此为增强小目标检测的准确率,而又不引入过多的计算量,本文仅将 CA 模块加在小目标检测头的位置,如图 7 所示.数值实验结果表明这种策略效果很好.而当将 CA 模块应用于其他尺度的检测头前时,不仅会导致模型的计算量显著增加,模型的检测性能也会下降.

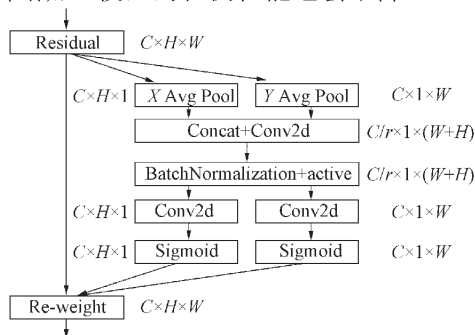


图 6 坐标注意力模块

Fig. 6 Coordinate attention module

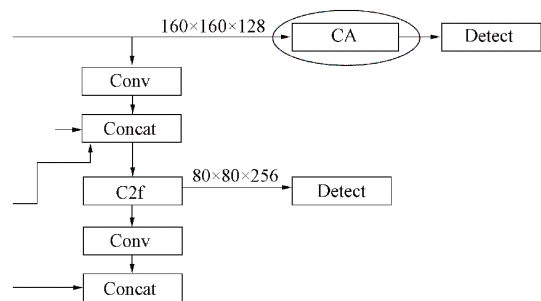


图 7 坐标注意力模块插入位置

Fig. 7 Insertion position of coordinate attention module

## 5 实 验

### 5.1 数据集及训练配置

实验基于北京大学公开的 PCB 缺陷数据集进行(数据集来源: <https://robot-ics.pkusz.edu.cn/resources/dataset/>). 该数据集包含缺孔(missing hole)、鼠咬(mouse bite)、开路(open circuit)、短路(short)、毛刺(spur)和余铜(spurious copper)6 种缺陷. 原始数据集共包含 693 张图片, 样本量较少. 为增强模型的泛化能力, 本文对该数据集进行数据增强. 先按 9:1 划分训练集和验证集, 再对训练集采用旋转、平移、裁剪、翻转、调整亮度和增加噪声 6 种方式随机组合进行数据增强, 共得到 3 120 张训练集图片.

实验使用的操作系统为 Linux Ubuntu, 显卡型号为 NVIDIA GeForce RTX 4090, 处理器为 Intel(R)Core(TM)i9-13900KF, 深度学习框架为 Pytorch, GPU 加速库为 CUDA12.2, 编程语言为 Python, 解释器为 Python 3.9. 训练时, 输入图像尺寸为 640×640, 每个模型训练 300 轮, batch size 为 16, 使用 SGD 优化器, 学习率设为 0.01.

### 5.2 评价指标

实验的评价指标主要是全类平均精度(mAP), 它与精确率(Precision)和召回率(Recall)有关. 精确率是指模型预测为正样本的实例中真正为正样本的比例, 召回率是指实际为正样本的实例中模型准确预测为正样本的比例, 其计算公式分别为

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}},$$

其中 TP 表示模型预测为正样本且实际是正样本的数量, FP 表示模型预测为正样本但实际是负样本的数量, FN 表示模型预测为负样本但实际是正样本的数量. 精确率和召回率之间的关系曲线称为 PR 曲线, PR 曲线的线下面积为平均正确率(AP). 全类平均精度(mAP)是对所有类别检测的平均正确率取平均数得到的, 它反映了模型在所有类别中的应用效果.

衡量模型计算复杂度的指标是 GFLOPs. GFLOPs 越高, 模型的计算量越大, 对硬件的要求也越高. 训练完成后得到的模型检测速度用每秒能检测的图像帧数(FPS)描述.

### 5.3 消融实验

用消融实验验证上述各种改进策略的效果, 将全类平均精度作为衡量指标, 预测效果列于表 1. 表 1 中比较了 YOLOv8 原始模型、SOH(small object head)+BiFPN 模型、SOH+BiFPN+C2fGC 模型和 SOH+BiFPN+C2fGC+CA 模型的应用效果, 其中第二列是交并比为 0.5 时的全类平均精度, 第三列是交并比为 0.5~0.95 时的全类平均精度. 使用 SOH+BiFPN+C2fGC+CA 模型对 PCB 进行检测时, mAP@0.5 值提高了 1.5 百分点, mAP@(0.5~0.95)值提高了 2.1 百分点. 由表 1 可见, 随着对 YOLOv8 模型的不断改进, 全类平均精度显著提高. 在 GFLOPs 指标上, SOH, BiFPN 和 CA 都在一定程度上增加了网络复杂度, 而 C2fGC 模块可以降低网络复杂度.

表 1 本文改进策略在数据集 PCB 上的应用效果

Table 1 Application effect of proposed improvement strategy on PCB dataset

模型	mAP@0.5/%	mAP@(0.5~0.95)/%	GFLOPs
原始 YOLOv8	93.9	49.5	6.8
SOH+BiFPN	94.2	49.8	9.4
SOH+BiFPN+C2fGC	94.9	50.3	8.5
SOH+BiFPN+C2fGC+CA	95.4	51.6	13.1

表 2 列出了 YOLOv8 原始模型和 SOH+BiFPN+C2fGC+CA 模型对各类缺陷的检测精度. 由表 2 可见, 虽然改进模型对于鼠咬和毛刺两种缺陷的检测精度略有下降, 但对于开路、短路和余铜 3 种缺陷的检测精度提升明显. 而且 FPS(帧/s)可以达到 105.4.

综上所述, 针对工业印刷电路板缺陷检测任务中, 小目标特征不明显且检测精度不足的问题, 本文提出了一种基于 YOLOv8 算法的改进算法. 该算法通过在原 YOLOv8 结构中增加小目标检测头,

引入 BiFPN 结构,设计 C2fGC 模块,添加注意力模块等方式大幅度改善了模型对 PCB 的检测效果.此外,本文还使用分组卷积降低了模型的复杂度.实验结果表明,改进后的模型可以很好地适应工业检测的精度及实时性要求,具有很好的应用前景.

表 2 模型改进前后各类别缺陷的检测精度

Table 2 Detection accuracy of each defect category before and after model improvement

模型	各类别 AP/%					
	缺孔	鼠咬	开路	短路	毛刺	余铜
原始 YOLOv8	99.5	91.9	94.5	92.5	93.6	91.5
SOH+BiFPN+C2fGC+CA	99.5	89.6	96.8	96.8	93.5	96.6

## 参 考 文 献

- [1] 谢宏威,张宪民,邝泳聪.基于图像特征统计分析的 PCB 焊点检测方法[J].仪器仪表学报,2011,32(2):401-407. (XIE H W, ZHANG X M, KUANG Y C. Solder Joint Inspection Method Based on Image Feature Statistica Analysis [J]. Chinese Journal of Scientific Instrument, 2011, 32(2): 401-407.)
- [2] 朱寒,林丽,王健华,等.基于改进模板匹配及图像差分法的 PCB 板缺陷多级检测方法[J].应用光学,2020,41(4):837-843. (ZHU H, LIN L, WANG J H, et al. Multi-level Detection Method for PCB Board Defects Based on Improved Template Matching and Image Difference [J]. Journal of Applied Optics, 2020, 41(4): 837-843.)
- [3] 胡兰兰,邓超.基于 YOLO-MCG 的 PCB 缺陷检测算法[J].仪表技术与传感器,2024(4):100-106. (HU L L, DENG C. PCB Defect Detection Algorithm Based on YOLO-MCG [J]. Instrument Technique and Sensor, 2024(4): 100-106.)
- [4] 刘天真,滕桂法,苑迎春,等.基于改进 YOLOv3 的自然场景下冬枣果实识别方法[J].农业机械学报,2021,52(5):17-25. (LIU T Z, TENG G F, YUAN Y C, et al. Winter Jujube Fruit Recognition Method Based on Improved YOLOv3 under Natural Scene [J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(5): 17-25.)
- [5] 江晟,张仲义,汪宗洋,等.基于改进 YOLOv7 的交通路口目标识别算法[J].吉林大学学报(理学版),2024,62(3):665-673. (JIANG S, ZHANG Z Y, WANG Z Y, et al. Target Recognition Algorithm of Traffic Intersection Based on Improved YOLOv7 [J]. Journal of Jilin University (Science Edition), 2024, 62(3): 665-673.)
- [6] 陈森,徐伟峰,王洪涛,等.基于改进 YOLOv7 的麦穗检测算法[J].吉林大学学报(理学版),2024,62(4):886-894. (CHEN S, XU W F, WANG H T, et al. Wheat Ear Detection Algorithm Based on Improved YOLOv7 [J]. Journal of Jilin University (Science Edition), 2024, 62(4): 886-894.)
- [7] 欧阳继红,王梓明,刘思光.改进多尺度特征的 YOLO\_v4 目标检测方法[J].吉林大学学报(理学版),2022,60(6):1349-1355. (OUYANG J H, WANG Z M, LIU S G. YOLOv4 Object Detection Method with Improved Multi-scale Features [J]. Journal of Jilin University (Science Edition), 2022, 60(6): 1349-1355.)
- [8] BOCHKOVSKI A, WANG C Y, LIAO H Y M. YOLOv4: Optimal Speed and Accuracy of Object Detection [EB/OL]. (2020-04-23)[2024-10-20]. <http://arxiv.org/abs/2004.10934>.
- [9] CUI B, LIANG L, JI B, et al. Exploring the YOLO-FT Deep Learning Algorithm for UAV-Based Smart Agriculture Detection in Communication Networks [J]. IEEE Transactions on Network and Service Management, 2024, 21(5): 5347-5360.
- [10] TAN M, PANG R, LE Q V. EfficientDet: Scalable and Efficient Object Detection [C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2020: 10781-10790.
- [11] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet Classification with Deep Convolutional Neural Networks [J]. Advances in Neural Information Processing Systems, 2017, 60(6): 84-90.
- [12] HOU Q B, ZHOU D Q, FENG J S. Coordinate Attention for Efficient Mobile Network Design [C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2021: 13713-13722.

(责任编辑:韩 啸)