

DOI: 10.13382/j.jemi.B2407757

基于改进 YOLOv8s 的轻量级 PCB 缺陷检测算法*

王天洋¹ 刘 路¹ 王太勇² 沙宗泰¹ 江 浩¹

(1. 天津职业技术师范大学机械工程学院 天津 300130; 2. 天津大学机械工程学院 天津 300130)

摘 要:YOLO 算法模型在轻量化与保证检测精度之间存在矛盾。针对印刷电路板小目标缺陷检测任务,提出一种基于改进 YOLOv8s 的轻量级目标检测算法。首先,删除主干网络中最后的卷积层与 C2f 层,然后引入轻量级跨尺度特征融合模块 CFFM,在实现模型轻量化的同时提升小目标检测精度。其次,引入分布移位卷积 DSCov,将 C2f 与 DSCov 结合生成 C2f_DSCov 模块,再与轻量级注意力机制 CBAM 集成,设计出 C2f_DSCov_CBAM 模块,分别替换骨干网络部分与颈部网络部分的 C2f,进一步减少模型参数量,增强模型特征提取能力。最后将辅助边界框损失函数 Inner-IoU、边界框聚焦损失函数 Focal IoU Loss、原边界框损失函数 CIoU 三者结合生成 Focal Inner-CIoU,引入大小可控的辅助边界框用以计算定位损失,提高较高的 IoU 边界框所占比重,最终实现检测精度提升。实验表明,改进模型较 YOLOv8s 原模型参数量降低 81.5%,计算量降低 21.3%,模型大小降低 72.3%,平均精度均值(mAP)提升 3.0%。有效降低了算法的计算成本,便于实际应用部署。

关键词:印刷电路板;计算机视觉;轻量化;小目标检测

中图分类号:TP391.4;TN911.73 **文献标识码:**A **国家标准学科分类代码:**540.4050

Lightweight PCB defect detection algorithm based on improved YOLOv8s

Wang Tianyang¹ Liu Lu¹ Wang Taiyong² Sha Zongtai¹ Jiang Hao¹

(1. School of Mechanical Engineering, Tianjin University of Technology and Education, Tianjin 300130, China;

2. School of Mechanical Engineering, Tianjin University, Tianjin 300130, China)

Abstract: There is a trade-off between the lightweight nature of YOLO algorithm models and maintaining detection accuracy. To address the task of detecting small defects in printed circuit boards, we propose a lightweight object detection algorithm based on an improved YOLOv8s. This approach significantly reduces the number of parameters and model size while enhancing detection accuracy. First, remove the last convolutional layer and the C2f layer from the backbone network. Then, introduce the lightweight cross-scale feature fusion module to achieve model lightweighting while enhancing the detection accuracy of small objects. Secondly, we introduce distribution shifting convolution, combining C2f and DSCov to create the C2f_DSCov module, which is then integrated with the lightweight attention mechanism CBAM to design the C2f_DSCov_CBAM module. This module replaces the C2f components in both the backbone and neck networks, further reducing the number of model parameters and enhancing feature extraction capability. Finally, by combining the auxiliary bounding box loss functions Inner-IoU, the bounding box focal loss function Focal IoU Loss, and the original bounding box loss function CIoU, we design the Focal Inner-CIoU. This introduces a controllable auxiliary bounding box to calculate localization loss, increasing the proportion of high IoU bounding boxes and ultimately enhancing detection accuracy. Experimental results show that compared to the original YOLOv8s model, the improved model reduces the number of parameters by 81.5%, computation by 21.3%, and model size by 72.3%, while increasing mAP by 3.0%. This effectively lowers the computational cost of the algorithm, making it more suitable for practical applications and deployment.

Keywords:PCB; computer vision; lightweight; small target detection

0 引言

近年来,电信、航空航天和医疗设备等行业对高质量印刷电路板 (printed circuit board, PCB) 的需求显著增加。PCB 作为电子产品的核心组成部分,其质量直接影响到产品的性能和可靠性。在 PCB 生产过程中,容易出现如穿孔、开路、毛刺等各种缺陷,任何未检测到的缺陷都可能导致产品故障,使得 PCB 自动缺陷检测成为一项关键任务。

在早期 PCB 生产流程中,缺陷检测主要依靠人工检查。工人通过肉眼或放大镜,检查 PCB 表面的焊点、走线和组件的安装情况。虽然这种方法适用于简单的 PCB 产品检测,但随着电路板设计复杂度和生产量的增加,人工检测效率低、成本高且易出错,难以满足大规模生产的需求。随着图像处理技术的发展,自动光学检测技术 (automated optical inspection, AOI)^[1] 开始被应用于 PCB 缺陷检测中。这种方法依赖于预设的规则,利用 2D 或 3D 摄像头拍摄 PCB 图像,并通过图像处理算法检测是否存在缺陷,如焊点不良、走线断裂或错位。尽管这一阶段检测效率得到了提高,人工干预逐渐减少,但基于规则的算法容易受到图像噪声、复杂光照条件等因素的影响,导致误报率和漏报率较高,且无法有效应对 PCB 设计日益复杂化的问题。随着计算机技术的发展,基于机器学习和深度学习的检测算法在图像识别领域取得了显著进展。这类方法通过训练模型来学习如何识别特定类型的缺陷,具有更强的泛化能力和自动化水平,机器学习算法 (如支持向量机、KNN 等),在 PCB 缺陷检测领域取得了一定成效,但仍面临数据依赖性强、特征提取方法复杂等问题^[2]。

随着卷积神经网络 (convolutional neural networks, CNN) 在图像处理领域的成功应用,基于深度学习的目标检测方法成为了 PCB 缺陷检测的主流。这一方法能够自动提取图像特征,并通过大规模数据训练来提高检测的精度和速度。依据检测流程的不同,目标检测算法分为以 Mask R-CNN^[3]、Faster-RCNN^[4] 为代表的双阶段算法和以 SSD^[5]、YOLO 系列^[6-7] 为代表的单阶段算法。Niu 等^[8] 在 Faster-RCNN 上进行改进,通过在 K-means 聚类算法中加入了遗传算法,生成与数据集更匹配的初始锚框,再将 Resnet50 网络中的标准卷积替换为深度可分离卷积以减少计算量,平均精度达到 95.6%,检测速度达到 0.125 s。Fan 等^[9] 使用 EfficientNet-B7 网络结构的前 8 个阶段替代 VGG-16 进行特征提取,并使用广义交并比 (GIoU) 作为边界框回归的损失函数,避免了预测框与真实框不相交时导致的非最优解问题,并使用性能更好的 Swish 激活函数替代 ReLU,检测精度达到 99%。但双

阶段算法仍面临模型较复杂,计算速度慢,硬件需求较高的问题,这对于资源有限的嵌入式系统或边缘设备并不友好。因此,模型结构更加简单,计算速度更快的单阶段算法 YOLO 成为 PCB 缺陷检测的主流算法。杨永跃等^[10] 以 YOLOv5x 为基础算法,通过增加小目标检测头,设计全新的 PANet 多特征融合结构基础算法,以及引入 CBAM 注意力机制,将 PCB 缺陷的检测精度提升 11.3% 达到 90.8%,YOLOv5x 作为 YOLOv5 系列中的轻量级版本,拥有较快的检测速度,但 90.8% 的检测精度无法满足工业检测的精度需求。许皓翔等^[11] 通过使用轻量化网络 Efficientnet-lite 替换 YOLOv5l 主干网络,降低模型复杂度,并通过加入小目标检测头提升模型对小目标特征的提取能力,提升识别精度,但由于 YOLOv5l 模型本身的局限性,其模型大小仍较为庞大。沈萍等^[12] 在 YOLOv8 基础上更换轻量级主干网络后,引入 Triplet Attention 模块,增强张量中不同维度间特征的捕捉能力,引入兼顾边界框定位精度与分类性能的损失函数 MPDIoU^[13],在一定程度上解决了模型简化之后带来的精度下降问题,但对于利用 YOLO 系列进行小目标识别存在缺陷的问题没有明显改善。YOLO 使用较大的网格单元来预测目标,对于 PCB 小目标缺陷识别精度较低,且有大量计算冗余,造成资源浪费。无论是双阶段还是单阶段深度学习方法,大多都只在提高小目标检测精度和降低模型复杂度其中一个问题上取得了一定成果,对于同时实现轻量化设计并提高检测精度方面,仍有改进空间。

YOLOv8 具备良好的模块化结构,易于进行进一步的优化和自定义。例如 C2f 模块可以根据需求被替换或增强,适用于不同的任务需求;其高度灵活的网络结构可以方便地结合其他注意力机制或特征融合模块,从而提高目标检测的准确性。此外,YOLOv8 引入动态锚框机制,在一定程度上提升了小目标检测能力。YOLOv8s 作为 YOLOv8 系列中的轻量级版本,具有较少的参数和较小的模型尺寸。这使其能够在嵌入式设备或计算资源有限的环境中高效运行,适合 PCB 缺陷检测任务中对实时性要求较高的场景。故本文选择以 YOLOv8s 作为基础模型进行研究。

目前 YOLOv8 尚存在网络结构中大网格、深层次预测对小目标检测不利,以及回归损失在定位精度上的不足带来的检测精度不高的问题。为解决这一问题,同时实现进一步轻量化模型结构设计以便于实际应用部署,本文提出一种基于改进 YOLOv8s 的轻量级 PCB 缺陷检测算法,通过改进网络结构,应用轻量级卷积与注意力机制,采用更高效准确的边界框损失函数,同时实现了模型的轻量化改造与识别精度的提升。

1 YOLOv8s 模型改进

本文对 YOLOv8s 模型进行了 3 个方面的改进,改进后的结构如图 1 所示。

1) 删除骨干网络最后一层卷积层与 C2f 层,然后引入跨尺度特征融合模块(cross-scale feature fusion module, CFFM),简化模型结构的同时降低感受野,提升小目标识别精度。

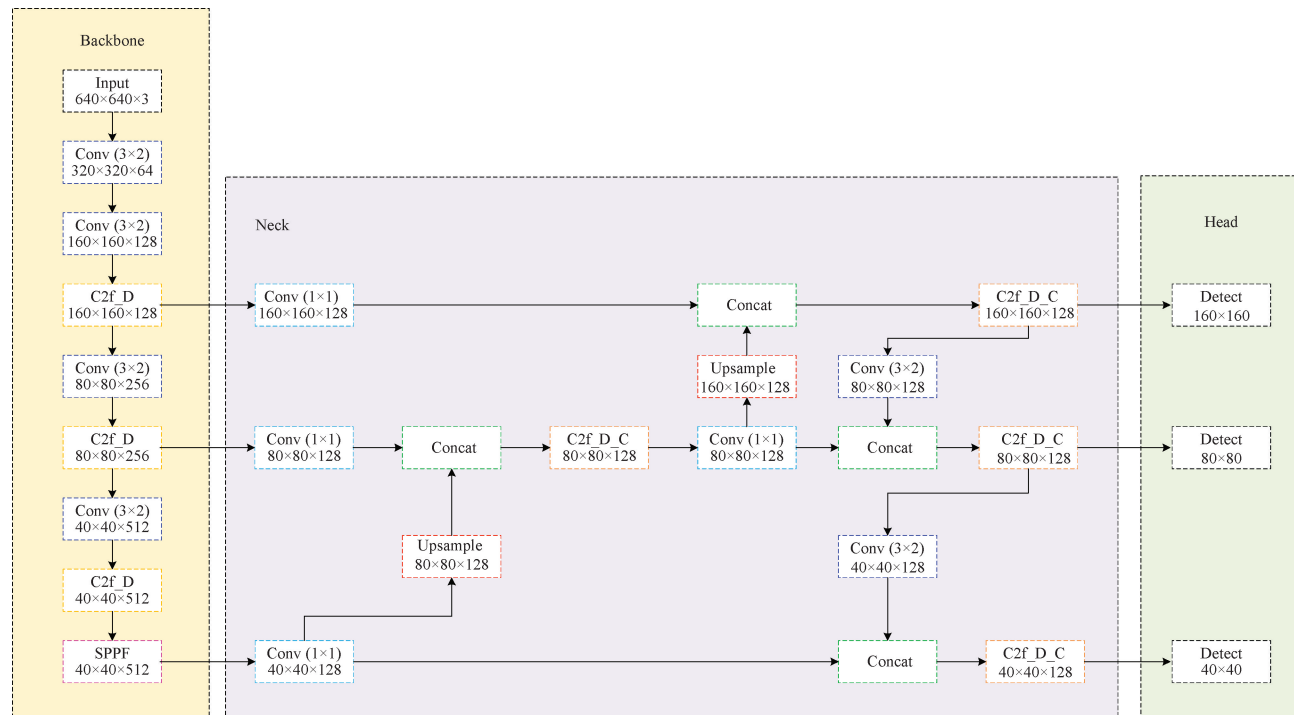


图 1 改进后的 YOLOv8 网络结构

Fig. 1 Structure of improved YOLOv8s network

1.1 网络结构改进

在骨干网络部分, YOLOv8 包含多次下采样操作, 可以扩大网络的感受野, 使后续层能捕捉更大范围的上下文信息, 能够提高中大型目标的特征表达, 但会导致小目标在较深层次的特征图上仅占据少量像素, 特征表达不够充分, 不利于小目标识别。同时, 深层卷积带来的大量通道数产生了过大的计算冗余, 造成资源浪费。其次, 在颈部网络部分, YOLOv8 采用 PAN 和 FPN 来进行特征融合, 但对于小目标, 其在低层特征图中的信息可能无法有效传播到最终的检测头。尤其是在目标过小且分布密集时, 这种特性融合显得不足。针对 PCB 缺陷检测这一小目标检测任务, 为提高检测精度, 同时实现轻量化模型结构设计, 需对 YOLOv8 网络结构进行改进。

在 YOLOv8s 原模型中, 骨干网络部分通过 5 个卷积层将 640×640 特征图下采样为 20×20 特征图, 经特征融合分别输出 80×80 , 40×40 , 20×20 三个图层, 感受野分别

别精度。

2) 设计 C2f_DSConv 与 C2f_DSConv_CBAM 模块, 分别替换掉骨干网络与颈部网络中的 C2f。减少模型参数的同时保证识别精度。

3) 将辅助边界框损失函数 Inner_IoU 与 Focal loss IoU 集成到原损失函数 CIoU 中, 设计出 Focal Inner_CIoU, 进一步提升识别精度。

为 8×8 , 16×16 , 32×32 。在本文模型中, 删去骨干网络最后一层卷积以及 C2f 模块, 使骨干网络输出 40×40 特征图, 最终 3 个输出图层感受野大小分别降低一倍, 分别为 4×4 , 8×8 , 16×16 。更小的感受野能够更精确的识别到小目标特征, 提高了小目标检测精度, 同时减小了模型深度, 降低了模型复杂度。

2023 年发布的 RT-DETR 算法^[14]中提出 CFFM 结构。与 YOLOv8s 相同, CFFM 仍以 PAN+FPN 结构为基础, 但在不同模块连接处加入了若干 1×1 卷积模块。 1×1 卷积模块有自由设置通道数的特性, 通过设置较低的通道数可以有效减少参数量。同时, 1×1 卷积通过在通道维度上进行卷积操作, 将不同通道之间的特征进行组合, 增强模型对于尺度变化的适应性, 有助于模型学习到更加丰富和抽象的特征, 提高检测精度。在本文模型中, 分别在改进后的骨干网络与颈部网络的 3 个连接处, 以及颈部网络的上采样模块与 C2f 模块连接处加入 1×1 卷

积模块,并将通道数设置为 128,减少了大量计算冗余,同时保证了检测精度。

1.2 C2f 模块改进

YOLOv8 采用 C2f 模块作为特征提取模块,与 YOLOv5 的 C3 模块不同,C2f 模块通过简化设计和压缩卷积,减少了计算复杂度和参数量。但在深层网络中,大量的卷积层仍会带来显著的计算开销和内存占用。同时,C2f 模块依赖卷积操作提取局部特征,难以有效建立 PCB 图像中小目标缺陷与背景或邻近结构的全局上下文关系,导致检测精度不高。为减小模型计算复杂度,同时保证识别精度,需对 C2f 模块进行改进。因此,引入分布偏移卷积(distribution shifting convolution, DSConv)^[15]与卷积块注意力模块(convolutional block attention module, CBAM)^[16]。

DSConv 使用量化和分布偏移来模拟卷积层的行为,替换原始架构中的卷积层,实现较低的内存使用和轻量化计算。DSConv 由可变量化核(VQK)和分布偏移组件(distribution shifts, DS)两部分组成,而 DS 由核分布偏移器(KDS)和通道分布偏移器(CDS)两部分组成。DSConv 结构如图 2 所示。

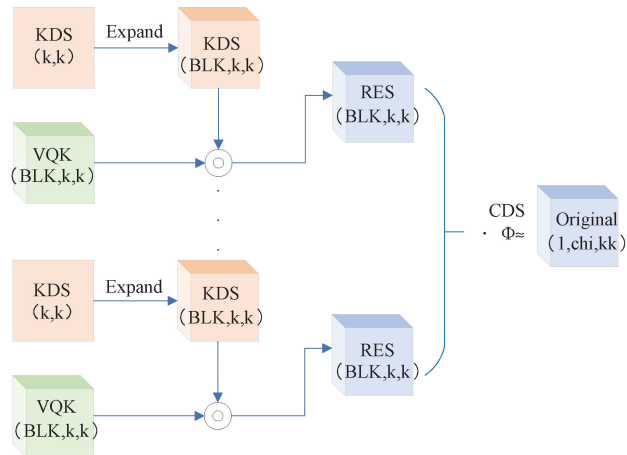


图 2 DSConv 结构

Fig. 2 DSConv structure diagram

VQK 将原始浮点模型参数量化为可变位数的整数参数,并使其保持与原始卷积相同的数据维度(输入通道数、输出通道数、卷积核高度、卷积核宽度)。在量化过程中,输入网络需要量化的位数,经过式(1)得到新的整数权重 wq ,并存储到 VQK 中以进行后续的训练和推断。相比于原始的浮点型模型参数,整数型参数更加轻量。

$$wq \in Z - 2(b-1) \leq wq \leq 2(b-1) - 1 \quad (1)$$

式中: wq 表示张量中每个参数的值; b 表示位数。

DS 通过移位 KDS 和 CDS 改变 VQK 的分布,输出与原始权重相匹配的张量的值,以保留与原始卷积相同的输出,从而保证模型的特征提取能力。

CBAM 是一种结合了通道和空间两种注意力机制的模块,旨在增强卷积神经网络的特征表示能力。CBAM 结构如图 3 所示,其工作原理可以概述为两个主要步骤。

步骤 1) 经过通道注意力模块(channel attention module, CAM),对输入特征图进行处理,通过全局平均池化和最大池化将每个通道在空间维度上的信息进行压缩,得到两个一维的特征图,实现了对全局空间信息的压缩和提取。经共享网络处理后将特征值相加并使用 sigmoid 激活函数将其限制在 0~1,减少共享网络参数量,降低计算成本。

步骤 2) 将经过调整的特征图与原始特征图 F 相乘得到 F' ,这一步可以调整每个通道的特征值,强调重要通道的信息。再将 F' 输入空间注意力模块(spatial attention module, SAM),在通道维度上对 F' 进行平均池化和最大池化,生成两个单通道特征图,将其连接并输入 1 个卷积层,计算出空间注意力权重。得到的这些权重用于调整特征图的每个空间位置,使模型能够关注图像中的重要位置。最后,将特征图 F' 与空间注意力权重相乘,得到最终输出结果。

CBAM 通过逐层处理特征图,分别计算通道和空间的注意力权重,从而使模型能够更有效地关注重要特征,忽略不相关的信息,提升检测精度。

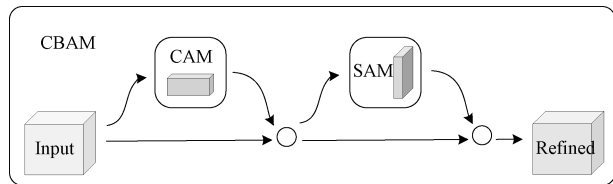


图 3 CBAM 结构

Fig. 3 CBAM structure diagram

在本文模型中,将 C2f 模块 Bottleneck 结构中的卷积替换成 DSConv,得到 C2f_DSConv 模块,再将 CBAM 注意力机制加入 Bottleneck 结构中,得到 C2f_DSConv_CBAM 模块,具体结构如图 4 所示(图中 C2f_DSConv 与 C2f_DSConv_CBAM 分别简称为 C2f_D 与 C2f_D_C)。如图 1 所示,将骨干网络部分和颈部网络部分的 C2f 模块分别替换为 C2f_DSConv 与 C2f_DSConv_CBAM。通过上述改进可降低模型计算量,帮助模型在更细粒度的层级上进行特征增强。同时在特征提取的不同阶段应用注意力机制,有助于在多个层级上学习到不同层次的特征信息,从而更好地捕捉图像中的全局和局部特征,有效应对在模型轻量化之后的精度下降问题。

1.3 损失函数改进

YOLOv8 采用 anchor-free 的方法,通过直接预测目标边界框的中心点和尺寸确定预测框,使得模型更能适应不同尺度和形状的目标,在一定程度上增强了模型的

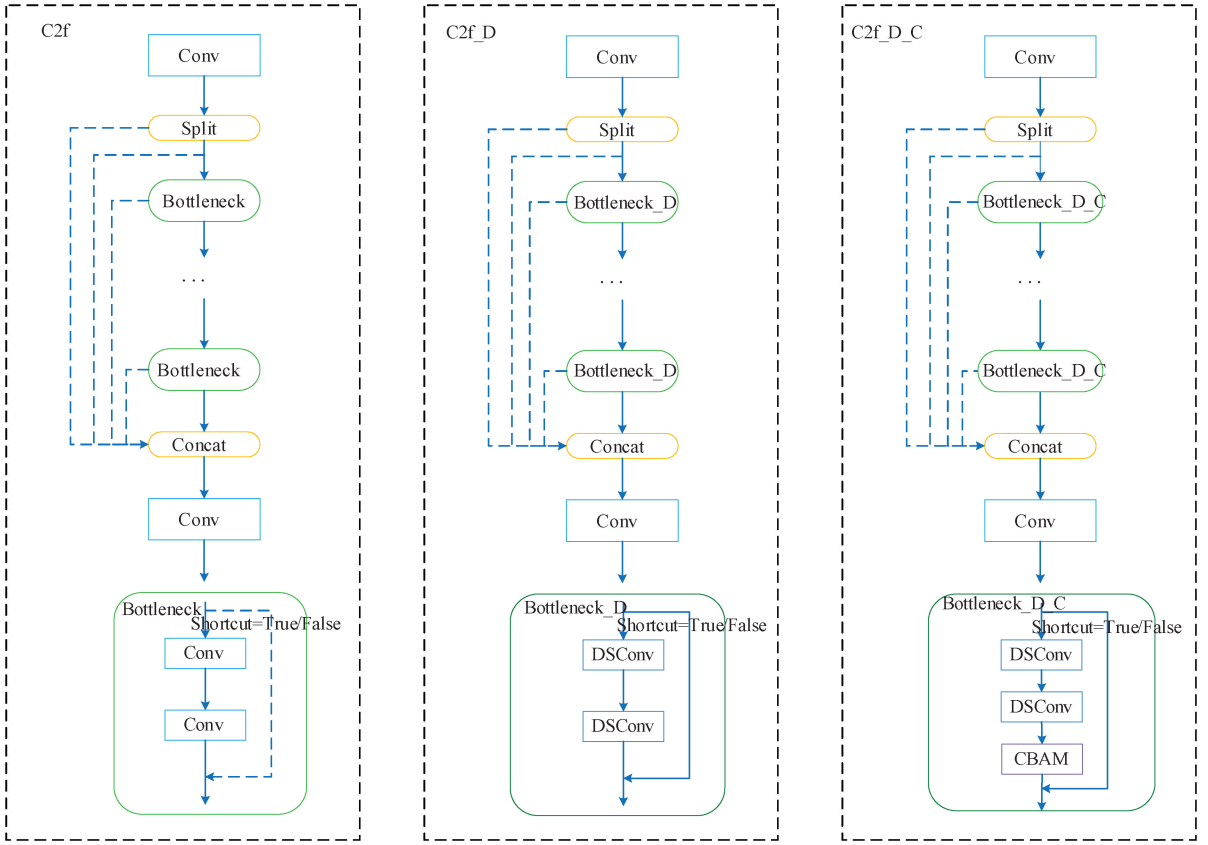


图 4 C2f、C2f_DSCConv、C2f_DSCConv_CBAM 结构

Fig. 4 C2f, C2f_DSCConv, C2f_DSCConv_CBAM structure diagram

泛化能力^[17]。但在小目标检测任务中,微小的定位误差也可能对结果造成较大影响。另一方面,边界框回归过程中存在训练样本不平衡的问题,即一张图像中回归误差较小的高质量锚框的数量远少于误差大的低质量锚框数量,质量较差的锚框会产生过大的回归梯度,影响训练结果。为解决上述两个问题。分别引入 Inner-IoU^[18]与 Focal IoU Loss^[19]。

为提高边界框的定位精度,引入基于辅助边框的 IoU 损失函数 Inner-IoU,其定义为:

$$b_l^{gt} = x_c^{gt} - \frac{w^{gt} \times ratio}{2}, b_r^{gt} = x_c^{gt} + \frac{w^{gt} \times ratio}{2} \quad (2)$$

$$b_t^{gt} = y_c^{gt} - \frac{h^{gt} \times ratio}{2}, b_b^{gt} = y_c^{gt} + \frac{h^{gt} \times ratio}{2} \quad (3)$$

$$b_l = x_c - \frac{w \times ratio}{2}, b_r = x_c + \frac{w \times ratio}{2} \quad (4)$$

$$b_t = y_c - \frac{h \times ratio}{2}, b_b = y_c + \frac{h \times ratio}{2} \quad (5)$$

$$inter = (\min(b_r^{gt}, b_r) - \max(b_l^{gt}, b_l)) \times (\min(b_b^{gt}, b_b) - \max(b_t^{gt}, b_t)) \quad (6)$$

$$union = (w^{gt} \times h^{gt}) \times (ratio)^2 + (w \times h) \times (ratio)^2 - inter \quad (7)$$

$$IoU^{inner} = \frac{inter}{union} \quad (8)$$

式中: b^{gt} 表示 GT 框; b 表示锚框; (x_c^{gt}, y_c^{gt}) 表示 GT 框内部中心点; (x_c, y_c) 表示锚框内部中心点; h, w 分别表示高度和宽度; 变量“ratio”表示尺度因子,取值范围为 $[0.5, 1.5]$ 。

Inner-IoU 可设置尺度因子 ratio 控制辅助边框大小,并通过辅助边框计算损失。与以往的损失函数改进不同,Inner-IoU 在预测边框尺度层面进行改进,通过调整 ratio 的值找到定位精度更高的边界框尺寸,进而提高识别精度。

针对边界框回归过程中训练样本不平衡的问题,引入 Focal IoU Loss,其定义为:

$$L_{Focal-IoU} = IoU^\lambda L_{IoU} \quad (9)$$

式中: λ 表示异常因子。

Focal IoU Loss 使用 IoU 的值对原始损失进行加权, IoU 更高的锚框能获得更大的权重,帮助 IoU 损失更专注于高质量锚框,从而提升检测精度。

本文模型中,将原边界框损失函数 CIoU 与 Inner-IoU 结合,生成 Inner-CIoU,通过 ratio 将边框略微调小,提高定位精度,加速模型收敛。再引入 Focal IoU Loss,生成

Focal Inner-CIoU, 对 Inner-CIoU 生成的锚框损失进行加权, 提高较高的 IoU 锚框所占权重, 进一步提升检测精度。其定义为:

$$L_{Focal-Inner-CIoU} = IoU^{\lambda} (L_{CIoU} + IoU - IoU^{inner}) \quad (10)$$

2 实验分析

2.1 数据集

实验数据集采用北京大学人工智能实验室公开的 PCB 缺陷数据集^[20], 该数据集包含 693 张图片, 共 6 类, 分别为缺孔、余铜、短路、鼠咬、开路 and 毛刺。每个类别图片张数分别为 115、115、116、116、115 和 116。各类别示例如图 5 所示。

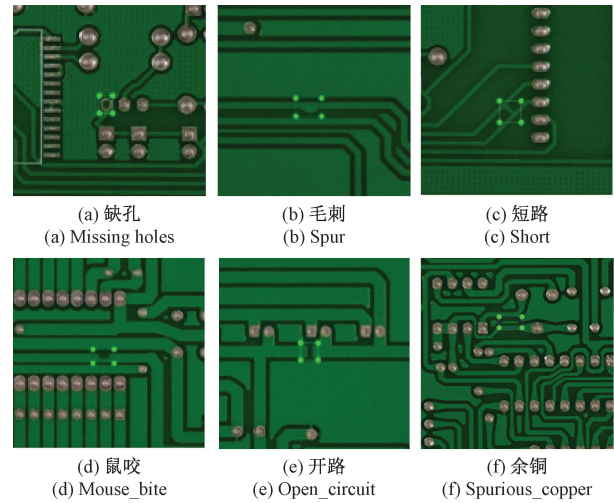


图 5 各类别示例

Fig. 5 Example diagram of each category

2.2 实验环境与参数配置

本文实验采用 windows10 操作系统, NVIDIA RTX A2000 显卡 (12 G), 实验环境为 python 1.8、CUDA11.3 和 Pytorch 1.10 深度学习框架。数据集按 7 : 3 比例划分为训练集和验证集, 未使用预训练权重。具体实验参数如表 1 所示。

表 1 实验参数

Table 1 Experimental parameter table	
参数名	参数值
学习率	0.01
动量	0.937
权重衰减	0.000 5
批大小	16
迭代次数	200
输出图片尺寸	640×640

2.3 评价指标

本文实验采用参数量、计算量、模型体积、平均精度均值 (mean average precision, mAP) 为评价指标。mAP 定义为:

$$Precision = \frac{TP}{TP + FP} \quad (11)$$

$$Recall = \frac{TP}{TP + FN} \quad (12)$$

$$AP = \int_0^1 P(R) dR \quad (13)$$

$$mAP = \frac{\sum_{i=0}^n AP(i)}{n} \quad (14)$$

式中: TP 表示正确预测的正样本数量; FP 表示负样本被错误预测为正样本的数量; FN 表示未检测出的正样本数量。

2.4 实验结果与分析

1) 网络结构改进对比实验

为验证调整网络结构与添加 CFFM 的有效性, 本小节对 YOLOv8s 原始模型、删除骨干网络最后两层模型、删除骨干网络最后两层并添加 CFFM 模型进行对比实验。实验结果如表 2 所示。调整网络层数后, 参数量大幅降低, 计算量小幅提高, mAP 提高。引入 CFFM 结构后, 最终参数量降低 1.22×10⁶, 计算量降低 6.4 GFLOPs, mAP 提高 1.3%。

表 2 模型结构对比试验

Table 2 Model structure comparison test table			
模型	参数量/ (×10 ⁶)	计算量/ GFLOPs	mAP/%
YOLOv8s	11.16	28.7	93.2
YOLOv8s+调整层数	3.49	30.6	93.9
YOLOv8s+调整层数+CFFM	2.27	24.2	94.2

2) C2f 模块改进对比实验

在调整网络结构的基础上, 进行 C2f 模块改进对比实验。实验结果如表 3 所示, 将网络中 C2f 模块替换为 C2f_DSConv 后, 参数量及计算量减少, mAP 略微降低。进一步将颈部网络部分 C2f_DSConv 改进为 C2f_DSConv_CBAM 后, 参数量与计算量基本不变, mAP 有所增长。最终通过上述改进方法使得 mAP 较原始模型有 0.7% 提升, 参数量降低了 0.16×10⁶, 计算量降低了 1.6 GFLOPs。

表 3 C2f 模块改进对比试验表

Table 3 C2f module improvement comparison test table			
C2f 模块	参数量/(×10 ⁶)	计算量/GFLOPs	mAP/%
C2f	2.27	24.2	94.2
C2f_DSConv	2.06	22.0	94.0
C2f_DSConv_CBAM	2.09	22.6	94.9

3) 损失函数改进对比试验

为验证损失函数改进对 mAP 增长的有效性,分别进行损失函数改进对比实验与 Inner 参数对比实验。实验结果如表 4 所示,单独引入 Inner-CIoU 与 Focal Ciou, mAP 分别增长 0.9% 和 0.2%,将二者结合改进, mAP 增长 1.3%。在 Inner 参数对比实验中,当设置 ratio 为 0.95 时, mAP 最高,达到了 96.2%。

4) 消融实验

为验证本文算法的有效性,对不同改进进行消融实验,实验结果如表 5 所示。经过对网络结构、C2f 模块以及损失函数的改进,相较于原始 YOLOv8 模型,参数量降

低 81.5%, 计算量降低 21.3%, 模型大小降低 72.3%, mAP 提升 3.0%。针对 PCB 小目标检测任务,同时实现了轻量化部署和检测精度的提高。

表 4 损失函数改进比试验表

损失函数改进对比实验		Inner 参数对比实验	
损失函数	mAP/%	ratio	mAP/%
Ciou	94.9	0.85	95.2
Inner-Ciou	95.8	0.90	95.7
Focal Ciou	95.1	0.95	96.2
Focal Inner-Ciou	96.2	1.05	95.1

表 5 消融实验对比

Table 5 Comparison table of ablation experiments								
YOLOv8	调整层数	CFFM	C2f 改进	损失函数改进	参数量/($\times 10^6$)	计算量/GFLOPs	模型体积/MB	mAP/%
√					11.16	28.7	24.2	93.2
√	√				3.49	30.6	7.4	93.9
√	√	√			2.27	24.2	4.9	94.2
√	√	√	√		2.09	22.6	6.7	94.9
√	√	√	√	√	2.09	22.6	6.7	96.2

5) 不同算法对比实验

为验证本文算法较于其他算法的优势,将算法与 YOLOv5、YOLOv7、YOLOv9、YOLOv10、Faster RCNN、Rt-

Detr、SSD 等算法进行比较,实验结果如表 6 所示,本文算法在轻量化指标与检测精度方面均体现出较高优越性。

表 6 不同算法对比实验

Table 6 Comparison of different algorithms in the experimental table						
模型	参数量/($\times 10^6$)	计算量/GFLOPs	模型体积/MB	mAP/%	帧率/FPS	
YOLOv5	7.1	16.0	54.3	93.6	72	
YOLOv7	37.2	105.2	74.9	92.5	45	
YOLOv8	11.6	28.7	24.2	93.2	140	
YOLOv9	9.7	39.6	20.3	95.0	96	
YOLOv10	8.04	24.5	15.8	90.7	100	
Faster RCNN	41.2	206.7	89.6	9.05	22	
Rt-Detr	32.8	108.0	66.2	0.87	32	
SSD	24.5	87.9	50.3	77.6	36	
本文	2.09	22.6	6.7	96.2	104	

3 结 论

在利用 YOLOv8 进行 PCB 缺陷检测过程中,一方面发现其深层次的网络结构会带来大量计算冗余,同时深层卷积带来的较大感受野对小目标检测不利。在损失函数设计方面,也仍有提升边界框回归精度的空间;另一方面,需要在确保检测精度不降低的前提下,降低模型复杂度以实现轻量化部署。针对这些问题提出一种基于改进

YOLOv8s 的轻量级 PCB 缺陷检测算法。首先通过删除骨干网络末尾的卷积层与 C2f 层,降低网络感受野,提升网络对小目标特征的提取能力,同时简化模型结构,减少模型参数量。引入 CFFM 结构,调整网络通道数,减少计算冗余,并加强颈部网络的特征融合。其次,设计 C2f_DSConv 与 C2f_DSConv_CBAM 模块,进一步降低模型参数量,增强模型特征提取能力。最后,在原边界框损失函数 Ciou 基础上,集成 Inner-IoU 与 Focal IoU Loss,提升边界框对小目标特征的定位精度,提高较高的 IoU 边界框

所占比重,从而提升模型检测精度。

在北京大学公开 PCB 缺陷检测数据集上,本文算法相较原始 YOLOv8 算法,平均精度均值提升 3.2% 达到 96.2%,参数量降低 81.5%,计算量降低 21.3%,模型大小降低 72.3%,实现了高精度与轻量化的网络设计。未来工作将进一步优化网络结构,降低模型复杂度以便更加利于轻量化部署。同时提升检测速度,以更好地满足实际工业生产的需要,并推动本文算法在更多类型工业视觉任务中的应用。

参考文献

- [1] LIAO H C, LIM Z Y, HU Y X, et al. Guidelines of automated optical inspection (AOI) system development[C]. 2018 IEEE 3rd International Conference on Signal and Image Processing (ICSIP). IEEE, 2018: 362-366.
- [2] 赵朗月, 吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. 仪器仪表学报, 2022, 43(1): 198-219.
ZHAO L Y, WU Y Q. Research progress of surface defect detection methods based on machine vision[J]. Chinese Journal of Scientific Instrument, 2022, 43(1): 198-219.
- [3] HE K M, GKIOXARI G, DOLLÁR P, et al. Mask R-CNN[C]. Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961-2969.
- [4] REN SH 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, 2016, 39(6): 1137-1149.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. Ssd: Single shot multibox detector[C]. Computer Vision - ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 - 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.
- [6] REDMON J. You only look once: Unified, real-time object detection [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 779-788.
- [7] TERVEN J, CÓRDOVA-ESPARZA D M, ROMERO-GONZÁLEZ J A. A comprehensive review of YOLO architectures in computer vision: From YOLOv1 to YOLOv8 and YOLO-nas [J]. Machine Learning and Knowledge Extraction, 2023, 5(4): 1680-1716.
- [8] NIU J, HUANG J, CUI L, et al. A PCB defect detection algorithm with improved faster R-CNN [C]. ICBASE, 2022: 283-292.
- [9] FAN F, WANG B, ZHU G, et al. Efficient faster R-CNN: Used in PCB solder joint defects and components detection[C]. 2021 IEEE 4th International Conference on Computer and Communication Engineering Technology (CCET). IEEE, 2021: 1-5.
- [10] 杨永跃, 夏远超. PCB 缺陷检测深度学习算法的精度改进[J]. 电子测量与仪器学报, 2023, 37(5): 11-19.
YANG Y Y, XIA Y CH. Accuracy improvement of deep learning algorithm for PCB defect detection[J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(5): 11-19.
- [11] 许皓翔, 段国华. 基于轻量化的 YOLOv5 的 PCB 缺陷检测算法[J]. 电气自动化, 2024, 46(2): 95-97, 102.
XU H X, SHU G H. PCB defect detection algorithm based on lightweight YOLOv5 [J]. Electrical Automation, 2024, 46(2): 95-97, 102.
- [12] 沈萍, 李想, 杨宁, 等. 基于三重注意力的轻量级 YOLOv8 印刷电路板缺陷检测算法[J]. 微电子学与计算机, 2024, 41(4): 20-30.
SHEN P, LI X, YANG N. et al. Lightweight YOLOv8 PCB defect detection algorithm based on triple attention [J]. Microelectronics & Computers, 2024, 41(4): 20-30.
- [13] MA S L, XU Y. MPDIoU: A loss for efficient and accurate bounding box regression [J]. ArXiv preprint arXiv. 2307. 07662, 2023.
- [14] ZHAO Y AN, LYU W Y, XU SH L, et al. Detrs beat yolos on real-time object detection [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 16965-16974.
- [15] NASCIMENTO M G, FAWCETT R, PRISACARIU V A. Dsconv: Efficient convolution operator [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 5148-5157.
- [16] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 3-19.
- [17] SOHAN M, SAI RAM T, REDDY R, et al. A review on YOLOv8 and its advancements [C]. International Conference on Data Intelligence and Cognitive Informatics. Springer, 2024: 529-545.
- [18] ZHANG H, XU C, ZHANG SH. Inner-IoU: More effective intersection over union loss with auxiliary bounding box [J]. ArXiv prints arXiv. 2311. 02877, 2023.
- [19] ZHANG Y F, REN W Q, ZHANG ZH, et al. Focal and efficient IOU loss for accurate bounding box regression [J]. Neurocomputing, 2022, 506: 146-157.
- [20] DING R W, DAI L H, LI G P, et al. TDD-Net: A tiny

defect detection network for printed circuit boards [J].
CAAI Transactions on Intelligence Technology, 2019,
4(2): 110-116.

作者简介

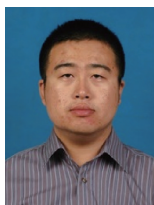


王天洋,现为天津职业技术师范大学硕士研究生,主要研究方向为机器视觉、智能制造。

E-mail: 1104116679@qq.com

Wang Tianyang now is a M. Sc. candidate of Tianjin University of Technology

and Education. His main research interests include machine vision and intelligent manufacturing.



刘路(通信作者),2004 年于燕山大学获得学士学位,2008 年于河北工业大学获得硕士学位,2011 年于天津大学获得博士学位,现为天津职业技术师范大学副教授,主要研究方向为机器学习、图像处理。

E-mail: lordman1982@163.com

Liu Lu (Corresponding author) received his B. Sc. degree from Yanshan University in 2004, B. Sc. degree from Hebei University of Technology in 2008 and Ph. D. from Tianjin University in 2011. He is currently an associate professor at Tianjin University of Technology and Education. His main research interests include machine learning and image processing.