

文章编号:1007-2780(2024)10-1332-09

基于改进 YOLOv7 的小目标焊点缺陷检测算法

刘兆龙^{1,3}, 曹伟², 高军伟^{1,3*}

(1. 青岛大学自动化学院, 山东青岛 266071;

2. 青岛国际机场集团有限公司, 山东青岛 266300;

3. 山东省工业控制技术重点实验室, 山东青岛 266071)

摘要:针对现有的小目标焊点缺陷检测方法存在错检、漏检等问题,提出一种改进 YOLOv7 的小目标焊点缺陷检测算法。考虑到焊点尺寸较小,添加小目标检测层和检测头以提取更多的浅层特征信息。引入无参注意力机制(SimAM)为特征图分配三维权重,提高模型特征提取能力。使用部分卷积(PCConv)重构 ELAN 模块,减少冗余运算和内存访问次数,在颈部利用长径特征网络(GiraffeDet)融合不同尺度特征,提高模型轻量化程度。最后,利用 NWD (Normalized Wasserstein Distance) 损失函数改进原有的 CIoU 损失函数,加快模型收敛速度并提高小目标检测精度。实验结果证明,改进后的 YOLOv7 算法平均检测精度达到 90.3%,相较于原算法提升了 5.1%,召回率提高了 3.2%,参数量降低了 36.3%,且在收敛速度方面有了较大的提升。本文算法为边缘设备检测小目标焊点缺陷提供了参考。

关键词:图像处理;缺陷检测;YOLOv7;SimAM;轻量化;NWD

中图分类号:TP391.41 文献标识码:A doi:10.37188/CJLCD.2024-0051

Defect detection of small object solder joints based on improved YOLOv7

LIU Zhaolong^{1,3}, CAO Wei², GAO Junwei^{1,3*}

(1. School of Automation, Qingdao University, Qingdao 266071, China;

2. Qingdao International Airport Group Co. Ltd., Qingdao 266300, China;

3. Shandong Key Laboratory of Industrial Control Technology, Qingdao 266071, China)

Abstract: Aiming at the problems of the existing small target solder joint defect detection methods, such as error detection and leakage detection, an improved YOLOv7 small target solder joint defect detection algorithm was proposed. Considering the small size of solder joints, a small target detection layer and detection head were added to extract more shallow feature information. The non-parametric attention mechanism (SimAM) was introduced to assign 3D weights to the feature graphs to improve the feature extraction ability of the model. Partial Convolution (PCConv) was used to reconstruct ELAN modules to reduce redundant operations and memory access, and GiraffeDet was used to integrate different scale features at the neck to improve the lightweight of the model. Finally, the NWD (Normalized Wasserstein

收稿日期:2024-02-26;修订日期:2024-04-10.

基金项目:山东省自然科学基金(No.ZR2019MF063)

Supported by Natural Science Foundation of Shandong Province (No.ZR2019MF063)

*通信联系人, E-mail: qdgao163@163.com

Distance) loss function was used to improve the original CIoU loss function, which sped up the convergence of the model and improved the detection accuracy of small targets. Experimental results show that the average detection accuracy of the improved YOLOv7 algorithm reaches 90.3%, which is 5.1% higher than that of the original algorithm. The recall rate is 3.2% higher, the number of parameters is 36.3% lower, and the convergence speed has been greatly improved. This algorithm provides a reference for detecting small target solder joint defects in edge equipment.

Key words: image processing; defect detection; YOLOv7; SimAM; lightweight; NWD

1 引 言

在当代电子制造产业中,印刷电路板(PCB)焊点的质量是保证整个电子设备可靠运行的关键因素。面对复杂的生产环境和有限计算资源的边缘设备,传统的计算模型往往难以达到理想的速度和精度,从而限制了这些方法在实际生产中的大规模应用。

随着深度学习技术的迅猛发展和各种硬件的不断更新换代^[1],各种新型焊点检测算法和轻量化模型方法被相继提出。传统深度学习模型算法,如二阶段检测算法 RCNN(Region with CNN Feature)^[2]、Fast-RCNN^[3]、Faster-RCNN^[4],双阶段检测算法 SSD(Single Shot MultiBox Detector)^[5]、YOLO(You Only Look Once)^[6]等。文献[7]针对焊膏缺陷尺寸差异大和图像采集的难题,开发了 D3PointNet 双层缺陷检测模型。实验结果显示,该模型有效提高了稳定性,其性能比传统 CNN 模型高出 10.2%。文献[8]在 YOLOv8 网络中加入 CA 注意力机制,使用 BiFPN 替换 PANet 特征金字塔,并将损失函数更换为 EIoU 损失函数,提高了检测精度,但其模型参数量和计算量大幅增加。文献[9]在 YOLOv5s 基础上利用 GhostConv 替换经典卷积,引入 ODCov 提高精度,其精度相较于原模型提高了 1%,参数量下降 23.89%,但总体精度较差。文献[10]将主干网络替换为 MobileNetv2 网络并利用 GIoU 损失函数替换原有损失函数,其精度约为 75%,检测精度较低。文献[11]用 GhostNet 代替 CSPDarknet53 作为主干网络并加入 SE 注意力机制,其平均检测精度为 86.68%,但对于数据集中特别小的缺陷难以检测。

针对目前焊点缺陷检测方法精度低且存在漏检、误检等问题,本文提出了一种基于改进

YOLOv7 的小目标焊点缺陷检测算法。焊点目标大多数为中小目标,本文添加一个小目标检测层,并用 K-means++ 算法聚类先验框加快模型收敛速度。在 Head 区域引入无参注意力 SimAM 模块,对密集分布且尺寸小的焊点图像分配三维权重,提高模型对关键特征的聚焦能力。为平衡模型参数量和计算量上的提升,用 PConv 替换卷积核为 3×3 、步长为 1 的经典 Conv,跳过部分卷积运算,在颈部利用 GiraffeDet 融合不同尺度特征,提高模型的轻量化程度。最后,利用 NWD(Normalized Wasserstein Distance)损失函数与原有的 CIoU 损失函数融合,更有利于小目标检测,获得了更快的收敛速度和更加准确的检测结果。

2 YOLOv7 算法

YOLOv7 是 2022 年提出的一种高效的单阶段目标检测网络,在提升检测精度和速度方面展现出显著优势^[12]。该网络架构由输入层(Input)、骨干网络(Backbone)、颈部(Neck)和预测头(Head)组成,其结构如图 1 所示。输入层主要进行数据的预处理,采用 Mosaic 数据增强等技术对图像进行随机组合、缩放、裁剪等操作,以增进数据的多样性并提升模型的泛化能力。骨干网络是 YOLOv7 的核心部分,由 CBS、ELAN(Effective Long-Rang Aggregation Network)和 MPC-B 模块组成,负责高效地提取图像特征。CBS 模块利用卷积层、批量归一化和 SiLU 激活函数进行特征提取,以提高网络的训练稳定性和非线性建模能力。MPC-B 模块结合池化和卷积,优化局部特征提取和下采样,增强了模型的感知能力。ELAN 模块通过其网络结构优化 GPU 效率,降低内存需求,并通过梯度技术和短连接技术加强特征提取和提高训练效率。

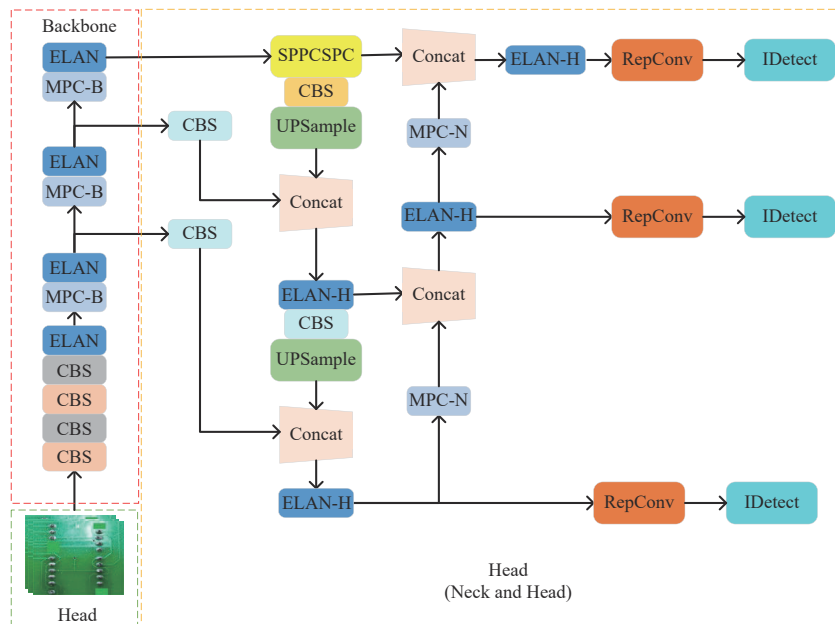


图1 YOLOv7网络结构图

Fig. 1 Diagram of YOLOv7 network architecture

颈部通过多路径方式融合不同尺度的特征,为预测头提供综合特征信息,使模型能够同时识别图像中的大目标和小目标。颈部由多个模块组成,包括 SPPCSPC (Spatial Pyramid Pooling, Cross Stage Partial Channel) 模块的空间金字塔池化^[13],用于扩展感受野和提高上下文信息利用;ELAN-H 模块和 MPC-N 模块进行更深层次的特征提取和融合;UPSample 模块通过插值减少计算量^[14];RepConv 模块通过重参数化改善网络性能并节省空间。

在模型的预测端,整体损失函数包含位置损失、置信度损失和分类损失 3 部分,通过 BCEWithLogitsLoss 处理分类问题,用 CIoU 损失函数优化定位框精度,以提高检测准确率^[15]。YOLOv7 的这些组件共同工作,形成了一个高效且准确的目标检测系统。

3 改进 YOLOv7

3.1 小目标检测层

YOLOv7 采用 3 个不同尺寸的检测头,分别对应 3 个不同尺寸的卷积层,每个卷积层产生不同尺度的特征图,大小分别为 $80 \times 80 \times 128$ 、 $40 \times 40 \times 256$ 、 $20 \times 20 \times 512$ 。浅层特征图具有较高的空间分辨率,适用于捕捉小目标的详细信息。本

文研究的焊点缺陷属于密集型小目标检测,部分焊点目标尺寸较小,检测效果较差,因此添加一个输出大小为 $160 \times 160 \times 64$ 的小目标检测层,其单个网格视野大小为 4×4 ,最小可检测尺度为宽高大于 4 像素的目标。4 个不同尺度的特征融合在模型准确性、模型鲁棒性方面均有不同程度的提高,更加适用于小目标焊点缺陷检测。工业 CCD 为 48 像素/mm,理论最小检测焊点宽高为 $1/12$ mm,根据正样本匹配策略及不同尺度特征融合效果,检测焊点宽高大于 $1/4$ mm,效果最好。

YOLOv7 的官方预设锚框参数相对于焊点数据集存在相当大的偏差,在训练过程中收敛速度较慢,检测精度较低。引入 K-means++ 算法可以有效避免局部最优解问题,与传统 K-means 算法相比,K-means++ 算法在第一步选取聚类中心时不再是一次性随机取样,而是使聚类中心保持相对较远距离以避免不同聚类中心位于同一类内,寻找全局最优解。

3.2 无参注意力机制

考虑到某些特征信息未被充分利用,因此在 Backbone 和 Head 连接处加入无参注意力 (SimAM) 模块。人体视觉中不同信息规模的神经元放电模式不同,信息丰富的神经元往往会抑制周围神经元,也就是具有更重要的关注度。Yang 等人

进一步推导出能量函数快速封闭形式的解,并基于此提出了 SimAM 注意力机制^[16]。与现有的空间注意力模块和通道注意力模块不同, SimAM 能够对不同位置、不同通道区别对待,即能够为特征图提供三维注意力权重,评估每个位置、每个通道的重要性。其最小能量函数如式(1)~(3)所示:

$$e_i^* = \frac{4(\sigma^2 + \lambda)}{4(t - u)^2 + 2\lambda + 2\sigma^2}, \quad (1)$$

$$u = \frac{1}{M} \sum_{i=1}^M x_i, \quad (2)$$

$$\sigma^2 = \frac{1}{M} \sum_{i=1}^M (x_i - u)^2, \quad (3)$$

其中: e_i^* 是每个神经元的最小能量函数, t 是输入特征 X 的值, λ 是正则化系数, u 是通道上所有神经元的均值, σ^2 是所有神经元的方差。代表神经元与周围神经元差异的 t 在分母,故最小能量函数数值越低,该神经元与周围神经元差异越大,重要性越高。而注意力机制需要为特征进行增强操作,用 Sigmoid 激活函数限制最小能量函数的数值大小,公式如式(4)所示:

$$\tilde{X} = \text{Sigmoid}\left(\frac{1}{E}\right) \odot X, \quad (4)$$

其中: X 为输入特征, \tilde{X} 为经过 Sigmoid 增强后的输出特征。在此基础上得到增强后的注意力特征图,且没有增添额外的网络参数,如图 2 所示。

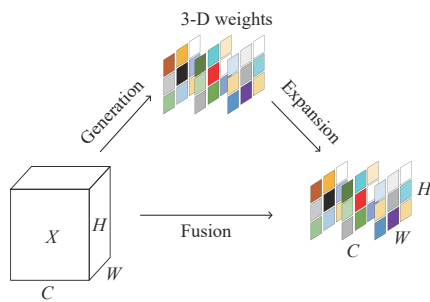


图 2 SimAM 注意力机制结构

Fig. 2 Structure of SimAM attention mechanism

3.3 轻量化

为削减冗余运算和减少内存访问次数,引入部分卷积 PConv 代替 ELAN 中的部分卷积核大小为 3×3 、步数为 1 的经典卷积。如图 3 所示,其基本原理是利用整个系统存在大量冗余,因此只对输入的部分通道进行卷积,剩下的通道保持不变,作

为整体模块的输出结果,从而达到更低的延时效果,更好地利用设备的计算能力^[17]。PConv 模块的浮点运算次数(FLOPs)为 $h \times w \times k^2 \times c_p^2$,内存访问次数为 $h \times w \times 2c_p$,其中 c_p 为输入通道数的 1/4,整个 PConv 模块的 FLOPs 是传统 Conv 模块的 1/16,内存访问次数是经典 Conv 模块的 1/4。

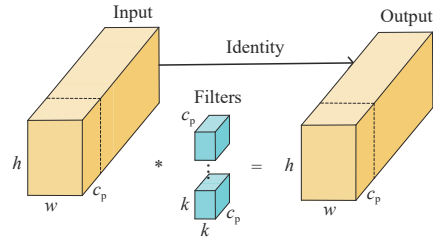


图 3 PConv 模块原理图

Fig. 3 Schematic diagram of PConv module

模型颈部对于从主干提取到的不同尺度特征图进行特征融合和提取。GiraffeDet 网络^[18]有一个非常长的颈部网络,更有利于不同空间尺度融合和密集信息交换,因此利用颈部网络 Generalized-FPN 中的特征提取融合网络替换原 YOLOv7 模型中进行深层次特征变换融合的 ELAN-H 模块,直接对检测头进行输出,有效减少了模型参数量和计算量,并提高了检测精度。

如图 4 所示,算法的具体流程如下:两个卷积层分别对输入特征图进行卷积操作,将输入通道分裂为两部分。定义的卷积块序列逐个处理输入特征图的每个部分,如多次卷积、激活函数、SPP

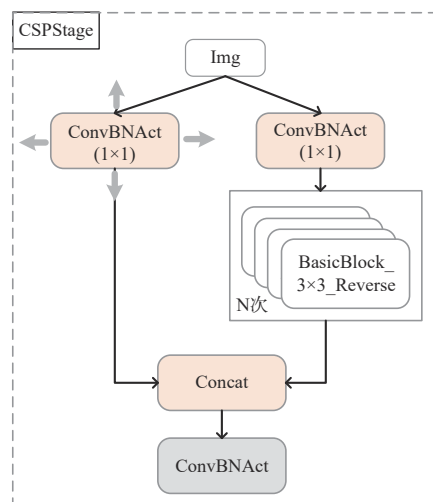


图 4 CSPStage 网络结构

Fig. 4 CSPStage network structure

池化等。最后经过一系列卷积操作后将所有分裂的特征图连接在一起形成一个深且大的特征图,并将连接后的特征图的通道数转换成输出通道数。CSP架构下的算法在保持高性能的同时具有较低的运算和内存消耗。通过CSP架构和SPP池化,能够有效地融合多尺度特征信息。

3.4 损失函数改进

NWD是一种基于优化理论中的 Wasserstein 距离的新型损失函数,旨在解决小目标检测中传统IoU度量的不足^[19]。NWD采用二维高斯分布来建模边界框,进而使用归一化的 Wasserstein 距离来度量这些分布之间的相似性。与IoU不同,Wasserstein距离即使在分布之间没有重叠的情况下也能有效衡量相似度。这种方法在处理非重叠或相互包含的边界框时尤为有效,提高了小目标检测的准确性和鲁棒性。

NWD通过结合 Wasserstein 距离和归一化操作,有效降低了对目标尺寸和位置差异的敏感性。这种归一化处理是NWD的关键,因为它将 Wasserstein 距离转换为介于0和1之间的相似度量,与传统IoU度量更为一致。NWD公式表达为:

$$W_2^2(N_a, N_b) = \left\| \begin{pmatrix} \left[cx_a, cy_a, \frac{w_a}{2}, \frac{h_a}{2} \right]^T \\ \left[cx_b, cy_b, \frac{w_b}{2}, \frac{h_b}{2} \right]^T \end{pmatrix} \right\|_2^2, \quad (5)$$

$$NWD(N_a, N_b) = \exp\left(-\frac{\sqrt{W_2^2(N_a, N_b)}}{C}\right), \quad (6)$$

$$Loss = \alpha \times CIoU + (1 - \alpha) \times NWD, \quad (7)$$

式中: C 为目标类别数6, N_a 和 N_b 表示 $A=(cx_a, cy_a, w_a, h_a)$ 和 $B=(cx_b, cy_b, w_b, h_b)$ 建模的高斯分布,Loss是位置损失函数, α 为CIoU损失函数的权重。当数据集中的中小微目标较多时,应该适当下调CIoU函数权重以取得更好的回归效果。本模型中 α 设置为0.5。

NWD的引入是小目标检测领域的一项进步,它在网络训练过程中提供了更稳定的梯度,加快了收敛速度,并减少了对训练数据选择的依赖。通过添加小目标检测层、引入SimAM注意力机制、利用PConv部分卷积重构ELAN模块,使用Generalized-FPN提取特征信息,融合NWD和CIoU损失函数,改进后的模型结构图如图5所示。

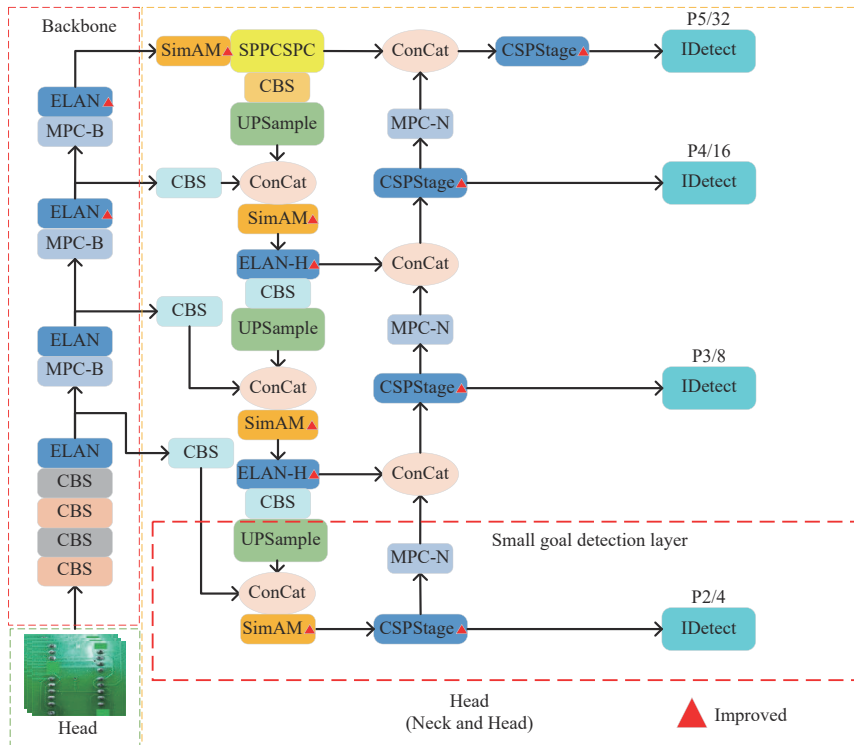


图 5 改进后的YOLOv7网络结构图
 Fig. 5 Diagram of improved YOLOv7 network structure

4 实验与结果分析

4.1 实验环境

实验选用云服务器平台,选取与 YOLO 系列算法发布官方相同的配置,实验采用的环境和训练基本参数如下:主机 CPU 为 Xeon(R) Platinum 8255C,显卡为 2080Ti (11 GB 显存),操作系统选用 Ubuntu18.04 版本,深度学习框架选用 Pytorch1.8.1,Cuda 版本为 11.1,实验迭代次数为 300,batch-size 为 8,img-size 为 640×640 ,初始学习率为 0.01,权重衰减系数为 0.0005。

4.2 数据集的准备

本研究中收集了在实习和各种竞赛中手工焊接的电路板图片。为了增加数据的多样性,对这些图片进行了多种处理,包括水平翻转、目标裁剪、曝光度增强、对比度调整、旋转缩放以及引入高斯噪声等方法,共获得了 1 250 张经过数据增强的图片。使用 LabelImg 软件对这些图片进行标注,创建了一个包含 35 089 个目标的数据集。图 6 展示了部分图片中放大后的焊点,表 1 详细列出了每类目标的具体数量。

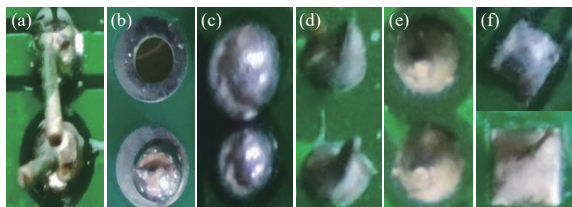


图 6 数据集图片示例。(a)桥连;(b)少锡;(c)多锡;(d)毛刺;(e)合格;(f)方形。

Fig. 6 Example of dataset images. (a) Bridge; (b) Insuff; (c) Excess; (d) Spike; (e) Accept; (f) Square.

表 1 数据集中目标分类及数量

Tab. 1 Classification and number of targets in the data set

| Defect type | Training set | Validation set | Total |
|-------------|--------------|----------------|--------|
| Bridge | 912 | 314 | 1 226 |
| Insuff | 10 408 | 1 515 | 10 559 |
| Excess | 2 040 | 593 | 2 633 |
| Spike | 4 530 | 1 230 | 5 760 |
| Accept | 10 136 | 3 091 | 13 227 |
| Square | 1 284 | 400 | 1 684 |

4.3 实验评估指标

为全面评价模型效果,本实验采用精确率 (Precision, P)、召回率 (Recall, R) 及平均精度均值 (mean Average Precision, mAP) 评价模型精准度,采用参数量 (Params)、浮点运算次数 (GFLOPs) 来评价模型复杂程度。计算公式如式 (8) 和式 (9) 所示:

$$P = \frac{TP}{TP + FP}, R = \frac{TP}{TP + FN}, \quad (8)$$

$$mAP = \frac{\sum_0^1 P(R) dR}{N}, \quad (9)$$

式中: P 为精确率, R 为召回率,TP 代表模型正确分类的焊点数量,FP 代表模型错误分类为特定类型的焊点但实际上不属于那个类型的数量,FN 代表模型未能正确分类的实际焊点数量。精确度表示在所有被模型分类为某一类焊点的案例中,确实属于那个类别的比例。召回率衡量了在所有真实属于某一类别的焊点中,被模型正确分类的比例。mAP 衡量了模型在处理多分类时的整体效果。

参数量指的是模型中可训练参数的总数,反映了模型的规模和复杂度。GFLOPs 是执行模型一次前向传递所需的浮点运算次数。通常情况下,在相同准确率的模型中,参数量和 GFLOPs 较小通常意味着模型更加轻量化和高效。

4.4 实验结果与分析

4.4.1 消融实验

本文提出了 5 种针对小目标焊点缺陷检测的改进 YOLOv7 模型方法:增加小目标检测层、引入 SimAM 无参注意力机制、使用 PConv 替换部分经典 Conv、使用 Generalized-FPN 提取多尺度特征信息、结合 NWD 改进 CIoU 损失函数。为进一步验证每个改进方法的有效性,在相同实验环境、相同数据集下进行消融实验,结果如表 2 所示。

由表 2 可知,单独加入小目标检测层后对焊点的识别精度大幅度提升,平均精度提升了 3.0%,精确率有所降低而召回率提高了 3.4%,这意味着更少的假阴性,更多的假阳性,一定程度上提高了模型的泛化能力。单独加入 SimAM 注意力机制后,模型的计算量并没有大幅度提升,说明该注意力模块足够轻量,同时在平均精度方面相比于小目标检测层提升得更多,提升了

表 2 消融实验

Tab. 2 Ablation experiments

| Small goal layer | SimAM | PConv | Generalized-FPN | NWD | mAP/% | P/% | R/% | Params/M | GFLOPs/G |
|------------------|-------|-------|-----------------|-----|-------|------|------|----------|----------|
| | | | | | 85.2 | 82.6 | 82.5 | 37.22 | 105.2 |
| ✓ | | | | | 88.2 | 82.4 | 85.9 | 37.78 | 119.6 |
| | ✓ | | | | 88.6 | 83.6 | 83.6 | 37.68 | 105.4 |
| | | ✓ | | | 86.6 | 82.4 | 83.1 | 32.08 | 87.2 |
| | | | ✓ | | 87.4 | 83.6 | 83.5 | 26.35 | 85.2 |
| | | | | ✓ | 86.0 | 82.1 | 83.7 | 37.22 | 105.2 |
| ✓ | ✓ | | | | 89.2 | 84.5 | 84.9 | 37.77 | 119.6 |
| ✓ | ✓ | ✓ | | | 89.6 | 83.3 | 86.1 | 32.95 | 107.9 |
| ✓ | ✓ | ✓ | ✓ | | 90.0 | 84.1 | 86.4 | 23.70 | 85.4 |
| ✓ | ✓ | ✓ | ✓ | ✓ | 90.3 | 84.3 | 85.7 | 23.70 | 85.4 |

3.4%。使用 PConv 模块代替部分经典 Conv 模块,使模型参数量降低了 5.14M,GFLOPs 降低了 18G。单独更换颈部的特征融合网络后,平均精度提升 2.2%,参数量和浮点运算量大幅降低。结合 NWD 损失函数改进后,模型在检测精度方面略有提升,但稳定性和对小目标的检测性能都有提高。

在多项改进中,加入小目标检测层和加入 SimAM 注意力机制后,平均精度进一步提升,并且得到了相对平衡的精确率和召回率,但模型参数量和计算量都有增加。加入 PConv 模块缩减模型参数量和计算量,精度仅提高 0.4%。在更换颈部的特征融合网络后,模型的平均检测精度达到了 90.0%,模型参数量降低了 36.3%,浮点运算量降低了 18.8%。加入 NWD 损失函数改进后,平均精度有了进一步提升,达到了 90.3%,相较于原模型提高了 5.1%;其收敛速度有了极大提升,且在训练过程中内存占用变低,平均内存占用从 9.98G 降低到 5.88G,降低了 41.4%。

4.4.2 损失函数对比

为体现结合 CIoU 与 NWD 损失函数在收敛速度和稳定性方面的提升,选取原始 YOLOv7 模型、只更换损失函数后的 YOLOv7 模型、改进但不更换损失函数的 YOLOv7 模型、改进且更换损失函数后的 YOLOv7 模型在训练过程中验证集的损失函数,并分别用曲线 A、B、C、D 表示,如图 7 所示。

由图 7 可知,在训练过程中,4 种算法皆在前 50 次迭代中收敛速度较快,在约 250 次趋于稳

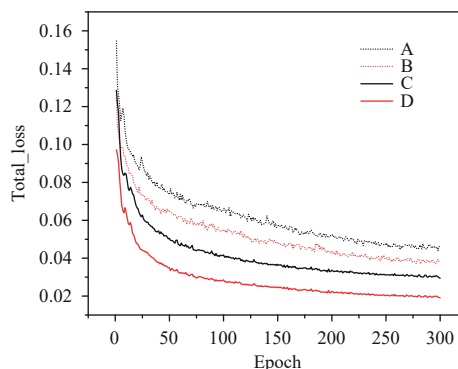


图 7 损失函数对比

Fig. 7 Loss function comparison

定,原模型最终的总损失稳定在约 0.046。只改进损失函数,模型的总损失稳定在约 0.038,加入小目标检测层但不改进损失函数后模型的总损失稳定在约 0.029,而改进损失函数后模型的总损失进一步降低到约 0.019。可见,利用 NWD 改进 CIoU 损失函数以及其他改进都能起到加快模型收敛速度的效果,且更加稳定,测试过程中未观察到明显的过拟合或欠拟合现象。

4.4.3 多种模型检测结果对比

改进后的 YOLOv7 模型相较于原模型在小目标检测精度方面有了明显的提升。为进一步验证改进的有效性,以 mAP、Precision(P)、Recall(R)、Params、GFLOPs 作为评价指标,将本文算法与 YOLO 系列算法在统一配置环境下对同一数据集进行训练验证,结果如表 3 所示。从表 3 可以看到,相较于其他模型,原 YOLOv7 模型虽然精度较高,但参数量和计算量都较大,且召回率较

表 3 不同算法模型的检测结果

Tab. 3 Detection results of different algorithm models

| Model | mAP/ % | P/% | R/% | Params/ GFLOPs/ M G | |
|----------------------|-------------|------|------|------------------------|-------|
| | YOLOv3-tiny | 66.8 | 68.7 | 60.6 | 8.68 |
| YOLOv3 | 82.8 | 80.9 | 80.7 | 61.55 | 155.3 |
| YOLOv5s | 80.4 | 75.8 | 78.2 | 7.36 | 16.0 |
| YOLOv5m | 82.3 | 77.2 | 81.9 | 20.89 | 48.3 |
| YOLOv7 (original) | 85.2 | 82.6 | 82.5 | 37.22 | 105.2 |
| YOLOv8s | 84.2 | 79.5 | 79.5 | 11.14 | 28.7 |
| YOLOv8m | 86.0 | 80.0 | 82.2 | 25.86 | 79.1 |
| YOLOv7 (improved) | 90.3 | 84.3 | 85.7 | 23.70 | 85.4 |

低。YOLOv8m 算法在平均精度上比原 YOLOv7 模型高 0.8%，但其精确率和召回率较低，其损失函数 DFL 在处理不同尺度输出特征图时存在不足。改进后的模型相较于原模型有了一定的轻量化，平均精度、精确率、召回率都有了较大提升，在轻量化程度和准确率上达到了一定平衡。

参 考 文 献:

- [1] SITU G H. Deep holography [J]. *Light: Advanced Manufacturing*, 2022, 3(2): 278-300.
- [2] AGRAWAL P, GIRSHICK R, MALIK J. Analyzing the performance of multilayer neural networks for object recognition [C]. 13th European Conference on Computer Vision. Zurich, Switzerland: Springer, 2014: 329-344.
- [3] GIRSHICK R. Fast R-CNN [C]//*Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago: IEEE, 2015: 1440-1448.
- [4] REN S Q, HE K M, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [C]//*Proceedings of the 28th International Conference on Neural Information Processing Systems*. Montreal: MIT Press, 2015: 91-99.
- [5] LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]. 14th European Conference on Computer Vision. Amsterdam: Springer, 2016: 21-37.
- [6] REDMON J, FARHADI A. YOLO9000: better, faster, stronger [C]//*Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017: 6517-6525.
- [7] PARK J M, YOO Y H, KIM U H, et al. D³PointNet: dual-level defect detection pointNet for solder paste printer in surface mount technology [J]. *IEEE Access*, 2020, 8: 140310-140322.
- [8] 卢子册,刘小芳,王德伟. 基于改进 YOLOv8 的 PCB 焊点语义分割方法[J]. *无线电工程*, 2024, 54(7): 1614-1621.
LU Z C, LIU X F, WANG D W. PCB solder joint semantic segmentation method based on improved YOLOv8 [J]. *Radio Engineering*, 2024, 54(7): 1614-1621. (in Chinese)
- [9] 严蓬辉,陈绪兵,彭伊丽,等. 基于改进 YOLOv5s 的激光软钎焊焊点缺陷检测算法[J]. *激光与光电子学进展*, 2024, 61(8): 0812005.
YAN P H, CHEN X B, PENG Y L, et al. Algorithm for detecting laser soldering point defect based on improved YOLOv5s [J]. *Laser & Optoelectronics Progress*, 2024, 61(8): 0812005. (in Chinese)

5 结 论

本文针对 YOLOv7 算法进行了一系列改进，旨在提高模型对小目标焊点缺陷的检测能力，增加实用性。通过添加小目标检测层，提高了对浅层特征信息提取能力，提高了小目标检测精度。在 Head 区引入轻量型 SimAM 注意力机制，为特征图分配三维注意力权重且不增加额外的参数，提高了模型的精度。将部分 Conv 模块替换为 PConv 模块，在减少模型参数量和计算量的同时避免了频繁的内存访问。利用 Generalized-FPN 网络对不同尺度特征信息进行深度融合提取，在轻量化的同时提高检测精度。结合 NWD 损失函数改进 CIoU 损失函数，加快了模型收敛速度，减少了内存占用。实验结果证明，改进后的模型 mAP 为 90.3%，相比于原始模型的 mAP 提高了 5.1%，召回率提高了 3.2%，参数量降低了 36.3%，平均内存占用降低了 41.4%。这些改进提高了模型的轻量化程度和检测精度，有效提升了模型在小目标焊点缺陷检测上的实用性。

- [10] 何智成,王振兴. 基于改进YOLOv2的白车身焊点检测方法[J]. 计算机工程,2020,46(11):246-254.
HE Z C, WANG Z X. Welding spot detection method for body in white based on improved YOLOv2 [J]. *Computer Engineering*, 2020, 46(11): 246-254. (in Chinese)
- [11] 刘涛,张涛. 基于GhostNet-YOLOv4算法的印刷电路板缺陷检测[J]. 电子测量技术,2022,45(16):61-70.
LIU T, ZHANG T. Defect detection of printed circuit board based on GhostNet-YOLOv4 algorithm [J]. *Electronic Measurement Technology*, 2022, 45(16): 61-70. (in Chinese)
- [12] 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*. Vancouver: IEEE, 2023: 7464-7475.
- [13] HU C P, BAI X, QI L, *et al.* Vehicle color recognition with spatial pyramid deep learning [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2015, 16(5): 2925-2934.
- [14] ZUO C, QIAN J M, FENG S J, *et al.* Deep learning in optical metrology: a review [J]. *Light: Science & Applications*, 2022, 11(1): 39.
- [15] 付惠琛,高军伟,车鲁阳. 基于改进YOLOv7的口罩佩戴检测[J]. 液晶与显示,2023,38(8):1139-1147.
FU H C, GAO J W, CHE L Y. Mask wearing detection based on improved YOLOv7 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2023, 38(8): 1139-1147. (in Chinese)
- [16] 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*. Online: ICML, 2021: 11863-11874.
- [17] CHEN J R, KAO S H, HE H, *et al.* Run, don't walk: chasing higher FLOPS for faster neural networks [C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Vancouver: IEEE, 2023: 12021-12031.
- [18] JIANG Y Q, TAN Z Y, WANG J Y, *et al.* GiraffeDet: a heavy-neck paradigm for object detection [C]. The Tenth International Conference on Learning Representations. Online: ICLR, 2022.
- [19] WANG J W, XU C, YANG W, *et al.* A normalized Gaussian Wasserstein distance for tiny object detection [J/OL]. *arXiv*,2021: 2110.13389.

作者简介:



刘兆龙,男,硕士研究生,2022年于青岛大学获得学士学位,主要从事计算机视觉与模式识别方面的研究。E-mail: 2896145142@qq.com



高军伟,男,博士,教授,2003年于中国铁道科学研究院获得博士学位,主要从事模式识别及智能控制方面的研究。E-mail: qdgao163@163.com