文章编号：1003－1251(2026)05－0031－10

基于 YOLOv8s 的手机元件表面小目标缺陷检测算法研究

彭 方，文 峰，田 闻

（沈阳理工大学 信息科学与工程学院，沈阳 110159）

摘 要：针对 YOLOv8 在手机元件表面缺陷检测中的实际应用，发现其在处理大量小目标时存在检测精度不足的问题。检测性能不佳的原因主要有两个：其一是小目标通常缺乏足够的外观信息，且目标与复杂背景之间对比度较低，导致模型提取特征时小目标信息丢失严重，检测头获得的有效信息不足，难以做出良好判断；其二是以 IoU 为主体的衡量边界框重叠度的指标对小目标较为敏感。为此，本文在 YOLOv8s 基础上提出一系列改进方法，以增强其对小目标的感知能力，提升模型在实际检测任务中的精度与鲁棒性。首先在主干网络中的卷积模块后添加具有切片操作的 SimAM 注意力模块（SWS），有效改善提取特征后小目标信息丢失严重的问题；其次在原有的 SPPF 中添加全局最大池化和全局平均池化分支，同时在 SPPF 块中嵌入 SWS，以给模型提供丰富的信息和重要区域聚焦能力，帮助模型更好地做出判断；最后在原有算法的 CIoU 损失函数中加入归一化高斯 Wasserstein 距离（NWD），降低 CIoU 对小目标敏感的问题，使模型能够更加稳定地收敛。在一种手机元件上的实验表明，改进后的算法相较原有算法其 F1-Score、mAP50、mAP50∶95 分别提升了 4.0%、3.9%、2.4%，证明了改进算法的有效性。此外改进后算法相较原算法其 GFLOPs 只增加了 0.2，保障了实际应用对检测速度的要求。

关 键 词：手机元件；小目标检测；缺陷检测；YOLOv8；损失函数；注意力机制

中图分类号：TP391.4 **文献标志码**：A **DOI**：10.3969/j.issn.1003－1251.2026.05.005

Research on a Defect Detection Algorithm for Small Targets on Mobile Phone Components Surface Based on YOLOv8s

PENG Fang，WEN Feng，TIAN Wen

（Shenyang Ligong University，Shenyang 110159，China）

Abstract：In the practical application of YOLOv8 for surface defect detection on mobile phone components，low detection accuracy for numerous small targets was identified. Two main reasons were found. Firstly，small targets often lack sufficient visual information and have low contrast with complex backgrounds，leading to significant information loss during feature extraction. Secondly，the Intersection over Union（IoU），a metric for measuring bounding box overlap，is sensitive to small targets. To address these issues，a series of improvements were proposed based on YOLOv8s to enhance its perception of small targets and improve the accuracy and robustness of the model in practical detection tasks. Firstly，a SimAM attention module with slicing（SWS）was added after the Conv block in the Backbone，effectively addressing the issue of severe information loss for small

收稿日期：2025－06－08

基金项目：国家重点研发计划“社会治理与智慧社会科技支撑”重点专项(2022YFC3302500)

作者简介：彭方(2000—)，男，硕士研究生；文峰(1977—)，通信作者，男，教授，博士。

targets after feature extraction. Secondly, global max pooling and global average pooling branches were added to the original SPPF, and SWS was embedded in the SPPF block to provide the model with rich information and improved focus on important areas for better decision-making. Lastly, Normal Wasserstein Distance (NWD) was added to the original CIoU loss function, reducing the sensitivity of CIoU to small targets and allowing the model to converge more stably. Experimental results on a real mobile phone component dataset showed that the improved algorithm achieved an increase of 4.0%, 3.9%, and 2.4% in F1-Score, mAP50, and mAP50-95, respectively, demonstrating the effectiveness of the improvements. Additionally, the improved algorithm's GFLOPs increased by only 0.2, ensuring that the detection speed met industrial real-time requirements.

Key words: mobile phone components; small object detection; defect detection; YOLOv8; loss function; Attention Mechanism

随着智能终端的广泛应用,手机元件在生产过程中的质量控制日益重要,其中表面缺陷检测作为保障产品可靠性与推动生产自动化的关键环节,具有重要的工程价值。近年来,深度学习驱动的目标检测算法在工业视觉检测中取得广泛应用,但在手机元件表面缺陷检测中受限于缺陷目标普遍存在尺寸小、形态不规则、对比度低等特性,现有算法在检测精度与鲁棒性方面仍难以满足实际需求[1]。其原因主要包括两方面:一是模型在特征提取过程中小目标信息损失严重,影响后续检测精度;二是当前普遍采用的交并比(IoU)及其变体作为回归损失函数对小目标的定位误差容忍度较低,影响其回归精度与定位性能。

近年来,目标检测算法发展迅速,YOLO 系列以其高效性与高精度被广泛应用于智能制造,并已持续迭代至 YOLOv12。

研究表明,YOLOv9 在小目标检测与效率上表现欠佳,YOLOv10 在复杂场景下准确性不足,YOLOv11 在精度与效率间表现均衡但在小目标密集场景中速度较慢,而 YOLOv12 因架构复杂带来额外计算开销,性能提升有限。

针对小目标提取特征后信息不足的问题,杨大为等[2]通过在 YOLOv7 中增加小目标检测层以增强小目标的感知能力;赵子艺等[3]在 YOLOv8 的快速空间金字塔(SPPF)模块中引入大型可分离核注意力(LSKA),以抑制背景干扰并提取更多对小目标有用的信息;于敖泽等[4]引入分块的通道注意力机制,计算通道注意力时考虑了空间不同位置的差异,增强特征表达能力;王晓青等[5]与 Li 等[6]采用 SimAM 分别增强 YOLOv7、YOLOv8 对关键区域的关注;Wang 等[7]则使用归一化注意力机制(NAM)与轻量化上采样器提升颈部的特征融合效率;Sun 等[8]在 YOLOv7 的颈部引入 GE 注意力,以增强上下文建模能力,提升小目标表达能力。

上述方法主要通过添加注意力机制、增强特征提取路径、增加小目标检测层来缓解小目标信息在特征提取时信息损失严重的问题,但仍存在以下不足:首先,增加检测层或引入复杂注意力结构虽能增强小目标表征能力,但会显著提升模型计算开销,影响效率;其次,注意力机制在小目标与背景相似情况下效果不显著;最后,如果主干提取的特征不足,在颈部加强特征融合或加入注意力机制的效果会被削弱。

针对手机元件表面缺陷的实测中,发现 YOLOv8n/s[9]在检测精度上优于 YOLOv11 和 YOLOv12。考虑到手机元件缺陷具有小目标多、数据难获取、样本量有限等特点,结合模型精度、速度与适配性等多方面因素,本文选用轻量级、性能均衡的 YOLOv8s 作为改进的基准模型。

针对 YOLOv8 特征提取阶段易造成小目标信息丢失的问题[9],本文提出两项改进措施以提升小目标检测性能:首先,在主干中引入具有切片机制的注意力模块 SimAM(SWS),以保留更多有利于小目标的细节信息;其次,设计改进的多尺度特征融合模块 AugSPPF_SWS(augmented SPPF with SWS),在原有 SPPF 基础上加入全局最大池化和全局平均池化,并嵌入 SWS,以提升模型的全局感知与重要区域聚焦能力。通过在主干部分做出改进,有助于在早期保留更多重要信息。整体结构在增强特征表达能力的同时仍能保持轻量性,有效提升小目标检测精度。

损失函数设计方面,陈扬等[10]通过改进 GIoU,缓解预测框与真实框无交集时的梯度消失问题;赵子艺等[3]采用 WIoU 以降低低质量样本的负面影响;王晓青等[5]与 Sun 等[8]使用归一化高斯 Wasserstein 距离(NWD)[11]衡量边界框的相

似度,提升回归精度并减弱对小目标误差的敏感性。然而上述做法仍存在以下不足:首先,部分损失函数对超参数依赖性强,且在预测框与真实框无交集情况下仍存在梯度缺失问题;其次,直接使用 NWD 作为预测框与真实框的相似度衡量,未显式考虑两者的面积重叠度,可能对最终定位精度造成影响。

因此,本文针对 YOLOv8 使用的 CIoU 边界框回归损失[12]在小目标检测中对定位误差敏感的问题,融合 NWD 与 CIoU 构建改进的边框回归损失函数,保留了 CIoU 对面积重叠的度量能力,同时利用 NWD 提升小目标检测的精度与鲁棒性。

1　YOLOv8 算法

YOLOv8 网络结构如图 1 所示。

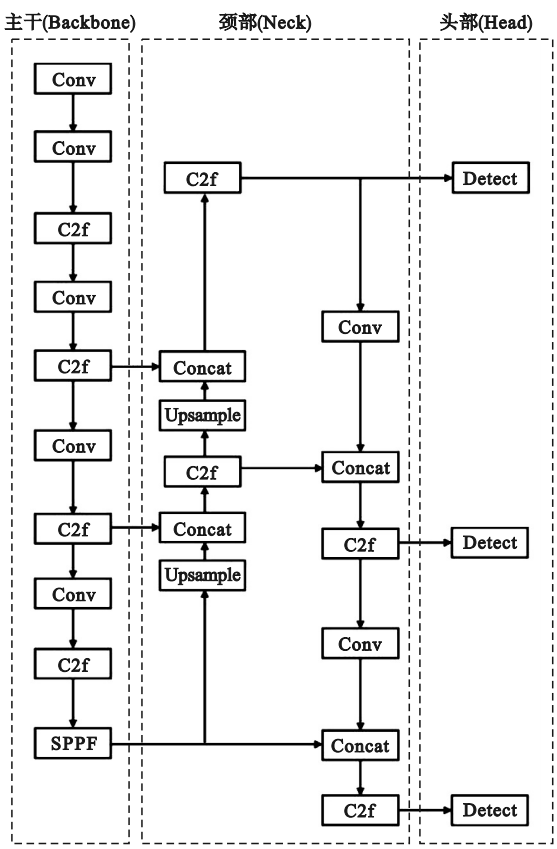


图 1　YOLOv8 网络结构图

Fig. 1　YOLOv8 network structure diagram

从图 1 可以看到,YOLOv8 由卷积(Conv)模块、C2f、SPPF 组合连接为主干,其颈部利用特征金字塔网络的思想,将主干提取到的多尺度特征图进行融合,其中 Upsample 表示上采样,Concat 表示拼接操作。然后将融合后的多尺度特征图送入检测头。其各个模块的结构如图 2 所示。

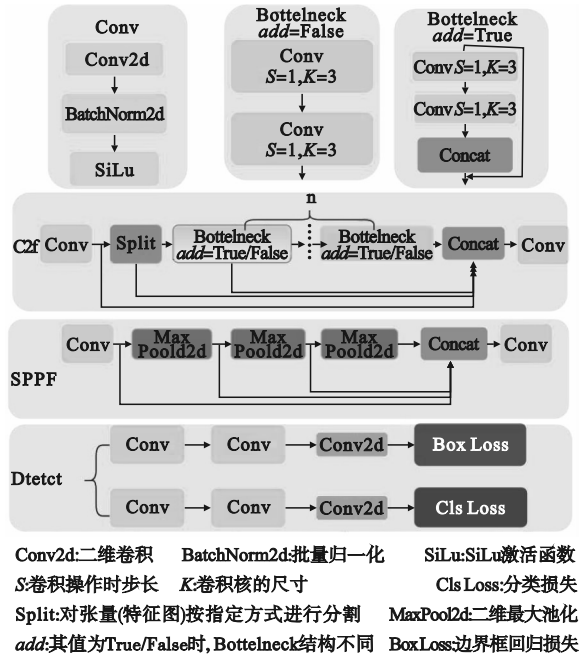


图 2　YOLOv8 各模块结构

Fig. 2　Structure of each module in YOLOv8

2　本文算法

本文对 YOLOv8 改进后的模型结构如图 3 所示,主要通过在主干中的 Conv 块后面添加 SWS,加强模型对重要信息的关注,改善提取特征时小目标信息丢失严重的问题;使用改进后的 AugSPPF_SWS 替换原有的 SPPF,使其在融合多尺度信息的基础上融合全局背景信息和全局视角信息,帮助模型更好地做出判断;最后,其头部使用对尺度不敏感且能更好度量重叠目标相似性的 NWD 以改进原有的 CIoU 损失,使模型更稳定地收敛。

2.1　SWS

SimAM 是一种无需引入额外参数的轻量级注意力机制,可以赋予特征图三维的权重。该方法基于神经科学中对神经元能量模型的启发,通过构建能量函数度量特征值的重要性,从而对特征图进行显著性加权。对特征图 $\boldsymbol{X} \in \mathbf{R}^{B\times C\times H\times W}$($B$、$C$、$H$、$W$ 分别表示特征图的批量、通道、高、宽)的 SimAM 加权操作如式(1)所示。

$$\widetilde{\boldsymbol{X}} = \sigma\left(\frac{1}{E}\right)\odot \boldsymbol{X} \tag{1}$$

式中,$\widetilde{\boldsymbol{X}}$ 表示 $\boldsymbol{X}$ 经过 SimAM 注意力加权后的特征图;σ 为 Sigmoid 函数;⊙表示张量逐元素相乘;$\boldsymbol{E}$ 为能量项,其计算公式如式(2)所示。

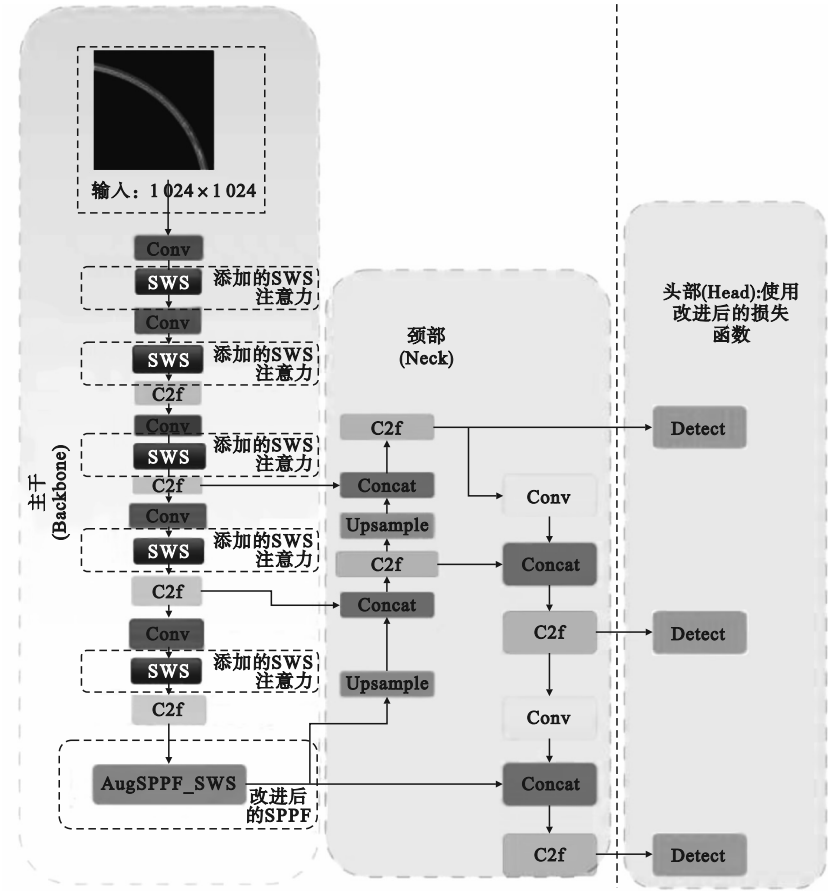


图 3 YOLOv8 改进后网络结构图

Fig. 3 Network architecture of improved YOLOv8

$$E=\frac{4(\hat{\sigma}^2+\lambda)}{(X-\hat{\mu})^2+2\hat{\sigma}^2+2\lambda} \tag{2}$$

式中：$\hat{\mu}\in\mathbf{R}^{B\times C\times 1\times 1}$、$\hat{\sigma}^2\in\mathbf{R}^{B\times C\times 1\times 1}$，分别为通道维度上的均值和方差；$\lambda$ 是一个常数，防止分母为零并稳定能量函数的数值计算。

尽管 SimAM 在多类目标检测任务中表现出良好的适应性，但其在小目标检测任务中存在显著局限。SimAM 的均值与方差统计是基于空间维度进行的，而小目标在空间上仅占据极小比例，其响应易被其通道内背景平均值所掩盖，导致能量函数计算中未能准确体现其相对差异性，最终导致小目标获得的注意力增强不足。

为缓解上述问题，本文受到于敖泽等[4]在 YOLOv5 中采用局部通道注意力机制以提升对无人机小目标检测性能的启发，提出一种 SWS 机制。在将特征图输入 SimAM 模块之前，先将其划分为 m^2 个不重叠的空间子块，每个子块的宽高为原特征图的 1/m，并对每一子块分别计算局部均值和方差，从而更精细地捕捉局部结构差异。该设计能够有效提升小目标与其邻域之间的特征差异性，使其在能量函数中获得更大的加权，增强对小目标区域的响应强度，且考虑局部差异性的思想也更符合卷积神经网络的局部连接性。SWS 算法流程如图 4 所示。

2.2 AugSPPF_SWS

空间金字塔池化（SPP）通过多尺度池化融合不同感受野信息，提升模型对多尺度目标的感知能力。其高效变体 SPPF 通过连续多次的 5×5 最大池化，显著降低计算复杂度并加快推理速度，其结构如图 2 所示。

尽管 SPPF 能有效捕捉局部多尺度特征，但存在以下局限：首先，缺乏对全局上下文信息的建模，限制了模型对复杂背景与目标关系的理解；其次，特征融合过程中未进行显式的注意力机制调整，导致小目标特征可能被多尺度池化稀释。

为此，本文提出一种改进的多尺度特征融合模块AugSPPF_SWS，其结构如图 5 所示。该模块在 SPPF 基础上引入以下两个关键改进。

1）全局信息增强：在 SPPF 的基础上，增加全局平均池化（Adaptive AvgPool2d）和全局最大池化（Adaptive MaxPool2d）分支，有效捕获图像的全局语义和背景信息，提升模型对复杂场景的适

应能力。

2)小目标特征强化:为充分发挥上一节提出的 SWS 模块在小目标特征增强中的优势,本文在 SPPF 中嵌入 SWS,确保多尺度融合前后均能对关键特征区域进行有效聚焦,显著提升对小目标区域的特征表达能力。

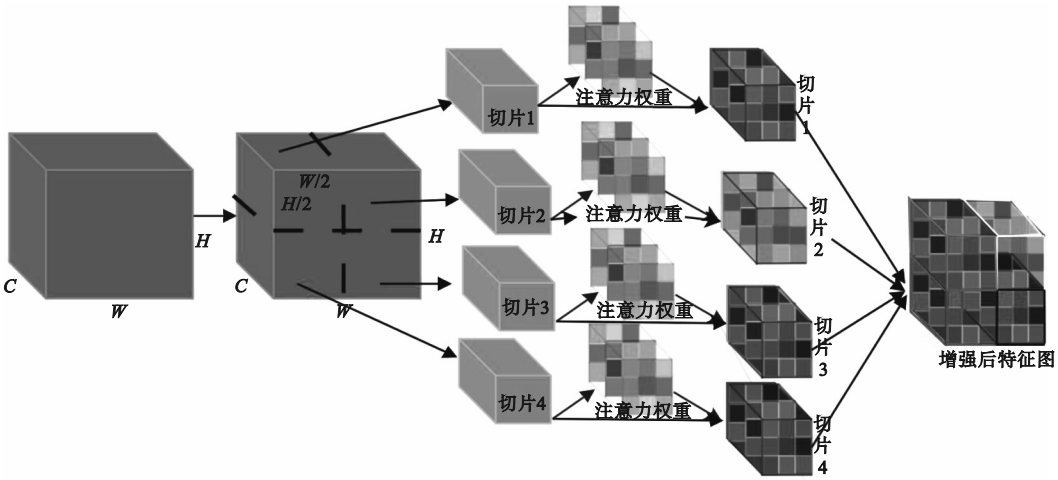


图 4　SWS 的流程

Fig. 4　The SWS process

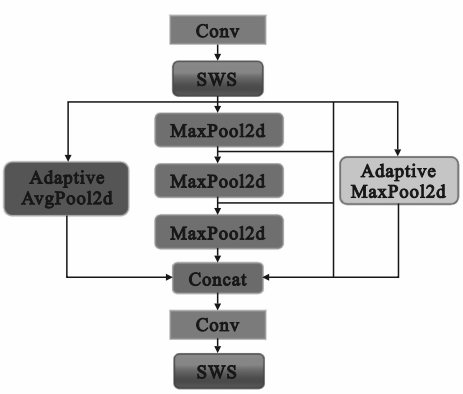


图 5　AugSPPF_SWS 结构图

Fig. 5　AugSPPF_SWS structure diagram

AugSPPF_SWS 结合了多尺度局部特征、全局语义信息与显著性注意力机制,克服了传统 SPPF 在全局上下文和小目标感知方面的不足,实现了对特征图的多维度信息强化,针对小目标检测场景表现出更强的特征区分能力。

2.3　损失函数的改进

在目标检测任务中,回归分支的关键在于学习预测框与真实框之间的几何关系。传统的 IoU 作为常用的匹配度量,仅反映边界框重叠面积的比例,存在两个显著问题:其一是当预测框与真实框无交集时,IoU 的值为 0,无法提供有效梯度;其二是对小目标偏移敏感,易导致损失波动,影响训练稳定性。

YOLOv8 使用了 CIoU 评判预测框定位的准确程度,考虑了预测框和真实边界框的重叠度、中心点距离、长宽比一致性因素[12],其计算过程如式(3)所示。

$$L_{\mathrm{CIoU}} = L_{\mathrm{IoU}} - \frac{d^2}{c^2} - \alpha v \tag{3}$$

式中:L_{IoU}是传统的边界框损失函数,即预测框和真实框的交集与并集面积之比;L_{CIoU}是 YOLOv8 原有的边界框损失函数;d 为预测框与真实框中心点的距离;c 为两者最小外接矩形的对角线距离;α 是一个平衡参数,不参与梯度计算;v 表示两者高宽比的一致性,计算表达式如式(4)所示。

$$v = \frac{4}{\pi^2}\left(\arctan\left(\frac{w_{\mathrm{g}}}{h_{\mathrm{g}}}\right) - \arctan\left(\frac{w_{\mathrm{p}}}{h_{\mathrm{p}}}\right)\right)^2 \tag{4}$$

式中 w_{g}、h_{g}、w_{p}、h_{p} 代表真实框与预测框的高和宽。

尽管 CIoU 增强了几何约束,但其对小目标仍较敏感[11],如图 6 所示,6×6 像素小目标偏移一格导致 L_{CIoU}从 0.51 降至 0.22,而 36×36 像素正常目标仅从 0.89 降至 0.80,反映了其对尺度变化的脆弱性。

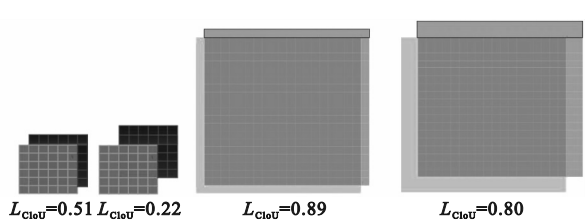


图 6　CIoU 对不同尺度目标的敏感差异

Fig. 6　The sensitivity difference of CIoU for targets of different scales

为缓解此问题,有学者使用 NWD 作为 IoU

的替代[6][9]。该方法首先将预测框与真实框分别建模为二维高斯分布 N_p 和 N_g。然后计算两个高斯分布的二阶 Wasserstein 距离并进行归一化以衡量预测框和真实框的相似度。使用 NWD 作为损失函数的计算表达式如式(5)所示。

$$L_{NWD} = \exp\left(-\sqrt{\frac{(W_2^2(N_p, N_g))}{D}}\right) \quad (5)$$

式中:L_{NWD}代表使用 NWD 作为损失函数;D 为平衡因子,是一个常数;$W_2^2(N_p, N_g)$表示二维高斯分布 N_p 和 N_g 的二阶 Wasserstein 距离,其计算公式如式(6)所示。

$$W_2^2(N_p, N_g) = z^2 + \left(\frac{h_p}{2} - \frac{h_g}{2}\right)^2 + \left(\frac{w_p}{2} - \frac{w_g}{2}\right)^2 \quad (6)$$

式中 z 为预测框与真实框的中心点距离。

NWD 的优势在于即使边界框无重叠亦能提供平滑梯度,且对目标尺度不敏感,适合小目标匹配。然而,由式(6)可以看出,NWD 未考虑边框面积重叠信息,难以准确反映边界框的定位质量。其对边界框评估的缺点如图 7 所示,预测框中心在不同位置时,L_{NWD}变化不大,但 L_{IoU}差异显著,说明其评估能力仍有不足。

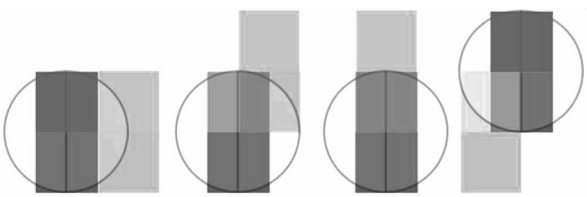


图 7 NWD 对边界框评估的缺点

Fig. 7 NWD's shortcomings for box evaluation

为此,本文在引入 NWD 的同时保留了 CIoU 对面积重叠的度量能力,但是降低了 CIoU 在边界框回归损失中的比重,以降低其对小目标过于敏感的影响,且 NWD 能更好评判预测框与真实边界框的相似程度。本文改进后的损失函数 NWD_CIoU 计算表达式如式(7)所示,其中 $\lambda \in [0,1]$,是一个超参数,表示 CIoU 所占的比重。

$$L_{NWD_CIoU} = \lambda \times L_{CIoU} + (1-\lambda) \times L_{NWD} \quad (7)$$

式中 L_{NWD_CIoU}表示 NWD_CIoU 的损失函数值。

3 实验结果与分析

3.1 数据集与评估指标

本文的实验数据集是名为 MT269 的手机元件,缺陷类别有长芽、牙缺、齿边、崩边、CD 纹不良等。CD 纹不良主要是 CD 纹区域的各种点子、异物、划伤等,长芽、牙缺、齿边均发生于油墨区边缘。其中长芽是边缘油墨多出,导致原有的白色玻璃位置为黑色油墨,环光图下透光玻璃为灰色,表现为深灰区域突出;牙缺是由于油墨区边缘油墨缺少,导致原来的灰色油墨位置为白色,环光图下表现为浅灰色有突出小块;齿边,表现为油墨印刷不良导致的油墨区与透光玻璃处形成长条形的齿牙交错状;崩边出现于产品最外侧油墨带边缘破损、裂纹等。各类缺陷如图 8 所示。数据集共有 1 619 张图,按 8∶2 比例划分为训练集和验证集;共有 535 个 CD 纹、562 个崩边、461 个长芽、493 个齿边、441 个牙缺。

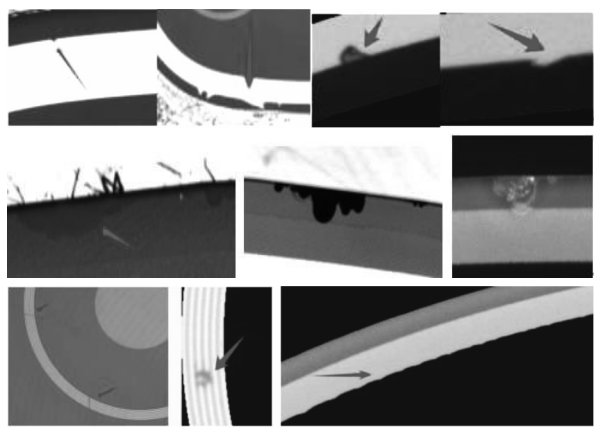


图 8 各类缺陷示例图

Fig. 8 Sample images with various defects

缺陷的尺度分布如图 9 所示,可以看到大部分目标的宽高不足输入图片尺度宽高的 0.03,本文输入图像尺度为 1 024 × 1 024,则大部分缺陷目标的尺度不到 30 × 30。小目标的定义有两种,绝对小目标和相对小目标,绝对小目标是指检测目标的尺度小于 32 × 32;相对小目标是指检测目标的宽高不到输入图像宽高的0.1。本文检测的缺陷明显满足

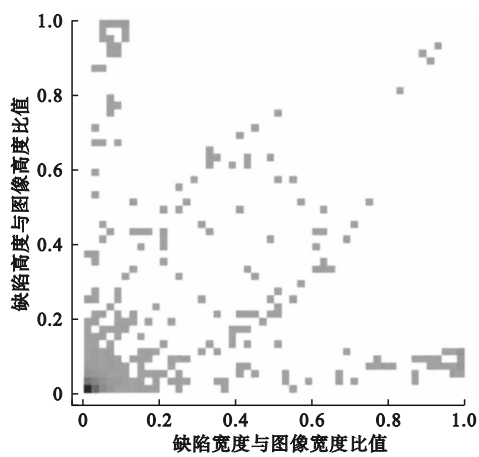


图 9 缺陷尺度分布

Fig. 9 Defect data scale distribution

小目标的定义。其中数据集中的中大目标只来自齿边缺陷,虽然中大目标占比不高,但是原始的 YOLOv8s 对中大目标已拥有不错的检测精度,且改进模型中未对其精度有明显影响,故不存在因为其占比较少导致模型对其关注度不够的情况。此外,YOLOv8s 自带的数据增强起到了一定的平衡样本作用。

为了验证改进算法的有效性,本文采用目标检测中常用的评估指标 F1-Score、mAP@50 和 mAP@50:95 对算法性能进行评估,这些指标能够综合反映算法对各缺陷类别的检测性能,对模型提供客观的性能评判。

3.2　实验环境

本文训练过程中使用的硬件设备为 16 GBx2NVIDIA Tesla V100 显卡,pytorch 框架为 2.0.0,Python 版本为 1.9.12,CUDA 版本为 11.7。具体实验的参数设置如表 1 所示。

表 1　实验参数设置表

Table 1　Experimental parameters settings

参数	设置	参数	设置
Epoch	500	Lr0	0.01
Imgsz	1 024	Lrf	0.01
Patience	100	Workes	8
Batch	8	Optermizer	Auto
aMP	False	Mosaic	0
λ	0.5	*m*	2

表 1 中 λ 为 CIoU_NWD 中 CIoU 所占的比重,为确定在本文数据集上 λ 的最优取值,通过不同 λ 取值的 CIoU_NWD 作为 YOLOv8 的边界框损失函数,其性能对比如表 2 所示。

表 2　不同 λ 取值的性能对比

Table 2　Comparison of results for different λ values

λ 取值	F1-Score/%	mAP@50/%	mAP@50:95/%
0.1	75.0	76.6	49.5
0.3	76.0	76.5	48.3
0.5	76.0	77.5	49.0
0.7	75.0	76.9	49.1
0.9	76.0	76.8	48.7

由表 2 可知,当 $\lambda = 0.5$ 时,其 mAP@50 最优,故后续实验中 λ 的设定均为 0.5。

本文通过将改进后算法的参数 *m* 分别设定为 2、4、8 进行实验,其性能对比如表 3 所示。

表 3　*m* 取不同值的性能对比

Table 3　Comparison of results for different *m* values

m 取值	F1-Score/%	mAP@50/%	mAP@50:95/%
2	79.0	80.7	50.6
4	77.0	80.6	50.4
8	78.0	80.3	51.1

由表 3 可知,当 $m = 2$ 时,其 mAP@50 最优,故后续实验中的 *m* 设定均为 2。

3.3　消融实验

为了验证本文各模块改进的有效性,将上文提到的模块 SWS、AugSPPF_SWS、NWD_CIoU 进行消融实验,并使用 F1-Score、mAP@50 与 mAP@50:95、计算量(GFLOPS)、参数量指标对实验结果进行评估。使用 YOLOv8s 作为基线模型,消融实验的结果如表 4 所示。

表 4　消融实验结果评价

Table 4　Ablation experiment results analysis

实验组号	SWS	AugSPPF_SWS	NWD_CIoU	参数量/10^6	GFLOPs	F1-Score/%	mAP@50/%	mAP@50:95/%
1				11.14	28.7	75.0	76.8	48.2
2	√			11.14	28.7	75.0	78.3	50.0
3		√		11.40	28.9	73.0	73.2	45.2
4			√	11.14	28.7	76.0	77.5	49.0
5	√		√	11.40	28.9	77.0	77.9	49.0
6		√	√	11.14	28.7	77.0	79.5	51.0
7		√	√	11.40	28.9	80.0	80.4	51.5
8	√	√	√	11.40	28.9	79.0	80.7	50.6

从表 4 可知,本文改进后的算法相较于原算法 GFLOPs 仅上涨了 0.2 的情况下,其 F1-Score、mAP@50、mAP@50:95 分别上升了 4.0、3.9、2.4 个百分点,达到了 79.0%、80.7% 与 50.6%。单独添加 SWS 时,其 mAP@50、mAP@50:95 较原始算法分别提高了 1.5、1.8 个百分点,说明了 SWS 对改善小目标特征提取的有效性。单独将 CIoU 替换为 NWD_CIoU,其 F1-Score、mAP@50、mAP@50:95 达到了 76.0%、77.5%、49.0%,说明了 NWD_CIoU 对小目标有着更好的梯度回传。

当单独加入 AugSPPF_SWS 时,虽然加入了全局视角信息和背景信息,然而由于 CIoU 对于小目标的敏感性,导致模型没有利用好其提供的丰富信息,反而干扰了小目标的检测。使用 NWD_CIoU 的同时加入 AugSPPF_SWS 比单独使用 NWD_CIoU 的 F1-Score、mAP@50、mAP@50:95 分别提升了 4.0、2.9、2.5 个百分点,说明使用 AugSPPF_SWS 提供的丰富信息和重要区域聚焦能力对小目标的有效性。最后,本文同时改进三个模块比其他所有实验的 mAP@50 都要高,表明三个改进模块在最终改进模型中都发挥了作用,证明了改进后算法的有效性。

为进一步验证 NWD_CIoU 对训练稳定性的影响,图 10 展示了 YOLOv8s 与使用 NWD_CIoU 改进后的模型在训练与验证阶段的边框损失曲线。

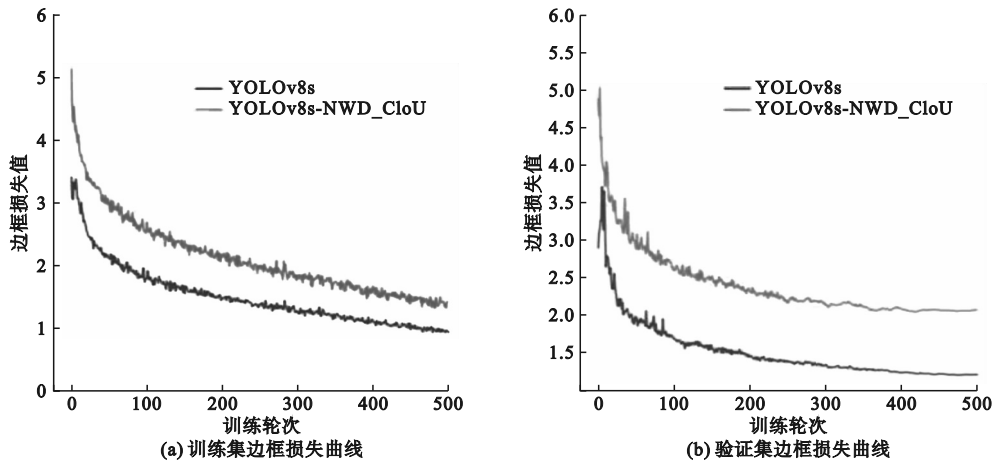


图 10 边框损失曲线对比

Fig. 10 Comparison of Box_loss-Epoch curves

由图 10 可以观察到,虽然 NWD_CIoU 的损失数值在整体上略高,但是因为 NWD_CIoU 保留了 CIoU 的情况下引入了 NWD。可以看到 NWD_CIoU 收敛速度相对较快,其下降趋势更加平稳,波动幅度显著小于基线模型,表明 NWD_CIoU 在训练过程中提供了更加平滑的梯度信号,有效减缓了因小目标带来的优化不稳定问题。

特别地,对于小目标占比较高的检测任务而言,稳定的优化过程有助于模型更充分地学习空间位置信息,从而提升最终的检测精度。

3.4 对比实验

为了检验本文算法改进的有效性,将本文算法与目前主流的工业缺陷检测模型进行对比,同时还将其与最新的小目标检测算法 SOD-YOLOv8[13]、LEAF-YOLO[14]进行对比,其结果如表 5 所示。

表 5 对比实验表

Table 5 Comparison experiments

模型	参数量/10^6	GFLOPs	F1-Score/%	mAP@50/%	mAP@50:95/%
Yolov8s	11.14	28.7	75.0	76.8	48.2
YOLOv8m	25.86	79.1	76.0	77.3	48.2
YOLOv11s	9.43	21.6	55.0	54.3	28.6
YOLOv12s	9.26	21.5	58.0	54.5	27.8
RT-DETR	32.82	108.0	52.0	51.2	27.1
YOLOX	8.94	34.107	63.0	64.1	38.2
SOD-YOLOv8	11.90	38.2	56.0	55.0	28.7
LEAF-YOLO	4.27	21.1	69.0	70.4	40.8
YOLOv8n	3.01	8.2	70.0	71.7	42.9
YOLOv8n + 本文改进	3.08	8.3	73.0	74.8	46.8
本文算法	11.40	28.9	79.0	80.7	50.6

由表 5 可知,本文改进算法的 GFLOPs 较原算法只提升了 0.2,其 F1-Score、mAP@50、mAP@50:95

却达到了 79.0%、80.7%、50.6%,比 GFLOPs 达到 79.1 的 YOLOv8m 高出了 3.0、3.4、2.4 百分点。同时本文使用了 RT-DETR、YOLOX、YOLOv11、YOLOv12 等为主流的实时性目标检测算法,从表 5 可以看到,本文算法的 F1-Score、mAP@50 与 mAP@50:95 都是最高的,且其 GFLOPs 比 YOLOX、RT-DETR 更低,比 YOLOv11s、YOLOv12s 只增加了不到 7.5。SOD-YOLOv8、LEAF-YOLO 为针对小目标检测的最新算法,可以看到,本文改进后算法的多个指标优于这两个算法。最后,本文在更小规模的 YOLOv8n 算法中进行了相同的改进,在 GFLOPs 只增加了 0.1 的情况下,F1-Score、mAP@50、mAP@50:95 分别提高了 3.0、3.1、3.9 个百分点。以上对比实验证明了本文算法改进的有效性。

为验证所提算法的鲁棒性,本文选取西北工业大学发布的 NWPU VHR-10 数据集,将改进前后算法的实验结果进行对比,结果如表 6 所示。

表 6　NWPU VHR-10 数据集上模型性能对比

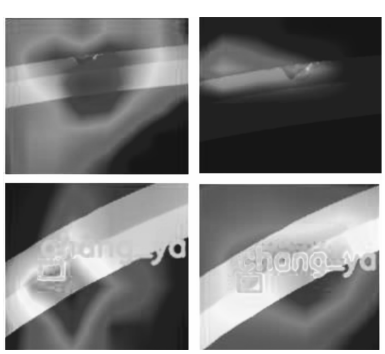
Table 6　Performance comparison of models on the NWPU VHR-10 dataset

模型	F1-Score/%	mAP@50/%	mAP@50:95/%
YOLOv8s	91.0	94.7	67.3
本文算法	91.0	95.5	68.6

由表 6 可知,本文改进后算法在 NWPU VHR-10 数据集上较原 YOLOv8s 在 mAP@50 与 mAP@50:95 方面均有提高。

3.5　热力图对比及检测结果示例

热力图是一种直观且强大的可视化方式,可以明显表示模型关注的重点区域。本文改进前后的模型热力图对比如图 11 所示,可以很明显地看到,改进后的模型能够更加关注缺陷位置及缺陷附近的区域。

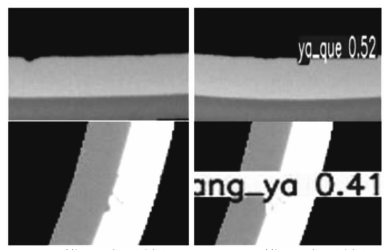


(a) 改进模型热力图　(b) 原模型热力图

图 11　热力图对比

Fig. 11　Heatmaps comparisons

图 12 为模型改进前后检测结果对比图,可以看到对于模型改进前无法检出的微小缺陷,改进后的模型能够将其检出,说明本文改进后的模型对小目标检测具有更良好的性能。



(a) 原模型检测结果　(b) 改进模型检测结果

图 12　检测结果图对比

Fig. 12　Comparison of detection results diagrams

4　结论

本文针对 YOLOv8 检测手机元件缺陷时检测性能不佳的问题,首先使用切片操作改进了轻量化的 SimAM,使其能够更好地聚焦于小目标所需要的重要特征,将其嵌入 Backbone 中的 Conv 块之后,有效保留了更多小目标的重要信息,避免小目标在提取特征后信息损失严重等问题;并设计了改进的多尺度特征融合模块 AugSPPF_SWS,其在原有 SPPF 的基础上加入了丰富的全局信息,并将 SWS 嵌入其中,使得模型能够更好地聚焦于丰富信息中的重要部分,帮助模型更好地做出判断;另外在对小目标敏感的 CIoU 损失函数中加入 NWD,缓解了 CIoU 对小目标敏感的问题,同时避免了简单使用 NWD 替换 CIoU 而导致的对边界框重叠度判断不精确的问题。

参考文献(References):

[1] Iqra, Giri K J, Javed M. Small object detection in diverse application landscapes: a survey[J]. Multimedia Tools and Applications, 2024, 83(41): 88645 – 88680.

[2] 杨大为,张成超. 基于改进 YOLOv7 的安全帽佩戴检测算法[J]. 沈阳理工大学学报,2024,43(1):16 – 21.
Yang D W, Zhang C C. Algorithm of safety helmet wearing detection based on improved YOLOv7[J]. Journal of Shenyang Ligong University, 2024, 43(1): 16 – 21. (in Chinese)

[3] 赵子艺,张德育,徐双成. 基于改进 YOLOv8 的铁路异物检测算法[J]. 沈阳理工大学学报,2025,44(3):32 – 38,46.
Zhao Z Y, Zhang D Y, Xu S C. Algorithm of railway foreign object intrusion detection based on improved YOLOv8[J]. Journal of Shenyang Ligong University, 2025, 44(3): 32 – 38, 46. (in Chinese)

[4] 于傲泽,魏维伟,王平,等. 基于分块复合注意力的无人机小目标检测算法[J]. 航空学报,2024,45(14):629148.
Yu A Z, Wei W W, Wang P, et al. Small target detection algorithm for UAV based on patch-wise co-attention[J]. Acta

Aeronautica et Astronautica Sinica, 2024, 45(14): 629148. (in Chinese)
[5] 王晓青,阎吉,张德育. 改进 Bot-SORT 的边坡落石监测方法[J]. 沈阳理工大学学报,2024,43(4):19 – 26.
Wang X Q, Yan J, Zhang D Y. Improved bot-SORT for slope rockfall monitoring[J]. Journal of Shenyang Ligong University, 2024, 43(4): 19 – 26. (in Chinese)
[6] Li N, Ye T R, Zhou Z H, et al. Enhanced YOLOv8 with BiFPN-SimAM for precise defect detection in miniature capacitors [J]. Applied Sciences, 2024, 14(1): 429.
[7] Wang H Z, Liu J H, Zhao J, et al. Precision and speed: LSOD-YOLO for lightweight small object detection[J]. Expert Systems with Applications, 2025, 269: 126440.
[8] Sun K J, Huo J, Liu Q, et al. An infrared small target detection model via Gather-Excite attention and normalized Wasserstein distance [J]. Mathematical Biosciences and Engineering, 2023, 20(11): 19040 – 19064.
[9] Varghese R M S. YOLOv8: a novel object detection algorithm with enhanced performance and robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). Chennai, India: IEEE, 2024: 1 – 6.
[10] 陈扬,吕艳辉. 基于改进 YOLOv5s 的头盔佩戴检测算法[J]. 沈阳理工大学学报,2023,42(5):11 – 17.
Chen Y, Lyu Y H. Algorithm of helmet wearing detection based on improved YOLOv5s[J]. Journal of Shenyang Ligong University, 2023, 42(5): 11 – 17. (in Chinese)
[11] Xu C, Wang J W, Yang W, et al. Detecting tiny objects in aerial images: a normalized Wasserstein distance and a new benchmark[J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2022, 190: 79 – 93.
[12] Zheng Z H, Wang P, Ren D W, et al. Enhancing geometric factors in model learning and inference for object detection and instance segmentation[J]. IEEE Transactions on Cybernetics, 2022, 52(8): 8574 – 8586.
[13] Khalili B, Smyth A W. SOD-YOLOv8-enhancing YOLOv8 for small object detection in aerial imagery and traffic scenes [J]. Sensors, 2024, 24(19): 6209.
[14] Nghiem V Q, Nguyen H H, Hoang M S. LEAF-YOLO: lightweight edge-real-time small object detection on aerial imagery [J]. Intelligent Systems with Applications, 2025, 25: 200484.

(责任编辑:和晓军)

(上接第 30 页)

[8] 戚翔宇,苏庆华,张智超,等. 改进 YOLOv11 的缺陷检测算法[J]. 计算机科学与应用,2025,15(1):108 – 116.
Qi X Y, Su Q H, Zhang Z C, et al. Improvement of the YOLOv11 defect detection algorithm[J]. Computer Science and Application, 2025, 15(1): 108 – 116. (in Chinese)
[9] Zhou P, Zhou G B, Wang S H, et al. Visual sensing inspection for the surface damage of steel wire ropes with object detection method[J]. IEEE Sensors Journal, 2022, 22(23): 22985 – 22993.
[10] 崔晶楠,黄春艳,李艳玲. 融合动态卷积的 YOLOv5s 钢材表面缺陷检测[J]. 制造业自动化,2024,46(11):126 – 132.
Cui J N, Huang C Y, Li Y L. Fusion of YOLOv5s and dynamic convolution for steel surface defect detection[J]. Manufacturing Automation, 2024, 46(11): 126 – 132. (in Chinese)
[11] 王红尧,韩爽,李勤怡. 改进 YOLOv5 的钢丝绳损伤图像识别实验方法研究[J]. 计算机工程与应用,2023,59(17):99 – 106.
Wang H Y, Han S, Li Q Y. Experimental research on image recognition of wire rope damage based on improved YOLOv5 [J]. Computer Engineering and Applications, 2023, 59(17): 99 – 106. (in Chinese)
[12] 赵佰亭,张晨,贾晓芬. ECC-YOLO:一种改进的钢材表面缺陷检测方法[J]. 电子测量与仪器学报,2024,38(4):108 – 116.
Zhao B T, Zhang C, Jia X F. ECC-YOLO: an improved method for detecting surface defects in steel[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4): 108 – 116. (in Chinese)
[13] Bao D C, Gao R J. YED-YOLO: an object detection algorithm for automatic driving[J]. Signal, Image and Video Processing, 2024, 18(10): 7211 – 7219.
[14] Xie X D, Xu B, Chen Z F. Real-time fall attitude detection algorithm based on iRMB[J]. Signal, Image and Video Processing, 2025, 19(2): 156.
[15] Zou L, Chen A Q, Yang X H, et al. An improved method of AUD-YOLO for surface damage detection of wind turbine blades[J]. Scientific Reports, 2025, 15: 5833.
[16] Liu T Q, Gu M M, Sun S H. RIEC-YOLO: an improved road defect detection model based on YOLOv8[J]. Signal, Image and Video Processing, 2025, 19(4): 285.
[17] Yao L, Zhao B, Wang X H, et al. A detection algorithm for surface defects of printed circuit board based on improved YOLOv8[J]. IEEE Access, 2024, 12: 170227 – 170242.
[18] 成顺,李建荣,王志乾,等. 基于 YOLOv8 轻量化水下光学图像识别算法[J]. 激光与光电子学进展,2025,62(4):392 – 402.
Cheng S, Li J R, Wang Z Q, et al. Lightweight underwater optical image recognition algorithm based on YOLOv8[J]. Laser & Optoelectronics Progress, 2025, 62(4): 392 – 402. (in Chinese)

(责任编辑:和晓军)