

改进YOLOv11的缺陷检测算法

戚翔宇, 苏庆华, 张智超, 徐升, 万开政

北京物资学院信息学院, 北京

收稿日期: 2024年12月20日; 录用日期: 2025年1月17日; 发布日期: 2025年1月26日

摘要

PCB缺陷检测是电子制造业质量控制的重要组成部分。PCB板在生产过程中可能遭遇各种环境与工艺因素的影响, 缺陷类型复杂多样, 准确检测这些缺陷是提高生产效率和电子产品质量的关键。针对PCB缺陷特点, 本文提出了一种结合多头混合自注意力机制改进YOLOv11n模型的PCB缺陷检测方法, 以提高检测精度和鲁棒性。通过将MMSA嵌入YOLOv11n框架的Backbone和Neck部分, 增加的小目标检测层, 增强对PCB缺陷特征的提取与融合能力, 提升了对PCB微小缺陷的识别能力。实验结果表明, 该方法在PKU-Market-PCB数据集上实现了94.8%的mAP@0.5, 相较于YOLOv11n提升了1.6%, 且优于其他主流检测算法, 显示了其在PCB缺陷检测中的显著优势。

关键词

YOLO, 缺陷检测, 多头混合自注意力机制(MMSA), 小目标检测

Improvement of the YOLOv11 Defect Detection Algorithm

Xiangyu Qi, Qinghua Su, Zhichao Zhang, Sheng Xu, Kaizheng Wan

School of Information, Beijing Wuzi University, Beijing

Received: Dec. 20th, 2024; accepted: Jan. 17th, 2025; published: Jan. 26th, 2025

Abstract

PCB defect detection is a crucial component of quality control in the electronics manufacturing industry. During production, PCBs may encounter a variety of environmental and process factors, leading to complex and diverse types of defects. Accurately detecting these defects is key to improving production efficiency and the quality of electronic products. In response to the characteristics of PCB defects, this study proposes a PCB defect detection method that improves the YOLOv11n

model by integrating a multi-head mixed self-attention mechanism (MMSA) to enhance detection accuracy and robustness. By embedding MMSA into the Backbone and Neck sections of the YOLOv11n framework, and adding a small target detection layer, this approach enhances the capability to extract and merge features of PCB defects, thus improving the recognition of minute PCB flaws. Experimental results demonstrate that this method achieves an mAP@0.5 of 94.8% on the PKU-Market-PCB dataset, which is a 1.6% improvement over the YOLOv11n baseline and surpasses other mainstream detection algorithms, highlighting its significant advantages in PCB defect detection.

Keywords

YOLO, Defect Detection, Multi-Head Mixed Self-Attention Mechanism (MMSA), Small Target Detection

Copyright © 2025 by author(s) and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

1. 引言

随着电子制造业的快速发展,印刷电路板(PCB)在现代电子设备中扮演着核心角色。确保 PCB 在生产过程中无缺陷对于提高最终产品的可靠性和性能至关重要。长久以来,PCB 缺陷检测主要依赖人工视觉检查,这种方法不仅效率低下,而且结果的准确性受到操作员疲劳和主观判断的影响,难以适应高精度和高效率的现代制造需求。

为了克服人工检测的局限性,传统的自动化方法,如通过像素对比的 XOR 操作来识别缺陷位置,已被广泛应用。尽管这些方法在简单应用中有效,但它们在处理复杂的或微小的缺陷时往往效果不佳。进一步地,研究者尝试通过定义一系列的图像处理规则,如欧拉数和连通数,来进行更精细的缺陷分类。然而,这些方法仍然面临速度慢和准确度不足的挑战。

相较于传统方法,深度学习提供了一种从数据中自动学习特征的能力,这对于处理 PCB 的多样性和复杂性至关重要。深度学习模型,尤其是卷积神经网络(CNN)类模型,能够有效地从图像中提取丰富的层次特征,即使这些特征的语义信息不十分明显。这一能力使得深度学习模型在识别微小或在复杂背景中的缺陷方面表现出色[1]。

本文采用了改进的 YOLOv11 算法,这是一种深度学习框架,专门针对实时对象检测进行优化。通过对 YOLOv11 算法进行改进,本研究旨在提升模型对 PCB 图像中缺陷的检测精度,并确保算法能够适应各种 PCB 设计的变化。这种改进不仅提高了缺陷检测的准确性,也显著增强了算法的泛化能力,使其能在多变的生产条件下维持稳定性能。

本文将详细介绍这种改进的 YOLOv11 算法在 PCB 缺陷检测中的应用,并比较其性能与传统方法和基本 YOLOv11 模型的差异。通过这些分析,本文旨在展示深度学习如何改进 PCB 缺陷检测,以及这些技术如何帮助制造业提高质量控制的标准。

2. 相关工作

2.1. 注意力机制

注意力机制(Attention Mechanism) [2]最早源于对人类视觉系统的研究,其灵感来自于人脑如何在处

理信息时，集中注意力于某些重要的部分，从而高效地完成任任务。它广泛应用于自然语言处理(NLP)、计算机视觉(CV) [3] [4]、语音识别等多个领域，成为深度学习模型尤其是神经网络中的一种重要机制。

在神经网络中，注意力机制通过动态地分配不同的权重，赋予输入的不同部分不同的关注度，从而使得网络能够自适应地聚焦于更有信息量的区域或特征。这种机制有效地缓解了传统卷积神经网络(CNN)和循环神经网络(RNN)在长距离依赖、局部信息处理和特征融合方面的限制。

在目标检测任务中，“混合局部通道注意力”(MLCA)机制[5]有效解决了传统注意力机制的一些局限性，尤其是在处理空间信息时的不足。传统的通道注意力机制，如 Squeeze-and-Excitation (SE)和 Efficient Channel Attention (ECA) [6]，通常将整个通道特征图压缩为单一值，忽略了每个通道内的空间信息，在某些情况下可能导致重要的空间特征信息丢失，尤其是当只有部分通道对最终结果具有重要作用时。

为了解决这一问题，MLCA 通过将特征图划分为多个斑块并在这些斑块上应用注意力机制，类似于局部 SE 机制的实现。这一做法确保了每个斑块中的空间特征能够得到充分保留，从而有效增强了网络对局部信息的关注。然而，过多的斑块可能导致参数数量的显著增加，影响计算效率。为了进一步平衡通道注意力机制性能和复杂度，提升目标检测网络的性能，产生了多头注意力，多头注意力(Multi-Head Attention)是将查询(Q)、键(K)和值(V)映射到多个子空间，在不同的表示子空间中捕获不同的信息，并行地计算注意力得分的注意力机制。本文将基于通道、空间和局部混合注意力，设计一种同时融合通道注意力和空间信息以及多头自注意力的多头混合自注意力机制适应自然界遮挡目标检测需求。

2.2. YOLOv11n 模型

YOLO (You Only Look Once)系列[7]-[10]是一系列高效且实时的目标检测算法，自 2015 年首次提出以来，经历了多次迭代与优化。YOLO 系列算法以其单阶段检测框架、高速推理能力和不断提升的检测精度而著称。各版本在目标检测、图像分类、语义分割、实时追踪和姿态检测等领域取得了显著成果，广泛应用于计算机视觉相关领域。YOLOv11 是目前最新的 YOLO 版本，提供了五种不同尺寸的版本，本文使用其中最小最快的 YOLOv11n 模型进行改进，YOLOv11n 的原模型结构见图 1。

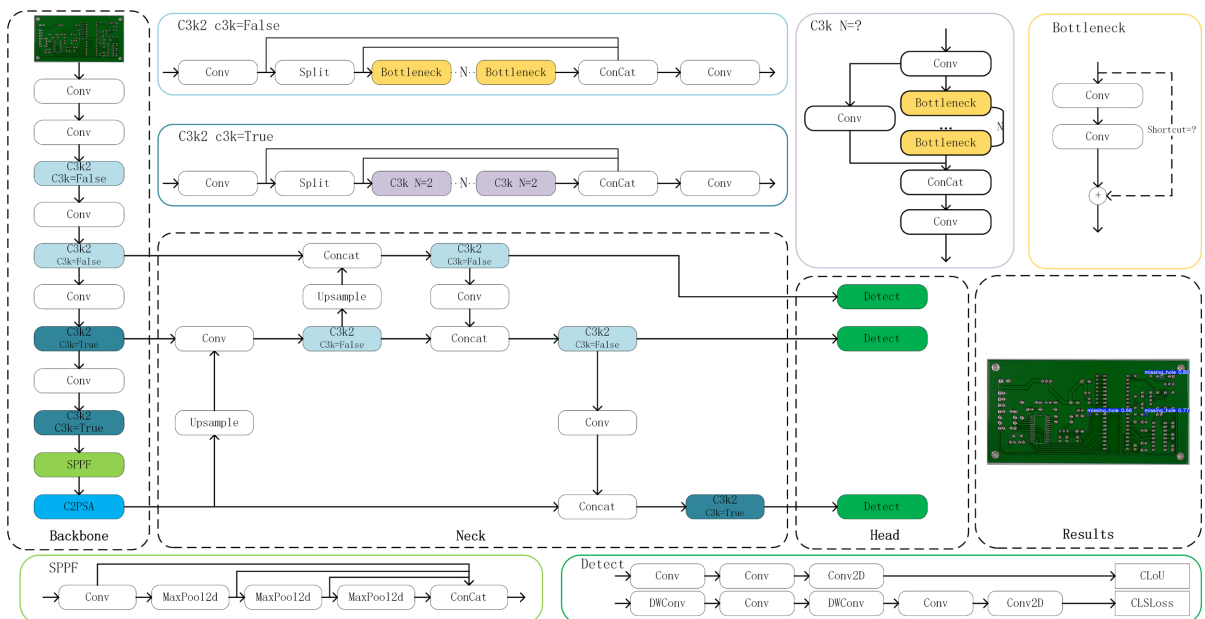


Figure 1. YOLOv11n model structure
图 1. YOLOv11n 模型结构

YOLOv11 由 Backbone、Neck 和 Head 组成。

Backbone 主要由卷积模块、C3k2 模块、SPPF 模块和 C2PSA 模块构成，卷积模块包含二维卷积层、BN(批量归一化)和激活函数，用于提取特征。YOLOv11 提出一种新的 C3k2 机制，传入一个参数为 $c3k$ ，其中在网络的浅层 $c3k$ 设置为 `False`，此时所谓的 C3k2 就相当于 YOLOv8 中的 C2f，其网络结构为一致的。SPPF (Spatial Pyramid Pooling-Fast)通过在特征图上执行不同大小的池化操作，并将结果进行整合，从而得到固定尺寸的输出，在减少计算量的同时最大限度地保留了下采样过程中的信息；C2PSA 模块是一个 C2 模块内部嵌入了一个多头注意力机制。YOLOv11 的 Head 部分使用了新的模块 DWConv (Depthwise Convolution)，只处理空间维度上的卷积，不再处理通道间的卷积，可以有效减少参数复杂度和参数量。为了弥补深度卷积缺乏跨通道信息整合的问题，通常深度卷积后会配合 1×1 的点卷积，这种组合被称为深度可分离卷积(Depthwise Separable Convolution)，这也是 YOLOv11 的做法。

3. 方法

本文参考相关工作[5] [6]，在 YOLOv11n 的基础上添加更多的注意力机制，结合了通道信息和空间信息，以及局部和全局信息，融合基于图像特征设计的多头混合自注意力机制，并将其融合进 YOLOv11n 的 Backbone 和 Neck 部分，增强模型进行缺陷检测的能力，改进后的 YOLOv11n 如图 2 所示。多头混合自注意力(MMSA)的具体实现步骤如下：首先，分别对输入特征图进行局部和全局的平均池化操作；接着，将池化后的特征图转换为适配于 Multi-Head Attention 机制的输入形状；随后，利用多头自注意力机制计算特征间的相关性；之后，将自注意力结果重新调整为与原始特征图一致的形状；接下来，基于局部和全局信息计算对应的注意力权重；为了将权重限制在 0 到 1 的范围内，采用 Hard Sigmoid 激活函数进行处理；随后，通过自适应平均池化对权重进行进一步的尺度调整；然后，将局部和全局注意力权重进行融合；随后，将融合后的注意力权重作用于原始特征图；最后，将经过自注意力处理的结果与原始特征图进行相加，得到最终的特征表示，具体的原理图见图 3。

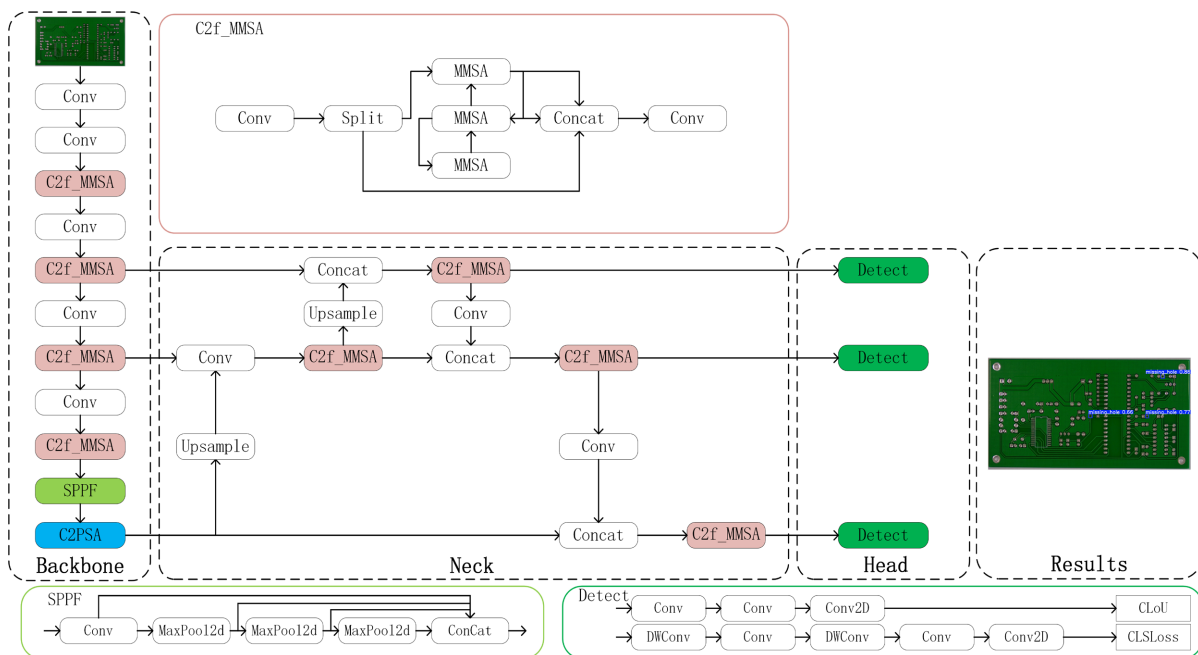


Figure 2. The improved YOLOv11 model structure
图 2. 改进的 YOLO11 模型结构

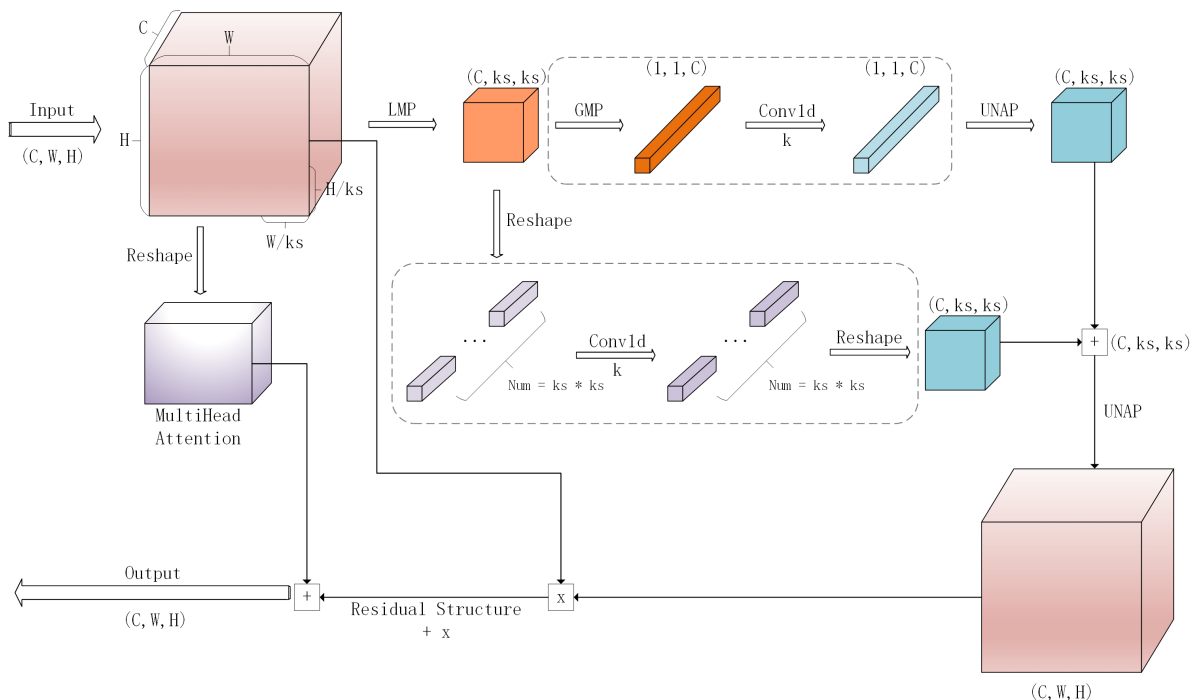


Figure 3. Structure diagram of MMSA
图 3. MMSA 结构图

图 4 展示了 MMSA 结构中 GMP、LMP 和 UNAP 之间的关系。GMP (Global Max Pooling)表示全局最大池化，其输出为大小为 1×1 的特征映射；LMP (Local Max Pooling)指局部最大池化，通过将整个特征图划分为 $k \times k$ 的片区，然后对每个片区执行 $k \times k$ 的最大池化操作；UNAP 则是反池化操作，主要用于关注特征图的属性并扩展至所需的大小。UNAP 可通过输出大小等于源特征映射大小的自适应池化机制来实现。在对 LMP 的特征图进行扩展时，若输出大小不是 1×1 ，则无法直接执行扩展操作，必须通过 UNAP 过程将特征映射还原至其原始大小。

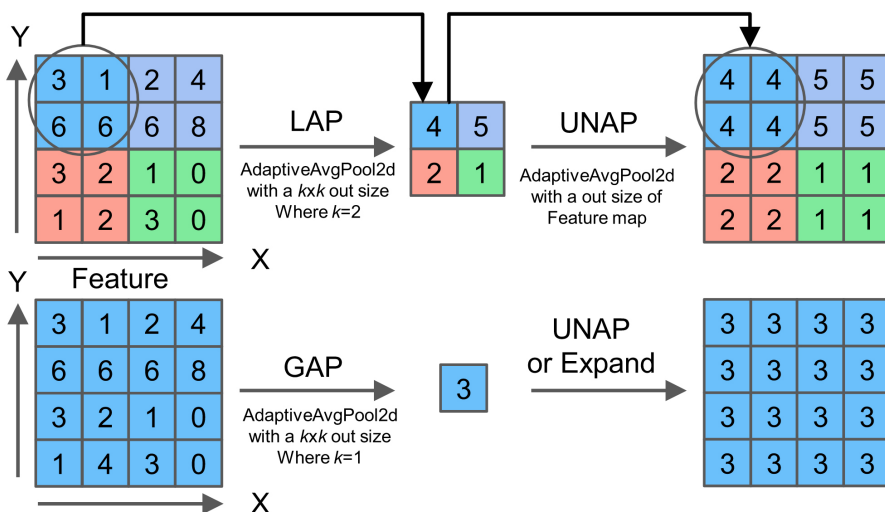


Figure 4. Schematic diagram of GAP, LAP and UNAP
图 4. GAP、LAP 和 UNAP 的原理图

4. 实验及结果分析

4.1. 数据集与评价指标

4.1.1. 数据集

本文选用的数据集为 PKU-Market 系列数据集。

PKU-Market-PCB 数据集是由北京大学智能机器人开发实验室制作了一个简单的 PCB 瑕疵数据集, 该数据集的开发旨在支持对 PCB 缺陷识别算法的评估和验证。数据集包括高质量的 PCB 图像, 这些图像专门展示了电路板的视觉特征及其潜在瑕疵。这些瑕疵是人为通过 Photoshop 软件生成的, 而非来自实际生产线, 因为实际生产过程中缺陷样本的获取相对较少。数据集中只包含单层裸板, 而非多层复合结构的 PCB。数据集里的图片分辨率比较高, 而且已经进行了一系列的数据增强操作, 包括裁剪, 改变亮度等。

本文将数据集划分为训练集和验证集, 其中 80%的数据作为训练集用于模型训练, 20%为验证集用于验证。

4.1.2. 评价指标

本文使用平均精度(AP)、平均精度均值(mAP)、参数量(Params)和每秒浮点运算次数(GFLOPs)作为模型的主要评估指标, 每秒帧数(FPS)作为参考的评估指标。

平均精度(AP)是通过召回率 - 准确率曲线下的面积进行积分计算得到的, 它反映了模型在不同召回率水平上的准确率表现。计算公式如下: $AP = \int_0^1 p(r) dr$ 其中, $p(r)$ 是在给定召回率 r 时的准确率。

平均精度均值(mAP)是所有类别的平均精度(AP)的平均值, 它提供了一个单一指标来评估模型在所有类别上的整体性能。对于不同的 IoU (交并比) 阈值, 我们计算多个 mAP 值来更全面地评估模型性能, 如 mAP 在 IoU 阈值为 0.5 时(通常表示为 mAP@0.5)和在 IoU 从 0.5 到 0.95 的范围内的平均值(表示为 mAP@0.5:0.95)。

模型的参数总量(Params)和运算性能(GFLOPs)也是衡量模型复杂度和效率的重要指标。参数总量直接关联到模型的存储和内存需求, 而 GFLOPs (十亿次浮点运算)则表明模型在每秒钟可以执行的浮点运算次数, 是衡量模型运算需求的指标。参数越多或 GFLOPs 越高的模型可能具有更高的预测精度, 但同时也可能导致计算成本和能耗增加。

每秒帧数(Frames Per Second, FPS)衡量的是模型处理输入数据的速度, 对于实时应用尤为重要。高 FPS 意味着模型可以更快地处理图像, 从而适用于需要即时反馈的应用场景。

4.2. 实验环境与设置

Table 1. The Impact of Batch_size on the Performance of YOLOv11n

表 1. Batch_size 对 YOLOv11n 性能的影响

Model	Batch_size	P	R	mAP@0.5	mAP@0.5:0.95
YOLOv11n	8	0.945	0.883	0.929	0.5
	16	0.948	0.884	0.931	0.501
	32	0.933	0.905	0.937	0.487

本文实验环境如下: NVIDIA GeForce RTX 4060 GPU、AMD Ryzen 7 8845HS CPU、pytorch 2.0.1、Python 3.9、cuda 11.8。本文在 PKU-Market-PCB 数据集上采用了相同的初始化实验参数设置, 实验参数设置如下: 输入图像的大小为 640×640 , 每个训练批次大小(batch size)为 16 (不同 Batch_size 对模型的

影响见表 1), 训练的世代数(epoch)为 300, 初始学习率为 0.01, 用于加速训练过程的动量参数为 0.937, 优化器选择 SGD 优化器, 采用多线程读取数据, 工作线程量(num workers)设置为 4。

4.3. 实验结果与分析

为了验证在模型中融合新的多头混合自注意力机制模块对模型性能的影响, 本文将融合了 MMSA 注意力机制的 YOLOv11n 模型与原始的 YOLOv11n 进行了对比实验。表 2 展示了引入多头混合自注意力机制后, YOLOv11n 模型的检测性能变化。实验结果表明, 在 PKU-Market-PCB 数据集中, 加入 MMSA 注意力机制后, 平均精度(mAP@0.5)从 0.932 提升至 0.948, 提高了 1.6%。此外, mAP@0.5:0.95 的性能从 0.5 提高至 0.515, 召回率也从 0.879 提升至 0.914, 说明模型在多尺度目标上的检测能力得到了进一步增强。

与此同时, 引入 MMSA 注意力机制后, 虽然计算复杂度(GFLOPs)有所增加, 但在目标检测的关键性能指标上取得的显著提升证明了该机制的有效性, 见图 5。

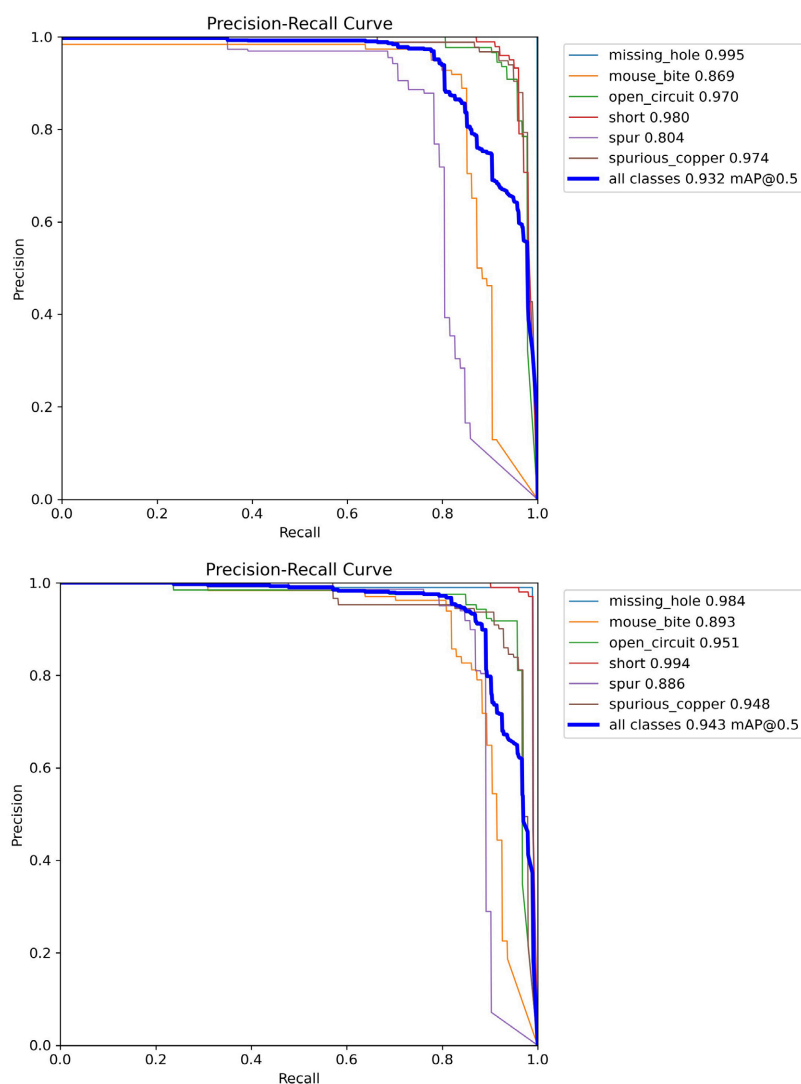


Figure 5. PR curve comparison before and after algorithm improvement on the PKU-Market-PCB dataset
图 5. PKU-Market-PCB 数据集中算法改进前后的 PR 曲线对比图

Table 2. Comparison of performance of the multihead mixed self-attention mechanism algorithm
表 2. 多头混合自注意力机制算法检测性能对比表

Model	P	R	mAP@0.5	MAP@0.5:0.95	FPS	GFLOPs
YOLOv11n	0.961	0.879	0.932	0.5	85	6.3
YOLOv11n + MMSA	0.949	0.914	0.948	0.515	70	10.4

为了进一步验证多头混合自注意力机制对模型性能的提升作用,进行了消融实验(见表 3)。实验通过分别移除 Backbone 和 Neck 中的 MMSA 模块,研究其对模型性能的具体影响。结果显示:引入 MMSA 模块后,尽管计算复杂度有所增加,但模型的检测性能显著提升,特别是在复杂场景下的检测精度和小目标检测率上体现出明显优势。

Table 3. Ablation experiment
表 3. 消融实验

Model	Backbone	Neck	mAP@0.5	MAP@0.5:0.95	FPS	GFLOPs
YOLOv11n	×	×	0.932	0.5	85	6.3
	×	√	0.942	0.494	75	6.8
	√	×	0.921	0.492	72	8.7
	√	√	0.948	0.515	70	10.4

5. 结语

本文设计了一种缺陷检测的方法,应用于 PCB 缺陷检测领域。实验结果表明,本文所提方法在 PKU-Market-PCB 数据集上取得了较高的检测精度,表现出了良好的泛化能力和鲁棒性。本文提出的方法不仅提高了表面缺陷检测的准确性,还为 PCB 生产中的表面缺陷检测应用提供了有价值的理论基础和实用指导。然而,所提方法仍有一些改进的空间。例如,可以尝试将所提方法应用于其他类型的缺陷检测任务中,以验证其泛化能力和实用性。

参考文献

- [1] Lin, G., Liu, K., Xia, X. and Yan, R. (2022) An Efficient and Intelligent Detection Method for Fabric Defects Based on Improved Yolov5. *Sensors*, **23**, Article 97. <https://doi.org/10.3390/s23010097>
- [2] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., et al. (2017) Attention Is All You Need. *Proceedings of the 31st International Conference on Neural Information Processing Systems*, Long Beach, 4-9 December 2017, 6000-6010.
- [3] Dosovitskiy, A. (2020) An Image Is Worth 16x16 Words: Transformers for Image Recognition at Scale. arXiv: 2010.11929.
- [4] Guo, M., Xu, T., Liu, J., Liu, Z., Jiang, P., Mu, T., et al. (2022) Attention Mechanisms in Computer Vision: A Survey. *Computational Visual Media*, **8**, 331-368. <https://doi.org/10.1007/s41095-022-0271-y>
- [5] Wan, D., Lu, R., Shen, S., Xu, T., Lang, X. and Ren, Z. (2023) Mixed Local Channel Attention for Object Detection. *Engineering Applications of Artificial Intelligence*, **123**, Article 106442. <https://doi.org/10.1016/j.engappai.2023.106442>
- [6] Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W. and Hu, Q. (2020). ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, 13-19 June 2020, 11531-11539. <https://doi.org/10.1109/cvpr42600.2020.01155>
- [7] Terven, J., Córdova-Esparza, D. and Romero-González, J. (2023) A Comprehensive Review of YOLO Architectures in

- Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. *Machine Learning and Knowledge Extraction*, **5**, 1680-1716. <https://doi.org/10.3390/make5040083>
- [8] Redmon, J., Divvala, S., Girshick, R. and Farhadi, A. (2016) You Only Look Once: Unified, Real-Time Object Detection. 2016 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, 27-30 June 2016, 779-788. <https://doi.org/10.1109/cvpr.2016.91>
- [9] Wang, C., Yeh, I. and Mark Liao, H. (2024) YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. *Computer Vision—ECCV 2024*, Milan, 29 September-4 October 2024, 1-21. https://doi.org/10.1007/978-3-031-72751-1_1
- [10] Wang, A., Chen, H., Liu, L., *et al.* (2024) YOLOv10: Real-Time End-to-End Object Detection. arXiv: 2405.14458.