

DOI:10.19651/j.cnki.emt.2518156

改进 YOLO11 的 PCB 表面缺陷检测方法^{*}

吴 葛 朱宇凡 贾泽宇

(华中科技大学电气与电子工程学院 武汉 430074)

摘要: 针对 PCB 表面缺陷检测准确率不足,无法充分平衡模型检测准确性与实时性,难以满足现代电子制造系统稳定运行要求的问题,提出了一种改进 YOLO11 的 PCB 表面缺陷检测方法 HDH-YOLO。该方法通过采用优化的 HGNetV2 替换原 YOLO11 的骨干网络,实现模型的轻量化;借鉴 Dynamicconv 的思想对 HGBlock 进行改进,用改进的 Dynamic_HGBlock 替换 HGNetV2 网络中后四层 HGBlock,在不增加过多计算量的前提下引入更多网络参数,增强网络对泛化特征的学习能力,进而提高检测精度;在骨干网络末端添加 DSM 注意力机制层,通过放大关键区域的空间域和频率域响应提升模型的小目标检测能力。在 PKU-Market PCB 和 DeepPCB 数据集进行对比实验和消融实验,实验结果表明,提出的 HDH-YOLO 模型较基线 YOLO11n 模型参数量降低 6.20%,计算量降低 12.70%,mAP50 和 mAP50-95 分别提升 2.6% 和 2.3%,较好地平衡了轻量化和检测精度,在现代电子制造系统中具有高可靠性和高实用性。

关键词: PCB 表面缺陷检测;YOLO11;HGNetV2;HGBlock;Dynamicconv;DSM

中图分类号: TP391.4; TN791 文献标识码: A 国家标准学科分类代码: 520.6040

Improved PCB surface defect detection method based on YOLO11

Wu Ge Zhu Yufan Jia Zening

(School of Electrical & Electronic Engineering, Huazhong University of Science and Technology, Wuhan 430074, China)

Abstract: Aiming at the problem that the detection accuracy of PCB surface defects is insufficient and it is difficult to balance the detection accuracy and real-time performance of the model, which cannot meet the stable operation requirements of modern electronic manufacturing systems, an improved PCB surface defect detection method named HDH-YOLO is proposed. This method replaces the backbone network of the original YOLO11 with an optimized HGNetV2 to achieve model lightweighting. It also improves the HGBlock by referring to the idea of Dynamicconv, and replaces the last four HGBlocks in the HGNetV2 network with the improved Dynamic_HGBlock. This approach introduces more network parameters without significantly increasing the computational load, thereby enhancing the network's ability to learn generalized features and improving detection accuracy. In addition, a DSM attention mechanism layer is added at the end of the backbone network to enhance the model's feature extraction capability by amplifying the spatial and frequency domain responses of key regions. Comparative experiments and ablation studies were conducted on the PKU-Market PCB and DeepPCB datasets. The results show that, compared with the baseline YOLO11n model, the proposed HDH-YOLO model reduces the number of parameters by 6.20% and the computational load by 12.70%, while increasing the mAP50 and mAP50-95 by 2.6% and 2.3%, respectively. HDH-YOLO thus achieves a better balance between lightweighting and detection accuracy and demonstrates high reliability and practicality in modern electronic manufacturing systems.

Keywords: PCB 表面缺陷检测;YOLO11;HGNetV2;HGBlock;Dynamicconv;DSM

0 前言

在现代电子制造领域,印刷电路板(printed circuit

board, PCB)的质量是决定电子产品性能和可靠性的关键因素。PCB 表面缺陷(如缺孔、短路、开路等)直接影响其质量,因此高效、精准的缺陷检测是提升电子产品可靠性的

收稿日期:2025-02-24

* 基金项目:教育部 2020 年第二批新工科研究与实践项目(E-NYDQHGC20202219)资助

核心环节^[1]。然而,传统方法如人工目检、自动光学检测(automated optical inspection,AOI)等^[2]因检测效率低、易错漏等问题,难以满足现代生产的高效率和高精度需求。深度学习技术凭借强大的特征学习能力,显著提升了 PCB 缺陷检测的自动化水平和精度,成为该领域的重要发展方向^[3]。

当前深度学习目标检测算法主要分为两大类:一阶段算法和二阶段算法^[4]。二阶段算法主要以 R-CNN^[5],Fast R-CNN^[6]和 Faster R-CNN^[7]等为代表,二阶段算法需要对预先生成的候选区域进行分类和定位,虽然精度较高,但是检测速度慢,无法满足 PCB 表面缺陷检测的实时性需求,不适合部署在相关系统中。一阶段算法主要以 SSD^[8-9],YOLO 系列^[10-11]为代表,由于直接在图像上预测类别和边界框,无需先生成候选区域,一阶段算法具有更快的检测速度和更好的实时性,适合部署在 PCB 表面缺陷检测场景中。

尽管已有研究对 PCB 表面缺陷检测提出了多种改进方法,但现有模型仍存在局限性。例如,王龙业等^[12]基于 YOLOv5 引入 SE(squeeze-and-excitation)注意力机制和加权双向特征金字塔网络(bidirectional feature pyramid network,BIFPN),并增加小目标检测层以检测更小的 4×4 目标缺陷,但 SE 注意力机制过度依赖通道特征的加权,可能导致对空间信息的忽略,从而降低模型的鲁棒性,并且新网络结构和检测层的引入增加了训练时间和模型参数,限制了算法的实际应用。郭艳等^[13]采用多维注意力机制、部分卷积(partial convolution, PConv)和 Alpha-SIoU 损失函数改进 YOLOv7,但多维注意力机制大幅增加了模型的计算复杂度,PConv 的引入无法保证特征提取的完整性。此外,Alpha-SIoU 损失函数降低了模型对低交并比(intersection over union, IoU)样本的优化能力,导致检测精度有限。胡兰兰等^[14]基于 YOLOv5s 引入多尺度加权通道融合和轻量化模块,但其检测精度与实时性仍无法匹配最新的 YOLO11。

总体而言,这些研究中存在以下不足:使用的基线模型相对陈旧,整体参数量过大,难以满足电子工业系统对实时性的要求,因此不适用于当前 PCB 表面缺陷检测。

在 YOLO11 改进模型方面,尽管已有研究取得了一定进展,但面向 PCB 表面缺陷检测暂无较好的模型。王国国等^[15]提出基于 YOLO11 s 的轻量化算法,用于绝缘子缺陷检测,该算法引入了轻量化的 GhostConv 模块、CA 注意力机制以及改进的 SioU 损失函数。然而,GhostConv 生成的特征图存在冗余,难以捕捉 PCB 表面复杂且微小的缺陷特征。此外,CA 注意力机制和 SioU 的引入增加了模型复杂度,不利于算法的高效部署。Huang 等^[16]针对合成孔径雷达(synthetic aperture radar,SAR)图像检测提出了一种改进的 YOLO11 模型,该模型引入了改进的 Swin Transformer 模块以减少冗余信息,并应用跨阶段部分连

接(cross-stage partial,CSP)思想改进原 SPPF 模块,增强多尺度特征提取能力。但这种改进后的模型对计算资源的需求较高,难以满足 PCB 表面缺陷检测的实时性要求。Li 等^[17]针对盲人障碍物检测改进 YOLO11,采用部分卷积前馈网络(partial convolutional feed-forward network,PCFN)减少传统卷积网络中对无效信息的重复计算,并引入跨尺度注意力融合(cross-scale attention fusion,CSAF)机制以增强模型对不同大小障碍的鲁棒性。然而,该模型的泛化能力尚未在更复杂实际场景中得到验证。

综上所述,上述基于 YOLO11 的改进模型虽然在特定场景下取得了进展,但未针对 PCB 表面缺陷检测场景进行特定的优化。因此,这些模型在当前 PCB 表面缺陷检测任务中存在无法在高准确性和高实时性之间取得良好平衡的局限性。

基于已有改进模型在 PCB 表面缺陷检测中准确度有限,难以同时满足轻量化、实时性与高检测准确率的问题,本文选取 YOLO11 作为基线模型,针对其骨干网络计算成本高、特征提取能力不足和小目标检测准确率有限的问题,提出一种改进 YOLO11 的 PCB 表面缺陷检测方法。本文贡献如下:1)骨干网络重新设计:采用优化的 HGNetV2^[18]替换原骨干网络,实现模型的轻量化;2)借鉴动态卷积(dynamic convolution, Dynamicconv)的思想^[19]对 HGNetV2 中的混合通用块(hybrid generalized block,HGBlock)进一步创新,增强模型特征提取能力,平衡轻量化与检测精度;3)引入双域选择注意力机制(dual-domain selection mechanism,DSM)^[20],放大关键区域的信息响应,增强模型的小目标检测能力。实验结果证明,本文方法在降低模型复杂度的同时实现了较高的检测准确率,有助于及时发现 PCB 表面缺陷,确保现代电子制造产业的高效稳定运行。

1 YOLO11 模型

YOLO11 是由 Ultralytics 团队创建并维护的 YOLO 系列最新算法,继承了单阶段检测的核心特点,能够在一次前向传递中预测图像中所有目标的类别和边界框。通过优化模型架构和训练策略,YOLO11 在检测精度和速度上进行了改进,实现了检测速度、精度和模型大小之间的更优平衡,被广泛应用于目标检测领域。

YOLO11 相比 YOLOv8、YOLOv10 等前作的创新主要体现在以下几个方面。首先,YOLO11 引入了 C3k2 模块,这是一种更快速的 CSP 模块变体,基于 C3 模块优化。C3k2 模块通过分割特征图并应用一系列较小的内核卷积来优化信息流,虽然计算效率比大内核卷积更高,但其多层次卷积操作仍需消耗较多计算资源。此外,C3k2 模块可以通过参数指定实现两个 C3k 模块的串联或退化为 YOLOv8 中的 C2f 模块,并且增加特征提取的深度,间接提升检测精度。

其次,YOLO11 新增了 C2PSA 模块,这是一个具有注

意力机制的特征提取和处理模块。C2PSA 模块结合了 CSP 结构和部分自注意力 (partial self-attention, PSA) 机制, 特征经过一次 1×1 卷积后, 一部分直接传递, 另一部分进入 PSA 模块处理。PSA 模块中的多种卷积核提取多尺度特征, 并使用 SE 模块为特征通道加权, 最后通过 Softmax 实现通道逐点加权, 提升对重要特征的关注, 增强模型在复杂场景下对多尺度物体的检测性能。

此外, YOLO11 还借鉴 YOLOv10 的思想, 将深度可分离卷积引入检测头, 减少冗余计算。这种优化进一步提高了模型的计算效率, 使其在保持高精度的同时具备更快的推理速度。

然而, YOLO11 基线模型仍存在一些需要改进的地方。首先, 骨干网络的计算成本较高。尽管 C3k2 模块通过分割特征图并应用小内核卷积优化了信息流, 但其多层次卷积操作仍需消耗较多计算资源。其次, 特征提取能力仍有提升空间, 骨干网络无法充分保留目标的细节特征, 导致特征提取不完整。最后, 小目标检测的准确率有限。小目标通常与复杂背景噪声相邻, 而 YOLO11 骨干网络生成的特征图分辨率较低, 难以捕捉小目标的细节信息, 使得检测准确率下降。

基于上述优缺点, 本文采用 YOLO11 作为改进的基线模型, 其网络结构如图 1 所示。

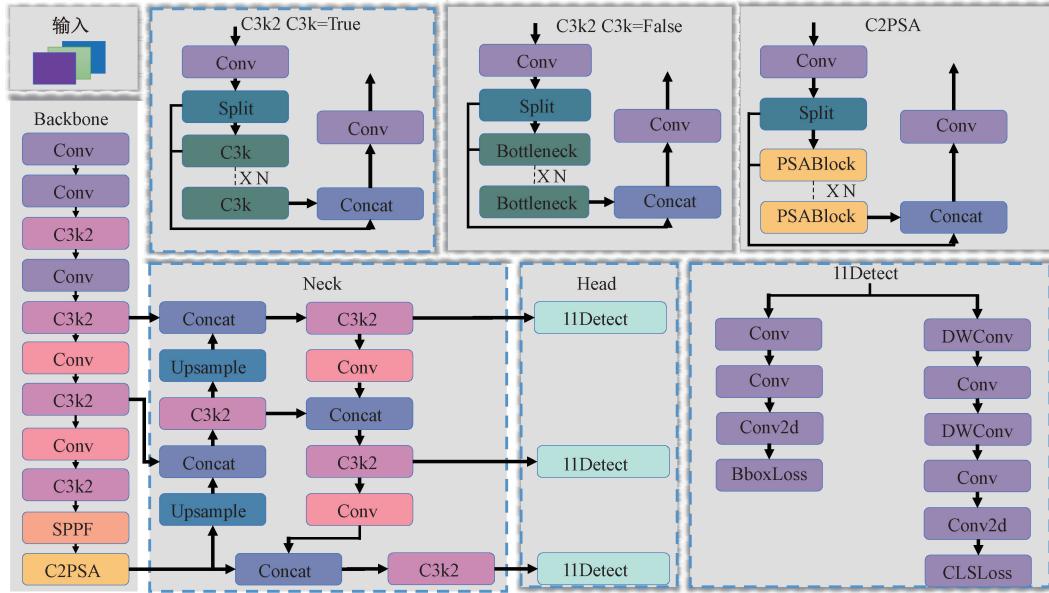


图 1 YOLO11 网络结构

Fig. 1 Architecture of YOLO11

2 改进 YOLO11 模型

本文针对 YOLO11 的骨干网络计算成本高、特征提取能力不足和小目标检测准确率有限 3 个重要缺点, 结合已有 YOLO11 改进鲁棒性弱、模型复杂和检测精度不足等问题做出如下改进。

2.1 HGNetV2 构建骨干网络

在 PCB 表面缺陷检测任务中, YOLO11 的原骨干网络因计算复杂度较高, 难以满足资源受限场景的部署需求。因此, 本文引入 RT-DETR 模型的 HGNetV2 骨干网络。HGNetV2 是一种轻量化设计的骨干网络, 采用层次化特征提取机制, 能够在不同尺度上学习 PCB 表面缺陷的特征表示。这种多尺度学习能力通过整合不同尺度的特征信息, 使模型能够有效处理从微观缺陷到宏观缺陷的检测任务, 从而提升模型对不同尺度目标的检测能力, 同时增强了对复杂背景的鲁棒性。例如, 在复杂工业环境中, PCB 表面的缺陷可能呈现多种尺寸和形状, 多尺度特征融

合能够有效捕捉这些差异, 提高模型对小目标和大目标的检测精度。此外, 多尺度特征融合还能增强模型对复杂背景的鲁棒性, 尤其是在面对多样化纹理和丝印等背景噪声时, 通过融合不同尺度的特征, 模型能够更好地分离目标与背景, 减少误检和漏检。

在具体实现上, HGNetV2 主要由 HGStem 预处理层模块、HGBlock 数据处理模块和深度可分离卷积 (depthwise separable conv, DWConv) 组成。

HGStem 作为 HGNetV2 网络的预处理模块, 采用多级卷积与池化操作的组合架构, 实现了对输入数据的高效特征提取与降维。该模块通过轻量化卷积层提取局部特征, 并利用最大池化操作进行渐进式跨层特征融合; 在每次下采样前通过特征拼接 (feature concatenation) 将不同尺度的局部特征进行融合, 使浅层高分辨率特征与深层语义特征形成互补。这种多尺度特征融合机制有效保留了原始输入的空间细节信息, 同时整合了经过非线性变换的高级语义特征。在降低特征图分辨率的过程中, 通过跨层跳

跃连接将底层特征注入深层网络,显著增强了模型对微小目标的特征响应能力。并在严格控制参数量和计算复杂度的前提下,实现了多尺度特征的有效提取,为后续网络层提供了丰富的特征表示基础。HGStem 模块结构如图 2 所示。

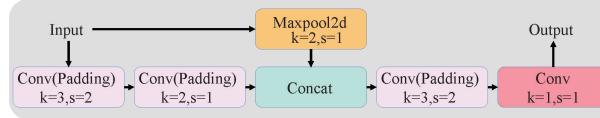


图 2 HGStem 结构

Fig. 2 Architecture of HGStem

HGBlock 作为 HGNetV2 网络的核心构建模块,采用并行化的层次化处理架构,实现了对输入数据的多尺度特征提取与融合。该模块由卷积子模块和通道压缩解压缩子模块并联组成,通过分层处理机制能够捕获不同层次的特征表示。具体而言,每个 HGBlock 包含多个特征处理层次,每个层次通过特定范围的感受野提取相应尺度的特征信息,从而实现对低级细节特征和高级语义特征的有效学习。模块内部集成多种尺寸的卷积滤波器,包括轻量化卷积(LightConv)和标准卷积(Conv),在保证特征提取效果的同时优化了计算效率。这种设计不仅增强了网络对细节特征的敏感度,还通过通道压缩解压缩机制实现了特征信息的有效筛选与增强,为后续任务提供了更具判别性的特征表示。HGBlock 模块结构如图 3 所示。

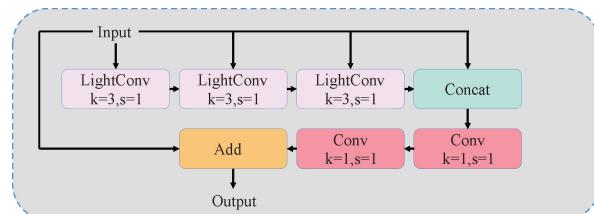


图 3 HGBlock 结构

Fig. 3 Architecture of HGBlock

HGNetV2 中使用 DWConv 替换普通卷积,参数量和计算量更低。对于普通卷积,可以用公式表示为:

$$\begin{cases} P = C_{in} \times C_{out} \times K \times K \\ F = C_{in} \times C_{out} \times H \times W \times K \times K \end{cases} \quad (1)$$

式中: H 、 W 和 C_{in} 为输入特征图的高、宽和输入通道数, $K \times K$ 为卷积核的宽高尺寸, C_{out} 为输出通道数, P 和 F 为普通卷积核的参数量和计算量。

对于 DWConv,公式为:

$$\begin{cases} P_{DW} = C_{in} \times K^2 + C_{in} \times C_{out} \\ F_{DW} = C_{in} \times H \times W \times K^2 + C_{in} \times C_{out} \times H \times W \end{cases} \quad (2)$$

式中: P_{DW} 和 F_{DW} 为 DWConv 的参数量和计算量。

由式(1)、(2)可以得到:

$$\frac{P_{DW}}{P} = \frac{F_{DW}}{F} = \frac{1}{C_{out}} + \frac{1}{K^2} \quad (3)$$

由式(3)可知,DWConv 的参数量和计算量仅有普通卷积的 $\frac{1}{C_{out}} + \frac{1}{K^2}$,实现了网络轻量化。DWConv 结构如图 4 所示。

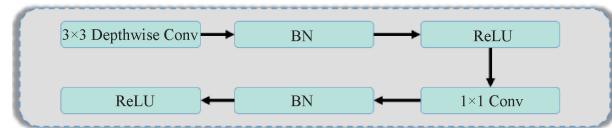


图 4 DWConv 结构

Fig. 4 Architecture of DWConv

原 HGNetV2 作为 RT-DETR 的骨干网络,表现出优秀的检测性能和充分的灵活性,本文结合 YOLO11 网络特点,优化 HGNetV2 并替换原 YOLO11 的骨干网络。降低模型的计算量和参数量,进而实现网络的轻量化,同时增强了模型的鲁棒性。

2.2 Dynamic convolution 改进 HGBlcok

通过精心设计的 HGNetV2 架构,本文成功构建了一种创新的 YOLO11 骨干网络。该网络在显著降低模型参数量和计算量的同时,实现了计算效率的显著提升。然而,这种轻量化设计不可避免地导致模型的检测精度出现一定程度的下降。为了有效提升检测精度,本文借鉴 Dynamicconv 思想,对骨干网络中的 HGBlock 进行了进一步的改进与创新。

Dynamicconv 的核心思想是在保持较低计算量的前提下,通过动态组合多个卷积核来引入更多的参数。这使得网络能够学习到更丰富、更具区分性的特征表示,进而显著增强模型的特征提取能力。同时,由于 Dynamicconv 的系数是根据输入动态生成的,网络通过动态专家针对不同输入生成卷积权重张量和动态系数,实现自适应地调整卷积核的组合方式,从而有效增强模型的鲁棒性。

在具体实现上,给定输入特征 $X \in \mathbb{R}^{C_{in} \times H \times W}$ 和权重张量 $W \in \mathbb{R}^{C_{out} \times C_{in} \times K \times K}$,传统卷积层可以用公式表示为:

$$Y = X * W \quad (4)$$

式中: $Y \in \mathbb{R}^{C_{out} \times H' \times W'}$ 为卷积层输出。

为了增加模型的参数量,Dynamicconv 引入参数增强函数,可以用公式表示为:

$$W' = f(W) \quad (5)$$

式中: W 为原权重张量, W' 为参数增强后的权重张量, f 为参数增强函数。该函数满足两条基本原则:1)不需要过多计算成本;2)能够增加模型容量或可训练参数。

Dynamicconv 通过多个动态专家(dynamic experts)实现,带有 M 个动态专家的 Dynamicconv 可以用公式表示为:

$$\begin{cases} Y = X * W \\ W' = \sum_{i=1}^M \alpha_i W_i \end{cases} \quad (6)$$

式中: $\mathbf{W}_i \in \mathbb{R}^{C_{out} \times C_{in} \times H \times W}$ 为第 i 个动态专家生成的卷积权重张量, α_i 为相应的动态系数, 且满足 $\sum_{i=1}^M \alpha_i = 1$ 。

系数 α_i 根据不同的输入样本动态生成, Dynamicconv 采用多层感知机(multi-layer perceptron, MLP)根据输入生成 α_i , 对于输入 \mathbf{X} 应用全局平均池化将信息融合为一个向量, 然后使用 softmax 函数激活的双层 MLP 模块生成动态系数, 该步骤可以用公式表示为:

$$\boldsymbol{\alpha} = \text{softmax}(\text{MLP}(\text{Pool}(\mathbf{X}))) \quad (7)$$

式中: $\boldsymbol{\alpha} \in \mathbb{R}^M$ 为动态生成的系数向量。

对于传统卷积层, 参数量和计算量为:

$$\begin{cases} \text{Params} = C_{out} \times C_{in} \times K \times K \\ \text{FLOPs} = W' \times H' \times C_{out} \times C_{in} \times K \times K \end{cases} \quad (8)$$

式中: Params 为参数量, FLOPs 为计算量。

对于 Dynamicconv, 参数量和计算量为:

$$\begin{cases} \text{Params}' = M \times C_{out} \times C_{in} \times K \times K + \\ C_{in}^2 + M \times C_{in} \\ \text{FLOPs}' = M \times C_{out} \times C_{in} \times K \times K + W' \times H' \times \\ C_{out} \times C_{in} \times K \times K + C_{in}^2 + M \times C_{in} \end{cases} \quad (9)$$

式中: Params' 为参数量, FLOPs' 为计算量。

通常情况下, M 远小于 $W' \times H'$, 因此与传统卷积层相比, Dynamicconv 在最大程度引入更多参数的同时将 FLOPs 的增加降至最低, 而参数量则扩大至 M 倍, 有效增强模型对泛化特征的学习能力。

本文使用 Dynamicconv 替换 HGBlock 中的原始卷积层得到 Dynamic_HGBlock 模块, 从而有效缓解了使用 HGNetV2 作为骨干网络结构在目标检测中的精度下滑问题, 并进一步减少模型的计算量。改进后的 Dynamic_HGBlock 结构如图 5 所示。

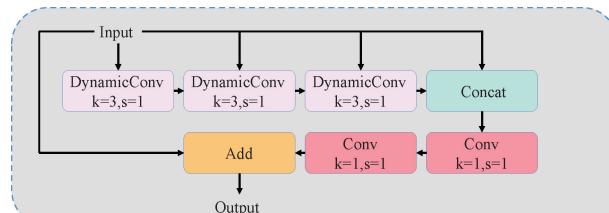


图 5 Dynamic_HGBlock 结构

Fig. 5 Architecture of Dynamic-HGBlock

2.3 DSM 注意力机制

PCB 表面输入图像的背景噪声大, 目标缺陷尺寸小, 为进一步增强 YOLOv11 的小目标检测能力, 本文引入一种新颖的注意力机制: DSM 注意力机制。该注意力机制通过空间域选择模块(spatial selection module, SSM)和频率域选择模块(frequency selection module, FSM)放大空间域和频率域的信息相应, 使网络聚焦于更重要的区域, 进而增强特征提取能力。SSM 结构如图 6 所示, FSM 结构如图 7 所示。

所示。

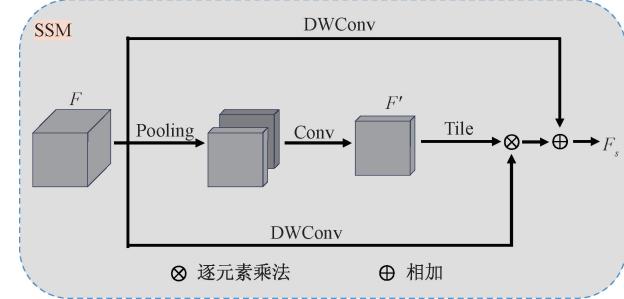


图 6 SSM 结构

Fig. 6 Architecture of SSM

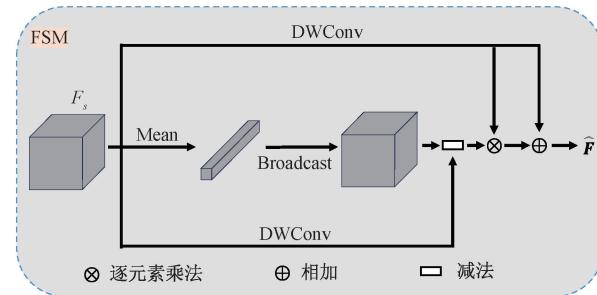


图 7 FSM 结构

Fig. 7 Architecture of FSM

在具体实现上, SSM 首先作用于输入特征, 为需要重点关注的位置生成一般特征表示。通过最大池化和平均池化对输入特征进行通道维度的压缩, 生成两个池化后的特征图, 将这两个特征图拼接, 再使用 3×3 卷积层生成通用特征图, 用于表示重点关注区域的位置信息。该步骤用公式表示为:

$$\mathbf{F}' = \text{Conv}_3([\text{AvgPool}(\mathbf{F}), \text{MaxPool}(\mathbf{F})]) \quad (10)$$

式中: $\mathbf{F}' \in \mathbb{R}^{H \times W \times 1}$ 为通用特征图, $\mathbf{F} \in \mathbb{R}^{H \times W \times C}$ 为输入特征图, $[,]$ 表示特征图的拼接操作, Conv_3 为 3×3 卷积层, AvgPool 为平均池化, MaxPool 为最大池化。

为进一步增强对每个通道关键区域的注意力, SSM 使用 DWConv 对输入特征进行通道分离操作通过卷积核大小为 5×5 和 7×7 的 DWConv 作用于输入特征生成高频特征, 卷积核大小为 3×3 的 DWConv 生成低频特征, 再将高频特征与通用特征图进行逐元素乘法, 并加上低频特征, 输出空间选择后的特征。该步骤可以用公式表示为:

$$\mathbf{F}_s = \text{DWConv}_{5,7}(\mathbf{F}) \otimes T(\mathbf{F}', \mathbf{C}) + \text{DWConv}_3(\mathbf{F}) \quad (11)$$

式中: $\mathbf{F}_s \in \mathbb{R}^{H \times W \times C}$ 为输出特征, $\text{DWConv}_{5,7}$ 为核大小为 5×5 和 7×7 的级联卷积层, \otimes 为逐元素乘法, $T(\mathbf{F}', \mathbf{C})$ 表示将 \mathbf{F}' 沿通道维度复制 C 次到 $\mathbf{F}' \in \mathbb{R}^{H \times W \times C}$ 的 Tile 函数。

FSM 接收 SSM 输出的特征 \mathbf{F}_s , 在频率域中进一步强调重要信息。通过全局平均池化对 \mathbf{F}_s 进行通道维度的全

局平均操作,生成低频特征。从 F_s 中减去低频特征得到高频特征,将高频特征与 F_s 进行逐元素乘法,再加上 F_s ,生成输出特征。该步骤可以用公式表示为:

$$\begin{cases} F_s^h = F_s - \text{Mean}(F_s) \\ \hat{F} = F_s^h \otimes F_s + F_s \end{cases} \quad (12)$$

式中: F_s 为 SSM 输出特征, F_s^h 为高频特征, Mean 为全局平均池化, \otimes 为逐元素乘法, \hat{F} 为输出特征。

本文在改进后的 HGNetV2 骨干网络的末端 C2PSA 模块后添加 DSM 注意力机制层,进一步提取 C2PSA 输出特征图中的关键特征,进而增强骨干网络对 PCB 图像的小目标检测能力。

2.4 HDH-YOLO 模型

本文提出了一种基于 YOLO11n 的改进模型:HDH-YOLO,实现了在轻量化基础上的性能提升,为 PCB 表面缺陷检测任务提供了一种高效且准确的解决方案。该模型的改进集中在以下 3 个方面:

1) 在 YOLO11n 的骨干网络中,本文将原有的骨干网络替换为优化后的 HGNetV2 架构。这一改进显著降低了模型的参数量和计算量,实现了网络结构的轻量化。通过这种轻量化设计,模型在保持较高运行效率的同时,能够更好地适应资源受限的场景。

2) 针对 HGNetV2 网络中的 HGBlock 模块,本文引入了 Dynamicconv 技术创新改进,设计了 Dynamic_HGBlock。将改进后的 Dynamic_HGBlock 替换 HGNetV2 中的后四层 HGBlock,从而增强网络对泛化特征的学习能力。这一改进不仅有助于提高模型对 PCB 表面缺陷的检测准确率,还能在保持较低计算量的前提下引入更多参数,进一步提升模型对复杂特征的提取能力。

3) 在 C2PSA 模块后,本文引入了 DSM 注意力机制层。该机制通过强调关键区域的信息响应,增强模型在小目标、多噪声背景条件下的鲁棒性。这一改进进一步提升了模型对 PCB 表面缺陷的检测准确率,使其在复杂工业场景中表现出色。

改进后的 HDH-YOLO 结构如图 8 所示。

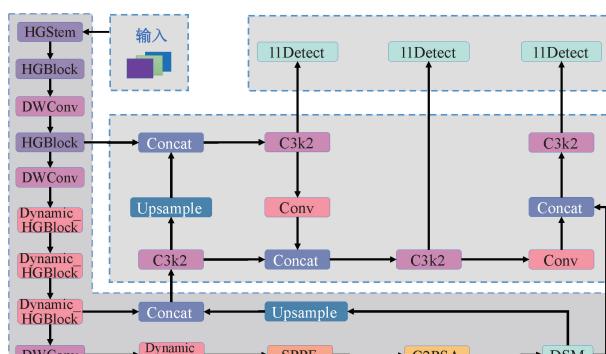


图 8 HDH-YOLO 结构

Fig. 8 Architecture of HDH-YOLO

3 实验结果与分析

3.1 实验环境

本文实验全部基于 Linux20.04 操作系统,采用 Python3.8.0 和 Pytorch2.0.0 构建深度学习模型。处理器为 15 vCPU Intel(R) Xeon(R) Platinum 8474C,硬件平台为 NVIDIA GeForce RTX 4090D GPU,并通过 CUDA 11.7 实现 GPU 加速,深度学习模型参数设置如表 1 所示。

表 1 模型参数

Table 1 Model parameters

参数	设置	参数	设置
类别数量	6	学习阈值	0.01
输入宽	640	WARMUP	True
输入高	640	批大小	32
学习率	0.001 25	训练轮数	300

3.2 实验数据集

本文实验采用北京大学智能机器人开放实验室公开的印刷电路板瑕疵数据集(PKU-Market PCB)进行实验。原数据集共有 1 386 张图像,本文根据标注完整性和数据平衡性选取适用于缺陷检测任务的 693 张裸板缺陷图片,其中包含漏孔、鼠咬、开路、短路、杂散、杂铜 6 类缺陷,共 2 953 个标注实例,且一张图片中只包含一个缺陷类别。实例分布如表 2 所示。本文实验按照 8:1:1 划分为训练集、验证集和测试集。

表 2 数据集实例分布

Table 2 Distribution of instances in the dataset

实例名称	实例数量	实例占比/%
漏孔	497	16.83
鼠咬	492	16.66
开路	482	16.32
短路	491	16.63
杂散	488	16.53
杂铜	503	17.03

3.3 评价指标

为验证本文改进模型对 PCB 表面缺陷检测的有效性,选取准确率(precision, P)、召回率(recall, R)、IoU 为 0.5 时的平均精度均值(mean average precision, mAP)mAP50 和 IoU 从 0.5 到 0.95(每隔 0.05 为一个步长)范围内计算的 mAP 值的平均值 mAP50-95 作为模型对 PCB 表面缺陷检测准确率的评价指标。其中,mAP50 和 mAP50-95 是本文检测准确率的关键评价指标。P、R、mAP 越接近 1 说明模型检测准确率越高,三者的计算公式如下:

$$TP = \frac{TP}{TP + FP} \quad (13)$$

$$R = \frac{TP}{TP + FN} \quad (14)$$

$$mAP = \frac{1}{n} \sum_{i=1}^N AP_i = \frac{1}{n} \sum_{i=1}^N P dR \quad (15)$$

式中: TP 为检测缺陷正确的目标数量; FP 为检测错误缺陷的目标数量; FN 为漏检缺陷的目标数量; AP 为平均精度, 即 PR 曲线下的面积; N 为目标检测类别的数量。

由式(13)可知, P 可以反映模型的误检率, 即 P 越高, 说明模型误检的目标越少。由式(14)可知, R 可以反映模型的漏检率, 即 R 越高, 说明模型漏检的目标越少。在电

子制造工业系统中, 为了避免瑕疵产品的出现, 漏检率的降低尤为重要。

同时, 本文选取模型中可训练参数的总数参数量与模型在推理过程中所需的浮点运算次数计算量作为模型效率的评价指标, 参数量与计算量越小说明模型越高效。

3.4 消融实验

为了验证本文提出的改进 PCB 表面缺陷检测方法所做的各项改进的有效性, 对各项改进进行消融实验。实验之间严格控制模型参数一致, 消融实验结果如表 3 所示。

表 3 消融实验结果

Table 3 Results of ablation experiment

模型	Params/ 10^6	GFLOPs	mAP50/%	mAP50-95/%	P/%	R/%
YOLO11n	2.58	6.3	90.3	48.0	93.5	86.0
+ HGNetV2	2.14	5.7	87.8	43.8	87.9	82.4
+ HGNetV2+Dynamic_HGBlock	2.39	5.5	92.7	49.2	96.7	88.9
+ HGNetV2+Dynamic_HGBlock+DSM	2.42	5.5	92.9	50.3	94.5	92.4

由表 2 可知, 使用优化后的 HGNetV2 替换原 YOLO11 骨干网络后, 模型参数量和计算量分别减少 17.05% 和 9.52%, 说明替换的骨干网络有效实现了轻量化的目标, 但是 mAP 下降过多, 无法满足 PCB 表面缺陷检测场景准确率的要求。而引入 Dynamic_HGBlock 后 mAP50 和 mAP50-95 较基线模型分别提升 2.4% 和 1.2%, 并且模型的误检率达到最低, 漏检率也有一定的下降, 而模型的参数量和计算量分别减少 7.36% 和 12.70%, 说明 Dynamic_HGNetV2 为网络引入更多参数, 增强了模型的特征提取能力。在二者基础上添加 DSM 注意力机制

层后, 通过放大关键区域信息响应进一步提升模型对 PCB 表面微小缺陷的检测能力, mAP50 和 mAP50-95 分别达到最高的 92.9% 和 50.3%, 且模型的漏检率达到最低, 确保了对缺陷 PCB 的及时发现。而模型参数量和计算量较基线模型分别下降 6.20% 和 12.70%, 说明本文模型能较好地平衡 PCB 表面缺陷检测场景下实时性和检测准确性的要求。

3.5 模型对比实验

为了证明本文提出的 HDH-YOLO 模型的先进性, 将 HDH-YOLO 与主流一阶段目标检测算法进行对比分析, 实验之间严格控制参数一致, 实验结果如表 4 所示。

表 4 不同模型实验结果对比

Table 4 Comparison of different models

模型	Params/ 10^6	GFLOPs	mAP50/%	mAP50-95/%	P/%	R/%
YOLOv5n	2.18	5.8	92.4	46.9	95.3	88.4
YOLOv6n	4.16	11.5	89.6	45.3	93.5	85.3
YOLOv8n	2.69	6.8	91.5	48.4	95.0	89.4
YOLOv10n	2.27	6.5	91.0	48.7	91.1	84.7
ASF-YOLO ^[21]	2.24	6.3	92.6	48.3	97.6	88.5
文献[15]	2.23	5.5	91.2	48.3	94.2	86.2
YOLO11n	2.58	6.3	90.3	48.0	93.5	86.0
HDH-YOLO	2.42	5.5	92.9	50.3	94.5	92.4

由表 3 可知, 本文模型与 YOLO 系列经典模型相比: 参数量仅略高于 YOLOv5n, 而计算量为最低; 与 YOLOv8n 相比参数量和计算量分别降低 10.03% 和 19.12%, 而 mAP50 和 mAP50-95 分别提升 1.4% 和 1.9%; 与 YOLO11n 相比, 参数量和计算量分别降低 6.20% 和 12.70%, 漏检率和误检率均有大幅下降, mAP50

和 mAP50-95 分别提升 2.6% 和 2.3%; 相比除本文模型外综合性能较好的 YOLOv5n, mAP50 和 mAP50-95 分别提升 0.5% 和 3.4%, 充分证明了本文改进的先进性。与其他 YOLO 改进模型相比: 虽然参数量和误检率略高于 ASF-YOLO, 但计算量降低 12.70%, 漏检率大幅下降, 且 mAP50 和 mAP50-95 分别提升 0.3% 和 2.0%, 综合性能

更优;除参数量略高于文献[15]中的 YOLO11 改进模型,其他指标均优于该改进模型。值得注意的是,HDH-YOLO 在要求更严格的 mAP50-95 指标和关键的漏检率上相比其他模型有着更大的优势,说明本文模型可以回收最多的缺陷 PCB 并满足更高精度的检测需求。结果表明,本文模型在保证轻量化的同时,显著提升 PCB 表面缺陷检测的准确率并大幅降低了对缺陷 PCB 的漏检率,较好地平

衡了模型轻量化和检测精度,在现代电子制造系统中具有高准确性和高实时性等优点。

3.6 实验结果可视化分析

为了直观的对比本文提出的 HDH-YOLO 模型与基线 YOLO11n 模型的检测效果,从测试集中随机抽取部分图像进行测试,测试结果如图 9 所示。

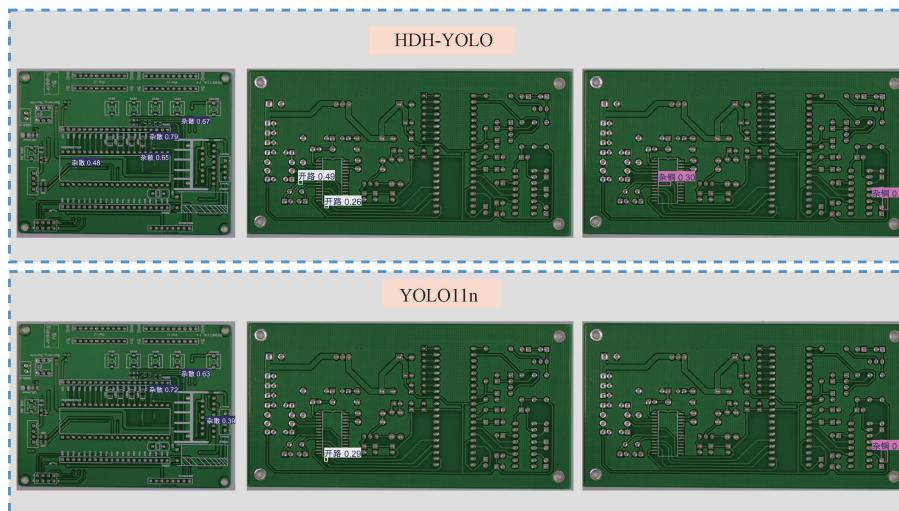


图 9 YOLO11n 与 HDH-YOLO 检测结果对比

Fig. 9 Comparison of detection results between YOLO11n and HDH-YOLO

由图 9 可知,在不同的缺陷种类下本文提出的 HDH-YOLO 模型相比基线模型 YOLO11n 均有检测精度的提升。具体而言,HDH-YOLO 能检测出原 YOLO11n 模型无法检测到的目标,有效降低了模型漏检率,这表明 HDH-YOLO 模型在复杂工业场景中具有更强的适应性和鲁棒性,尤其是在处理小目标、低对比度缺陷以及复杂背景下的检测任务时,其性能优势更为明显。

3.7 公开数据集对比

PKU-Market PCB 数据集规模相对较小,增加了模型出现过拟合的风险,严重限制模型的实际应用。为了验证 HDH-YOLO 的泛化能力,本文采用更具挑战性的

DeepPCB 数据集进行对比试验。该数据集包含 6 种常见的 PCB 缺陷:开路、短路、鼠咬、毛刺、铜和针孔,共 1 500 张图像,按照 8:1:1 随机划分为训练集、验证集和测试集。不同于 PKU-Market PCB 数据集,DeepPCB 数据集每张缺陷图中包含若干个不同种类的缺陷,这种复杂的缺陷分布情况更加贴近实际生产环境中的 PCB 缺陷场景。并且所有图像都是从线性扫描电荷耦合器件(charge-coupled device,CCD)中获得的,具有较高的分辨率和清晰度,能够真实地反映 PCB 缺陷的细节特征。两个数据集区别较大,可以很好的检验模型的泛化性能。实验设置与 3.1 节保持一致,实验结果如表 5 所示。

表 5 DeepPCB 数据集实验结果对比

Table 5 Comparison of DeepPCB dataset experimental results

模型	Params/ 10^6	GFLOPs	mAP50/%	mAP50-95/%	P/%	R/%
YOLOv5n	2.18	5.8	98.4	80.2	96.8	95.2
YOLOv6n	4.16	11.5	98.0	77.2	96.0	93.3
YOLOv8n	2.69	6.8	97.9	75.3	95.9	93.4
YOLOv10n	2.27	6.5	98.3	78.6	97.7	94.8
YOLO11n	2.58	6.3	98.6	77.7	95.7	96.2
ASF-YOLO ^[21]	2.24	6.3	98.5	80.3	96.9	95.3
文献[15]	2.23	5.5	98.3	78.1	97.2	95.0
HDH-YOLO	2.42	5.5	98.6	80.4	97.7	95.4

由表 4 可知,本文提出的 HDH-YOLO 模型相比 YOLO11n 在模型参数量、计算量分别降低 6.20%、12.70% 的同时,有效降低误检率和漏检率,并在要求更严格的 mAP50-95 指标提升 2.7%,且只有参数量略高于 YOLOv5n,其他指标均优于对比模型。这说明本文提出的 HDH-YOLO 模型相比 YOLO11n 及其改进,在 DeepPCB 数据集上综合性能更优,充分证明了改进模型的泛化性。

4 结 论

针对 PCB 表面缺陷检测准确率不足,无法充分平衡模型检测准确性与实时性,难以满足现代电子制造系统稳定运行要求的问题,提出了一种改进 YOLO11n 的 PCB 表面缺陷检测方法。主要结论为:引入 HGNetV2 替换骨干网络,实现模型轻量化;采用 Dynamicconv 的思想改进 HGNetV2 中的 HGBLOCK,在不增加过多计算量的前提下引入更多参数,增强模型的特征提取能力;在骨干网络末端引入 DSM 注意力机制层,放大关键区域的空间与频率响应,进而提高模型对微小缺陷的检测能力。综上所述,本文提出的改进 YOLO11 的 PCB 表面缺陷检测方法相较于原始 YOLO11n 模型及其现有改进具有更高的检测准确性与实时性,有助于 PCB 表面缺陷检测智能化的实现。

参考文献

- [1] 姜媛媛,蔡梦南.轻量化的印刷电路板缺陷检测网络 Multi-CR YOLO[J].电子测量与仪器学报,2023,37(11):217-224.
- [2] JIANG Y Y, CAI M N. Lightweight PCB defect detection network Multi-CR YOLO [J]. Journal of Electronic Measurement and Instrumentation, 2023, 37(11): 217-224.
- [3] 陈锦妮,拜晓桦,李云红,等.基于 PA-YOLOv5 的印制电路板缺陷检测[J].红外技术,2024, 46 (6): 654-662.
- [4] CHEN J N, BAI X H, LI Y H, et al. PCB defect detection based on PA-YOLOv5 [J]. Infrared Technology, 2024, 46(6): 654-662.
- [5] 孙铁强,于洪健,张灿,等.基于 YOLO 提示与 ICT-ViT 的 PCB 缺陷高效分割技术研究[J/OL].激光与光电子学进展,1-19[2025-02-10].<http://kns.cnki.net/kcms/detail/31.1690.TN.20250103.1411.096.html>.
- [6] SUN T Q, YU H J, ZHANG C, et al. Research on efficient PCB defect segmentation technology based on YOLO prompts and ICT-ViT [J/OL]. Laser & Optoelectronics Progress, 1-19[2025-02-10].<http://kns.cnki.net/kcms/detail/31.1690.TN.20250103.1411.096.html>.
- [7] [4] 吕秀丽,杨昕升,曹志民.改进 YOLOv8 的 PCB 表面缺陷检测算法[J].电子测量技术,2024, 47(12): 100-108.
- [8] LYU X L, YANG X SH, CAO ZH M. Improved PCB surface defect detection algorithm for YOLOv8 [J]. Electronic Measurement Technology, 2024, 47 (12): 100-108.
- [9] [5] YASIR M, LIU S W, XU M M, et al. ShipGeoNet: SAR image-based geometric feature extraction of ships using convolutional neural networks [J]. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62(1):1-13.
- [10] GIRSHICK R. FastR-CNN [C]. IEEE International Conference on Computer Vision, 2015:1440-1448.
- [11] REN SH Q, HE K M, ROSS G, et al. FasterR-CNN: Towards real-time object detection with region proposal networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39 (6): 1137-1149.
- [12] [6] 李波,李志康,周钰彬.结合特征融合和注意力机制的 SAR 舰船检测[J].电子测量技术,2024, 47 (10): 134-140.
- [13] LI B, LI ZH K, ZHOU Y B. SAR ship detection algorithm combining feature fusion and attention mechanism[J]. Eleclronic Measurement Technology, 2024, 47(10): 134-140.
- [14] [7] LI X S, WANG C L, ZENG Z. WS-SSD: Achieving faster 3D object detection for autonomous driving via weighted point cloud sampling[J]. Expert Systems with Applications,2024,249(1):123805.
- [15] [8] [9] [10] 余浪,苗鸿宾,苏赫朋,等.基于改进 YOLOv3 的轴承端面缺陷检测算法[J].机床与液压,2024, 52(9): 209-214.
- [16] YU L, MIAO H B, SU H P, et al. Defect detection algorithm of bearing end face based on improved YOLOv3 [J]. Machine Tool & Hydraulics, 2024, 52(9): 209-214.
- [17] [11] WANG J X, LIU M, DU Y R, et al. PG-YOLO: An efficient detection algorithm for pomegranate before fruit thinning [J]. Engineering Applications of Artificial Intelligence,2024,134(1):108700.
- [18] 王龙业,黄鋆,曾晓莉.基于 YOLO-PCB 的印刷电路板裸板缺陷检测[J].科学技术与工程,2024, 24(15): 6338-6345.
- [19] WANG L Y, HUANG Y, ZENG X L. Bare board defect detection of printed circuit board based on YOLO-PCB[J]. Science Technology and Engineering, 2024, 24(15): 6338-6345.

- [13] 郭艳,王智文,赵润星. YOLO-POD: 基于多维注意力机制的高精度 PCB 微小缺陷检测算法[J]. 电子学报, 2024, 52(7): 2515-2528.
- GUO Y, WANG ZH W, ZHAO R X. YOLO-POD: High-precision PCB tiny-defect detection algorithm based on multi-dimensional attention mechanism[J]. Acta Electronica Sinica, 2024, 52(7): 2515-2528.
- [14] 胡兰兰,邓超. 基于 YOLO-MCG 的 PCB 缺陷检测算法 仪表技术与传感器[J]. 仪表技术与传感器, 2024(4): 100-106.
- HU L L, DENG CH. PCB defect detection algorithm based on YOLO-MCG[J]. Instrument Technique and Sensor, 2024(4): 100-106.
- [15] 王国国,龚小玉,袁发庭. 基于 YOLO11s 绝缘子缺陷检测的轻量化算法[J/OL]. 重庆工商大学学报(自然科学版), 1-10 [2025-02-12]. <http://kns.cnki.net/kcms/detail/50.1155.N.20241226.1428.003.html>.
- WANG G G, GONG X Y, YUAN F T. Lightweight algorithm based on YOLO11s insulator defect detection[J/OL]. Journal of Chongqing Technology and Business University(Natural Science Edition), 1-10 [2025-02-12]. <http://kns.cnki.net/kcms/detail/50.1155.N.20241226.1428.003.html>.
- [16] HUANG Y Y, WANG D, WU B X, et al. NST-YOLO11: ViT merged model with neuron attention for arbitrary-oriented ship detection in SAR images [J]. Remote Sensing, 2024, 16(24), DOI:10.3390/rs16244760.
- [17] LI J CH, ZHENG M L, DONG D Y, et al. PC-CS-YOLO: High-precision obstacle detection for visually impaired safety [J]. Sensors, 2025, 25 (2), DOI: 10.3390/s25020534.
- [18] ZHAO Y, LYU W Y, XU SH L, et al. DETRs beat YOLOs on real-time object detection[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024.
- [19] HAN K, WANG Y H, GUO J Y, et al. ParameterNet: Parameters are all you need for large-scale visual pretraining of mobile networks[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR), 2024.
- [20] CUI Y N, REN W Q, CAO X CH, et al. Focal network for image restoration[C]. 2023 IEEE/CVF International Conference on Computer Vision (ICCV), 2023.
- [21] KANG M, TING C M, TING F F, et al. ASF-YOLO: A novel YOLO model with attentional scale sequence fusion for cell instance segmentation [J]. Image and Vision Computing, 2024, 147: 105057.

作者简介

- 吴葛(通信作者),工程师,主要研究方向为自动控制原理、电力电子学、信号与控制。
E-mail: gewu@hust.edu.cn
- 朱宇凡,本科,主要研究方向为深度学习、机器视觉。
E-mail: zhuyufan@hust.edu.cn
- 贾泽宁,本科,主要研究方向为深度学习、机器视觉。
E-mail: jiazening@hust.edu.cn