

文章编号:1007-2780(2026)05-0700-17

基于改进 DEIM 的印刷电路板缺陷检测算法

高雪^{1*}, 袁飞², 霍艳飞³

- (1. 常州工业职业技术学院 智能控制学院, 江苏 常州 213164;
2. 常州工业职业技术学院 邓建军工匠学院, 江苏 常州 213164;
3. 大连海洋大学 应用技术学院, 辽宁 大连 116300)

摘要:针对印刷电路板(Printed Circuit Board, PCB)表面缺陷目标尺寸小、背景复杂、分布密集以及现有检测算法难以兼顾精度与速度的问题,本文提出基于OL-HAIM(OverLoCK-HAFB-APKConv with Improved Matching)的PCB缺陷检测算法。该算法首先采用先概览后细看(OverLoCK)主干网络,通过深层级分解策略(DDS)和上下文混合动态卷积(ContMix)降低计算复杂度并增强小目标特征表达;其次设计了分层注意力融合模块(HAFB)替代基于CNN的跨尺度特征融合模块(CCFF)结构,通过局部-全局注意力机制提升微小缺陷检测能力;最后引入自适应逐点卷积(APKConv),通过可学习偏移量实现感受野自适应变形,精准捕获不同形态缺陷特征。实验结果表明,OL-HAIM在PCB-DEFECT数据集上的IoU阈值为0.5时的平均精度均值(mAP@0.5)达97.6%,较基线模型提升4.8个百分点;该模型参数量仅为13.8M,每秒浮点运算次数(FLOPS)降至23.6G,在RK3588嵌入式平台上的推理速度达55.7帧/s(FPS)。在NEU-DET(Northeastern University Surface Defect Detection Dataset)和MS COCO(Microsoft Common Objects in Context)数据集上的跨域实验验证了算法的泛化能力。在某PCB生产线实地采集样本上的现场测试中,模型的mAP@0.5达95.1%,于RK3588嵌入式平台的推理速度达55.4FPS,兼顾了实地场景下的高精度检测与实时性需求。本文方法为工业现场PCB缺陷实时检测提供了兼顾高精度与轻量化部署的有效方案。

关键词:PCB缺陷检测;DEIM;轻量化目标检测;边缘部署

中图分类号:TP391 **文献标识码:**A **doi:**10.37188/CJLCD.2026-0045 **CSTR:**32172.14.CJLCD.2026-0045

PCB defect detection algorithm based on improved DEIM

GAO Xue^{1*}, YUAN Fei², HUO Yanfei³

- (1. Intelligent Control College, Changzhou Vocational Institute of Industry Technology,
Changzhou 213164, China;
2. Deng Jianjun Craftsman College, Changzhou Vocational Institute of Industry Technology,
Changzhou 213164, China;
3. Applied Technology College of Dalian Ocean University, Dalian 116300, China)

Abstract: To address the challenges of small-sized targets, complex background, dense distribution of

收稿日期:2026-03-19;**修订日期:**2026-04-15.

基金项目:江苏省高等教育教改立项研究课题(重点)(No. 2023JSJG238);江苏省高等学校基础科学(自然科学)研究面上项目(No. 23KJB460002);常州市应用基础研究计划(No. CJ20241057)

Supported by Jiangsu Provincial Higher Education Teaching Reform Research Project (Key Project) (No. 2023JSJG238); Jiangsu Provincial Higher Education Basic Science (Natural Science) Research General Project (No. 23KJB460002); Changzhou Applied Basic Research Program (No. CJ20241057)

*通信联系人, E-mail:13585317253@163.com

Printed Circuit Board(PCB) surface defects, and the difficulty of existing detection algorithms in balancing accuracy and speed, this paper proposes a PCB defect detection algorithm based on OverLoCK-HAFB-APKConv with Improved Matching(OL-HAIM). First, an Overview-first-Look-Closely-next(OverLoCK) backbone network is adopted, which leverages a Deep-level Decomposition Strategy(DDS) and Context-Mixing Dynamic Convolution(ContMix) to reduce computational complexity while enhancing feature representation for small targets. Second, a Hierarchical Attention Fusion Block(HAFB) is designed to replace the CNN-based Cross-scale Feature Fusion(CCFF) structure, exploiting a local-global attention mechanism to improve the detection capability for minute defects. Finally, an Adaptive Pointwise Kernel Convolution (APKConv) is introduced, which achieves adaptive deformation of the receptive field via learnable offsets to accurately capture defects of diverse morphologies. Experimental results demonstrate that OL-HAIM attains a mean Average Precision at IoU threshold of 0.5(mAP@0.5) of 97.6% on the PCB-DEFECT dataset, representing a 4.8 percentage-point improvement over the baseline, with only 13.8M parameters and Floating Point Operations Per Second(FLOPS) reduced to 23.6G, while achieving an inference speed of 55.7 frames/s(FPS) on the RK3588 embedded platform. Cross-domain experiments on the Northeastern University Surface Defect Detection Dataset(NEU-DET) and Microsoft Common Objects in Context(MS COCO) datasets further verify the generalization capability of the proposed algorithm. On field-collected samples from an actual PCB production line, the model achieves an mAP@0.5 of 95.1% and an inference speed of 55.4 FPS on the RK3588 embedded platform, concurrently meeting the requirements of high-accuracy detection and real-time performance in practical scenarios. The proposed method provides an effective solution that reconciles high accuracy with lightweight deployability for real-time PCB defect inspection in industrial settings.

Key words: PCB defect detection; DEIM; lightweight object detection; edge deployment

1 引言

印刷电路板(Printed Circuit Board, PCB)作为现代电子设备的核心载体,其制造质量直接决定终端产品的性能与可靠性^[1]。近年来,三维集成光子封装^[2]与拓扑光子超材料^[3]等先进光学技术的持续突破,为工业精密视觉检测提供了更高灵敏度的感知手段,进一步凸显了对电子制造质量实施精密光学检测的迫切需求。随着电子产品向小型化和高集成度方向发展,PCB元器件布局日益密集,但在其生产过程中极易产生缺孔、短路、断路、伪铜等微小缺陷^[4],一旦漏检将引发电路功能失效乃至重大安全事故。因此,发展高效、准确的PCB缺陷自动检测方法具有重要的工程价值。

传统PCB缺陷检测依赖人工目视或模板匹配^[5]、边缘检测^[6]等图像处理方法,此类方法效率低且对光照与噪声敏感,而机器学习方法虽具备一定自适应能力,但受限于人工特征,其在复杂纹理与小尺寸缺陷场景下表现不佳^[7]。近年来,以

YOLO(You Only Look Once)为代表的单阶段检测器凭借实时性成为主流^[8],相关改进围绕特征增强^[9]、主干轻量化^[10]和细节特征优化^[11]展开,但YOLO系列依赖NMS(Non-Maximum Suppression)后处理,对超参数敏感且在小目标密集的情况下易漏检^[12],难以在精度、速度与资源消耗间取得平衡。为此,Carion等^[13]提出的DETR(Detection Transformer)摒弃锚框与NMS,实现了端到端检测,但该类算法存在收敛慢、计算开销大等问题。RT-DETR^[14]通过高效混合编码器提升了推理速度,并在PCB检测中衍生出多尺度增强^[15]与主干轻量化^[16]等改进,但训练效率瓶颈仍未解决。Huang等^[17]提出的DEIM(DETR with Improved Matching)通过Dense O2O(One to One)匹配策略将训练时间缩短50%,为端到端检测器的高效训练提供了新路径。

除上述基于YOLO和DETR架构的改进工作外,近年来国内外学者还从多元化技术路线探索了PCB缺陷检测的新方案。在国际研究方

面,Zhao等^[18]提出PSDDNet(PCB Surface Defect Detection Network)模型,通过多分支卷积和聚集-分发特征融合架构实现了极低参数量下的高精度检测,但该模型计算量偏高,在密集缺陷场景下鲁棒性不足。为了改善这种缺陷,Gao等^[19]提出DECNet(DMA-EADS-CDPN Network),集成可变形多头注意力和聚集扩散金字塔网络优化多尺度特征融合,有效降低了参数量,但模型的推理速度下降,边缘端部署受限。基于此,Zhu等^[20]提出了SDD-Net(Subtle Defect Detection Network),结合感受野增强骨干网络与全维动态卷积预测头,在精度与速度间取得了较好平衡,但参数量和模型体积偏高,轻量化程度有限。在国内研究方面,朱黎颖等^[21]针对复杂背景下PCB表面焊接缺陷的视觉检测问题开展了系统研究,在抑制复杂电路纹理背景干扰方面取得了有效进展,但对模型轻量化与边缘端部署的关注仍有不足。欧宇枫等^[22]提出了DES-Det(DualBranch Feature Extraction-Spatial Attention Detection)模型,该方法采用双分支特征提取结构与可变形注意力增强小目标判别能力,在降低参数量的同时提升了准确率,但整体均值精度偏低,高精度场景适应性不足。魏嘉敏等^[23]提出YOLO-HorL,利用递归门控卷积高阶交互模块和幻影卷积实现了模型体积的显著压缩,检测精度较基线模型有明显提升,但复杂背景下的泛化能力有待验证。综上,现有方法在轻量化设计、边缘端实时推理与高精度检测三者之间尚未实现有效统一,亟需一种能够同时兼顾并具有嵌入式部署能力的新方案。

近年来,计算机视觉领域在轻量化主干网络、自适应卷积和多尺度特征融合等方面取得重要进展,为解决上述问题提供了新的技术路径。Lou等^[24]提出的OverLoCK架构通过深层级分解策略和上下文混合动态卷积实现自上而下的注意力机制,在ImageNet分类和COCO检测任务上显著优于同等规模的主流主干网络。Zhang等^[25]提出的AKConv(Adaptive Kernel Convolution)突破传统方形卷积核限制,通过任意数量采样点和可学习偏移量实现自适应感受野变形。此外,基于注意力的特征融合策略也已在多个视觉任务中展现优势^[26]。然而,上述技术主要面向通用视觉任务设计,尚未在工业缺陷检测领域得到充分验证,

缺乏对密集小目标与复杂纹理背景等特定场景的针对性适配。

综合上述分析,当前在PCB缺陷检测领域,无论是基于YOLO的单阶段方法、基于DETR的端到端方法,还是其他多元化技术路线,这些方法仍面临3方面核心挑战:首先,主干网络未能充分平衡计算效率与检测精度,难以适应边缘设备的资源约束;其次,特征融合策略对复杂电路纹理背景的干扰抑制能力有限,影响小目标缺陷的定位精度;最后,传统融合方式未能充分利用不同尺度特征间的互补信息,制约多尺度检测性能的进一步提升。针对上述问题,本文提出一种基于改进DEIM的PCB缺陷检测算法OL-HAIM,通过对现有先进技术进行面向任务的系统性集成与针对性适配,为工业现场PCB缺陷实时检测提供兼顾高精度与轻量化部署的有效方案。具体工作如下:

(1)采用先概览后细看卷积网络(Overview-first-Look-Closely-next, OverLoCK)替换原主干网络,通过引入更加轻量化的架构在提升小目标检测准确率的同时显著降低推理时间;

(2)将基于CNN的跨尺度特征融合模块(CNN-based Cross-scale Feature Fusion, CCFF)中拼接融合(Concat)机制替换为分层注意力融合模块(Hierarchical Attention Fusion Block, HAFB),通过层次化自适应特征混合强化不同尺度特征语义一致性;

(3)在CCFF结构中嵌入自适应逐点卷积(Adaptive Pointwise Kernel Convolution, APKConv)模块,利用自适应金字塔卷积核增强复杂背景判别能力。

实验结果表明,改进后的算法在PCB缺陷数据集上取得了优于现有先进方法的检测性能。

2 OL-HAIM 网络模型

DEIM作为新一代端到端检测器,在实时性和检测精度方面均优于传统YOLO系列,其采用主干网络、混合编码器以及Transformer解码器的三段式架构,通过消除NMS后处理实现稳定推理。然而该算法在PCB缺陷检测中仍面临模型冗余、计算开销大、边缘部署困难等挑战。针对上述问题,本文提出OL-HAIM检测算法,通过系统性架构优化实现精度与效率平衡。核心改进包

括:(1)将主干网络替换为 OverLoCK,通过深层分解和上下文混合动态卷积降低计算复杂度,同时增强复杂电路纹理背景的特征表达;(2)引入 HAFB 模块替代 CCFF 结构中的 Concat 模块,通过局部-全局注意力机制和多路径融合策略提

升微小缺陷检测能力;(3)采用 APKConv 替代 CCFF 结构中的卷积(Conv)模块,通过自适应采样和逐点卷积高效聚合优化不同 PCB 形态缺陷的检测性能。改进后的 OL-HAIM 网络架构如图1所示。

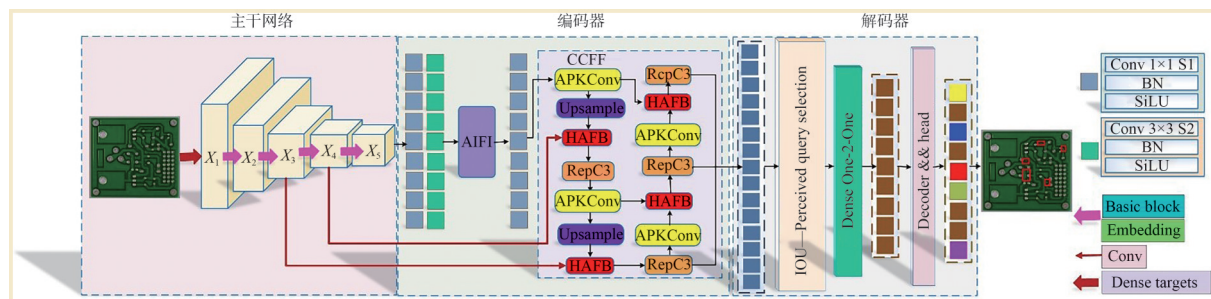


图1 OL-HAIM网络结构

Fig. 1 OL-HAIM network architecture

2.1 OverLoCK

DEIM作为基于Transformer的检测器虽在通用目标检测中表现出色,但在PCB缺陷检测中面临缺乏针对密集走线的自适应机制、计算开销大以及特征提取缺少语义引导等局限,这使得其难以在边缘设备上实现实时检测。针对上述问题,本文采用OverLoCK架构改进DEIM主干网络。OverLoCK的架构特性与PCB缺陷检测需求高度契合。PCB板上缺陷尺度差异显著,从仅占数个像素的微小缺孔到横跨数百像素的延展性划痕,要求主干网络同时具备细粒度局部感知和宏观语义理解能力。OverLoCK通过深层级分解策略(Deep-level Decomposition Strategy, DDS)和

上下文混合动态卷积(Context-Mixing Dynamic Convolution, ContMix)两种核心机制实现高效检测,其结构如图2所示。DDS策略将网络分解为Base-Net、Overview-Net和Focus-Net三个功能互补的子网络,形成自上而下的特征提取范式。Base-Net负责提取PCB图像的中层通用特征,为后续处理提供基础表征;随后Overview-Net捕获电路板的整体布局信息作为语义先验;最后Focus-Net在该先验指导下有针对性地细化缺陷区域特征,从而避免传统自下而上范式在复杂走线背景中将正常纹理误判为缺陷。

OverLoCK中的ContMix结构(图3)有利于充分发挥自上而下注意力,其通过动态生成卷积

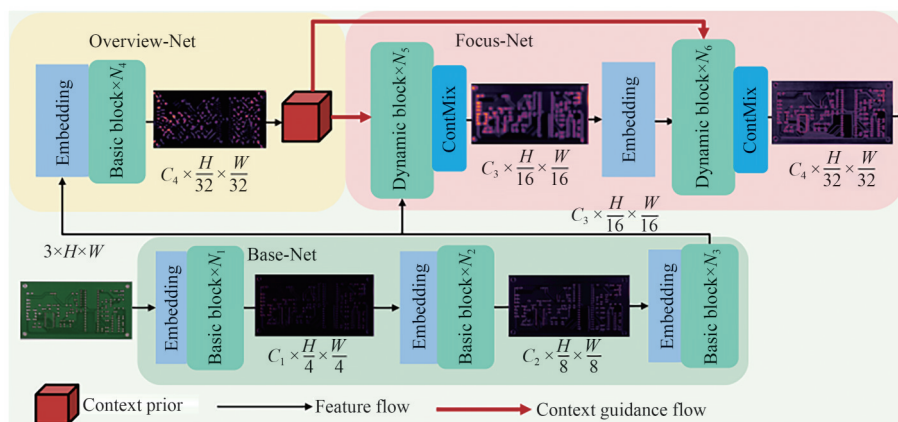


图2 OverLoCK整体架构

Fig. 2 Overall architecture of OverLoCK

核建模长距离依赖,同时以混合核设计兼顾全局上下文与局部细节。该机制以纯卷积方式实现,相比 Transformer 类全局注意力机制更适合 RK3588 等 NPU 架构的高效推理部署。此外,上下文更新机制在每个动态块中保留初始上下文

防止信息稀释,确保输出特征同时包含缺陷表示和电路布局语义。这些协同设计使 OverLoCK 在扩大有效感受野的同时保持局部特征的精细表达,有效适应 PCB 板上不同尺度缺陷的形态变化。

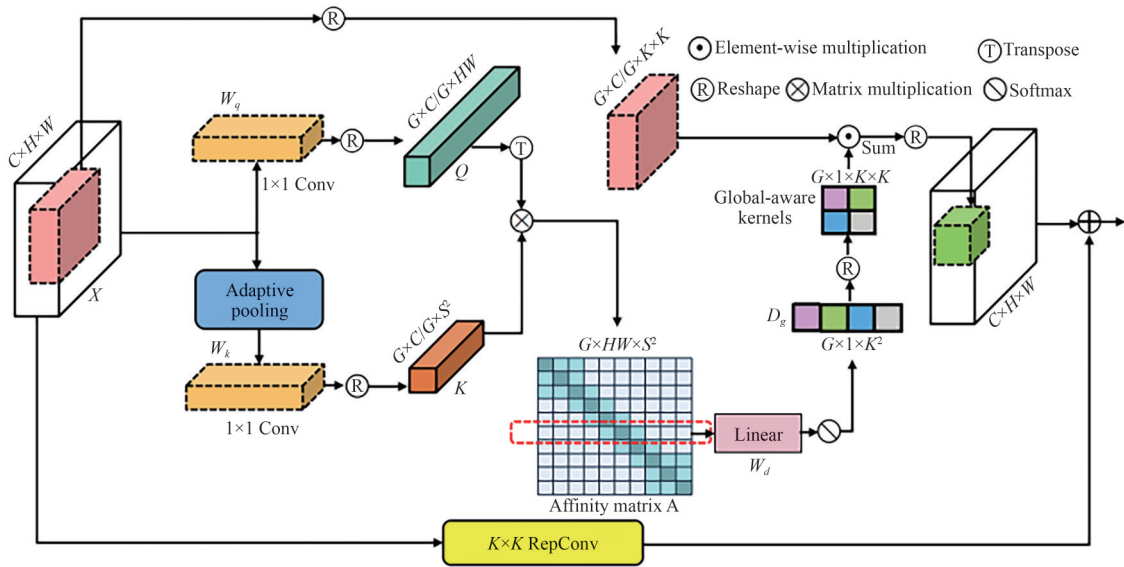


图 3 ContMix 结构

Fig. 3 ContMix architecture

2.2 HAFB

PCB 缺陷检测面临光照不均所致特征模糊、批次工艺差异引发尺度剧变以及密集电路布局导致目标重叠等多重挑战,且缺陷形态包括微小缺陷、细长划痕与不规则伪铜等各异类型,对模型的多尺度特征提取能力提出极高要求。然而 DEIM 中的 CCFF 结构在处理小尺寸缺陷时难以有效拓展特征空间与捕捉细粒度信息,其多次卷积操作同时加重了计算负担。针对上述问题,本文设计 HAFB^[27] 模块替代 CCFF 结构中的 Concat 模块,其整体结构如图 4 所示。HAFB 采用双路径多尺度注意力机制,集成 4 个局部-全局注意力 (Local Global-Attention, LGA) 模块,并配置不同感受野参数,通过两条注意力增强路径和一条基础融合路径实现多层次特征提取。注意力增强路径结合分组卷积强化局部结构差异感知,能够有效区分密集电路背景中的缺陷目标与正常纹理。全局注意力分支捕获更大范围的电路布局上下文依赖。3 路特征经融合机制整合后使模型能从可见区域推断被遮挡的缺陷结构。

HAFB 的核心机制 LGA 旨在同时捕获细粒度局部模式和上下文全局信息。LGA 将输入特征图划分为非重叠的块,对每个块通过平均池化聚合局部特征,随后计算特征与可学习提示向量的余弦相似度生成注意力掩码,再经可学习变换矩阵进行特征映射,最后通过卷积精炼输出。LGA 通过使用不同大小的块处理输入特征图,能够同时捕获不同粒度的特征信息,有效处理 PCB 板上不同尺寸的缺陷目标。针对大小为 P 的块,特征聚合可以表示为:

$$F^P = \{f_i^P | i \in [1, (H/P) \times (W/P)]\}, \quad (1)$$

其中: F^P 表示块大小为 P 的特征集合; f_i^P 表示第 i 个块的特征表示; P 表示块的大小,取值为 2 或 4; H 和 W 分别表示特征图的高度和宽度; i 表示块的索引。通过组合不同尺度下的特征表示, LGA 能够有效处理 PCB 板上不同尺寸的缺陷目标。LGA 的核心优势在于其注意力分配机制,通过余弦相似度计算特征与可学习提示向量之间的关联度,其数学过程如下:

$$\cos_{\text{sim}}(\mathbf{f}, \mathbf{p}) = (\mathbf{f} \cdot \mathbf{p}) / (\|\mathbf{f}\| \cdot \|\mathbf{p}\|) \in [-1, 1], \quad (2)$$

其中: $\mathbf{f}$ 是特征向量, $\mathbf{p}$ 是可学习的提示向量, $\|\cdot\|$ 表示向量的模(L2范数)。这种相似度计算方式能够有效识别PCB缺陷的重要特征,并通过掩码操作突出关键信息:

$$\text{Attention}(\mathbf{f}) = \mathbf{f} \odot \text{clamp}(\text{Similarity}(\mathbf{f}, \mathbf{p}), 0, 1), \quad (3)$$

其中: $\odot$ 表示Hadamard积;clamp表示截断函数,将值限制在 $[0, 1]$ 区间。在自适应特征变换中,LGA模块采用可学习的变换矩阵 T 实现特征空

间的自适应映射,增强PCB缺陷特征的判别能力;在局部-全局信息融合过程中,LGA模块通过块内信息聚合和块间关系建模,实现了局部缺陷细节和全局电路布局的有效融合,特别适合处理PCB板上密集走线背景中的复杂缺陷场景。在模块协同方面,HAFB接收OverLoCK主干网络输出的多尺度特征图,通过双路径注意力机制融合不同层级的语义信息,为APKConv提供包含丰富上下文的特征表示,三者形成主干特征提取、多尺度特征融合、自适应边界捕获的级联处理流程。

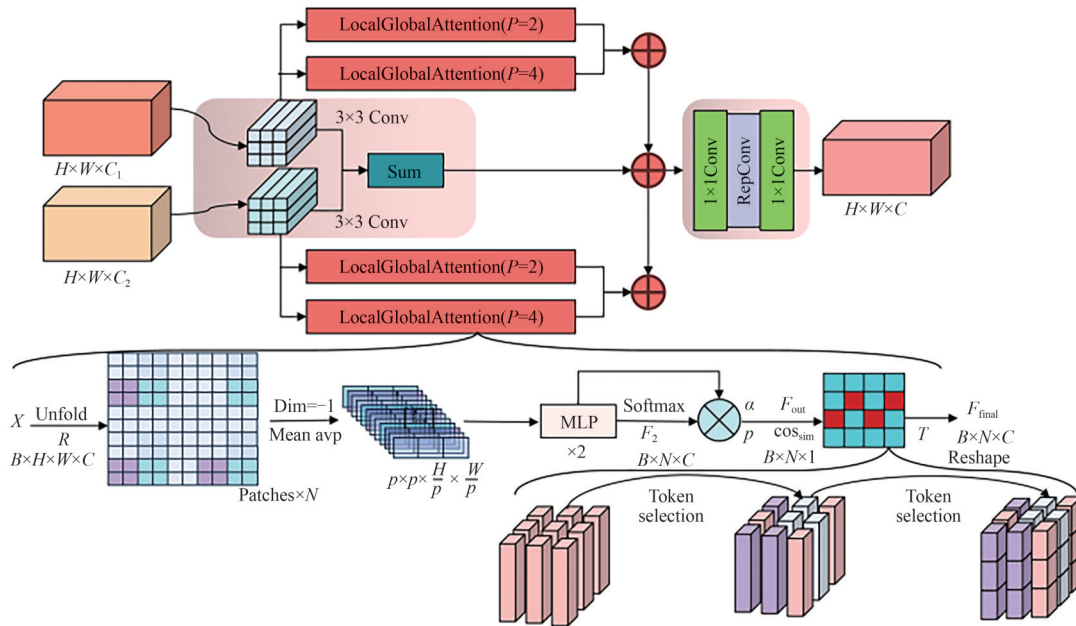


图4 HAFB模块结构

Fig. 4 HAFB module structure

2.3 APKConv

PCB缺陷呈现极大尺度差异性,其通常是从仅占数个像素的微小缺陷到横跨数百像素的延展性划痕,并且短路、断路等线性缺陷以及伪铜等不规则面状缺陷的边界均是传统方形卷积核难以精确捕获的。DEIM的卷积参数数量随核尺寸呈平方级增长,这严重制约了大尺度特征提取能力。针对上述问题,本文引入APKConv模块替代CCFF结构中的标准卷积。APKConv的自适应采样特性与PCB缺陷几何异质性高度匹配。PCB生产过程中产生的缺陷形态差异显著,从仅占数个像素的微小缺陷到横跨数百像素的延展性划痕,从沿走线方向延伸的短路、断路到形状不规

则的伪铜氧化,不同类型缺陷对采样策略的需求截然不同。APKConv突破了传统方形卷积核的限制,通过可学习偏移量机制实现感受野的自适应变形,能够根据缺陷几何特征动态调整采样点位置,其详细结构如图5所示。

为了适配PCB缺陷检测场景,本文对原始AKConv进行了两项关键改进。(1)引入逐点卷积(Pointwise Convolution, PWConv)^[28]实现跨通道特征融合与维度变换,通过 1×1 卷积融合多尺度空间特征生成更具判别力的缺陷表征。(2)采用任意参数数量的卷积核设计使参数增长从平方级降至线性级。以等效 7×7 感受野为例,传统卷积需要49倍通道参数量而APKConv仅需14

倍,参数量减少 71.4%。APKConv 针对不同形态的 PCB 缺陷采用差异化采样模式。对于划痕类线性缺陷,采样点沿延伸方向拉伸形成细长感受野以完整捕获缺陷轮廓。对于短路、断路等沿走

线分布的缺陷,采样点自适应调整为与走线方向一致分布模式。对于伪铜氧化等面状缺陷,采样点呈均匀区域分布以覆盖完整边界。对于微小缺陷,采样点集中于缺陷中心以增强局部特征响应。

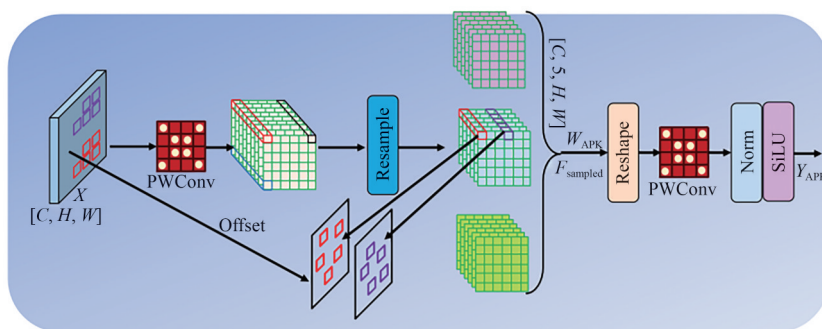


图 5 APKConv 结构

Fig. 5 APKConv structure

在整个模型结构中,OverLoCK 提供包含全局语义的多尺度特征图,HAFB 融合不同层级的特征表示,APKConv 在此基础上精准捕获各类缺陷的几何边界,三者协同实现从全局语义理解到局部边界定位的完整检测流程。

3 实验与分析

3.1 实验环境及数据集

本研究的实验平台构建于 Windows 10 操作系统,并采用 Intel Core i9-14900KF 处理器和 NVIDIA GeForce RTX 4090 图形处理器作为硬件基础,软件环境配置为 Python 3.8、PyTorch 2.2.0 和 CUDA (Compute Unified Device Architecture)

11.8 并行计算架构。为全面验证 OL-HAIM 模型的检测性能和泛化能力,本实验采用两个具有代表性的工业缺陷检测数据集,分别如图 6(a)和图 6(b)所示。图 6(a)所示的 PCB-DEFECT 印刷电路板缺陷数据集作为主要实验数据集,其源自北京大学智能机器人开放实验室,包含 1 386 张高分辨率图像,覆盖缺陷、鼠咬、开路、短路、毛刺和杂铜 6 种关键制造缺陷。本文将图 6(b)所示的 NEU-DET (Northeastern University Defect Detection) 钢材表面缺陷数据集作为泛化性验证数据集,其包含龟裂、斑块、夹杂物、麻面、轧入氧化皮和划痕 6 类缺陷共 1 800 张原始图像,该类工业表面缺陷的视觉检测研究已有充分积累^[29]。NEU-DET 数据集与 PCB 缺陷在成像特征上存在显著差异,

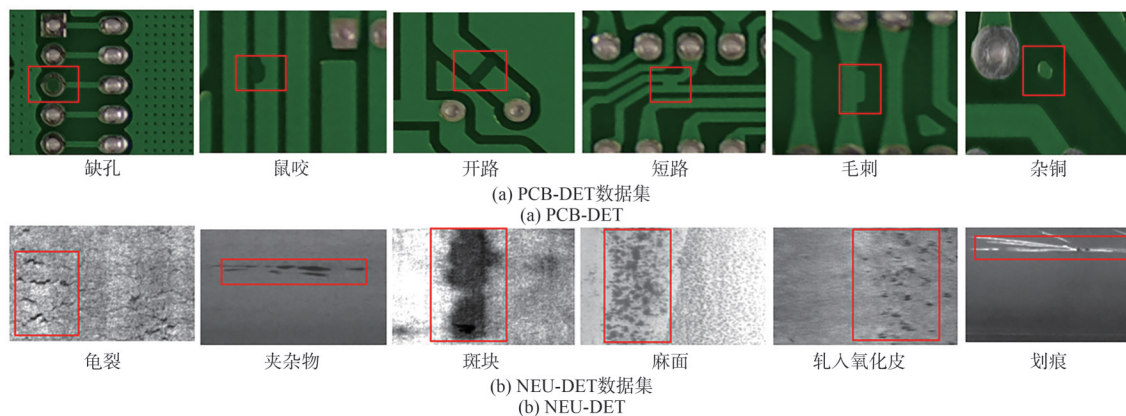


图 6 各数据集缺陷图片

Fig. 6 Defect images from each dataset

钢材表面呈现金属光泽和轧制纹理,而PCB板具有复杂电路走线背景,两者的缺陷形态、纹理特征和背景干扰模式截然不同,能够有效检验算法在跨场景下的适应能力。针对工业缺陷检测领域样本稀缺的挑战,本研究实施了系统化数据增强策略,包括几何变换增强、光度变换增强、工业场景模拟以及缺陷特征增强,最终将PCB-DEFECT和NEU-DET分别扩充至5544张和7200张图像,均按8:1:1划分训练集、验证集和测试集。训练配置采用AdamW(Adam with Decoupled Weight Decay)优化器,初始学习率0.0001配合余弦退火调度策略,训练批次大小为16,权重衰减为0.05,训练175轮,输入图像尺寸统一为640像素×640像素。

3.2 RK3588平台部署

本研究采用瑞芯微RK3588异构计算平台实现深度学习模型的边缘端部署。该平台搭载专用神经网络处理单元(Neural Processing Unit, NPU),具备6 TOPS(Tera Operations Per Second)的AI(Artificial Intelligence)算力性能,并支持INT4(4-bit Integer)/INT8/INT16多精度混合量化推理机制,为工业现场的实时缺陷检测提供了高效的硬件加速支撑。为充分发挥RK3588平台NPU的计算潜力,需将训练完成的深度学习模型转换为该平台原生的RKNN(Rockchip Neural Network)格式。该格式通过算子融合、内存池化和数据流优化等技术实现推理加速。本研究构建了标准化的模型迁移流程,首先利用PyTorch框架的原生导出接口将训练模型转换为开放神经网络交换(Open Neural Network Exchange, ONNX)中间表示格式,随后通过rknn-toolkit2转换工具链将ONNX模型映射至RKNN格式。该转换流程涵盖计算图优化、算子映射、量化校准和性能调优等关键环节。在模型转换阶段,工具链执行自动化的计算图分析,实施冗余节点消除、算子重排序和内存分配优化等操作,从而确保模型在目标硬件上的高效执行。为保障模型在嵌入式环境中的稳定运行,转换过程采用混合精度量化策略。通过对数据集激活值分布进行校准,实现权重和激活值的动态量化。同时,借助NPU的并行计算特性,在算子层面实现卷积、批归一化和激活函数的融合操作,显著降低内存访问开销和计算延迟,

实现从云端训练到边缘推理的高效模型迁移。图7展示了RK3588平台的完整部署流程。

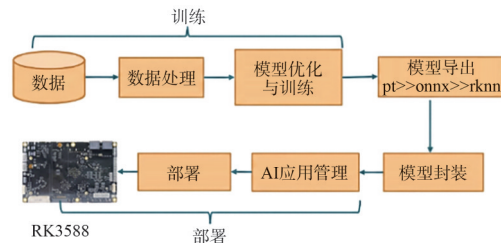


图7 RK3588平台部署流程

Fig. 7 RK3588 platform deployment process

3.3 评价指标

本文采用参数数量衡量模型大小(Size),通过浮点运算量(Floating Point Operations, FLOPs)衡量计算复杂度,同时利用平均精度均值(mean Average Precision, mAP)、精确率(Precision, P)和召回率(Recall, R)等指标对模型性能进行综合评估。各指标具体计算公式如下:

$$P = \frac{TP}{TP + FP}, \quad (4)$$

$$R = \frac{TP}{TP + FN}, \quad (5)$$

$$AP = \int_0^1 P(r) dr, \quad (6)$$

$$mAP = \frac{\sum_{i=1}^N AP_i}{N}, \quad (7)$$

在上述公式中:TP(True Positive)代表正确检测到的目标数量,FP(False Positive)代表错误检测到的目标数量,FN(False Negative)代表未检测到的目标数量, $P(r)$ 代表精确率-召回率曲线, N 代表类别数量。由于实验均涉及6个类别,因此本实验中 N 设为6。

3.4 消融实验

OL-HAIM算法在PCB数据集上的检测结果如表1所示。OL-HAIM算法在PCB数据集上将mAP@0.5从92.8%提升至97.6%,增长4.8个百分点,6类缺陷均实现显著改善。其中缺陷提升6.1个百分点,得益于OverLoCK增强了微小目标的特征表达。鼠咬缺陷提升5.3个百分点,归功于APKConv自适应贴合其不规则锯齿状边界。开路 and 短路缺陷分别提升4.5和4.6个百分点,源于HAFB融合了局部细节与全局电路布局信息。毛

表 1 PCB数据集上各缺陷检测精度 mAP@0.5 对比

Tab. 1 Comparison of mAP@0.5 for various defects on the PCB dataset

模型	缺陷检测精度/%						
	缺孔	鼠咬	开路	短路	毛刺	杂铜	总体
DEIM	89.7	91.6	92.9	93.5	94.2	95.1	92.8
OL-HAIM	95.8	96.9	97.4	98.1	98.3	98.7	97.6

刺和杂铜缺陷分别提升 4.1 和 3.6 个百分点,改进后均超过 98%。值得注意的是,基线模型中检测难度越大的类别改进幅度越显著,各类别精度离散程度从 5.4 个百分点缩小至 2.9 个百分点,表明 3 个模块协同弥补了原始 DEIM 在小尺寸和不规则形态缺陷检测上的不足,实现了更加均衡的多类别检测性能。

为验证算法的泛化能力,本文在 NEU-DET

钢材表面缺陷数据集上进行了相同的对比实验。如表 2 所示,OL-HAIM 在 NEU-DET 数据集上同样取得显著性能提升,整体 mAP@0.5 从 85.2% 提升至 89.7%,提升 4.5 个百分点。值得注意的是,对于形态差异较大的龟裂和轧入氧化皮缺陷,改进算法分别实现 2.1 和 5.2 个百分点的提升,证明了所设计模块在不同缺陷形态下的有效性和跨域泛化能力。

表 2 NEU-DET数据集上各缺陷检测精度 mAP@0.5 对比

Tab. 2 Comparison of mAP@0.5 for various defects on the NEU-DET dataset

模型	缺陷检测精度/%						
	裂纹	斑点	夹杂物	麻点表面	轧入氧化皮	划痕	总体
DEIM	82.5	84.2	84.8	86.1	86.4	87.3	85.2
OL-HAIM	84.9	88.8	89.0	91.1	92.0	92.4	89.7

为验证 OverLoCK、HAFB 和 APKConv 三个模块对检测性能的具体贡献,本文在 PCB 数据集上设计了 8 种不同模块组合的消融实验。如表 3 所示,单独引入各模块均实现不同程度的性能提升:OverLoCK 模块使 mAP@0.5 提升 1.4 个百分点,同时 FLOPs 从 61.1G 大幅降至 31.8G;HAFB 模块使 mAP@0.5 提升 1.7 个百分点,召回率提升 1.8 个百分点;APKConv 模块使 mAP@0.5 提升 1.3 个百分点,对不规则形态缺陷的检测能

力明显增强。当 3 个模块全部集成时,mAP@0.5 达到 97.6%,模型大小仅为 13.8M,FLOPs 降至 23.6G,PC 端 FPS 达到 102.3,RK3588 嵌入式平台 FPS 达到 55.7,实现了精度与效率的双重优化。

为进一步验证各模块的泛化能力,在 NEU-DET 数据集上进行了相同的消融实验。如表 4 所示,3 个模块在钢材表面缺陷检测任务上同样展现出良好的性能提升效果,完整模型在 NEU-

表 3 PCB数据集上的消融实验对比

Tab. 3 Comparison of ablation experiments on the PCB dataset

模型	OverLoCK	HAFB	APKConv	召回率/ %	精度/ %	mAP@ 0.5/%	mAP@0.5: 0.95/%	mAP@ 0.75/%	模型 尺寸/M	计算量/ G	PC 端 速度/ FPS	RK3588 速度/ FPS
1	×	×	×	86.9	95.7	92.8	77.6	63.8	25.2	61.1	64.3	30.8
2	√	×	×	88.2	96.4	94.2	78.5	64.7	18.7	31.8	81.6	42.9
3	×	√	×	88.7	96.8	94.5	78.9	65.1	23.1	47.9	72.5	35.8
4	×	×	√	88.0	96.2	94.1	78.4	64.5	24.5	51.4	69.1	33.7
5	√	√	×	89.5	97.6	95.7	79.6	65.8	17.4	27.9	89.4	48.5
6	√	×	√	89.1	97.3	95.3	79.3	65.4	18.0	29.5	85.3	46.1
7	×	√	√	89.8	97.9	96.0	79.8	66.1	21.7	42.5	77.8	40.4
8	√	√	√	90.8	99.1	97.6	80.7	67.2	13.8	23.6	102.3	55.7

表 4 NEU-DET 数据集上的消融实验对比

Tab. 4 Comparison of ablation experiments on the NEU-DET dataset

模型	OverLoCK	HAFB	APKConv	召回率/ %	精度/ %	mAP@ 0.5/%	mAP@0.5: 0.95/%	mAP@ 0.75/%	模型 尺寸/M	计算量/ G	PC 端 速度/ FPS	RK3588 速度/ FPS
1	×	×	×	79.8	90.3	85.2	72.6	61.4	25.2	61.1	67.4	32.7
2	✓	×	×	81.1	91.0	86.5	73.7	62.4	18.7	31.8	84.1	44.6
3	×	✓	×	81.4	91.3	86.8	74.0	62.7	23.1	47.9	75.2	37.9
4	×	×	✓	80.9	90.8	86.3	73.5	62.2	24.5	51.4	71.6	35.5
5	✓	✓	×	82.4	92.1	87.9	74.9	63.8	17.4	27.9	91.5	50.8
6	✓	×	✓	82.0	91.7	87.5	74.5	63.4	18.0	29.5	87.9	48.1
7	×	✓	✓	82.6	92.3	88.1	75.1	64.0	21.7	42.5	80.3	42.2
8	✓	✓	✓	83.8	93.1	89.7	76.1	65.2	13.8	23.6	106.7	58.3

DET 上实现 89.7% 的 mAP@0.5, 验证了所设计模块在跨域场景下的有效性。实验结果表明, OverLoCK 模块通过深层级分解策略有效降低了计算复杂度, 同时保持了较高的检测精度; HAFB 模块通过分层注意力机制显著提升了对小尺寸缺陷的召回率; APKConv 模块通过自适应采样策略增强了对不规则形态缺陷的检测能力。3 个模块在 PCB 和 NEU-DET 两个差异显著的数据集上均取得一致的性能提升, 验证了所设计模块的普适性和泛化能力。

3.5 算法对比实验

为验证 OverLoCK 作为主干网络的优越性, 本文在 PCB 数据集上设计了不同主干网络的对比实验。如表 5 所示, OverLoCK 实现了 94.2% 的 mAP@0.5 和 64.7% 的 mAP@0.75, 相比基线模型分别提升 1.4 和 0.9 个百分点。同时 FLOPs 从 61.1G 大幅降至 31.8G, 降幅达 48.0%。相比其他轻量级网络, OverLoCK 在保持较低计算复

杂度的同时实现了最高的检测精度, 验证了深层级分解策略和上下文混合动态卷积的有效性。

为验证 APKConv 卷积结构的优越性, 本文在 PCB 数据集上将 CCFF 结构中的标准卷积替换为多种主流卷积变体进行对比实验。如表 6 所示, DSConv 和 GhostConv 虽有效降低了计算量, 但通道间信息交互不充分导致 mAP@0.5 分别下降 1.1 和 1.4 个百分点。PConv 仅对部分通道施加空间卷积, mAP@0.75 降至 62.7%, 表明选择性计算在需要全通道空间感知的缺陷定位任务中存在局限。DCNv3 通过可变形卷积将 mAP@0.5 提升至 93.5%, 但额外的偏移学习网络使 FLOPs 达 54.2G, 效率提升有限。APKConv 以 51.4G 的 FLOPs 实现了 94.1% 的 mAP@0.5 和 64.5% 的 mAP@0.75, 在所有对比方法中取得最优精度-效率平衡, 验证了线性级参数增长策略和自适应采样机制在 PCB 缺陷检测中的有效性。

为全面评估 OL-HAIM 算法在 PCB 缺陷检

表 5 不同主干网络性能对比

Tab. 5 Comparison of performance across different backbone networks

主干网络	mAP@ 0.5/%	mAP@ 0.75/%	计算量/ G	PC 端 速度/FPS
Baseline	92.8	63.8	61.1	64.3
LSNet	91.5	62.5	38.2	85.1
FasterNet	91.1	62.1	28.9	91.7
MobileNetV4	90.8	62.3	37.9	84.6
EfficientFormerV2	92.1	63.1	44.5	78.3
OverLoCK	94.2	64.7	31.8	81.6

表 6 不同卷积结构性能对比

Tab. 6 Performance comparison of various convolution architectures

卷积结构	mAP@ 0.5/%	mAP@ 0.75/%	计算量/ G	PC 端 速度/FPS
Conv	92.8	63.8	61.1	64.3
DSConv	91.7	62.9	39.1	82.4
GhostConv	91.4	62.5	35.6	86.7
PConv	92.0	62.7	36.8	85.2
DCNv3	93.5	64.1	54.2	66.8
APKConv	94.1	64.5	51.4	69.1

测任务上的综合性能,本文与现有主流检测算法进行了对比实验。所有对比实验均在3.1节所述的不同条件下进行。对于YOLOv5m至YOLOv13s、RT-DETR-R18、D-FINE-L^[30]和DEIM等开源算法,本文直接使用其官方代码库在上述配置下从初始化训练至收敛;对于相关文献改进算法,本文根据文献描述的网络结构在对应基线框架上复现,并在相同条件下重新训练以确保公平性。如表7所示,与YOLO、DETR等主流开源系列算法相比,OL-HAIM在保持13.8M极低参数量和23.6G计算量的同时,实现了跨级别的精度突破,其99.1%的精度和97.6%的mAP@0.5远超最新YOLOv13s等同级轻量化模型。此外,该算法在推理效率上优势显著,其在PC端高达102.3FPS,

在边缘设备RK3588上也达到了55.7FPS,运行速度均大幅领先于绝大多数主流开源模型,真正兼顾了极致的检测性能与高效的部署能力。在基于YOLO的改进算法中,文献[9]通过HorNet模块和改进MCBAM注意力机制使得mAP@0.5达到93.4%,但参数量达26.3M且FLOPs高达62.1G。在基于DETR的改进算法中,文献[15]引入级联注意力机制使mAP@0.5达94.5%,但模型增至23.8M;文献[16]通过使用GhostNet替换原主干使得推理速度达到82.3FPS,FLOPs仅为38.5G,但mAP@0.5仅达到92.6%,在所有对比算法中较低;OL-HAIM相比于精度相当的D-FINE-L、OL-HAIM的参数量和计算量分别降低49.1%和74.1%,边缘端推理速度提升181.3%。

表 7 PCB数据集上不同算法性能对比

Tab. 7 Performance comparison of different algorithms on the PCB dataset

算法	召回率/%	精度/%	mAP@ 0.5/%	mAP@ 0.75/%	模型尺寸/M	计算量/G	PC端速度/ FPS	RK3588速度/ FPS
YOLOv5m	81.9	90.8	88.2	56.1	21.7	64.4	70.5	28.1
YOLOv8s	83.7	92.6	90.1	57.6	21.8	28.4	89.2	31.8
YOLOv9s	84.6	93.3	90.8	58.3	24.1	27.4	87.1	29.2
YOLOv10s	85.8	94.1	91.4	58.9	21.4	24.8	92.0	30.9
YOLOv11s	86.5	94.8	91.7	57.2	20.8	22.3	94.5	32.3
YOLOv13s	87.3	95.4	92.1	57.8	19.6	21.5	97.2	34.8
RT-DETR-R18	85.1	93.8	91.1	59.6	20.0	60.0	71.1	27.5
D-FINE-L ^[30]	91.2	98.9	97.8	68.5	27.1	91.0	51.4	19.8
文献[9]	85.2	94.6	93.4	60.7	26.3	62.1	65.8	25.4
文献[15]	88.3	96.7	94.5	63.9	23.8	55.7	73.4	29.6
文献[16]	84.7	93.5	92.6	60.1	15.2	38.5	82.3	38.4
文献[18]	88.1	97.2	95.0	63.4	20.9	55.0	65.4	26.3
文献[19]	87.9	96.5	94.3	61.0	7.6	8.3	81.3	33.6
文献[20]	87.8	96.1	93.8	60.8	26.43	105.3	53.0	21.2
文献[22]	86.3	94.9	92.7	60.5	4.2	5.3	109.7	57.1
文献[23]	87.1	95.1	92.6	60.8	10.2	15.2	69.7	30.2
OL-HAIM	90.8	99.1	97.6	67.2	13.8	23.6	102.3	55.7

在与国内外先进算法的进一步对比中,文献[18]虽以极低参数量将mAP@0.5提升至95.0%,但计算量高达55.0G,在RK3588平台上的推理速度仅达26.3FPS,反映出高计算量与边缘端实时性之间的突出矛盾。文献[19]在模型尺寸和计算量上较为紧凑,但mAP@0.5仅为94.3%,表明单纯依赖注意力机制增强特征融合难以充分挖掘小目标的判别信息。文献[20]的mAP@

0.5达到93.8%,然而参数量高达26.43M、计算量为105.3G,在RK3588平台上的推理速度仅为21.2FPS,边缘端部署能力明显不足。文献[22]的模型尺寸和计算量在所有对比算法中最低,但mAP@0.5仅为92.7%,说明过度压缩模型容量会削弱对复杂缺陷特征的特征能力。文献[23]实现了一定程度的轻量化,但mAP@0.5仅为92.6%且PC端的推理速度仅69.7FPS,在精度

和速度上均存在不足。

综合以上分析,无论是基于 YOLO 的改进方法还是基于 DETR 的改进方法,现有算法普遍面临检测精度与模型轻量化难以兼顾的瓶颈,高精度模型往往伴随高计算开销和边缘端部署困难,而轻量化模型则不可避免地牺牲检测精度。相比之下,本文所提 OL-HAIM 通过协同优化骨干网络、特征融合和卷积结构,以 13.8M 参数量和 23.6G 计算量实现了 97.6% 的 mAP@0.5 以及 RK3588 平台 55.7 FPS 的推理速度,在检测精度、模型轻量化和边缘端实时性 3 个维度上均全面优于上述国内外算法,有效突破了现有方法在精度与效率之间的权衡瓶颈。

为进一步验证 OL-HAIM 算法架构的普适性,本文在 MS COCO 通用目标检测数据集上进行了对比实验。如表 8 所示,OL-HAIM 实现了 66.5% 的 mAP@0.5 和 53.8% 的 mAP@0.75,模型大小为 16.8M,FLOPs 为 24.3G。相比 DEIM-RT-DETR-S,mAP@0.5 提升 0.4 个百分点,同时模型大小减少 16%,FLOPs 降低 59.5%。实验表明,OL-HAIM 不仅在工业缺陷检测任务上表现优异,在通用目标检测任务上同样具有良好的性能,证明了算法架构的普适性和泛化能力。

图 8 展示了 PCB 数据集上 DEIM 基线模型

表 8 MS COCO 数据集上不同模型检测性能对比

Tab. 8 Detection performance comparison of different models on the MS COCO dataset

模型	mAP@0.5/%	mAP@0.75/%	模型尺寸/M	计算量/G
RT-DETR-R18	63.8	50.4	20.1	60.1
RT-DETR-R34	66.8	52.9	31.1	92.1
DEIM-RT-DETR-S	66.1	53.3	20.0	60.0
DEIM-D-FINE-S	65.9	53.1	10.2	24.9
DEIM-D-FINE-M	70.0	57.3	19.1	57.1
DEIM-D-FINE-L	72.4	59.4	31.2	90.8
YOLOv11S	62.6	49.8	9.4	21.5
YOLOv13S	65.2	52.0	9.0	20.8
OL-HAIM	66.5	53.8	16.8	24.3

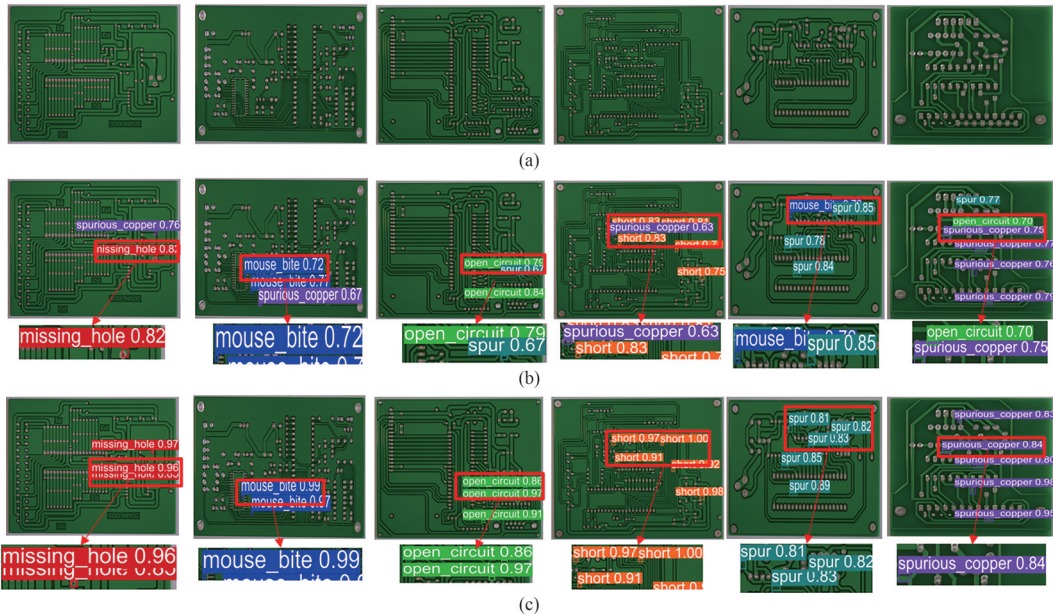


图 8 PCB 数据集上的测试结果。(a)原始数据集图像;(b)原模型测试结果;(c)改进后的模型测试效果。

Fig. 8 Test results on the PCB dataset. (a) Original dataset images; (b) Detection results of the baseline model; (c) Detection performance of the improved model.

与本文 OL-HAIM 算法的检测效果对比。其中图 8(a) 表示原始图像, 图 8(b) 表示基线模型的检测结果, 图 8(c) 表示 OL-HAIM 的检测效果, 从图 8 可以看出, 基线模型对各类缺陷的检测置信度普遍偏低, 且存在明显漏检和误检现象; OL-HAIM 则在 6 类缺陷上均实现了显著提升, 缺陷检测置信度从 0.78 提升至 0.96, 开路从 0.79 提升至 0.86, 短路从 0.83 提升至 0.91, 鼠咬从 0.72 提升至 0.99, 毛刺从稳定在 0.85 左右, 伪铜从 0.75 提升至 0.84。同时, OL-HAIM 的检测框定位更加精准, 漏检和误检现象明显减少, 验证了 OverLoCK 主干网络增强小目标特征表达、HAFB 模块融合多尺度信息以及 APKConv 模块自适应贴合不规则缺陷边界的协同增效

作用。

图 9 展示了 NEU-DET 钢材表面缺陷数据集上的检测效果对比, 用以验证 OL-HAIM 的跨域泛化能力。其中图 9(a) 表示原始图像, 图 9(b) 表示基线模型的检测结果, 图 9(c) 表示 OL-HAIM 的检测效果。相比基线模型, OL-HAIM 在 6 类缺陷上均实现了更高的置信度和更准确的定位。龟裂置信度从 0.80 提升至 0.96, 夹杂置信度从 0.82 提升至 0.95, 斑块置信度从 0.73 提升至 0.91 以上且密集目标识别更完整, 点蚀置信度从 0.76 提升至 0.97, 轧入氧化皮置信度从 0.76 提升至 0.86, 划痕置信度从 0.75 提升至 0.93。结果表明, OL-HAIM 在不同形态和尺度的工业缺陷检测中均具备良好的泛化性能。

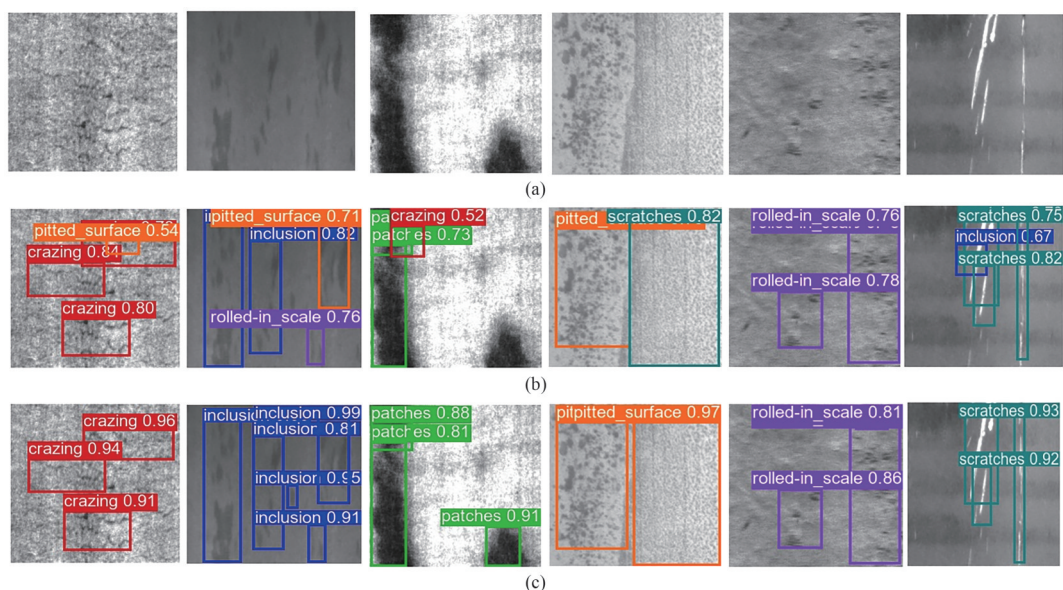


图 9 NEU-DET 数据集上测试结果。(a)原始数据集图像;(b)原模型测试结果;(c)改进后的模型测试效果。

Fig. 9 Test results on the NEU-DET dataset. (a) Original dataset images; (b) Detection results of the baseline model; (c) Detection performance of the improved model.

为验证 OL-HAIM 算法在实际生产环境中的适用性, 本研究与常州某电子制造企业合作, 通过从其 PCB 生产线的自动光学检测工位实地拍摄采集了真实缺陷样本。采集设备为工业面阵相机, 分辨率为 2 592 像素 $\times$ 1 944 像素, 配合环形 LED 光源和传送带同步触发系统, 在线采集了包含 4 层和 6 层 PCB 板的生产图像。经人工标注和质检专家复核, 共获得 428 张包含有效缺陷的图像, 覆盖缺陷、开路、短路、毛刺和杂铜 5 种缺陷类型。与公开数据集相比, 真实产线图像存在显著差异。首先, 光照

条件不均匀, 部分区域因 LED 反射角度差异出现局部过曝或阴影。其次, PCB 板面存在焊膏残留、丝印油墨等非缺陷干扰物。此外, 部分缺陷与正常焊盘、过孔的视觉特征高度相似, 增加了判别难度。

实验采用两种策略评估模型的真实场景适应能力。策略一为直接迁移测试, 将在 PCB-DEFECT 数据集上训练的模型直接应用于真实产线图像, 评估零样本迁移能力; 策略二为微调测试, 从 428 张真实图像中随机抽取 60% 用于微调, 仅训练 20 轮且学习率降至 0.000 02, 其余

40%用于测试。如表9所示,直接迁移测试中OL-HAIM的mAP@0.5为95.1%,相比公开数据集下降了2.5个百分点,召回率降至87.3%。性能下降主要源于真实产线环境中光照不均匀降低了缺陷与背景的对比度,焊膏残留和氧化斑点等干扰物与真实缺陷特征高度相似,以及多层板走线透印模糊了缺陷边界,使得检测难度显著增大。这表明成像光场条件是制约真实场景检测性能的重要因素,针对复杂多层结构的光场优化设计^[31]

是后续提升系统鲁棒性的重要方向。尽管如此,OL-HAIM的mAP@0.5仍大幅优于基线模型DEIM的89.2%,体现了较强的泛化能力。经少量真实数据微调后,mAP@0.5提升至97.8%,精度和召回率分别达到99.7%和91.1%,RK3588平台的推理速度保持在53.1 FPS,验证了该轻量化架构可通过少量领域适配有效部署于真实生产场景。需指出当前实验样本规模有限且仅覆盖单一产线,结论的普适性有待进一步验证。

表9 真实产线场景验证实验

Tab. 9 Verification experiments in a real production line environment

测试策略	mAP@0.5/%	mAP@0.75/%	精度/%	召回率/%	RK3588速度/FPS
DEIM直接迁移	89.2	60.5	91.3	81.4	30.1
OL-HAIM直接迁移	95.1	65.2	97.1	87.3	55.4
微调后	97.8	67.7	99.7	91.1	53.1

图10为所提模型在PCB真实生产线数据集上的检测结果,其中图10(a)表示原始图像,图10(b)表示基线模型的检测结果,图10(c)表示OL-

HAIM的检测效果。从图10可以看出,面对更复杂的真实生产线上的PCB缺陷,所提出的改进模型仍然实现了优秀的检测效果。

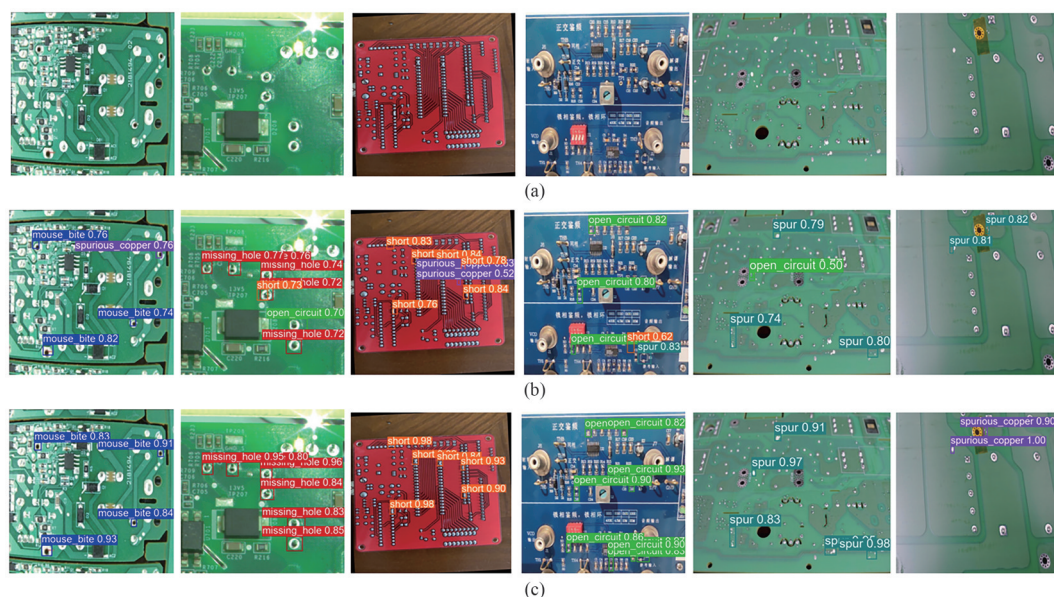


图10 真实产线数据集测试结果。(a)原始数据集图像;(b)原模型测试结果;(c)改进后的模型测试效果。

Fig. 10 Test results of real production line dataset. (a) Original dataset images; (b) Detection results of the baseline model; (c) Detection performance of the improved model.

4 结 论

印刷电路板缺陷检测面临目标尺寸微小、背

景纹理复杂、缺陷分布密集以及检测速度-精度难以平衡等多重挑战。针对上述问题,本文提出了基于改进DEIM的OL-HAIM检测算法:(1)采

用 OverLoCK 作为主干网络,通过深层级分解策略和上下文混合动态卷积建模长距离依赖关系,在扩大感受野的同时保持局部特征精细表达。(2)设计 HAFB,通过双路径多尺度注意力机制和 LGA 模块实现局部细节与全局上下文的有效融合。(3)引入 APKConv,通过可学习偏移量动态调整采样位置,使卷积核自适应贴合不规则缺陷边界。3项改进协同作用,最终模型 mAP@0.5 达到 97.6%,较基线模型提升 4.8 个百分点;参数量仅为 13.8M;FLOPs 为 23.6G;在 RK3588 嵌入式平台上的推理速度达到 55.7 FPS。与国内外主流方法相比,OL-HAIM 在精度与效率方面均取得最优平衡,在 NEU-DET 和 MS COCO 数据集上的跨域实验进一步验证了算法的泛化能力。

尽管 OL-HAIM 取得了较好的检测性能,但仍存在局限性。在算法鲁棒性方面,模型在极端光照和严重缺陷重叠场景下的稳定性有待提升,上下文混合机制在大面积高光区域可能出现特征响应衰减。在数据依赖性方面,模型依赖充足的高质量标注数据,在标注稀缺的新产品线或小批量定制化场景中的快速适应能力尚未充分验证。在硬件适配性方面,本文仅在 RK3588 单一平台上完成部署验证,不同算力和架构的边缘设备上的性能表现及移植成本尚未系统评估。在

真实场景适应性方面,产线实验表明,公开数据集与实际生产环境间存在显著域差异,少量微调虽可有效恢复性能,但当前实验仅涉及单一产线和有限工艺条件,不同厂商、板型及工艺参数下的泛化能力仍需进一步验证。

基于上述局限性分析,未来研究将从以下方向展开。第一,探索元学习、对比学习等小样本学习策略,结合基于风格迁移的跨产品线域适应方法,降低算法对大规模标注数据的依赖,提升模型在新产品线上的快速适应能力。第二,研究结构化剪枝、知识蒸馏、混合精度量化等模型压缩技术,开展针对不同边缘计算平台的算子优化和部署适配工作。第三,将 OL-HAIM 的核心模块抽象为可复用的功能组件,扩展至钢材表面、纺织品、玻璃面板等更多类型的工业缺陷检测任务,推动研究成果向通用化方向发展。第四,深化与电子制造企业的合作,建立大规模真实缺陷数据库,利用域适应和迁移学习技术缩小实验室环境与真实产线之间的性能差距,推动算法向实际部署落地。

综上所述,本文提出的 OL-HAIM 算法为轻量化目标检测器设计提供了新思路,为工业现场 PCB 缺陷实时检测提供了兼顾高精度与轻量化部署的有效方案。

参 考 文 献:

- [1] 杨彦萍,高军伟,刘兆龙,等. 基于改进 YOLO11 的印刷电路板表面缺陷检测算法[J]. 液晶与显示,2025,40(10): 1545-1556.
YANG Y P, GAO J W, LIU Z L, *et al.* Surface defect detection on printed circuit boards based on improved YOLO11 [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(10): 1545-1556. (in Chinese)
- [2] 李江,骆炜,陈豪,等. 基于改进 RT-DETR 的 PCBA 管脚焊点缺陷检测方法[J]. 液晶与显示,2025,40(10):1532-1544.
LI J, LUO W, CHEN H, *et al.* PCBA pin solder defect detection method based on improved RT-DETR [J]. *Chinese Journal of Liquid Crystals and Displays*, 2025, 40(10): 1532-1544. (in Chinese)
- [3] WENINGER D, SERNA S, RANNO L, *et al.* Advances in waveguide to waveguide couplers for 3D integrated photonic packaging [J]. *Light: Science & Applications*, 2026, 15(1): 17.
- [4] HE C S, ZHAO L, ZHANG S, *et al.* Intrinsic topological hinge states induced by boundary gauge fields in photonic metamaterials [J]. *eLight*, 2025, 5(1): 19.
- [5] ANNABY M H, FOU DA Y M, RUSHDI M A. Improved normalized cross-correlation for defect detection in printed-circuit boards [J]. *IEEE Transactions on Semiconductor Manufacturing*, 2019, 32(2): 199-211.
- [6] TSAI D M, HUANG C K. Defect detection in electronic surfaces using template-based Fourier image reconstruction [J]. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 2019, 9(1): 163-172.
- [7] HU B, WANG J H. Detection of PCB surface defects with improved faster-RCNN and feature pyramid network [J].

- IEEE Access*, 2020, 8: 108335-108345.
- [8] REDMON J, DIVVALA S, GIRSHICK R, *et al.* You only look once: unified, real-time object detection [C]//*Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas: IEEE, 2016: 779-788.
- [9] YUAN M H, ZHOU Y B, REN X Y, *et al.* YOLO-HMC: An improved method for PCB surface defect detection [J]. *IEEE Transactions on Instrumentation and Measurement*, 2024, 73: 2001611.
- [10] 周著国,鲁玉军,吕利叶. 基于改进YOLOv5s的印刷电路板缺陷检测算法[J]. *浙江大学学报(工学版)*, 2025, 59(8): 1608-1616.
- ZHOU Z G, LU Y J, LV L Y. Improved YOLOv5s-based algorithm for printed circuit board defect detection [J]. *Journal of Zhejiang University (Engineering Science)*, 2025, 59(8): 1608-1616. (in Chinese)
- [11] ZHOU Z G, LU Y J, LV L Y. Real-time surface defect detection algorithm for printed circuit boards based on improved YOLOv11n [J]. *The Journal of Supercomputing*, 2025, 81(10): 1148.
- [12] 孟伟君,安雯,马素刚,等. 基于特征增强与自适应阈值非极大值抑制的目标检测算法[J]. *北京航空航天大学学报*, 2025, 51(7): 2349-2359.
- MENG W J, AN W, MA S G, *et al.* An object detection algorithm based on feature enhancement and adaptive threshold non-maximum suppression [J]. *Journal of Beijing University of Aeronautics and Astronautics*, 2025, 51(7): 2349-2359. (in Chinese)
- [13] CARION N, MASSA F, SYNNAEVE G, *et al.* End-to-end object detection with transformers [C]//*Proceedings of the 16th European Conference on Computer Vision*. Glasgow: Springer, 2020: 213-229.
- [14] ZHAO Y A, LV W Y, XU S L, *et al.* DETRs beat YOLOs on real-time object detection [C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2024: 16965-16974.
- [15] WANG Y S, WU B B, ZHANG L H, *et al.* Enhanced PCB defect detection via HSA-RTDETR on RT-DETR [J]. *Scientific Reports*, 2025, 15: 31783.
- [16] 李鹏,余珺泽,于涛,等. 基于轻量化RT-DETR的PCB缺陷检测算法[J]. *计算机工程与设计*, 2025, 46(9): 2714-2721.
- LI P, YU J Z, YU T, *et al.* PCB defect detection algorithm based on lightweight RT-DETR [J]. *Computer Engineering and Design*, 2025, 46(9): 2714-2721. (in Chinese)
- [17] HUANG S H, LU Z C, CUN X D, *et al.* DEIM: DETR with improved matching for fast convergence [C]//*Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2025: 15162-15171.
- [18] ZHAO X L, ZHANG H F, SONG C H, *et al.* Lightweight intelligent detection algorithm for surface defects in printed circuit board [J]. *Computers & Industrial Engineering*, 2025, 203: 111030.
- [19] GAO Z W, LI Y Y, YUAN S H. DECNet: an efficient improved model for printed circuit board surface defect detection [J]. *Signal, Image and Video Processing*, 2026, 20(4): 188.
- [20] ZHU L Y, WANG S, CHEN M F, *et al.* Subtle defect detection network: more accurately detect subtle defects on the printed circuit board surface [J]. *Engineering Applications of Artificial Intelligence*, 2025, 152: 110765.
- [21] 朱黎颖,王森,沈爱萍,等. 复杂背景下的电路板表面焊接缺陷视觉检测[J]. *光学精密工程*, 2024, 32(14): 2256-2271.
- ZHU L Y, WANG S, SHEN A P, *et al.* Visual inspection of soldering defects on board surfaces against complex backgrounds [J]. *Optics and Precision Engineering*, 2024, 32(14): 2256-2271. (in Chinese)
- [22] 欧宇枫,王嘉宏. 基于双分支特征提取的小目标印刷电路板缺陷检测方法[J/OL]. *计算机应用*, 2026: 1-9. <https://link.cnki.net/urlid/51.1307.TP.20251107.1547.014>.
- OU Y F, WANG J H. Small target printed circuit board defect detection method based on dual-branch feature extraction [J/OL]. *Journal of Computer Applications*, 2026: 1-9. <https://link.cnki.net/urlid/51.1307.TP.20251107.1547.014>. (in Chinese)
- [23] 魏嘉敏,续欣莹,韩晖,等. 基于YOLO-HorL的轻量级PCB缺陷检测算法[J]. *计算机工程与设计*, 2025, 46(11): 3283-3291.

- WEI J M, XU X Y, HAN H, *et al.* Lightweight PCB defect detection algorithm based on YOLO-HorL [J]. *Computer Engineering and Design*, 2025, 46(11): 3283-3291. (in Chinese)
- [24] LOU M, YU Y Z. OverLoCK: An overview-first-look-closely-next ConvNet with context-mixing dynamic kernels [C]//*Proceedings of the 2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2025: 128-138.
- [25] ZHANG X, SONG Y Z, SONG T T, *et al.* AKConv: Convolutional kernel with arbitrary sampled shapes and arbitrary number of parameters [J/OL]. *arXiv*, 2024: 2311.11587.
- [26] WANG A, CHEN H, LIN Z J, *et al.* Rep ViT: Revisiting mobile CNN from ViT perspective [C]//*Proceedings of the 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2024: 15909-15920.
- [27] YE T, WANG Z, QIN Y, *et al.* A hypergraph computing and knowledge-enhanced framework for forklift pallet pose estimation [J]. *Annals of the New York Academy of Sciences*, 2026, 1556(1): e70219.
- [28] HOWARD A G, ZHU M, CHEN B, *et al.* MobileNets: Efficient convolutional neural networks for mobile vision applications [J/OL]. *arXiv*, 2017: 1704.04861.
- [29] 陈名渝, 谢玥辰, 吕雄涛, 等. 冷轧钢表面与内部缺陷检测研究[J]. *中国光学(中英文)*, 2024, 17(4): 823-833.
CHEN M Y, XIE Y C, LV X T, *et al.* Detection of surface and internal defects in cold rolled steel [J]. *Chinese Optics*, 2024, 17(4): 823-833. (in Chinese)
- [30] PENG Y S, LI H B, WU P X, *et al.* D-FINE: Redefine regression task in DETRs as fine-grained distribution refinement [C]//*Proceedings of the 13th International Conference on Learning Representations*. Singapore: OpenReview.net, 2025: 1-23.
- [31] 蒋仕飞, 张兆国, 王法安, 等. 纸膜双层缺陷检测的视觉成像光场设计[J]. *中国光学(中英文)*, 2024, 17(2): 354-365.
JIANG S F, ZHANG Z G, WANG F A, *et al.* Design of optical field of vision imaging for defect detection of paper and transparent film [J]. *Chinese Optics*, 2024, 17(2): 354-365. (in Chinese)

作者简介:



高 雪,女,硕士,副教授,2010年于南京理工大学获得硕士学位,主要从事电子信息与嵌入式系统方面的教学与研究工作。E-mail:13585317253@163.com