

DOI: 10.19650/j.cnki.cjsi.J2514097

# MFSF-DETR: 一种基于多尺度特征移位融合的 PCB 缺陷与元件检测算法\*

张永宏<sup>1</sup>, 许鑫豪<sup>1</sup>, 尹贺峰<sup>2</sup>, 李子奇<sup>2</sup>

(1. 南京信息工程大学自动化学院 南京 210044; 2. 无锡学院自动化学院 无锡 214105)

**摘 要:**随着电子产品向高性能、小型化方向发展,印刷电路板(PCB)作为电子系统的核心载体,其设计与制造日趋复杂,元件排列更加紧密,结构也更加精细,从而对元件检测与缺陷检测提出了更高的要求。虽然以 YOLO 为代表的基于卷积神经网络的目标检测模型已获得大量研究,但这些模型只针对单一的缺陷或元件检测场景进行设计,且在小目标和密集场景的检测上效果有限,而 RT-DETR 的出现使得基于 Transformer 的端到端检测模型在实时检测领域有了出色的表现。为此,在 RT-DETR 模型基础上,针对 PCB 场景提出了一种基于 Transformer 的端到端实时目标检测模型 MFSF-DETR。首先,采用 Faster-CGLU Block 替换主干网络中的 Block 层,细化通道注意力机制,引入了纠缠 Transformer 模块(ETB)整合频域与空间域,丰富深层语义。然后,设计了自适应加权跨尺度特征融合网络(RAWCFF)代替了基于 CNN 的跨尺度特征融合网络(CCFN),并与跨尺度特征移位融合网络(CFSF)组成新的特征融合编码器,实现邻层特征与非邻层特征的深度交互。最后,分别使用 PCB 缺陷数据集 DsPCBSD+ 与 PKU-Market-PCB、PCB 元件数据集 PCB\_WACV、PCB 与无人机数据集 VisDrone2019 评估提出的模型在 PCB 场景下的检测效果与泛化能力。实验结果表明,MFSF-DETR 模型在缺陷与元件检测上达到了 85.6%、98.1%与 89.9%的最高精度,相比基线模型提高 3.1%、1.0%与 3.8%,同时 FPS 指标也达到了 120.2、57.1 与 71.8,实现了 PCB 背景下的高效、高精度检测。

**关键词:** 特征融合;深度学习;注意力机制;缺陷检测;元件检测

中图分类号: TH701

文献标识码: A

国家标准学科分类代码: 520.20

## MFSF-DETR: A PCB defect and component detection algorithm based on multi-scale feature shift fusion

Zhang Yonghong<sup>1</sup>, Xu Xinhao<sup>1</sup>, Yin Hefeng<sup>2</sup>, Li Ziqi<sup>2</sup>

(1. School of Automation, Nanjing University of Information Science & Technology, Nanjing 210044, China;

2. School of Automation, Wuxi University, Wuxi 214105, China)

**Abstract:** With the development of electronic products in the direction of high performance and miniaturization, printed circuit boards (PCBs), as the core carrier of electronic systems, are becoming more and more complex in design and manufacturing, with more closely arranged components and a finer structure, which puts forward higher requirements for component detection and defect detection. Although the target detection models based on convolutional neural networks represented by YOLO have received a lot of research, these models are only designed for a single defect or component detection scenario, and have a limited effect on the detection of small targets and dense scenarios. The emergence of RT-DETR has enabled Transformer-based end-to-end detection models to perform excellently in real-time detection. Therefore, based on the RT-DETR model, this article proposes an end-to-end real-time target detection model MFSF-DETR, based on a Transformer for PCB scenarios. Firstly, the Faster-CGLU Block is used to replace the Block layer in the backbone network, the channel attention mechanism is refined, and the entanglement transformer block (ETB) is introduced to integrate the frequency domain with the spatial domain to enrich the deep semantics. Then, the rep adaptive weighted cross-scale feature fusion (RAWCFF) is designed to replace the CNN-based cross-scale feature fusion and form a new feature fusion encoder with the cross-scale feature shift fusion (CFSF) to realize the deep interaction between neighboring and non-neighboring features. Finally, the proposed

收稿日期: 2025-05-29 Received Date: 2025-05-29

\* 基金项目: 国家自然科学基金项目(42175157)、江苏省高等学校基础科学研究面上项目(23KJB520036)、无锡市科技发展资金项目(K20231003)资助

model is evaluated using the PCB defect dataset DsPCBSD+, the PKU-Market-PCB dataset, the PCB component dataset PCB\_WACV, and the PCB and drone dataset VisDrone2019 to assess its detection performance and generalization ability in PCB scenarios. The experimental results show that the MFSF-DETR model achieved the highest accuracy of 85.6%, 98.1%, and 89.9% in defect and component detection, respectively, which is 3.1%, 1.0%, and 3.8% higher than the baseline model. Meanwhile, the FPS indicators also reached 120.2, 57.1, and 71.8, respectively, achieving efficient and high-precision detection in the PCB background.

**Keywords:** feature fusion; deep learning; attention mechanism; defect detection; component detection

## 0 引言

信息技术的快速发展推动了电子产品的普及,而电子产品的快速升级与功能拓展也不断提高对作为其核心组件的印制电路板(printed circuit board, PCB)的性能和质量的要求。由于元件尺寸的缩小、排布密度的提升以及制造工艺的复杂性与多样性<sup>[1]</sup>,对PCB元件与缺陷的检测提出了更高的要求。元件检测主要用于识别并定位PCB上的各类电子元器件,是实现自动化装配和质量控制的基础环节<sup>[2]</sup>,尤其在表面贴装技术(surface mounted technology, SMT)中,对于及时发现拾放过程中出现的元件偏移、缺失等问题具有重要意义。而缺陷检测则聚焦于识别PCB在生产过程中产生的开路、短路、焊接不良、异物等微小缺陷,这些缺陷往往隐藏在复杂背景或细微结构中,却能直接影响电子设备的正常运行,缩减其使用寿命,甚至导致严重的安全事故<sup>[3]</sup>。因此,在PCB的生产制造过程中,实现对元件与缺陷的精准检测与定位显得尤为重要。这不仅是保证产品质量的关键环节,也是提升生产效率、降低成本并满足现代电子工业高可靠性要求的重要技术保障<sup>[4]</sup>。

基于人工的传统PCB检测方法存在效率低下、漏检率高与易对成品造成损伤的问题。而自动光学检测(automated optical inspection, AOI)技术使用视觉算法对目标物体进行非接触式检测,有效避免了对成品的损伤,提高了检测的精度与效率,因此成为了一个重要的研究领域。

为实现PCB自动检测,研究者提出了多种基于机器视觉的PCB检测方法。Baygin等<sup>[5]</sup>将Hough变换与Canny边缘检测算子相结合提取图像特征,通过将测试图与参考图进行匹配,精确识别出缺孔类缺陷。Liu等<sup>[6]</sup>采用图像平滑、对比度增强与锐化等预处理手段对PCB图像进行增强,并结合数字形态学与阈值分割技术,实现了对缺陷的有效识别。Melnyk等<sup>[7]</sup>提出了一种基于聚类算法的PCB缺陷检测方法,通过对标准样本和缺陷样本进行聚类分析,实现对正常与异常情况的有效区分。Onshaanjit等<sup>[8]</sup>提出一种可同时检测并分类多种PCB缺陷的算法。该方法先对标准样本与缺陷样本提取特征向量并进行K-means聚类,然后计算欧氏距离,通过差值的正负判断铜是否过多、过少或无缺陷。Li等<sup>[9]</sup>基于图像

像素信息开发了一种表面贴装器件分割技术,通过提取小型器件如电阻和电容的印刷边缘特征,以及识别集成电路均质的大面积区域,实现了对不同元器件的有效分割。Chigateri等<sup>[10]</sup>提出了一种融合先验知识的目标检测方法,首先通过获取元件的长宽比和引脚数量进行初步筛选,随后计算器件的圆度特征,最终利用相关性匹配算法完成电子元件的识别与定位。Ghosh等<sup>[11]</sup>通过使用随机森林重要性和方差阈值方法对k-size进行特征选择,结合颜色特征、形状特征与纹理特征实现元件识别。但基于机器视觉的PCB检测方法受环境光线、成像条件、图像精度等因素以及算法泛化能力的限制<sup>[12]</sup>,鲁棒性欠佳,易引发漏检、误检等问题,需要人工后期校正。

随着深度学习方法的兴起,深度学习模型更好的特征提取能力和泛化能力,使得基于深度学习的PCB检测方法通过学习大量样本后,能够显著提升缺陷与元件的检测效果<sup>[13]</sup>。但基于深度学习的PCB检测中仍存在的挑战为:1)PCB中缺陷尺寸微小,各种缺陷分布缺乏规律,元件种类繁多,尺寸与形态差异大,存在大量密集分布的情况,且PCB图像背景复杂多变,为实现高精度检测带来挑战;2)检测模型通常仅针对缺陷检测或元件检测设计,虽然在缺陷或元件单一场景具有适应能力,但在统一的PCB背景下缺乏泛化能力;3)在实际检测中,深度学习模型的复杂性使得平衡计算成本、推理效率与检测精度存在困难。

针对以上挑战,本文通过改进RT-DETR(real-time detection Transformer)引入了一种新颖的多尺度PCB检测模型,主要贡献为:

1)基于RT-DETR提出了一种针对PCB场景的实时端到端目标检测模型—多尺度特征移位融合DETR(multiscale feature shift fusion detection Transformer, MFSF-DETR),确保模型实时检测效率的基础上,实现了对PCB背景中缺陷与元件的精确定位。

2)主干网络部分,设计了快速门控线性单元模块(faster convolutional gated linear unit block, Faster-CGLU Block),利用细粒度通道注意力机制增强特征提取能力,并减少参数量与计算量,引入了纠缠Transformer模块(entanglement Transformer block, ETB)模块,整合频域信息与空间域信息,有效捕捉全局上下文信息,丰富深层语义特征。

3) 设计了由自适应加权跨尺度特征融合网络 (rep adaptive weighted cross-scale feature fusion, RAWCFF) 与跨尺度特征移位融合网络 (cross-scale feature shift fusion, CFSF) 组成的特征融合编码器, RAWCFF 网络简化特征融合路径, 并利用自适应权重有效结合邻层特征图, CFSF 网络通过特征移位建立非邻层特征的直接交互, 增强了浅层细节信息以及深层语义信息的传播与融合, 再利用递进式的多尺度感受野提取目标的中心特征, 实现精确定位。

4) 针对所提出模型在 PCB 缺陷数据集 DsPCBSD+ 与 PKU-Market-PCB、PCB 元件数据集 PCB\_WACV 以及无人机数据集 VisDrone2019 上进行实验验证, 并与主流实时检测模型进行对比分析, 证明了 MFSF-DETR 模型的有效性与泛化能力。

## 1 相关工作

随着深度学习在 PCB 检测领域的广泛应用, 研究者们提出了多种基于深度学习模型的 PCB 缺陷与元件检测方法。

Lan 等<sup>[14]</sup> 通过引入 K-means++ 聚类方法优化 YOLOv3 的锚框设置, 以提升定位精度; 同时加入批归一化层以加快模型的推理速度, 并利用 GIoU 损失函数进一步增强检测性能。Xiao 等<sup>[15]</sup> 对 YOLOv7-tiny 的骨干网络进行了改进, 加入坐标注意力机制以增强模型的特征提取能力, 从而有效提升检测精度。Ran 等<sup>[16]</sup> 在 (single shot multibox detector, SSD) 深度卷积神经网络框架基础上, 设计了基于多尺度特征图的边界框机制, 以适应不同尺寸目标的检测需求, 并通过 3×3 小卷积核对类别及边界框信息进行预测。姜媛媛等<sup>[17]</sup> 提出了一个名为 Multi-CR YOLO 的轻量化印刷电路板缺陷检测网络, 通过增加小目标检测头, 显著提高了缺陷的检测精度和实时性。张莹等<sup>[18]</sup> 基于 YOLOv7 网络, 删除融合层中的扩展高效层聚合网络 (extended-efficient layer aggregation network, E-ELAN)、CBS、UpSample 与 20×20 的预测头, 增加高分辨率检测头, 并使用跨通道信息连接模块串联精简网络, 提升了小目标缺陷检测能力。王军等<sup>[19]</sup> 将压缩与激励 (squeeze-and-excitation, SE) 注意力机制融入 C2f (cross stage partial bottleneck with 2 convolutions) 中, 新增小目标检测尺度, 舍弃大目标检测尺度, 提出一种改进的 YOLOv8 小目标缺陷检测方法。王天洋等<sup>[20]</sup> 在 YOLOv8 基础上引入了分布移位卷积 (distribution shifting convolution, DSConv) 与轻量级跨尺度特征融合模块 (cross-scale feature fusion module, CFFM), 有效降低了算法的计算成本, 并提升了缺陷检测精度。Yuan 等<sup>[21]</sup> 提出了一种改进的 YOLOv5 检测模型 YOLO-HMC, 引入改进

的多卷积块注意模块, 增强模型在相似背景下的缺陷突出能力, 并使用了 (content-aware reassembly of features, CARAFE) 上采样机制, 有效整合图像上下文信息, 扩大感受野, 提升了对微小缺陷的定位精度。

Ling 等<sup>[22]</sup> 提出了一种基于改进的 YOLOv8 的元件检测模型, 引入了轻量级的 Ghost 卷积, 有效降低计算成本, 结合 Sig-IoU 损失函数, 实现对元件的高精度定位。Chi 等<sup>[23]</sup> 使用特征提取网络 EfficientNetv2-L 替换 YOLOv4 原有主干网络, 融合数据增强技术, 在复杂的 PCB 布局与变化的光照条件下展现出优势。Liu 等<sup>[24]</sup> 设计了高斯交并边界框回归损失函数 GsIoU (Gaussian-IoU), 以 YOLOv4 为基线模型, 使用高斯函数合并同一位置不同锚点下的元件预测框来计算框回归损失, 从而提高最终框回归的精度。Luo 等<sup>[25]</sup> 基于 YOLOv7 使用兼具自注意力机制与卷积优势的 ACmix (a mixed model that enjoys the benefits of both self-attention and convolution) 模块进行改进, 设计了 EC-YOLOv7 网络用于检测电子元件。Li 等<sup>[26]</sup> 基于 YOLOv3 通过扩展不同深度层中有效感受野大小, 结合锚框大小匹配算法设计了轻量级的实时检测网络。Mallaiyan Sathiseelan 等<sup>[27]</sup> 以基于颜色的背景扣除算法为基础设计了基于两阶段检测模型的电子元件定位和检测网络 (electronic component localization and detection network, ECLAD Net), 实现对电阻与电容元件的精确检测。

现有工作主要针对缺陷检测场景或元件检测场景中的一种场景进行研究, 在泛化实验中也未对另一场景进行深入探索。同时, 所使用的深度学习检测方法主要为以 YOLO (you only look once, YOLO) 系列为代表的基于卷积神经网络 (convolutional neural networks, CNN) 的检测模型为主, 而 RT-DETR 的出现使得基于 Transformer 的端到端检测模型在实时检测领域有了出色的表现。因此, 本文旨在研究基于 Transformer 的实时检测模型在 PCB 检测场景的应用, 通过改进 RT-DETR 模型, 使其针对 PCB 场景下的目标检测 (包括缺陷检测与元件检测) 均有较好适应能力, 从而设计出一种能够实现高效、高精度检测的 PCB 场景泛用的实时检测模型。

## 2 检测模型

### 2.1 背景理论

与 YOLO 系列基于 CNN 的检测模型相比, DETR<sup>[28]</sup> 是基于 Transformer 的端到端检测模型, 通过将二分匹配损失与并行解码 Transformer 结合, 舍弃非极大值抑制 NMS 等后处理步骤, 避免了锚框等先验知识的约束, 从而简化检测流程。RT-DETR<sup>[29]</sup> 则在 DETR 基础上, 使用高效的混合编码器取代 Transformer 编码器, 大大提高了

模型的推理速度,使得 RT-DETR 在 COCO 数据集上的检测精度与推理速度超越了同等规模的 YOLO 系列模型,成为极具竞争力的实时端到端目标检测模型,RT-DETR 模型结构如图 1 所示。

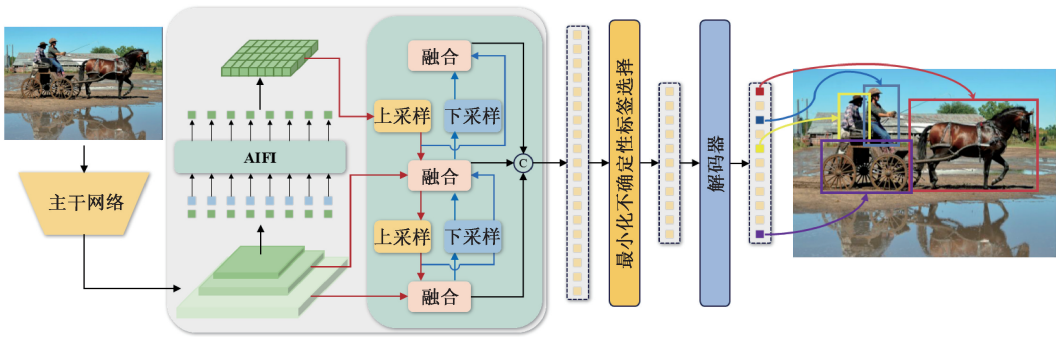


图 1 RT-DETR 结构  
Fig. 1 RT-DETR architecture

RT-DETR 模型由主干网络、高效混合编码器和带有辅助预测头的 Transformer 解码器组成,输入尺寸 640×640 特征图通过主干网络进行特征提取,生成 80×80、40×40、20×20 由浅层到深层的三级特征图。编码器部分,RT-DETR 舍弃了 DETR 中使用 vanilla Transformer 编码器对所有的三级特征都进行尺度内特征交互与尺度间特征融合的策略,仅在尺寸为 20×20 的深层特征图上使用基于注意力机制的尺度内特征交互 (attention-based intra-scale feature interaction, AIFI),对于尺度间特征融合则采用基于 CNN 的跨尺度特征融合 (CNN-based cross-scale feature fusion, CCFF),避免了注意力编码机制因过长的三级特征序列而导致的计算成本与推理延迟大增,大大提升了特征混合编码的效率。特征编码输入 Transformer 解码器后,通过最小化不确定性标签选择 (uncertainty-minimal query selection) 提供更具

于准确分类与精确定位的特征,再由解码头解码为坐标框与类标签,实现对目标最终的检测。

2.2 改进模型 MFSF-DETR

针对 PCB 图像中缺陷与元件的检测,本文在 RT-DETR 基础上提出了一种高效的多尺度特征移位融合 DETR (MFSF-DETR) 模型,其结构如图 2 所示。主干网络部分,使用 Faster-CGLU Block 替换 Resnet18 中基本的 Block 层,减少计算冗余的同时以细粒度通道注意力优化特征表示。再引入 ETB 模块替换 AIFI 模块,充分交织融合频域信息与空间信息,得到更丰富的深层语义特征。提出的 RAWCFF 网络与 CFSF 网络取代 CCFF 作为新的卷积编码器,RAWCFF 网络简化了多层级特征融合路径,并灵活利用通道与空间注意力加强相邻层间特征交互, CFSF 网络则建立非相邻层间的信息互补,以多尺度感受野丰富特征融合编码,实现对于检测目标的精确捕获。

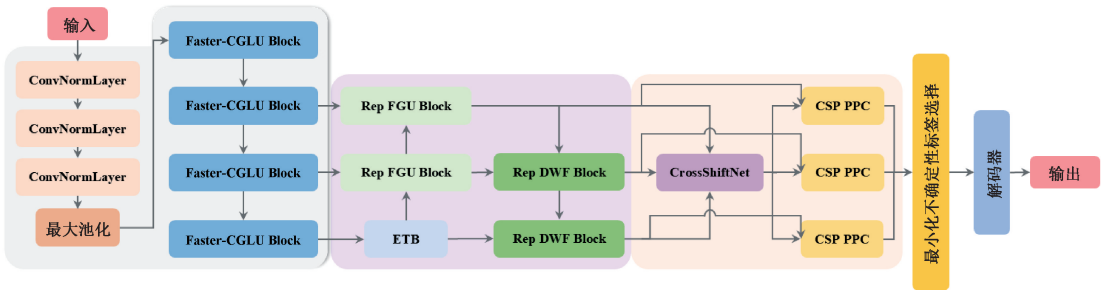


图 2 MFSF-DETR 结构  
Fig. 2 MFSF-DETR architecture

2.3 Faster-CGLU Block 模块

针对主干网络,本文结合 FasterNet Block 模块与 CGLU 机制设计了 Faster-CGLU Block 模块,旨在保证主干特征提取能力的同时减轻参数量与计算量,其结构如图 3 所示。

在阶段 1 中利用 3×3 卷积与平均池化操作对输入特征图进行降维,紧接 Faster CGLU 模块和 1×1 卷积进行特征提取与通道扩展,在阶段 2 和 3 中,保持特征图维度不变,一侧经过 3×3 卷积与 Faster CGLU 模块处理,另一侧经过 1×1 卷积处理,最后两者相加并经过 ReLU 函数

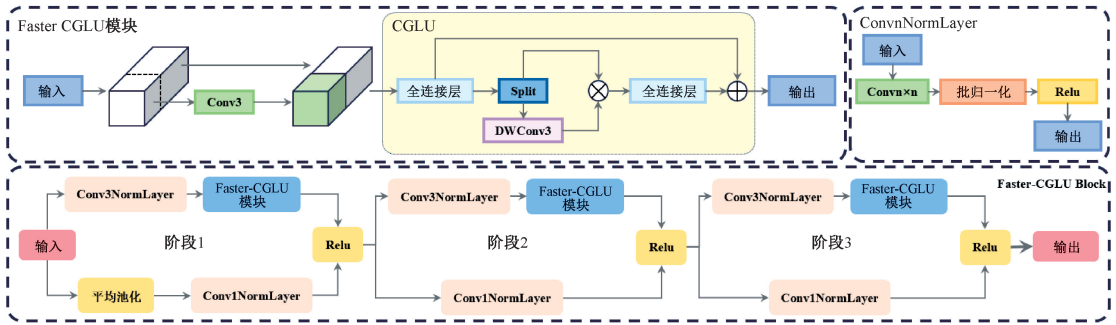


图3 Faster-CGLU block 结构  
Fig. 3 Faster-CGLU block architecture

激活,形成残差结构。而Faster CGLU 模块沿用FasterNet<sup>[30]</sup>中的局部卷积(partial convolution),其只在部分输入通道上进行常规的 $3\times 3$ 卷积操作来提取空间特征,其余的通道保持不变,相比直接使用常规卷积大大减少了计算冗余。

基于常规通道注意力机制的压缩的SE模块使用全局平均池化聚合全局信息,再通过激活函数生成门控信号来控制输出信息流,全局池化方法使得每个通道得到了不同的通道门控信号,但同一通道的特征图上所有的token享有同样的空间门控信号,导致其机制不够灵活且粒度过粗。而Faster CGLU模块中使用的CGLU<sup>[31]</sup>门控模块实现了细粒度通道注意力机制,输入特征图经过全连接层处理后被拆分成两部分,其中一部分经过 $3\times 3$ 深度可分离卷积处理得到门控信号,深度可分离卷积对每个通道的特征图使用一个独立的单通道卷积进行卷积操作,生成每个通道的通道门控信号,同时,相比全局平均池化将特征图降维至 $1\times 1$ 生成门控信号,深度可分离卷积操作可生成原尺寸的特征图作为空间门控信号,特征图上的每个token都能得到基于最近邻特征生成的门控信号,从而使得每个通道与特征图中的每个token都能拥有唯一的门控信号。将基于最近邻特征的细粒度门控信号与另一部分特征图相乘得到注意力信息,再经过全连接层处理,与原输入相加构成残差连接后输出。

Partial Convolution有效地减少了冗余计算,节省了计算开销,使得在Faster-CGLU Block模块中可以堆叠更多的卷积层,对特征图进行更深层次的处理,而CGLU门控模块充分利用了所有通道以及特征图的空间信息,生成了粒度更细的通道门控信号,实现了更高质量的全局信息聚合,有效增强了模型的鲁棒性。

#### 2.4 纠缠Transformer模块

AIFI模块使用标准的Transformer编码器结构,使用多头注意力机制处理沿空间维度展平的张量,并在其基础上利用正弦、余弦函数生成二维位置编码,以便Transformer层能够捕捉到空间位置信息。AIFI模块利用

注意力机制在尺度内进行全局特征融合,但其仅在空间中实现了特征的整合表示。

本文使用ETB<sup>[32]</sup>替换AIFI模块,ETB模块分别在频域与空间域中对原特征图进行处理,融合频域与空间域中生成的注意力信息以实现丰富而全面的特征表示,其结构如图4所示,输入 $X_0$ 经过频率自注意力(frequency self-attention, FSA)和空间自注意力(spatial self-attention, SSA)处理得到频域特征图 $X_f$ 与空间特征图 $X_s$ , $X_f$ 、 $X_s$ 与原始输入 $X_0$ 相加后得到 $X_1$ ,经过纠缠前馈网络(entanglement feed-forward network, EFFN)进行深度交织融合得到频域-空间混合特征 $X_{fs}$ ,与 $X_1$ 进行残差连接后,得到更具鲁棒性的特征表示 $X_2$ 。FSA利用快速傅里叶变换(fast Fourier transform, FFT)将输入特征转至频域,生成 $Q_f$ 、 $K_f$ 、 $V_f$ ,通过 $Q_f$ 和 $K_f$ 生成频域注意力图作为权重,优化 $V_f$ 的特征表示,然后使用逆变换(inverse fast Fourier transform, IFFT)得到频域注意力特征,与经过自适应权重校正的输入特征进行融合后,得到频域特征 $X_f$ 。在SSA中,利用 $1\times 1$ 卷积生成位置编码,再通过拼接2个深度可分离卷积的输出得到 $Q_s$ 、 $K_s$ 、 $V_s$ ,同样通过自注意力机制得到空间注意力特征,与输入特征进行融合得到空间特征 $X_s$ 。融合了频域信息 $X_f$ 与空间信息 $X_s$ 的输入特征 $X_1$ 进入EFFN后,分别使用快速傅里叶变换与深度可分离卷积处理得到频域门控信号 $G_f$ 与空间域门控信号 $G_s$ ,利用门控机制控制输入信息流,得到频域分支与空间分支,将两个分支拼接后,再分别经过空间与频域的独立处理后融合输出频域-空间混合特征 $X_{fs}$ 。

ETB模块通过频域与空间域的多次融和、全局特征与局部特征的多次交互,使得取自不同状态、不同区域的特征相互交织、相互适应,从而能够稳定的得到丰富而全面的特征。

#### 2.5 自适应加权跨尺度特征融合网络

CCFF网络使用多个由卷积构成的融合模块,逐层级将相邻尺度的特征融合成新的特征。CCFF网络对于特征表示的优化主要依靠由卷积组成的融合模块,在输入

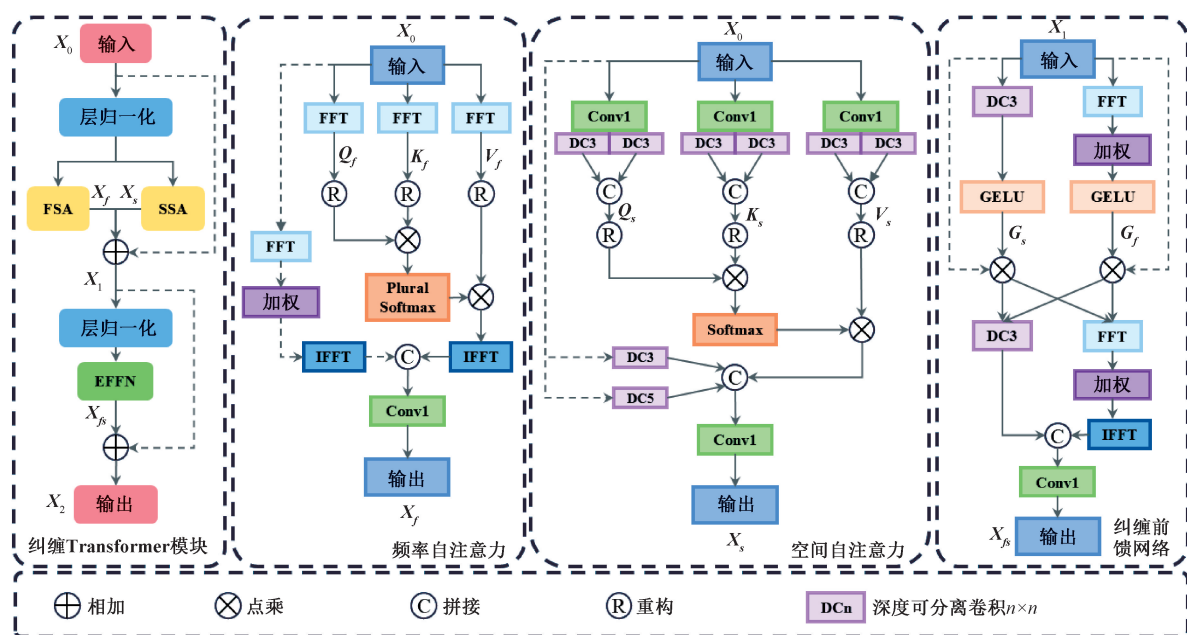


图 4 ETB 结构

Fig. 4 ETB architecture

融合模块前, 只对相邻层级特征图升维或降维之后进行简单拼接, 这些匹配维度与简单拼接输入的操作占据了网络的参数量与计算量, 但并未作用于提取浅层特征中的细节信息或深层特征中的语义信息, 导致实际对于优化特征表示的贡献有限。因此, 本文设计了自适应加权跨尺度特征融合网络 RAWCFF, 其主要由基于特征引导的上采样模块 (rep feature-guided upsample block, Rep

FGU Block) 与动态加权融合模块 (rep dynamic weighted fusion block, Rep DWF Block) 构成, Rep FGU Block 与 Rep DWF Block 都使用自适应加权融合处理相邻层级的输入, 然后通过基于跨阶段部分融合 (cross stage partial, CSP) 结构的 RepConv 卷积网络进一步优化特征表示。CCFF 网络与 RAWCFF 网络的融合路径以及 Rep FGU Block 与 Rep DWF Block 的结构如图 5 所示。

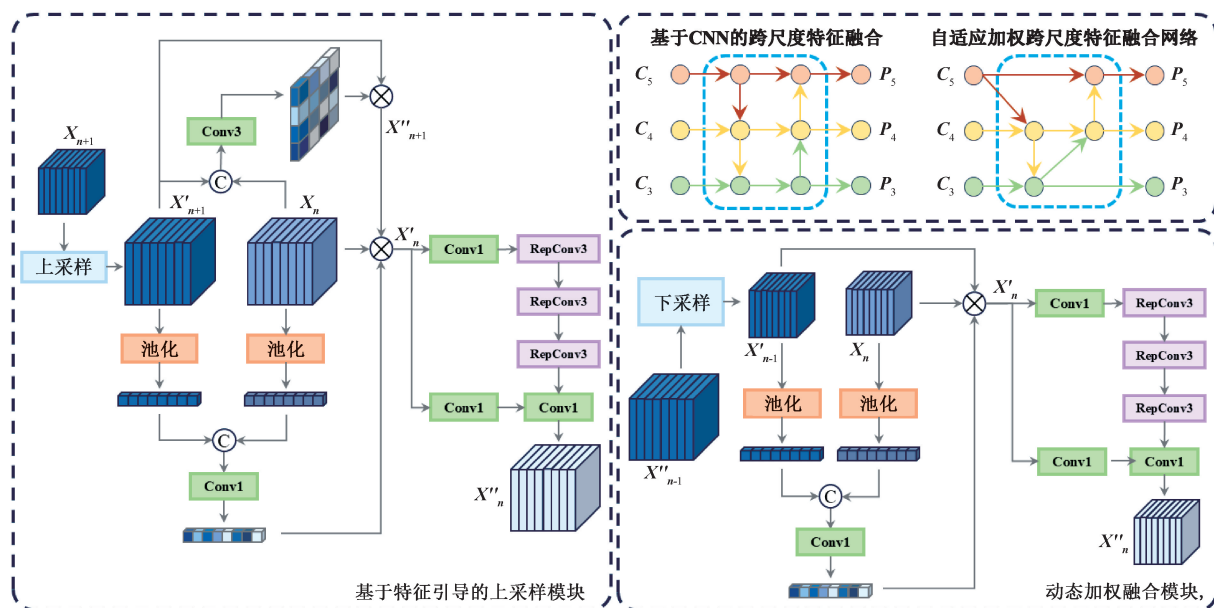


图 5 CCFF 网络与 RAWCFF 网络融合路径及 Rep FGU Block 与 Rep DWF Block 结构

Fig. 5 CCFF network and RAWCFF network convergence path and structure of rep FGU block and rep DWF block

自上而下的融合路径使用了 Rep FGU Block, 先将深层输入特征  $X_{n+1}$  通过双线性差值上采样至与浅层特征  $X_n$  同尺寸, 记作  $X'_{n+1}$ 。然后对相同尺寸的  $X'_{n+1}$  与  $X_n$  进行拼接操作, 通过  $3 \times 3$  卷积将特征图通道缩减至 1, 生成空间注意力图, 通过 softmax 函数归一化作为空间权重, 与来自深层特征的  $X'_{n+1}$  相乘, 得到由浅层特征引导生成的深层特征图  $X''_{n+1}$ 。另一边, 对  $X'_{n+1}$  与  $X_n$  分别进行全局平均池化, 把空间维度将至 1, 生成通道注意力图, 将两部分注意力图拼接, 通过  $1 \times 1$  卷积得到通道权重, 浅层特征  $X_n$  和深层特征  $X''_{n+1}$  与通道权重相乘后进行结合, 得到充分融合浅层细节信息与深层语义信息的特征图  $X'_n$ 。随后, 特征图  $X'_n$  将被输入基于 CSP 结构的 RepConv 卷积网络进行进一步优化。首先经过两个  $1 \times 1$  卷积将通道减半, 一部分不做操作保留原始信息, 另一部分连续进行 3 次 RepConv 操作对特征进行优化, 将两部分的输出进行拼接, 通过  $1 \times 1$  卷积处理后输出最终的融合特征图  $X''_n$ 。

自下而上的融合路径中则使用 Rep DWF Block, 先将浅层输入特征  $X_{n-1}$  下采样至与深层特征  $X_n$  同尺寸, 记作  $X'_{n-1}$ 。与 Rep FGU Block 不同, Rep DWF Block 中删去了基于相邻层特征引导的空间注意力机制部分, 只保留通道注意力机制部分。然后利用通道权重, 结合  $X'_{n-1}$  与  $X_n$  计算得到特征  $X'_n$ , 再经过卷积网络处理后, 输出融合的特征图  $X''_n$ 。

Rep FGU Block 同时使用基于相邻层特征引导的空间注意力机制与通道注意力机制, 利用浅层特征中利于精确定位的细节信息, 优化由上采样得到的深层特征, 使其中的语义信息在加权融合之前更具辨别力。而 Rep DWF Block 仅使用通道注意力机制, 在将浅层特征融入深层特征的过程中, 尽可能多的保留原始的浅层细节信息, 丰富特征表示。RepConv 卷积网络训练时内部使用多分支卷积层, 而推理时将分支部分重参数化到主分支, 进而减少参数量与计算量。CSP 结构使得一部分输入经过卷积层深度处理, 另一部分保留原始信息, 两部分拼接后输出, 能够保留更多的特征信息。RAWCFF 网络简化了特征融合路径, 将维度匹配与拼接输入过程与卷积模块集成, 并在其中融入了空间与通道注意力机制, 使相邻层特征得以充分融合, 基于 CSP 结构的 RepConv 卷积网络以较低的计算量与参数量, 有效利用了融合后的原始信息, 优化了特征表示, 增强了网络的特征表达能力。

## 2.6 跨尺度特征移位融合网络

CCFF 网络中, 融合操作只在相邻层间进行, 因此得到的输出特征图与相邻层级的特征图之间有较高的相关性, 而非相邻层级的特征图相关性很低, 改进后的 RAWCFF 网络中融合操作也更侧重于相邻层级特征图之间。为了使各层级特征之间能够直接进行信息交互, 本

文设计了跨尺度特征移位融合网络 (cross-scale feature shift fusion, CFSF), 引入了跨尺度移位网络 (cross-scale shift network, CSN) [33] 模块, 并对其进行修改调整, 使其能够适用于 3 层特征金字塔。CSN 模块利用移位操作, 直接移动各层级特征通道, 以此来与非相邻层级进行信息交互。辅以基于跨阶段部分融合的平行风车卷积 (cross stage partial parallel pinwheel-shaped conv, CSP PPC) 模块对移位后的各层级特征进行处理, 利用递进式大尺度感受野捕获不同尺度目标的特征信息, 实现非相邻层级特征之间的互补。融合网络整体结构如图 6 所示。

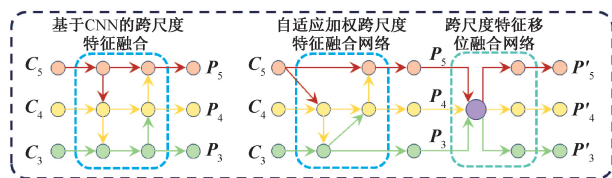


图6 CFSF 网络融合路径

Fig. 6 CFSF network convergence path

### 1) 跨尺度移位网络

CSN 模块由多尺度移位模块 (multi-scale shift module, MSM) 与双全局上下文模块 (dual global context module, DGCM) 两部分组成, 如图 7 所示。输入的特征金字塔通过升维与降维调整至相同尺寸, 并且按尺度拼接为一个特征序列。随后, 在 MSM 中进行跨层级的特征移位, 实现非相邻层之间的信息交互, DGCM 则在尺度与空间维度上整合全局上下文信息, 对移位后特征进行加权融合。最后, 按尺度维度将各层级特征调整回原尺寸, 与原始输入结合后输出一个新的融合特征金字塔。

MSM 部分接受按尺度排序的特征序列后, 在每个尺度的特征图中选取部分通道的特征并沿尺度维度移动, 生成移位特征, 再将移位特征与原特征序列拼接, 得到包含非相邻层级信息的特征序列。经过 2 次  $1 \times 1$  卷积融合信息并调整通道后, 添加回原始特征序列。具体的移位特征生成过程如图 8 所示, 原始特征序列中包含 3 个尺度的特征图, 分别为特征图  $a$ 、 $b$ 、 $c$ , 从特征图  $a$  中选取  $1/4$  的通道, 并将其等分成 4 部分, 记作  $a_1$ 、 $a_2$ 、 $a_3$ 、 $a_4$ , 特征图  $b$ 、 $c$  依次类推。然后, 将选中通道中的第 1 与 4 部分向前移动 1 个尺度, 将第 2 与第 3 部分向前移动 2 个尺度, 以此来生成移位特征, 例如, 将  $a_1$  与  $a_4$  部分移动至特征图  $c$  处, 作为特征图  $c$  的补充信息, 将  $a_2$  与  $a_3$  部分移动至特征图  $b$  处, 作为特征图  $b$  的补充信息。最后将移位特征与对应尺度的原始特征拼接, 得到融合了跨层特征信息的特征序列。

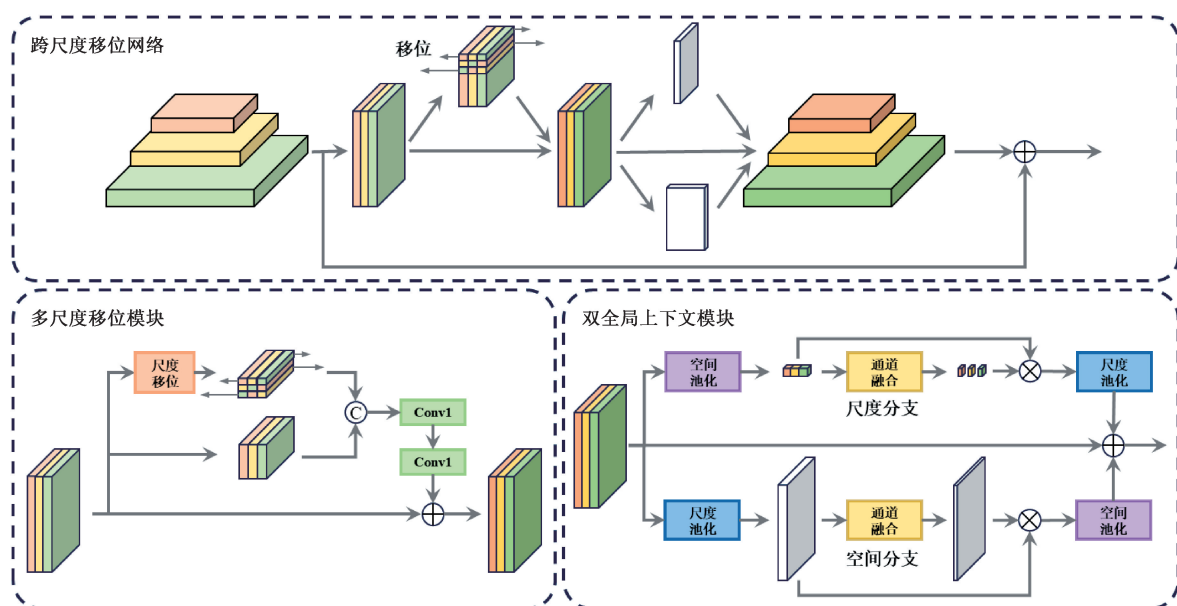


图 7 CSN 结构

Fig. 7 CSN architecture

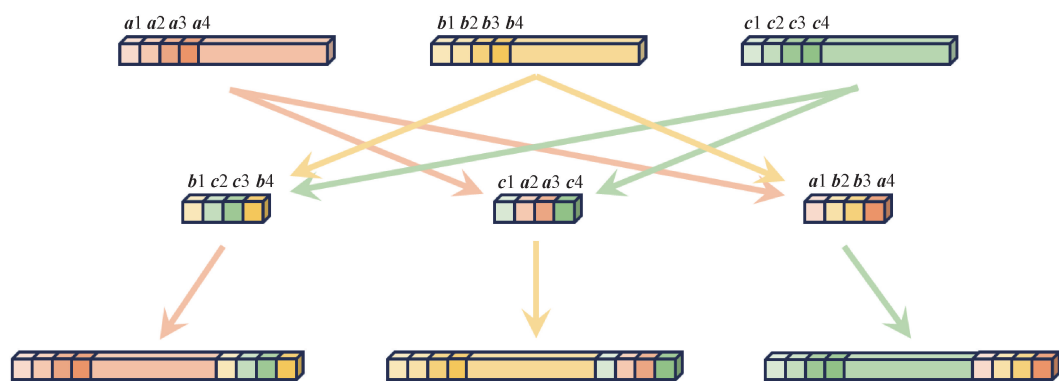


图 8 Scale shift 示意图

Fig. 8 Scale shift diagram

DGCM 部分主要对移位后特征进行加权融合。尺度分支中,先利用全局平均池化将特征图空间维度降至 1,然后使用  $1 \times 1$  卷积对通道上下文信息进行聚合,通过 softmax 生成空间维度和通道维度均为 1 的对应尺度维度的注意力权重,利用生成的权重对去除空间维度的特征图进行加权,最后在尺度维度上应用全局平均池化聚合尺度上下文信息,并将输出添加回主分支。空间分支中,先在尺度维度上进行平均池化,去除尺度维度,然后同样使用  $1 \times 1$  卷积对通道上下文信息进行聚合,通过 softmax 生成单通道的空间注意力权重,利用生成的权重对去除尺度维度的特征图进行加权,最后在空间维度上应用全局平均池化聚合空间上下文信息,并将输出添加回主分支。通过引入尺度分支与空间分支整合的全局尺度与空间上下文信息,加强了不同层级移位特征的相关性,优化

了网络的融合效果。

## 2) 基于跨阶段部分融合的平行风车卷积

为了进一步增强移位后特征的表达能力,本文在风车卷积 (pinwheel-shaped conv, PConv)<sup>[34]</sup> 基础上设计了 CSP PPC 模块,如图 9 所示。模块整体采用 CSP 结构,其中一部分图像特征输入平行风车卷积 (parallel pinwheel-shaped conv, PPC) 网络处理,另一部分特征保留,从而得到更丰富的特征表示。

PPC 网络平行地使用 3 个 PConv 卷积,3 个 PConv 分别选择 3、4、5 作为卷积核大小,以此来扩大模块的感受野。PConv 结构如图 9 左下所示,图示为卷积核设置为 3 的情况,首先使用 4 组不对称填充为条形卷积创建不同的卷积域,Padding(1,0,0,3) 表示在原特征图的左、右、上、下方向填充的像素个数。然后根据填充的不同,

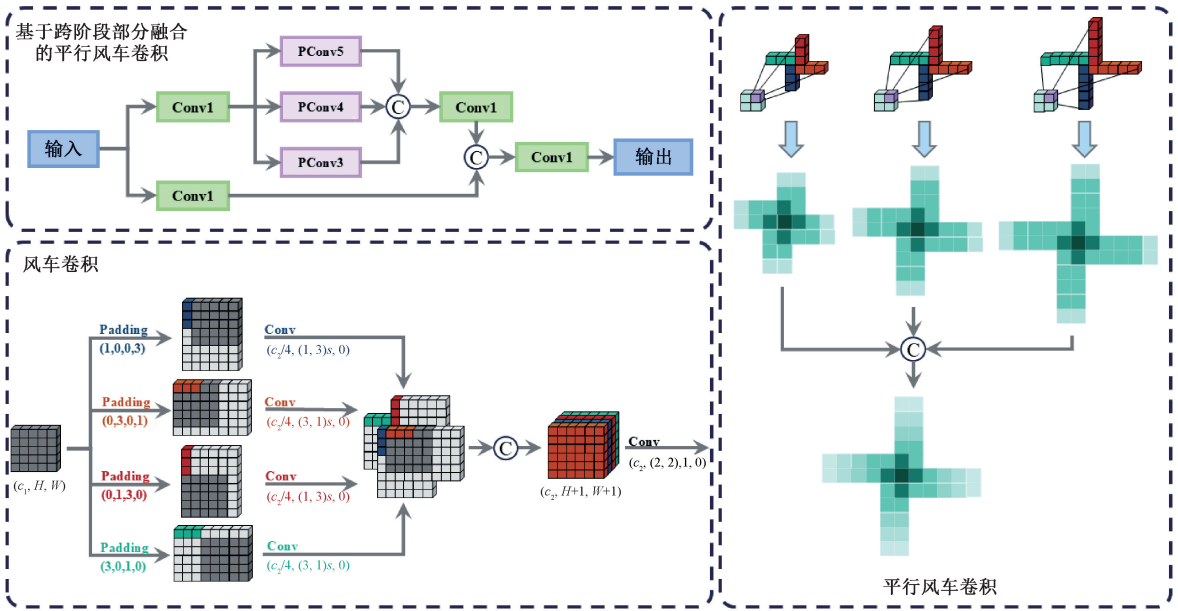


图9 CSP PPC 结构  
Fig. 9 CSP PPC architecture

使用垂直或水平的条形卷积对特征图进行处理, Conv(1,3)表示1×3的垂直卷积,Conv(3,1)表示3×1的水平卷积。最后将4组卷积得到的特征图沿通道维度拼接,通过不填充的2×2卷积调整特征图通道与尺寸后输出。卷积核为3的PConv的感受野为25,相比3×3卷积增加了177%,而在输入通道为 $c_1$ 输出通道为 $c_2$ 时,3×3卷积计算量为 $9c_1c_2$ ,PConv仅为 $4 \times (3 \times 1 \times c_1 \times (c_2/4)) + 4c_1c_2 = 7c_1c_2$ ,PConv使用更少的参数有效地扩展了模块的感受野。

在PPC网络中,3个PConv平行放置,同时对输入特征进行处理,并将结果拼接,实际等效感受野如图9右侧所示,像素点经过卷积处理的次数由中心像素向外递减,即感受野的有效性向外递减,类似高斯分布。目标越小,提取的特征也越集中,利于突出小目标物体;而面对大目标,也能高效的提取中心特征,实现基于目标中心的精确定位。

3 实验与结果分析

3.1 数据集

为了评估提出方法的检测效果,本文在4个公开数据集上进行了实验,分别是PCB缺陷数据集DsPCBSD<sup>[35]</sup>与PKU-Market-PCB<sup>[36]</sup>、PCB元件数据集PCB\_WACV<sup>[37]</sup>与无人机数据集VisDrone2019<sup>[38]</sup>。

1) DsPCBSD+数据集

DsPCBSD+数据集是由华南农业大学工程学院收集

并公开的PCB缺陷数据集,在残铜、缺铜、导体划伤、异物4种主要缺陷类别的基础上,进一步细分为短路(SH)、杂散(SP)、杂铜(SC)、开路(OP)、鼠咬(MB)、破孔(HB)、导体划伤(CS)、导体异物(CFO)和基材异物(BMFO)9种类别。因PCB微型化,图像微小细节相应增多,缺陷尺寸整体较小,易与背景混淆,CS、CFO和BMFO等缺陷类别内部颜色、大小和形态上都有较大的差异,具有强类内可变性。该数据集共计包含10 259张缺陷图像和20 276个人工标注的缺陷目标,本文将该数据按8:1:1的比例随机分成训练集、验证集与测试集,DsPCBSD+数据集详细信息如表1所示。

2) PKU-Market-PCB数据集

PKU-Market-PCB数据集是由北京大学智能机器人开放实验室制作的PCB缺陷数据集,包括缺孔(missing hole)、鼠咬(mouse bite)、断路(open circuit)、短路(short)、毛刺(spur)、残铜(spurious copper)6种缺陷类型,共计693张缺陷图像和2 953个目标缺陷。每张图像上均只存在一种类型的缺陷,每张图像均有2~6个缺陷标签,每种类型缺陷所占图像张数与缺陷数大致相同,类别之间较为均衡,避免了类别间差异的影响。本文将该数据集按7:1:2的比例随机分成训练集、验证集与测试集,PKU-Market-PCB数据集详细信息如表2所示。

3) PCB\_WACV数据集

PCB\_WACV数据集是由Georgia Institute of Technology公开的PCB元件数据集,共计包含47张高分率图像,31种组件类型,62 000个元件实例。该数据集中元件类内差异很大,标签相同,但是由于元件具体类

表 1 DsPCBSD+数据集详细信息  
Table 1 DsPCBSD+ dataset details

























缺陷类别	缺陷样例	标签名称	缺陷数
短路		SH	915
尖刺		SP	4 584
杂铜		SC	1 593
断路		OP	1 770
鼠咬		MB	2 529
孔崩		HB	2 883
导体划痕		CS	2 490
导体异物		CFO	1 832
基材异物		BMFO	1 680
总计			20 276

表 2 PKU-Market-PCB 数据集详细信息  
Table 2 PKU-Market-PCB dataset details

缺陷类别	缺陷样例	标签名称	缺陷图像数	缺陷数
缺孔		missing hole	115	497
鼠咬		mouse bite	115	492
断路		open circuit	116	482
短路		short	116	491
毛刺		spur	115	488
残铜		spurious copper	116	503
总计			693	2 953

型繁多,差异较大,而元件类间差异较小,由于封装相似,不同元件相似度很高。本文在原数据集基础上,将每一张图像随机以  $224\times 224$ 、 $256\times 256$ 、 $320\times 320$ 、 $384\times 384$ 、 $448\times 448$ 、 $512\times 512$  中一种分辨率进行裁剪,得到共计 1 263 张图像。原数据集中不同类别目标个数差异较大,经过筛选后保留标签个数超过 200 个的类别,分别为接线端子 (connector)、电阻 (resistor)、焊盘 (pads)、电容 (cap)、管脚 (pins)、测试点 (test\_point)、发光二极管 (led)、芯片 (ic)、电解电容 (electrolytic\_cap) 9 个类别的标签,并且同样按 8:1:1 的比例随机分成训练集、验证集与测试集,筛选后的 PCB\_WACV 数据集详细信息如表 3 所示。

表 3 PCB\_WACV 数据集详细信息  
Table 3 PCB\_WACV dataset details

元件类别	元件样例	标签名称	元件数
接线端子		connector	609
电阻		resistor	2 249
焊盘		pads	359
电容		cap	2 674
管脚		pins	360
测试点		test_point	301
发光二极管		led	238
芯片		IC	413
电解电容		electrolytic-cap	277
总计			7 480







4) VisDrone2019 数据集

VisDrone2019 数据集是由天津大学公开的大型无人机视角数据集,共计 8 629 张图片与 457 066 个人工注释的边界框,其中训练集 6 471 张、验证集 548 张、测试集 1 610 张。由无人机配备的相机在不同场景、天气与照明条件下采集,取自 14 个不同地区,包括城市与乡村环境,

数据集中存在大量且密集的小目标,主要包含行人、车辆与自行车等目标,具体为行人 ( pedestrian )、人 ( people )、自行车 ( bicycle )、私家车 ( car )、货车 ( van )、卡车 ( truck )、三轮车 ( tricycle )、遮阳三轮车 ( awning-tricycle )、公交车 ( bus )、摩托车 ( motor ),其余信息如表 4 所示。

表 4 VisDrone2019 数据集详细信息

Table 4 VisDrone2019 dataset details

标签类别	标签样例	标签名称	标签数
行人		pedestrian	109 187
人 (非步行)		people	38 560
自行车		bicycle	13 069
私家车		car	187 005
货车		van	32 702
卡车		truck	16 284
三轮车		tricycle	6 387
遮阳三轮车		awning-tricycle	4 377
公交车		bus	9 117
摩托车		motor	40 378
总计			457 066

3.2 实验环境与评估指标

为了综合评价模型的性能,本文采用多种目标检测领域中被广泛使用的评估指标,包括精度 ( Precision )、召回率 ( Recall )、平均精度 ( average precision, AP )、平均精度均值 ( mean average precision, mAP )、参数量 ( Params )、千兆浮点运算 ( giga floating-point operations per second, GFlops )、每秒帧数 ( frames per second, FPS )。精度是指

所有预测为正类的样本中正确预测为正类的数量占比,召回率则是指所有实际为正类的样本中正确预测为正类的数量占比,平均精度则以精度与召回率曲线所占面积来衡量,平均精度均值则是由综合所有类别的平均精度得出,具体定义如式 (1) ~ (4) 所示。

$$Precision = \frac{TP}{TP + FP}$$
 (1)

$$Recall = \frac{TP}{TP + FN}$$
 (2)

$$AP = \int_0^1 P(R) dR$$
 (3)

$$mAP = \frac{1}{n} \sum_{i=1}^n AP_i$$
 (4)

其中,  $TP$  表示模型正确预测为正类的样本,  $FP$  表示模型错误地将负类预测为正类的样本,  $FN$  表示模型错误地将正类预测为负类的样本。  $P(R)$  表示精度与召回率的函数关系,  $AP_i$  为类别  $i$  的平均精度。

指标 Params 与 GFlops 则反应了模型的计算空间复杂度与时间复杂度,计算模型 FPS 时本文选取 Batch size 为 4,并将前处理、模型处理与后处理过程都计入推理时间,该指标用作评估模型推理速度的指标,这些指标对于模型的实际部署与运行效率有重要意义。

本文实验环境基于 Ubuntu22.04 操作系统,硬件条件:CPU 为 Xeon ( R ) Platinum 8352V, GPU 为 NVIDIA RTX 4090,模型构建基于 Python 3.10、Pytorch 2.1.0、CUDA 12.1 环境,输入图像尺寸为 640 × 640,使用 AdamW 优化器,初始学习率设置为 0.000 125, Batch size 设置为 4。详细信息参见表 5。

表 5 实验环境信息

Table 5 Experimental environment information

实验环境	详细配置	
操作系统	Ubuntu 22. 04	
编程环境	Python 3. 10	
深度学习框架	Pytorch 2. 1. 0+cu121	
GPU	NVIDIA RTX 4090	
CPU	Intel(R) Xeon(R) Platinum 8352V	
训练周期	DsPCBSD+	100
	PKU-Market-PCB	300
	PCB_WACV	300
	VisDrone2019	100
批量大小	4	
优化器	AdamW	
初始学习率	0. 000 125	
输入图像尺寸	640×640	

3.3 实验结果

为了验证改进后模型的性能,本文分别在 DsPCBSD+、PCB\_WACV 与 VisDrone2019 数据集上与其余 8 种检测模型进行了对比实验,并在 PCB\_WACV 数据集上进行了消融实验。8 种对比模型包括经典的二阶段检测模型 Faster R-CNN<sup>[39]</sup>与一阶段检测模型 Retinanet<sup>[40]</sup>,实时检测模型代表 YOLO 系列中的 YOLOv8、YOLOv10<sup>[41]</sup>、YOLOv11、YOLOv12<sup>[42]</sup>与 YOLOX<sup>[43]</sup>,以及 MMDetection 推出的实时检测模型 RTMDet<sup>[44]</sup>和基准模型 RT-DETR。后续实验中仅计算实时检测模型的 FPS,而 Faster R-CNN 与 Retinanet 并非实时检测模型,所以并未计算其 FPS。

1) DsPCBSD+对比实验

为了验证所提出模型在 PCB 缺陷检测任务中的性

能,本文在 PCB 缺陷数据集 DsPCBSD+上进行了对比实验,结果如表 6 所示。MFSF-DETR 模型以最低的参数量 19.86 M 与计算量 59.6 G 取得了最高的 mAP50 与 mAP50-95,对比基线模型 RT-DETR 高出了 3.1% 与 1.9%。主干网络中 Faster-CGLU Block 与邻层融合网络 RAWCFF 以更低的参数量与计算量实现了优异的特征提取与融合效果,使得 MFSF-DETR 模型在增加了移位融合网络 CFSF 后,仍然能保持相对最低的计算成本。相比最新的 YOLO 系列实时检测模型 YOLOv11m 与 YOLOv12m 有不小的优势,比表现出色的 YOLOXm 也高出了 1.1%,而在推理速度 FPS 上改进后 MFSF-DETR 模型的推理速度略有下降,但 120.2 fps 也超越了 YOLOv12m 与 RTMDet,接近 YOLOv10 与 YOLOv11 的推理速度,达到实时检测模型的要求。

表 6 DsPCBSD+对比实验结果  
Table 6 Comparative experimental results on the DsPCBSD+ dataset

模型	精度/%	召回率/%	mAP50/%	mAP50-95/%	参数量/M	GFlops/G	FPS <sub>bs=4</sub>
Faster R-CNN	79.7	74.9	81.6	46.0	41.39	208.0	—
Retinanet	80.5	78.9	82.4	48.3	36.52	210.0	—
YOLOv8m	79.7	80.4	82.8	50.2	25.85	78.7	<b>131.1</b>
YOLOv10b	82.3	78.1	82.2	49.8	20.46	98.7	121.4
YOLOv11m	80.1	<b>81.3</b>	83.6	51.5	20.04	67.7	123.9
YOLOv12m	80.7	79.1	83.3	50.9	20.11	67.2	97.1
YOLOXm	83.2	79.1	84.5	50.8	25.3	73.8	73.9
RTMDet	80.6	79.4	82.8	47.8	24.71	76.2	85.8
RT-DETR-R18	83.7	77.2	82.5	49.6	20.0	60.0	130.9
MFSF-DETR	<b>84.8</b>	81.2	<b>85.6</b>	<b>51.5</b>	<b>19.86</b>	<b>59.6</b>	120.2

本文进一步对 DsPCBSD+数据集的各类缺陷检测结果进行了分析,各类缺陷检测精度如表 7 所示。MFSF-DETR 仅在短路缺陷上稍低于基线模型 RT-DETR,其余类型缺陷的检测精度均有提升。MFSF-DETR 在尖刺、杂铜、断路、鼠咬、导体异物缺陷上实现了最高的精度,尤其在缺陷杂铜、断路、鼠咬、导体异物上与其余模型拉开较大的差距,而 Faster R-CNN、YOLOv11m、YOLOv12m、YOLOXm 与 RTMDet 也在单一类型的缺陷检测上展现出小幅优势。尖刺、杂铜、断路、鼠咬、导体异物几类缺陷常以微小的尺寸和明显的颜色变化出现在铜线上或附近,RAWCFF 网络自适应的特征融合方式使网络能够更好地保留并提炼细小目标的细节特征与位置信息,而 CFSF 网络进行非邻层特征直接交互的方式也促进了浅层细节信息与深层语义信息的传播与互补,使得 MFSF-DETR 能对目标进行更精确的定位识别。RT-DETR 模型与 MFSF-DETR 模型检测效果对比如图 10 所示。

2) PKU-Market-PCB 对比实验

为了进一步验证所提出模型在 PCB 缺陷检测任务中的性能,本文在 PCB 缺陷数据集 PKU-Market-PCB 上也进行了对比实验,实验结果如表 8 所示。MFSF-DETR 模型同样在 mAP50 与 mAP50-95 上取得了最高的 98.1% 与 55%,实现了最高精度,相比基线模型 RT-DETR 提高了 1.0% 与 1.9%。RAWCFF 网络与 CFSF 网络中对于多层级特征的充分融合,弥补了跨层信息的缺失,使各层特征信息更丰富,CSP PPC 模块带来的多尺度感受野使得模型对于目标整体的捕捉更为精确,得出的目标检测框更为贴近真实标签,从而取得了更高的 mAP50-95。在推理速度上也达到了 57.1,仅次于 YOLOv10b、YOLOXm 与 RT-DETR。满足实时检测的要求。

本文进一步对 PKU-Market-PCB 数据集的各类缺陷检测结果进行了分析,如表 9 所示。相比基线模型 RT-DETR,所提出的 MFSF-DETR 模型在各个类别的缺陷

表 7 DsPCBSD+对比实验各类缺陷检测结果

Table 7 Various defect detection results of the comparative experiment on the DsPCBSD+ dataset

(%)

模型	AP50								
	短路	尖刺	杂铜	断路	鼠咬	孔崩	导体划痕	导体异物	基材异物
Faster R-CNN	85.5	80.5	82.4	85.9	77.9	95.9	66.5	68.8	<b>90.8</b>
Retinanet	84.3	81.8	82.8	84.8	81.0	96.1	72.8	69.0	88.9
YOLOv8m	85.9	82.0	84.6	86.6	82.2	96.6	74.1	67.0	86.4
YOLOv10b	83.7	81.8	81.5	86.1	81.7	96.9	73.8	66.4	88.3
YOLOv11m	87.1	83.1	84.2	84.9	83.4	96.9	77.6	69.1	86.3
YOLOv12m	87.2	82.2	83.1	87.0	82.7	<b>97.0</b>	76.3	67.4	87.0
YOLOXm	87.3	<b>83.4</b>	84.3	90.4	83.0	94.8	77.1	74.2	85.7
RTMDet	<b>88.0</b>	81.7	80.4	86.8	81.7	96.3	75.8	68.2	86.3
RT-DETR-R18	87.5	82.3	85.0	90.1	78.3	96.3	65.9	69.7	87.1
MFSF-DETR	86.8	<b>83.4</b>	<b>87.3</b>	<b>91.2</b>	<b>86.1</b>	96.9	74.3	<b>75.9</b>	88.5

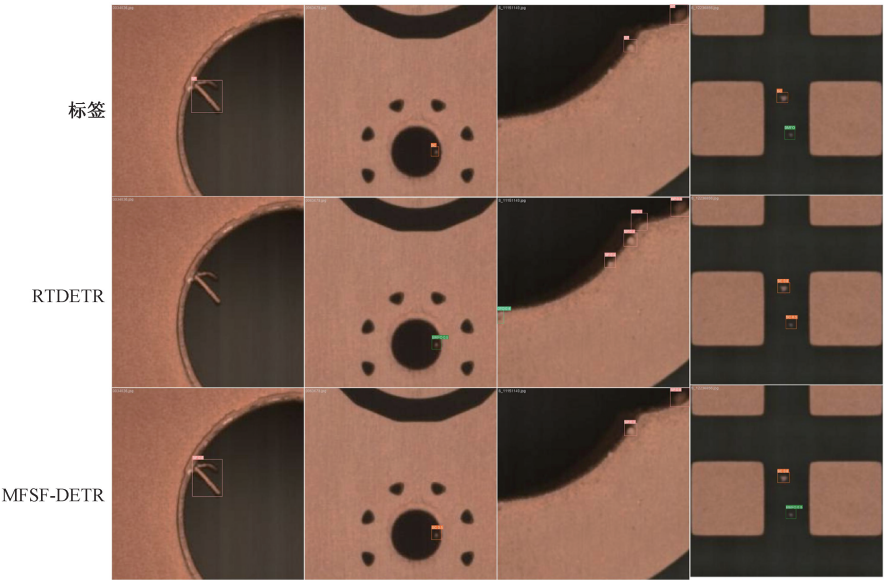


图 10 RT-DETR 与 MFSF-DETR 检测结果对比

Fig. 10 Comparison of RT-DETR and MFSF-DETR detection results

表 8 PKU-Market-PCB 对比实验结果

Table 8 Comparative experimental results on the PKU-Market-PCB dataset

模型	精度/%	召回率/%	mAP50/%	mAP50-95/%	参数量/M	GFlops/G	FPS <sub>bs=4</sub>
Faster R-CNN	85.3	84.4	89.2	45.5	41.39	208.0	—
Retinanet	84.4	83.5	89.4	45.0	36.52	210.0	—
YOLOv8m	<b>97.2</b>	95.1	96.6	52.9	25.85	78.7	54.5
YOLOv10b	94.5	94.0	96.8	52.3	20.46	98.7	<b>73.4</b>
YOLOv11m	96.9	96.4	97.2	53.5	20.04	67.7	55.6
YOLOv12m	95.0	96.3	96.2	51.4	20.11	67.2	53.6
YOLOXm	95.7	96.6	97.3	51.5	25.30	73.8	61.5
RTMDet	96.1	94.1	96.7	51.2	24.71	76.2	50.9
RT-DETR-R18	94.3	97.7	97.1	53.1	20.00	60.0	65.6
MFSF-DETR	97.0	<b>98.3</b>	<b>98.1</b>	<b>55.0</b>	<b>19.86</b>	<b>59.6</b>	57.1

表 9 PKU-Market-PCB 对比实验各类缺陷检测结果

Table 9 Various defect detection results of the comparative experiment on the PKU-Market-PCB dataset (%)

模型	mAP50					
	缺孔	鼠咬	断路	短路	毛刺	残铜
Faster R-CNN	0.995	0.931	0.926	0.831	0.812	0.859
Retinanet	0.995	0.869	0.765	0.935	0.866	0.843
YOLOv8m	0.995	<b>0.987</b>	0.966	0.945	0.962	0.941
YOLOv10b	0.994	0.970	0.960	0.965	0.969	0.949
YOLOv11m	0.995	0.986	0.966	0.950	0.977	0.956
YOLOv12m	0.995	0.974	0.918	<b>0.975</b>	0.966	0.946
YOLOXm	0.995	0.981	0.959	0.958	0.976	0.971
RTMDet	0.995	<b>0.987</b>	0.955	0.969	0.967	0.930
RT-DETR-R18	0.995	0.970	0.960	0.961	0.968	0.974
MFSF-DETR	<b>0.995</b>	<b>0.987</b>	<b>0.967</b>	0.967	<b>0.995</b>	<b>0.978</b>

上都实现了检测精度的提升。同时,MFSF-DETR 模型在 5 种缺陷类型取得了最高的检测精度,仅在短路类型的缺陷上落后于模型 YOLOv12m,在残铜与毛刺上领先其他模型较多,这两类缺陷特征与图像背景相似,易与背景信息混淆,

Faster-CGLU Block 中门控注意力机制与 RAWCFF 网络中基于相邻层特征引导的空间和通道注意力机制使得模型聚焦于目标缺陷本身,利于提取目标的细节信息。RT-DETR 模型与 MFSF-DETR 模型检测效果对比如图 11 所示。

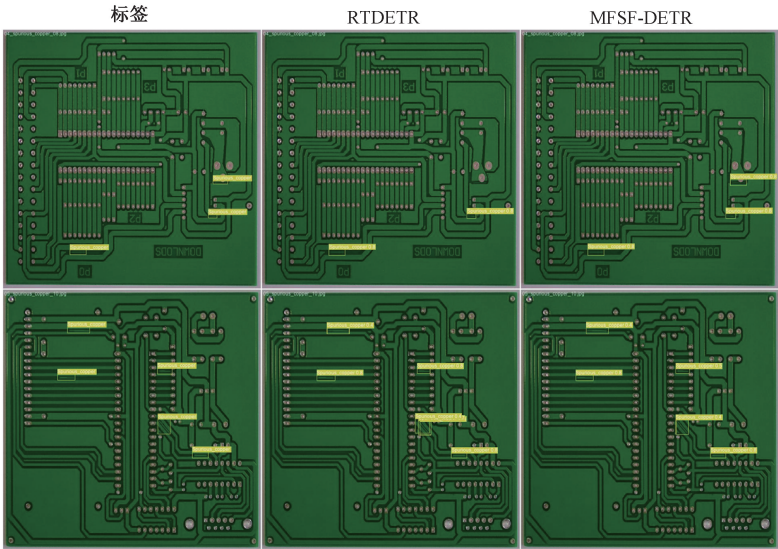


图 11 RT-DETR 与 MFSF-DETR 检测结果对比

Fig. 11 Comparison of RT-DETR and MFSF-DETR detection results

3) PCB\_WACV 对比实验

为检验 PCB 元件检测效果,本文在数据集 PCB\_WACV 上与经典模型和主流实时检测模型进行了对比实验,实验结果如表 10 所示。MFSF-DETR 模型同样以最低参数量与计算量实现了最优的精度表现,达到了 89.9%的 mAP50 与 72.5%的 mAP50-95,超越基线模型 3.8%与 3.7%。在数据量偏小的 PCB\_WACV 数据集上 YOLO 系列模型均表现出不错的检测精度,对比基线模

型 RT-DETR 有着一定优势。但在推理速度上,精度表现优异的 YOLOv8m、YOLOv11m 与 YOLOv12m 却低于精度稍差的 YOLOv10b,而 RT-DETR 达到了最高的 81.3 fps,所提出的 MFSF-DETR 模型也达到了 71.8 fps,仅次于 RT-DETR 与 YOLOv10b。由于 PCB\_WACV 数据集中部分图片中有着密集分布的元件,存在大量待检测的密集排布的目标,这使得 YOLO 系列模型为这样场景中的目标生成了过多的备选目标框,使得后处理过程中 NMS 机

表 10 PCB\_WACV 对比实验结果

Table 10 Comparative experimental results on the PCB\_WACV dataset

模型	精度/%	召回率/%	mAP50/%	mAP50-95/%	参数量/M	GFlops/G	FPS <sub>bs=4</sub>
Faster R-CNN	84.6	79.5	85.5	64.4	41.39	208.0	—
Retinanet	83.3	78.8	84.1	65.9	36.52	210.0	—
YOLOv8m	86.9	79.1	88.3	70.7	25.85	78.7	67.3
YOLOv10b	84.7	80.1	85.9	68.8	20.46	98.7	78.3
YOLOv11m	<b>87.7</b>	80.2	87.5	69.5	20.04	67.7	70.8
YOLOv12m	87.3	80.7	88.3	70.5	20.11	67.2	57.1
YOLOXm	79.7	81.7	84.7	62.5	25.30	73.8	68.3
RTMDet	79.8	84.1	86.3	67.0	24.71	76.2	70.9
RT-DETR-R18	79.2	84.6	86.1	68.7	20.00	60.0	<b>81.3</b>
MFSF-DETR	86.7	<b>87.2</b>	<b>89.9</b>	<b>72.4</b>	<b>19.86</b>	<b>59.6</b>	71.8

制需要消除大量的备选框从而保留最优的检测结果,造成了模型推理的延迟,而 MFSF-DETR 则回避了 NMS 机制,使得在推理速度上实现了反超。

本文进一步对 PCB\_WACV 数据集的各类元件检测结果进行了分析,各类元件检测精度如表 11 所示。MFSF-DETR 在基线模型 RT-DETR 基础上,整体精度都得到了提升,尤其是在焊盘检测上实现了较大的提升,同时在接线端子、焊盘、电容、芯片、电解电容上取得了最高的检测精度,YOLO 系列中较出色的 YOLOv12 m 则在管脚、测试点、发光二极管的检测上有着小幅优势,但其在

接线端子上的检测精度仅高于 Retinanet 的 75.6,与其余模型有着较大差距,且模型的推理速度也是实时检测模型中最慢的。数据集中,焊盘、电容、电解电容通常是出现在密集的元素场景中的小型目标,而接线端子、芯片则有着最大的类内尺寸与形态差异,得益于 CSP PPC 的渐进式的多尺度感受野,中心特征的集中提取使得 MFSF-DETR 模型在密集场景中也能更好地聚焦于小目标,同时也对于目标的尺寸与形态变化有着更好的适应能力。RT-DETR 模型与 MFSF-DETR 模型检测效果对比如图 12 所示。

表 11 PCB\_WACV 对比实验各类元件检测结果

Table 11 Various component detection results of the comparative experiment on the PCB\_WACV dataset (%)

模型	AP50								
	接线端子	电阻	焊盘	电容	管脚	测试点	发光二极管	芯片	电解电容
Faster R-CNN	77.1	97.4	65.8	97.4	68.7	88.6	97.6	83.3	93.8
Retinanet	75.6	94.2	63.9	95.1	67.2	88.5	96.8	83.5	92.1
YOLOv8m	81.5	97.4	70.4	97.1	76.1	89.2	98.5	87.2	97.2
YOLOv10b	80.2	96.7	63.6	97.3	71.0	87.7	95.5	84.9	96.2
YOLOv11m	80.6	97.3	69.9	96.8	70.1	88.9	96.7	88.6	<b>98.4</b>
YOLOv12m	76.9	98.0	70.6	95.3	<b>82.5</b>	<b>91.1</b>	<b>98.6</b>	85.7	96.0
YOLOXm	67.5	97.6	68.6	98.1	73.7	82.2	94.7	81.4	98.2
RTMDet	81.7	<b>98.7</b>	68.9	97.2	69.3	86.1	91.7	84.6	<b>98.4</b>
RT-DETR-R18	84.9	94.6	58.2	95.2	77.8	86.6	94.1	86.6	96.7
MFSF-DETR	<b>88.8</b>	97.1	<b>72.9</b>	<b>97.5</b>	80.4	89.7	95.6	<b>88.6</b>	<b>98.4</b>

4) 消融实验

PCB\_WACV 数据集上的消融实验结果如表 12 所示。本文对 MFSF-DETR 模型进行了全面的模块消融实验,消融实验中将对应的改进模块单独替换为基准模型

RT-DETR 中的原始模块,w/o FCB 表示 Faster-CGLU Block 部分替换为 Resnet18 中的 Block,w/o ETB 表示 ETB 替换为 AIFI,w/o RAWCFF 表示 RAWCFF 替换为 CCFF,w/o CSP PPC 表示将 CSP PPC 中的 PConv 替换普

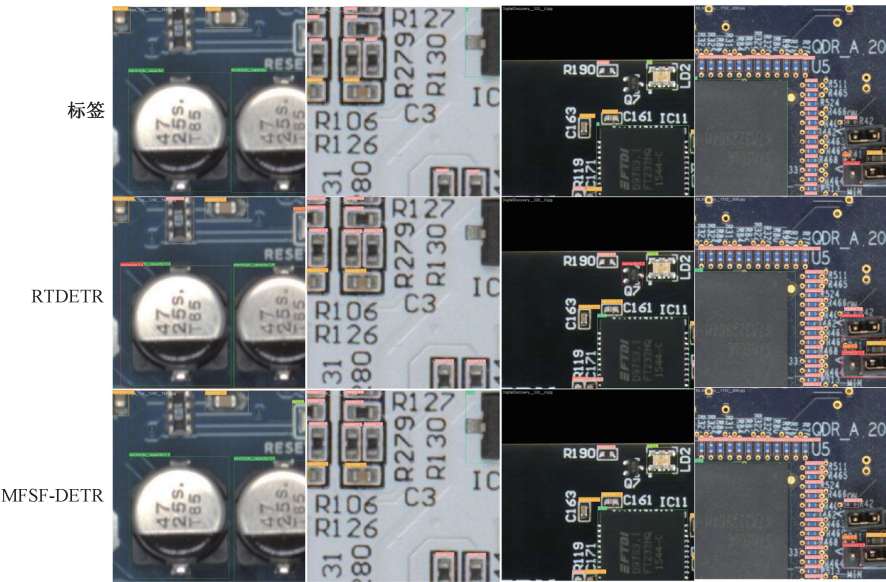


图 12 RT-DETR 与 MFSF-DETR 检测结果对比

Fig. 12 Comparison of RT-DETR and MFSF-DETR detection results

表 12 各部分消融实验结果

Table 12 Results of ablation experiment

模型	$mAP_{50}/\%$	$mAP_{50-95}/\%$	参数量/M	$GFlops/G$	$FPS_{bs=4}$
RT-DETR	86.1	68.7	20.00	60.0	81.3
MFSF-DETR	89.9	72.5	19.86	59.6	71.8
MFSF-DETR w/o FCB	88.7	70.8	20.34	63.1	73.7
MFSF-DETR w/o ETB	88.5	71.8	19.74	59.3	70.6
MFSF-DETR w/o RAWCFF	88.8	71.5	21.30	63.0	71.1
MFSF-DETR w/o CSN	87.7	70.5	19.09	57.1	71.6
MFSF-DETR w/o CSP PPC	88.0	70.9	20.31	63.2	75.6
MFSF-DETR w/o CFSF	87.4	69.2	18.12	50.6	80.5

通的扩张卷积, w/o CSN 与 w/o CFSF 则表示单独省略 CSN 与 CFSF 网络。

单独替换掉 Faster-CGLU Block 与 RAWCFF 后, 模型的精度均有小幅度的下降, 而模型的参数量与计算量都有明显的增加, 从原本的 19.86 M、59.6 G 分别增加到了 20.34 M、63.1 G 与 21.3 M、63 G, Faster-CGLU Block 中细粒度通道注意力 CGLU 的引入对模型的  $mAP_{50-95}$  有较大的提升, 使得模型对于目标的定位更精确, RAWCFF 网络通过相邻层自适应加权融合简化特征融合路径, 保证精度的同时减少了参数量与计算量。以 AIFI 模块替换掉 ETB 模块后, 模型的参数量与计算量几乎不变, 但精度损失较大, AIFI 中的空间域信息对于提取深层语义并不充足, 而 ETB 中补充的频域信息对于深层语义信息的丰富起到了重要作用。

省略 CFSF 网络后, 模型的精度损失最为严重, 证明

非邻层级间直接的特征交互与递进式的多尺度感受野能够大大提高密集小目标的检测性能, 但该部分对于模型的推理速度影响达到了 10 fps, 是各个模块中最高的。进一步将 CFSF 网络拆解为 CSN 与 CSP PPC 两部分后进行消融实验, 单独省略 CSN 与将 CSP PPC 中的 PConv 换为扩张卷积对于精度都有较大的影响, CSN 模块以较低的计算量与参数量实现了高效的非邻层间特征融合, 建立浅层细节与深层语义的直接联系, CSP PPC 中的分组卷积与条形卷积操作使得其能以相比普通的扩张卷积更低的参数量与计算量实现感受野的扩大, 着重提取中心特征的方式也使网络对于不同尺度的目标有更好的适应性。

5) VisDrone2019 对比实验

为了测试 MFSF-DETR 模型的泛化能力, 本文在 VisDrone2019 上也进行了对比实验, VisDrone2019 数据

集中存在大量的小目标以及检测对象密集分布的场景,与 PCB 场景中的缺陷与元件检测所面临的挑战一致,VisDrone2019 数据集上的对比实验结果如表 13 所示。在总体精度 AP 与 AP50 上,本文所提出的 MFSF-DETR 模型达到了 21.3%与 37.5%,比基线模型 RT-DETR 高出 1.6%与 2.7%,与 YOLO 中表现最出色的 YOLOXm 相比,MFSF-DETR 也以相似的参数量与更低的计算量实现了更高的精度。在代表小物体检测精度的 APs 上,基线模型 RT-DETR 本就具有一定优势,而 MFSF-DETR 进一步提升了小目标的检测效果,达到最高的 11.9%,而在中

型物体的检测上,改进后的 MFSF-DETR 模型以 31.4%的 APm 超越了 YOLOv11m 的 31.2%,展现了在中小物体检测方面的优势。大型物体的检测方面,YOLO 系列模型与基线模型 RT-DETR 表现都较为一般,而经典的二阶段检测模型 Faster R-CNN 与一阶段检测模型 Retinanet 的表现则较为优秀,RTMDet 在 API 上达到了最高的 44.8%,MFSF-DETR 在基线 41.3%的基础将 API 提升到了 42.1%,缩小了在大型物体检测精度上的差距。MFSF-DETR 模型在 VisDrone2019 数据集上的检测效果如图 13 所示。

表 13 VisDrone2019 对比实验结果

Table 13 Comparative experimental results on the VisDrone2019 dataset

模型	输入图像尺寸	主干网络	GFlops/G	参数量/M	AP/%	AP50/%	APs/%	APm/%	API/%
Faster R-CNN	(768,1 344)	R50	208.0	41.39	19.4	32.9	9.5	30.9	42.9
Retinanet	(768,1 344)	R50	210.0	36.52	16.4	27.6	6.0	27.4	42.7
YOLOv8m	(640,640)	-	78.7	25.85	19.0	33.2	9.0	29.4	41.7
YOLOv10b	(640,640)	-	98.7	20.46	19.5	34.5	9.7	30.0	41.4
YOLOv11m	(640,640)	-	67.7	20.04	20.3	35.0	9.8	31.2	41.3
YOLOv12m	(640,640)	-	67.2	20.11	19.2	33.6	9.4	29.8	38.4
YOLOXm	(640,640)	CSPNeXt	73.8	25.30	19.8	36.2	11.2	30.8	41.3
RTMDet	(640,640)	CSPNeXt	76.2	24.71	18.4	33.5	8.3	30.9	<b>44.8</b>
RT-DETR	(640,640)	R18	60.0	20.00	19.7	34.8	11.3	30.5	41.3
MFSF-DETR	(640,640)	-	<b>59.6</b>	<b>19.86</b>	<b>21.3</b>	<b>37.5</b>	<b>11.9</b>	<b>31.4</b>	42.1



图 13 MFSF-DETR 检测结果

Fig. 13 MFSF-DETR detection results

4 结 论

针对 PCB 组件与缺陷检测提出了一种基于 Transformer 的实时目标检测模型 MFSF-DETR,在 RT-DETR 网络架构基础上,使用 Faster-CGLU Block 替换原有的 Block 层,降低计算量与参数量的同时,通过细粒度通道注意力机制提取细节信息,引入了 ETB 模块,通过融合空间与频域的双域特征信息提炼深层语义,设计了自适应加权特征融合网络 RAWCFF 简化特征融合路径,

增强邻层特征融合效果,利用 CFSF 网络建立非邻层特征的直接交互,加强细节信息与语义信息在特征间的流动,优化特征表示,并通过递进的多尺度感受野提取中心特征,实现目标的精确定位。实验证明,所提出的 MFSF-DETR 检测模型提高了对微小目标的精确定位能力,增强了对复杂背景与目标密集分布的场景的适应能力,面对种类繁多与类内差异性强的目标时也更具鲁棒性,同时兼具较高的推理速度与优秀的泛化能力。  
未来的工作将进一步探索模型的架构,在加入更多种类缺陷与元件的场景下,提升模型的检测效果,尝试知

识蒸馏与模型剪枝技术,以进一步提高模型的推理速度,减小模型的规模,降低实际部署的成本,提升检测效率。同时,目前PCB场景下的目标检测数据集主要分为缺陷检测数据集与元件检测数据集,缺乏综合两者的PCB检测数据集,后续的研究重点将落在构建全面的PCB检测数据集并在其基础上探究基于深度学习模型的检测技术。

## 参考文献

- [1] YU J B, ZHAO L X, WANG Y SH, et al. Defect detection of printed circuit board based on adaptive key-points localization network[J]. Computers & Industrial Engineering, 2024, 193: 110258.
- [2] STARODUBOV D, DANISHVAR S, ABU EBAYYEH A A R M, et al. Advancements in PCB components recognition using WaferCaps: A data fusion and deep learning approach[J]. Electronics, 2024, 13(10): 1863.
- [3] LEI L, LI H X, YANG H D. Reliable and lightweight adaptive convolution network for PCB surface defect detection[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-8.
- [4] CHEN CH L, WU Q M, ZHANG J, et al. U<sup>2</sup>D<sup>2</sup>PCB: Uncertainty-aware unsupervised defect detection on PCB images using reconstructive and discriminative models[J]. IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-10.
- [5] BAYGIN M, KARAKOSE M, SARIMADEN A, et al. Machine vision based defect detection approach using image processing[C]. Proceedings of the International Artificial Intelligence and Data Processing Symposium, 2017: 1-5.
- [6] LIU ZH CH, QU B D. Machine vision based online detection of PCB defect[J]. Microprocessors and Microsystems, 2021, 82: 103807.
- [7] MELNYK R A, TUSHNYTSKYY R B. Detection of defects in printed circuit boards by clustering the etalon and defected samples[C]. 15th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering, 2020: 961-964.
- [8] ONSHAUNJIT J, SRINONCHAT J. Algorithmic scheme for concurrent detection and classification of printed circuit board defects[J]. Computers, Materials & Continua, 2022, 71(1): 355-367.
- [9] LI W, JIANG CH, BREIER M, et al. Localizing components on printed circuit boards using 2D information[C]. Proceedings of the International Conference on Industrial Technology, 2016: 769-774.
- [10] CHIGATERI M K, MANJUVANI K M, MANJUNATH K M, et al. The detection of electrical and electronics components using K nearest neighbour (KNN) classification algorithm[J]. International Research Journal of Engineering and Technology, 2016, 3: 169-175.
- [11] GHOSH S, MOSTAFIZ M T, GURUDU S R, et al. PCB component detection for hardware assurance: A feature selection-based approach[C]. Proceedings of the International Symposium on Hardware Oriented Security and Trust, 2022: 109-112.
- [12] 吴一全, 赵朗月, 苑玉彬, 等. 基于机器视觉的PCB缺陷检测算法研究现状及展望[J]. 仪器仪表学报, 2022, 43(8): 1-17.  
WU Y Q, ZHAO L Y, YUAN Y B, et al. Research status and the prospect of PCB defect detection algorithm based on machine vision[J]. Chinese Journal of Scientific Instrument, 2022, 43(8): 1-17.
- [13] ZHANG L X, CHEN J SH, CHEN J G, et al. LDD-Net: Lightweight printed circuit board defect detection network fusing multi-scale features[J]. Engineering Applications of Artificial Intelligence, 2024, 129: 107628.
- [14] LAN ZH, HONG Y, LI Y. An improved YOLOv3 method for PCB surface defect detection[C]. Proceedings of the International Conference on Power Electronics, Computer Applications, 2021: 1009-1015.
- [15] XIAO G SH, HOU SH L, ZHOU H Y. PCB defect detection algorithm based on CDI-YOLO[J]. Scientific Reports, 2024, 14(1): 7351.
- [16] RAN G Z, LEI X, LI D SH, et al. Research on PCB defect detection using deep convolutional neural network[C]. Proceedings of the 5th International Conference on Mechanical, Control and Computer Engineering, 2020: 1310-1314.
- [17] 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.  
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.
- [18] 张莹, 邓华宣, 王耀南, 等. 基于多通道特征融合学习的印制电路板小目标缺陷检测[J]. 仪器仪表学报, 2024, 45(5): 10-19.  
ZHANG Y, DENG H X, WANG Y N, et al. Small defects detection of PCB based on multi-channel feature

- fusion learning [J]. Chinese Journal of Scientific Instrument, 2024, 45(5): 10-19.
- [19] 王军, 伍毅, 陈正超. 基于 SMT-YOLOv8 的 PCB 缺陷检测研究[J]. 电子测量技术, 2024, 47(11): 131-137.
- WANG J, WU Y, CHEN ZH CH. Research on PCB defect detection based on SMT-YOLOv8[J]. Electronic Measurement Technology, 2024, 47(11): 131-137.
- [20] 王天洋, 刘路, 王太勇, 等. 基于改进 YOLOv8s 的轻量级 PCB 缺陷检测算法[J]. 电子测量与仪器学报, 2025, 39(3): 44-52.
- WANG T Y, LIU L, WANG T Y, et al. Lightweight PCB defect detection algorithm based on improved YOLOv8s[J]. Journal of Electronic Measurement and Instrumentation, 2025, 39(3): 44-52.
- [21] 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: 1-11.
- [22] LING Q, ISA N A M, ASAARI M S M. Precise detection for dense PCB components based on modified YOLOv8[J]. IEEE Access, 2023, 11: 116545-116560.
- [23] CHI T S, AB WAHAB M N, MOHAMED A S A, et al. Enhancing EfficientNet-YOLOv4 for integrated circuit detection on printed circuit board (PCB) [J]. IEEE Access, 2024, 12: 25066-25078.
- [24] LIU X, HU J SH, WANG H X, et al. Gaussian-IoU loss: Better learning for bounding box regression on PCB component detection[J]. Expert Systems with Applications, 2022, 190: 116178.
- [25] LUO SH Y, WAN F, LEI G B, et al. EC-YOLO: Improved YOLOv7 model for PCB electronic component detection[J]. Sensors, 2024, 24(13): 4363.
- [26] LI J, LI W Y, CHEN Y Q, et al. A PCB electronic components detection network design based on effective receptive field size and anchor size matching [J]. Computational Intelligence and Neuroscience, 2021, 2021(1): 6682710.
- [27] MALLAIYAN SATHIASEELAN M A, PARADIS O P, TAHERI S, et al. Why is deep learning challenging for printed circuit board (PCB) component recognition and how can we address it? [J]. Cryptography, 2021, 5(1): 9.
- [28] CARION N, MASSA F, SYNNAEVE G, et al. End-to-end object detection with transformers[C]. Proceedings of the European Conference on Computer Vision, 2020: 213-229.
- [29] ZHAO Y AN, 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, 2024: 16965-16974.
- [30] CHEN J R, KAO S H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [31] SHI D. TransNeXt: Robust foveal visual perception for vision transformers[C]. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024: 17773-17783.
- [32] SUN Y G, XU CH Y, YANG J, et al. Frequency-spatial entanglement learning for camouflaged object detection[C]. 2024 European Conference on Computer Vision, 2024: 343-360.
- [33] ZONG ZH F, CAO Q G, LENG B. RCNet: Reverse feature pyramid and cross-scale shift network for object detection[C]. 29th ACM International Conference on Multimedia, 2021: 5637-5645.
- [34] YANG J N, LIU SH L, WU J J, et al. Pinwheel-shaped convolution and scale-based dynamic loss for infrared small target detection[C]. 2025 AAAI Conference on Artificial Intelligence, 2025, 39(9): 9202-9210.
- [35] LYU SH P, OUYANG B, DENG ZH H, et al. A dataset for deep learning based detection of printed circuit board surface defect[J]. Scientific Data, 2024, 11(1): 811.
- [36] HUANG W B, WEI P. A PCB dataset for defects detection and classification[J]. ArXiv preprint arXiv: 1901.08204, 2019.
- [37] KUO C W, ASHMORE J D, HUGGINS D, et al. Data-efficient graph embedding learning for PCB component detection [C]. 2019 IEEE Winter Conference on Applications of Computer Vision, 2019: 551-560.
- [38] DU D W, ZHU P F, WEN L Y, et al. VisDrone-DET2019: The vision meets drone object detection in image challenge results [C]. 2019 IEEE/CVF International Conference on Computer Vision Workshops, 2019: 213-226.
- [39] REN SH Q, HE K M, GIRSHICK R, et al. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 39(6): 1137-1149.
- [40] LIN T Y, GOYAL P, GIRSHICK R, et al. Focal loss for

dense object detection [ C ]. 2017 IEEE International Conference on Computer Vision, 2017: 2980-2988.

[ 41 ] WANG AO, CHEN H, LIU L H, et al. YOLOv10: Real-time end-to-end object detection[ J ]. Advances in Neural Information Processing Systems, 2024, 37: 107984-108011.

[ 42 ] TIAN Y J, YE Q X, DOERMANN D. YOLOv12: Attention-centric real-time object detectors [ J ]. ArXiv preprint arXiv:2502.12524, 2025.

[ 43 ] GE ZH, LIU S T, WANG F, et al. YOLOX: Exceeding YOLO series in 2021 [ J ]. ArXiv preprint arXiv: 2107.08430, 2021.

[ 44 ] LYU CH Q, ZHANG W W, HUANG H AN, et al. RTMDet: An empirical study of designing real-time object detectors[ J ]. ArXiv preprint arXiv:2212.07784, 2022.

作者简介



张永宏(通信作者),1998 年于景德镇陶瓷学院获得学士学位,2002 年于南京林业大学获得硕士学位,2005 年于上海交通大学获得博士学位,现为无锡学院教授,主要研究方向为遥感大数据分析 with 深度学习。

E-mail:yhzhang@cw Xu.edu.cn

**Zhang Yonghong**(Corresponding author) received his B. Sc. degree from Jingdezhen Ceramic University in 1998, his M. Sc. degree from Nanjing Forestry University in 2002, and his Ph. D. degree from Shanghai Jiao Tong University in 2005. He is currently a professor at Wuxi University. His main research interests include remote sensing, big data analysis, and deep learning.



许鑫豪,2023 年于南京工程学院获得学士学位,现为南京信息工程大学硕士研究生,主要研究方向为计算机视觉与深度学习。

E-mail:xinhao\_xu@nuist.edu.cn

**Xu Xinhao** received his B. Sc. degree from Nanjing Institute of Technology in 2023. Now he is a master's student at Nanjing University of Information Science and Technology. His main research interests include computer vision and deep learning.



尹贺峰,2011 年于许昌学院获得学士学位,2020 年于江南大学获得博士学位,现为无锡学院讲师,主要研究方向为图像处理与计算机视觉。

E-mail:yinhefeng@cw Xu.edu.cn

**Yin Hefeng** received his B. Sc. degree from Xuchang University in 2011, and his Ph. D. degree from Jiangnan University in 2020. He is currently a lecturer at Wuxi University. His main research interests include image processing and computer vision.



李子奇,2012 年于扬州大学获得学士学位,2020 年于江南大学获得博士学位,现为无锡学院副教授,主要研究方向为模式识别与人工智能。

E-mail:ziqu\_li@cw Xu.edu.cn

**Li Ziqi** received his B. Sc. degree from Yangzhou University in 2012, and his Ph. D. degree from Jiangnan University in 2020. He is currently an associate professor at Wuxi University. His main research interests include pattern recognition and artificial intelligence.