

基于多尺度上下文增强的轻量级工业缺陷检测网络^{*}

李 帆 周 迅 张 岩 顾海明

(青岛科技大学数理学院 青岛 266061)

摘 要: 针对工业缺陷目标小、边界模糊以及背景复杂等问题,提出了一种基于多尺度上下文增强的轻量级工业缺陷检测网络 MSCENet。首先,提出多尺度注意力增强特征提取(MAFE)模块,通过多尺度空洞卷积并行提取不同感受野的缺陷特征。其次,设计残差增强融合(REF)模块,利用双注意力特征增强模块与残差连接自适应融合多级特征,提升解码器的缺陷边界与细节重建质量。此外,提出全局注意力聚合(GAA)模块,聚焦缺陷区域并抑制背景干扰,进一步提升检测精度与鲁棒性。在 3 类工业缺陷数据集上的实验结果表明,相比主干网络 FasterNet-T1,所提方法在参数量仅为 9.418 M 的情况下平均交并比(IoU)、平均像素精度、准确率均有显著提升,同时提高了工业检测效率。

关键词: 缺陷检测;多尺度注意力特征提取;残差增强融合;双注意力特征增强;全局注意力聚合

中图分类号: TP391.4; TN99 **文献标识码:** A **国家标准学科分类代码:** 520.2060

Lightweight industrial defect detection network using multi-scale context enhancement

Li Fan Zhou Xun Zhang Yan Gu Haiming

(School of Mathematics and Physics, Qingdao University of Science and Technology, Qingdao 266061, China)

Abstract: Aiming at the challenges of small industrial defect targets, blurred boundaries and complex backgrounds, this paper proposes a lightweight industrial defect detection network named MSCENet based on multi-scale context enhancement. First, a multi-scale attention-enhanced feature extraction module (MAFE) is introduced, which employs parallel multi-scale dilated convolutions to capture defect features under varying receptive fields. Next, a residual enhancement fusion (REF) module is designed to adaptively integrate multi-level features using a dual-attention feature enhancement mechanism alongside residual connections, thereby improving the reconstruction quality of defect boundaries and details in the decoder. Furthermore, a global attention aggregation (GAA) module is proposed to focus on defect regions while suppressing background interference, further enhancing detection accuracy and robustness. Experimental results on three industrial defect datasets demonstrate that, compared to the backbone network FasterNet-T1, the proposed method achieves significant improvements in mean intersection over union (IoU), mean pixel accuracy, and overall accuracy, with only 9.418 million parameters, while improving industrial inspection efficiency.

Keywords: defect detection; multi scale attention enhanced feature extraction; enhanced residual fusion; dual attention feature enhancement; global attention aggregation

0 引 言

工业缺陷检测是保障产品质量的关键环节,其检测精度与效率直接影响生产效益,高效精准的工业表面缺陷检测具有重要意义。传统机器视觉方法主要依赖人工设计的特征进行分析,在金属表面缺陷检测等领域应用广泛^[1]。传统机器视觉方法具有较强的可解释性、较低的计算资源

需求,在小样本或特定场景下常表现出稳定性能。然而,这类方法也存在鲁棒性较差、泛化能力不足以及难以自动学习复杂特征等明显局限性。

近年来,深度学习方法因其良好的泛化和复杂特征学习能力在很大程度上克服了传统方法的局限,并在工业缺陷检测领域表现突出。早期基于深度学习的方法多采用图像分类方法。例如,Zheng 等^[2]利用稀疏编码代替卷积神

经网络的卷积层,进行无监督训练,实现对轮胎缺陷的分类。该模型结构相对简单、标注成本低,但对局部小缺陷敏感度低、易漏检和受背景干扰。为解决上述问题,研究人员提出基于目标检测的方法,实现在图像中识别并定位特定对象。赵伯亭等^[3]提出一种基于 YOLOv7 的钢材表面缺陷检测模型 ECC-YOLO,该方法可以提高钢材表面缺陷检测效率和精度。吴葛等^[4]提出一种改进 YOLO11 的印刷电路板 (printed circuit board, PCB) 表面缺陷检测方法 HDH-YOLO,可用于 PCB 表面的缺陷检测。虽然此方法能够定位缺陷、区分不同实例,但在处理密集小目标时仍易出现重叠与特征提取困难的问题。为实现缺陷的像素级定位与形状描述,研究者提出基于语义分割方法。语义分割的本质是传统图像分割与目标识别融合,以各像素点分类的方式把图像分割成具有特定语义类别的区域^[5]。Shelhamer 等^[6]提出全卷积网络 (fully convolutional networks, FCN) 将图像分类扩展至像素级分类。Chen 等^[7]在 DeepLabV3 基础上提出 DeepLabV3+,引入解码器优化边界分割效果。语义分割方法能够精确刻画缺陷边缘与内部结构,对复杂形状或边界模糊缺陷具有较强的描述能力。但该方法仍存在模型复杂度高、对极小结构分割准确性有限等问题^[8]。

本文提出一种基于多尺度上下文增强的轻量级工业缺陷检测网络 MSCENet。该模型使用 FasterNet-T1^[9] 作为主干网络,设计并引入多尺度注意力增强特征提取 (multi-scale attention-enhanced feature extraction, MAFE)、残差增强融合 (residual enhancement fusion, REF) 和全局注意力聚合 (global attention aggregation, GAA) 这 3 个核心模块,旨在增强特征表达能力、提升小目标检测效果与整体分割精度。

1 相关工作

1.1 特征提取

特征提取是人工智能和计算机视觉中的关键环节。特征提取模块通常基于卷积神经网络自下而上地生成图像的多层级特征图,实现对微小、低对比、弱边缘且形态多变缺陷的检测与分类。在缺陷检测领域,空洞空间金字塔池化 (atrous spatial pyramid pooling, ASPP) 模块是一种广泛采用的特征提取结构,能够有效捕获多尺度上下文信息。但该模块在处理对边缘细节等高频特征上表现不足。针对该问题,Xie 等^[10]提出一种多尺度上下文提取模块,通过级联 4 个混合膨胀卷积分支以捕获更广更深层次的特征。然而,该方式未能充分建模不同尺度特征之间的交互关系,在复杂背景下仍存在语义信息丢失问题。

针对现有方法在局部关联性弱、多尺度覆盖不足等局限导致的边缘定位不准与小目标漏检等问题,本文提出 MAFE 模块。该模块在特征提取阶段实现了感受野多样性、特征质量、计算效率之间的良好平衡,能够在复杂工业缺陷检测任务中表现出更强的特征表征能力和检测性能。

1.2 注意力机制

随着深度神经网络在视觉任务中的广泛应用,注意力机制已成为增强特征表达能力的关键技术之一。为增强模型对通道特征的敏感度,Hu 等^[11]提出挤压激励网络 (squeeze-and-excitation networks, SENet),通过压缩-激励结构建模通道间的依赖关系,但其降维操作可能会导致特征表达能力受限。针对该问题,Wang 等^[12]提出高效的通道注意力 (efficient channel attention, ECA) 模块,通过引入一维卷积避免维度缩减,但难以适应缺陷在空间分布上的多样性。为进一步整合空间与通道注意力,Woo 等^[13]设计了卷积注意力模块 (convolutional block attention module, CBAM),将通道和空间注意力机制进行结合。但该机制在处理小特征缺陷时易导致局部语义丢失。

为缓解上述方法在复杂缺陷检测环境中存在的注意力推断不精确、细节特征易被抑制等问题,本文提出 GAA 模块。该模块采用并行协同的注意力生成机制与残差连接,在增强对真实缺陷区域聚焦能力的同时,保留细微结构与弱缺陷特征,避免了过度抑制或误抑制的问题。

1.3 注意力融合

近年来,注意力融合机制研究受到广泛关注。Huang 等^[14]提出一种结合注意力机制和多尺度特征融合网络的时空融合方法,以获得更高质量时空融合图像。但该方法过度强化结构信息,对细节缺陷特征的保持能力有限。Zhou 等^[15]探索了一种由空间多轴信息融合、独立分类和空间信息融合以及多尺度特征融合组成的独立特征信息融合模型,用于缺陷检测任务,但该模型在稳健性、推理速度及其在复杂的实际缺陷检测场景中的性能方面的能力还有待加强。Shao 等^[16]设计了 BEIC-DeepLabv3+ 模型,引入类间特征注意力模块 (interclass feature attention module, I-CFAM) 和边界增强模块 (boundary enhancement module, BEM),以缓解绝缘子缺陷检测中的边界模糊和误分类问题。但该方法存在易导致过拟合、计算速率难以满足实时检测需求等问题。

为解决上述方法在细节保持、泛化能力与计算效率方面的局限,本文将双注意力特征增强 (dual attention feature enhancement, DAFE) 模块分别嵌入 MAFE 与 REF 模块中,实现对多尺度特征的层级化校准与强化。DAFE 模块并非仅作用于单一层级,而是贯穿于整个特征传播过程,实现多层次注意力协同作用机制,能有效增强对低对比度、微小缺陷及复杂背景的检测能力。

2 本文方法

2.1 模块优势对比

ASPP 是一种增强特征提取结构,包含不同空洞率的卷积和全局平均池化,在捕捉上下文信息方面有优势,但仍存在一些结构局限性。例如,ASPP 模块参数量大,计算复杂度高,不利于工业实时部署,而且分支结构简单,特征表

征能力较弱。与 ASPP 相比,MAFE 模块通过引入深度可分离卷积的双重设计($3 \times 3 + 5 \times 5$ 卷积组合),在每个分支内部实现了更丰富的尺度变化,构建了感受野更为连续的多尺度特征金字塔。同时,在每个分支末端集成 DAFE,实现自适应注意力的重新校准,能够显著增强缺陷相关特征通道的响应强度。并且,MAFE 使用深度可分离卷积代替标准卷积,在降低计算开销的同时保持了特征提取能力。通过这种设计,MAFE 能够更精确地捕获从局部细节(像素级裂纹)到全局结构(大面积缺陷)的多尺度缺陷特征,提升缺陷区域激活强度,在保持模型性能的同时实现模型的轻量化。

ECA 是一种轻量级的注意力机制模块,在不增加显著计算开销的情况下,用于强调重要特征通道。CBAM 也是一种注意力机制模块,它通过通道、空间两种独立的注意力机制自适应地细化特征图。两种模块均主要用于卷积神经网络中,通过各自的方式提升模型的整体性能。但 ECA 只考虑了通道间的注意力,忽略了空间位置上的重要性,而空间注意力的缺失会导致模型无法聚焦于这些区域,从而影

响分割和检测的精度。CBAM 先应用通道注意力,将通道注意力调整后的特征再输入空间注意力,这使得空间注意力是在通道维度聚合后的特征上进行计算的,可能会损失一些信息。与 ECA 只考虑通道维度、CBAM 按顺序操作不同,GAA 模块将通道和空间注意力以相乘的方式并行融合,更有利于同时捕捉通道和空间上的重要特征,保留更多有用信息。同时,残差连接的设计使得网络更容易训练,确保训练稳定性和特征复用。此外,适中的压缩比平衡了计算效率和特征保持,大感受野空间注意力也更适合缺陷的定位需求。这些特性使得 GAA 模块在工业缺陷检测这种需要精确位置信息和丰富特征表达的任务中具有明显优势。

针对工业表面缺陷检测任务高精度、资源受限工业部署需求与复杂检测场景之间的矛盾,提出一种基于多尺度上下文增强的轻量级工业缺陷检测网络 MSCENet。如图 1 所示,网络采用编码器-解码器结构,选用 FasterNet-T1 为主干网络,在维持较强表征能力的同时具备优异的计算效率。

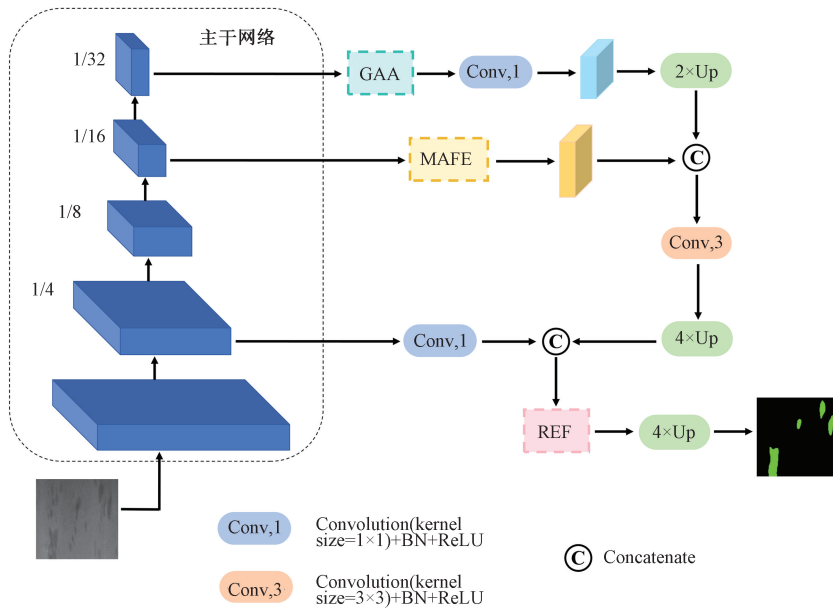


图 1 MSCENet 整体结构

Fig. 1 The overall architecture of MSCENet

2.2 MAFE 模块

在工业缺陷检测任务中,缺陷尺寸差异显著。为解决多尺度缺陷检测挑战,本文提出 MAFE 模块,通过引入多分支结构并相应配置空洞率序列,实现对从局部细节到全局上下文的多尺度信息的高效捕获。此外,通过在模块末端嵌入 DAFE 子模块,进一步强化缺陷相关特征的响应,有效增强模型对微小缺陷的敏感性。DAFE 模块结构如图 2 所示,其通过通道-空间协同注意力机制与轻量化特征增强操作,实现对缺陷区域的精准聚焦。

具体地,给定输入特征图 $\mathbf{X}$, 首先进行全局平均池化

压缩通道维度,获取全局通道信息;随后进行深度可分离卷积操作,采用自适应动态卷积核 k 用于跨通道的交互,并经 Sigmoid 激活函数生成通道注意力权重:

$$\mathbf{A}_c = \sigma(\text{DSConv}_{k \times 1}(\text{GAP}(\mathbf{X}))) \quad (1)$$

其中,GAP 为全局平均池化, σ 为 Sigmoid 激活函数,DSConv $_{k \times 1}$ 为动态卷积核是 $k \times 1$ 的深度可分离卷积。将通道注意力权重 $\mathbf{A}_c$ 与输入特征 $\mathbf{X}$ 逐元素相乘,得到通道注意力加权后的特征:

$$\mathbf{X}_c = \mathbf{X} \circ \mathbf{A}_c \quad (2)$$

其中, $\circ$ 为逐通道相乘。对通道维度加权后的特征进

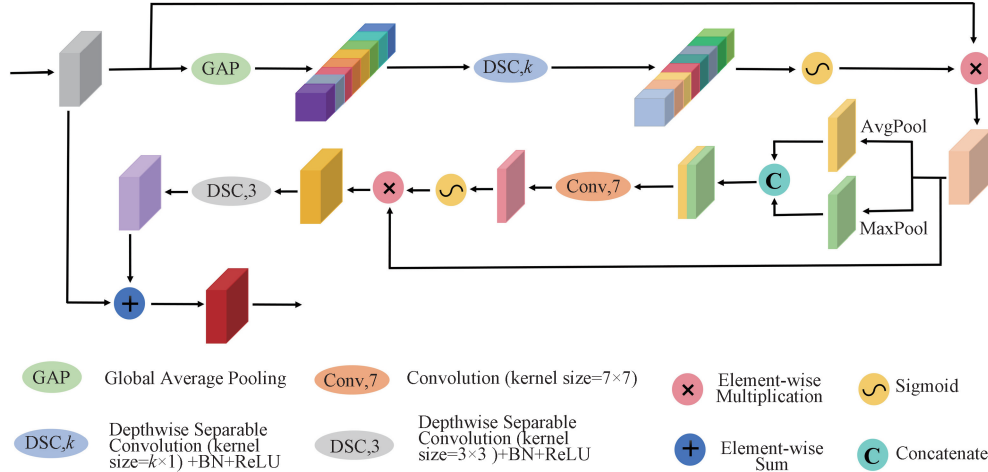


图 2 DAFE 模块结构

Fig. 2 The architecture of the DAFE module

行双池化特征提取,即分别进行平均池化和最大池化,分别得到两个空间特征图,即:

$$\mathbf{M}_{\text{avg}} = \text{Avg}(\mathbf{X}_c) \quad (3)$$

$$\mathbf{M}_{\text{max}} = \text{Max}(\mathbf{X}_c) \quad (4)$$

其中,Avg 为平均池化,Max 为最大池化。拼接这两个特征图形成一个双通道的特征图,再通过一个 7×7 卷积层处理,经 Sigmoid 函数激活生成空间注意力权重:

$$\mathbf{A}_s = \sigma(\text{Conv}_{7 \times 7}(\text{Con}(\mathbf{M}_{\text{avg}}, \mathbf{M}_{\text{max}}))) \quad (5)$$

其中, σ 为 Sigmoid 激活函数,Conv $_{7 \times 7}$ 为卷积核是 7×7 的卷积操作,Con 为拼接。将空间注意力权重与通

道分支输出的特征逐位置相乘,再采用深度可分离卷积将加权后的特征输入特征增强部分进行特征增强。最后,通过进行残差连接与输入 $\mathbf{X}$ 相加得到最终输出,即:

$$\mathbf{X}_{\text{out}} = \text{DSConv}_{3 \times 3}(\mathbf{X}_c \circ \mathbf{A}_s) + \mathbf{X} \quad (6)$$

其中, $\circ$ 为逐位置相乘,DSConv $_{3 \times 3}$ 为卷积核是 3×3 的深度可分离卷积, $+$ 为逐元素相加。

如图 3 所示,MAFE 采用多分支并行结构,通过各分支不同感受野的卷积操作捕获特征,并在每个分支末端集成注意力机制。

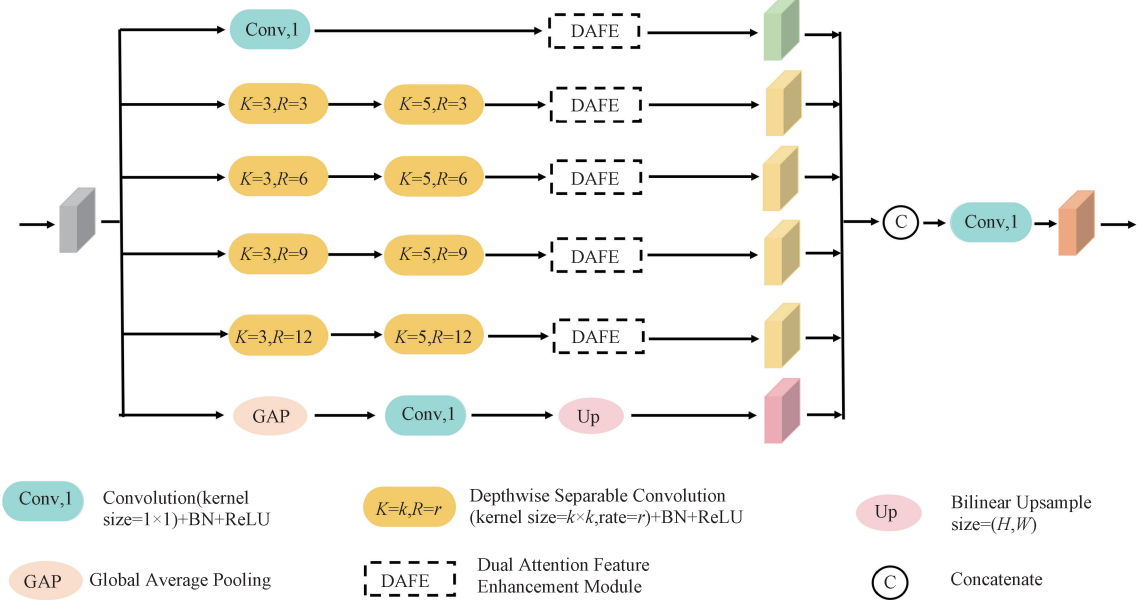


图 3 MAFE 模块结构

Fig. 3 The architecture of the MAFE module

MAFE 模块采用 6 分支并行结构,以充分提取多尺度特征并增强缺陷相关表示。第 1 条分支中输入的是来自

编码器的次高级特征 $\mathbf{X}_{\text{sub-senior}}$,通过 1×1 卷积捕获原始特征的局部信息,同时调整通道数:

$$\mathbf{Y}_1 = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(\mathbf{X}_{\text{sub-senior}}))) \quad (7)$$

其中, $\text{Conv}_{1 \times 1}$ 为卷积核是 1×1 的卷积操作, BN 为批量归一化层, ReLU 为激活函数。随后, 将 $\mathbf{Y}_1$ 输入 DAFE 模块, 增强关键通道响应。先对输入的特征进行加权增强, 定义通道权重 $\mathbf{N}_c$ 和空间权重 $\mathbf{N}_s$:

$$\mathbf{N}_c = \sigma(\text{DSConv}_{k \times 1}(\text{GAP}(\mathbf{Y}_1))) \quad (8)$$

$$\mathbf{N}_s = \sigma(\text{Conv}_{7 \times 7}(\text{Con}(\text{Avg}(\mathbf{Y}_1), \text{Max}(\mathbf{Y}_1)))) \quad (9)$$

其中, GAP 为全局平均池化, σ 为 Sigmoid 激活函数, $\text{DSConv}_{k \times 1}$ 为动态卷积核是 $k \times 1$ 的深度可分离卷积, Avg 为平均池化, Max 为最大池化, Con 为拼接, $\text{Conv}_{7 \times 7}$ 为卷积核是 7×7 的卷积。在经过注意力分支得到加权后的特征之后再采用深度可分离卷积将加权后的特征输入特征增强部分进行特征增强, 最后与输入的 $\mathbf{Y}_1$ 相加:

$$\mathbf{Z}_1 = \text{DSConv}_{3 \times 3}(\mathbf{Y}_1 \circ \mathbf{N}_c \circ \mathbf{N}_s) + \mathbf{Y}_1 \quad (10)$$

其中, $\circ$ 表示逐元素乘法, $\text{DSConv}_{3 \times 3}$ 为卷积核是 3×3 的深度可分离卷积, $+$ 为逐元素相加。分支 2~5 采用不同空洞率的深度可分离卷积结构, 用于在扩展感受野的同时控制计算复杂度, 特征先后进入两个不同卷积进行提取。每条分支的空洞率不同, 分别设为 $\mathbf{R}_i = \{3, 6, 9, 12\}$ ($i = \{2, 3, 4, 5\}$):

$$\mathbf{F}_i = \text{PW}_{1 \times 1}(\text{DW}_{3 \times 3}(\mathbf{X}, \mathbf{R}_i)) \quad (11)$$

$$\mathbf{Y}_i = \text{PW}_{1 \times 1}(\text{DW}_{5 \times 5}(\mathbf{F}_i, \mathbf{R}_i)) \quad (12)$$

其中, $\text{DW}_{3 \times 3}$ 为卷积核是 3×3 的逐深度卷积, $\text{DW}_{5 \times 5}$ 为卷积核是 5×5 的逐深度卷积, $\text{PW}_{1 \times 1}$ 为卷积核是 1×1 的逐点卷积。然后, 在每个分支的末端集成 DAFE 模块进行特征优化。先对输入的特征进行加权增强, 定义通道权重 $\mathbf{Q}_c$ 和空间权重 $\mathbf{Q}_s$:

$$\mathbf{Q}_c = \sigma(\text{DSConv}_{k \times 1}(\text{GAP}(\mathbf{Y}_i))) \quad (13)$$

$$\mathbf{Q}_s = \sigma(\text{Conv}_{7 \times 7}(\text{Con}(\text{Avg}(\mathbf{Y}_i), \text{Max}(\mathbf{Y}_i)))) \quad (14)$$

随后经过注意力分支得到使用重加权后的特征, 再将其进行特征加强:

$$\mathbf{Z}_i = \text{DSConv}_{3 \times 3}(\mathbf{Y}_i \circ \mathbf{Q}_c \circ \mathbf{Q}_s) + \mathbf{Y}_i \quad (15)$$

第 6 条分支通过全局平均池化获取图像级全局特征, 并使用 1×1 卷积进行特征变换:

$$\mathbf{Y}_6 = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(\text{GAP}(\mathbf{X}_{\text{sub-senior}})))) \quad (16)$$

其中, GAP 为全局平均池化, $\text{Conv}_{1 \times 1}$ 为卷积核是 1×1 的卷积操作。

为恢复空间维度并保持与其它分支尺度一致, 对 $\mathbf{Y}_6$ 进行双线性上采样恢复空间维度, 在提供全局上下文信息的同时增强模型语义理解能力, 为多尺度特征融合提供全局视角:

$$\mathbf{Z}_6 = \text{Upsample}(\mathbf{Y}_6, \text{size} = (H, W)) \quad (17)$$

其中, Upsample 为双线性插值上采样。

最后, 将 6 个分支的输出特征沿通道维度拼接, 并通过 1×1 卷积进行特征压缩与融合, 在批量归一化和激活函数完成后最终得到输出 $\mathbf{Z}$:

$$\mathbf{Z}_{\text{cat}} = \text{Con}(\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3, \mathbf{Z}_4, \mathbf{Z}_5, \mathbf{Z}_6) \quad (18)$$

$$\mathbf{Z} = \text{ReLU}(\text{BN}(\text{Conv}_{1 \times 1}(\mathbf{Z}_{\text{cat}}))) \quad (19)$$

其中, Con 为拼接, $\text{Conv}_{1 \times 1}$ 为卷积核 1×1 的卷积操作。

本文提出的 MAFE 模块通过构建 6 分支并行与异构卷积设计, 有效整合多尺度空洞卷积、深度可分离卷积与 DAFE 模块。该结构设计兼顾了感受野多样性、计算复杂度和特征表达能力, 为工业视觉检测系统提供了强大的特征提取基础。

2.3 REF 模块

如图 4 所示, 本文 REF 模块由 3 部分组成: 特征增强层、残差卷积层和正则化层。在特征增强层, 采用 3×3 卷积核提取局部空间特征并保持分辨率, 引入 DAFE 模块以自适应增强缺陷相关通道响应。残差卷积部分采用双重

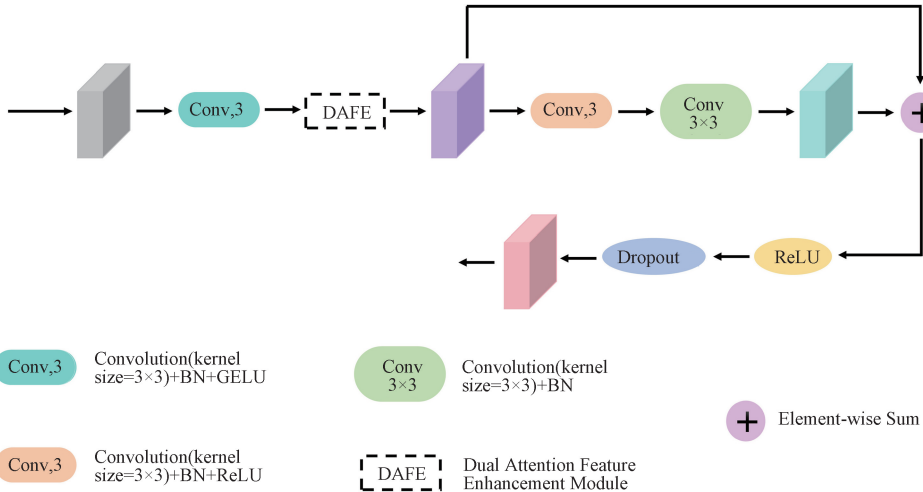


图 4 REF 模块结构

Fig. 4 The architecture of the REF module

卷积和残差连接进行,在增强网络的非线性表达能力的同时确保梯度直接传播、缓解梯度消失,而且保留了原始特征信息。正则化层通过 Dropout 机制随机丢弃部分神经元输出,打破神经元间的协同适应性,从而提升模型泛化能力。通过这 3 部分有机整合,REF 模块在保持空间分辨率的前提下实现了特征表达能力、训练稳定性与泛化性能的均衡。

具体地,如图 4 所示,对于输入的融合特征 $\mathbf{X}_{\text{fusion}}$,首先采用 3×3 卷积进行特征变换与通道调整,再经批归一化稳定训练过程:

$$\mathbf{H}_1 = \text{GELU}(\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{X}_{\text{fusion}}))) \quad (20)$$

其中, $\text{Conv}_{3 \times 3}$ 为卷积核是 3×3 的卷积操作, BN 为批量归一化层, GELU 为激活函数。其次,将 $\mathbf{H}_1$ 输入 DAFE 模块以增强相关特征通道并抑制噪声。先对输入的特征进行加权增强,定义通道权重 $\mathbf{P}_c$ 和空间权重 $\mathbf{P}_s$:

$$\mathbf{P}_c = \sigma(\text{DSConv}_{k \times 1}(\text{GAP}(\mathbf{H}_1))) \quad (21)$$

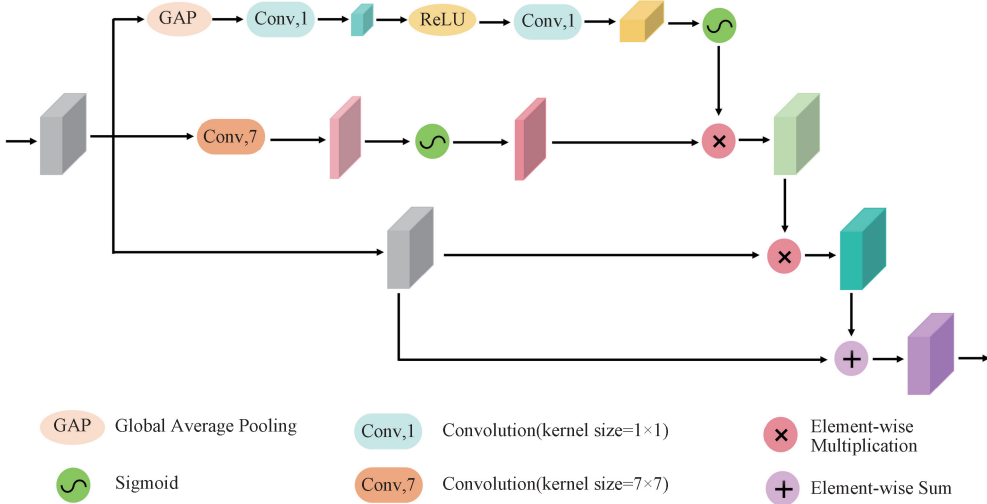


图 5 GAA 模块结构

Fig. 5 The architecture of the GAA module

在通道注意力分支,对来自编码器的高级特征 $\mathbf{X}_{\text{high}}$ 首先通过全局平均池化压缩空间维度获取通道级全局信息。其次,使用 1×1 卷积将通道数降至原始数量的 $1/4$,并经 ReLU 激活函数引入非线性:

$$\mathbf{L}_1 = \text{ReLU}(\text{Conv}_{1 \times 1}(\text{GAP}(\mathbf{X}_{\text{high}}))) \quad (26)$$

其中, GAP 为全局平均池化, $\text{Conv}_{1 \times 1}$ 为卷积核是 1×1 的卷积操作, ReLU 为激活函数。

再通过 1×1 卷积恢复原始通道维度,并经 Sigmoid 函数生成通道注意力权重,实现通道级的软注意力选择:

$$\mathbf{B}_c = \sigma(\text{Conv}_{1 \times 1}(\mathbf{L}_1)) \quad (27)$$

其中, σ 为 Sigmoid 激活函数, $\text{Conv}_{1 \times 1}$ 为卷积核是 1×1 的卷积操作。

在空间注意力分支,采用较大尺寸卷积核 (7×7) 对输入特征 $\mathbf{X}_{\text{high}}$ 进行卷积操作,以捕获像素间的长程空间依赖

$$\mathbf{P}_s = \sigma(\text{Conv}_{7 \times 7}(\text{Con}(\text{Avg}(\mathbf{H}_1), \text{Max}(\mathbf{H}_1)))) \quad (22)$$

再经过注意力分支加权和特征增强:

$$\mathbf{F}_1 = \text{DSConv}_{3 \times 3}(\mathbf{H}_1 \circ \mathbf{P}_c \circ \mathbf{P}_s) + \mathbf{H}_1 \quad (23)$$

随后通过两个 3×3 卷积层构成深层特征提取路径:

$$\mathbf{F}_2 = \text{BN}(\text{Conv}_{3 \times 3}(\text{ReLU}(\text{BN}(\text{Conv}_{3 \times 3}(\mathbf{F}_1)))) \quad (24)$$

其中, $\text{Conv}_{3 \times 3}$ 为卷积核为 3×3 的卷积操作。然后再与原始特征进行元素级相加,经过 ReLU 激活之后再行 Dropout 正则化,避免网络对特定训练样本的过拟合:

$$\mathbf{F}_3 = \text{Dropout}(\text{ReLU}(\mathbf{F}_1 + \mathbf{F}_2)) \quad (25)$$

2.4 GAA 模块

如图 5 所示,本文提出了 GAA 模块。该模块通过并行处理通道与空间两个注意力分支,实现对关键缺陷特征的协同增强与选择,显著提升模型在复杂工业场景中的特征判别能力。残差结构的设计在保障模型训练稳定性的同时,缓解潜在特征丢失问题。

关系并建立不同空间位置间的上下文关联,再通过 Sigmoid 函数生成空间维度上的注意力权重图,产生空间域的注意力分布:

$$\mathbf{B}_s = \sigma(\text{Conv}_{7 \times 7}(\mathbf{X}_{\text{high}})) \quad (28)$$

其中, σ 为 Sigmoid 激活函数, $\text{Conv}_{7 \times 7}$ 为卷积核是 7×7 的卷积操作。

将两条分支生成的注意力权重与原始输入进行元素级加权,最后将加权后的特征与 $\mathbf{X}_{\text{high}}$ 进行相加,确保原始特征信息不丢失、缓解梯度消失问题:

$$\mathbf{X}_{\text{final}} = \mathbf{X}_{\text{high}} \circ \mathbf{B}_c \circ \mathbf{B}_s + \mathbf{X}_{\text{high}} \quad (29)$$

其中, $\circ$ 表示逐元素乘法, $+$ 为逐元素相加。

该模块中,通道、空间注意力二者协同作用从两个维度优化特征表达,显著提升复杂工业场景缺陷特征的区分度,降低误检率。残差结构在保障模型训练稳定性的同

时,缓解了潜在特征丢失问题,确保网络能够有效传递未被注意力机制充分响应的缺陷特征,提升网络训练的收敛速度。

3 实验与分析

3.1 实验设置

实验在 Inter Xeon Platinum 8260,32 G Nvidia Tesla, Python3.7,Pytorch1.7.1,Cuda11.0,Cudnn 环境中运行。采用 Adam 优化器与交叉熵损失函数对模型进行训练。输入图像分辨率设为 256×256 ,初始学习率设为 5×10^{-4} ,权重衰减系数为 1×10^{-4} 。采用余弦退火衰减策略,批量大小设为 2,训练迭代 200 个 epoch。

3.2 数据集

为全面验证模型在复杂工业场景下的泛化能力,选用 3 类典型工业缺陷数据集评估模型的有效性:SD-saliency-900 数据集、磁砖(MT)缺陷数据集和轮胎缺陷图像数据集。

SD-saliency-900 数据集^[17]是基准钢材缺陷数据集,共有 900 幅图像,包含夹杂物、划痕和斑块这 3 类典型缺陷。每类缺陷包含 300 张图像及其对应像素级标签。数据集中随机选取 720 幅作为训练样本、180 幅用于测试。

磁砖(MT)缺陷数据集^[18]包含 1 344 幅图像,涵盖 115 幅气孔、57 幅裂纹、32 幅磨损、85 幅断裂以及 103 幅凹凸不平共 5 种缺陷。为提高数据多样性并缓解过拟合,采用颜色变换(亮度、对比度、饱和度和清晰度)和随机角度旋转(90° 、 180° 和 270°)进行数据增强。从增强后的 3 530 个样本中随机选取 3 140 和 390 幅用于训练和测试。

轮胎缺陷图像数据集包含胎面异物 245 幅、胎侧异物

337 幅、胎侧帘线重叠 132 幅、胎面帘线裂纹 218 幅、胎侧帘线裂纹 331 幅和气泡缺陷 187 幅,共 1 450 幅缺陷图像及其像素级标注,经数据增强处理后共计 6 800 幅,其中随机选取 5 800 幅用于训练,1 000 幅用于测试。

3.3 评价指标

为全面评估模型性能,本文采用以下指标:平均交并比(mIoU)、平均像素精度(mPA)、准确率(Accuracy)、检测时间(Time)和参数量(Params)。各项指标定义为:

$$mIoU = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FP_i + FN_i}$$

$$mPA = \frac{1}{k} \sum_{i=1}^k \frac{TP_i}{TP_i + FN_i}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

其中, k 表示类别总数, TP_i 表示类别 i 的真阳性, FP_i 表示类别 i 的假阳性, FN_i 表示类别 i 的假阴性。TP 表示所有类别中预测正确的正样本像素数,TN 表示所有类别中预测正确的负样本像素数,FP 表示所有类别中预测错误的正样本像素数,FN 表示所有类别中预测错误的负样本像素数。

3.4 消融实验与统计显著性分析

为验证本文 MAFE、REF、GAA 模块及其核心组件 DAFE 模块的有效性,在 SD-saliency-900 数据集上基于 FasterNet-T1 主干网络进行了系统的消融实验,如表 1 所示。实验从基线模型开始,依次引入各模块,并特别对 MAFE 与 REF 模块中 DAFE 组件的作用进行了单独分析,以全面评估各组件对模型性能的贡献。

表 1 消融实验结果
Table 1 Ablation experiment results

方法	MAFE-DAFE	MAFE	REF-DAFE	REF	GAA	mIoU/%
主干网络						65.49
主干网络+MAFE		✓				78.35
主干网络+MAFE-DAFE	✓					77.10
主干网络+REF				✓		78.61
主干网络+REF-DAFE			✓			76.62
主干网络+MAFE+REF		✓		✓		79.57
主干网络+MAFE+REF+GAA		✓		✓	✓	80.62

实验结果表明,各模块均对模型性能提升有显著贡献。仅使用主干网络时,mIoU 为 65.49%,表明其基础特征提取能力有限。引入 MAFE 模块后,mIoU 显著提升至 78.35%(+12.86%),证明其多尺度注意力机制有效增强了对不同尺度的缺陷特征提取与融合能力。而当移除 MAFE 中的 DAFE 组件后(MAFE-DAFE),性能降至 77.10%,虽仍显著优于基线,但相较于完整 MAFE 模块下降了 1.25%,说明 DAFE 通过其通道与空间双重注意力机制,在 MAFE 框架内进一步优化了多尺度特征的权重分配

与选择性增强,对提升特征表征质量起到了关键作用。单独引入 REF 模块后,mIoU 达到 78.61%(+13.12%),该模块通过残差连接与 DAFE 的有效融合,在减少信息丢失的同时强化了语义细节的保留,从而提升了缺陷边界的识别精度。与之对比,当 REF 模块中未包含 DAFE 组件时(REF-DAFE),mIoU 为 76.62%,性能出现明显下降(降幅达 1.99%)。这一结果凸显了 DAFE 在 REF 模块中的重要性,DAFE 增强关键特征的响应并抑制冗余信息,使得融合过程更加高效。在 MAFE 与 REF 模块协同工作的基

础上(主干网络 + MAFE + REF), mIoU 进一步提升至 79.57%,体现了模块间的互补性。最终引入 GAA 模块, mIoU 达到 80.62%,该模块借助全局上下文建模机制强化了对长程依赖关系的捕捉能力,使网络更加聚焦于缺陷区域并抑制背景干扰,全面增强了模型的判别性能。

综上,MAFE、REF 和 GAA 模块分别从多尺度特征提取、特征融合优化和全局上下文建模 3 个不同方面显著提升了缺陷检测的性能,共同构建了高效而高精度的 MSCENet 架构,证明了其有效性。

另外,为了判断模型性能提升是否具有统计意义,本研究在磁砖(MT)缺陷数据集上通过 10 次重复实验并报告平均值与标准差,对模型性能进行了严格的统计评估。结果如表 2 所示, MSCENet 模型展现了较为稳定的性能,其核心分割指标分别达到了 mIoU: $72.38\% \pm 0.31\%$, mPA: $84.26\% \pm 0.31\%$, Accuracy: $97.77\% \pm 0.14\%$ 。而且所有关键指标的标准差均保持在较低水平,这充分证明了该模型具有良好的鲁棒性。同时,模型保持了高效的推理速度与较低的参数量,实现了精度与效率的良好平衡。

表 2 磁砖(MT)缺陷数据集上 MSCENet 的 10 次独立运行结果
Table 2 Results of ten independent runs of MSCENet on magnetic tile (MT) defect dataset

运行次数	mIoU/%	mPA/%	Accuracy/%	Time/s	Params/M
1	72.35	84.45	97.85	0.009	9.418
2	72.75	84.51	97.89	0.009	9.418
3	72.23	84.23	97.81	0.010	9.418
4	72.09	83.89	97.59	0.009	9.418
5	72.65	84.48	97.87	0.010	9.418
6	72.51	84.42	97.84	0.009	9.418
7	71.95	83.82	97.60	0.009	9.418
8	72.42	84.31	97.78	0.010	9.418
9	72.81	84.63	97.89	0.009	9.418
10	72.01	83.82	97.54	0.009	9.418
平均值±标准差	72.38 ± 0.31	84.26 ± 0.31	97.77 ± 0.14	0.0093 ± 0.0004	9.418

综上所述,结合消融实验与统计显著性分析,有力地证明了 MSCENet 中各模块的有效性以及模型整体的优越性和鲁棒性。

3.5 对比实验

为验证本文提出的 MSCENet 模型在工业缺陷检测任务中的性能,分别在 3 个具有代表性的数据集上进行对比实验。对比模型包括 U-Net^[19]、PSPNet^[20]、DeepLabv3+、DANet^[21]、DDRnet-39^[22]、BiSeNetV2^[23]、

FDSNet^[24]和 FFNet^[25]、LEDNet^[26]、PIDNet^[27]、HRNet^[28]、STDC^[29]。为保证实验结果,所有对比实验均在相同实验环境进行。

1) SD-saliency-900 对比实验

在 SD-saliency-900 数据集上的对比实验结果如表 3 所示,在分割精度方面, MSCENet 取得了 80.62%的 mIoU、87.23%的 mPA 以及 97.17%的 Accuracy。除了 PIDNet, MSCENet 都优于 U-Net 等其他先进算法。

表 3 SD-saliency-900 数据集上各方法的对比实验结果
Table 3 Comparative experiments results of various methods on the SD-Saliency-900 dataset

方法	mIoU/%	mPA/%	Accuracy/%	Time/s	Params/M
U-Net	77.98	83.65	96.86	0.010	18.272
DANet	73.41	81.78	96.11	0.007	7.084
DeepLabv3+	77.02	85.19	96.75	0.008	11.092
PSPNet	57.98	64.36	94.27	0.007	7.108
DDRNet-39	76.61	84.82	96.52	0.012	32.357
FDSNet	78.49	85.26	96.89	0.007	0.970
BiSeNetV2	68.10	76.13	95.46	0.010	5.100
FFNet	77.83	84.09	96.83	0.010	9.657
LEDNet	72.81	80.74	95.96	0.015	0.897
STDC	67.13	75.28	95.25	0.005	8.570
PIDNet	81.11	87.86	97.20	0.010	7.624
HRNet	79.89	84.75	97.13	0.038	29.541
MSCENet	80.62	87.23	97.17	0.009	9.418

虽然 MSCENet 的 mIoU 结果比 PIDNet 低 0.49%，但这一差距并未全面反映模型在不同缺陷类别上的细微表现。进一步根据表 4 对 3 类缺陷进行逐类分析发现，MSCENet 在斑块与划痕这两类关键缺陷上的交并比 IoU 均优于 PIDNet，表明 MSCENet 对于形态不规则、边界模糊及较大尺寸跨度的缺陷具有更精准的定位与分割能力。同时由表 5、6 可以看出 MSCENet 的各项指标均全面超越 PIDNet，有力验证了其网络架构在面对不同工业场景、不

同缺陷特性时具备更优的适应性与鲁棒性。在效率方面，MSCENet 的单幅图像处理时间为 0.009 s，其推理速度与 BiSeNetV2、FFNet 等轻量模型相当，满足工业检测中对实时性的要求。此外，MSCENet 的参数量为 9.418 M，处于中等水平，远低于 U-Net 和 DDRNet-39，说明模型在未显著增加计算负担前提下实现了精度提升。综合来看，MSCENet 在分割精度、推理效率与模型复杂度之间实现了较优的平衡，表现出显著的综合性能。

表 4 PIDNet 与 MSCENet 的 IoU、召回率和精确率的定量结果比较

Table 4 Comparison of quantitative results of IoU, recall, and precision between PIDNet and MSCENet %

缺陷类别	PIDNet			MSCENet		
	IoU ₁	IoU ₂	Recall ₁	Recall ₂	Precision ₁	Precision ₂
背景	96.92	96.90	98.63	98.69	98.25	98.16
夹杂物	65.38	61.13	73.08	67.97	86.13	85.33
斑块	81.99	82.76	90.25	90.08	89.96	91.06
划痕	80.15	81.69	89.47	92.16	88.49	87.79

为更直观展示 MSCENet 的分割性能，图 6 给出了在 SD-saliency-900 数据集上的可视化分割结果对比。从前两行所示夹杂物缺陷检测结果可以看出，DeepLabv3+、DANet 和 BiSeNetV2 网络在边缘处都存在模糊或误判现象；U-Net 与 DDRNet-39 虽可基本定位缺陷区域，但在细微结构与轮廓准确度方面仍有不足。在第 3、4 行中，DANet、BiSeNetV2 检测结果均存在漏检和断裂现象，

DDRNet-39 和 PIDNet 在检测完整性方面有改善，但在边界细节还原上仍有差距。对于具有挑战性的划痕缺陷（第 5、6 行），由于目标形态细长且对比度低，DeepLabv3+、PIDNet 等出现断裂和噪声响应，DANet 和 BiSeNetV2 则存在平滑过度的问题。相比之下，本文方法 MSCENet 在缺陷形态保持、边缘清晰度方面均具有最佳表现，结构完整性最好，与真实值吻合度最高。

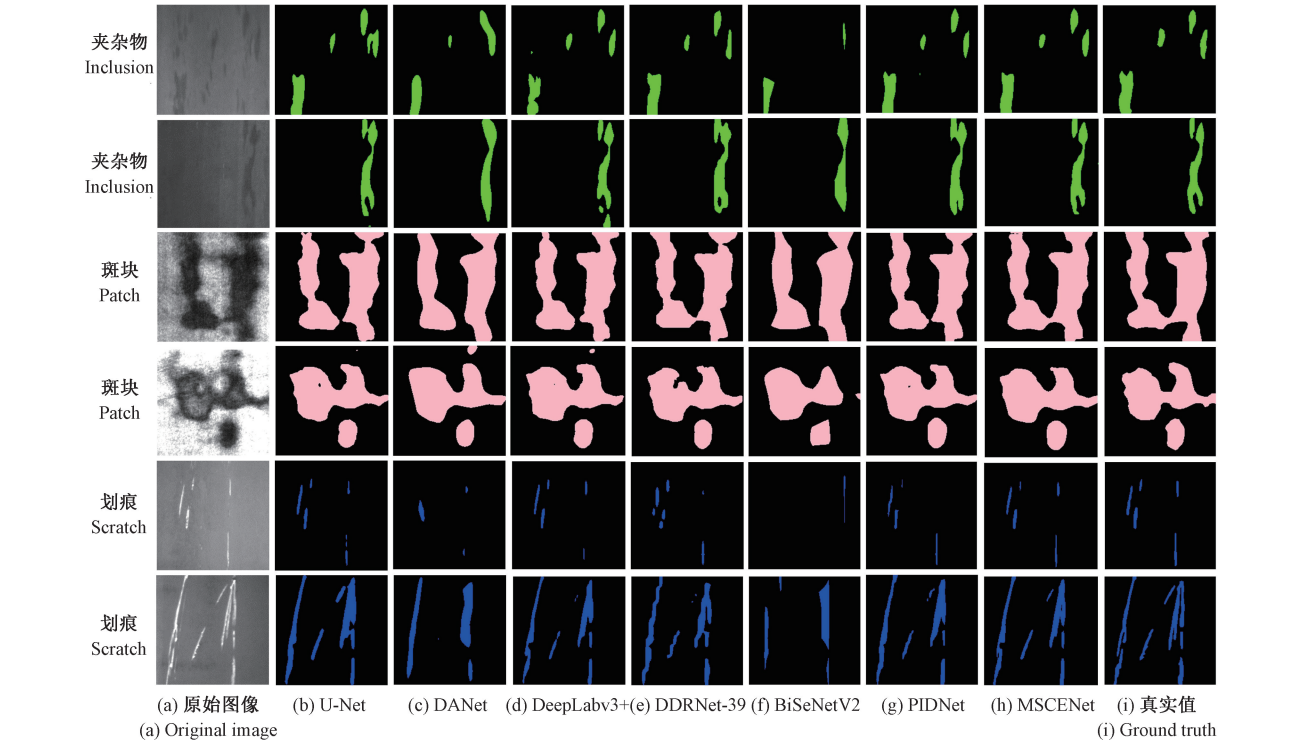


图 6 SD-saliency-900 数据集上的对比实验结果

Fig. 6 The comparative experimental results on the SD-salince-900 dataset

2) MT 数据集对比实验

为进一步验证 MSCENet 在工业缺陷检测任务中的综合性能,在磁砖(MT)缺陷数据集上进行了对比实验。如

表 5 所示, MSCENet 在多项性能指标上实现了最优的分割精度,其 mIoU 达到 72.75%,显著优于 U-Net 和 DeepLabv3+。

表 5 磁砖(MT)缺陷数据集上各方法的对比实验结果

Table 5 Comparative experiments results of various methods on magnetic tile (MT) defect dataset

方法	mIoU/%	mPA/%	Accuracy/%	Time/s	Params/M
U-Net	66.96	81.95	97.83	0.010	18.272
DANet	50.94	55.62	96.88	0.007	7.084
DeepLabv3+	64.32	73.82	96.79	0.008	11.092
PSPNet	46.04	49.11	96.51	0.007	7.108
DDRNet-39	69.37	80.72	97.72	0.012	32.357
FDSNet	64.78	72.61	97.53	0.007	0.970
BiSeNetV2	52.05	55.39	97.30	0.010	5.100
FFNet	62.63	68.74	97.12	0.010	9.657
LEDNet	59.26	66.77	97.59	0.015	0.897
STDC	48.02	54.23	96.51	0.005	8.570
PIDNet	69.41	79.52	97.69	0.010	7.624
HRNet	71.93	84.82	97.56	0.038	29.541
MSCENet	72.75	84.51	97.89	0.009	9.418

图 7 给出了 MT 数据集上的可视化检测结果。在第 1 行所示气孔缺陷中,DeepLabv3+、DDRNet-39 以及 PIDNet 均出现严重类别混淆,将气孔错误判为断裂缺陷;DANet 与 BiSeNetV2 可基本定位缺陷,但边界模糊、完整性差。在第 2、3 行的裂纹和断裂缺陷样本中,DDRNet-39 与 BiSeNetV2 虽可检测到部分目标,但存在边界断裂或过度平

滑问题。对于第 4、5 行中形态多变的磨损和不均匀缺陷,传统方法易受背景干扰,出现漏检(如 DeepLabv3+、DANet)或欠分割或过拟合(如 DANet、BiSeNetV2、PIDNet)的问题。对比之下, MSCENet 在类别预测和边界细节还原方面均更接近真实标注,表现出更好的结构感知能力和更强的鲁棒性,可以有效识别出缺陷区域的细节变化。

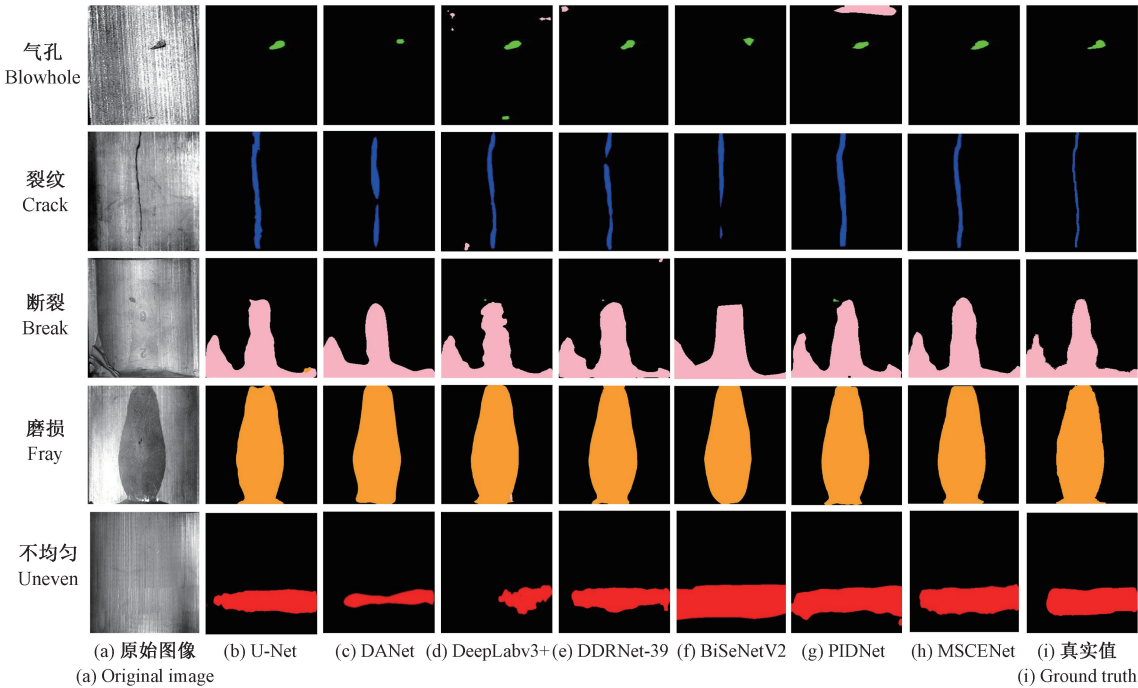


图 7 磁砖(MT)缺陷数据集上的对比实验结果

Fig. 7 The comparative experimental results on the magnetic tile (MT) defect dataset

综合而言, MSCENet 不仅在缺陷分类准确性方面显著优于其他对比方法, 有效缓解类间混淆问题, 同时在高精度定位与边界完整性方面也表现出竞争力, 在 MT 数据集上展现出优异的综合性能, 为复杂工业缺陷检测任务提供了更可靠的解决方案。

3) 轮胎缺陷图像数据集对比实验

在轮胎缺陷图像数据集上的对比结果如表 6 所示。结果表明, MSCENet 的 mIoU、mPA 和 Accuracy 显著优于其他对比模型, 在精度与轻量化之间实现了更优的平衡。

表 6 轮胎缺陷图像数据集上各方法的对比实验结果

Table 6 Comparative experiments results of various methods on tire defect image dataset

方法	mIoU/ %	mPA/ %	Accuracy/ %	Time/ s	Params/ M
U-Net	82. 78	90. 84	98. 41	0. 010	18. 272
DANet	78. 02	82. 56	97. 93	0. 007	7. 084
DeepLabv3+	78. 03	87. 85	97. 98	0. 008	11. 092
PSPNet	57. 12	64. 00	96. 60	0. 007	7. 108
DDRNet-39	82. 05	91. 67	98. 36	0. 012	32. 357
FDSNet	79. 82	90. 16	98. 09	0. 007	0. 970
BiSeNetV2	72. 14	80. 13	97. 44	0. 010	5. 100
FFNet	80. 82	89. 10	98. 16	0. 010	9. 657
LEDNet	75. 89	89. 36	97. 68	0. 015	0. 897
STDC	76. 81	86. 49	97. 61	0. 005	8. 570
PIDNet	83. 16	92. 12	98. 34	0. 011	7. 624
HRNet	84. 06	92. 72	98. 51	0. 038	29. 541
MSCENet	84. 52	93. 86	98. 63	0. 009	9. 418

轮胎缺陷图像数据集上的可视化检测结果如图 8 所示。在第 1 行胎面异物缺陷中, DDRNet-39 和 DeepLabv3+ 将背景误判为前景, BiSeNetV2 虽可定位目标, 但对异物轮廓的还原较差。在第 2 行胎侧帘线裂纹小目标缺陷中,

DeepLabv3+、DDRNet-39 和 PIDNet 均出现显著漏检现象。在第 3 行中, DeepLabv3+、BiSeNetV2 等难以准确识别帘线开裂结构。在第 5、6 行中, U-Net、DeepLabv3+ 和 PIDNet 对边缘的识别能力较弱, 导致边界模糊与过分割现象。

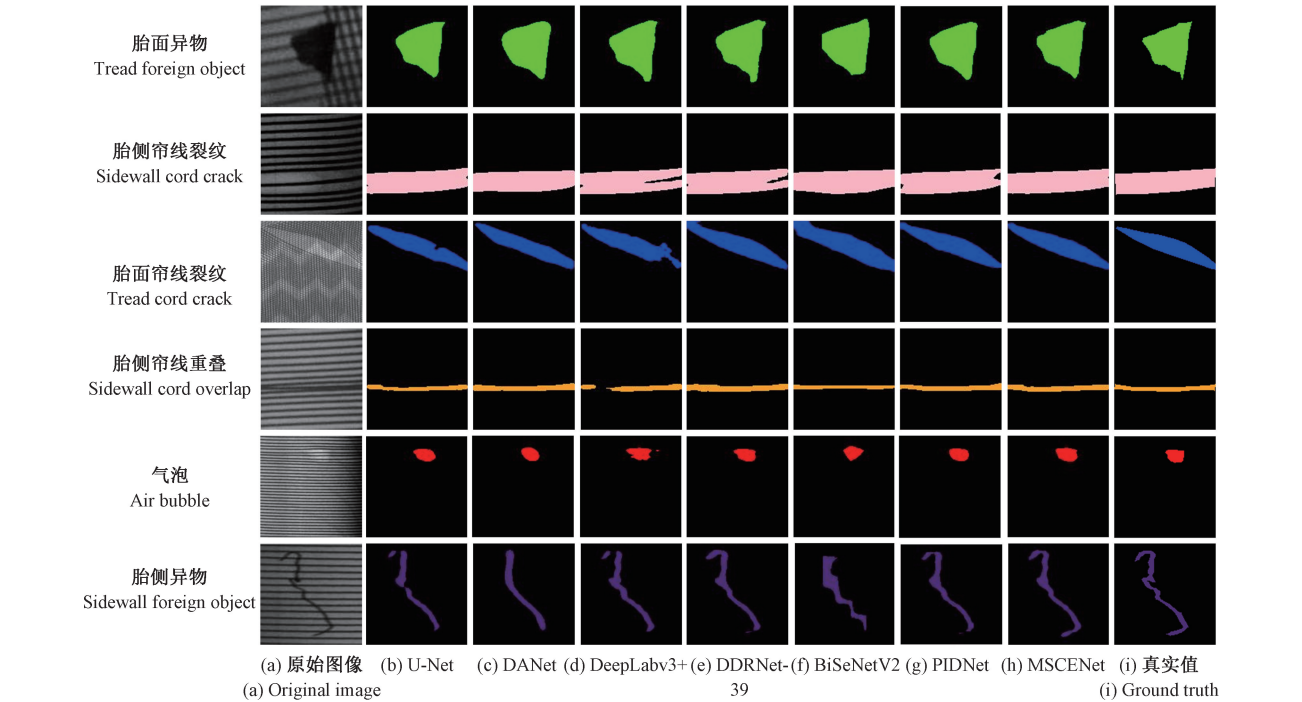


图 8 轮胎缺陷图像数据集上的对比实验结果

Fig. 8 The comparative experimental results on the tire defect image dataset

相比之下, MSCENet 在不同缺陷类型上均展现出更精确的分割边界、更好的区域一致性及更高的细节还原度。综上, MSCENet 在轮胎缺陷图像数据集中能够准确判别不同缺陷类别, 缓解类间混淆问题, 同时在细节还原、边界拟合和小目标检测方面表现优异, 验证了其在复杂工业场景下的鲁棒性和实用性。

4 结 论

本研究提出了一种基于多尺度上下文增强的轻量级工业缺陷检测网络 MSCENet, 并在多个典型工业缺陷数据集上进行了系统验证。首先, 该方法通过 3 个核心模块实现性能提升: 多尺度注意力增强特征提取(MAFE)模块有效聚合多尺度上下文信息, 显著增强模型对从微小裂纹到大面积磨损等多尺度缺陷的感知能力; 残差增强融合(REF)模块通过注意力引导与残差连接实现了编码器-解码器间的特征高效融合, 有效保留了细节与边缘结构; 全局注意力聚合(GAA)模块通过通道与空间注意力协同优化, 有效聚焦缺陷区域并抑制复杂背景干扰。实验结果表明, MSCENet 以参数量仅为 9.418 M 的条件下, 可实现单幅图像 0.009 s 的推理速度, 并在多个工业缺陷数据集上取得了领先的 mIoU、mPA 与 Accuracy 指标性能, 在精度、速度与轻量化三者间实现了优异平衡, 展现出显著的工业应用潜力。未来工作可从以下两方面展开: 一方面, 探索更极致的轻量化网络结构, 在维持精度的同时进一步压缩模型规模、提升推理速度, 以满足更高实时性要求的工业场景; 另一方面, 可尝试将所提出的模块(如 MAFE 等)与弱监督学习方法相结合, 降低对大量精细标注数据的依赖, 提升在标注资源受限的工业环境中的实用性与泛化能力。

参考文献

- [1] CAO B, LI CH H, SONG Y F, et al. Network intrusion detection model based on CNN and GRU[J]. Applied Sciences, 2022, 12(9):4184.
- [2] ZHENG ZH ZH, SHEN J Y, SHAO Y Y, et al. Tire defect classification using a deep convolutional sparse-coding network [J]. Measurement Science and Technology, 2021, 32(5):055401.
- [3] 赵佰亭, 张晨, 贾晓芬. ECC-YOLO: 一种改进的钢材表面缺陷检测方法[J]. 电子测量与仪器学报, 2024, 38(4):108-116.
ZHAO B T, ZHANG CH, JIA X F. ECC-YOLO: An improved method for detecting surface defects in steel [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(4):108-116.
- [4] 吴葛, 朱宇凡, 贾泽宁. 改进 YOLO11 的 PCB 表面缺陷检测方法 [J]. 电子测量技术, 2025, 48 (14): 136-145.

WU G, ZHU Y F, JIA Z N. Improved PCB surface defect detection method based on YOLO11 [J]. Electronic Measurement Technology, 2025, 48 (14): 136-145.

- [5] 刘传洋, 吴一全, 刘景景. 基于视觉的输电线路金具锈蚀缺陷检测方法研究进展[J]. 仪器仪表学报, 2024, 45(3):286-305.
LIU CH Y, WU Y Q, LIU J J. Research progress of vision-based rust defect detection methods for metal fittings in transmission lines[J]. Chinese Journal of Scientific Instrument, 2024, 45(3): 286-305.
- [6] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2015: 3431-3440.
- [7] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-decoder with atrous separable convolution for semantic image segmentation [C]. European Conference on Computer Vision, 2018: 833-851.
- [8] SHI X W, ZHANG SH B, CHENG M, et al. Few-shot semantic segmentation for industrial defect recognition[J]. Computers in Industry, 2023: 148.
- [9] CHEN J R, KAO SH H, HE H, et al. Run, don't walk: Chasing higher FLOPS for faster neural networks[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 12021-12031.
- [10] XIE X W, PAN X P, SHAO F, et al. MCI-Net: Multi-scale context integrated network for liver CT image segmentation [J]. Computers and Electrical Engineering, 2022, 101:108085.
- [11] HU J, SHEN L, SUN G, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018:7132-7142.
- [12] WANG Q L, WU B G, ZHU P F, et al. ECA-Net: Efficient channel attention for deep convolutional neural networks [C]. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2020:11531-11539.
- [13] WOO S, PARK J, LEE J Y, et al. CBAM: Convolutional block attention module[C]. European Conference on Computer Vision, 2018:3-19.
- [14] HUANG ZH Q, LI Y J, BAI M H, et al. A multiscale spatiotemporal fusion network based on an attention mechanism [J]. Remote Sensing, 2022, 15(1):182.
- [15] ZHOU X, ZHANG Y CH, LIU ZH, et al. IFIFusion: A independent feature information fusion model for surface defect detection [J]. Information

- Fusion, 2025,120:103039.
- [16] SHAO N, ZOU G F, WEI L Y, et al. Weak feature defect detection of insulators considering class boundaries and inter-class dependencies [J]. IEEE Transactions on Instrumentation and Measurement, 2024,73:1-13.
- [17] SONG G R, SONG K CH, YAN Y H, et al. EDRNet: Encoder-decoder residual network for salient object detection of strip steel surface defects[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12):9709-9719.
- [18] HUANG Y B, QIU C Y, YUAN K. Surface defect saliency of magnetic tile[J]. The Visual Computer, 2018:612-617.
- [19] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]. Medical Image Computing and Computer Assisted Intervention Society, 2015: 234-241.
- [20] ZHAO H SH, SHI J P, QI X J, et al. Pyramid scene parsing network[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017:6230-6239.
- [21] FU J, LIU J, TIAN H J, et al. Dual attention network for scene segmentation[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2019: 3141-3149.
- [22] HONG Y D, PAN H H, SUN W CH, et al. Deep dual-resolution networks for real-time and accurate semantic segmentation of road scenes [J]. ArXiv preprint arXiv:2101.06085, 2021.
- [23] YU CH Q, GAO CH X, WANG J B, et al. Bisenet V2: Bilateral network with guided aggregation for real-time semantic segmentation [J]. International Journal of Computer Vision, 2021, 129(11):3051-3068.
- [24] ZHANG J, DING R W, BAN M J, et al. FDSNeT: An accurate real-time surface defect segmentation network[C]. ICASSP 2022 IEEE International Conference on Acoustics, Speech and Signal Processing(ICASSP), 2022:3803-3807.
- [25] MEHTA D, SKLIAR A, YAHIA H B, et al. Simple and efficient architectures for semantic segmentation [C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2022: 2627-2635.
- [26] WANG Y, ZHOU Q, LIU J, et al. Lednet: A lightweight encoder-decoder network for real-time semantic segmentation[C]. 2019 IEEE International Conference on Image Processing (ICIP), 2019: 1860-1864.
- [27] XU J C, XIONG Z X, BHATTACHARYYA S P. PIDNet: A real-time semantic segmentation network inspired by PID controllers[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023: 19529-19539.
- [28] SUN K, XIAO B, LIU D, et al. Deep high-resolution representation learning for human pose estimation[C]. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5686-5696.
- [29] FAN M Y, LAI SH Q, HUANG J SH, et al. Rethinking bisenet for real-time semantic segmentation[C]. IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2021: 9716-9725.

作者简介

李帆, 硕士研究生, 主要研究方向为机器视觉、无损检测。

E-mail:lf_2426923409@163.com

周迅, 硕士研究生, 主要研究方向为机器视觉、无损检测。

E-mail:Zhouxunn_n@163.com

张岩, 博士, 教授, 主要研究方向为无损检测, 计算机视觉、深度学习、智能制造、红外线、超声波等数据可视化及其应用。

E-mail:zy@qust.edu.cn

顾海明(通信作者), 博士, 教授, 主要研究方向为偏微分方程数值解的理论及应用软件。

E-mail:ghm@qust.edu.cn