

DOI: 10.19650/j.cnki.cjsi.J2413194

基于层次化多尺度特征融合的金属缺陷分类模型^{*}

李季桐, 刘 杰, 杨 娜, 王子宁

(沈阳工业大学机械工程学院 沈阳 110870)

摘 要:金属缺陷检测作为工业质量控制的关键环节,其检测精度直接影响制造业智能化进程。针对现有特征融合模块存在特征信息丢失、跨尺度信息交互不足以及识别准确率低等问题,提出一种基于层次化多尺度特征融合的分类模型。该模型通过融合 Swin Transformer 与 ConvNeXt 两种网络架构的互补优势,构建了具有层次化感知能力的特征学习网络。其中, Swin Transformer 采用移位窗口机制和多级自注意力机制有效捕获全局特征, ConvNeXt 通过深度可分离卷积和高效卷积操作精准提取局部特征。为实现全局与局部的高效融合,创新性地设计自适应层次特征融合层,该层采用通道注意力机制、空间注意力机制和多尺度融合策略,实现全局与局部特征在多层次上的有效融合,同时在该层中增加多层倒残差融合模块,通过动态调整提取特征信息,以确保特征融合的精准性与可靠性。为验证模型的有效性,在公开 NEU-DET 和 GC10-DET 数据集上进行实验,准确率分别达到 99.6% 和 96.9%。为验证模型的泛化性,在自建数据集上进行实验,准确率达到 99.8%,与目前主流算法 ConvNeXt、Swin transformer、VGG16、ResNet34 模型相比,准确率分别提升 3.4%、2.3%、4.3%、2.7%。实验结果表明, HMFF 模型在金属缺陷检测领域具有更显著的分类准确性和鲁棒性,为工业场景下的高精度缺陷检测提供了新的研究方法。

关键词: Swin Transformer; ConvNeXt; 缺陷分类; 多尺度特征提取; 自适应层次特征融合; 多层倒残差连接

中图分类号: TH164

文献标识码: A

国家标准学科分类代码: 510.40

Enhanced hierarchical multi-scale feature fusion model for metal defect classification

Li Jitong, Liu Jie, Yang Na, Wang Zining

(School of Mechanical Engineering, Shenyang University of Technology, Shenyang 100870, China)

Abstract: Metal defect detection, as a critical component of industrial quality control, directly determines the advancement of intelligent manufacturing. To address existing issues in feature fusion modules including feature information loss, insufficient cross-scale interaction, and low recognition accuracy, a hierarchical multi-scale feature fusion-based classification model is proposed. By integrating complementary advantages of Swin Transformer and ConvNeXt architectures, a hierarchical perception-enabled feature learning network is constructed. Specifically, the Swin Transformer employs shifted window mechanisms and multi-stage self-attention to effectively capture global features, while ConvNeXt utilizes depth separable convolution and efficient convolutional operations for precise local feature extraction. To achieve efficient global-local fusion, an innovative adaptive hierarchical feature fusion layer is designed, incorporating channel attention mechanisms, spatial attention mechanisms, and multi-scale fusion strategies to enable effective multi-level feature integration. Additionally, a multi-layer inverted residual fusion module is incorporated to dynamically adjust feature extraction, ensuring precise and reliable feature fusion. Experimental validation on public NEU-DET and GC10-DET datasets demonstrates superior performance with accuracy rates of 99.6% and 96.9%, respectively. To verify generalization capability, evaluations on a self-constructed dataset achieve an accuracy of 99.8%, outperforming mainstream models including ConvNeXt, Swin Transformer, VGG16, and ResNet34 by 3.4%, 2.3%, 4.3%, and 2.7% respectively. The results confirm that the HMFF model exhibits enhanced classification accuracy and robustness in metal defect detection, providing a novel methodological framework for high-precision industrial defect inspection.

Keywords: Swin Transformer; ConvNeXt; defect classification; multi-scale feature extraction; adaptive hierarchical feature fusion; multi-layer inverted residual connections

0 引言

在现代工业与制造业中,金属制品表面缺陷,如腐蚀、氧化、裂纹等,不仅影响设备性能,还可能导致事故。而传统的人工视觉和无损检测方法由于检测速度慢、人工成本高、依赖专家经验,已无法满足当前行业需求。因此,实现高精度和高效率的金属表面缺陷检测对保证产品质量和提升生产效率至关重要。工业上对金属缺陷检测关注度逐年增加,迫切需要更先进的检测手段取代人工检测,以提高检测效率和精度^[1-2]。

目前基于深度学习的表面缺陷分类检测方法按照模型可以分为卷积神经网络(convolutional neural networks, CNN)系列^[3]和视觉转换器(vision transformer, ViT)系列^[4],其中CNN系列分类模型有VGG^[5]、ResNet^[6]、ConvNeXt^[7]等,ViT系列分类模型有Swin Transformer^[8]等。Zhang等^[9]以CNN为主干,提出一种结合域自适应(domain adaptive, DA)和自适应卷积神经网络(adaptive convolutional neural network, ACNN)的检测新方法DA-ACNN,利用邻阈分类器和损耗优化实现精准分类,但该网络结构复杂,训练速度慢成本高,同时具有过拟合风险。Wang等^[10]对ResNet50进行改进,使用可变卷积网络,并对分类结构进行改进,使钢材表面缺陷识别准确率进一步提升,但该模型较复杂,算力需求大。付军军等^[11]基于EfficientNet模型,引入非对称卷积用来识别较小缺陷,然后将混合空间通道注意力机制引入BiFPN特征融合网络中以提高检测性能。Wan等^[12]采用VGG19结合迁移学习,提升小样本和不平衡数据集上钢表面缺陷识别性能,但在不平衡数据问题上分类效果仍需加强。Bai等^[13]采用ConvNeXt模型,通过伪标签和聚类的联合优化实现钢材表面缺陷的无监督检测,显著提升检测性能,但在应对复杂背景和微小缺陷方面仍有改进空间。Wang等^[14]引入动态上下文特征提取模块,结合多尺度空洞卷积(atrous spatial pyramid pooling, ASPP),增强模型对局部纹理和多尺度缺陷的捕捉能力,并嵌入高效通道注意力模块,抑制噪声干扰,以提高检测精度。

近年,Transformer因其擅长学习序列信息在自然语言处理领域备受关注。部分学者将Transformer引入计算机视觉领域,提供了新的发展方向。Zhu等^[15]提出基于Swin Transformer的高效网络,利用其分层结构和移动窗口注意力机制提取钢铁表面的多尺度特征,从而实现缺陷检测和分类,但在高噪声环境下检测性能受限。Gao等^[16]提出Cas-VSwin Transformer,通过引入局部自适应卷积模块和变尺度窗口机制,增强其在表面缺陷检测中的表现。Ding等^[17]结合ViT和混合窗口注意力机制,提出一种动态视觉转换器(dynamic vision transformer,

DHT)的新型方法,但其在处理高分辨率图像时速度较慢。陈俊英等^[18]提出多尺度ViT特征提取和注意力特征融合的端到端缺陷检测算法,利用层级多头注意力对不同尺度的特征图执行自适应的注意力操作使网络更好的提取特征,但该模型较复杂。Li等^[19]将CNN和Transformer编码器合并,提出一种混合网络架构CNN-t,该网络具有较强的局部和全局感知能力,但识别准确率较低。在此基础上,唐东林等^[20]在浅层CNN中融合Transformer,并基于此加入SE模块。该网络利用浅层CNN和Transformer多头注意力机制的特点,可有效学习图像局部和全局信息,而SE模块能够实现对某些通道中特点的重点关注。陈晓雷等^[21]提出全局和局部多尺度特征融合分类网络(multi-scale local-global network, MLG-Net),该网络全局分支采用空间金字塔池化和自注意力机制,局部分支采用密集连接卷积网络,参数量较大。

在金属缺陷检测中,现有深度学习模型结构复杂、计算复杂度高,且特征融合模块存在信息损失、准确率低以及检测速度慢,针对上述问题,提出一种基于层次化多尺度特征融合(hierarchical multi-scale feature fusion, HMFF)分类模型。该模型首先通过Swin Transformer层提取全局特征信息,通过ConvNeXt层提取局部特征信息,同时在模型中加入自适应层次特征融合层(adaptive hierarchical feature fusion, AHFF),该层采用通道注意力机制(channel attention, CA)、空间注意力机制(spatial attention, SA)和多尺度融合策略,实现全局与局部特征在多层次上的高效融合,并在该层增加多层倒残差融合模块(multi-layer inverted residual block, ML-IRB)动态调整提取的特征信息,从而提高分类精度。新构建的HMFF分类模型,在保证高准确、低损失的情况下,具有较少的参数,较快的检测速度,能够有效地实现金属缺陷分类。

1 算法介绍

HMFF模型网络结构如图1所示,详细参数如表1所示。该模型能够有效获取不同尺度下的局部和全局空间信息,采用并行结构从局部和全局特征层提取缺陷图像典型特征,再通过AHFF层融合不同层次的特征,最终由全局平均池化进行降维,由线性层实现金属缺陷分类。

1.1 Swin Transformer 全局特征提取层

微软研究院在2021年ICCV会议上首次提出Swin Transformer^[8],即一种基于改进ViT的视觉模型。该模型采用与CNN类似的层次化搭建方法,通过对特征图下采样的方式使其具备多尺度检测能力。该模型采用窗口多头自注意力(window multi-head self-attention, W-MSA)机制以降低算力要求,同时为解决多头自注意力(multi-

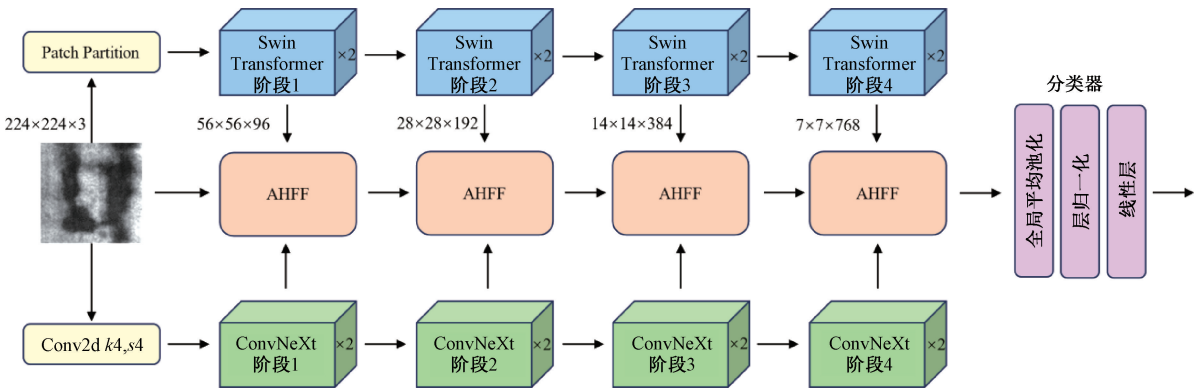


图 1 HMFF 网络结构
Fig. 1 HMFF architecture

表 1 HMFF 网络具体参数
Table 1 HMFF specific parameters

阶段	输出	局部分支		AHFF	全局分支	
stem	56×56,96	4×4,96, stride 4		4×4,96, stride 4		
→spatial attention channel attention←						
1	56×56,96	d7×7,96 1×1,96	×2	d3×3,96 Downsample,3×3 1×1,384 1×1,96 Upsample	win,sz,7×7, dim96,head3 1×1,96	×2
→spatial attention channel attention←						
2	28×28,192	d7×7,192 1×1,192	×2	d3×3,192 Downsample,3×3 1×1,768 1×1,192 Upsample	win,sz,7×7, dim192,head6 1×1,192	×2
→spatial attention channel attention←						
3	14×14,384	d7×7,384 1×1,384	×2	d3×3,384 Downsample,3×3 1×1,1 536 1×1,384 Upsample	win,sz,7×7, dim384,head12 1×1,384	×2
→spatial attention channel attention←						
4	7×7,768	d7×7,768 1×1,768	×2	d3×3,768 Downsample,3×3 1×1,3 072 1×1,768 Upsample	win,sz,7×7, dim768,head24 1×1,768	×2

head self-attention, MSA)导致的窗口间信息隔绝问题,在网络中加入移动窗口多头自注意力(shifted window multi-head self-attention, SW-MSA)。经过广泛验证,分层滑动窗口机制使自注意力计算更加高效,显著提升模型在图像分类、目标检测和语义分割等任务中的性能。具体网络结构如图 2 所示。

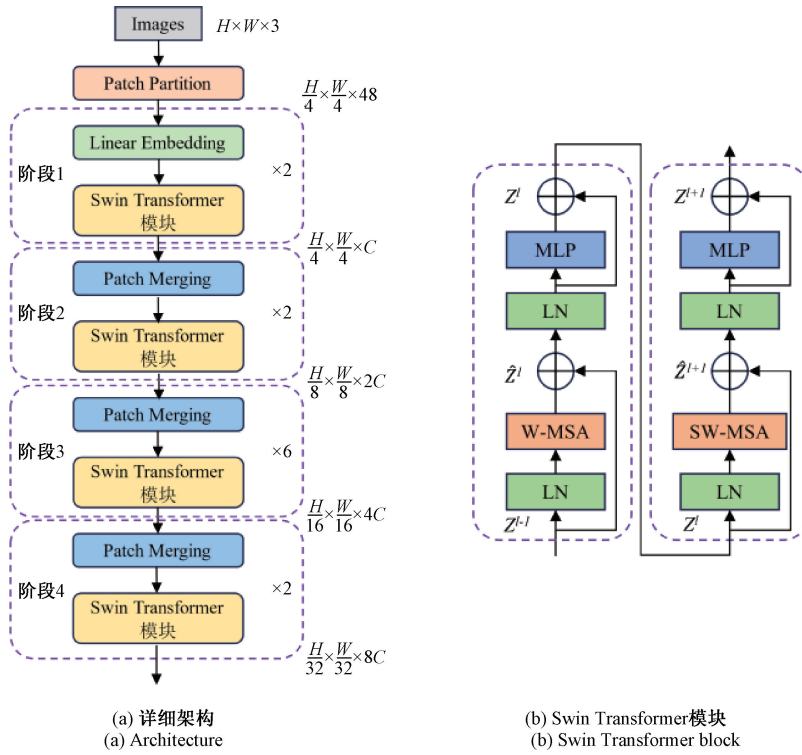


图 2 Swin Transformer 网络结构

Fig. 2 Swin Transformer architecture

图 2(a)所示为 Swin Transformer 体系结构,首先使用 4×4 的图像分割模块,将输入的图像分割成不重叠的图像块,然后应用线性嵌入层提升向量的特征表示能力,最后将分割好的图像块输入到 Swin Transformer Block 模块中,该阶段一起被称为“第 1 阶段”。在后个阶段中,首先使用图像块融合进行下采样实现图像尺寸缩减,同时,在每个阶段中重复堆叠 Swin Transformer Block 模块进行深度特征提取。图 2(b)为 Swin Transformer Block 模块的结构,其由一个基于移位窗口的 MSA 模块、2 个多层感知机 (multi-layer perceptron, MLP) 和 GELU 非线性层组成。在每个 MSA 模块和每个 MLP 模块前后分别有一个归一化层 (layer norm, LN) 和一个残差连接。

具体计算方法如式 (1) ~ (4) 所示。

$$\Omega(MSA) = 4hwC^2 + 2(hw)^2C \quad (1)$$

$$\Omega(W - MSA) = 4hwC^2 + 2M^2hwC \quad (2)$$

$$\mathbf{g}_i = f^{1 \times 1}(W - MSA(LN(\mathbf{G}_{i-1}))) + \mathbf{G}_{i-1} \quad (3)$$

$$\mathbf{G}_i = f^{1 \times 1}(SW - MSA(LN(\mathbf{g}_i))) + \mathbf{g}_i \quad (4)$$

其中, h 代表 feature map 的高度; w 代表 feature map 的宽度; C 代表 feature map 的深度; M 代表每个窗口 (Windows) 的大小; $\mathbf{G}_i$ 和 $\mathbf{g}_i$ 表示 SW-MSA 和 W-MSA 的全局输出特征; $f^{1 \times 1}$ 是大小为 1×1 的卷积运算,相当于线性运算; LN 是 LayerNorm 操作。

1.2 ConvNeXt 局部特征提取层

ConvNeXt^[7]是 2022 年由 Facebook AI Research 提出的一种卷积神经网络模型。在卷积神经网络的基础上,结合 Transformer 思想,通过简化卷积操作和引入大卷积核来提升模型的表现力,并优化其网络结构使得计算更高效。具体网络结构图如图 3 所示。

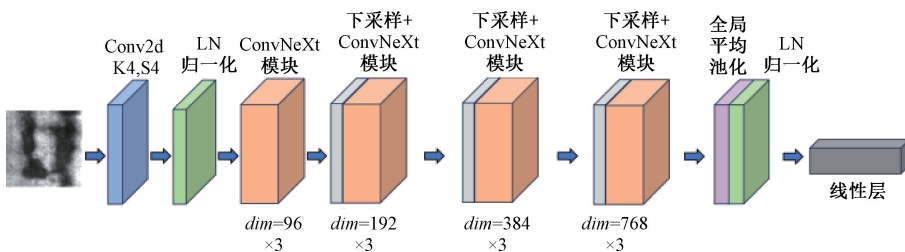


图 3 ConvNeXt 网络结构

Fig. 3 ConvNeXt architecture

在 ConvNeXt 网络中,借鉴了 ResNeXt 网络(如图 4(a))中的分组卷积结构,并引入深度卷积,最后通过 1×1 卷积压缩通道,确保分组卷积的组数与通道数一致。图 4(b)将 ResNeXt 网络中的瓶颈层构造改为逆瓶颈层构造,即将输入的形态配置从 $(384, 96, 96)$ 调整为 $(96, 384, 384)$ 。图 4(c)在大尺寸卷积核优化阶段,调整深度卷积模块的位置,将其上移,并调整其卷积核大小,从 $(3, 3)$ 修改为 $(7, 7)$ 。这种调整旨在优化特征提取过程,通过这些不同的结构修改,从而确定了最优的网络设计方案。具体 ConvNeXt 模块结构如图 5 所示。

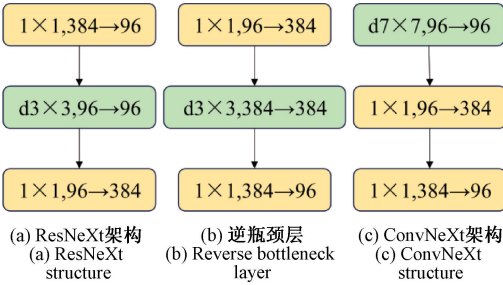


图 4 3 种模块对比

Fig. 4 Comparison of three modules

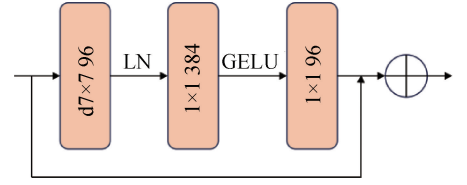


图 5 ConvNeXt 模块

Fig. 5 ConvNeXt block

1.3 AHFF 自适应层次特征融合层

AHFF 层能够自适应地将不同层次的局部特征、全局特征以及前一层次融合的特征信息整合至当前输入特征中。如图 6 和 7 所示, G_i 为全局特征模块输出的特征矩阵, L_i 为局部特征模块输出的特征矩阵, F_{i-1} 为前一阶段 AHFF 输出的特征矩阵, 而 F_i 则是当前阶段 AHFF 融合后生成的特征矩阵。通过这种方式, AHFF 层能够高效融合多层次的特征信息, 提升模型对复杂金属缺陷的表达和分类能力。

由于全局特征层中的自注意可以在一定程度上捕获全局信息, 因此 AHFF 层将传入的全局特征信息输入到

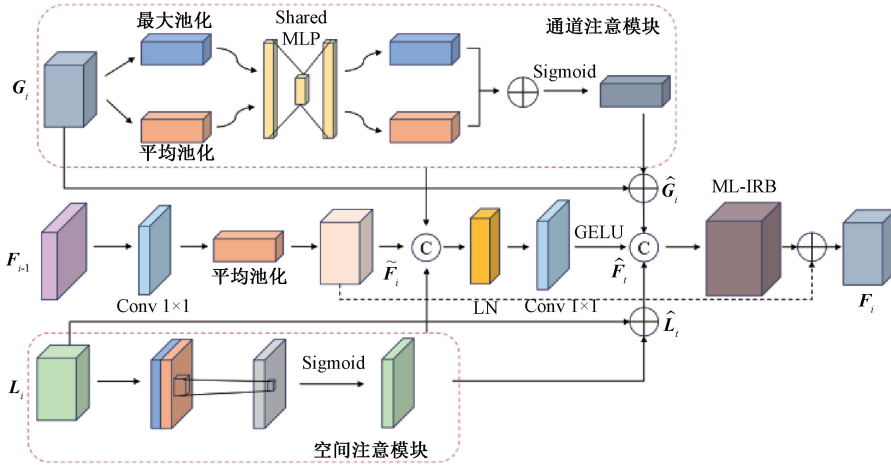


图 6 AHFF 详细架构

Fig. 6 AHFF architecture

CA 模块中,该模块对通道映射间的依赖关系加以利用,对特定含义的特征进行表示,将局部特征输入到 SA 模块中,局部细节得以加强,不相关区域被抑制。最后,将每个注意力和融合路径产生的结果进行特征融合,并连接一个多层倒残差融合块,如图 8 所示。将 IRB 升级为 ML-IRB 增加上下采样层,在一定程度上减少计算量,防止过拟合,同时增大感受野,使后面的卷积核能够有效地捕获各层次的全局和局部特征信息。CA 与 SA 模块的相关计算如式(5)~(7)所示。

$$CA(x) = \sigma(MLP(Avgpool(x)) +$$

$$MLP(Maxpool(x))) \quad (5)$$

$$SA(x) = \sigma \left(f^{7 \times 7} (Concat[AvgPool(x), MaxPool(x)]) \right) \quad (6)$$

$$IRMLP(x) = f^{1 \times 1} (f^{1 \times 1} (f^{3 \times 3} (LN(x)) + LN(x))) \quad (7)$$

其中, x 为输入特征矩阵, σ 为 Sigmoid 函数, $f^{7 \times 7}$ 为 7×7 的卷积运算。特征融合相关计算如式(8)~(12)所示。

$$\hat{G}_i = CA(G_i) \otimes G_i \quad (8)$$

$$\hat{L}_i = SA(L_i) \otimes L_i \quad (9)$$

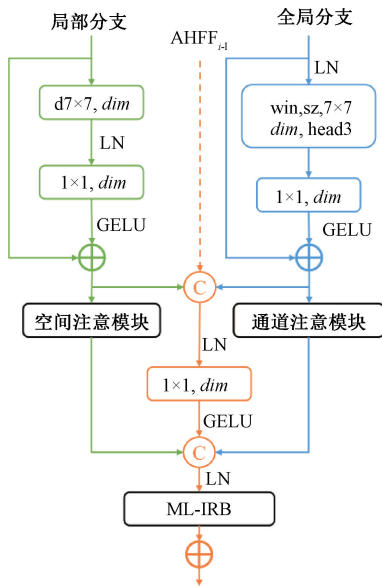


图 7 AHFF 设计原理

Fig. 7 Design schematic of AHFF

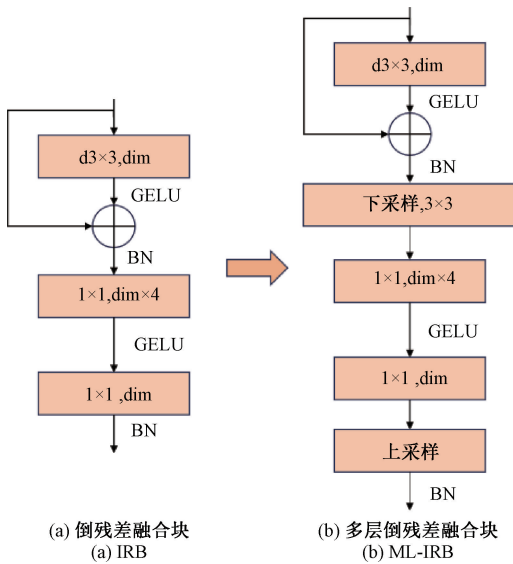


图 8 多层倒残差融合块

Fig. 8 Multi-layer inverted residual block

$$\tilde{F}_i = \text{Avgpool}(f^{1 \times 1}(x_{f,i-1})) \quad (10)$$

$$\hat{F}_i = f^{3 \times 3}(\text{Concat}[\mathbf{G}_i, \mathbf{L}_i, \tilde{F}_i]) \quad (11)$$

$$\mathbf{F}_i = \text{IRMLP}(\text{Concat}[\hat{\mathbf{G}}_i, \hat{\mathbf{L}}_i, \hat{\mathbf{F}}_i]) + \tilde{F}_i \quad (12)$$

其中, $\otimes$ 表示各元素逐一相乘, $\hat{\mathbf{G}}_i$ 由通道注意力组合生成, $\hat{\mathbf{L}}_i$ 由空间注意力组合生成, $\tilde{F}_i$ 由 AHFF 块的上一级通过下采样生成, $\hat{F}_i$ 是全局 - 局部特征和上一级融合的结果。最后, $\hat{F}_i$ 、 $\hat{\mathbf{G}}_i$ 和 $\hat{\mathbf{L}}_i$ 被连接起来, 并通过 ML-IRB 生成特征 F_i 。

2 实验结果与分析

2.1 数据集介绍

为验证所提 HMFF 模型有效性, 引入公开数据集 NEU-DET^[22] 和 GC10-DET^[23] 来进行验证, 同时使用自建数据集来进行模型泛化能力验证。

1) NEU-DET 数据集, 如图 9 所示。该数据集 6 种带钢表面缺陷为裂纹 (crazing)、夹杂 (inclusion)、斑块 (patches)、麻点 (pitted_surface)、轧制氧化皮 (rolled-in_scale) 和划痕 (scratches), 分别如图 9(a) ~ (f) 所示。每类缺陷包含 300 张图片, 共 1 800 张钢表面缺陷图像, 每张图片像素 200 pixels×200 pixels。

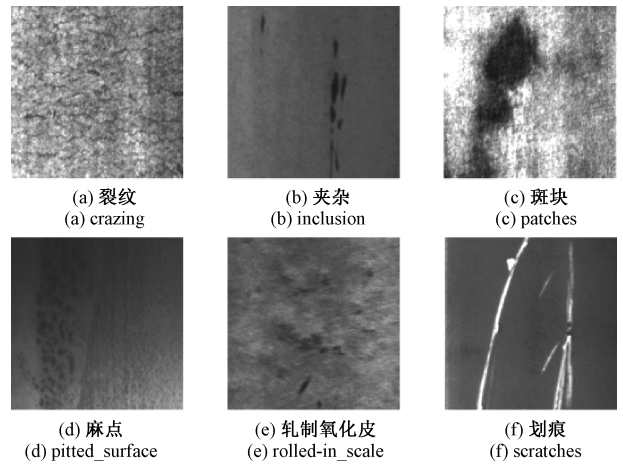


图 9 NEU-DET 数据集相关缺陷

Fig. 9 NEU-DET dataset related defects

2) GC10-DET 数据集, 如图 10 所示。该数据集包含 10 种钢板表面缺陷。包括冲孔 (punching_hole)、焊缝 (welding_line)、水斑 (water_spot)、丝斑 (silk_spot)、轧坑 (rolling_pit)、腰部折痕 (waist_fold)、夹杂物 (inclusion)、油斑 (oil_spot)、新月形缝隙 (crescent_gap) 和折痕 (crease), 分别如图 10(a) ~ (j) 所示。共 3 570 张钢表面缺陷图像。

3) 自建数据集, 如图 11 所示。本数据集采用 CMOS 成像系统 (索尼 IMX 传感器, 单张最大分辨率 5 320×7 968), 并将实验室环境色温调整至 5 600 k 以模拟日光, 基于此构建多角度光学成像平台。以此实现 304 钢样本表面全覆盖成像, 单样本最大采集面积达 200 mm×200 mm。使用以上成像系统捕捉真实使用环境下 304 不锈钢表面的微观缺陷。其包括水斑 (water_spot)、杂质 (impurity)、锈迹 (rust_stains) 和划痕 (scratches), 分别如图 11(a) ~ (d) 所示。每类缺陷包含 300 张图片, 共 1 200 张 304 不锈钢缺陷图像。

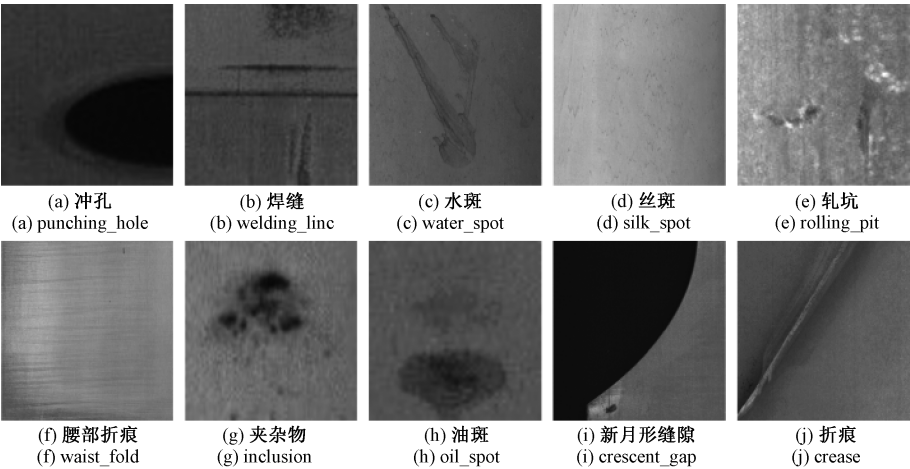


图 10 GC10-DET 数据集相关缺陷
Fig. 10 GC10-DET dataset related defects

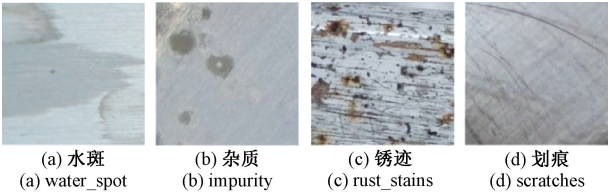


图 11 自建数据集相关缺陷
Fig. 11 Self-constructed dataset related defects

为确保所提模型验证过程的严谨性与有效性,采用 9:1 的比例精确划分训练集与测试集,以此确保所有待评估模型均基于相同且独立的数据集进行训练与测试,从而保障评估结果的一致性与客观性。

2.2 缺陷分类实验

实验环境为 CPU 为 3 700 X, GPU 为 2 060 super, RAM 为 16 G, Python 版本为 3.9。在模型训练期间,将学习率设置为 0.001。batchsize 为 8, epoch 根据不同数据集的收敛程度不同分别设置为 100 和 200。同时,选择个常见的多分类任务评价指标^[24]进行评估:准确率、召回率和 $F1$ 值。

首先引入 NEU-DET 数据集验证该模型对缺陷的分类能力,由图 12 可以看出,由于迭代初期学习率较大,导致损失函数和准确率变化浮动较大。在迭代 50 次左右后,模型趋于平稳接近收敛状态,准确率达到 99.6%。同时,模型性能稳定,未出现过拟合现象。

通过图 13 的混淆矩阵也能清晰看出,6 类缺陷有 5 类均达到了 100% 的准确率,说明该模型整体识别准确率较高,只有麻点 (pitted_surface) 类的少量样本被错误分类为轧制氧化皮 (rolled-in_scale), 准确率仍达到 96.67%。整体看来,该模型在检测钢板表面缺陷方面具有较高的准确性和鲁棒性。

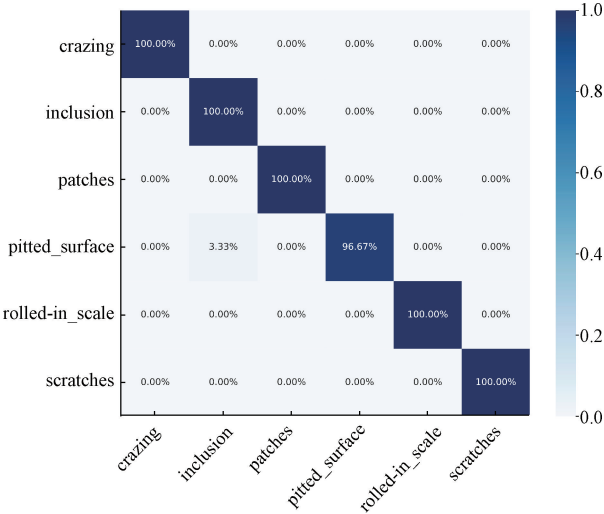
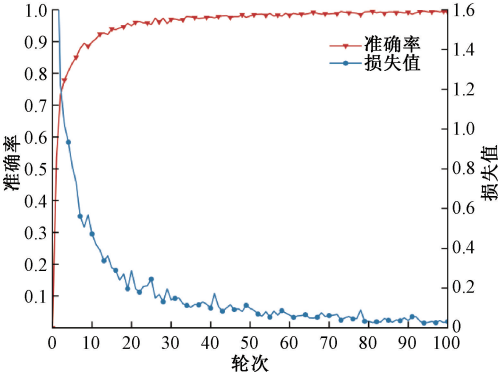


图 13 NEU-DET 数据集混淆矩阵
Fig. 13 NEU-DET dataset confusion matrix

为验证该模型的通用性,引入 GC10-DET 数据集进行实验,由图 14 可以看出,该数据集由于缺陷类别更多,缺陷特征间界限模糊、难以精准区分,故模型训练相较于 NEU-DET 数据集收敛速度较慢,但在迭代 180 轮左右稳定的达到了收敛状态,准确率达到 96.9%。这一结果有力证明该模型在复杂多变的工业检测任务时,依然能够保持良好的性能。

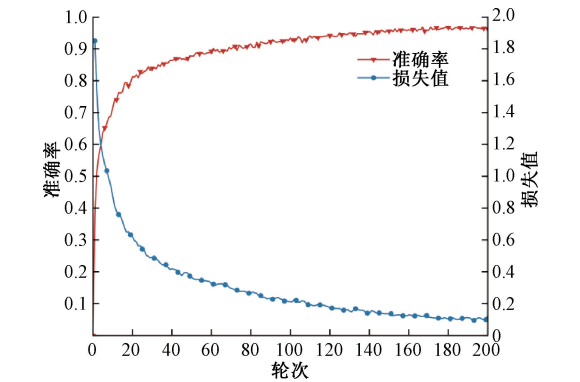


图 14 GC10-DET 数据集学习曲线

Fig. 14 GC10-DET dataset learning curve

通过图 15 的混淆矩阵可以清晰看出,10 类缺陷均达到 95% 以上的准确率,其中轧坑 (rolling_pit) 和折痕 (crease) 达到 100%,冲孔 (punching_hole) 这一类有 4.62% 识别为新月形缝隙 (crescent_gap),其原因为 2 类缺陷大体形状和颜色均相似,进而导致误差率最大,但整体识别效果较好。

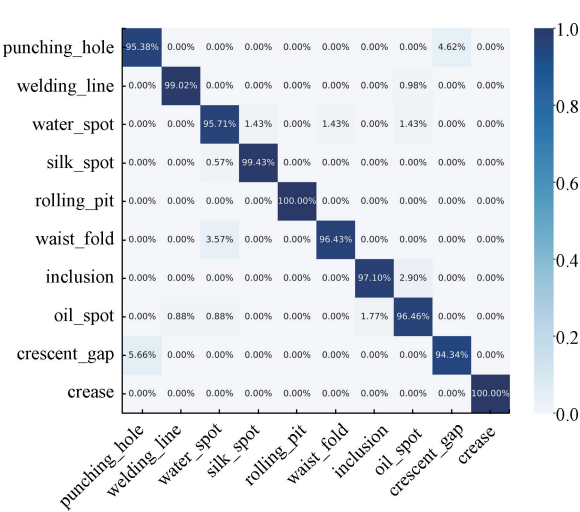


图 15 GC10-DET 数据集混淆矩阵

Fig. 15 GC10-DET dataset confusion matrix

为验证该模型的泛化能力,又引入自建数据集进行实验,由图 16 可以看出,在训练初期,模型准确率迅速上升,并趋于稳定,最终达到 99.8%。同时,损失值逐渐下

降,并趋于极低值,这表明该模型在自建数据集上性能较好,具备良好的泛化能力。

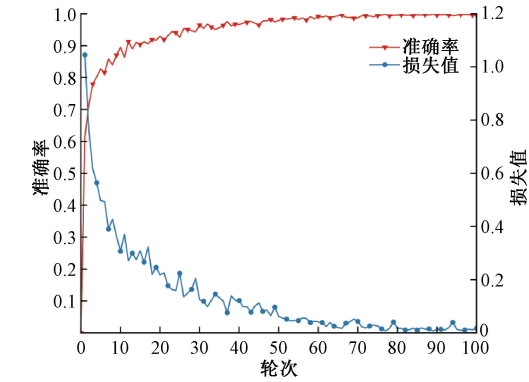


图 16 自建数据集学习曲线

Fig. 16 Self-constructed dataset learning curve

通过图 17 的混淆矩阵可以清晰看出,4 类缺陷中有 3 类准确率均达到 100%,所有样本均被正确分类且无交叉误分现象。仅 water_spot 类存在 3.33% 的误分率,原因是存在少量边界特征与 impurity 类别的重叠现象。该模型整体展现出高度可靠性,平均准确率为 99.8%,说明该模型泛化能力较好。

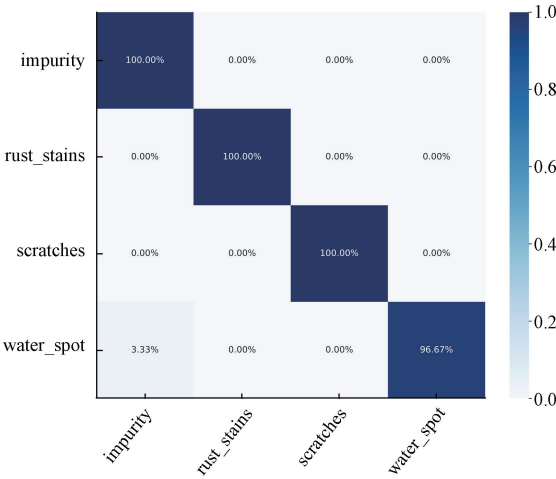


图 17 自建数据集混淆矩阵

Fig. 17 Self-constructed dataset confusion matrix

2.3 热力图可视化实验

为直观验证该模型捕捉缺陷特征的能力,在公开数据集和自建数据集上分别进行可视化实验,采用类激活热力图技术,通过反向传播梯度信号生成高分辨率热力图,对模型缺陷区域进行可视化分析,具体如图 18~20 所示。

由图 18 热力图可以看出,由于 NEU-DET 数据集缺陷特征较为明显,模型能够有效捕捉到缺陷区域与背景间的显著差异。在检测过程中,缺陷区域均识别正

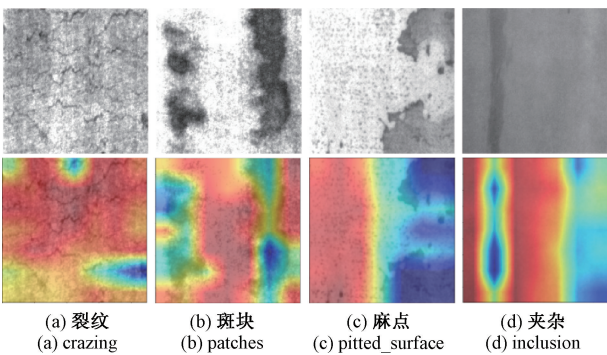


图 18 部分 NEU-DET 数据集热力图
Fig. 18 Partial NEU-DET dataset heatmap

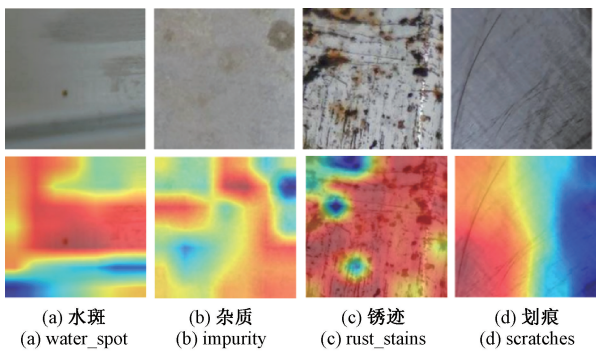


图 20 部分自建数据集热力图
Fig. 20 Partial self-constructed dataset heatmap

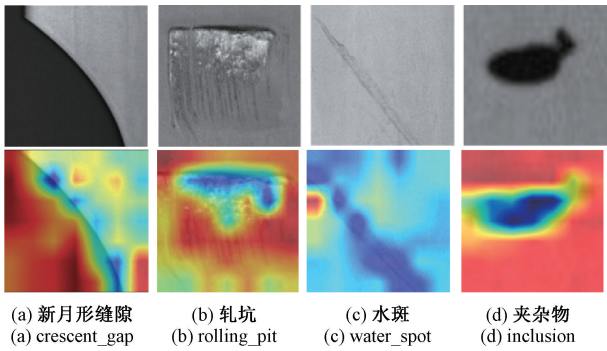


图 19 部分 GC10-DET 数据集热力图
Fig. 19 Partial GC10-DET dataset heatmap

确,轮廓划定较为清晰,整体效果较好,有效证明该模型性能较好。

由图 19 热力图可以看出,GC10-DET 数据集缺陷类

别较多,缺陷之间相似度较高,但该模型对特征复杂且不明显的缺陷类型仍展现出优异的检测性能,其定位也较为准确且缺陷轮廓分割清晰,尤其在弱特征缺陷的识别中具有显著优势。

由图 20 热力图可以看出,该模型在自建数据集上的整体检测效果较好,轮廓较为清晰,针对划痕这类缺陷,虽准确定位出缺陷区域,但轮廓较模糊,后续将进行深入研究。

2. 4 网络模型对比实验

为验证该模型的综合性能,选取 ConvNeXt、Swin transformer、VGG16 和 ResNet344 种模型作为基线,在 NEU-DET、GC10-DET 和自建数据集上分别进行对比实验。

在公开数据集上,模型综合指标对比如表 2 所示,实验对比曲线如图 21 和 22 所示。

表 2 公开数据集模型综合指标对比

Table 2 Comparison of comprehensive indicators of public dataset models

模型	NEU-DET				GC10-DET			
	准确率/%	召回率/%	F1 值/%	损失	准确率/%	召回率/%	F1 值/%	损失
ConvNeXt	97.5	95.6	95.5	0.076	92.1	87.9	88.5	0.238
Swin Transformer	95.8	95.3	95.5	0.110	90.9	85.1	87.1	0.282
VGG16	96.5	97.2	96.8	0.106	94.3	86.9	88.1	0.167
ResNet34	97.4	96.2	94.8	0.077	94.1	90.5	91.9	0.174
文献[20]	97.8	—	—	0.067	—	—	—	—
文献[25]	90.3	—	—	—	74.1	—	—	—
文献[26]	98.8	—	—	—	90.2	—	—	—
HMFF	99.6	99.4	99.4	0.019	96.9	92.5	92.7	0.098

由表 2 可以看出, HMFF 模型在识别准确率、召回率和 F1 值均优于其他模型,与 ConvNeXt、Swin Transformer、VGG16、ResNet34 模型相比,其准确率在 NEU-DET 和 GC10-DET 数据集上分别提升 2.1%、3.8%、3.1%、2.2%

和 4.8%、6.0%、2.6%、2.8%,同时与目前该领域文献进行对比,准确率也有所提升,在损失值方面,较其他 4 种模型相对更低,足可以看出, HMFF 模型收敛效果更好,稳定性更强。

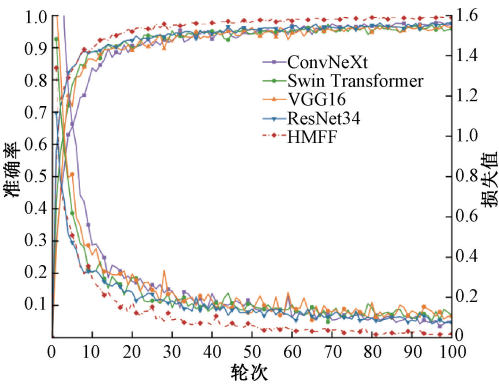


图 21 NEU-DET 数据集模型对比曲线
Fig. 21 NEU-DET dataset model comparison curve

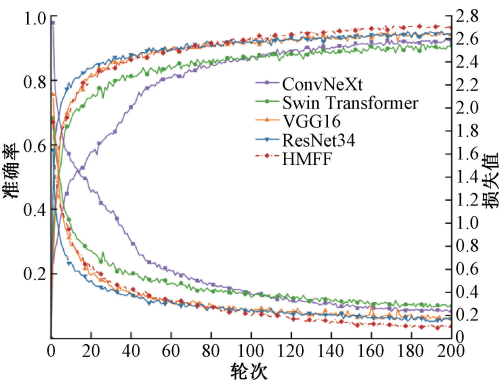


图 22 GC10-DET 数据集模型对比曲线
Fig. 22 GC10-DET dataset model comparison curve

如图 21 所示,由于 NEU-DET 数据集缺陷之间差异明显,5 种模型收敛速度大体相同,波动较小,而 HMFF 模型相较于其他 4 种模型,迭代 50 轮后,损失值更稳定、波动更小,损失最低且准确率最高,在每一轮都更优于其他 4 种模型,最终准确率达到 99.6%。

如图 22 所示,GC10-DET 数据集缺陷类别较多,缺陷特征更难区分,故模型性能之间存在较大差异。ConvNeXt 模型在训练初期波动较大,极不稳定;Swin Transformer 模型整体损失值和准确率相比于其他 4 种模型都较差;迭代 100 轮后,HMFF 模型充分展现其独特优势,损失值呈现出平稳的收敛趋势,同时模型准确率逐步达到最高,充分证明该模型的有效性。

在自建数据集上,模型综合指标对比如表 3 所示,实验对比曲线如图 23 所示。

由表 3 可以看出, HMFF 模型在准确率、召回率和 $F1$ 值个指标上均达到最优,与 ConvNeXt、Swin Transformer、VGG16、ResNet34 模型相比,准确率分别提升 3.4%、2.3%、4.3% 和 2.7%,同时,其最低的损失值进一步验证了该模型泛化能力。

表 3 自建数据集模型综合指标对比
Table 3 Comparison of comprehensive indicators for self-constructed dataset models

数据集	模型	准确率/%	召回率/%	$F1$ 值/%	损失
自建数据集	ConvNeXt	96.4	94.7	95.2	0.103
	Swin Transformer	97.5	96.2	96.4	0.054
	VGG16	95.5	94.2	95.3	0.133
	ResNet34	97.1	96.2	94.6	0.099
	HMFF	99.8	99.6	99.6	0.012

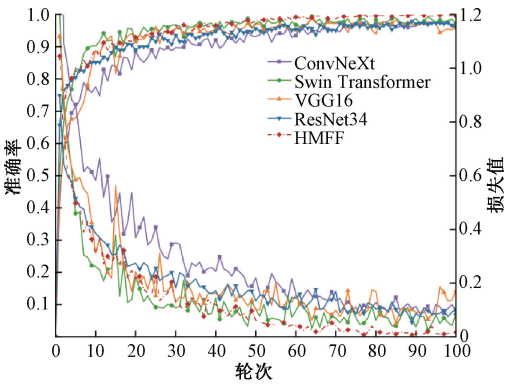


图 23 自建数据集模型对比曲线
Fig. 23 Self-constructed dataset model comparison curve

如图 23 所示,由于自建数据集类别少,各模型训练曲线均呈现显著波动。VGG16 模型在整个训练过程中,波动最大,极不稳定;ConvNeXt 模型次之,而 HMFF 模型相较于其他 4 种模型展现出鲁棒收敛特性,且准确率也达到最高,并最终趋于稳定,更充分地证明了该模型泛化能力较好。

2.5 模型复杂度及效率分析对比实验

为更全面地验证 HMFF 模型的优势,将其模型参数量、浮点计算量和参数文件大小作为模型复杂度的评价指标。不同模型帧率和参数数量比较如表 4 所示, HMFF 模型在保持较低参数数量的同时,识别帧率最高,为识别帧率最低的 Swin Transformer 的 3 倍,说明其推理速度最快。

表 4 不同模型帧率和参数量比较
Table 4 Comparison of frame rates and parameter quantities among different models

模型	模型参数量/M	浮点计算量/GFLOPs	参数文件大小/MB	帧率/fps
ConvNeXt	111.27	4.455	106.0	103.85
Swin Transformer	110.00	4.371	105.0	47.46
VGG16	537.14	15.466	512.0	92.78
ResNet34	81.21	7.334	81.3	148.52
HMFF	82.49	6.130	95.0	150.21

图 24 直观地展示了不同方法的模型复杂度和准确率的比较。所提出的 HMFF 模型不仅在识别准确率方面表现出色,而且在模型复杂度上也优于其他对比模型。具体而言,HMFF 模型的数量和浮点计算量分别为 82.49 M 和 6.130 GFLOPs,并且参数文件大小为 95 MB,同时实现了 99.6% 的识别准确率。与准确率较低的 Swin Transformer 模型相比,参数量减少了

27.51 M,准确率提高了 3.8%。与准确率次之的 VGG16 模型相比,HMFF 在减少超过 6 倍的模型参数量、2 倍以上的浮点计算量和 5 倍以上的参数文件大小的同时,准确率提高了 3.1%。与模型复杂度最低的 ResNet34 相比,虽然在参数量略高于该模型,但在浮点计算量方面较该模型更低,并在识别准确率方面提高了 2.2%。

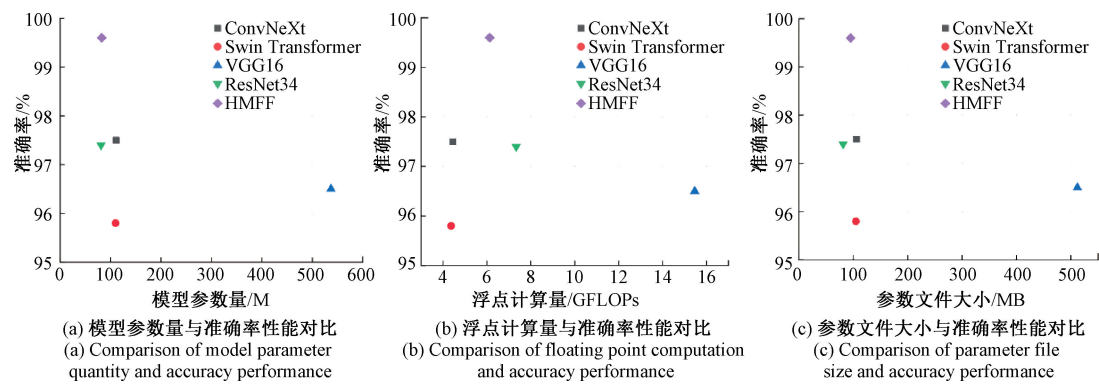


图 24 模型复杂度与准确率的比较
Fig. 24 Comparison of model complexity and accuracy

综上所述,HMFF 模型在金属缺陷任务上不仅实现了更高的准确率,而且在模型参数量、浮点计算量和参数文件大小上也展现了显著优势,在工业场景中具有更强的实用性和部署价值。

3 结 论

金属表面缺陷检测在工业领域中具有重要意义。HMFF 模型通过结合 Swin Transformer 的全局特征提取能力与 ConvNeXt 的局部特征提取能力,克服传统金属表面缺陷检测模型的局限性。同时,设计 AHFF 层,实现对不同尺度特征的高效融合。为验证模型性能,在公开的 NEU-DET 和 GC10-DET 数据集分别进行实验,结果表明,HMFF 模型准确率、召回率、F1 值分别达到 99.6%、99.4%、99.4% 和 96.9%、92.5%、92.7%。为验证模型泛化能力,在自建数据集上进行实验,该模型准确率、召回率、F1 值分别达到 99.8%、99.6%、99.6%。同时该模型在处理复杂背景和微小缺陷时表现出参数量和计算复杂度低、训练速度快等特点,对于未来工业领域金属缺陷检测的高精度分类具有重要的实际意义。此外,HMFF 模型在其他领域的应用也值得进一步深入研究。

参考文献

[1] LYU X M, DUAN F J, JIANG J J, et al. Deep metallic surface defect detection: The new benchmark and detection network[J]. Sensors, 2020, 20 (6): 1562-

1572.
[2] 赵朗月, 吴一全. 基于机器视觉的表面缺陷检测方法研究进展[J]. 仪器仪表学报, 2022, 43 (1): 198-219.
ZHAO L Y, WU Y Q. Research progress on surface defect detection methods based on machine vision[J]. Chinese Journal of Scientific Instrument, 2022, 43 (1): 198-219.
[3] JHA S B, BABICEANU R F. Deep CNN-based visual defect detection: Survey of current literature[J]. Computers in Industry, 2023, 148: 103911-103923.
[4] KHAN S, NASEER M, HAYAT M, et al. Transformers in vision: A survey[J]. ACM Computing Surveys, 2022, 54 (10s): 1-41.
[5] SENGUPTA A, YE Y T, WANG R, et al. Going deeper in spiking neural networks: VGG and residual architectures[J]. Frontiers in Neuroscience, 2019, 13: 95-106.
[6] WU Z F, SHEN CH H, VAN DEN HENGEL A. Wider or deeper: Revisiting the resnet model for visual recognition[J]. Pattern Recognition, 2019, 90: 119-133.

- [7] CONG SH, ZHOU Y. A review of convolutional neural network architectures and their optimizations[J]. Artificial Intelligence Review, 2023, 56(3): 1905-1969.
- [8] WANG J P, XU G L, YAN F J, et al. Defect transformer: An efficient hybrid transformer architecture for surface defect detection[J]. Measurement, 2023, 211: 112614.
- [9] ZHANG S Y, ZHANG Q J, GU J F, et al. Visual inspection of steel surface defects based on domain adaptation and adaptive convolutional neural network[J]. Mechanical Systems and Signal Processing, 2021, 153: 107541.
- [10] WANG SH, XIA X J, YE L Q, et al. Automatic detection and classification of steel surface defect using deep convolutional neural networks[J]. Metals, 2021, 11(3): 388.
- [11] 付军军, 陈林, 童华敏, 等. 基于改进 EfficientNet 的海上风机叶片早期缺陷检测及分类[J]. 电子测量技术, 2024, 47(16): 110-119.
- FU J J, CHEN L, TONG H M, et al. Early defect detection and classification of offshore wind turbine blades based on improved EfficientNet[J]. Electronic Measurement Technology, 2024, 47(16): 110-119.
- [12] WAN X, ZHANG X Y, LIU L L. An improved VGG19 transfer learning strip steel surface defect recognition deep neural network based on few samples and imbalanced datasets[J]. Applied Sciences, 2021, 11(6): 2606.
- [13] BAI D X, LI G F, JIANG D, et al. Unsupervised method for detecting surface defects in steel based on joint optimization of pseudo-labeling and clustering[J]. Measurement Science and Technology, 2024, 35(10): 106131.
- [14] WANG J J, BI L, MA X D, et al. An efficient YOLOX-based method for photovoltaic cell defect detection[J]. Instrumentation, 2024, 11(2): 83-94.
- [15] ZHU W, ZHANG H, ZHANG CH, et al. Surface defect detection and classification of steel using an efficient Swin Transformer[J]. Advanced Engineering Informatics, 2023, 57: 102061.
- [16] GAO L F, ZHANG J X, YANG CH H, et al. Cas-VSwin transformer: A variant swin transformer for surface-defect detection[J]. Computers in Industry. 2022, 140: 103689.
- [17] DING CH, TANG D L, ZHENG X H, et al. DHT: Dynamic vision transformer using hybrid window attention for industrial defect images classification[J]. IEEE Instrumentation & Measurement Magazine, 2023, 26(2): 19-28.
- [18] 陈俊英, 李朝阳, 席月芸, 等. ViT 和注意力融合的分类不平衡 PCB 缺陷检测方法[J]. 仪器仪表学报, 2024, 45(4): 294-306.
- CHEN J Y, LI ZH Y, XI Y Y, et al. ViT and attention fusion for class-imbalanced PCB defect detection[J]. Chinese Journal of Scientific Instrument, 2024, 45(4): 294-306.
- [19] LI SH F, WU CH X, XIONG N X. Hybrid architecture based on CNN and transformer for strip steel surface defect classification[J]. Electronics, 2022, 11(8): 1200.
- [20] 唐东林, 杨洲, 程衡, 等. 浅层卷积神经网络融合 Transformer 的金属缺陷图像识别方法[J]. 中国机械工程, 2022, 33(19): 2298-2305, 2316.
- TANG D L, YANG ZH, CHENG H, et al. Metal defect image recognition method using shallow convolutional neural network fusion with Transformer[J]. China Mechanical Engineering, 2022, 33(19): 2298-2305, 2316.
- [21] 陈晓雷, 李正成, 杨富龙, 等. 全局与局部多尺度特征融合晶圆缺陷分类网络[J]. 电子测量与仪器学报, 2024, 38(10): 159-169.
- CHEN X L, LI ZH CH, YANG F L, et al. Wafer defect classification network with global and local multi-scale feature fusion[J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(10): 159-169.
- [22] BAO Y Q, SONG K CH, LIU J, et al. Triplet-graph reasoning network for few-shot metal generic surface defect segmentation[J]. IEEE Transactions on Instrumentation and Measurement, 2021, 70: 5011111.
- [23] 姜晓恒, 段金忠, 卢洋, 等. 融合先验知识推理的表

面缺陷检测[J]. 图学学报, 2024, 45(5): 957-967.

JIANG X H, DUAN J ZH, LU Y, et al. Surface defect detection based on reasoning with prior knowledge[J]. Journal of Graphics, 2024, 45(5): 957-967.

[24] 吕游, 郑茜, 齐欣宇, 等. 基于改进 EfficientNet 的红外图像光伏组件故障识别研究[J]. 仪器仪表学报, 2024, 45(4): 175-184.

LYU Y, ZHENG X, QI X Y, et al. Research on infrared image photovoltaic module fault recognition based on improved EfficientNet[J]. Chinese Journal of Scientific Instrument, 2024, 45 (4): 175-184.

[25] ELLEUCH J F, MEHDI M Z, KAMOUN H M, et al. New possibilistic formalism based on confidence levels for metallic surface defect classification[J]. IEEE Access, 2024, 12, 93010-93019.

[26] 胡坤, 吴国庆, 胡祖辉, 等. 基于改进的 VGG16 网络金属表面缺陷图像分类研究[J]. 计算机应用与软件, 2024, 41(6): 175-180.

HU K, WU G Q, HU Z H, et al. Research on image classification of metal surface defects based on improved VGG16 network[J]. Computer Applications and Software, 2024, 41(6): 175-180.

作者简介



李季桐, 2023 年于沈阳工业大学获得学士学位, 现为沈阳工业大学硕士研究生, 主要研究方向为机器视觉检测。
E-mail: 1114221315@ qq. com

Li Jitong received her B.Sc. degree from Shenyang University of Technology in 2023.

Now she is a master's student at Shenyang University of Technology. Her main research interest includes machine vision inspection.



刘杰 (通信作者), 2002 年于鞍山科技大学获得学士学位, 2005 年于东北大学获得硕士学位, 2008 年于东北大学获得博士学位, 现为沈阳工业大学副教授, 主要研究方向为机械系统动力学、机械设备智能运维等。

E-mail: liuj@ sut. edu. cn

Liu Jie (Corresponding author) received her B.Sc. degree from Anshan University of Science and Technology in 2002, received her M. Sc. degree from Northeastern University in 2005, received her Ph. D. degree from Northeastern University in 2008. Now she is an associate professor at Shenyang University of Technology. Her main research interests include mechanical system dynamics and intelligent operation and maintenance of mechanical equipment, etc.