

基于边缘引导与局部-全局 Mamba 网络的 工业缺陷分割模型^{*}

陈国国 张笃振 王斯豪 杨昌昌

(江苏师范大学人工智能与计算机学院 徐州 221116)

摘要: 针对工业表面缺陷检测中存在的缺陷形态多变、对比度低及小目标易漏检等挑战,提出了一种基于边缘引导与局部-全局 Mamba(GLMamba)网络的语义分割模型,构建了一个由边缘引导的编码-解码分割网络,实现了“边缘增强-全局局部编码-双路径解码”的协同。具体地,首先构建了(GLMamba)编码器,通过在浅层与深层分别部署局部与全局 Mamba 块,协同建模图像的局部细节与全局上下文。另外,设计了边缘特征提取模块(EGFM),通过融合 Sobel 梯度与多尺度特征,显式生成边缘特征图,并利用边缘注意力模块(EAM)对其进行增强。为进一步将边缘信息注入主干网络,引入了多尺度边界门控机制(MSBG),动态调制编码器与解码器特征。解码器则采用双路径注意力模块,并行融合空间细节与通道上下文,进一步优化了多尺度特征的融合效率与语义一致性。在 NEU-Seg、MT-Defect、FSSD-12 这 3 个数据集上的实验结果证明了该模型的有效性和泛化能力。

关键词: 缺陷分割;状态空间模型;边缘引导;注意力机制

中图分类号: TP391;TN911 **文献标识码:** A **国家标准学科分类代码:** 520.6040

Edge-guided local-global Mamba network for industrial defect segmentation

Chen Yuanyuan Zhang Duzhen Wang Sihao Yang Changchang

(School of Artificial Intelligence and Computer Science, Jiangsu Normal University, Xuzhou 221116, China)

Abstract: To address the challenges in industrial surface defect detection, such as variable morphology, low contrast and the frequent missing of small targets. A novel framework for semantic segmentation is presented, which leverages an edge-guided encoder-decoder architecture and local-global Mamba (GLMamba) to realize the synergy of "edge enhancement, global-local encoding and dual-path decoding". Firstly, a local-global Mamba encoder is built. Local details and global context are collaboratively modeled through the deployment of local and global Mamba blocks at shallow and deep layers, respectively. Secondly, an edge feature extraction module (EGFM) is designed, where edge feature maps are explicitly generated by fusing Sobel gradients with multi-scale features and are subsequently enhanced by an edge attention module (EAM). To further inject edge information into the backbone network, a multi-scale boundary gating (MSBG) mechanism is introduced to dynamically modulate encoder and decoder features. The decoder employs a dual-path attention module, where spatial details and channel context are fused in parallel to optimize multi-scale feature fusion efficiency and semantic consistency. Experimental results on the NEU-Seg, MT-Defect and FSSD-12 datasets demonstrate the effectiveness and strong generalization capability of the proposed model.

Keywords: defect segmentation; state space models; edge guidance; attention mechanism

0 引言

表面缺陷检测是现代工业制造自动化生产体系中的重要环节,其检测精度与效率直接关系到产品的可靠性和安全性。在各类工业产品制造的过程中难以避免会存在表面

的缺陷,例如划痕、裂纹、污渍等,缺陷的存在不仅会影响产品的外观,也会影响其相关产品的使用性能,甚至会成为潜在的安全隐患。传统的表面缺陷检测主要依赖于人们肉眼检测,这种检测方式效率底下,成本高昂,并且容易出现人为错误,难以满足大规模、高速度、高精度的现代工业生产

需求。随着深度学习的迅猛发展,基于机器视觉的自动化缺陷检测的方法逐渐取代传统的人工和基于规则的机器视觉方法。

在众多的基于深度学习的缺陷检测算法中,表面缺陷分割能够提供像素级的缺陷定位和精确的轮廓信息从而成为工业质检领域的关键技术。与仅判断图像中是否存在缺陷的图像级别分类和仅提供缺陷大致位置的区域级检测相比,分割任务输出的精确掩码能够为后续的缺陷尺寸测量、形态分析、严重程度评估提供关键的数据基础。然而,在工业场景下的缺陷分割仍面临着一些挑战。1)缺陷的形态复杂,包括尺度差异大,形状不规则,即使是同种缺陷之间也存在巨大差异;2)缺陷区域与背景之间的对比度低,且容易受到光照不均、反光等干扰因素的影响,导致难以从背景中准确分离出目标缺陷;3)对于小尺寸缺陷,因其本身像素占比极小,在下采样过程中易造成漏分割。

表面缺陷的特点要求分割模型具备局部纹理感知能力、捕获全局上下文信息能力的同时,还要保持对细粒度边界的敏感性。针对这些挑战,早期方法主要依赖卷积神经网络(convolutional neural network, CNN),其中文献[1-3]提出的 U-Net 及其系列变体凭借编码-解码与跳跃连接机制,在局部特征提取和空间细节恢复方面具有优势。但由于其感受野有限,在处理大尺度结构或弱对比缺陷时容易出现全局语义不足、边界不完整等问题。为了增强长距离依赖建模能力, Dosovitskiy 等^[4]提出的 ViT (vision transformer),将 Transformer 引入到分割任务中,其自注意力机制直接建模像素级别的依赖关系,单层网络即可捕捉全局的上下文信息。然而,Transformer 的计算开销随输入分辨率呈平方增长,内存资源消耗巨大;同时,缺乏局部归纳偏置的特性使其在细节的感知能力较弱,容易忽略微小缺陷。研究者们开始尝试结合 CNN 和 Transformer,现有的方法多采用并行或串行结构融合两者特征,但普遍存在特征融合方式粗糙、全局建模效率不足以及边界信息丢失等问题。在提升全局建模效率的探索中,近年来的文献[5]提出的状态空间模型(state space models, SSMs)为视觉任务提供了新的方向。Zhu 等^[6]提出基于 Mamba 选择性扫描机制的 VisionMamba,结合参数共享设计,在保持线性计算复杂度的同时,实现对图像空间和通道维度的动态建模,显著降低了全局建模的计算成本,但纯 Mamba 架构在局部纹理与精细边界的捕捉能力仍弱于卷积结构。另外,研究者也开始探索多特征融合策略,如多尺度特征融合^[7]和基于边缘约束^[8]等方法,以增强模型对复杂缺陷形态的表达能力。现有特征融合方法虽能结合局部和全局信息,但对边缘几何的利用多停留于浅层监督或简单拼接,难以在弱对比、小尺度缺陷中提供稳定的结构约束。

针对上述挑战,本文提出一种基于局部-全局 Mamba (global-local mamba, GLMamba)与边界感知融合的表面缺陷分割模型。将状态空间模型与显式边缘引导机制深度

融合,构建了“边缘增强-全局局部编码-双路径解码”的协同框架。

1)构建 GLMamba 编码器,引入 GLMamba 混合建模,在编码器中结合 LocalMamba 与 GlobalMamba 模块,分别建模局部结构与长程依赖,从而实现缺陷的细节捕捉与全局上下文的统一。

2)设计了边缘特征提取模块(edge guidance feature module, EGFM),基于 Sobel 算子提取梯度信息,结合可变形卷积与深度可分离残差块,在浅层网络实现边界特征的自适应增强。并通过边缘注意力模块(edge attention module, EAM)增强边缘信息和利用门控机制将边界图作为注意力权重动态引导编码器与解码器特征响应,显著提升模型对缺陷边缘的感知能力。

3)构建双路径注意力融合解码机制,采用门控增强策略协同调制跳跃连接与上采样特征,结合双路径注意力实现通道与空间交互建模,通过深度可分离卷积完成高效特征融合,显著提升边界分割精度。

4)提出基于边界与 Mamba 感知的表面缺陷分割网络:通过上述模块的有机整合,构建了一个高效、精准的缺陷分割模型。在多个公开缺陷数据集上的实验结果表明,所提方法在保持轻量化的同时,显著提升了缺陷分割的综合性能。

1 相关工作

1.1 基于卷积神经网络的语义分割模型

CNN 在计算机视觉领域展现出强大的特征提取能力,被广泛应用于医学图像分割、工业缺陷分割等任务。Stern 等^[9]提出全卷积网络首次实现了端到端的像素级预测,文献[1]设计了 U-Net 通过编码器-解码器结构和跳跃连接,有效融合了浅层细节与深层语义信息,成为缺陷分割的基准模型。随后,Chen 等^[10]提出的 DeepLab 系列模型通过引入空洞卷积扩大感受野,进一步提升了模型性能。传统卷积操作具有局部性,难以建模长距离依赖,导致大尺度缺陷的全局结构难以捕捉,小缺陷也容易丢失。为缓解这些问题,研究者提出了多特征融合、注意力机制等策略。Chen 等^[11]引入空间位置注意模块与跨尺度特征融合模块,缓解 U-Net 在空间信息保持与多尺度语义特征融合上的不足;陈光耀等^[12]则在太阳能电池片缺陷分割中引入跨层注意力模块,有效捕获长程依赖关系;周炜杰等^[13]通过像素级注意力模块和金字塔池化模块实现局部与全局特征的协同提取。Zhang 等^[14]设计了空间、上下文与边界三支,协同分割复杂路面缺陷;邹凯鑫等^[8]结合空间通道注意力与循环交叉模块,进一步增强模型对边界与上下文的建模能力。这些研究从不同角度验证了注意力机制与多尺度信息融合在提升复杂缺陷刻画精度方面的有效性。

然而,虽然这些办法在性能上都有一定的提升,但本质上仍然是对局部区域操作的增强,未能从根本上解决 CNN 对全局依赖建模不足的问题,在应对背景复杂、形态多变且

需全局语义理解的缺陷时,其性能上限逐渐显现。

1.2 Transformer 与全局建模视觉模型

Transformer 中自注意力机制能够建模序列中任意位置之间的依赖关系,具有强大的建模全局依赖能力,弥补了 CNN 感受野的局限。ViT 通首次将纯 Transformer 架构引入图像分类,Liu 等^[15]提出 Swin Transformer 则通过移位窗口机制提升了在密集预测任务中的效率与性能。文献[16-17]采用 ViT 替代 CNN 作为编码器进一步推动了 Transformer 在语义分割中的应用。由于单一依赖 Transformer 在局部细节捕捉能力较弱,研究者开始尝试将其与 CNN 进行结合。Zhang 等^[18]提出双分支局部引导全局自注意力网络,利用交叉分支特征交互实现了 CNN 与 Transformer 特征的深度融合;Liu 等^[19]采用交叉形窗口注意力增强感受野交互能力;针对复杂工业场景,Liu 等^[20]提出 TransUNet 框架将 Transformer 作为编码器引入缺陷检测;Zhu 等^[21]引入多物理信息并结合多路径学习与多模态特征融合,有效提升了图像分割精度。

尽管已有研究取得显著进展,但 Transformer 仍面临计算复杂度高、对边界信息不敏感等问题。现有融合方法在特征交互与边界信息利用方面仍显粗糙,如何在保持效率的同时实现全局上下文与局部信息的高效融合,仍是当前研究的重点与难点。

1.3 状态空间模型与 Mamba

近年来,状态空间模型的长序列建模中能力也开始显现,逐渐被应用到计算机视觉领域。文献[22]提出的 S4 模型(structured state space sequence model)引入具有特殊数学结构的状态矩阵,将连续时间状态空间进行离散化。Gu 等^[23]提出其视觉版本 VisionS4 通过结构化状态空间的设计在保持线性复杂度的同时实现对长距离依赖的有效建模,为视觉任务提供了新的基础架构。Mamba 作为 S4 的突破性进展,与传统状态空间模型对所有输入内容采用统一建模方式不同,它能根据当前输入动态调整参数,从而有选择地关注关键信息。其视觉版本 Vision Mamba 在效率与表达能力之间取得了均衡,尤其适合处理表面缺陷分割中常见的复杂形态、多尺度缺陷等问题,为缺陷检测任务提供了新的可行方案。

在医学影像分割领域,Liu 等^[24]提出的 Swin-UMamba 融合了视觉基础模型与 Mamba 的优势,在多个数据集上表现优异;Li 等^[25]通过引入多尺度特征融合机制,实现了轻量化的高效分割模型;Sun 等^[26]则通过小波变换与多尺度 Mamba 模块,增强了特征建模能力。在工业缺陷检测方面,文献[27-28]将 Mamba 与 CNN 结合,有效提升了跨模态鲁棒性与长程依赖建模能力;Zuo 等^[29]应用创新扫描模块在裂缝分割中取得领先性能。

尽管 Mamba 在视觉任务中已经展现出良好潜力,但现有研究大多将其作为通用骨干网络,尚未充分挖掘其在局部与全局上下文协同建模方面的优势,特别是在缺陷分

割这种对细节和上下文要求极高的任务中。本文针对这一空白,提出局部-全局 Mamba 混合编码器,目的是充分发挥 Mamba 在精度与效率上的综合优势。

2 本文方法

2.1 总体框架

本文提出的网络的整体架构如图 1 所示,在经典 U 型编解码结构的基础上,融合全局与局部建模、边界引导以及轻量化特征解码的多模块分割框架,实现对工业表面缺陷的高精度分割。整体结构主要包括 4 个主要模块:GLMamba 编码模块、EGFM 模块、混合瓶颈结构 HybridBottleneck 以及轻量级特征融合解码器 LightFusionDecoder。

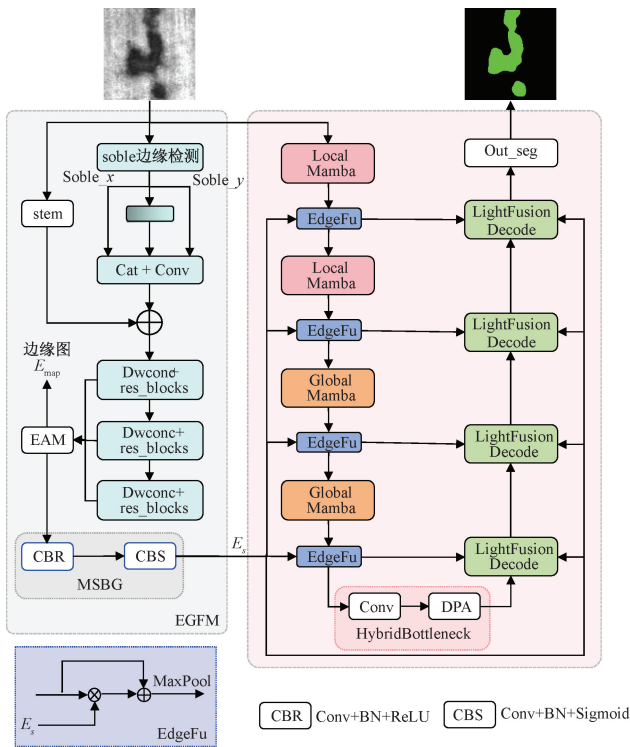


图 1 模型结构

Fig. 1 Network structure

在特征提取阶段,图像 X 通过两个并行分支进入网络。一个分支进入编码器部分,该部分通过将局部 Mamba 模块与全局 Mamba 模块由浅至深地协同嵌入到特征提取路径中,构建出一种分层的特征提取体系,在不同层级上逐步实现局部细节与全局上下文的建模。另一个分支进入 EGFM 提取边缘信息。该模块融合 Sobel 梯度与可变形卷积,以建模对方向敏感的边界特征,其输出包括用于监督的边缘预测图和用于后续融合的边缘信息。提取的边缘信息通过门控机制,与编码器-解码器的各层级特征进行融合,以此增强边缘信息在整个特征提取过程中的引导作用。

解码阶段采用 LightFusionDecoder,通过反卷积进行

上采样,并利用门控调制与双路径注意力机制(DualPath attention,DPA),逐级融合编码器输出的多尺度特征和边界引导信息,在逐步恢复空间信息的同时高效融合高低层信息,兼顾边缘细节的精确性与整体分割的一致性。通过 1×1 卷积输出多通道分割结果,与EGFM边缘分支输出的边缘预测图共同构成主分割任务与边界辅助任务的联合监督信号。

总体而言,该网络在保持结构轻量与计算效率的同时,通过Mamba的动态特征建模与边界引导机制实现了“语义-结构-边缘”三维度的特征协同建模,使模型在捕获全局上下文和局部细节信息方面实现平衡,能够在复杂表面缺陷分割任务中表现出更优的分割性能。

2.2 GLMamba 编码模块

为了同时捕获钢材表面缺陷的局部细节与全局上下文信息,受文献[30]的启发,在标准U-Net编码器基础上引入了GLMamba编码模块,如图2所示。该模块在浅层使用局部Mamba模块,通过对特征图进行局部划分,精确地提取缺陷的边缘、纹理等细节特征;深层使用全局Mamba模块,通过4个方向的扫描机制对特征图进行长程依赖建模,获取全局上下文信息以构建缺陷的整体形状与结构等高层语义结构。为进一步提升边界定位精度,各层还融合了EGFM模块提供的边缘先验信息。该设计实现了局部细节与全局语义的协同建模,显著增强了模型对缺陷区域的分割准确性及边界刻画能力。

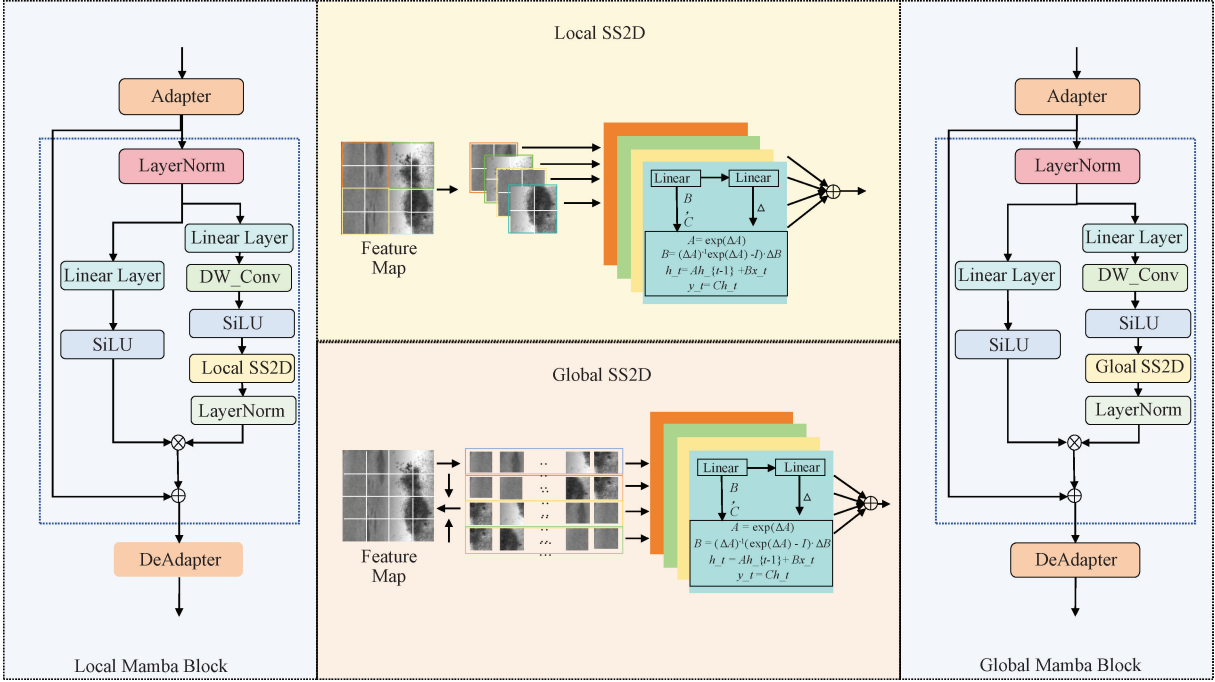


图 2 GLMamba 模型

Fig. 2 GLMamba model

1) 局部 Mamba 模块

局部Mamba模块通过对特征图进行区域划分,并逐区域进行二维选择性扫描(selective scan in 2D,SS2D)扫描建模,获取缺陷特征的细粒度特征信息。特别是对于金属表面中的小缺陷和对比度低的缺陷来说,可以对局部特征进行更精确地建模,一定程度上避免小缺陷的漏检并增强了其在复杂或干扰背景中准确分离缺陷区域的能力。

由图2可知,该模块使用双路径进行特征提取,路径1使用线性映射和激活函数增强通道表达能力;路径2首先利用线性变换和深度可分离卷积提取局部结构特征,接着输入到LocalSS2D进行分块细节信息提取。LocalSS2D将输入特征划分为4个象限 $\{Q_i\}$, $i \in \{1,2,3,4\}$,并在每个象限上独立进行序列建模。对每个象限通过SS2D进行局部时空建模,处理结果再重新拼接,形成局部增强后的特征

图。最后,两条路径的输出通过相乘进行融合,并与原始特征进行残差连接,以稳定训练并保留底层信息。此过程可用式(1)~(4)表示。

$$P_1 = \text{SiLu}(\text{Linear}_1(\mathbf{X}_{\text{norm}})) \quad (1)$$

$$P_2 = \text{SiLu}(\text{DwConv}(\text{Linear}_2(\mathbf{X}_{\text{norm}}))) \quad (2)$$

$$Q_i = \text{QuadrantSplit}(P_2), i \in \{1,2,3,4\} \quad (3)$$

$$\mathbf{X}_{\text{LM}} = \mathbf{X} + (P_1 \odot \text{LN}(\text{Merge}(\text{SS2D}(Q_i)))) \quad (4)$$

其中, P_1 为通道上下文向量,由路径1生成,用作门控信号, P_2 为空间特征图,由路径2生成,作为局部序列建模的输入, Q_i 为由 P_2 划分得到的第*i*个图像象限。DWConv($\cdot$)为分组可变形卷积操作,LN为层归一化,用于稳定训练过程,SS2D($\cdot$)为二维选择性状态空间模型,对每个象限进行局部序列建模, $\odot$ 表示逐元素乘积,SiLU($\cdot$)为线性激活函数。

2) 全局 Mamba 模块

全局 Mamba 模块通过双路径特征提取与四向扫描机制建立长程依赖关系,在全局尺度上捕获上下文信息。与局部特征相互补充,这种全局感知能力能够帮助网络更好地理解缺陷的整体形态与空间分布,缓解了由于感受野有限而导致的形状不完整或预测模糊等问题。如图 2 右侧所示,全局 Mamba 模块与局部 Mamba 模块一样采取双路径进行特征提取,不同的是,在线路 2 中使用的是 GlobalSS2D 对特征分别在 4 个方向上展开扫描:从左到右(LR)、从右到左(RL)、从上到下(TB)、从下到上(BT),最后再将 4 个方向的特征进行合并。这种全局感知能力与局部特征形成互补,使网络能精准理解缺陷的整体形态与空间分布,从而有效缓解因感受野有限导致的预测形状不完整或边界模糊等问题。此过程可用式(5)~(8)表示。

$$P_1 = \text{SiLu}(\text{Linear}_1(\mathbf{X}_{\text{norm}})) \quad (5)$$

$$P_2 = \text{SiLu}(\text{DwConv}(\text{Linear}_2(\mathbf{X}_{\text{norm}}))) \quad (6)$$

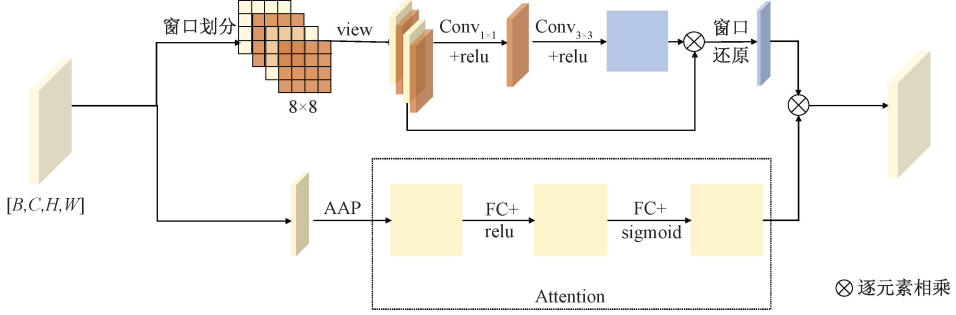


图 3 EAM 模型

Fig. 3 EAM model

EGFM 模块具体如图 1 左侧所示,该模块接收输入图像 $\mathbf{X} \in \mathbf{R}^{B \times C \times H \times W}$,其中 B 为 batch 大小, C 为通道数, H 和 W 分别为高和宽。首先输入特征经灰度化后通过固定参数的 Sobel 算子 ($\mathbf{K}_x, \mathbf{K}_y$) 分别计算水平方向梯度 $\mathbf{G}_x$ 与垂直方向梯度 $\mathbf{G}_y$,并进一步计算其梯度幅值 M ,以捕获底层的边缘强度与方向性。这些梯度信息拼接后通过 1×1 卷积形成边缘特征 $\mathbf{E}_{\text{sobel}}$ 。与此同时,输入图像经由一个 7×7 卷积层提取初始浅层特征 $\mathbf{E}_{\text{stem}}$ 。将边缘特征 $\mathbf{E}_{\text{sobel}}$ 与初始特征 $\mathbf{E}_{\text{stem}}$ 进行逐元素相加,形成边缘增强的初始特征 $\mathbf{E}$ 。为了进一步提升特征对不规则缺陷边缘的感知能力,在特征增强阶段引入 3 层分组可变形卷积,其偏移量可根据局部结构自适应调整,以感知不规则边缘的空间位置有效建模复杂形态。最终,3 个尺度的特征被送入 EAM 模块,对所提取的边界特征进行权重再校准,突出关键边缘响应。此过程可用式(9)~(13)表示。

$$\mathbf{G}_x = \mathbf{K}_x * \mathbf{X}_g, \mathbf{G}_y = \mathbf{K}_y * \mathbf{X}_g \quad (9)$$

$$\mathbf{E}_{\text{sobel}} = \text{Conv}_{1 \times 1}(\text{Concat}(\mathbf{G}_x, \mathbf{G}_y, \sqrt{\mathbf{G}_x^2 + \mathbf{G}_y^2})) \quad (10)$$

$$\mathbf{E} = \mathbf{E}_{\text{sobel}} + \text{Conv}_{7 \times 7}(\mathbf{X}) \quad (11)$$

$$\mathbf{E}_i = \text{Res}(\text{DWConv}(\mathbf{E}_{i-1})) \quad (12)$$

$$\mathbf{E} = \text{EAM}(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3) \quad (13)$$

$$\mathbf{S}_d = \text{ScanExpanding}(\mathbf{P}_2), d \in \{\text{LR}, \text{RL}, \text{TB}, \text{BT}\} \quad (7)$$

$$\mathbf{X}_{\text{GM}} = \mathbf{X} + (\mathbf{P}_1 \odot \text{LN}(\frac{1}{4} \text{Merge}(\text{SS2D}(\mathbf{S}_d)))) \quad (8)$$

其中, $\text{ScanExpanding}(\cdot)$ 表示四向序列扫描扩展操作, $\text{Merge}(\cdot)$ 表示多方向特征聚合操作。

2.3 边缘引导特征增强模块

在表面缺陷分割中,尤其是在边界特征里,包含着较多的几何与结构信息。传统的卷积特征提取方法,通常容易丢失边缘梯度方向和局部结构细节,导致模型对于缺陷边缘的预测不够准确。受文献[31]的启发,本文设计并实现了 EGFM 模块。该模块利用图像梯度信息来增强对缺陷边缘特征的感知能力。该模块通过 Sobel 算子增强边缘梯度处理能力,结合 3 层可变卷积自适应地生成多尺度的边缘特征信息,得到多尺度的缺陷边界信息。使用 EAM 增强对边界信息的选择,如图 3 所示。

其中, $\mathbf{X}_g$ 表示灰度处理后的图像, $\text{Concat}(\cdot)$ 表示通道维度拼接操作, $\text{DWConv}(\cdot)$ 表示分组可变形卷积操作, $\text{Res}(\cdot)$ 表示深度可分离残差块。

EAM 模块通过“多尺度特征拼接+方向注意力机制”对边界区域进行重点建模,从而提升分割结果在缺陷边缘区域的精确性与连续性。具体结构如图 3 所示,三尺度特征 $\{\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3\}$,首先沿通道维度拼接,利用 1×1 卷积进行通道压缩。为增强对边缘的感知能力,该模块引入了方向敏感的 StripAttention,通过自适应平均池化分别提取水平方向和垂直方向上的长程依赖信息。接着,将二者拼接并通过 1×1 卷积和 Sigmoid 激活得到方向注意力权重。利用该注意力对特征进行方向加权增强。经过方向增强后的特征 $\mathbf{E}_{\text{att}}$,最后经过 3×3 卷积提炼信息,得到边缘信息 $\mathbf{E}$ 。该特征一方面流入后续网络,另一方面通过一个 1×1 卷积与 Sigmoid 函数回归生成边缘预测图 $\mathbf{E}_{\text{map}}$ 。边缘预测图 $\mathbf{E}$ 将与真实边缘标注计算辅助损失,该损失作为显式的边界监督信号,引导主干网络学习更具判别力的边缘特征,从而最终提升分割结果的边界定位精度。此过程可用式(14)~(16)表示:

$$\mathbf{E} = \text{Conv}_{1 \times 1}(\text{Concat}(\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3)) \quad (14)$$

$$\mathbf{A} = \sigma(\text{Conv}_{1 \times 1}(\text{AvgPool}_h(\mathbf{E}) \parallel \text{AvgPool}_w(\mathbf{E}))) \quad (15)$$

$$\mathbf{E}_s = \text{Conv}_{3 \times 3}(\mathbf{E} \odot \mathbf{A}) \quad (16)$$

其中, $\mathbf{E}_1, \mathbf{E}_2, \mathbf{E}_3$ 分别表示 3 个不同尺度的输入边缘特征, AvgPool_h 和 AvgPool_w 分别表示高度和宽度方向的自适应平均池化, $\parallel$ 表示拼接操作, σ 为 Sigmoid 函数。

最后,多尺度边界门控机制(multiscale boundary gating, MSBG)接收 EAM 生成的多尺度边缘特征,通过插值与轻量卷积适配至多个尺度,利用深度可分离卷积提取边界结构信息,最终生成门控权重对编码器各阶段特征进行自适应调制,确保网络在下采样过程中始终保持对目标边缘的敏感性,其结构如图 2 左下所示,此过程可用式(17)~(18)表示。

$$\mathbf{B}_s = \text{BN}(\delta(\text{DWConv}_{3 \times 3}(\mathbf{E}_s))) \quad (17)$$

$$\mathbf{E}_s = \sigma(\text{BN}(\text{Conv}_{1 \times 1}(\mathbf{B}_s))) \quad (18)$$

其中, δ 为 ReLU 函数, σ 为 Sigmoid 函数。

2.4 双路径注意力解码模块

1) 双路径注意力机制

在获取下采样特征 $\mathbf{F}$ 后,本文设计了一种双路径注意力机制(dual-path attention, DPA),由局部窗口注意力与通道注意力两条并行路径构成,分别建模缺陷特征的空间显著性与通道依赖关系。DPA 输出的特征可以看作两条互补的信息,一条是局部窗口内的空间关联信息,另一条是包含全局通道的建模信息,实现对特征响应在空间域与通道域的协同优化。该模块被嵌入网络的瓶颈层及解码

器各上采样阶段,在关键层级实现特征响应的协同优化。

如图 4 所示,在具体实现中,局部路径采用“窗口划分-权重生成-重组还原”的 3 步策略,特征图被均匀划分为互不重叠的局部窗口,每个窗口的尺寸为 8×8 ,每个窗口 $\mathbf{F}_{ij}$ 内部独立执行 1×1 卷积、ReLU 激活函数与 3×3 卷积,随后通过 Sigmoid 激活生成空间注意力图,所有窗口的权重图经重组后与原始输入特征逐元素相乘,实现局部显著性增强,有效抑制复杂背景噪声。通道路径则借鉴经典的压缩-激励思想,在通道维度上进行信息重标定。先对整幅特征图进行全局平均池化以获得通道信息 $\mathbf{Z} \in \mathbf{R}^{B \times C \times 1 \times 1}$,再经过两层全连接构成的瓶颈结构生成通道权重向量,该向量通过广播机制作用于原特征,完成跨通道的自适应校准。最终,两条路径的输出通过逐元素乘法融合,形成兼具局部精细与全局语义的精炼特征,此过程可用式(19)~(21)表示。

$$\mathbf{S}_{ij} = \mathbf{F}_{ij} \odot \sigma(\text{Conv}_{3 \times 3}(\delta(\text{Conv}_{1 \times 1}(\mathbf{F}_{ij})))) \quad (19)$$

$$\mathbf{A} = \mathbf{F} \odot \sigma(\mathbf{W}_2(\delta(\mathbf{W}_1(\text{AAP}(\mathbf{F})))))) \quad (20)$$

$$\mathbf{F}_{\text{out}} = \mathbf{A} \odot \mathbf{S} \quad (21)$$

其中, $\mathbf{F}_{ij} \in \mathbf{R}^{B \times C \times 8 \times 8}$ 为第 (i, j) 个 8×8 局部窗口特征, $\text{AAP}(\cdot)$ 为自适应平均池化, $\mathbf{W}_1 \in \mathbf{R}^{C/r \times C}$, $\mathbf{W}_2 \in \mathbf{R}^{C \times C/r}$ 为全连接层权重, r 为通道压缩比, $\mathbf{S}$ 为所有局部窗口权重图经重组后的完整空间注意力图, $\mathbf{A}$ 为通道注意力加权后的特征图。

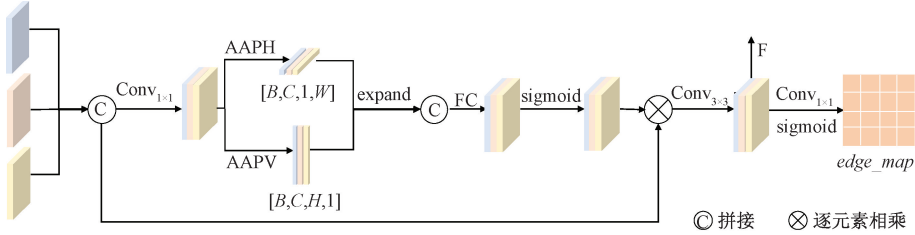


图 4 DPA 模块

Fig. 4 DPA model

2) 轻量化融合解码器

瓶颈层是编码器与解码器之间的关键连接组件,它包含了特征压缩,语义增强以及上下文信息建模的功能。在以往的研究中,瓶颈层通常由一系列的卷积层组成,这些简单的卷积层难以捕捉特征信息中较长的依赖关系,并且在识别关键语义方面也存在缺陷。为了解决上述问题,本文提出了一种名为 HybridBottleneck 的处理方式,通过将局部卷积增强与双路径注意力机制进行结合,实现了对局部特征信息提取和全局上下文信息的有效整合。

如图 1 所示,HybridBottleneck 主要由两个核心模块组成:局部卷积增强模块和 DPA 模块。首先,输入特征经过 DoubleConv 模块处理,获得空间表征能力。DoubleConv 包含两层卷积运算,以实现通道维度的扩展及局部特征的强化,从而提升特征的空间表征能力。随后,经过初步增强的特征传入双路径注意力模块,将局部细节

与全局上下文信息进行有效融合,此过程可用式(22)表示。

$$\mathbf{F}_{\text{out}} = \text{DPA}(\text{DoubleConv}(\mathbf{X})) \quad (22)$$

3) 解码器双路径注意力融合机制

解码器在特征重建过程中面临着多尺度语义信息融合与空间细节恢复的双重挑战。为此,本文设计了轻量化融合解码器(LightFusionDecoder),该模块通过门控增强、双路径注意力与深度可分离卷积的级联操作,实现编码器跳跃连接特征与解码器上采样特征的高效融合,该融合机制的关键在于将门控增强前置于注意力计算,使得双路径注意力能够在已经过边缘强化的特征空间中运作,从而避免注意力机制在原始低对比度特征上的判别困难。具体结构如图 5 所示。

在解码路径中,输入的深层特征 $\mathbf{F}_{\text{up}}$ 经过转置卷积进行上采样后与跳跃连接传递的浅层特征 $\mathbf{F}_{\text{skip}}$,分别和来自

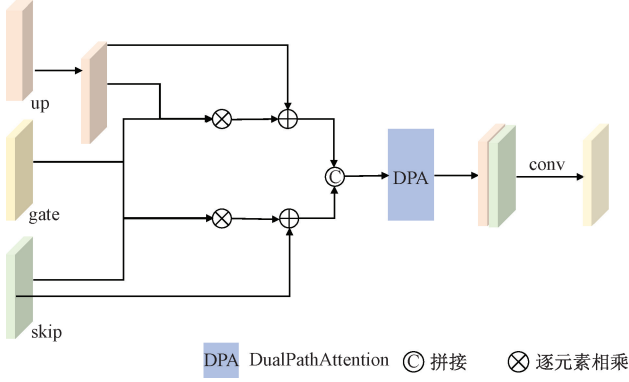


图 5 LightFusionDecoder 模型

Fig. 5 LightFusionDecoder model

EGFM 模块的边界门控信号 $\mathbf{G}$ 以残差形式进行融合,同步增强两类输入在边缘区域的响应。该设计在保留原始特征中高层语义信息完整性的同时,强化了边缘区域的响应,有助于模型更稳定地定位缺陷边界。随后,增强后的两路特征沿通道维度进行拼接,形成融合表征,并输入到 DPA 模块进行特征信息处理,实现了语义信息与空间特征之间的有效衔接,此过程可用式(23)~(25)表示。

$$\mathbf{F}_{\text{skip}'} = \mathbf{F}_{\text{skip}} \odot (1 + \mathbf{G}) \quad (23)$$

$$\mathbf{F}_{\text{up}'} = \mathbf{F}_{\text{up}} \odot (1 + \mathbf{G}) \quad (24)$$

$$\mathbf{F}_{\text{seg}} = \text{Conv}(\text{DPA}(\text{cat}(\mathbf{F}_{\text{skip}'}, \mathbf{F}_{\text{up}'}))) \quad (25)$$

2.5 损失函数

在缺陷分割任务中,为了同时监督缺陷区域的分割精度和边界定位的准确性,本研究采用了分割损失和边界损失两种损失,具体可用式(26)~(27)表示。

$$\mathcal{L}_{\text{seg}} = 0.6 \times \mathcal{L}_{\text{Dice}} + 0.4 \times \mathcal{L}_{\text{Dice}} \quad (26)$$

$$\mathcal{L}_{\text{total}} = \alpha \cdot \mathcal{L}_{\text{seg}} + \beta \cdot \mathcal{L}_{\text{edge}} \quad (27)$$

其中, α 为 1.0, β 为 0.7。分割损失采用 BCE-Dice 混合损失函数。边界损失采用与分割损失中 Dice 损失相同的数学形式,但仅针对边界区域进行计算。具体可用式(28)~(29)表示。

$$\mathcal{L}_{\text{dice}}(p, g) = 1 - \frac{2 \sum_i^{H \times W} p_i g_i}{\sum_i^{H \times W} p_i^2 + \sum_i^{H \times W} g_i^2} \quad (28)$$

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \ln(p_i) + (1 - y_i) \ln(1 - p_i)] \quad (29)$$

其中, p_i 和 g_i 分别为预测值和真实值。

3 实验结果和分析

3.1 数据集和实验设置

工业表面缺陷分割面临的挑战可直观地从典型缺陷图像中体现,包括形态复杂、对比度低及微小缺陷等情况,如图 6 所示。为验证本研究所提模型的有效性,本研究采

用 3 个公开的数据集对所提出的缺陷分割模型进行训练与验证。按照 7:2:1 的比例将数据集随机划分为训练集、验证集、测试集,并将输入图像像素统一转变为 256×256 。3 个数据集分别是 NEU-Seg 数据集^[32]、MT-Defect 数据集^[33]、FSSD-12 数据集^[34]。

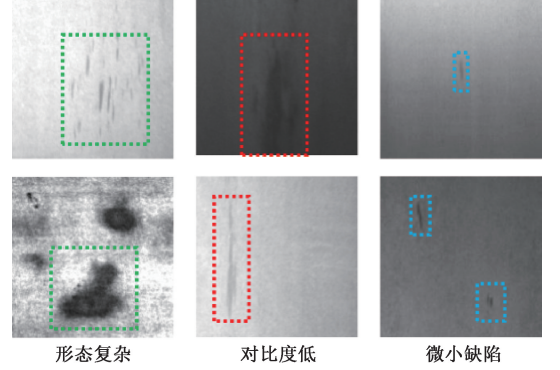


图 6 缺陷分割挑战

Fig. 6 Challenges in defect segmentation

NEU-Seg 数据集共包含 3 600 幅由热轧带钢产生的缺陷图像,所有图像尺寸均为 200×200 pixels。数据集涵盖了 3 种典型的表面缺陷类型夹杂物、斑块和划痕,且每种缺陷均提供了 1 200 个标注样本。

MT-Defect 数据集包含了磁性瓦片表面的 392 幅有缺陷图像以及 952 幅无缺陷图像,其分辨率大小不一。该数据集涵盖了 5 种缺陷,分别为气孔、裂纹、磨损、断裂和不均匀。

FSSD-12 数据集是一个新发布的数据集,用于纯带钢表面缺陷小样本分割,包含了 12 个种类的缺陷类型:磨损遮蔽、铁皮灰、液体、氧化皮、油斑、水斑、斑块、冲孔、红铁皮、滚印、划痕和夹杂物。每种缺陷包含了 50 幅图像,像素为 200×200 pixels。

在实验设计方面,本研究使用 AdamW 优化器进行模型训练,其初始学习率设置为 1×10^{-3} ,权重衰减系数为 1×10^{-4} 。以有效防止过拟合,采用基于余弦退火策略的学习率调度器,并设置了动态调整机制。所有实验的批量大小(BatchSize)统一设置为 8。为了提升模型的泛化能力和鲁棒性,在训练阶段实施了多种的数据增强方法,具体包括随机水平翻转、90°旋转、亮度与对比度调整等方法。

3.2 评价指标

在本实验中,为了全面、客观地评估所提出模型在缺陷分割任务上的性能,主要使用了 4 个核心指标,交并比(IoU)、Dice 系数、准确率(Precision)和召回率(Recall),另外使用参数量(Params)来衡量模型的复杂性。本文模型 Params 是在输入图像像素为 256×256 pixels 分辨率的情况下进行计算的。指标计算可用式(30)~(33)表示。

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (30)$$

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}} \tag{31}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{32}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{33}$$

其中, TP、FP、FN 分别代表真阳性、假阳性、假阴性。

3.3 对比实验

为客观评估该模型的性能,本研究在 3 个数据集上进行了对比实验。在相同的环境配置下,与主流分割模型 (U-Net, AttUNet, UNet++, TransUNet, RS³ mamba, Vmunet, MANet, DBRNet, DDSNet) 进行了对比实验,所有实验采取了 MIOU、Dice, Recall, Precision 性能指标进行对比评价。实验结果如表 1、2 所示。

表 1 NEU-Seg 数据集上各模型指标数据
Table 1 Segmentation results of each model on the NEU-Seg dataset (%)

模型	Params	MIoU	Dice	Recall	Precision
U-Net ^[1]	20.23	75.81	84.92	85.52	85.52
AttUNet ^[2]	34.88	76.09	84.88	88.04	84.49
unet++ ^[3]	47.20	75.95	84.85	88.28	84.75
transunet ^[20]	54.86	76.05	85.23	81.38	87.21
DBRNet ^[35]	3.33	76.90	85.77	88.04	87.05
RS ³ mamba ^[36]	165.27	77.66	86.08	85.13	88.77
Vmunet ^[37]	35.90	76.6	85.41	88.07	85.48
MANet ^[38]	35.86	79.36	87.13	86.96	89.85
DDSNet ^[39]	0.30	75.98	86.55	83.48	86.80
本文	26.51	79.80	87.21	87.07	90.02

表 2 MT-Defect 数据集和 FSSD-12 数据集上各模型指标数据
Table 2 Segmentation results of each model on the MT-Defect and FSSD-12 datasets (%)

模型	Params	MT-Defect		FSSD-12	
		MIoU	Dice	MIoU	Dice
U-Net	20.23	72.21	83.18	61.97	75.53
AttUNet	34.88	51.23	64.99	69.61	81.03
unet++	47.20	63.18	74.77	71.87	82.73
transunet	54.86	73.93	76.81	64.85	77.59
DBRNet	3.33	68.78	75.24	50.53	65.42
RS ³ mamba	165.27	70.74	82.54	74.65	84.80
Vmunet	35.90	61.71	72.69	68.45	80.38
MANet	35.86	65.91	66.53	62.35	75.31
DDSNet	0.30	73.83	85.47	62.46	75.43
本文	26.51	79.34	87.49	75.87	85.71

从表 1 可以看出本文模型在 NEU-Seg 数据集上的各

项分割性能指标展现出显著优越性。本文模型在 MIoU、Dice、Recall、Precision 这 4 项指标分别达到了 79.80%、87.21%、87.07%、90.02%。本文模型通过融合 EGFM 模块提供的边界信息与 GLMamba 编码器混合建模,结合双路径注意力机制,对细节和全局信息进行联合精细化处理,对缺陷区域的定位更加精确,对边缘信息的识别更加敏锐,从而整体提升了分割性能并减少了小缺陷的漏检。

与基于 Transformer 架构的模型相比,本文模型参数量大大降低,并在所有评价指标上均显著领先,MIoU 高出 3.75%,Dice 系数提升了 1.44%。这一结果表明,本文提出的 GLMamba 编码器在保持轻量化的同时,有效地捕获缺陷细节信息,避免了 Transformer 架构因高复杂度导致的计算冗余问题。在与同样采用 Mamba 架构的模型对比中,本文模型以较少的参数取得更好的分割效果。与效果较好的 RS³ mamba 相比较,在参数量减少的同时 mIoU 提升 3.82%,精确率提升 1.13%,表明解码器的双路径注意力机制能够有效抑制背景噪声的影响,减少误检现象。总体而言,本文方法在模型参数量与分割性能之间实现了较好的平衡;以中等规模的参数量,获得了更高的分割精度。

根据表 2 所示的实验结果,本文模型在 MT-Defect 与 FSSD-12 数据集上分别取得 79.34%与 75.87%的 IoU 值,以及 87.49%与 85.71%的 Dice 值。其分割的精度略低于 RS³ mamba,但是其参数量大大减少。相较于同等规模的主流模型,本文模型以更少的参数量在 MIoU 和 Dice 指标上均远远领先 AttUNet 和 MANet。此外,与参数量相近的 U-Net 相比,MIoU 提升了 13.43%。实验数据进一步表明,本文提出的模型在保证参数量的同时,有效提升了缺陷的分割精度。此外,模型在多个数据集上保持稳定的性能,在一定程度上证明了该模型的泛化能力。

为评估本模型在不同缺陷尺度下的分割性能,将测试集样本按照缺陷所占像素面积进行分类:<200 pixel 的缺陷为 Small,其面积占比不足 0.31%;200~700 pixel 的缺陷为 Medium,>700 pixel 的缺陷为 Large。该划分方式客观反映了缺陷尺寸差异,且不会影响模型训练过程,仅用于评价阶段的细粒度性能分析。实验结果如表 3 所示,本文模型在 3 个尺度上均取得了最优的分割性能,尤其在 Small 级别上相较于基线 UNet 分别提升了 10.67%的 MIoU 和 12.27%的 Dice,表明该模型在微小缺陷与弱边界区域的表征上具有显著优势。

为系统评估所提模型与主流方法在特征表达能力上的差异,本文对网络最后一层输出的特征进行了可视化分析。实验从 NEU-Seg 数据集中每类缺陷随机选取一张样本,并引入 3 幅包含多类别缺陷的复杂场景图像以增强对比的挑战性。如图 7 所示,不同方法的特征图对比表明:本文方法在缺陷边界区域呈现出更清晰、连续的响应分布,尤其在划痕类缺陷及小尺度目标的分割中,能够更准确地还原细节结构并保持区域内部的一致性,体现出更强

表 3 NEU-Seg 数据集多尺度缺陷各模型指标数据

Table 3 Multi-scale defect index data of each model in the NEU-Seg dataset (%)

模型	Small		Medium		Large	
	MIoU	Dice	MIoU	Dice	MIoU	Dice
U-Net	34.25	43.86	57.97	68.04	78.89	86.03
AttUNet	34.42	44.46	58.51	68.81	79.34	86.16
unet++	34.62	44.42	59.67	69.97	79.17	86.10
transunet	34.55	44.03	58.92	68.27	79.18	86.19
DBRNet	34.77	45.69	60.16	69.09	80.58	86.47
RS ³ mamba	36.01	46.71	60.68	70.86	81.31	87.45
Vmunet	35.19	42.15	58.84	69.23	80.73	87.22
MANet	41.76	52.29	63.43	73.18	82.54	88.35
DDSNet	34.52	44.95	58.10	68.61	77.84	85.22
本文	44.92	56.13	65.95	75.72	82.71	88.58

的局部敏感性与语义判别能力。

3.4 消融实验

为验证本文所提出各模块的有效性,本研究在 NEU-Seg 数据集上基于基线模型(Baseline)逐步添加各模块进行消融实验,结果如表 4 所示。基线模型的 MIoU 为 75.09%,Dice 系数为 84.22%。在单独引入各模块后,模型性能均有稳定提升,验证了各模块设计的合理性与有效性。进一步地,当同时集成 GLMamba 编码器与 EGFM 模块时,模型性能取得显著提升,MIoU 提高了 3.94%,Dice 系数提升了 2.8%,表明全局-局部上下文建模与边缘特征引导机制具备良好的互补性。在此基础上继续引入混合瓶颈结构与融合解码器,模型性能得到进一步优化,在多个评价指标上均实现均衡提升。

实验结果表明,本文所提出的各模块均能有效增强模型对工业缺陷的感知与分割能力,尤其在全部模块协同工

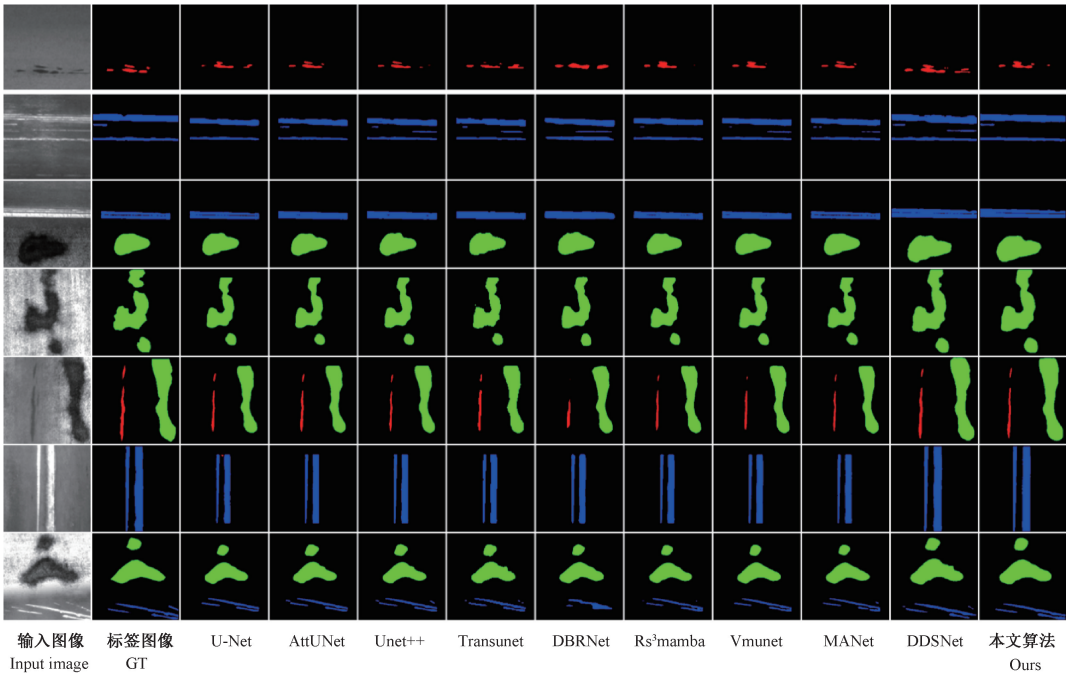


图 7 NEU_seg 分割结果可视化

Fig. 7 Visualisation of NEU_seg segmentation results

Table 4 Modular ablation experiments at NEU_seg (%)						(%)			
Baseline	GLmamba	EGFM	MSGb	HybridBottleneck	Fusiondecoder	MIoU	Dice	Precision	Recall
✓						75.09	84.22	87.77	84.06
✓	✓					76.72	85.41	88.47	85.23
✓		✓				75.91	84.94	84.09	84.38
✓			✓			75.36	84.37	83.62	84.47
✓		✓	✓			76.70	85.64	89.83	84.77
✓				✓		75.63	84.79	87.94	84.82
✓	✓	✓	✓			79.03	87.02	89.46	87.01
✓		✓	✓	✓		77.03	85.64	88.73	85.36
✓	✓	✓	✓		✓	79.39	87.13	90.31	86.74
✓	✓	✓	✓	✓	✓	79.80	87.21	90.02	87.07

作时,模型在保持较高精确率的同时,也展现出优异的召回性能,在缺陷边界的分割精度和内部区域的语义一致性方面均取得了显著改善。

4 结 论

本研究针对工业表面缺陷检测中缺陷形态多变、对比度低及小目标易漏检等挑战,提出了一种基于边缘引导与Mamba的语义分割网络。该网络采用编码-解码架构,在编码阶段设计了GLMamba编码器,通过分层部署局部Mamba块与全局Mamba块,有效协同建模图像的全局上下文信息与局部边界结构。为进一步增强模型对缺陷边界的感知能力,构建了边缘特征提取模块EGFM,融合多尺度特征与Sobel梯度信息,显式生成边缘特征与边界图,并引入MSBG将边缘特征转化为门控权重,动态调制编码器特征以保持边缘敏感性。在解码阶段,设计了双路径注意力解码模块,通过并行处理空间细节与通道依赖,在EGFM边缘引导的基础上实现编码器跳跃连接特征与解码器上采样特征的高效融合。在NEU-Seg钢材缺陷数据集、MT-Defect磁片数据集与FSSD-12数据集上的实验结果表明,本文方法在分割精度上优于多种主流分割模型,尤其在复杂缺陷边界保持与小目标缺陷识别方面表现出显著优势,验证了所提网络架构与各模块的有效性。未来的研究工作将聚焦于进一步优化模型的计算效率,探索轻量化设计,以促进其在实时工业检测场景中的实际应用。

参考文献

- [1] Ronneberger O, Fischer P, Brox T. U-Net: Convolutional networks for biomedical image segmentation [C]//Medical Image Computing and Computer-Assisted Intervention (MICCAI). Cham: Springer, 2015: 234-241.
- [2] Oktay O, Schlemper J, Folgoc L L, et al. Attention U-Net: Learning where to look for the pancreas[PP/OL]. V7. (2018-04-11) [2025-11-25]. <https://doi.org/10.48550/arXiv.1804.03999>.
- [3] Zhou Z W, Rahman S M M, Tajbakhsh N, et al. UNet++: A nested U-Net architecture for medical image segmentation [C]//Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. Cham: Springer, 2018: 3-11.
- [4] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16×16 words: Transformers for image recognition at scale [C]//International Conference on Learning Representations (ICLR), 2021.
- [5] Gu A, Dao T. Mamba: Linear-time sequence modeling with selective state spaces [C]//First Conference on Language Modeling (COLM), 2024.
- [6] Zhu L H, Liao B C, Zhang Q, et al. Vision mamba: Efficient visual representation learning with bidirectional state space model [C]//41st International Conference on Machine Learning. Proceedings of Machine Learning Research, 2024, 235: 62429-62442.
- [7] 王俞霏, 杜弘志, 胡蕴博, 等. 基于多尺度特征融合网络的电阻点焊缺陷精确定位与分割方法 [J]. 仪器仪表学报, 2025, 46(7): 202-213.
Wang Yufei, Du Hongzhi, Hu Yunbo, et al. Precise localization and segmentation method for resistance spot welding defects based on multi-scale feature fusion network [J]. Chinese Journal of Scientific Instrument, 2025, 46(7): 202-213.
- [8] 邹凯鑫, 张自嘉, 孙伟, 等. 改进 U 型网络的路面缺陷图像分割算法 [J]. 电子测量与仪器学报, 2024, 38(8): 15-25.
Zou Kaixin, Zhang Zijia, Sun Wei, et al. Improving the road defect image segmentation algorithm of U-Net [J]. Journal of Electronic Measurement and Instrumentation, 2024, 38(8): 15-25.
- [9] Stern M L, Schellenberger M. Fully convolutional networks for chip-wise defect detection employing photoluminescence images: Efficient quality control in LED manufacturing [J]. Journal of Intelligent Manufacturing, 2021, 32(1): 113-126.
- [10] Chen L, Papandreou G, Kokkinos I, et al. DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.
- [11] Chen H, Min B W. A high-precision segmentation network for industrial surface defect detection [J]. AIP Advances, 2025, 15(5), DOI:10.1063/5.0274903.
- [12] 陈光耀, 陈田, 高学海, 等. 跨层融合的轻量化太阳能电池片缺陷分割方法 [J]. 电子测量技术, 2025, 47(24): 118-127.
Chen Guangyao, Chen Tian, Gao Xuehai, et al. C2LA-U2-Net: Lightweight defect segmentation method for solar cells with cross-layer fusion [J]. Electronic Measurement Technology, 2025, 47(24): 118-127.
- [13] 周炜杰, 李智, 张绍荣, 等. 融合多级注意力与多尺度信息的轨枕缺陷分割网络 [J]. 电子测量与仪器学报, 2025, 39(7): 140-150.
Zhou Weijie, Li Zhi, Zhang Shaorong, et al. Railway defect segmentation network with multi-level attention and multi-scale information fusion [J]. Journal of

- Electronic Measurement and Instrumentation, 2025, 39(7): 140-150.
- [14] Zhang Y J, Wu J X, Li Q Z, et al. Beyond crack: Fine-grained pavement defect segmentation using three-stream neural networks[J]. IEEE Transactions on Intelligent Transportation Systems, 2021, 23(9): 14820-14832.
- [15] Liu Z, Lin Y T, Cao Y, et al. Swin transformer: Hierarchical vision transformer using shifted windows[C]//IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 10012-10022.
- [16] Zheng S X, Lu J C, Zhao H S, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers [C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021: 6881-6890.
- [17] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: Simple and efficient design for semantic segmentation with transformers[J]. Advances in Neural Information Processing Systems, 2021, 34: 12077-12090.
- [18] Zhang G W, Lu Y, Jiang X H, et al. LGGFormer: A dual-branch local-guided global self-attention network for surface defect segmentation [J]. Advanced Engineering Informatics, 2025, 64: 103099.
- [19] Liu X, Gao P, Yu T, et al. CSWin-UNet: Transformer UNet with cross-shaped windows for medical image segmentation[J]. Information Fusion, 2025, 113: 102634.
- [20] Liu G Y, Chen Y, Ye J, et al. A transformer neural network based framework for steel defect detection under complex scenarios [J]. Advanced Engineering Software, 2025, 202: 103872.
- [21] Zhu H J, Huang J H, Chen K, et al. MultiPI-TransBTS: A multi-path learning framework for brain tumor image segmentation based on multi-physical information[J]. Computers in Biology and Medicine, 2025, 191: 110148.
- [22] Gu A, Goel K, Ré C. Efficiently modeling long sequences with structured state spaces[C]//International Conference on Learning Representations (ICLR), 2022.
- [23] Gu A, Johnson I, Goel K, et al. Combining recurrent, convolutional, and continuous-time models with linear state space layers[J]. Advances in Neural Information Processing Systems, 2021, 34: 572-585.
- [24] Liu J R, Yang H, Zhou H Y, et al. Swin-UMamba: Adapting Mamba-based vision foundation models for medical image segmentation [J]. IEEE Transactions on Medical Imaging, 2024, 44(10): 3898-3908.
- [25] Li G J, Huang Q H, Wang W, et al. Selective and multi-scale fusion Mamba for medical image segmentation[J]. Expert Systems with Applications, 2025, 261: 125518.
- [26] Sun J D, Chen K X, Wu X S, et al. MSM-UNet: A medical image segmentation method based on wavelet transform and multi-scale Mamba-UNet [J]. Expert Systems with Applications, 2025, 288: 128241.
- [27] Tang K, Li Y, Zi B, et al. A multimodal unsupervised coating defect detection method based on dual-branch hybrid CNN-Mamba U-Net [J]. IEEE Transactions on Instrumentation and Measurement, 2025, 74: 1-17.
- [28] Zhang J M, Li D W, Zeng Z G, et al. Dual-branch crack segmentation network with multi-shape kernel based on convolutional neural network and Mamba[J]. Engineering Applications of Artificial Intelligence, 2025, 150: 110536.
- [29] Zuo X, Sheng Y, Shen J F, et al. Topology-aware Mamba for crack segmentation in structures [J]. Automation in Construction, 2024, 168: 105845.
- [30] Ji Z X, Zou B J, Kui X Y, et al. Global and local Mamba network for multi-modality medical image super-resolution [J]. Pattern Recognition, 2026, 173: 112888.
- [31] Ni Y, Liu J H, Cui J, et al. Edge guidance network for semantic segmentation of high-resolution remote sensing images[J]. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2023, 16: 9382-9395.
- [32] Bao Y Q, Song K C, 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: 1-11.
- [33] Huang Y B, Qiu C Y, Yuan K, et al. Surface defect saliency of magnetic tile[J]. The Visual Computer, 2020, 36: 85-96.
- [34] Feng H, Song K C, Cui W Q, et al. Cross position aggregation network for few-shot strip steel surface defect segmentation [J]. IEEE Transactions on Instrumentation and Measurement, 2023, 72: 1-10.
- [35] Zhang T P, Wei X M, Wu X M, et al. DBRNet: Dual-branch real-time segmentation network for metal defect detection [C]//Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2023: 422-434.
- [36] Ma X P, Zhang X K, Pun M O. RS3-Mamba: Visual state space model for remote sensing image semantic

segmentation [J]. IEEE Geoscience and Remote Sensing Letters, 2024, 21: 1-5.

[37] Ruan J C, Li J C, Xiang S C. VM-UNet: Vision Mamba UNet for medical image segmentation[J]. ACM Transactions on Multimedia Computing, Communications and Applications, 2025. <https://doi.org/10.1145/3767748>.

[38] Fan T L, Wang G L, Li Y, et al. MA-Net: A multi-scale attention network for liver and tumor segmentation [J]. IEEE Access, 2020, 8: 179656-179665.

[39] Yin Z Y, Qin L, Han G J, et al. DDSNet: Deep dual-branch networks for surface defect segmentation[J].

IEEE Transactions on Instrumentation and Measurement, 2024, 73: 1-16.

作者简介

陈圆圆, 硕士研究生, 主要研究方向为图像分割、表面缺陷。

E-mail: chen_yuan18@163.com

张笃振(通信作者), 博士, 副教授, 硕士生导师, 主要研究方向为图像处理、模式识别、机器学习。

E-mail: zhduzhen@jsnu.edu.cn

王斯豪, 硕士研究生, 主要研究方向为图像分割。

E-mail: 1826653404@qq.com

杨昌昌, 硕士研究生, 主要研究方向为缺陷检测。

E-mail: yangchangchang88@163.com