

DOI: 10.20271/j.cnki.1673-9833.2026.1014

基于快速傅里叶变换的裂缝分割算法

王永明, 胡仕刚, 伍绍兵, 梁思奇

(湖南科技大学 信息与电气工程学院, 湖南 湘潭 411201)

摘要: 裂缝分割是现代民用基础设施维护的关键一环, 快速准确地检测出裂缝至关重要。故针对基于卷积模型在远距离建模方面的不足, 以及基于 Transformer 模型在局部特征提取不足和计算复杂度较高的缺点, 构建了一个基于频率增强的 U 型结构网络 (FE-UNet)。首先, 提出一个频率增强注意力模块 (FEA), 通过快速傅里叶变换将图像从空间域映射到频域, 降低计算复杂度, 并在频域中引入卷积模块以突出裂缝特定频率成分, 从而增强特征表示。此外, 提出了一个局部增强前馈网络 (LE-FFN), 通过整合来自多个分支下的多尺度语义信息, 提高特征的局部交互能力。实验结果表明, 所提出网络模型的 $mIoU$ 和 F_1 值最高达 82.38% 和 79.48%, 为道路健康监测提供了有力支持。

关键词: 裂缝分割; 快速傅里叶变换; 多尺度语义信息; 局部交互

中图分类号: TP391.41

文献标志码: A

文章编号: 1673-9833(2026)01-0102-07

引文格式: 王永明, 胡仕刚, 伍绍兵, 等. 基于快速傅里叶变换的裂缝分割算法 [J]. 湖南工业大学学报, 2026, 40(1): 102-108.

Crack Segmentation Algorithm Based on Fast Fourier Transform

WANG Yongming, HU Shigang, WU Shaobing, LIANG Siqi

(School of Information and Electrical Engineering, Hunan University of Science and Technology,
Xiangtan Hunan 411201, China)

Abstract: Crack segmentation is a critical task in the maintenance of modern civil infrastructure, where rapid and accurate crack detection is essential. In view of the limitations of convolution-based models in long-range modeling and the flaws of Transformer-based models in local feature extraction and high computational complexity, a frequency-enhanced U-shaped network (FE-UNet) has thus been proposed. First, a frequency-enhanced attention module (FEA) is introduced, which adopts fast Fourier transform to map the image from the spatial domain to the frequency domain, thus reducing computational complexity. A convolution module is subsequently introduced in the frequency domain to highlight crack specific frequency components, thereby enhancing feature representation. In addition, a local-enhanced feed-forward network (LE-FFN) is proposed to enhance the local interaction capability of features by integrating multi-scale semantic information from multiple branches. Experimental results show that the proposed network achieves a maximum $mIoU$ of 82.38% and F_1 of 79.48%, providing strong support for road health monitoring.

Keywords: crack segmentation; fast Fourier transform; multi-scale semantic information; local interaction

收稿日期: 2025-03-16

基金项目: 湖南省自然科学基金资助项目 (2025JJ70638, 2024JJ6224); 湖南省教育厅科学研究基金资助项目 (24C0233)

作者简介: 王永明, 男, 湖南科技大学硕士生, 主要研究方向为深度学习与图像分割,

E-mail: 23020402012@mail.hnust.edu.cn

通信作者: 胡仕刚, 男, 湖南科技大学教授, 硕士生导师, 主要研究方向为深度学习与图像处理,

E-mail: hsg99528@126.com

1 研究背景

在现代工程实践中, 裂缝检测是确保基础设施质量和安全的重要环节, 尤其在地震和其他极端条件后, 准确识别和评估裂缝的结构健康显得尤为迫切。传统的裂缝检测方法往往依赖手动检查, 不仅费时费力, 也增加了安全风险。经典的图像分割方法有基于边缘检测、区域生长、阈值分割等, 然而这些方法很大程度上依赖于先验知识。而基于机器学习, 如随机森林和支持向量机等方法, 只能适应特定环境条件, 鲁棒性较差。

随着计算机视觉和深度学习技术的不断发展, 基于深度学习的方法已经成为裂缝分割领域的研究热点。卷积神经网络 (convolutional neural network, CNN) 由于其归纳偏置优势, 尤其擅长提取局部特征, 因此被广泛用于裂缝分割任务中。例如, Cheng J. R. 等^[1]提出了一种端到端的 U 型编码器-解码器裂缝分割网络, 其先对编码器部分进行深度特征提取, 再由解码器部分逐渐进行特征重建, 实现了端到端的裂缝分割, 省去了传统方法中繁琐的预处理步骤。Wang Y. Y. 等^[2]通过引入矩形卷积金字塔和边缘增强机制, 有效融合了图像边缘信息, 提高了检测精度。然而, 尽管卷积结构对局部特征提取效果显著, 但因其感受野受限, 难以建模远距离依赖关系, 导致在处理细长或结构复杂的裂缝时, 存在一定的局限。为了解决这一问题, 越来越多的研究者将 Transformer 引入裂缝分割网络, 以增强全局建模能力。Liu H. J. 等^[3]提出了 CrackFormer 网络, 将 Transformer 编码器嵌入 U 型结构中, 提升了对全局上下文的建模能力。Xiang C. 等^[4]进一步设计了双分支结构, 其融合了卷积和 Transformer 模块, 兼顾局部与全局建模能力。然而, 引入 Transformer 虽能拓展感受野, 却带来了参数量大、计算复杂度高等问题, 尤其在轻量化部署场景中, 存在较大挑战。此外, 一些研究者尝试通过引入多尺度上下文信息和注意力机制来提升分割鲁棒性。Wen X. 等^[5]提出结合多尺度特征提取与交叉注意力机制的方法, 有效提升了对边缘模糊区域的裂缝识别能力。综上所述, 现有方法在局部特征提取或远距离依赖建模方面各有优势, 但也存在性能瓶颈、计算开销高和细节信息提取不足等局限。为应对这些问题, 本文提出了一种基于频率增强的 U 型结构网络 (FE-UNet), 融合频域处理机制与编码器-解码器架构的优点, 以在保持模型轻量化的同时提升裂缝分割的准确性和分割效率。

尽管已有多种深度学习模型被应用于裂缝分割

任务中, 但在实际应用中, 裂缝图像的获取与处理常常伴随着复杂背景干扰, 例如环境噪声、光照变化以及裂缝形态的多样性。这些不确定因素加剧了分割任务的难度, 也对模型的泛化能力和稳定性提出了更高要求。最近, 频域特征在深度学习和图像处理领域受到广泛关注。快速傅里叶卷积利用基于快速傅里叶变换 (fast Fourier transform, FFT) 的特征学习概念, 进行全局到全局的特征映射, 提供令人印象深刻的性能^[6]。通过傅里叶变换, 可以将输入图像的频谱信息拆分为高频和低频成分。高频信息蕴含着图像边缘、纹理等细节信息, 低频信息包含了整体结构和全局信息。频谱分析可以揭示哪些频率成分对裂缝检测至关重要, 并且可以设计合适的滤波策略。而裂缝通常具有特殊的空间模式, 其频谱通常包含一些特定的方向分量。基于 FFT 的注意力机制可以突出这些方向性的特征, 从而使得模型对裂缝结构的学习更加有效。

受上述研究的启发, 本文旨在提升裂缝分割模型在复杂背景下的分割能力, 同时兼顾模型的参数规模与计算效率。为此, 本研究基于 U 型结构构建了轻量化分割网络, 并引入频域感知机制与局部增强模块, 以提升网络对全局结构与局部细节的感知能力。频域处理具有天然的全局表达优势, 本文拟通过 FFT 获取频域信息, 并引入卷积模块以增强裂缝相关频率分量的响应。在此基础上, 为增强模型的细节建模能力, 设计局部增强前馈网络 (locally enhanced feedforward network, LE-FFN), 通过可分离的大核卷积整合多尺度语义信息, 提升特征交互效率。

本文主要工作如下:

1) 设计了一个频率增强注意力机制 (frequency-enhanced attention, FEA)。本研究通过将空间域特征映射至频域, 引入频率建模, 以弥补卷积在全局建模方面的局限。同时, 结合空间卷积强化高低频信息融合, 提升了模型在复杂背景下的全局感知与结构敏感性。

2) 针对传统前馈网络在细粒度特征建模上的不足, 设计一个局部增强前馈网络。融合多尺度语义的可分离卷积模块, 可在增强局部纹理与边缘响应能力的同时, 有效控制模型参数量。

2 方法原理

2.1 网络整体架构

如图 1a 所示, FE-UNet 是一种创新的图像分割网络, 旨在有效处理复杂背景下的裂缝分割任务。

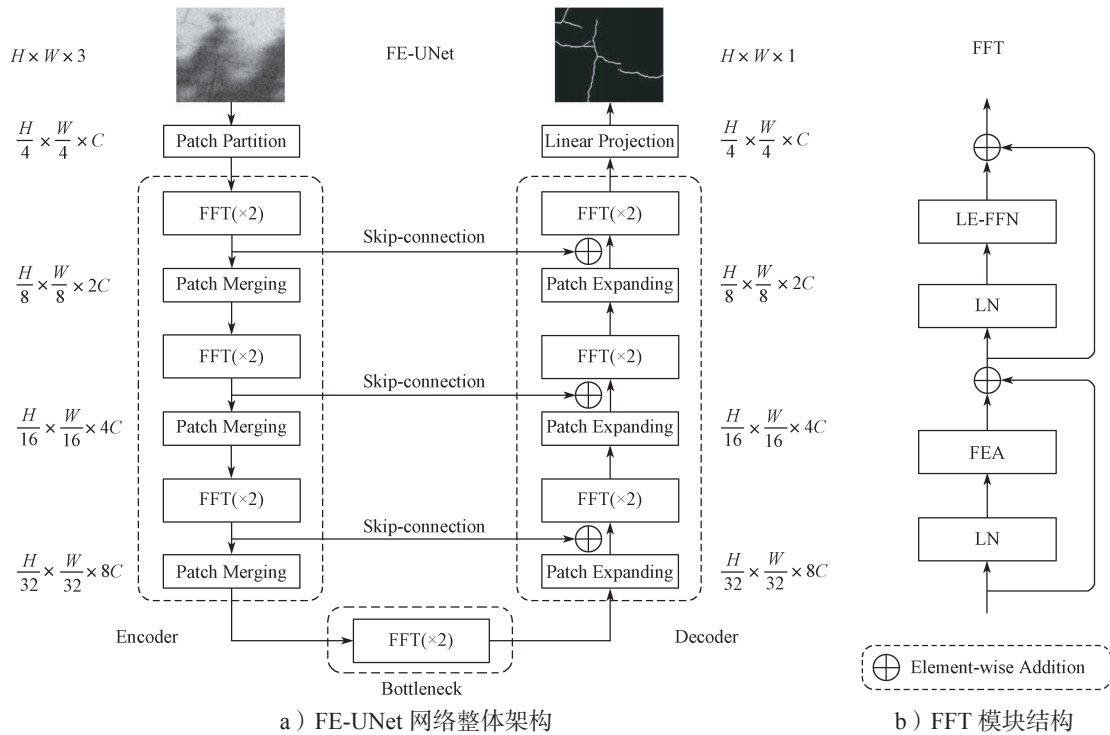


图 1 网络整体架构以及核心模块 FFT 的结构

Fig. 1 Overall network architecture and core module FFT structure

图 1 所示网络融合了频域信息处理和多层特征集成,通过层次化的特征提取和整合^[7],有效提高了在复杂场景下裂缝分割的准确性。首先,网络接收一个尺寸为 $H \times W \times 3$ 的彩色图像作为输入,然后经过补丁嵌入层 (Patch Partition), 被分为 16 个大小为 $\frac{H}{4} \times \frac{W}{4}$ 的补丁,同时将特征通道数由 3 维映射到 C 维。在编码层的每个阶段中,信息首先经过 FFT 变换进行频域特征提取,然后进行特征合并 (Patch Merging), 逐步降低空间分辨率,增加特征通道数。每个阶段重复执行两次傅里叶变换,以增强对频域信息的捕捉,从而有效提取图像的全局特征。在瓶颈层 (Bottleneck), 网络聚合编码器的特征,进一步应用傅里叶变换以捕捉深层特征信息。解码器部分同样由多个阶段组成,通过补丁扩展 (Patch Expanding) 逐步恢复图像的空间分辨率。每个解码阶段同样应用 FFT 进行频域特征处理,强化对细节的重建,确保在恢复空间分辨率时兼顾图像的边缘和结构特征。此外,通过结合来自编码器的跳跃连接 (Skip-connection), 有效地整合细节与上下文信息,从而减少解码层与编码层之间的语义差距。最终的输出经过线性投影层转换为 $H \times W \times 1$ 的单通道二值图像,完成裂缝分割任务。

2.2 频域增强注意力机制

仅依靠傅里叶变换的全局处理难以获得最优性能,

而通过在傅里叶变换后引入卷积操作,有利于强化特征表达,从而增强特征提取效果。为此,设计了一种融合卷积模块的频率增强注意力机制,如图 2 所示。

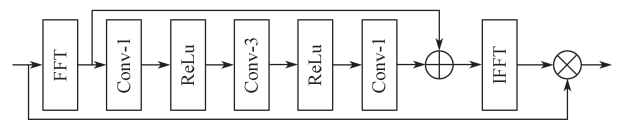


图 2 频率增强注意力机制结构图

Fig. 2 Frequency enhanced attention mechanism structural diagram

在频率增强注意力机制中,输入图像首先经过 FFT, 将其从空间域转换到频域。这样有利于提取图像中的频率特征,从而揭示图像中潜在的结构和模式。 1×1 和 3×3 卷积层对频域特征进行深度卷积处理。通过应用 ReLu 激活函数,模型能够引入非线性特性,从而增强网络对复杂背景下的识别能力。这些卷积操作使得模型能够提取不同层次的特征,同时增强了特征的表达能力。经过卷积处理后,再次利用 1×1 卷积将提取的特征做进一步处理,确保最终输出时保留关键的信息。该过程中所得到的特征与经过 FFT 的特征进行逐元素相加,残差连接以实现信息的有效整合和增强。最后,经过频域增强后的特征通过快速傅里叶逆变换 (inverse fast Fourier transform, IFFT) 转换回空间域,再与输入进行逐元素相乘,对逆变换后的特征进行加权操作,最终形成可以应用于后续分割任务或图像重建的输出。设输入特征

为 $X \in \mathbf{R}^{H \times W \times C}$, 其中 H 和 W 分别为空间维度的高度和宽度, C 为通道数, 最终输出为 X_{output} , 上述过程可描述如下:

$$X_{\text{FFT}} = \text{FFT}(X), \quad (1)$$

$$X' = \text{Conv}_{\text{block}}(X_{\text{FFT}}), \quad (2)$$

$$X_{\text{IFFT}} = \text{IFFT}(X' + X_{\text{FFT}}), \quad (3)$$

$$X_{\text{output}} = X_{\text{IFFT}} \otimes X. \quad (4)$$

式中: $\text{FFT}(\cdot)$ 为快速傅里叶变换; $\text{Conv}_{\text{block}}(\cdot)$ 为中间的卷积模块; $\text{IFFT}(\cdot)$ 为快速傅里叶逆变换; $\otimes$ 为逐元素相乘操作。

2.3 局部增强前馈网络

在传统的 Transformer 中, 前馈网络通常由几个线性变换和几个非线性激活函数组成, 网络只学习到简单的线性变换, 但是在捕获本地信息方面的能力有限^[8]。而图像细节的精确恢复是至关重要的。由于卷积核的归纳偏置特性, 卷积对局部信息表现出更高的敏感性。因此, 为了学习裂缝局部细节, 本文在前馈网络中引入可分离的多尺度卷积核^[9]。相较于普通卷积核, 可分离卷积核将二维卷积核分成水平和垂直的一维核, 并产生更低的计算复杂度和内存消耗。

所构建的局部增强前馈网络如图 3 所示, 在该模块中, 同时应用了多个深度可分离卷积 (depth-wise convolution, DWConv) 操作, 这是一种高效的卷积操作。首先在空间维度上进行深度卷积, 然后通过 1×1 卷积进行特征增强, 随后分成 4 个分支进行局部细节提取, 每个分支分别对应不同的卷积核尺寸。每个分支的卷积核又被分解为级联的水平和垂直方向的一维卷积。这种结构不仅在参数上具有高效性, 还能够从不同的方向提取特征。使得网络能够在不同尺度和方向上理解特征, 从而更充分地捕获局部信息的多样性。在深度可分离卷积输出后, 网络将多个尺度特征进行融合, 通过逐元素相加的方式, 将不同卷积层提取的信息整合起来。这种加法融合的方式有助于保留重要的特征并减少冗余, 使得最终的输出更加丰富和准确。最后通过 1×1 卷积进行特征融合, 并整合输出。对于输入特征 $X \in \mathbf{R}^{H \times W \times C}$, 模块最终的输出 X_{output} 公式化描述如下:

$$X' = \text{Conv-1}(X), \quad (5)$$

$$X_h = \sum_i \text{DWConv-}i\text{-}h(X'), \quad (6)$$

$$X_v = \sum_i \text{DWConv-}i\text{-}v(X_h), \quad (7)$$

$$X_{\text{output}} = \text{Conv-1}(X_v + X'). \quad (8)$$

式 (5) ~ (8) 中: $\text{Conv-1}(\cdot)$ 为 1×1 卷积; i 为卷积核大小, 其值分别为 1, 3, 5, 7; $\text{DWConv-}i\text{-}h(\cdot)$ 和 $\text{DWConv-}i\text{-}v(\cdot)$ 分别为水平和垂直的一维卷积核。

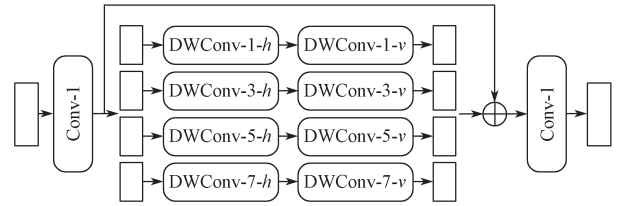


图 3 局部增强的前馈网络结构图

Fig. 3 Locally enhanced feedforward network structural diagram

3 模型训练

3.1 实验细节和数据集

本文所有实验均在 Ubuntu 20.04 操作系统上进行, 并使用 Pytorch 2.0.0 框架和 Python 3.8、Cuda 版本为 11.8 来实现所提出的网络。训练采用 Adam 优化器对损失函数进行优化, 所有网络训练 40 轮。

为了评估所提出网络模型的性能, 本研究在 CrackTree260 和 CrackForest-Dataset (CFD) 两个公开数据集上进行了实验验证。

1) CrackTree260。该数据集包含 260 张分辨率为 $800 \text{ dpi} \times 600 \text{ dpi}$ 的已标注裂缝图片。这些图片来自阴影遮挡和复杂路况下的裂缝图片。随机选择 210 张图片用于训练, 剩余 50 张图片用于测试。

2) CFD。该数据集包含 118 张分辨率为 $480 \text{ dpi} \times 320 \text{ dpi}$ 的已标注裂缝图片, 该数据集中裂缝受光照明暗变化的影响。随机选择 100 张图片用于模型训练, 剩余 18 张图片用于模型测试。

为防止模型过拟合, 提高网络的泛化能力。在训练之前, 对数据集采取随机旋转、随机翻转等数据增强手段, 扩充数据集。

3.2 损失函数

损失函数用于描述模型预测值与真实值之间的差距。损失函数随模型的训练不断下降, 最终达到收敛。二元交叉熵损失函数通常用于二分类任务, 可以定义如下:

$$BCE_{\text{Loss}} = -[y \log(\hat{y}) + (1 - y) \log(1 - \hat{y})], \quad (9)$$

式中: $\hat{y}$ 为预测值; y 为真实值。

不同于其它分割任务, 裂缝图像中裂缝像素通常占整个背景像素的比例较低, 因此, 裂缝分割任务属于严重的类不平衡分割任务。这会导致传统的交叉熵损失在训练过程中对小面积裂缝像素的分类缺乏敏感性。而 Dice 损失通过考虑真正例的比例, 能更好地优化小区域的分割效果, 在改善小目标检测性能方面发挥重要作用, Dice 损失函数定义如下:

$$Dice_{\text{Loss}} = 1 - 2 \sum_{i=1}^N y_i \hat{y}_i / \left(\sum_{i=1}^N y_i + \sum_{i=1}^N \hat{y}_i \right), \quad (10)$$

式中: N 为图像像素总数; y_i 为第 i 个图像像素的真实值; $\hat{y}_i$ 为第 i 个图像像素的预测值。

在模型训练过程中, 通过交叉熵损失优化每个像素的分类概率, 同时利用 Dice 损失优化最终预测的形状和结构特征。最终的损失函数可定义如下:

$$Loss(\hat{y}, y) = BCE_{Loss}(\hat{y}, y) + Dice_{Loss}(\hat{y}, y) \quad (11)$$

3.3 评价指标

为能更全面地评价模型的表现, 本研究采用准确率 (precision) P_r , 召回率 (recall) R_e , $F_{1-score}$ 和平均交并比 (mean intersection over union) $mIoU$ 作为模型的评价指标。各指标定义如下:

$$P_r = \frac{T_p}{T_p + F_p} \quad (12)$$

$$R_e = \frac{T_p}{T_p + F_N} \quad (13)$$

$$F_{1-score} = \frac{2 \cdot P_r \cdot R_e}{P_r + R_e} \quad (14)$$

$$mIoU = \text{mean} \left(\frac{T_p}{T_p + F_p + F_N} \right) \quad (15)$$

式 (12) ~ (15) 中: T_p 为真正, 表示裂缝像素被正确地预测为裂缝像素; F_p 为假阳性, 表示模型错误地将非裂纹像素预测为裂纹像素; F_N 为假阴性, 表示裂纹像素被模型错误地预测为非裂纹像素。

此外, 为量化模型的规模与计算效率, 本文采用参数量 (params) 和十亿次浮点运算量 (giga floating-point operations per second, GFLOPs) 分别表

征模型的存储开销与计算复杂度。

4 实验结果与分析

4.1 对比分析

将本研究中所提出的网络模型与 SegNet^[10], Deeplabv3+^[11], DeepCrack^[12], Swin-UNet^[13] 4 款经典分割网络模型在 CrackTree260 和 CFD 两个数据集集中进行定量和定性对比以及效率分析, 所得实验结果如表 1、表 2 和图 4 所示。

表 1 不同模型在 CrackTree260 中的性能比较

Table 1 Performance comparison of different models in CrackTree260

models	P_r /%	R_e /%	$F_{1-score}$ /%	$mIoU$ /%	GFLOPs/ $10^9 s^{-1}$	Params/ 10^6
SegNet	76.19	81.79	78.89	82.30	160.52	29.44
Deeplabv3+	65.79	<u>83.33</u>	73.53	78.35	69.14	40.29
DeepCrack	75.17	81.54	78.22	81.45	547.22	30.91
Swin-UNet	43.34	56.09	48.80	66.48	56.04	27.24
Ours	<u>76.42</u>	82.79	<u>79.48</u>	<u>82.38</u>	<u>20.91</u>	<u>17.47</u>

注: 下划线数据为最佳结果, 下同。

表 2 不同模型在 CFD 的性能比较

Table 2 Performance comparison of different models in CFD

models	P_r /%	R_e /%	$F_{1-score}$ /%	$mIoU$ /%	GFLOPs/ $10^9 s^{-1}$	Params/ 10^6
SegNet	73.95	75.35	74.64	79.21	160.52	29.44
Deeplabv3+	74.56	76.70	75.61	79.95	69.14	40.29
DeepCrack	72.00	71.55	71.77	76.22	547.22	30.91
Swin-UNet	66.04	64.29	65.15	71.34	56.04	27.24
Ours	<u>75.95</u>	<u>79.81</u>	<u>77.83</u>	<u>81.27</u>	<u>20.91</u>	<u>17.47</u>

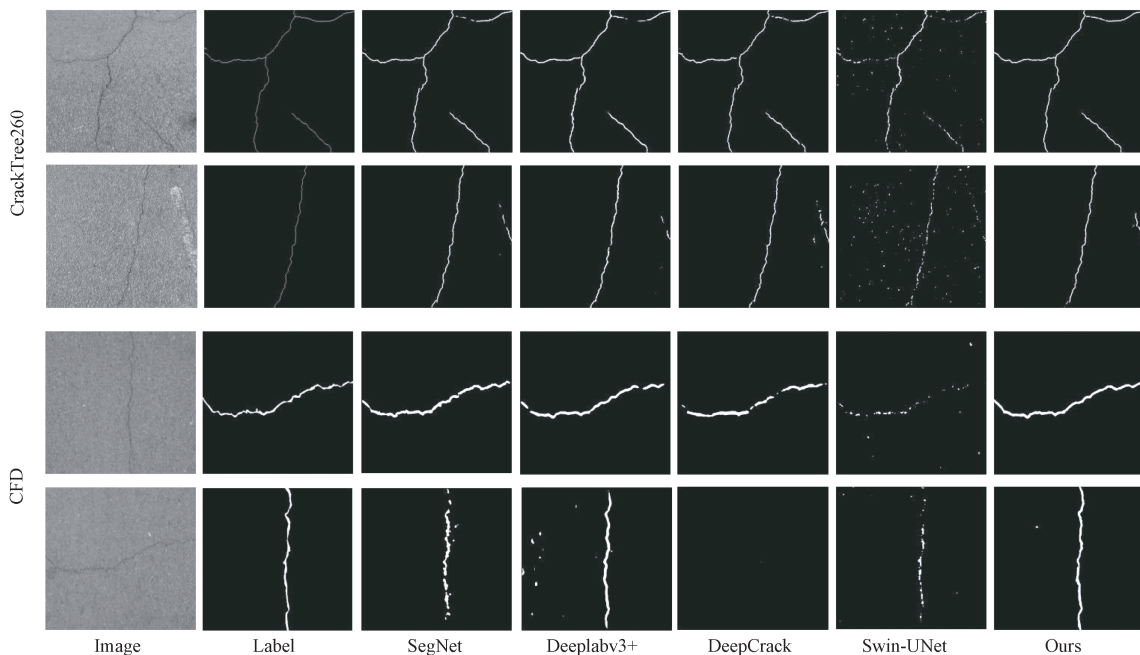


图 4 各模型在 CrackTree260 和 CFD 上的可视化分割结果

Fig. 4 Visualization of segmentation results of various models on CrackTree260 and CFD

1) 定量分析。在两个数据集中, 各模型的性能指标均表现出显著不同。如表1所示, 对于 P_r 、 R_c 、 $F_{1-score}$ 和 $mIoU$ 等重要指标, 本文所提模型 (Ours) 在 CrackTree260 数据集中分别达 76.42%, 82.79%, 79.48%, 82.38%。相比其他模型, 如 SegNet 以及 DeepCrack, 所提模型在多个指标上均有所提高。Deeplabv3+ 在召回率上表现出最好的性能, 相比所提模型高 0.54%。然而其整体性能有所欠缺, 且有较高的模型参数和计算复杂度。如表2所示, 在 CFD 数据集上, 本文所提模型全面领先。在4个指标上分别达 75.95%, 79.81%, 77.83%, 81.27%。尤其在 $mIoU$ 指标上较其他对比模型有明显提升, 表明该模型在更复杂的背景环境中, 对裂缝特征的捕捉能力更强。从定量分析结果看, 无论是在 CrackTree260 还是 CFD 数据集上, 本文的网络整体在评分指标中始终处于领先地位, 验证了其在裂缝检测任务中的有效性。

2) 定性分析。如图4所示, 在 CrackTree260 数据集中, 所提模型成功捕捉到了细微裂缝的结构, 且分割边缘更平滑, 明显优于其他模型。且在一些图像中, 其他模型未能完全识别裂缝, 导致漏检或分割不完整, 而本文模型则展现出较高的精度和完整性。在 CFD 数据集中, 尽管一些模型能在一定程度上识别出裂缝, 但本文模型再次显示出更加精确的轮廓重建, 减少了误检区域, 并在复杂背景下保持了高质量的分割效果。

3) 效率分析。参数数量和计算复杂度是衡量模型复杂度的重要指标, 会影响网络训练时间和推理速度。如表1所示, 所提网络输入尺寸为 512×512 时, 相比其他模型, 表现出最低的模型参数数量和计算复杂度, 分别为 17.47×10^6 和 $20.91 \times 10^9 s^{-1}$ 。图5为不同模型的计算复杂度随输入图像尺寸的变化趋势。

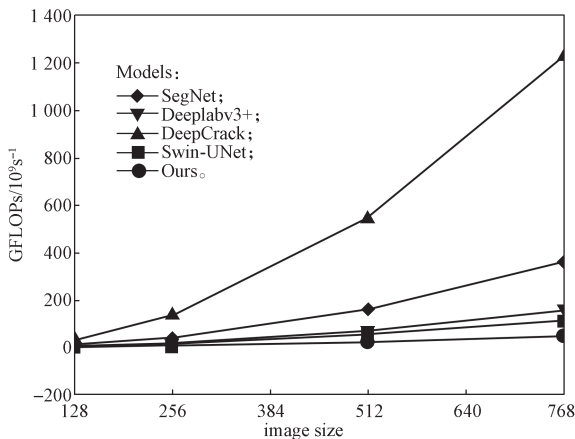


图5 各模型计算复杂度随输入尺寸大小的变化

Fig. 5 Model computational complexity variation with input sizes

如图5所示, 随着输入图像尺寸增大, 所有模型的计算复杂度均呈现出上升趋势, 但整体上所提出模型的增长速度相对较慢。在相同输入尺寸下, 所提出模型的计算复杂度始终低于其他主流模型。值得注意的是, 在输入尺寸为 768×768 时, 所提出网络的计算复杂度约为 $46 \times 10^9 s^{-1}$, 远远低于 DeepCrack 模型的对值, 在较低计算成本下仍能实现高分辨率的裂缝分割。这对于模型在实时处理和资源受限场景中的表现至关重要。

整体而言, 通过对 CrackTree260 和 CFD 两个数据集的定量和定性分析, 可知所提模型在裂缝检测任务中均展现出显著的优势。定量结果显示本文方法在 $F_{1-score}$ 和 $mIoU$ 等关键性能指标上超过了现有的模型, 而定性结果则进一步佐证了这些指标的有效性, 强调了模型在处理复杂的裂缝图像时所具备的优势。另外, 结合模型的参数量与计算复杂度, 本文所提模型在分割性能和效率上展现出了良好的平衡。

4.2 消融研究

通过分别将所提出的 FEA 和 LE-FFN 两个模块, 替换为多头自注意力 (MSA) 和多层感知机 (MLP), 在 CrackTree260 数据集上进行消融分析, 所得实验结果如表3所示。

表3 FE-UNet 在 CrackTree260 数据集上的消融实验结果
Table 3 Experimental results of FE-UNet ablation on CrackTree260 dataset

FEA	LE-FFN	P_r / %	R_c / %	$F_{1-score}$ / %	$mIoU$ / %	GFLOPs/ $10^9 s^{-1}$	Params/ 10^6
✓		72.29	77.79	74.94	79.36	28.95	25.18
	✓	75.00	81.42	78.08	81.42	22.06	18.90
✓	✓	76.42	82.79	79.48	82.38	20.91	17.47

具体来说, 替换掉 FEA 后, 模型整体性能下降, 综合指标 $F_{1-score}$ 和 $mIoU$ 分别降低了 1.4% 和 0.96%, 而参数量和模型计算复杂度分别增加了 1.43×10^6 和 $1.15 \times 10^9 s^{-1}$ 。这表明多头自注意力机制虽然在建模远距离依赖方面具有优势, 但基于快速傅里叶变换的注意力机制, 通过将空域信息转换到频域, 从而具有处理全局特征的优势, 并且具有更低的参数量和计算复杂度。融合的卷积模块进一步增强了特征提取能力。局部增强前馈网络替换为 MLP 后, 网络总体性能显著下降。综合指标 $F_{1-score}$ 和 $mIoU$ 分别减少了 4.54% 和 3.02%。参数量和计算复杂度分别增加了 8.04×10^6 和 $7.71 \times 10^9 s^{-1}$ 。这验证了传统多层感知机无法学习到局部信息特征的弊端, 同时隐藏层的数量较大, 造成了高昂的计算开销。相反, 本研究采用深度可分离卷积的 LE-FFN 在保持计算效率的同时, 极

大地增强了网络对局部特征提取的能力。通过多尺度和多方向的特征提取,提高了网络在复杂背景下对局部特征的敏感性,使得在图像分割任务中能够更好地捕获局部细节信息。

综上所述,消融实验清晰地表明了 FEA 和 LE-FFN 模块在模型性能提升中的重要性。通过对比不同模块的替代方案,验证了采用频率特征增强和局部增强策略所取得的显著性能优势,同时增加两个模块,模型达到最优性能。

5 结语

本文针对先前的裂缝分割网络在远距离依赖建模和局部特征提取上的不足,提出了一种面向裂缝图像精细分割的轻量化网络 FE-UNet。通过引入频域增强机制与局部特征建模策略,在保证模型效率的前提下实现了高精度分割性能。实验结果验证了所提方法在多个场景下的鲁棒性,尤其在复杂背景与微小裂缝处理方面具有明显优势。本方法具有良好的泛化能力与计算效率,有望被应用于道路健康监测、桥梁结构检测及地下管网维护等自动巡检场景中。同时,未来将进一步聚焦模型在实时部署中的轻量化设计,以及提升多源数据融合背景下的鲁棒裂缝识别能力。

参考文献:

- [1] CHENG J R, XIONG W, CHEN W Y, et al. Pixel-Level Crack Detection Using U-Net[C]//TENCON 2018-2018 IEEE Region 10 Conference. Jeju: IEEE, 2018: 462-466.
- [2] WANG Y Y, SONG K C, LIU J, et al. RENet: Rectangular Convolution Pyramid and Edge Enhancement Network for Salient Object Detection of Pavement Cracks[J]. Measurement, 2021, 170: 108698.
- [3] LIU H J, YANG J, MIAO X Y, et al. CrackFormer Network for Pavement Crack Segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2023, 24(9): 9240-9252.
- [4] XIANG C, GUO J J, CAO R, et al. A Crack-Segmentation Algorithm Fusing Transformers and Convolutional Neural Networks for Complex Detection Scenarios[J]. Automation in Construction, 2023, 52: 104894.
- [5] WEN X, LI S, YU H, et al. Multi-Scale Context Feature and Cross-Attention Network-Enabled System and Software-Based for Pavement Crack Detection[J]. Engineering Applications of Artificial Intelligence, 2024, 127: 107328.
- [6] CHI L, JIANG B, MU Y. Fast Fourier Convolution[J]. Advances in Neural Information Processing Systems, 2020, 33: 4479-4488.
- [7] 周秋艳, 肖满生, 范双南. 基于注意力机制和多层次特征融合的目标检测算法 [J]. 湖南工业大学学报, 2023, 37(1): 61-68.
ZHOU Qiuyan, XIAO Mansheng, FAN Shuangnan. Target Detection Algorithm Based on Attention Mechanism and Multi-Level Feature Fusion[J]. Journal of Hunan University of Technology, 2023, 37(1): 61-68.
- [8] LI Y W, ZHANG K, CAO J Z, et al. LocalViT: Analyzing Locality in Vision Transformers[EB/OL]. [2025-03-12]. <https://arxiv.org/abs/2104.05707v2>.
- [9] LAU K W, PO L M, REHMAN Y A U. Large Separable Kernel Attention: Rethinking the Large Kernel Attention Design in CNN[J]. Expert Systems with Applications, 2024, 236: 121352.
- [10] BADRINARAYANAN V, KENDALL A, CIPOLLA R. SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(12): 2481-2495.
- [11] CHEN L C, ZHU Y K, PAPANDREOU G, et al. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation[C]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 833-851.
- [12] LIU Y H, YAO J, LU X H, et al. DeepCrack: A Deep Hierarchical Feature Learning Architecture for Crack Segmentation[J]. Neurocomputing, 2019, 338: 139-153.
- [13] CAO H, WANG Y Y, CHEN J, et al. Swin-Unet: Unet-Like Pure Transformer for Medical Image Segmentation[M]//Computer Vision - ECCV 2022 Workshops. Cham: Springer Nature Switzerland, 2023: 205-218.

(责任编辑: 廖友媛)