中北大学学报(自然科学版)
JOURNAL OF NORTH UNIVERSITY OF CHINA(NATURAL SCIENCE EDITION)
2026年 第47卷 第5期(总第229期)
Vol. 47 No. 5 2026(Sum No. 229)

文章编号：1673-3193(2026)05-0749-13 J. North Univ. China, Nat. Sci. Ed., 2026, 47(5): 749-761.

面向稀疏前景与复杂背景的隧道裂缝图像分割方法

赵凯艺[1]，赵银燕[1]，秦品乐[1]，李 磊[2]

（1. 中北大学 计算机科学与技术学院，山西 太原 030051；2. 深眸远见（山西）技术有限公司，山西 太原 030051）

摘 要： 作为交通基础设施的核心枢纽，隧道的安全运营直接关系到公共安全的有效保障。隧道衬砌早期病害的快速识别，对保障隧道安全运营具有重要意义，但隧道裂缝自动化分割任务仍面临两大核心难题：一是裂缝形态呈现狭长不规则、随机断续的前景稀疏特性；二是易受电网线、维修痕迹等复杂背景噪声的干扰。对此，本文面向稀疏前景与复杂背景的特性提出了一种双路径引导的隧道裂缝图像分割网络模型。该模型协同主辅路径融合实现细微裂缝与低对比度区域特征的稳定表示，其中主路径结合空间、频域和通道注意力机制，构建具备多维感知能力的编码网络，精准捕捉裂缝结构的全局特征，实现高质量特征表示。辅路径基于前景引导式特征传播机制，持续强化裂缝前景特征并保持跨尺度一致性，从而有效抑制背景干扰噪声。在公开的DeepCrack数据集与高速公路隧道衬砌病害自建数据集上的对比实验结果显示，本文所提方法的IoU值最高达0.688 5，*F*1值最高达0.815 5，精确率最高达0.693 3，召回率最高达0.773 2，综合性能优于现有对比方法。

关键词： 隧道裂缝分割；稀疏前景；复杂背景；深度学习

中图分类号： TP391.41；U456.3 **文献标识码：** A **doi:** 10.62756/jnuc.issn.1673-3193.2026.04.0002

引用格式： 赵凯艺，赵银燕，秦品乐，等. 面向稀疏前景与复杂背景的隧道裂缝图像分割方法[J]. 中北大学学报(自然科学版)，2026，47(5)：749-761.

Zhao Kaiyi, Zhao Yinyan, Qin Pinle, et al. A precise tunnel crack image segmentation method for sparse foreground and complex background[J]. Journal of North University of China(Natural Science Edition), 2026, 47(5): 749-761.

A Precise Tunnel Crack Image Segmentation Method for Sparse Foreground and Complex Background

Zhao Kaiyi[1], Zhao Yinyan[1], Qin Pinle[1], Li Lei[2]

(1. School of Computer Science and Technology, North University of China, Taiyuan 030051, China;
2. DeepSight Foresight (Shanxi) Technology Co., Ltd., Taiyuan 030051, China)

Abstract: As the core hub of transportation infrastructure, the safe operation of tunnels is directly related to the effective guarantee of public safety. The rapid identification of early defects in tunnel lining is of great significance for ensuring the safe operation of tunnels. However, the automated segmentation task of tunnel cracks still faces two core challenges: firstly, the crack morphology exhibits sparse foreground characteristics of narrow, irregular, and randomly intermittent shapes. Secondly, it is easily interfered by complex background noises such as power grid lines and maintenance marks. In response, we proposed a dual-path guided tunnel crack image segmentation network model tailored to the characteristics of sparse foreground and complex background. The model achieves stable representation of subtle cracks and low-contrast regions through collaborative fusion of primary and auxiliary pathways. The primary pathway integrates spatial, frequency,

收稿日期： 2026-04-08

作者简介： 赵凯艺(1993—)，男，讲师，博士，主要从事不确定信息处理、多模态融合的研究。E-mail: zhaokaiyi@nuc.edu.cn。

and channel attention mechanisms to construct an encoding network with multidimensional perception capability, accurately capturing global features of crack structures and enabling high-quality feature representation. The auxiliary pathway leverages a foreground-guided feature propagation mechanism to continuously enhance foreground crack features and maintain cross-scale consistency, effectively suppressing background interference and noise. Comparative experimental results on the public DeepCrack dataset and a self-built dataset of highway tunnel lining defects show that the IoU value of the proposed method reaches up to 0.688 5, the $F1$ score reaches up to 0.815 5, the precision reaches up to 0.693 3, and the recall reaches up to 0.773 2, outperforming existing comparative methods in comprehensive performance.

Key words: tunnel crack segmentation; sparse foreground; complex background; deep learning

0 引 言

隧道作为复杂地形条件下交通组织的关键支撑，是交通基础设施体系的重要组成部分，其结构安全直接关系到交通运输的安全与畅通。在长期外力作用及湿热、雨雪、雾霾等复杂环境因素耦合影响下，隧道衬砌易产生裂缝病害，这不仅会改变结构受力状态，降低衬砌的刚度与整体承载能力，还会进一步诱发混凝土碳化、钢筋锈蚀等耐久性问题，对隧道长期安全运营构成了严重威胁[1-2]。因此，为确保隧道结构持续处于安全健康状态，开展常态化、规范化的安全巡检工作，是保障隧道安全运营的必要技术手段与管理措施。

传统隧道衬砌裂缝检测主要依赖现场目视巡查、无损检测等人工方式，这类方法在人力投入、检测周期与经济成本上均存在显著局限，且易受设备布设、作业空间条件及交通临时管控等现场因素制约，不仅整体检测效率偏低，检测结果也易受人员主观判断影响而产生偏差[3-4]。为此，早期学者基于图像处理与机器学习技术开展了大量研究探索，但受限于传统方法特征提取角度单一、表征能力有限，且分类器参数依赖经验调优等不足，相关模型在高曲率、不连续等复杂背景下的裂缝图像识别中，往往难以取得理想效果[5]。得益于深度学习在图像处理中能够自主学习目标多尺度深层语义特征，且具备较强的泛化性能等优势，基于深度学习开展的衬砌裂缝识别研究已取得了一定进展[6]。

随着算力水平的大幅提升，深度学习在复杂场景下实现特征表示的自动化建模方面展现出显著优势。典型方法方面，Liu等[7]通过多尺度特征提取与深度监督机制改进传统全卷积网络，同时捕捉裂缝的全局结构信息与细粒度纹理特征，实现了更高精度的裂缝分割。Lee等[8]采用基于深度卷积编解码结构的SegNet模型，实现了高速铁路隧道衬砌裂纹的智能识别，并结合图像处理技术完成了裂缝最大宽度测量。邝先验等[9]在MobileViT基准模型基础上引入改进的注意力机制，在保证分类精度的基础上提高了裂缝分割的推理效率。综上，现有模型虽在隧道裂缝分割、定量评估及推理效率等方面已取得良好性能，但在复杂背景与低占比裂缝前景并存的复杂场景下，隧道裂缝的多层级连续表征与融合、像素级精细化分割等方面仍存在显著提升空间。

针对隧道裂缝前景分布稀疏、像素占比极低，以及背景噪声空间随机多样、占比较大的特点，本文提出一种主-辅双路径引导的隧道裂缝图像分割网络。其中，主路径通过构建面向裂缝区域的语义与结构多维特征编码机制，提升模型在稀疏断续分布场景下的前景感知能力，从而强化稀疏裂缝的特征表达。辅路径则通过建立跨层级特征交互的前景引导传播机制，持续增强前景信息传递与表达，有效抑制复杂背景干扰，进一步保障稀疏裂缝分割的可分辨性与结构完整性。在DeepCrack公开数据集和山西省多座高速隧道的自建数据集上开展验证，实验结果表明，所提方法相比对比算法取得了更优的分割性能，验证了双路径协同机制可有效提高裂缝图像的分割精度。

1 相关工作

目前，裂缝分割相关研究主要采用各类深度学习方法实现，根据网络架构可分为三类技术路线：基于卷积神经网络的方法、基于Transformer的方法，以及二者融合的混合模型。

1.1 基于卷积神经网络的方法

卷积神经网络具有局部感受野、权重共享与分层特征提取等优势，因此早期涌现出大量基于卷积

结构的语义分割网络并被应用于裂缝检测任务，其中全卷积网络、U-Net 和 DeepLab 是最具代表性的经典架构。王森等[10]将全卷积网络应用于图像裂缝检测任务，提出 Crack-FCN 模型，有效实现了图像裂缝的精准识别。Liu 等[7]通过引入多尺度特征提取机制与深度监督机制，针对性改进了传统全卷积网络，使模型能够同步捕获裂缝的全局结构信息与细粒度纹理特征，有效提升了裂缝分割的精度与可靠性。常惠等[11]采用嵌入挤压激励模块的改进 U-Net 网络来提升隧道裂缝图像的分割精度，实现了针对开裂严重且光照条件良好的高质量裂缝图像的精准分割。Liu 等[12]采用 U-Net 模型开展了混凝土裂纹检测研究，所用方法相较于传统基于卷积神经网络的方法显示了一定优势。Sun 等[13]在 DeepLabv3 模型的基础上，通过在低层与高层特征图之间嵌入注意力模块，对不同层级的特征图赋予差异化权重，提升了裂缝图像的分割精度。然而，卷积神经网络对细长弱边缘目标的形态特征建模能力有限，在复杂背景或数据分布变化时易出现裂缝识别不稳定、漏检等问题[14]，制约了其在裂缝分割任务中的性能上限。

1.2　基于 Transformer 的方法

凭借强大的长程依赖建模与全局上下文捕获能力，Transformer 已被成功应用于裂缝图像分割任务中。Xu 等[15]将 Transformer 与跳跃连接相结合，对高分辨率道路裂纹图像进行检测，实现了高精度的道路裂缝分割目标。Lin 等[16]为同时解决裂缝检测中全局上下文建模与局部细节恢复的难题，利用 Transformer 编码器捕获裂缝像素间的全局关联，并设计了轻量化解码器，通过跨层特征聚合实现高分辨率裂缝细节的精准恢复。邝先验等[9]提出了 CrackViT 系列轻量化网络，在 MobileViT 基线模型中引入改进的注意力机制，在提升分类性能的同时保持了较快的推理速度。Wang 等[17]基于 Transformer 架构，采用分层编码器提取多尺度特征，借助带横向连接的路径解码器完成特征融合，再结合在线难例挖掘策略，在混凝土裂缝像素级分割任务中实现了高精度检测。综上，尽管 Transformer 可借助自注意力机制建模全局依赖关系，有效提升裂缝分割任务中的特征表达能力，但在裂缝前景极度稀疏、像素占比极低的场景下，模型有限的注意力资源易被大量无关背景区域分散，致使其对前景关键特征的学习不充分[18]。

1.3　基于混合模型的方法

为充分结合卷积神经网络（Convolutional Neural Network，CNN）在局部特征提取上的优势与 Transformer 在全局依赖建模上的能力，现有研究多采用 CNN 与 Transformer 相融合的混合架构。Liu 等[19]在 CrackFormer 基础上，引入包含局部自注意力层、局部前馈层与跳跃连接的 Transformer 编码器模块，优化了通道配置并保留缩放融合策略，为路面裂缝分割提供了更为高效的解决方案。Xiang 等[20]构建了以 CNN 和 Transformer 双编码器为核心的 DTrC-Net 模型，通过跨分支特征融合与残差路径强化局部与全局信息交互，使其在复杂场景下的裂缝分割任务中实现了精度与速度的兼顾。Zhou 等[21]在 DeepLabv3＋框架基础上，融合 Swin Transformer 与 CNN 构建了新型混合网络 SC-DeepLab，并进一步引入通道注意力机制与迁移学习策略，实现了对隧道衬砌裂缝的有效分割。Zhang 等[22]采用 ViT 与 ResNet 分别对裂缝缺陷进行特征表征，并在解码器中通过注意力机制实现二者融合，有效兼顾了全局与局部信息，该方法在细长裂缝缺陷检测任务中表现出显著优势。综上，当前语义分割研究多围绕 CNN 与 Transformer 的架构融合展开，相关模型在特征表达能力与复杂场景适应性上持续提升。然而，在复杂背景与稀疏前景共存的场景下，现有方法在小尺度结构的连续刻画、多级特征的有效协同及空间细节的精细恢复等方面仍存在明显不足。

2　方　法

现有文献和实验分析表明，裂缝前景的稀疏性会降低网络对裂缝语义的显式感知能力，而其空间分布的高度随机性则进一步加剧了前景特征在空间建模过程中的不确定性。若模型缺乏对全局上下文关系的有效建模，且未对前景特征构建显式的强化机制，在特征提取与传递过程中，裂缝特征响应极易在深层语义表示阶段被复杂背景噪声所淹没。因此，本文从编码阶段的特征建模视角出发，提出了一种主-辅双路径引导的隧道裂缝图像分割网络。其核心思路是在编码阶段引入双路径协同编码机制，将特征建模过程划分为全局多维特征构建（Dual-domain Wavelet-aware Former，DWFormer）与前景特征引导传播（Guided Foreground Feature Propagation，GFFP）两条主辅

路径，二者相互关联且各有侧重，协同缓解裂缝前景稀疏性与空间分布随机性所导致的裂缝特征弱化问题。本文所提出的网络框架如图1所示。

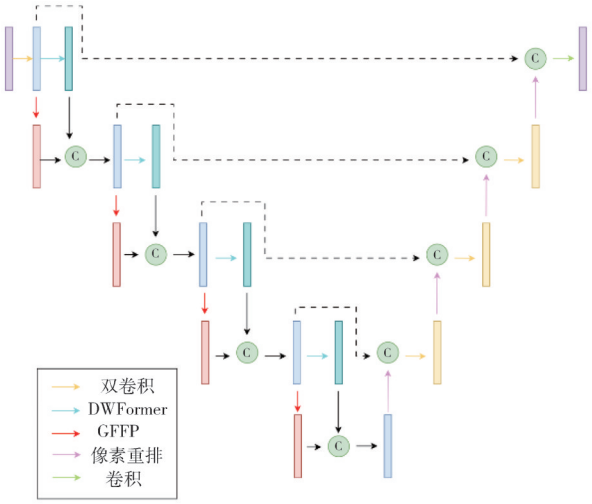


图 1　本文方法的网络架构示意图

Fig. 1　Schematic diagram of network architecture for the proposed method

2.1　全局多维特征构建

裂缝在空间形态上无明显规律，且与复杂背景纹理高度相似，导致编码阶段随机分布的裂缝片段难以在特征空间形成稳定、一致的结构表征，又易受背景纹理干扰产生特征混叠，进一步削弱了裂缝语义在编码阶段的显著性。因此，从建模机制分析，裂缝特征在编码阶段难以稳定表达，主要源于三方面的问题：

1）裂缝多表现为细长边缘或局部突变，其特征在频域中通常更为显著，若编码过程仅在空间域进行特征建模，忽略频域信息，将无法充分刻画裂缝的精细结构信息；

2）对于裂缝这类语义强度较弱的目标，其对应通道特征在整体通道分布中占比较低，若缺乏有效的通道重组与语义增强机制，稀疏裂缝信息极易被冗余背景通道特征所稀释；

3）裂缝空间分布高度随机，其结构信息难以通过固定感受野或局部上下文关系完成稳定建模，即裂缝片段间无法在特征空间建立有效关联，从而导致编码阶段的裂缝特征呈现断裂、不连续的问题。

基于上述分析，为缓解裂缝前景稀疏、空间分布随机所导致的编码建模难题，本文在特征表示阶段设计了全局多维特征构建模块DWFormer。

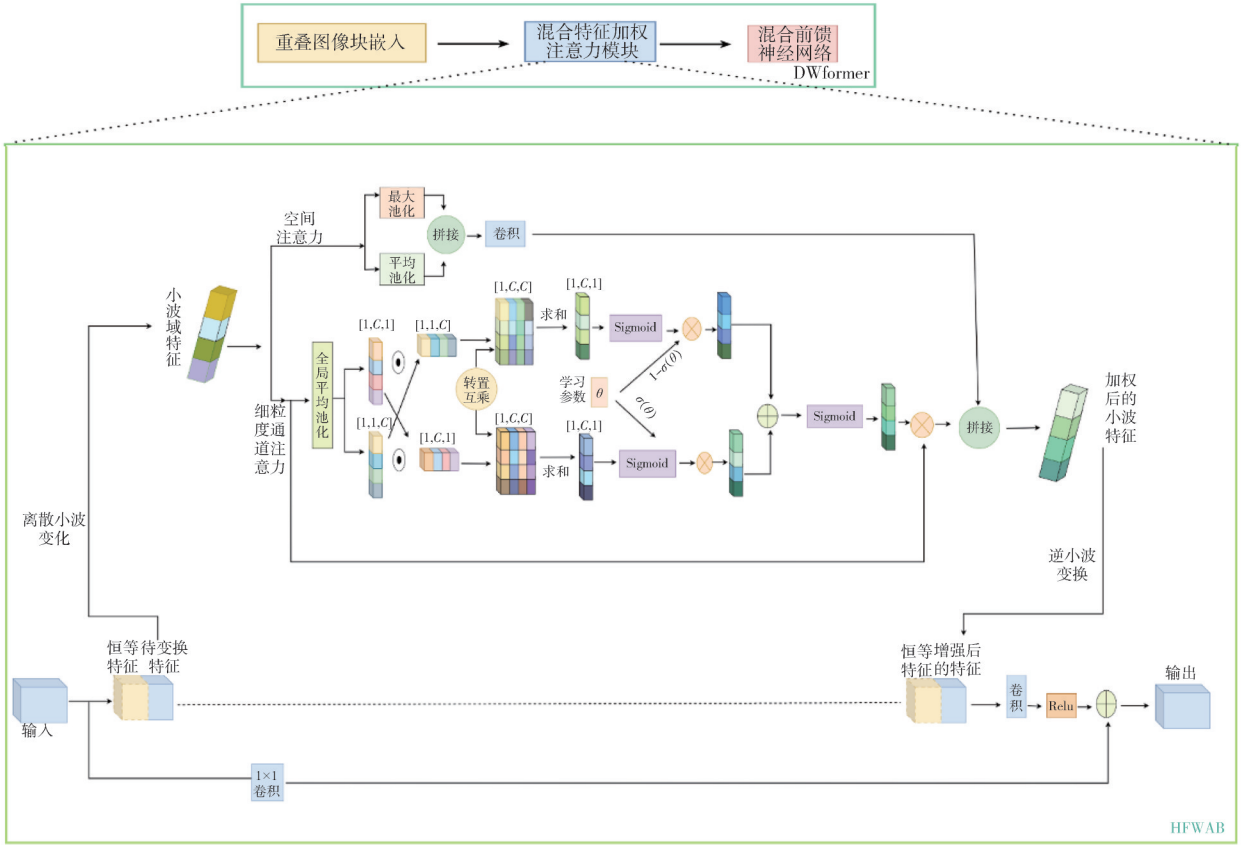


图 2　全局多维特征构建网络架构示意图

Fig. 2　Schematic diagram of DWFormer network architecture

该模块以裂缝特征精准表达为核心，结合裂缝的空间结构、频域特性和通道语义分布规律，通过对编码内部建模机制的重构与协同设计，引导裂缝相关信息在高维特征空间中形成更稳定、连续的特征表示，有效提升裂缝特征的显著性与鲁棒性，缓解随机分布裂缝在建模过程中的不确定性，从而为后续前景特征传导与解码重建提供坚实可靠的高层语义基础。全局多维特征构建网络架构示意如图 2 所示，该模块由重叠图像块嵌入（Overlapping Patch Embedding，OPE）、混合特征加权注意力模块（Hybrid Feature Weighted Attention Block，HFWAB）及混合前馈网络（Mixed Feed-Forward Network，MixFFN）三个结构模块按串联方式组合而成。

2.1.1　重叠图像块嵌入[23]

通过滑窗重叠划分在相邻图像块间建立显式关联，在将特征转换为词元表示的同时增强特征嵌入阶段的空间连续性，为后续长程依赖建模提供更为连贯的裂缝结构表示基础。该重叠图像块嵌入模块（Overlapping Patch Embedding，OPE）的数学表示为

$$x_0 = OPE(x_{\text{in}})。\tag{1}$$

2.1.2　混合特征加权注意力模块

为保留裂缝结构的局部连续性，并强化其长程结构与关键频域特征的建模能力，本文基于半小波注意力块[24]提出了混合特征加权注意力模块（Hybrid Feature Weighted Attention Block，HFWAB）。该模块通过构建空间、通道与频域三重注意力机制，有效提升了裂缝特征表示的判别性与鲁棒性。该模块的数学表示为

$$[X_w, X_i] = Chunk(X_0, 2, dim = c),\tag{2}$$

式中：$Chunk(\cdot)$表示沿指定维度对特征进行均匀分块操作。所得特征X_w被送入小波变换与注意力建模分支，以提升特征表达能力。X_i则作为恒等映射分支保留原始信息，在后续阶段与增强后的特征完成融合。

$$F_{\text{dwt}} = DWT(X_w),\tag{3}$$

式中：DWT 为基于离散小波变换的 Transformer 编码网络模块（Discrete Wavelet Transformer-based Transformer）；$DWT(X_w)$表示对输入特征图X_w进行离散小波变换，将其从空间域转换至频域，转换结果表示为F_{dwt}。该过程用于提取裂缝图像的多尺度频率信息，具体包括水平低频、垂直高频、水平高频和垂直低频分量，可以有效表示输入特征图的频率特性。频域分支对不同尺度裂缝具有不同程度的响应特性。其中，高频分量更加关注裂缝边缘、细小纹理以及局部结构变化，对于细裂缝与弱裂缝区域具有较强的敏感性；低频分量则保留更多整体结构与区域分布信息，有助于维持裂缝整体走向与连续性表达。频域分解能够从不同频率层面对裂缝结构进行多尺度表征，使模型在关注细小裂缝边缘信息的同时兼顾不同尺度下的结构特征，从而提升模型对复杂尺度变化裂缝的适应能力。

$$F_{\text{conv}} = Conv_1(ReLU(Conv_0(F_{\text{dwt}}))),\tag{4}$$

式中：$Conv_0$和$Conv_1$表示普通卷积操作（3×3 卷积）；$ReLU(\cdot)$表示所采用的激活函数。

$$\begin{aligned} F_{\text{smax}} &= MaxPool_c(F_{\text{conv}}),\\ F_{\text{savg}} &= AvgPool_c(F_{\text{conv}}),\\ F_s &= Conv_s(F_{\text{smax}}, F_{\text{savg}}),\\ M_s &= Sigmoid(F_s), \end{aligned}\tag{5}$$

式中：$AvgPool_c$表示沿通道维度的全局平均池化操作；$MaxPool_c$表示沿通道维度的全局最大池化操作；$Conv_s$表示用于生成空间注意力的卷积层；$Sigmoid(\cdot)$表示所采用的激活函数。

$$\begin{aligned} \boldsymbol{M}_r &= Conv_1^r(GAP(F_{\text{conv}})),\\ \boldsymbol{M}_d &= Conv_1^d(GAP(F_{\text{conv}})), \end{aligned}\tag{6}$$

式中：$Conv_1^r$表示通道注意力模块中的相关性卷积分支，其输入为小波卷积特征，能够有效编码通道间的全局依赖关系；$Conv_1^d$表示通道注意力模块中的差异性卷积分支，其目的是通过引入对比信息增强特征间的区分度；$GAP(\cdot)$表示全局平均池化（Global Average Pooling）。

$$\begin{aligned} A_r &= Sigmoid(\boldsymbol{M}_r \boldsymbol{M}_d),\\ A_d &= Sigmoid(\boldsymbol{M}_r^{\mathrm{T}} \boldsymbol{M}_d^{\mathrm{T}}), \end{aligned}\tag{7}$$

式中：$\boldsymbol{M}_r^{\mathrm{T}}$，$\boldsymbol{M}_d^{\mathrm{T}}$分别表示中间变量$\boldsymbol{M}_r$和$\boldsymbol{M}_d$的矩阵转置结果。通过矩阵交互操作分别生成相关性交互注意力A_r与差异性交互注意力A_d，以进一步强化通道间关系建模能力。

$$M_c = Sigmoid(\alpha A_r + (1-\alpha) A_d) \odot F_{\text{conv}},\tag{8}$$

式中：α为可学习的混合权重因子，一般通过$Sigmoid$激活得到，用于对A_r与A_d进行加权融合；$\odot$为逐元素乘法，用于对输入特征进行自适应重标定，从而增强关键裂缝特征表达并抑制无关背景响应。式（6）～式（8）共同组成了优化裂缝分割性能的细粒度通道注意力模块（Fine-Grained Channel Attention，FGCA）。

$$F_{\text{att}}=IWT([M_s,M_c]),\tag{9}$$

式中：$IWT(\cdot)$表示逆小波变换，其作用是将融合后的多频特征重构至空间域，从而实现原始空间信息的有效重建；F_{att}表示对FGCA模块提取所得多维特征进行逆小波变换的结果。

$$F_{\text{HFWA}B}=ReLU\left(Conv(Identity(X_i),F_{\text{att}})\right)+Conv_{1\times1}(X_0),\tag{10}$$

式中：$Identity(\cdot)$表示恒等映射分支，用于保留重叠图像块的原始嵌入特征，并与注意力增强分支的输出进行特征融合；F_{HFWAB}表示混合特征加权注意力模块对原始输入的特征提取结果。

2.1.3 混合前馈神经网络[25]

本文引入该模块并通过卷积操作实现隐式空间建模，从而减少对位置编码的依赖，提高模型对分辨率变化的适应能力。该混合前馈神经网络模块(Mixed Feed-Forward Network，MixFFN)由1×1卷积、深度可分离卷积和非线性激活单元构成，可进一步提升词元表示与局部建模能力。该模块的数学表示为

$$F_{\text{out}}=MixFFN(F_{\text{HFWAB}}),\tag{11}$$

式中：F_{out}表示混合特征加权注意力模块提取的特征经混合前馈神经网络模块处理后的结果。

2.2 前景特征引导传播

全局多维特征构建模块有效实现了裂缝频域信息表征与裂缝语义的局部增强，但要完成高精度裂缝分割，仍需保证跨尺度特征的持续传递与结构一致性约束。现有编码主干通常缺少显式的前景引导与特征补偿机制，裂缝相关特征在多尺度传递过程中多依赖隐式学习，难以对空间连续性与几何一致性形成稳定约束。尤其在低对比度或背景纹理复杂的场景下，裂缝分割模型易出现前景响应弥散、裂缝结构断裂等问题。其根本原因在于：裂缝分割任务中全局语义建模与细粒度结构保持之间存在内在耦合冲突，单一特征流在进行跨尺度语义聚合时，往往难以同时维持对局部结构细节的稳定表达。

基于上述分析，本文在主干编码分支之外，提出了一种面向稀疏前景保持的引导式特征传播机制，整体网络框架如图3所示。该模块融合高效非局部Transformer模块(Efficient Non-Local Transformer Block，ENLTB)[26]与轻量级自适应提取模块(Light-weight Adaptive Extraction，LAE)[27]，对特征传播过程进行显式调控，从而更好地建模跨区域空间依赖关系，构建适配稀疏前景的引导式特征传播结构。前景特征引导传播模块由引导分支与传播分支协同互补实现。

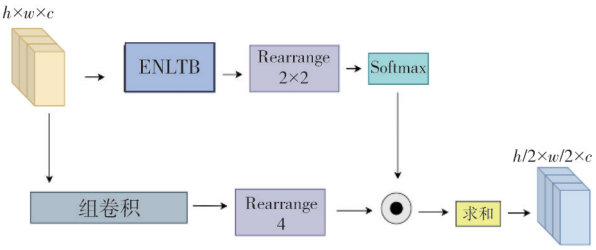


图3 前景特征引导传播网络架构示意图

Fig. 3 Schematic diagram of GFFP network architecture

2.2.1 引导分支

该分支以ENLTB为核心，借助高效非局部注意力机制在全局范围内聚合上下文信息，显式建立空间上离散裂缝片段间的长程关联，生成具备前景指向性的显著性权重以引导后续特征传播。其中，ENLTB模块的数学表示为

$$A_{\text{raw}}=ENLTB(X)$$

$$\Downarrow$$

$$\boldsymbol{Q}=\Theta(X),\boldsymbol{K}=\Phi(X),\boldsymbol{V}=\Gamma(X),$$

$$S_b(p,q)=\frac{\exp\left(\frac{(Q_{b,:,p},K_{b,:,p})}{\sqrt{d}}\right)}{\sum_{q'}\exp\left(\frac{(Q_{b,:,p},K_{b,:,p})}{\sqrt{d}}\right)},$$

$$\tilde{Y}=\boldsymbol{S}\boldsymbol{V},A_{\text{raw}}=\boldsymbol{W}_o(\tilde{Y})+X,\tag{12}$$

式中：$X\in\mathbf{R}^{B\times C\times H\times W}$表示输入的特征图，其中，$B$为批大小，$C$为通道数，$H\times W$为空间尺寸；$\boldsymbol{Q}$、$\boldsymbol{K}$和$\boldsymbol{V}$分别为经线性函数$\Theta(\cdot)$、$\Phi(\cdot)$和$\Gamma(\cdot)$映射获得的特征图的查询、键、值的矩阵；$(Q_{b,:,p},K_{b,:,p})$表示第$b$个样本中位置$p$与位置$q$向量的点积；$S_b(p,q)$表示第$b$个样本中位置$p$对位置$q$经$Softmax$函数归一化后的注意力权重；$\tilde{Y}$表示经注意力加权后的特征图，是全局依赖建模的结果；$\boldsymbol{W}_o$表示线性层输出的映射矩阵；A_{raw}表示加入残差连接后的注意力增强特征。

引导信息的生成公式可表示为

$$A=reshape(A_{\text{raw}},(B,C,H,W,4)),\tag{13}$$

式中：$reshape(\cdot,(B,C,H,W,4))$表示将特征图按照2×2划分子块后形成的4组子像素通道。

$$\tilde{A}(b,c,i,j,r)=\frac{\exp(A(b,c,i,j,r))}{\sum_{r=1}^{4}\exp(A(b,c,i,j,r))},\tag{14}$$

式中：$\tilde{A}$ 表示每个 2×2 子块通道经 *Softmax* 函数归一化后得到的权重。

2.2.2　传播分支

该分支主要负责对原始编码特征进行细粒度传播与重构，通过权重调控的方式引入引导信息，对不同空间位置与语义通道实现选择性增强。在跨尺度传递过程中，与裂缝结构相关的特征响应被优先保留并强化，而背景区域的冗余及干扰响应则得到有效抑制，从而在保持裂缝前景层级结构的同时，实现有效特征的持续聚合与增强。引导分支指导下的传播分支特征传递公式为

$$X_{ds}=GroupConv_{k=3,s=2,g=G}(X), \tag{15}$$

式中：X_{ds} 表示输入特征 X 经分组卷积下采样和扩展通道后所得的特征。

$$\tilde{X}(b,c,i,j,r)=X_{ds}(b,(r-1)C+c,i,j), \tag{16}$$

式中：$\tilde{X}$ 表示将 X_{ds} 按照子块位置重新排列后的特征。

$$Y(b,c,i,j,r)=\sum_{r=1}^{4}\tilde{X}(b,c,i,j,r)\cdot\tilde{A}(b,c,i,j,r)。 \tag{17}$$

GFFP 通过引导分支与传播分支的协同工作，将结构感知以及特征传播结合起来，并且利用自适应权重进行调节，在不破坏原始特征分布的前提下，增强前景表达的稳定性，从而弥补主干特征不足。

综上，全局多维特征构建模块从特征表示层面建模裂缝的整体结构与长程依赖，使高层语义特征具备良好的判别性与结构一致性。前景特征引导传播模块依托稳定的裂缝语义传递路径，确保裂缝相关区域在跨层传递过程中维持较高的语义连贯性与响应强度。二者在裂缝特征编码过程中高效配合，协同生成具备全局结构感知能力的高层特征，并在整个编码路径中稳定聚焦于稀疏裂缝前景。

3　实　验

为验证所提方法的有效性与性能优势，本文分别在自建数据集与公开数据集上开展了对比实验和消融实验。

3.1　数据集

3.1.1　自建隧道数据集

为应对复杂隧道环境下的裂缝分割难题，本文构建了一套专用隧道裂缝数据集。该数据集图像采集自太原—常家庄、离石—军渡等多座高速公路隧道，覆盖不同施工工艺与维护状态下的衬砌表面。

所采集的高分辨率图像尺寸为 4 096×1 024 像素，包含宽度在 0.1～1 mm 范围内的多种裂缝形态。在数据预处理阶段，首先，对现场采集图像进行统一拼接与筛选，接着，采用固定窗口方式将高分辨率图像裁剪为 512×512 像素的子图，以适应网络训练输入需求。数据集划分以裁剪后的子图为基本样本进行，训练集、验证集与测试集按照 7∶1∶2 的比例构建。由于各子图对应的空间区域及裂缝分布存在差异，因此划分后的数据在纹理背景、裂缝形态及结构特征方面具有一定多样性。该数据集共包含 553 张高分辨率图像，全部采用 LabelMe 工具完成像素级精细标注（见图 4），标注过程中对细小裂缝、分支结构及边缘区域进行了逐像素校验，以保证标签的准确性与结构完整性，从而为后续模型训练与性能评估提供可靠的数据基础。

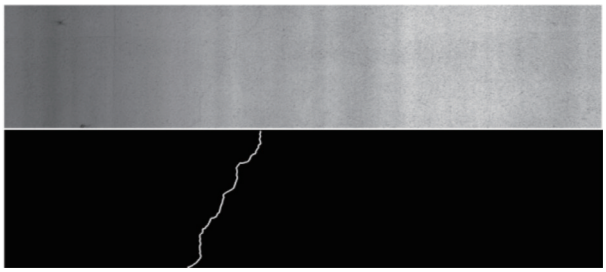


图 4　细小裂缝与其标注示例图

Fig. 4　Fine crack and its marking example

如图 4 所示，数据集中的裂缝普遍细窄，分布稀疏无序，前景占比极低，为隧道裂缝精确分割带来显著挑战。如图 5 所示，该数据集中背景纹理复杂多变，存在防火涂层、衬砌瓷砖、阴影、污渍、混凝土碎屑、渗水区域及人工修补痕迹等干扰。

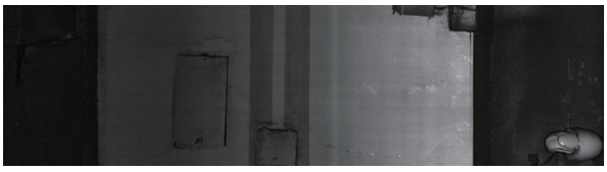


图 5　隧道裂缝图像的复杂背景示例图

Fig. 5　Example diagram of complex background of tunnel crack image

3.1.2　DeepCrack 数据集

作为裂缝分割领域广泛采用的公开基准数据集，DeepCrack 数据集包含 537 张高质量裂缝图像，覆盖多种裂缝类型、光照条件与背景场景。该数据集在裂缝边缘清晰度与结构多样性上具有较强代表性，可有效评估模型对裂缝细节的区分能力及泛化性能。实验中随机选取 300 张图像作为训练集，剩余 237 张

按7∶3的比例划分为验证集与测试集。

3.2 实验设计

本文实验运行于搭载NVIDIA Tesla V100 GPU的工作站上，模型训练总次数设置为300，Batch Size设为2，输入图像尺寸统一调整为512×512像素。初始学习率为1×10^{-3}，并通过Poly策略进行动态调整，动量与权重衰减系数分别设为0.9和5×10^{-4}，同时引入早停策略以兼顾训练效率与模型泛化性能。

为提升模型在复杂场景下的泛化能力，训练阶段采用了多种数据增强策略，包括随机尺度缩放、随机长宽比扰动、随机水平翻转、随机空间平移以及颜色扰动。其中，颜色扰动主要对图像的色调、饱和度与亮度进行随机变化，以增强模型对复杂光照条件及背景纹理干扰的适应能力。

考虑到裂缝分割任务中前景区域占比较低、类别分布不均衡的问题，本文采用Focal Loss和Dice Loss加权组合的损失函数进行联合优化。其中，Focal Loss能够降低易分类样本的影响，提升模型对难区分裂缝区域的关注能力。Dice Loss则从区域重叠角度约束预测结果，有助于增强细长裂缝结构的完整性与连续性。

3.3 评价指标

针对裂缝分割任务，本文选取精确率R_{Pre}(Precision)、召回率R_{Rec}(Recall)、$F1$值与交并比R_{IoU}(Intersection over Union，IoU)作为模型评估指标。其中，精确率衡量预测结果的可靠程度，对降低误检和提升分割可靠性具有重要意义。召回率表征实际裂缝区域被成功检出的比例，体现模型对裂缝目标的识别完整性。$F1$值综合平衡了精确率与召回率，能够有效评价模型在裂缝前景稀疏、样本不均衡及细微裂缝区域的分割性能。交并比则用于量化预测区域与真实标注区域的重叠程度，直观反映裂缝分割的空间精度。上述指标相互补充，共同实现对裂缝分割质量与工程应用价值的综合评价。

上述指标的计算公式为

$$R_{\mathrm{Pre}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}},$$

$$R_{\mathrm{Rec}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FN}}},$$

$$F1=2\frac{R_{\mathrm{Pre}}\cdot R_{\mathrm{Rec}}}{R_{\mathrm{Pre}}+R_{\mathrm{Rec}}},$$

$$R_{\mathrm{IoU}}=\frac{N_{\mathrm{TP}}}{N_{\mathrm{TP}}+N_{\mathrm{FP}}+N_{\mathrm{FN}}},\tag{18}$$

式中：真阳性N_{TP}(True Positive，TP)表示被正确预测为裂缝区域的样本数量；假阳性N_{FP}(False Positive，FP)表示被错误预测为裂缝区域的背景样本数量；假阴性N_{FN}(False Negative，FN)表示被错误预测为背景区域的实际裂缝样本数量。

在图像分割任务中，$F1$值与R_{IoU}是常用的核心评价指标。针对实际场景中背景干扰复杂、易引发误检的问题，本文引入R_{Pre}以评估模型对非裂缝区域的抑制能力。同时，结合R_{Rec}分析其漏检情况，从而实现对各方法更全面的综合性能评估。

3.4 对比实验

为验证所提方法在裂缝分割任务中的综合性能，与裂缝分割领域的专用网络及通用经典语义分割网络开展了横向对比，对比方法包括Unet[28]、CrackSegFormer[29]、segdec-net[30]和TransUnet[31]。为保证结果的公平性与可重复性，对比实验在相同数据集与统一评价标准下进行，分为定量性能评估与定性可视化分析两个维度，重点考察模型在复杂背景条件下对细长裂缝结构的精准识别能力、结构连续性保持效果以及整体分割质量。

本文方法与对比方法在两个数据集上对比试验的定量分析结果如表1和表2所示。本文所提方法在$F1$值和R_{IoU}的核心评价指标上均取得了明显优势，验证了所提方法在裂缝分割任务中的有效性与优越性。尽管表1结果显示本文方法的R_{Rec}指标略低于对比方法，但R_{Pre}、$F1$值及R_{IoU}三项指标均为最优，充分证明基于该方法构建的分割模型只是裂缝像素点预测较为保守，但整体上具备精准且可靠的像素级裂缝分割性能。

表1 隧道裂缝自建数据集上的对比实验结果

Tab. 1 Comparative experimental results on the self-built dataset of tunnel cracks

方法	R_{Pre}	R_{Rec}	$F1$	R_{IoU}
segdec-net	0.581 7	0.698 1	0.634 6	0.457 4
segformer	0.528 4	0.674 8	0.592 7	0.424 8
Unet	0.557 9	0.898 9	0.688 5	0.525 0
TransUnet	0.677 2	0.791 9	0.730 1	0.579 0
本文方法	0.693 3	0.781 6	0.734 8	0.616 5

从整体趋势可以看出，现有分割网络在裂缝前景稀疏、空间分布高度随机且无明显规律的场景下，裂缝区域易受大面积背景干扰，导致特征响应不连续，使得分割结果在区域覆盖完整性与

结构连续性上存在明显局限。

综合分析表 1 和表 2 可知，本文方法在两个数据集上均呈现出更为稳定的分割性能，尤其在裂缝占比极低或分布随机的样本中，仍能保持完整且连续的裂缝区域响应，有效缓解了因前景稀疏与背景复杂带来的分割挑战。

为进一步直观对比不同方法在复杂裂缝场景下的分割效果，本文给出各方法分割的可视化结果，如图 6 和图 7 所示。

表 2　DeepCrack 数据集上的对比实验结果

Tab. 2　Comparative experimental results on the public DeepCrack dataset

方法	R_{Pre}	R_{Rec}	$F1$	R_{IoU}
segdec-net	0.746 8	0.749 2	0.748 0	0.594 8
segformer	0.803 9	0.750 5	0.776 3	0.659 1
Unet	0.785 6	0.723 5	0.753 3	0.655 3
TransUnet	0.821 3	0.766 0	0.792 7	0.675 9
本文方法	0.862 7	0.773 2	0.815 5	0.688 5

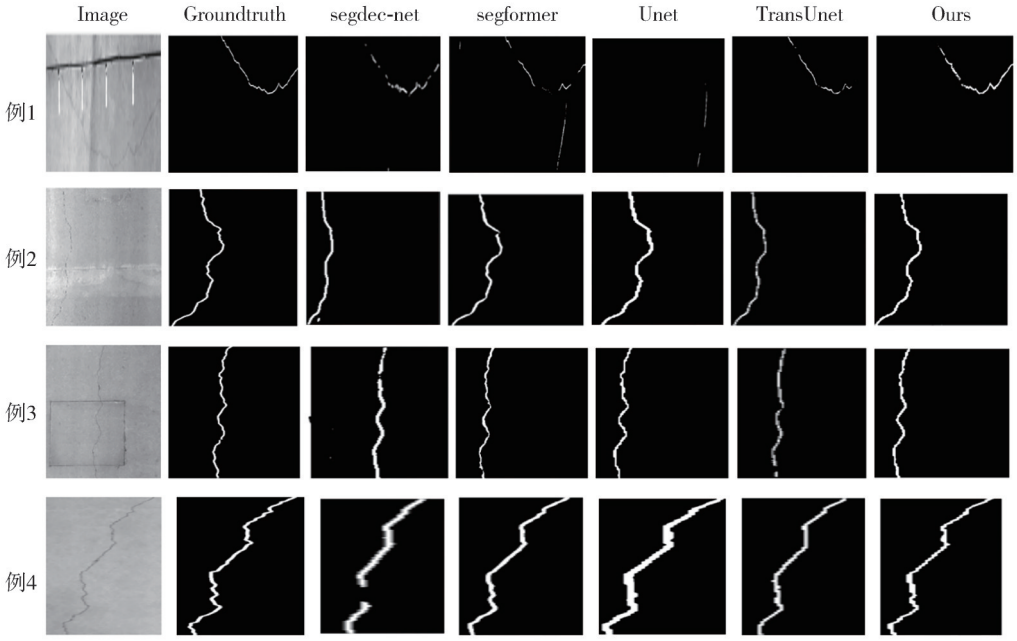


图 6　隧道裂缝自建数据集的裂缝分割效果

Fig. 6　Crack segmentation effect for tunnel crack self-built dataset

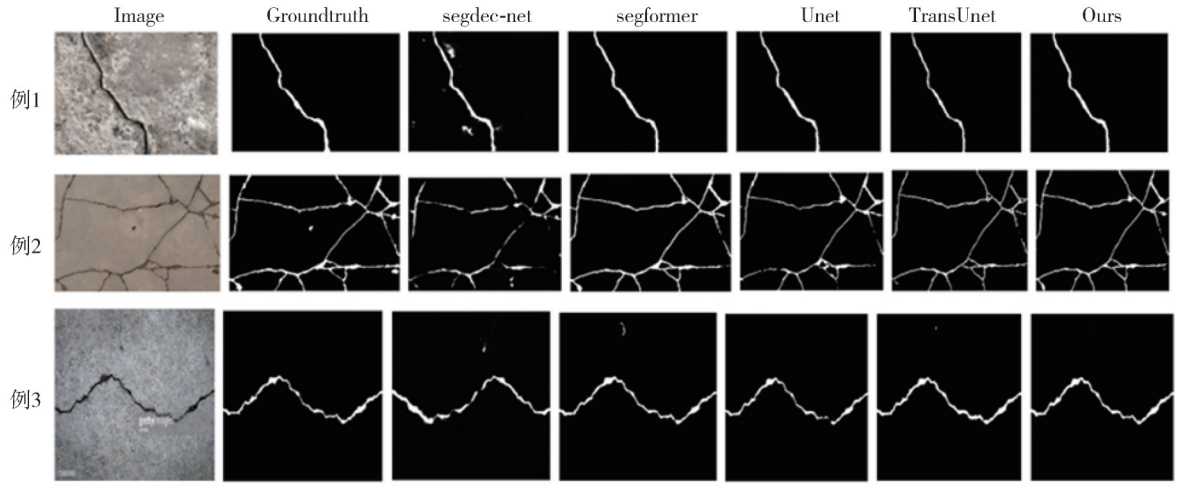


图 7　DeepCrack 数据集的裂缝分割效果

Fig. 7　Crack segmentation effect for DeepCrack dataset

图 6 展示了不同裂缝分割方法在自建数据集上的可视化结果对比。对于第一行中的低对比度细小裂缝场景，segdec-net 存在明显断裂现象，SegFormer 和 UNet 出现漏检，对细小分支结构响应不足，TransUnet 虽能够检测主体区域，但末端细节恢复不完整，而本文方法能够较完整地保留

裂缝整体走向及弯曲细节。第二行中，各方法均能够识别主体裂缝，但部分方法存在边缘形变或宽度不一致的问题，其中TransUnet结果偏细、局部连续性较弱，本文方法则在结构平滑性与边缘稳定性方面更接近真实标注。第三行复杂背景场景下，segdec-net存在局部断裂与噪声干扰，TransUnet出现局部断裂，而本文方法能够有效抑制背景干扰并保持裂缝主干结构连续。第四行中，segdec-net与UNet存在明显过分割现象，SegFormer与TransUnet在局部拐角区域存在一定平滑问题，而本文方法能够较好地恢复裂缝弯折特征与边缘细节。

图7展示了不同裂缝分割方法在DeepCrack数据集上的可视化结果对比。第一行中，裂缝整体较细且背景纹理较复杂，segdec-net存在局部噪声与边缘不稳定问题，其余方法都完整地提取了裂缝主体；第二行为多分支裂缝场景，各方法均能够识别主要裂缝区域，但segdec-net和Unet在交叉区域存在局部缺失，部分方法在细小分支处出现结构不连续或边缘偏粗的问题；第三行中裂缝呈明显弯曲形态，且背景存在较强纹理干扰，segdec-net出现较明显的断裂现象，SegFormer存在少量孤立误检区域，UNet在裂缝中间存在轻微断裂现象，TransUnet虽然能够保持整体结构，但裂缝上面存在轻微的误检，而本文方法在保持裂缝整体走向的同时，能够较好地恢复弯曲边缘细节，结果与真实标注更加接近。

综合定量与定性实验结果表明：本文所提方法在裂缝前景稀疏、背景复杂的分割场景中具备更优异的整体适应性，其优势不仅体现在分割结果的稳定性，更体现在对随机分布裂缝区域的完整提取与精准分割能力方面。

本文分割模型的推理速度可达12.5帧/s（Frames Per Second，FPS）。结合隧道图像采集密度（约96张/m），模型可支持0.5 km/h的实时处理速度，且能稳定输出裂缝长度、宽度等关键量化特征。若进一步采用多进程并发策略提升处理效率，该方法可充分满足实际工程对时效性与可靠性的双重需求。

3.5 消融实验

为验证本文编码阶段改进策略的有效性，以UNet作为基线模型，围绕全局多维特征构建（DWFormer）与前景特征引导传播（GFFP）两个核心模块，设计并开展了层层递进的消融实验。为保障实验结果的可靠性与可比性，消融实验采用与主实验完全一致的数据集、训练参数及运行环境，确保不同模型结构之间的性能对比具有严谨的可信性。不同结构配置下，模型对裂缝结构的建模性能如表3和表4所示，各模型的具体配置如下：

1）Baseline：不引入DWFormer或GFFP的基础网络；

2）Baseline＋OPE：仅引入DWFormer中的OPE模块（包含MixFFN）；

3）Baseline＋OPE＋HFWAB：引入DWFormer中的OPE与HFWAB模块（无FGCA）；

4）Baseline＋OPE＋HFWAB＋FGCA：完整引入DWFormer，即OPE＋HFWAB＋MixFFN；

5）Baseline＋LAE：仅引入GFFP中的局部特征增强模块（LAE）；

6）Baseline＋LAE＋ENLTB：引入GFFP的全部子模块，包括LAE与ENLTB；

7）Baseline＋DWFormer：单独引入完整DWFormer模块进行消融实验；

8）Baseline＋GFFP：单独引入完整GFFP模块进行消融实验；

9）Baseline＋DWFormer＋GFFP：同时引入DWFormer与GFFP，即本文所提方法。

表3 隧道裂缝自建数据集上的消融实验结果

Tab. 3 Ablation results on the self-built dataset of tunnel cracks

消融项	R_{IoU}	R_{Rec}	$F1$	R_{Pre}
Baseline	0.525 0	0.898 9	0.688 5	0.557 9
Baselin＋OPE	0.536 2	0.870 5	0.692 7	0.575 2
Baseline＋OPE＋HFWAB	0.560 1	0.832 1	0.710 3	0.619 6
Baseline＋OPE＋HFWAB＋FGCA	0.580 7	0.797 8	0.718 0	0.652 7
baseline＋LAE	0.586 4	0.816 3	0.705 0	0.620 4
Baseline＋LAE＋ENLTB	0.604 2	0.772 4	0.717 4	0.669 7
Baseline＋DWFormer	0.580 7	0.797 8	0.718 0	0.652 7
Baseline＋GFFP	0.604 2	0.772 4	0.717 4	0.669 7
Baseline＋DWFormer＋GFFP	0.616 5	0.781 6	0.734 8	0.693 3

表4 DeepCrack数据集上的消融实验结果

Tab. 4 Ablation results on the public DeepCrack dataset

消融项	R_{IoU}	R_{Rec}	$F1$	R_{Pre}
Baseline	0.655 3	0.723 5	0.753 3	0.785 6
Baselin＋OPE	0.663 1	0.732 2	0.769 5	0.810 8
Baseline＋OPE＋HFWAB	0.669 8	0.748 5	0.772 4	0.797 8
Baseline＋OPE＋HFWAB＋FGCA	0.678 2	0.768 0	0.794 1	0.822 0
baseline＋LAE	0.669 3	0.736 0	0.763 6	0.793 3
Baseline＋LAE＋ENLTB	0.676 7	0.752 9	0.785 4	0.820 8
Baseline＋DWFormer	0.678 2	0.768 0	0.794 1	0.822 0
Baseline＋GFFP	0.676 7	0.752 9	0.785 4	0.820 8
Baseline＋DWFormer＋GFFP	0.688 5	0.773 2	0.815 5	0.862 7

首先，对DWFormer模块的内部结构开展递进式消融分析。DWFormer模块由OPE、HFWAB及FGCA三个子模块逐步构建而成，其内部设计呈现出从空间连续性建模、频域判别增强到通道语义重组的递进逻辑。实验结果表明，随着各子模块的逐步加入，模型对裂缝结构的整体感知能力持续提升，充分验证了DWFormer内部多层次建模策略在增强裂缝高层语义稳定性和提升裂缝结构一致性方面的有效性。

其次，对GFFP模块的内部结构进行递进式消融分析。GFFP模块以LAE为基础引导结构，通过进一步嵌入高效非局部建模模块ENLTB，对特征传递过程进行结构化重构与优化。实验结果表明：仅依赖LAE局部引导机制时，裂缝前景特征在多尺度编码过程中的保持能力仍存在明显局限；引入ENLTB模块后，模型能够在更大范围内建模裂缝结构间的关联关系，使裂缝前景响应在特征传递过程中更具稳定性，充分验证了GFFP模块的引导式特征传播结构对于隧道裂缝分割任务的有效性。

再次，从整体层面考察DWFormer与GFFP两个核心模块独立引入时的效果。实验结果表明，无论是单独引入DWFormer模块，还是单独嵌入GFFP模块，模型的裂缝分割性能均较基线UNet模型实现稳定提升。这一结果说明，在编码阶段增强模型对裂缝特征的全局结构感知能力，或是在特征传递过程中对稀疏的裂缝前景特征进行显式引导，这两种改进思路均能有效缓解裂缝像素占比极低、空间分布随机所带来的建模难题，进一步印证了引入这两个核心模块进行隧道裂缝分割的有效性。

最后，本文所提方法（即同时将DWFormer与GFFP模块引入基线模型）在裂缝分割整体性能上取得了优异的表现。该结果表明，两个核心模块在编码阶段形成了清晰且互补的功能分工：DWFormer模块侧重于构建具备全局结构感知能力的高层裂缝特征表示，精准捕捉裂缝的整体结构与关联关系；GFFP模块则在特征传递过程中对稀疏分布的裂缝前景特征进行持续引导与强化，有效避免了前景信息的流失。通过这种协同建模策略，网络在编码端能够同时兼顾特征表达的精准性与裂缝前景信息的完整性，进而提高了隧道裂缝分割的准确性与可靠性。

4　结　论

针对裂缝前景稀疏与背景复杂所带来的图像分割难题，本文提出了一种基于双路径的裂缝分割网络模型。一方面，在全局多维特征建模路径中引入空间、频域和通道信息的联合建模策略，有效提升了随机分布裂缝在特征空间中的表达一致性。另一方面，在前景特征引导传播路径中，构建引导式特征传播机制，对裂缝相关特征信息进行持续强化与稳定传递，缓解前景裂缝特征在多尺度编码过程中逐步弱化的问题。两条路径相互配合、协同作用，在编码阶段实现裂缝特征的稳定表达与持续强化，进而显著提高了隧道裂缝分割的准确性与可靠性。对比实验与消融实验结果表明，所提方法在隧道裂缝分割任务中，相较于其他主流方法及各消融模型均表现出更优异的分割性能。

利益冲突声明/Conflict of Interests

所有作者声明不存在利益冲突。

All authors declare no relevant conflict of interests.

参考文献：

[1] 刘宇飞，樊健生，聂建国，等. 结构表面裂缝数字图像法识别研究综述与前景展望[J]. 土木工程学报，2021，54(6)：79-98.
Liu Yufei，Fan Jiansheng，Nie Jianguo，et al. Review and prospect of digital-image-based crack detection of structure surface[J]. China Civil Engineering Journal，2021，54(6)：79-98.（in Chinese）

[2] 高建章，钟泰枫，齐建波，等. 长期服役管线隧道衬砌结构病害特征与规律分析[J]. 中国安全生产科学技术，2024，20(7)：78-86.
Gao Jianzhang，Zhong Taifeng，Qi Jianbo，et al. Analysis on structural disease characteristics and laws of lining in long-serving pipeline tunnel[J]. Journal of Safety Science and Technology，2024，20(7)：78-86.（in Chinese）

[3] Dong C Z，Catbas F N. A review of computer vision - based structural health monitoring at local and global levels [J]. Structural Health Monitoring，2021，20(2)：692-743.

[4] 路耀邦，刘永胜，樊晓东. 地铁隧道结构表观病害快速检测方法与应用[J]. 隧道建设(中英文)，2021，41(S2)：655-663.

Lu Yaobang, Liu Yongsheng, Fan Xiaodong. Rapid detection method for surface defect for metro tunnel structure and its application[J]. Tunnel Construction, 2021, 41(S2): 655-663. (in Chinese)

[5] Kalfarisi R, Wu Z Y, Soh K. Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization [J]. Journal of Computing in Civil Engineering, 2020, 34(3): 04020010.

[6] 王康, 刘健, 吕高航, 等. 基于红外图像增强的衬砌裂损渗漏水识别方法及工程应用[J]. 应用基础与工程科学学报, 2023, 31(6): 1444-1460.
Wang Kang, Liu Jian, Lü Gaohang, et al. Method for identifying leakage water in the tunnel lining based on infrared image enhancement and engineering application [J]. Journal of Basic Science and Engineering, 2023, 31(6): 1444-1460. (in Chinese)

[7] 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.

[8] Lee J S, Hwang S H, Choi I Y, et al. Estimation of crack width based on shape-sensitive kernels and semantic segmentation[J]. Structural Control and Health Monitoring, 2020, 27(4): e2504.

[9] 邝先验, 徐姚明, 雷卉, 等. 基于轻量级Transformer的隧道裂缝分割[J]. 铁道科学与工程学报, 2024, 21(8): 3421-3433.
Kuang Xianyan, Xu Yaoming, Lei Hui, et al. Tunnel crack segmentation based on lightweight Transformer[J]. Journal of Railway Science and Engineering, 2024, 21(8): 3421-3433. (in Chinese)

[10] 王森, 伍星, 张印辉, 等. 基于深度学习的全卷积网络图像裂纹检测[J]. 计算机辅助设计与图形学学报, 2018, 30(5): 859-867.
Wang Sen, Wu Xing, Zhang Yinhui, et al. Image crack detection with fully convolutional network based on deep learning [J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30(5): 859-867. (in Chinese)

[11] 常惠, 饶志强, 赵玉林, 等. 基于改进U-Net网络的隧道裂缝分割算法研究[J]. 计算机工程与应用, 2021, 57(22): 215-222.
Chang Hui, Rao Zhiqiang, Zhao Yulin, et al. Research on tunnel crack segmentation algorithm based on improved U-Net network[J]. Computer Engineering and Applications, 2021, 57(22): 215-222. (in Chinese)

[12] Liu Z Q, Cao Y W, Wang Y Z, et al. Computer vision-based concrete crack detection using U-Net fully convolutional networks[J]. Automation in Construction, 2019, 104: 129-139.

[13] Sun X Z, Xie Y C, Jiang L M, et al. DMA-net: DeepLab with multi-scale attention for pavement crack segmentation[J]. IEEE Transactions on Intelligent Transportation Systems, 2022, 23(10): 18392-18403.

[14] Zhao H S, Shi J P, Qi X J, et al. Pyramid scene parsing network [C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017: 6230-6239.

[15] Xu Z S, Guan H Y, Kang J, et al. Pavement crack detection from CCD images with a locally enhanced transformer network[J]. International Journal of Applied Earth Observation and Geoinformation, 2022, 110: 102825.

[16] Lin C, M Tian D X, Duan X T, et al. TransCrack: revisiting fine-grained road crack detection with a transformer design [J]. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 2023, 381(2254): 20220172.

[17] Wang W J, Su C. Automatic concrete crack segmentation model based on transformer [J]. Automation in Construction, 2022, 139: 104275.

[18] Wang C, Liu H B, An X Y, et al. SwinCrack: Pavement crack detection using convolutional swin-transformer network [J]. Digital Signal Processing, 2024, 145: 104297.

[19] 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.

[20] 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, 152: 104894.

[21] Zhou Z, Zhang J J, Gong C J. Hybrid semantic segmentation for tunnel lining cracks based on Swin Transformer and convolutional neural network[J]. Computer-Aided Civil and Infrastructure Engineering, 2023, 38(17): 2491-2510.

[22] Zhang E H, Shao L H, Wang Y. Unifying transformer and convolution for dam crack detection[J]. Automation in Construction, 2023, 147: 104712.

[23] Xie E Z, Wang W H, Yu Z D, et al. SegFormer: simple and efficient design for semantic segmentation with transformers[PP/OL]. V3. arXiv (2021-10-28)[2026-04-08]. https://doi.org/10.48550/arXiv.2105.15203.

[24] Fan C M, Liu T J, Liu K H. Half wavelet attention on M-net+ for low-light image enhancement [C]//2022

IEEE International Conference on Image Processing (ICIP), 2022: 3878-3882.

[25] Chu X X, Tian Z, Zhang B, et al. Conditional positional encodings for vision transformers[PP/OL]. V3. arXiv (2023-02-13)[2026-04-08]. https://doi.org/10.48550/arXiv.2102.10882.

[26] Hu J T, Chen S Y, Pan Z Y, et al. Perspective+ Unet: enhancing segmentation with bi-path fusion and efficient non-local attention for superior receptive fields[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2024: 499-509.

[27] Yu Z W, Guan Q, Yang J M, et al. LSM-YOLO: a compact and Effective ROI detector for medical detection [C]//Neural Information Processing. 2025: 30-44.

[28] Ronneberger O. Invited talk: U-Net convolutional networks for biomedical image segmentation [C]//Bildverarbeitung für die Medizin 2017. 2017: 3.

[29] Li H B, Zhang H W, Zhu H, et al. Automatic crack detection on concrete and asphalt surfaces using semantic segmentation network with hierarchical Transformer[J]. Engineering Structures, 2024, 307: 117903.

[30] Tabernik D, Šuc M, Skočaj D. Automated detection and segmentation of cracks in concrete surfaces using joined segmentation and classification deep neural network[J]. Construction and Building Materials, 2023, 408: 133582.

[31] Chen J N, Mei J R, Li X H, et al. TransUNet: Rethinking the U-Net architecture design for medical image segmentation through the lens of transformers[J]. Medical Image Analysis, 2024, 97: 103280.