文章编号：1674-2974（2026）04-0019-10 DOI：10.16339/j.cnki.hdxbzkb.2026263

动态蛇形卷积结合可变形注意力Transformer增强的裂缝检测

陈永[1,2†]，周建宇[1]，安卓奥博[1]
（1. 兰州交通大学 电子与信息工程学院，甘肃 兰州 730070；
2. 甘肃省人工智能与图形图像处理工程研究中心，甘肃 兰州 730070）

摘 要：针对高速铁路无砟轨道板裂缝复杂多样，现有裂缝检测方法存在裂缝目标特征提取不充分、检测不连续及检测精度不高的问题，提出了一种动态蛇形卷积结合可变形注意力Transformer增强的裂缝检测方法.首先，在Mask2Former分割模型的基础上，提出动态蛇形卷积改进的ResNet-50作为裂缝特征提取主干网络，利用动态蛇形卷积的连续预测能力，提高特征提取网络对于复杂多样无砟轨道裂缝几何特征的拟合效果，克服裂缝检测不连续的问题.然后，设计可变形注意力Transformer解码器模块，使模型能够动态适应无砟轨道裂缝复杂多样的结构特征变化，增强捕获全局上下文信息的能力，解决裂缝识别不准确的问题.其次，在Transformer解码器中设计改进前馈神经网络（feedforward network，FFN），通过学习裂缝周围的局部信息，使其能够更准确地捕捉裂缝局部细节特征，提高检测精度.最后，将Transformer解码器输出与像素解码器输出融合，得到裂缝检测结果.裂缝检测实验结果表明，所提方法可以准确地检测出不同形状的裂缝，较原始的Mask2Former模型平均准确率提升了6.34个百分点，平均召回率提升了4.70个百分点，F_1达到了94.30%.所提方法对于铁路无砟轨道板表面裂缝检测具有较好的性能，提高了裂缝的检测精度，主客观评价均优于对比方法.

关键词：无砟轨道；裂缝检测；动态蛇形卷积；Transformer；缺陷检测
中图分类号：TP391.4 **文献标志码**：A

Crack detection enhanced by combining dynamic snake convolution and deformable attention Transformer

CHEN Yong[1,2†]，ZHOU Jianyu[1]，AN Zhuoaobo[1]
（1. School of Electronic and Information Engineering，Lanzhou Jiaotong University，Lanzhou 730070，China；
2. Gansu Provincial Engineering Research Center for Artificial Intelligence and Graphics & Image Processing，Lanzhou 730070，China）

Abstract：In response to the diverse and complex nature of cracks in ballastless track slabs，the existing crack detection methods suffer from insufficient extraction of crack target features，discontinuity detection and low accuracy. A crack detection enhanced method is proposed based on dynamic snake convolution and deformable attention Transformer. Initially，the approach enhances the ResNet-50 feature extraction backbone by using a

* **收稿日期**：2024-04-15
基金项目：国家自然科学基金资助项目（62462043，61963023），National Natural Science Foundation of China（62462043，61963023）
作者简介：陈永（1979—），男，甘肃武威人，兰州交通大学教授，博士
† 通信联系人，E-mail：edukeylab@126.com

dynamic snake convolution on the basis of the Mask2Former segmentation model. The dynamic snake convolution, with its continuous prediction capability, improves the feature extraction network's ability to fit the geometric features of diverse ballastless track cracks, enhancing the extraction of crack features and overcoming discontinuity detection issues. Subsequently, a deformable attention Transformer decoder module is designed to enable the model to dynamically adapt to changes in local features of ballastless track cracks, in order to enhance the ability to capture global contextual information and improve the accuracy of crack recognition. Furthermore, an improved feedforward network (FFN) is incorporated into the Transformer decoder to learn local information around the cracks, facilitating more accurate capture of local details and improving detection accuracy. Finally, the output of the Transformer decoder is fused with the pixel decoder output to obtain the crack detection results. Experimental results demonstrate that the proposed method accurately detects cracks of various shapes. It achieves a 6.34 percentage points increase in average accuracy, a 4.70 percentage points increase in average recall, and an F_1-score of 94.30% compared with the original Mask2Former model. The proposed method exhibits superior performance in the detection of surface cracks on ballastless track slabs, enhancing detection accuracy and outperforming comparative methods in both subjective and objective evaluations.

Key words: ballastless track; crack detection; dynamic snake convolution; Transformer; defect detection

随着我国高速铁路建设的日益发展，轨道作为高速铁路的行车基础，直接承受着机车车辆巨大的荷载压力，其安全性、平顺性对于保障高速列车行车安全至关重要.无砟轨道是高速铁路轨道的重要组成部分，由混凝土或沥青混合料构成，取代了传统的散粒碎石道床.无砟轨道因其稳定性高及维护工作量少等特点，被广泛应用于高速铁路施工建设中[1].然而，受混凝土材料脆性、路基不均匀沉降等因素影响，无砟轨道板易产生复杂多样的裂缝[2].此外，裂缝持续受到列车的冲击振动以及疲劳载荷作用的影响很容易扩展，逐渐演变为尺寸及形状差异更大的裂缝，会大幅降低轨道耐久性[3]，严重时危及行车安全.因而，实现对无砟轨道裂缝的准确检测对于保障铁路线路的安全运维具有重要意义.

目前，基于深度学习的检测方法被广泛应用于裂缝检测.Ren 等[4]提出了一种改进的深度全卷积 CrackSegNet 裂缝检测网络，该方法通过空洞卷积来获取更大的感受野检测范围，但该方法多次叠加多个相同空洞率的卷积核对裂缝进行检测，会造成部分裂缝像素点始终未参与运算，出现裂缝漏检的问题. Lau 等[5]提出了一种基于 U-Net 的裂缝检测模型来提高裂缝检测精度，但该方法中编码器采用局部特征提取操作，限制了裂缝全局特征的提取，检测结果易出现裂缝检测不连续的问题.Fu 等[6]基于 DeepLab V3+网络，通过引入密集连接的空间金字塔池化模块对桥梁裂缝进行语义分割，但该方法双线性上采样忽略了像素之间预测的相关性，导致出现裂缝分割精度不高的问题.Tang 等[7]提出了一种用于裂缝分割的编解码器网络 EDNet，但是该方法未考虑不同裂缝尺度差异的问题，存在误分割的问题.Yuan 等[8]提出了一种并行注意力机制和多尺度特征融合的裂缝分割方法，一定程度上提高了裂缝特征提取能力，但该方法因采用普通卷积操作，噪声对其影响较大，结果易出现误分割的问题. Li 等[9]提出了一种基于卷积神经网络的 SCCDNet 裂纹分割网络，该方法通过密集连接和空间维度注意力模块提高对裂缝的识别性能，但该方法中采用深度可分离卷积进行逐点卷积，导致检测结果易出现裂缝不连续、漏检的问题.Shi 等[10]提出了一种多级对比学习的分割网络来提升裂缝检测能力，但该方法中过多依赖坐标信息，会使模型忽略局部特征，导致细小裂缝检测精度不佳.Beyene 等[11]提出一种基于注意力机制的密集连接网络来进行裂缝检测，但该方法忽视了裂缝局部信息，从而影响了裂缝分割的连续性.Xiong 等[12]提出了一种融合金字塔池化和 Transformer 的双编码路径裂缝检测方法，解决了裂

缝分割间断的问题，但该方法中卷积核在特征图上的跳跃式操作会平滑掉微小裂缝的信息，导致分割精度下降.

综上所述，针对现有裂缝分割方法存在裂缝目标特征提取不充分、检测不连续及检测精度不高的问题，提出了一种动态蛇形卷积结合可变形注意力 Transformer 增强的裂缝检测网络.

1　裂缝检测网络

1.1　整体网络模型

在 Mask2Former 语义分割模型的基础上，本文提出了一种动态蛇形卷积结合可变形注意力 Transformer 增强的裂缝检测模型. 原始 Mask2Former 模型[13]通过使用 Mask 掩膜注意力，以及引入 Transformer 解码器来提高目标识别效果和分割精度，但 Mask 掩膜注意力采用普通卷积操作，像素之间的上下文关联较为薄弱，难以适应无砟轨道裂缝复杂多样的特点，因而本文基于 Mask2Former 提出了一种动态蛇形卷积结合可变形注意力 Transformer 增强的裂缝检测模型，整体模型如图 1 所示. 所提模型工作时，首先，采用设计的动态蛇形卷积 ResNet-50 网络作为主干网络，通过变形卷积核适应无砟轨道复杂多样的裂缝形状，充分提取裂缝特征. 然后，像素解码器将特征提取网络的输出生成多尺度特征图，并输入到 Transformer 解码器中. 其次，提出改进的可变形注意力 Transformer 解码器，增强捕获全局上下文信息的能力，克服裂缝识别不准确的问题. 最后，将像素解码器和 Transformer 解码器的输出融合，得到最终的裂缝检测结果.

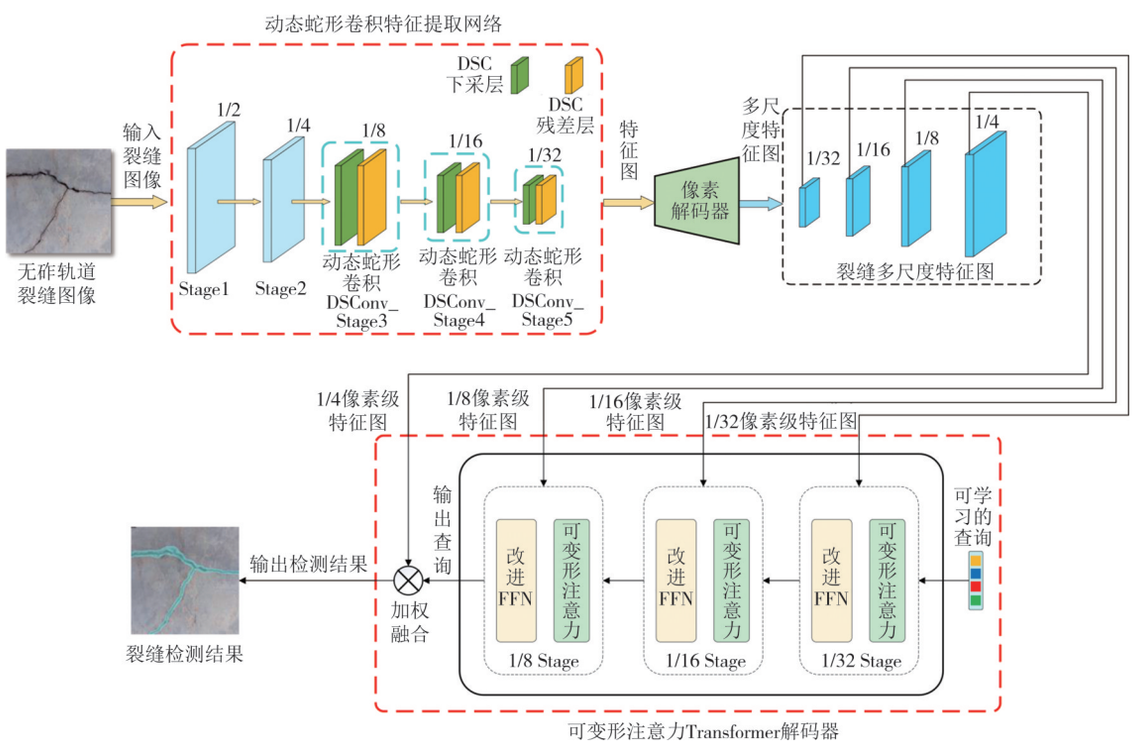


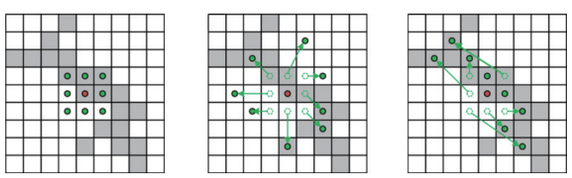
图 1　网络整体模型结构图

Fig.1　Overall network architecture diagram

1.2　动态蛇形卷积特征提取网络

在图 1 中，首先采用动态蛇形卷积特征提取网络进行裂缝特征提取. 原 Mask2Former 以 ResNet-50[14]作为特征提取网络，它使用标准的二维卷积对图像中的目标进行特征提取. 然而，由于无砟轨道裂缝形状复杂多样，标准卷积使用固定大小的二维卷积核，在裂缝检测时，其感受野大小是固定的，无法有效捕捉裂缝形变特征，如图 2(a)所示. 可变形卷积相对于标准卷积进行了改进，其通过二维偏移量来学习目标的几何变化，但偏移量采用无规律踩点方式获得，会导致可变形卷积的感受野出现偏离检测目标区域的情况，上述缺点导致可变形卷积在裂缝检测时会出现检测间断、不连续的情况[15]，如图 2(b)所示. 为了更好地拟合复杂多样的无砟轨道裂

缝,本文采用动态蛇形卷积作为裂缝特征提取主干网络的卷积操作,动态蛇形卷积[16]是在可变形卷积的基础上,将连续性约束策略加入卷积核的设计中,每一个卷积位置都由其前一个位置作为基准,从而保证特征提取的连续性,如图2(c)所示.



(a)普通卷积 (b)可变形卷积 (c)本文动态蛇形卷积

图2 不同卷积感受野对比

Fig.2 Comparison of receptive fields of different convolutions

动态蛇形卷积首先将标准卷积核分别沿着x轴和y轴方向展平.然后,将拉直后的卷积核的每个网格的位置表示为$K_{i\pm c}=\left(x_{i\pm c},y_{i\pm c}\right)$,其中,$c=\{0,1,2,3,4\}$表示距离中心网格的水平距离.接着,开始迭代计算卷积核K变形后各个网格的位置,从中心网格位置开始,其他网格的位置基于其前一个网格的位置.接着,每个网格的偏移量需要与其之前所有网格的偏移量Δ求和,于是卷积核沿着x轴方向上的变形见式(1):

$$K_{i\pm c}=\begin{cases}\left(x_{i+c},y_{i+c}\right)=\left(x_i+c,y_i+\sum_i^{i+c}\Delta y\right)\\ \left(x_{i-c},y_{i-c}\right)=\left(x_i-c,y_i+\sum_{i-c}^{i}\Delta y\right)\end{cases}\tag{1}$$

同样,卷积核沿着y轴方向的变形,见式(2):

$$K_{j\pm c}=\begin{cases}\left(x_{j+c},y_{j+c}\right)=\left(x_j+\sum_j^{j+c}\Delta x,y_i+c\right)\\ \left(x_{j-c},y_{j-c}\right)=\left(x_j+\sum_{j-c}^{j}\Delta x,y_i-c\right)\end{cases}\tag{2}$$

由于式(1)和式(2)中的偏移量Δx和Δy通常是小数,需要进一步采用双线性插值转化为真实无砟轨道裂缝像素的位置坐标,插值过程见式(3):

$$K=\sum_{K'}B\left(K,K'\right)\cdot K'\tag{3}$$

式中:K为式(1)和式(2)计算出的卷积核网格的非整数位置;K'则是输入特征图中像素点的整数位置;$B(\cdot)$为双线性插值函数,其包含沿x轴方向插值函数$b\left(K_x,K_x'\right)$和y轴方向插值函数$b\left(K_y,K_y'\right)$,如式(4)所示:

$$B\left(K,K'\right)=b\left(K_x,K_x'\right)\cdot b\left(K_y,K_y'\right)\tag{4}$$

在动态蛇形卷积特征提取过程中,动态蛇形卷积通过学习输入裂缝的几何形状等参数,然后调整卷积核的形状,以便更好地适应裂缝特征,并输出相应的裂缝特征提取图,该过程如图3所示.

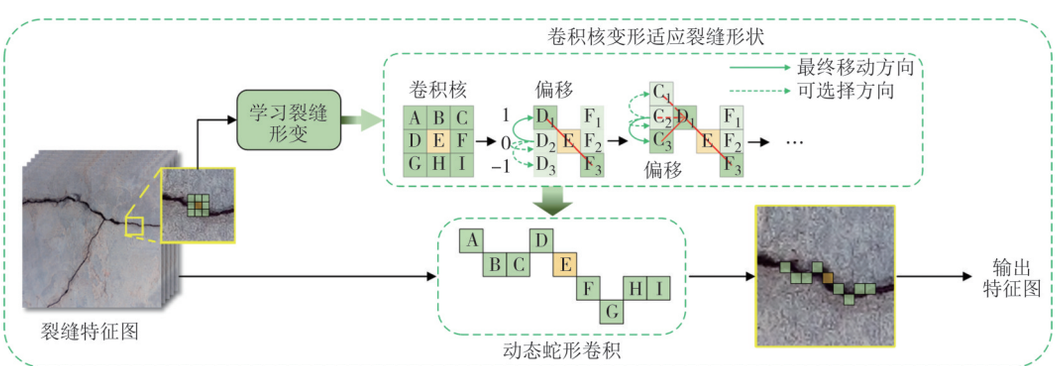
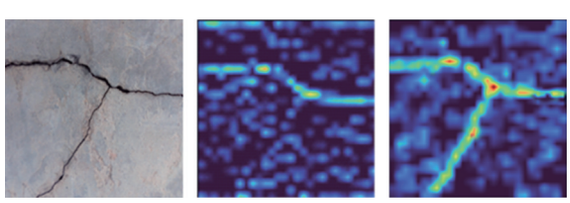


图3 动态蛇形卷积核适应裂缝形状示意图

Fig.3 Diagram of dynamic snake convolution kernel adapted to crack shape

为了直观说明改进后引入动态蛇形卷积主干特征提取的有效性,下面进行裂缝特征热力图可视化比较,如图4所示.图中,颜色越偏黄绿色,表明特征网络对于裂缝的关注度越大,而蓝紫色表示非裂缝区域.图4(b)是改进ResNet-50网络前的提取结果,可以看出原始网络对于裂缝无法全面聚焦.图4(c)为本文所提动态蛇形卷积网络结果,可以看出所提方法对于裂缝关注度更高,能够有效拟合裂缝几何特征.



(a)裂缝原图 (b)原特征提取网络 (c)本文方法

图4 特征提取可视化热力图比较

Fig.4 Comparison of feature extraction visualization heat map

1.3 可变形注意力Transformer解码器

在完成无砟轨道裂缝特征提取后,通过像素解码器,将提取的特征图生成多尺度特征图,然后采用本文改进的可变形注意力Transformer解码器输出检测结果.原始Mask2Former使用Transformer解码

器的自注意力机制[17]来理解全局上下文信息.然而Transformer自注意力是一种静态机制,当输入数据变化时,会对模型的输出结果产生较大影响[18].为了克服上述原始Mask2Former模型缺点,本文将可变形注意力机制[19]引入Transformer解码器中,取代原始的自注意力机制,来增强多头自注意力的性能,使模型更好地适应无砟轨道裂缝结构特征的变化,增强捕获全局上下文信息的能力,改善裂缝识别不准确的问题.改进后,可变形注意力Transformer解码器中可变形注意力机制如图5所示.

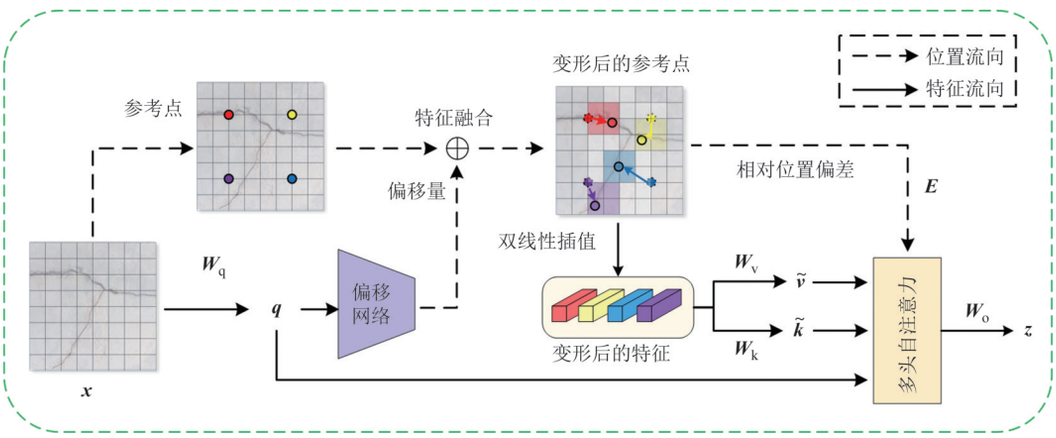


图5　可变形注意力示意图

Fig.5　Diagram of deformable attention

图5中可变形注意力模块,首先通过特征提取网络获得无砟轨道裂缝特征图$\boldsymbol{x}\in\mathbf{R}^{H\times W\times C}$,在特征图上生成由参考点组成的大小为$H_G\times W_G$的均匀网格,并通过输入特征图的降采样率$r$计算得到$H_G=H/r$,$W_G=W/r$.然后根据网格的形状$H_G\times W_G$将参考点坐标归一化.之后,将特征图线性地映射到查询标记$\boldsymbol{q}=\boldsymbol{x}\boldsymbol{W}_q$,并结合每个参考点的偏移量$\Delta\boldsymbol{p}$,得到变形后的特征$\tilde{\boldsymbol{x}}$,如式(5)所示.然后,变形后的特征$\tilde{\boldsymbol{x}}$通过映射后得到变形的key和value,如式(6)所示:

$$\tilde{\boldsymbol{x}}=\phi\left(\boldsymbol{x};\ \boldsymbol{p}+\Delta\boldsymbol{p}\right) \tag{5}$$

$$\boldsymbol{q}=\boldsymbol{x}\boldsymbol{W}_q,\tilde{\boldsymbol{k}}=\tilde{\boldsymbol{x}}\boldsymbol{W}_k,\tilde{\boldsymbol{v}}=\tilde{\boldsymbol{x}}\boldsymbol{W}_v \tag{6}$$

式中:$\tilde{\boldsymbol{k}}$和$\tilde{\boldsymbol{v}}$分别表示变形后的键和值.然后,使用式(7)所示的双线插值$\phi(\cdot)$作为采样函数.

$$\phi\left(\boldsymbol{z};\ \left(p_x,p_y\right)\right)=\sum_{(r_x,r_y)}g\left(p_x,r_x\right)g\left(p_y,r_y\right)z\left[r_y,r_x,:\right] \tag{7}$$

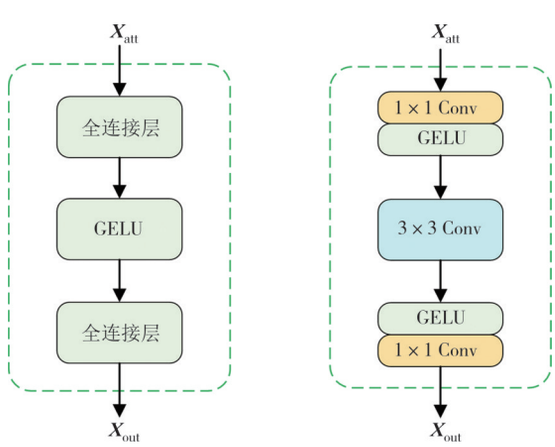
式中:点(r_x,r_y)表示$\boldsymbol{z}\in\mathbf{R}^{H\times W\times C}$上的所有整数像素点的位置;$g(p_x,r_x)$和$g(p_y,r_y)$分别表示在$x$轴方向上和$y$轴方向上的线性插值;$z[r_y,r_x,:]$表示点$(p_x,p_y)$邻域中的四个点的整数像素位置.然后,将$\boldsymbol{q}$、$\tilde{\boldsymbol{k}}$、$\tilde{\boldsymbol{v}}$送入到多头自注意力中,并使用相对位置偏移$\boldsymbol{E}$来增强多头自注意力,注意力头输出为:

$$z^{(m)}=\sigma\left[\boldsymbol{q}^{(m)}\tilde{\boldsymbol{k}}^{(m)\top}/\sqrt{d}+\phi\left(\hat{B};\ \boldsymbol{E}\right)\right]\tilde{\boldsymbol{v}}^{(m)} \tag{8}$$

式中:$\phi\left(\tilde{B};\ \boldsymbol{E}\right)\in\mathbf{R}^{HW\times H_GW_G}$与位置嵌入相对应,每个注意力头的特征被串联在一起,并通过$\boldsymbol{W}_o$进行特征映射,得到输出结果z.

1.4　改进Transformer前馈神经网络

在Transformer模型的解码器部分,裂缝的特征被送入前馈神经网络(feedforward network,FFN)模块中.原始Mask2Former模型FFN结构如图6(a)所示,由两个全连接层和一个GELU激活层组成.然而,原始FFN结构中全连接层会破坏图像的空间结构[20].本文对原始FFN进行改进,如图6(b)所示.



(a)原始FFN前馈模块　　(b)本文改进后FFN

图6　改进前后的FFN对比

Fig.6　Comparison of FFN before and after improvement

改进后FFN前馈模块在两个非线性激活函数GELU层之间插入了一个3×3深度卷积来减少参数量[21]. 然后，使用两个1×1的卷积层替换了原来的全连接层，避免了原始FFN全连接层会破坏图像空间结构的问题. 经过自注意力处理的特征序列$\boldsymbol{X}_{\text{att}}$被转换为二维特征图$\boldsymbol{X}_{\text{in}}$，转换过程见式(9)，然后通过改进的FFN输出，如式(10)所示：

$$\boldsymbol{X}_{\text{in}} = \text{Seq2Img}\left(\boldsymbol{X}_{\text{att}}\right) \tag{9}$$

$$\boldsymbol{X}_{\text{out}} = C^1\left(\text{dwC}^3\left(C^1\left(\text{BN}\left(\boldsymbol{X}_{\text{in}}\right)\right)\right)\right) + \boldsymbol{X}_{\text{in}} \tag{10}$$

式中：$C^1(\cdot)$表示1×1的卷积；$\text{dwC}^3(\cdot)$表示3×3深度卷积；$\text{BN}(\cdot)$表示批归一化处理.

最后，在完成上述迭代处理后，融合像素解码器的输出序列，从而得到最终的裂缝检测结果.

1.5 模型损失函数

所提模型使用的损失函数由两部分组成，分别是掩模损失L_{mask}和分类损失L_{cls}，计算公式为：

$$\text{Loss} = L_{\text{mask}} + \lambda_{\text{cls}} L_{\text{cls}} \tag{11}$$

式中：掩模损失L_{mask}为二元交叉熵损失和dice损失的组合；λ_{cls}是分类损失L_{cls}的权重系数，L_{cls}表示分类损失，分别定义为：

$$L_{\text{mask}} = \lambda_{\text{ce}} L_{\text{ce}} + \lambda_{\text{dice}} L_{\text{dice}} \tag{12}$$

$$L_{\text{cls}} = -\text{In}\left[p_{\sigma(j)}\left(c_j^{\text{gt}}\right)\right] \tag{13}$$

式中：λ_{ce}是二元交叉熵的权重系数；L_{ce}表示二元交叉熵损失，它度量模型对于每个样本的预测与实际标签之间的差异，L_{ce}定义如下：

$$L_{\text{ce}} = -\frac{1}{N} \sum_{i=1}^{N}\left[y_i \cdot \text{In}\left(p_i\right) + \left(1 - y_i\right) \cdot \text{In}\left(1 - p_i\right)\right] \tag{14}$$

式中：N为样本个数；y_i是第i个样本的真实标签；p_i是对第i个类别为1的预测概率.

λ_{dice}为dice损失的权重，L_{dice}是dice损失，用于缓解样本中前景和背景差异，定义如下：

$$L_{\text{dice}} = 1 - D \tag{15}$$

$$D = \frac{2\sum_i^N p_i g_i}{\sum_i^N p_i^2 + \sum_i^N g_i^2} \tag{16}$$

式中：D表示dice系数，取值范围是[0,1]，用来评估两个样本的相似度；p表示模型预测像素的类别；g表示真实像素的类别.

式(13)中的σ表示预测值与真值对应关系，$p_{\sigma(j)}\left(c_j^{\text{gt}}\right)$是算法预测的第$\sigma(j)$个掩膜片段为类别$c_j$的概率，将所有预测片段的分类损失相加，就得到了整个图像的分类损失.

2 实验结果与分析

2.1 实验数据与环境配置

本文实验中使用的数据集由高速铁路现场实际拍摄及网络无砟轨道裂缝图像构成，通过Labelme工具进行数据标注并通过图像数据增强进行扩充，扩充后的数据集共计有5 000张图片，包含横纵向裂缝、交叉裂缝等不同形状的裂缝. 数据集的70%作为训练集，30%作为测试集. 实验硬件配置为Intel(R) Core (TM) i7-10700K CPU @ 3.80 GHz、32.0 GB RAM、NVIDIA GeForce RTX 2060 SUPER. 模型进行训练时，使用AdamW优化器，初始学习率0.000 1、权重衰减为0.05，batch-size为2，epoch设置为120.

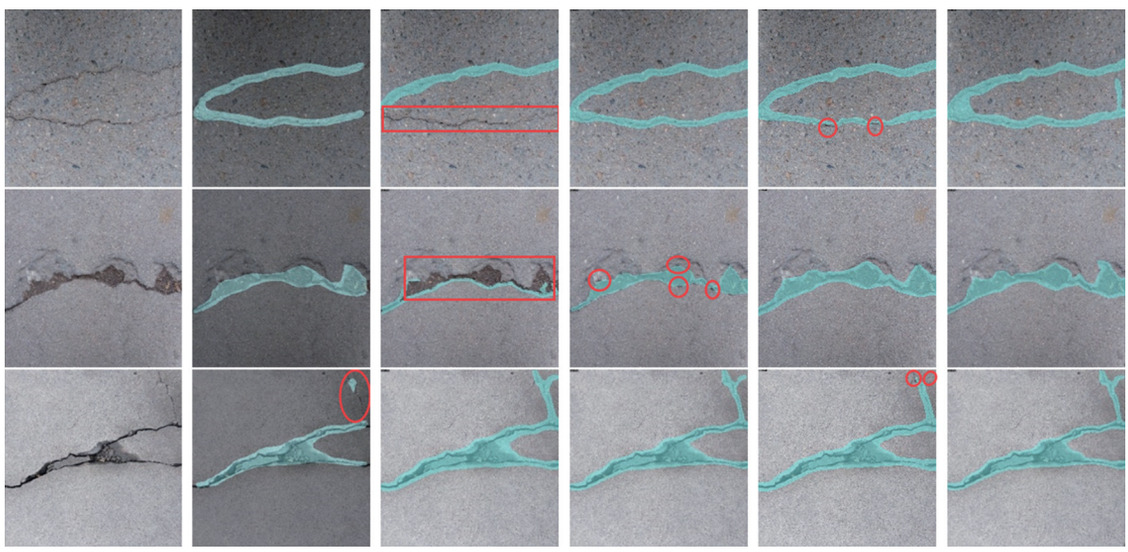
2.2 实验结果

为了验证本文方法的有效性，分别对无砟轨道横向裂缝、纵向裂缝、交叉裂缝进行检测实验，并与Mask2Former、Mask R-CNN[22]、Swin-Transformer[23]和CSTF[24]算法进行对比.

2.2.1 横向裂缝

首先，进行横向裂缝分割实验. 横向裂缝通常由水平载荷、轨道变形和位移引起. 不同方法的检测结果如图7所示. 其中第二列为Mask R-CNN对横向裂缝的检测分割结果，从第二列第3幅图可以看出，Mask R-CNN对于横向裂缝存在检测不连续的问题，其主要原因是Mask R-CNN采用的普通卷积固定的感受野无法有效拟合横向裂缝，从而产生断点. 第三列为Swin-Transformer检测结果，可以看出Swin-Transformer存在漏检的问题，对于细小裂缝不能很好地识别，原因是原始Swin-Transformer内部采用基于局部窗口的Self-Attention，窗口关联性较小，导致感受野变小，从而出现漏检问题. 第四列为Mask2Former检测结果，可以看出该方法较Mask R-CNN和Swin-Transformer裂缝检测结果都有了一定的提升，但仍存在检测不完整的问题，如裂缝边缘部分，这是因为Mask2Former模型中Mask掩膜注意力采用普通卷积操作，像素之间的上下文关联较为薄弱，因而对裂缝部分细节不能实现有效检测. 从第五

列CSTF方法的横向分割结果中可以看出,其对于裂缝的边缘部分不能较好地完成分割.最后一列为本文所提方法检测结果,可以看出,较对比方法,本文方法解决了分割不连续问题,对于横向裂缝的分割更加完整,其原因是所提方法利用动态蛇形卷积的连续性特征提取,克服了裂缝检测不连续的问题.



(a)原图　(b)Mask R-CNN　(c)Swin-Transformer　(d)Mask2Former　(e)CSTF　(f)本文方法

图7　横向裂缝分割结果对比

Fig.7　Comparison of transverse crack segmentation results

2.2.2　纵向裂缝

然后,进行无砟轨道纵向裂缝分割实验,如图8所示.从第二列中可以看出,Mask R-CNN对于裂缝的分割不连贯,如第1幅和第2幅图中,分割结果中均出现了裂缝右侧支路断连的问题.第三列和第四列分别是Swin-Transformer和Mask2Former检测结果,可以看出对于第1幅纵向裂缝,两种方法均存在检测分割不完整的问题,造成这种结果的主要原因是标准的自注意力机制在处理裂缝局部结构变化时缺乏足够的敏感性,不能较好地适应裂缝分支处的形状差异.从第五列CSTF的分割结果中可以发现,其也存在纵向裂缝分支处分割不完整的问题.在最后一列本文方法的检测结果中,相比于其他比较方法,本文方法将可变形注意力机制引入Transformer解码器中来增强多头自注意力的性能,增强了全局上下文信息的能力,裂缝的识别结果更加准确.

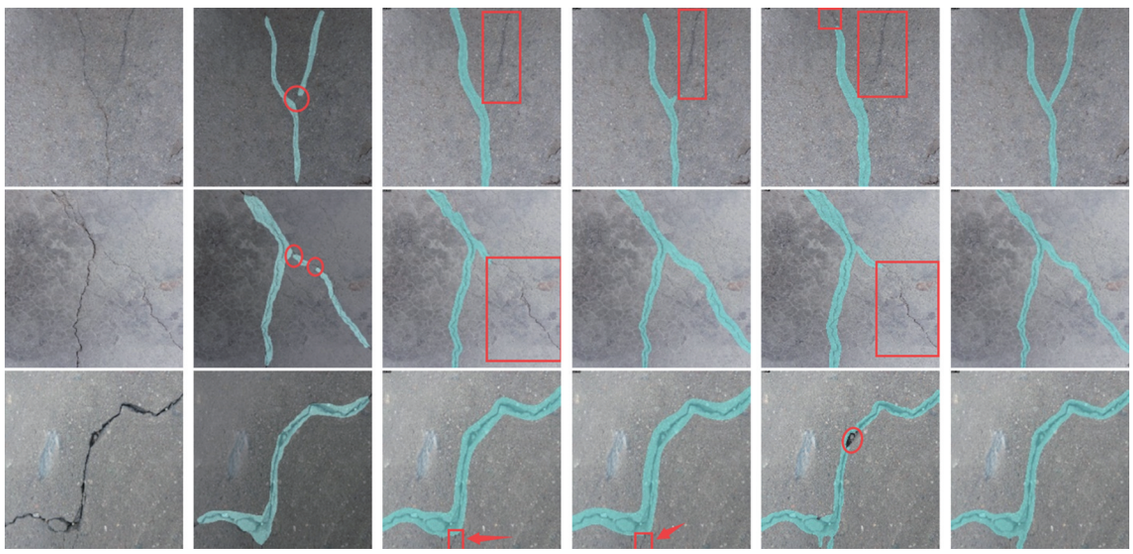
2.2.3　交叉裂缝

在完成横向和纵向裂缝对比实验后,接着进行交叉裂缝分割实验.交叉裂缝主要是由轨道及其连接部件变形或者受到不均匀的力所产生的.不同检测模型的检测结果如图9所示.

图9中,第二列是Mask R-CNN对交叉裂缝的分割结果,可以看出Mask R-CNN对于图像边缘的裂缝分割效果不佳,存在漏检的情况.第三列和第四列分别是Swin-Transformer和Mask2Former检测结果,可以看出对于第3幅图,两种方法对于右侧细小裂缝均无法实现有效检测分割.而对于第五列CSTF的分割结果,可以看出其还存在漏分割的情况,如第3幅图左上裂缝未分割,这是因为CSTF采用池化特征金字塔模块提供先验信息,该操作易造成小目标裂缝特征信息丢失.最后一列为本文检测结果,可以看出对于交叉裂缝,所提方法分割性能更优,边缘裂缝分割效果更好.

2.3　性能定量比较分析

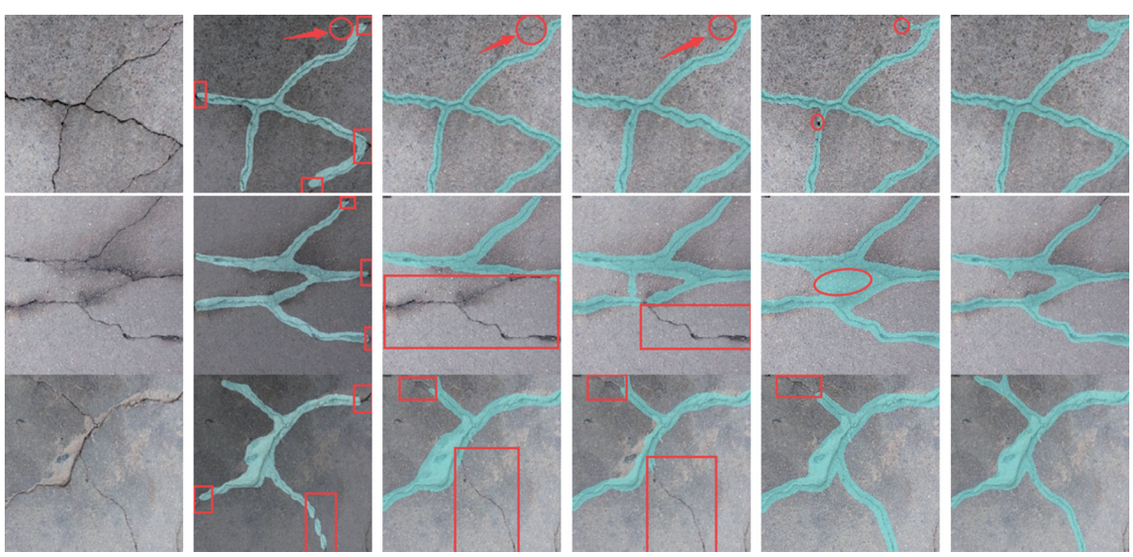
为定量评估本文方法对于裂缝目标的检测性能,采用平均准确率(average precision, AP),平均召回率(average recall, AR)、调和平均数(F_1-score, F_1)进行定量评价.其中,AP越高,表示模型对于目标的误检率越低;AR越高,表示模型能够检测到的目标越多,漏检率越低;调和平均数F_1综合了准确率和召回率,该值越高,表示模型分割结果越好.不同方法对比指标如表1所示.由表1中数据可知,Mask R-CNN模型分割裂缝的平均准确率最低,为



(a)原图 (b)Mask R-CNN (c)Swin-Transformer (d)Mask2Former (e)CSTF (f)本文方法

图 8 纵向裂缝分割结果对比

Fig. 8 Comparison of longitudinal crack segmentation results



(a)原图 (b)Mask R-CNN (c)Swin-Transformer (d)Mask2Former (e)CSTF (f)本文方法

图 9 交叉裂缝分割结果对比

Fig.9 Comparison of cross crack segmentation results

81.15%, CSTF 模型的平均准确率高于 Mask2Former 和 Swin-Transformer. 在五种方法中,本文所提方法的平均准确率最高,较 Mask R-CNN 提高了 12.48 个百分点,较 Swin-Transformer 提高了 8.93 个百分点,较 Mask2Former 提高了 6.34 个百分点,较 CSTF 提高了 5.32 个百分点;从平均召回率上可以看出,本文所提方法的平均召回率为 94.97%,较其他方法分别提高了 14.47 个百分点,8.80 个百分点,4.70 个百分点,4.08 个百分点;F_1 值可以更全面地评价模型,本文所提方法的 F_1 值达到了 94.30%. 因此,本文所提方法对于多形态的无砟轨道裂缝分割性能更好.

表 1 对比实验结果

Tab.1 Comparative experimental results %

检测方法	平均准确率	平均召回率	F_1
Mask R-CNN[22]	81.15	80.50	80.82
Swin-Transformer[23]	84.70	86.17	85.49
Mask2Former[13]	87.29	90.27	88.75
CSTF[24]	88.31	90.89	89.58
本文方法	**93.63**	**94.97**	**94.30**

2.4 消融实验

为了验证所提模块的有效性,采用消融实验进行验证. 基础网络为 Mask2Former,动态蛇形卷积特

征提取网络称为C，可变形注意力Transformer 解码器称为T，改进Transformer前馈神经网络称为F. 将平均准确率AP作为评价指标，在原Mask2Former模型中依次加入对应的改进模块进行实验，实验结果如表2所示.

表2　消融实验比较

Tab.2　Ablation study results

Mask2Former	C	T	F	AP/%
✓				87.29
✓	✓			89.89
✓	✓	✓		92.15
✓	✓	✓	✓	93.63

从表2可以看出,加入了动态蛇形卷积特征提取网络后,模型的平均准确率由原来的87.29%提高到了89.89%,提高了2.6个百分点,这表明动态蛇形卷积特征提取网络增强了模型对于裂缝特征的提取能力.在此基础上嵌入可变形注意力Transformer解码器,使模型适应无砟轨道裂缝特征结构的变化,平均准确率提高了2.26%.最后,加入改进Transformer前馈神经网络,可以更好地学习到裂缝周围的局部位置信息,在实验过程中所提方法的平均准确率达到了93.63%.消融实验验证了所提方法各改进部分的有效性.

3　结 论

1）提出了一种动态蛇形卷积结合可变形注意力Transformer的裂缝检测方法，通过构建动态蛇形卷积主干网络，能够更好地适应裂缝的不规则形变，克服了裂缝检测不连续的问题.

2）设计了可变形注意力Transformer解码器和改进的前馈神经网络FFN模块,加强了模型对无砟轨道裂缝的细节特征感知能力,提高了无砟轨道裂缝的检测精度.

3)实验结果表明,所提方法能够对不同形状的无砟轨道裂缝进行准确分割,平均准确率较Mask R-CNN、Swin-Transformer、Mask2Former、CSTF分别提高了12.48个百分点、8.93个百分点、6.34个百分点、5.32个百分点,裂缝检测更加准确.

参考文献

[1]　宋慧来，赵一馨，刘钰，等. 有砟与无砟轨道系统隧底上拱变形传递规律[J]. 中国铁道科学，2024，45(2)：123-133.
SONG H L，ZHAO Y X，LIU Y，et al. Deformation transmission law of tunnel bottom heave of ballasted and ballastless track system [J]. China Railway Science，2024，45（2）：123-133.（in Chinese）

[2]　YE M X，ZENG Z P，LI P C，et al. Research and prediction of early-age loads in double-block ballastless track structure [J]. Transportation Geotechnics，2024，49:101426.

[3]　CHEN W，LI S Q，WANG W D，et al. Analysis on crack propagation of CRTS Ⅲ slab ballastless track under temperature loads and freeze-thaw deterioration [J]. Theoretical and Applied Fracture Mechanics，2024，129:104206.

[4]　REN Y P，HUANG J S，HONG Z Y，et al. Image-based concrete crack detection in tunnels using deep fully convolutional networks [J]. Construction and Building Materials，2020，234：117367.

[5]　LAU S L H，CHONG E K P，YANG X，et al. Automated pavement crack segmentation using U-net-based convolutional neural network [J]. IEEE Access，2020，8:114892-114899.

[6]　FU H X，MENG D，LI W H，et al. Bridge crack semantic segmentation based on improved Deeplabv3+ [J]. Journal of Marine Science and Engineering，2021，9(6)：671-671.

[7]　TANG Y Z，ZHANG A A，LUO L，et al. Pixel-level pavement crack segmentation with encoder-decoder network [J]. Measurement，2021，184:109914.

[8]　YUAN J W，SONG X L，PU H J，et al. Bridge crack segmentation method based on parallel attention mechanism and multi-scale features fusion [J]. Computers，Materials & Continua，2023，74(3)：6485-6503.

[9]　LI H T，YUE Z，LIU J Y，et al. SCCDNet:a pixel-level crack segmentation network[J]. Applied Sciences，2021，11(11):5074.

[10]　SHI P F，SHAO S，FAN X N，et al. MCL-CrackNet:a concrete crack segmentation network using multilevel contrastive learning [J]. IEEE Transactions on Instrumentation and Measurement，2023，72:5030415.

[11]　BEYENE D A，HUANGRUI，TOLA K D，et al. DCNAM:automatic detection of pixel level fine crack using a densely connected network with attention mechanism [J]. Structures，2024，68:107073.

[12]　XIONG B，HONG R，LIU R，et al. FCT-Net:a dual-encoding-path network fusing atrous spatial pyramid pooling and transformer for pavement crack detection [J]. Engineering Applications of Artificial Intelligence，2024，137:109190.

[13]　CHENG B W，MISRA I，SCHWING A G，et al. Masked-attention mask transformer for universal image segmentation [C]// 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition（CVPR），June 18-24，2022，New Orleans，LA，USA. IEEE，2022:1280-1289.

[14]　HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition [C]//2016 IEEE Conference on Computer

Vision and Pattern Recognition (CVPR), June 27-30, 2016, Las Vegas, NV, USA. IEEE, 2016:770-778.

[15] 谭兆, 王保宪, 秦守鹏, 等. 基于边缘增强感知的混凝土裂缝病害检测方法[J]. 铁道科学与工程学报, 2023, 20(8):3172-3180.
TAN Z, WANG B X, QIN S P, et al. Crack damage detection method based on edge feature reinforcement learning[J]. Journal of Railway Science and Engineering, 2023, 20(8): 3172-3180. (in Chinese)

[16] QI Y L, HE Y T, QI X M, et al. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation [C]//2023 IEEE/CVF International Conference on Computer Vision (ICCV), October 1-6, 2023, Paris, France. IEEE, 2024:6047-6056.

[17] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[EB/OL]. (2017-06-12)[2024-04-15]. https://arxiv.org/abs/1706.03762.

[18] DOSOVITSKIY A, BEYER L, KIM A, et al. An image is worth 16×16 words: transformers for image recognition at scale [EB/OL]. (2020-10-22)[2024-04-15]. https://doi. org/10.48 550/arXiv.2010.11929.

[19] XIA Z F, PAN X R, SONG S J, et al. Vision transformer with deformable attention [C]//2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 18-24, 2022, New Orleans, LA, USA. IEEE, 2022:4784-4793.

[20] LIU Q Y, KAUL C, WANG J, et al. Optimizing vision transformers for medical image segmentation[C]//ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), June 4-10, 2023, Rhodes Island, Greece. IEEE, 2023:1-5.

[21] FU L H, TIAN H Y, ZHAI X P, et al. IncepFormer: efficient inception transformer with pyramid pooling for semantic segmentation[EB/OL]. (2022-12-06)[2024-04-15].https://doi.org/10.48550/arXiv.2212.03035.

[22] HE K M, GKIOXARI G, DOLLAR P, et al. Mask R-CNN[EB/OL]. (2017-03-20) [2024-04-15]. https://doi. org/10.48550/arXiv.1703.06870.

[23] LIU Z, LIN Y T, CAO Y, et al. Swin transformer: hierarchical vision transformer using shifted windows [C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV), October 10-17, 2021, Montreal, QC, Canada. IEEE, 2022: 9992-10002.

[24] YU Z W, CHEN Q Y, SHEN Y G, et al. Robust pavement crack segmentation network based on transformer and dual-branch decoder [J]. Construction and Building Materials, 2024, 453: 139026.