

基于旋转自注意力改进 Mask RCNN 的桥梁裂缝检测方法

陈 永^{1,2}, 安卓奥博¹, 张娇娇¹

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

摘 要: 针对现有桥梁裂缝检测方法对桥梁裂缝旋转特征提取不充分, 检测分割精度低的问题, 提出了一种基于旋转自注意力改进 Mask RCNN 的桥梁裂缝检测方法。首先, 在 Mask R-CNN 实例分割网络的基础上, 采用基于 Transformer 学习的 ViTAE 网络作为主干特征提取网络, 提高对裂缝的检测和分割精度; 然后, 设计旋转可变窗口自注意力机制融入桥梁裂缝检测网络, 提升特征提取网络对裂缝旋转特征的检测能力; 最后, 通过可变形卷积进一步拟合裂缝不规则几何形体, 强化对裂缝特征信息的识别能力。实验结果表明: 本文方法相比于原始 Mask R-CNN 检测分割方法准确率提高了 4.85%, 召回率提高了 13.95%, F_1 -score 可达 91.66%。本文方法能够更加充分地提取裂缝特征, 实现了更加准确的裂缝检测, 在主客观评价方面均优于对比方法。

关键词: 桥梁工程; 混凝土桥梁裂缝; 裂缝病害检测; 旋转可变窗口自注意力; Transformer 学习
中图分类号: U446.3 **文献标志码:** A **文章编号:** 1671-5497(2025)11-3660-13

DOI: 10.13229/j.cnki.jdxbgxb.20240122

Bridge crack detection method based on rotation self-attention improved Mask RCNN

CHEN Yong^{1,2}, ANZHUO Ao-bo¹, ZHANG Jiao-jiao¹

(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: Aiming at the problem that the existing bridge crack detection methods do not fully extract the rotation feature of bridge cracks and have low detection and segmentation accuracy, a bridge crack detection method based on improved Mask RCNN with rotation self-attention was proposed. Firstly, on the basis of the Mask R-CNN instance segmentation network, the ViTAE network based on Transformer learning is used as the backbone feature extraction network to improve the detection and segmentation accuracy of cracks. Then, a rotating variable window self-attention mechanism was designed to integrate into the bridge crack detection network to improve the detection ability of the feature extraction network for

收稿日期: 2024-01-30.

基金项目: 国家自然科学基金项目(62462043, 61963023); 兰州交通大学重点研发项目(ZDYF2304).

作者简介: 陈永(1979-), 男, 教授, 博士. 研究方向: 缺陷检测, 智能信息处理. E-mail: edukeylab@126.com

crack rotation features. Finally, the deformable convolution was used to further fit the irregular geometry of cracks to strengthen the recognition ability of crack feature information. Experimental results show that compared with the original Mask R-CNN detection and segmentation method, the accuracy of the proposed method is improved by 4.85%, the recall rate is increased by 13.95%, and the F_1 -score can reach 91.66%. The proposed method can extract crack features more fully, achieve more accurate crack detection, and is superior to the comparison methods in subjective and objective evaluation.

Key words: bridge engineering; concrete bridge crack; crack detection of diseases; rotational variable window attention; transformer learning

0 引言

桥梁是道路网络中至关重要的国家核心基础设施之一,其安全健康运营是路网畅通的重要保障^[1]。然而,由于桥梁混凝土材料的劣化、恒活负载、基础不均匀沉降和长期疲劳服役,易引起桥梁裂缝病害。桥梁混凝土裂缝也会导致内部钢筋腐蚀,钢筋腐蚀后,其体积增大,从而胀裂砼保护层,诱发裂缝开裂并降低结构的耐久性,严重威胁桥梁的安全性,因此,开展桥梁裂缝检测对于保障道路安全具有重要意义^[2]。

桥梁裂缝检测方法中,人工裂缝识别方法难以满足目前需求,传统检测存在任务量大、服务效率低、人为检测主观性强等缺点,对混凝土剪切裂缝等严重病害裂缝易造成遗漏和错误。目前基于计算机视觉的裂缝检测由于高精度、快速、自动化分析等优点,被国内外学者广泛研究。传统图像裂缝检测方法一般采用边缘检测、数学形态学、阈值分割等方法获取裂缝特征,通过边缘检测算子检测出裂缝边缘位置,然后利用设定阈值或腐蚀、膨胀等数字形态学运算来完成裂缝的分割。Zoubir等^[3]针对桥梁裂缝在采集和拍摄过程中出现的噪声问题,采用梯度方向直方图的图像处理方法提取裂缝特征,同时对横向和纵向裂缝的断点连接算法进行了改进;Vivekananthan等^[4]利用裂缝区域和背景区域的灰度差,通过阈值分割进行裂缝的提取,但分割阈值的选择存在主观性强的问题,其分割精度较低;卢印举等^[5]提出融合多尺度特征的马尔科夫随机场裂缝分割算法,增强了对不同尺度裂缝几何结构的特征提取能力,但该方法抗噪声能力较弱,易出现误分割的问题;Avendaño等^[6]提出一种基于中值绝对偏差的自适应阈值裂缝检测方法,根据像素的灰度自适应调整阈值完成裂缝边缘检测,但阈值易受背景影响,裂

缝检测性能受限。综上,传统图像裂缝检测方法普遍存在噪声干扰多、特征提取不准确、裂缝细节少、抗干扰能力差的问题。

当前基于深度学习的目标检测方法被广泛应用于各种工业领域,如何通过深度学习方法精准检测桥梁病害也成为热点研究问题。深度学习技术在检测裂缝时,以深度神经网络自动提取和学习裂缝特征为主,借助特征提取网络实现对裂缝结构的端到端学习,通过设定损失函数优化模型参数,从而完成对裂缝的检测。进行裂缝检测模型性能评价时,通常以平均准确度、平均召回率、调和平均数等作为主要指标。Wan等^[7]基于单发多框检测器和滑动窗口技术进行裂缝识别,但该方法在滑动窗口后易造成全局语义特征丢失;朱苏雅等^[8]针对桥梁裂缝检测细节丢失的问题,提出了基于U-Net卷积神经网络的像素级裂缝检测方法,将深层网络和浅层网络结合,提升裂缝检测的准确度,但该方法易受背景噪声的影响;谭国金等^[9]针对裂缝大小尺度不一所带来的多样性问题进行改进,提出了基于DeepLab的改进方法,提高对裂缝边缘的检测性能,但该方法对多方向延伸的裂缝检测性能较差;Xiao等^[10]改进了YOLOv5模型的主干网络,将双向特征金字塔和注意力机制融入模型进行裂缝检测,但基于锚点框的检测模型对细小裂缝的识别精度存在一定的影响;Deng等^[11]将可变形卷积模块和Faster RCNN进行融合,完成对混凝土裂缝的检测,但可变形卷积只能强化局部裂缝信息,难以捕捉全局裂缝信息,其检测结果存在裂缝检测间断的问题;Liu等^[12]提出一种基于Mask2former的裂缝检测模型,利用跳跃连接增强机制来提升模型的全局特征提取能力,但由于解码器输出的查询序列存在不连续的问题,裂缝检测结果易出现粘连的问题。

综上所述,现有桥梁裂缝检测方法存在对桥

梁裂缝旋转特征提取不充分,检测分割精度低的问题,因此,本文提出了一种基于旋转自注意力改进的Mask RCNN桥梁裂缝检测方法。为了提高桥梁裂缝的检测性能,本文以原始Mask R-CNN实例分割模型^[13],作为基线模型,以ViTAE(Vision transformer advanced by exploring)模型^[14]为特征提取网络,并对其进行改进后提出了本文模型。其中,Mask R-CNN是一种实例分割的深度学习模型,由主干网络、区域候选网络、分类分支、边框回归分支和掩码分支构成,其广泛应用于实例分割场景中。本文主要工作有:①在Mask R-CNN的基础上,构建基于ViTAE的特征提取网络,加强对裂缝特征的提取;②设计融合旋转可变窗口自注意力机制的网络框架,提升特征提取对裂缝旋转特征的检测能力;③通过可变形卷积进一步增强拟合裂缝能力,使可变形卷积对裂缝的细长型特征可以更加准确地进行分割;④通过桥梁裂缝数据集进行消融实验和对比实验分析,证明本文方法对桥梁裂缝分割的有效性。

1 桥梁裂缝图像多方向性特征分析

对于无人机或者爬壁机器人拍摄的图像,裂缝图像特征最重要的信息是裂缝边缘信息。混凝土桥梁裂缝的一个特征是方向延伸,裂缝延伸方向的图像像素灰度值变化平缓,其他方向变化剧烈,整体连续,存在局部断裂。

无人机在桥梁裂缝图像采集过程中,由于其稳定性差,导致不同角度采集的裂缝图像具有多方向性。此外,桥梁裂缝本身类型较多,不仅具有水平裂缝、垂直裂缝等,而且存在着诸如旋转、X形、网状等裂缝类型。对于最常见的水平裂缝图像,如果将裂缝平移后,采用深度学习中的传统卷积对桥梁裂缝进行检测,由于传统卷积具有平移等变性,所以其依然能提取到桥梁的裂缝特征信息,如图1(a)所示。由于普通卷积采用局部卷积操作滤波的模式,所以独立于它们的空间位置。当输入的桥梁裂缝图像水平移动后,采用普通卷积仍能预测其移动,该特性被称为普通卷积操作的平移等变性^[15]。

但是,对于旋转类型的裂缝,其裂缝特征呈现出在不同角度延伸的特点,如果采用传统卷积进行特征提取,则无法进行有效的裂缝特征提取。这是因为传统卷积采用滑动滤波的形式进行点卷积计

算得到,其不具备旋转等变性。对于旋转后的裂缝,传统卷积无法有效提取桥梁旋转裂缝特征,如图1(b)所示。因而采用传统卷积并不能保证输出的实际轨迹与理论轨迹一致,无法有效提取到多方向裂缝特征信息,导致裂缝检测精度下降。

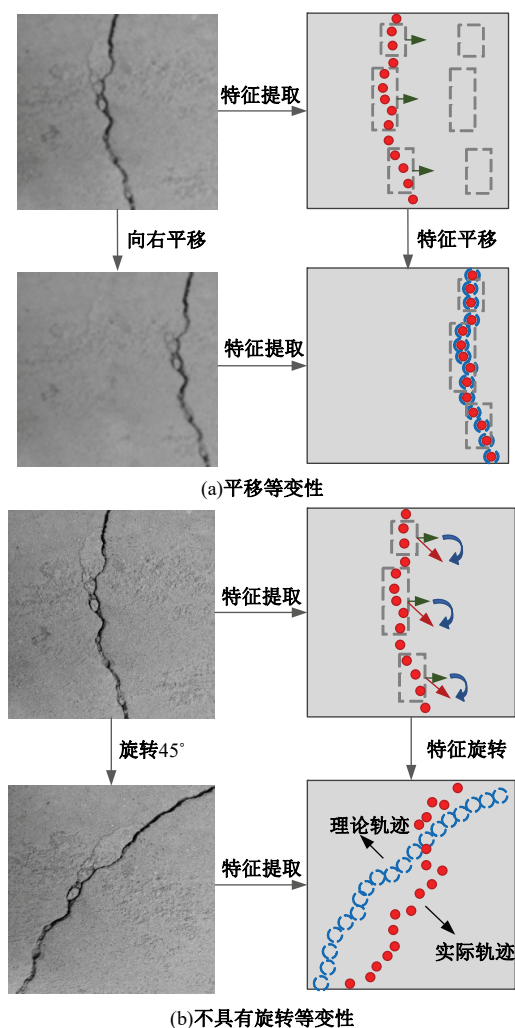


图1 裂缝旋转示意图

Fig. 1 Fracture rotation diagram

2 本文方法

2.1 网络整体结构

针对混凝土桥梁裂缝特征在现有的目标分割网络中特征提取不充分,旋转特征分割效果差,检测精度低的问题,考虑裂缝多方向的特点,在桥梁裂缝分割过程中,原始Mask R-CNN网络模型采用传统卷积进行特征提取,无法有效获取混凝土桥梁裂缝特征信息,因而本文在Mask R-CNN的基础上对其进行改进,将旋转等变性引入ViTAE特征提取网络中,通过学习多类型裂缝的不同角度延伸方向特征,以适应不同混凝土桥梁裂缝检

测,实现对桥梁复杂裂缝的检测和分割。

本文模型网络整体结构如图 2 所示,该模型工作时,首先将双分支 ViTAE 网络作为模型的特征提取网络。其次,利用旋转可变窗口自注意力机制充分提取裂缝旋转特征。然后,融入可变形

卷积增强对裂缝特征的拟合能力。通过区域候选网络对感兴趣区域进行对齐操作,并对裂缝候选框进行预选。最后,由全连接层及其对应的边框回归分支、分类分支和掩膜分支,对桥梁裂缝进行像素级分割和检测输出。

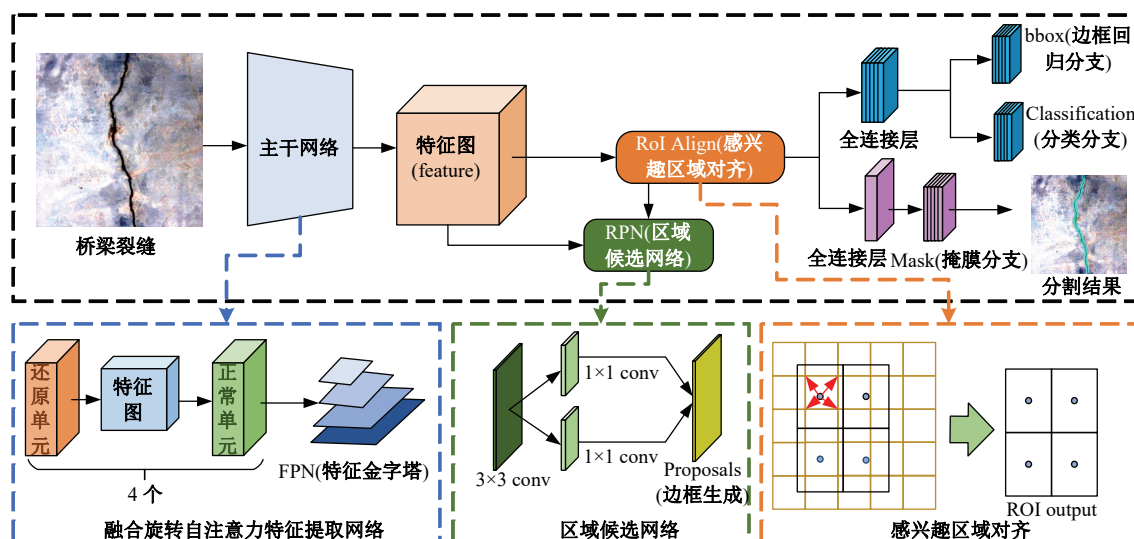


图 2 网络整体结构

Fig. 2 Model frame work

2.2 融合旋转自注意力的特征提取网络

原始 Mask R-CNN 在主干特征提取网络中采用卷积神经网络模块进行特征提取,但传统卷积神经网络关注的是图像的特征细节,擅长提取浅层边缘和角点等局部特征,对于类型多样的混凝土桥梁裂缝无法有效提取其多方向特征。为克服上述桥梁裂缝特征提取的不足,在本文网络模型中,主干网络采用 ViTAE 特征提取网络模型。ViTAE 模型是一种基于 Transformer 的改进模型,其具有较强的长程捕获上下文信息的能力,在进行裂缝分割时,首先将待检测桥梁裂缝图像进行位置编码,划分为多个查询序列(token)。

然后,通过编码器中本文提出的旋转可变窗口自注意力机制(Rotated varied size attention, RVSA)^[16],对输入序列的桥梁裂缝几何位置进行加权运算,并建立几何位置之间的旋转关联。最后,通过多个编码器构成的 stage 块,根据输出的 token 序列得到裂缝图像特征图。

ViTAE 桥梁裂缝特征提取模块由一系列还原单元模块(Reduction cell, RC)和正常单元模块(Normal cell, NC)构成的 stage 块组成,层次结构如图 3 所示。该模块中的还原单元 RC 模块分别对裂缝特征进行 4、2、2 和 2 倍 4 个阶段的下

采样,来增强桥梁裂缝多尺度上下文特征提取能力。其中,正常单元 NC 模块主要控制模型的长程语义提取能力,促进对裂缝特征的全局感知。此外,相比传统卷积神经网络,ViTAE 特征提取网络的自注意力机制提高了模型对裂缝像素的关注程度,同时还提升了对整体裂缝全局建模的能力,对于模型中多阶段的设计,采用逐级别降低分辨率方式,相比传统自注意力机制,其计算程度进一步降低。此外,由于传统卷积神经网络的填充操作,导致图像边缘裂缝存在漏检现象,而 ViTAE 中的注意力机制可以弥补传统卷积中填充操作的缺点。

由图 3 可知,ViTAE 桥梁裂缝特征提取模块首先堆叠多个还原单元 RC 模块和正常单元 NC 模块,然后,还原单元 RC 模块将特征图缩小为 4 倍后用 token 序列转换为图像(Sequence to image, Seq2Img)运算进行展平,随后输入正常单元 NC 模块。正常单元 NC 模块不改变图像分辨率,模型通过 RC 模块和 NC 模块的堆叠增加模型容量,两种模块详细结构如图 4 所示。最后,将多个阶段提取得到的不同尺度的特征图进行多尺度特征融合,进一步提升模型的尺度等变性能力。

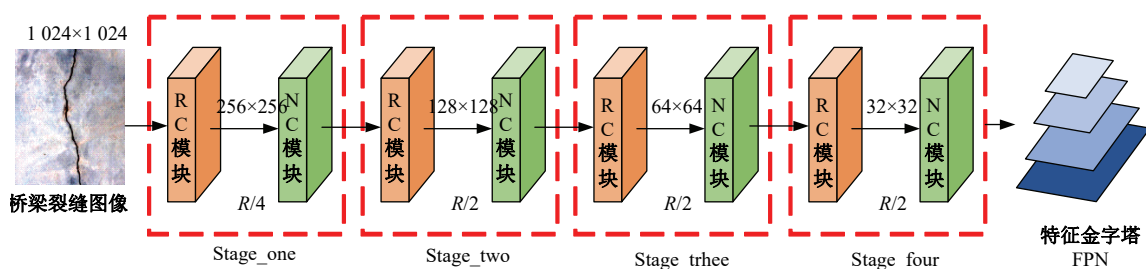
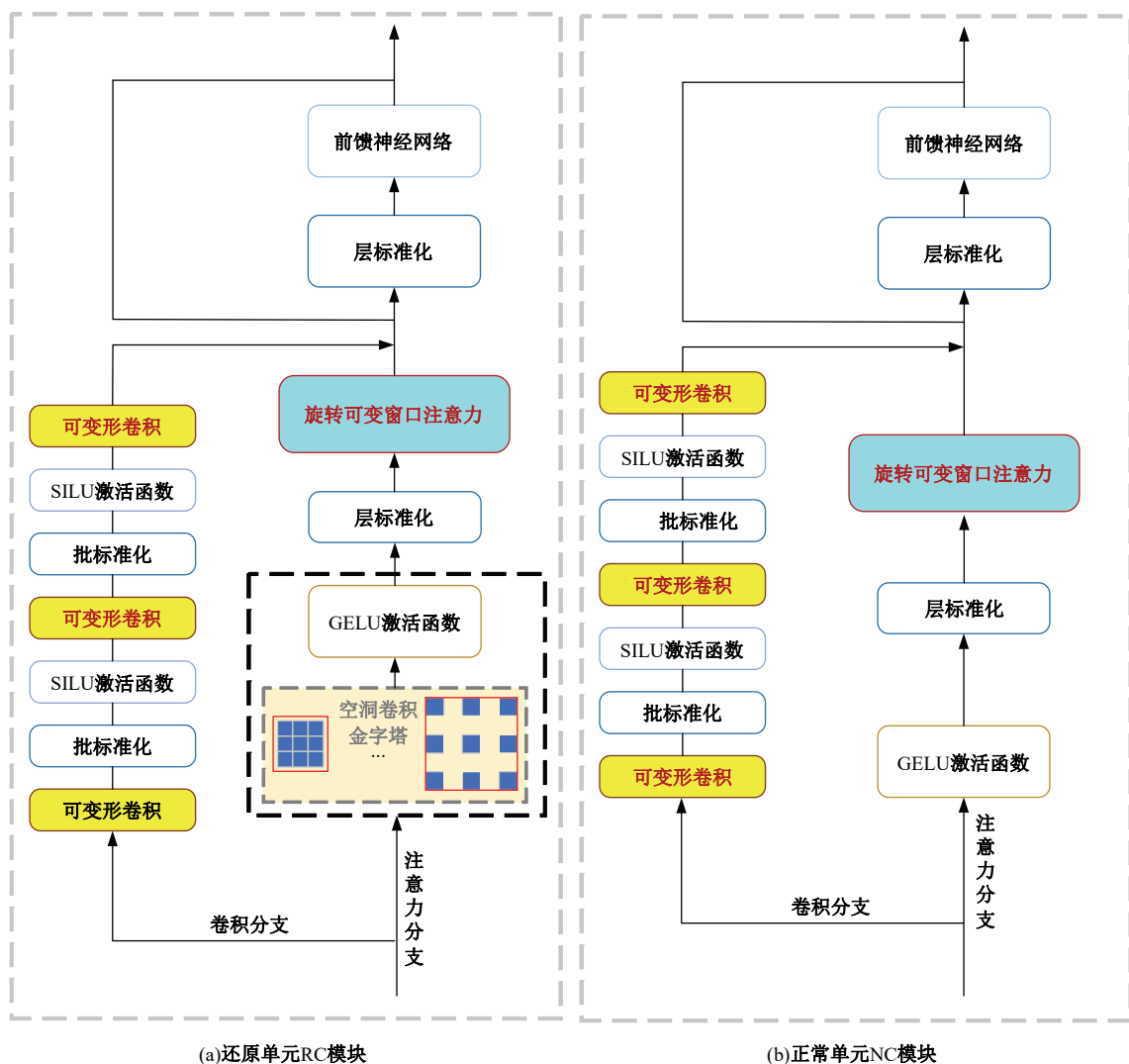


图3 ViTAE桥梁裂缝提取模块结构图

Fig. 3 ViTAE bridge crack extraction module structure diagram



(a)还原单元RC模块

(b)正常单元NC模块

图4 RC模块和NC模块

Fig. 4 Reduction cell and Normal cell

在图4中,裂缝特征通过RC模块和NC模块进行旋转特征提取,裂缝特征输入后,分别进入注意力分支和卷积分支。注意力分支主要负责建立全局特征,首先通过空洞卷积金字塔进行多尺度特征提取,加强对尺度等变性的能力。然后,输入将混凝土桥梁裂缝图像输入全局建模分支中,由空洞卷积金字塔(Pyramid reduction module,

PRM)提取多尺度信息,特征提取公式如下:

$$\text{PRM}(f_i) = \text{Cat}([\text{Conv}(f_i, s_i, r_i) | s_i \in S, r_i \in R]) \quad (1)$$

式中: f_i 为第*i*个RC模块输入的特征; s_i 为空洞卷积的扩张速率; r_i 为缩减比率;Conv为空洞卷积金字塔中的卷积层;Cat为拼接操作。

接着,送入可变窗口自注意力模块,注意力计

算方式如下:

$$X = \text{RVSA}(\text{token}) \quad (2)$$

式中: token 为全局建模分支输入;RVSA 为旋转可变窗口自注意力; X 为变换后输出的目标窗口。

图 4 中,左侧卷积分支主要负责建立局部特征,通过多个卷积层的堆叠,对裂缝局部进行特征提取。最后,将卷积分支和注意力分支进行融合,使得模块可以感知局部和全局感受野,之后由前馈神经网络重新生成特征图,并传递到下一个 NC 或 RC 模块中,之后采用序列转换图像 Seq2Img 函数进行运算:

$$\text{feature}(X) = \text{Seq2Img}(\text{token}) \quad (3)$$

正常单元模块相比还原单元 RC 模块缺少空洞卷积金字塔结构,因为输入正常单元 NC 模块的特征图较小,所以可直接进行特征融合后输出。

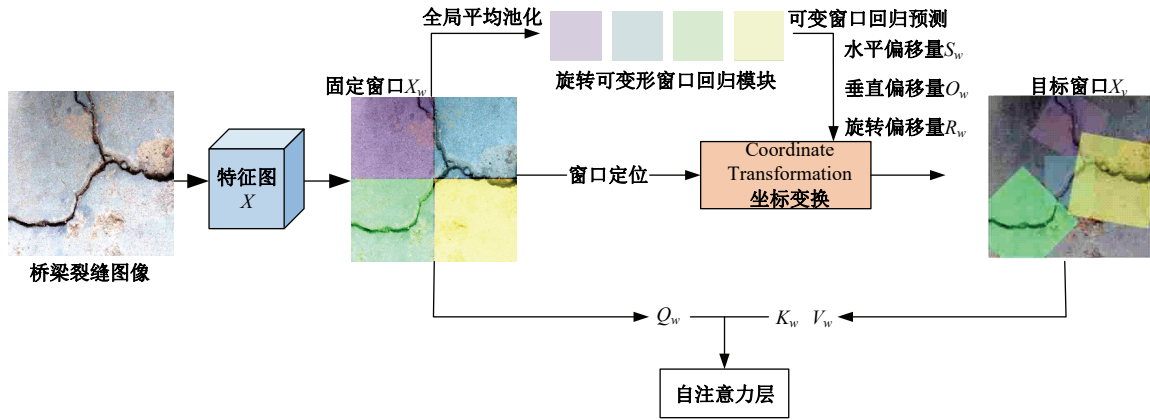


图 5 旋转可变窗口自注意力示意图

Fig. 5 Rotated varied-size attention schematic diagram

在图 5 旋转可变窗口自注意力模型中,为了增强对混凝土桥梁旋转裂纹的检测能力,本文将原始 Transformer 划分的固定窗口设计为适用裂缝的旋转窗口形式。通过旋转窗口来学习不同方向的裂缝特征,以实现对不同角度裂缝的检测。

对于输入桥梁裂缝特征 X ,首先将其按照固定窗口划分的方式得到查询 Q_w :

$$Q_w = \text{Linear}(X_w) \quad (4)$$

式中: X_w 为固定窗口;Linear 为线性操作。

然后,通过对每个窗口 X_w 的大小和位置预测各种偏移量,利用全局平均池化得到不同窗口对应的特征,并加入 LeakyReLU 激活层,进一步使用 1×1 的卷积层后,得到水平偏移量 S_w 、垂直偏移量 O_w 和旋转偏移量 R_w ,对应公式如下:

$$S_w, O_w, R_w = \text{Linear}(\text{LeakyReLU}(\text{GAP}(X_w))) \quad (5)$$

2.3 旋转可变窗口自注意力机制

为克服传统卷积运算在裂缝检测时不具备旋转可变性的缺点,本文提出旋转可变窗口自注意力机制对 ViTAE 桥梁裂缝提取模块进行改进。原始 ViTAE 网络中虽然使用多头自注意力 (Multi-head self-attention, MHSA)^[17] 进行特征提取,但无人机拍摄导致大量角度变化的裂缝图像在使用多头自注意力机制时无法有效聚焦^[18]。本文在 ViTAE 网络的基础上,通过设计旋转可变窗口自注意力 RVSA 替换原始 ViTAE 网络中的多头自注意力,以促进裂缝旋转信息在窗口的交换,提升裂缝分割对角度的敏感性以增强对桥梁裂缝旋转特征的检测能力。可变窗口根据混凝土桥梁裂缝学习的内容自适应改变窗口的大小和位置,计算过程如图 5 所示。

式中:LeakyReLU 为激活函数;Linear 为线性变换;GAP 为全局平均池化。

然后根据固定窗口并结合水平偏移量、垂直偏移量和旋转偏移量进行坐标旋转变换,计算公式如下:

$$X_v = X_w + S_w + O_w + R_w \quad (6)$$

$$K_w, V_w = \text{Reshape} \cdot \text{Linear}(X_v) \quad (7)$$

式中:Reshape 为矩阵变换;Linear 为线性变换; K_w 为键; V_w 为值。

通过式(6)(7)进行旋转变换后,将得到的旋转后的特征 K_w 和 V_w 以及固定窗口 Q_w 查询序列送入自注意力层进行 Attention 计算。

为了直观对比该过程,进行可视化后的效果如图 6 所示。从图 6(a)可以看出,原始 Transformer 采用固定窗口,其无法适应方向多变的裂缝特征,而采用本文方法进行窗口旋转后,如图 6(b)所示,

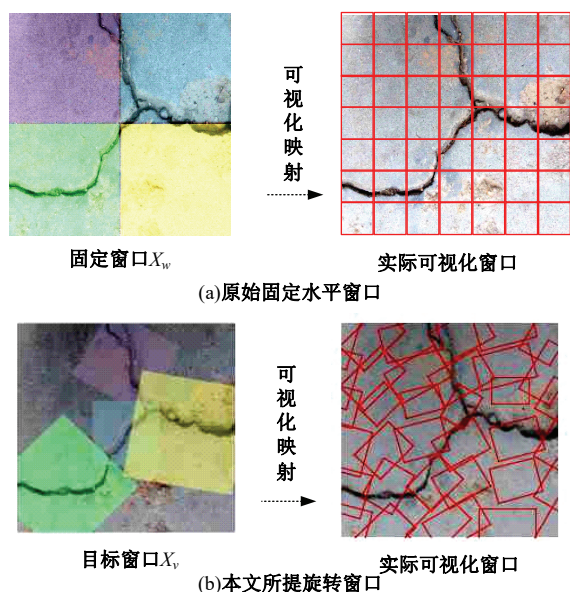


图 6 目标窗口的可视化映射

Fig. 6 Visualization mapping of target windows

旋转窗口能更符合裂缝方向多样性的特点,可以较好地分割区域进行选择。

为验证本文旋转可变窗口自注意力模块对混凝土桥梁裂缝分割的有效性,在不同角度的裂缝图像上进行分割检测实验,热力图检测结果如图 7 所示。在图 7 中,热力图的高亮暖色区域表示特征提取的区域,冷色蓝色区域表示对关注区域的贡献较小。从图 7 可以看出,原始 Transformer 方

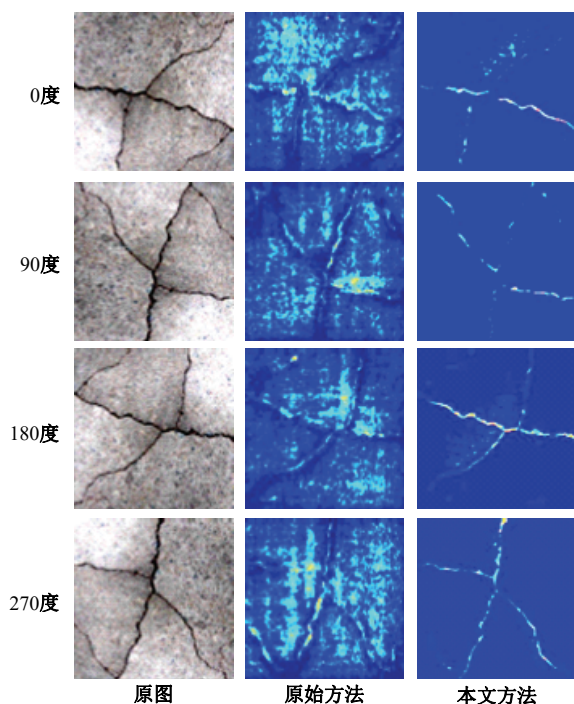


图 7 旋转特征提取性能比较

Fig. 7 Comparison of different rotation characteristics

法在进行特征提取时,无法有效聚焦裂缝特征,裂缝周围桥梁背景对其干扰较大,存在较多的背景杂波干扰,难以有效聚焦裂缝特征。而采用本文方法后,可以发现能够聚焦裂纹主体特征,有效抑制背景干扰。针对不同角度、大小、位置和尺度的旋转裂缝,本文方法取得了更好的聚焦性能。

2.4 可变形卷积

此外,在原始 ViTAE 特征提取模块中,采用传统卷积进行特征提取,然而传统卷积核采用固定几何形状,很大程度上限制了卷积操作的特征提取能力,难以有效区分混凝土桥梁背景与裂缝,因此,本文进一步设计了可变形卷积(Deformable convolution network, DCN)^[19],增强对裂缝特征的拟合能力,将 ViTAE 旁路卷积中的传统卷积层设计为可变形卷积。

可变形卷积通过为感受野添加偏移量 Δp_n ,从而捕捉裂缝的形变信息,可变形卷积通过计算特征偏移权重获得偏移量,随后由接收偏移量的卷积核进行偏移,解决了传统卷积层感受野受限问题,可变形卷积由以下映射公式计算:

$$y(p_0) = \sum_{p_n \in R} w(p_n) \cdot X(p_0 + p_n + \Delta p_n) \quad (8)$$

在检测混凝土桥梁裂缝时,可变形卷积能够通过动态调整采样点,动态自适应裂缝的形状,可以看出可变形卷积采样点更适应裂缝本身的真实形状,如图 8 所示。

为了验证本文引入可变形卷积对裂缝检测的效果,下面对加入可变形卷积后的特征提取网络进行可视化热力图比较实验,如图 9 所示。从比较实验可以看出,未加入可变形卷积之前,在进行特征提取时存在大量的背景杂波干扰,未能有效聚焦于裂缝本身特征,而加入本文可变形卷积后

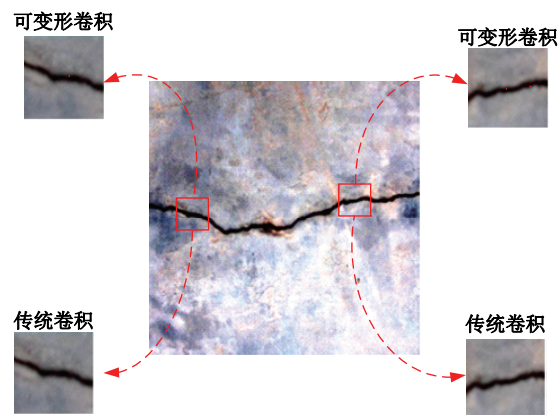


图 8 可变形卷积原理示意图

Fig. 8 Schematic diagram of deformable convolution

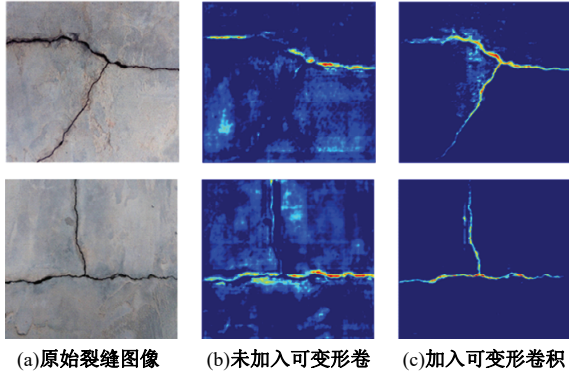


图 9 热力图比较实验

Fig. 9 Thermal map comparison experiment

在进行特征提取时,其能够覆盖更多的混凝土桥梁裂缝部分,能有效对裂缝特征进行聚焦。比较结果说明可变形卷积对裂缝的变形聚焦能进一步关注裂缝边缘部分,对裂缝的背景信息进行了有效抑制。

2.5 区域候选网络

在主干网络完成对混凝土桥梁裂缝的提取后,下一步采用区域候选网络(Region proposal network, RPN)对裂缝进行标记框选。将待检测桥梁裂缝候选框进行二值分类并过滤部分候选框,得到适合的候选框。然后遍历剩余候选框和最优候选框的重叠度,剔除重叠度超过阈值的候选框。最后,RPN通过双线性插值对筛选后的每个候选框计算固定的坐标像素值,从而实现对裂缝的候选框和像素点输出,RPN结构如图10所示。

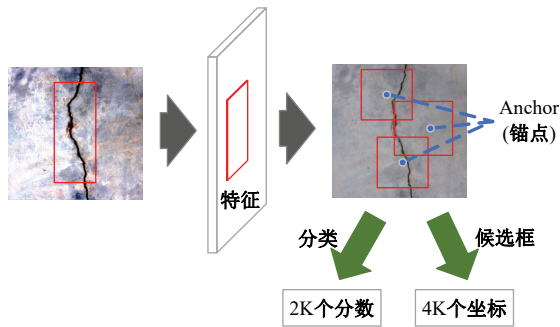


图 10 区域候选网络示意图

Fig. 10 RPN structure diagram

2.6 损失函数

本文损失函数设计与Mask R-CNN类似,由分类损失 L_{cls} 、边界框回归损失 L_{box} 、掩膜损失 L_{mask} 3部分构成,其计算公式如下:

$$L = L_{cls} + L_{box} + L_{mask} \quad (9)$$

分类损失 L_{cls} 和边界框回归损失 L_{box} 的计算

公式如下:

$$L_{cls} = \frac{1}{N_{cls}} \sum_i L_{cls}(p_i, p_i^*) \quad (10)$$

$$L_{box} = \lambda \frac{1}{N_{box}} \sum_i p_i^* L_{box}(t_i, t_i^*) \quad (11)$$

式中: i 为锚点索引; N_{cls} 和 N_{box} 为分类和回归数量; p_i 为候选框预测概率,用于候选框质量评估; p_i^* 表示候选框预测为前景的概率,用于判断候选框是否包含目标并给出类别标签; t_i 为候选框坐标偏移量; t_i^* 为目标候选框真实坐标偏移量; λ 为 N_{cls} 和 N_{box} 的平衡参数; L_{mask} 为掩膜损失函数,主要用于指导裂缝分割像素输出,其计算式如下:

$$L_{mask} = -\frac{1}{m^2} \sum_{k=1}^K \sum_{i=1}^{m^2} \log_2 [p_{ki}^* p_{ki} + (1 - p_{ki}^*)(1 - p_{ki})] \quad (12)$$

式中: m 为掩膜的长度和宽度; k 为检测目标总数; p_{ki}^* 为第 k 个目标在真实掩码图像中第 i 个像素的值; p_{ki} 为第 k 个目标在预测掩码图像中第 i 个像素的值。

3 实验验证

3.1 性能评价指标

为评估本文对混凝土桥梁裂缝的检测性能的定量评价,采用平均准确率(mean Average precision, mAP)、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,平均准确率AP越高表示模型误检程度越低;平均召回率AR越高表示模型漏检率越低;调和平均数 F_1 综合了准确率和召回率,该值越高表示模型整体检测性能越好,评价指标对应公式如下:

$$AP = \frac{TP}{TP + FP} \quad (13)$$

$$AR = \frac{TP}{TP + FN} \quad (14)$$

$$F_1 = \frac{2 \times P \times R}{P + R} \quad (15)$$

式中: TP 表示预测为正实际也为正的桥梁裂缝样本数量; FP 表示预测为正但应该预测为错的桥梁裂缝样本数量; FN 表示预测为错但应该预测正确的数量。

3.2 桥梁裂缝数据集及训练环境

为获取足够的桥梁裂缝数据集,本文在 Bridge_Crack_Image_Data 桥梁裂缝数据集^[20]的基础上,进一步结合网络桥梁裂缝图像,以及现场

对不同混凝土桥梁拍摄采集的图像,通过 Labelme 工具制作了扩充后的桥梁裂缝数据集,共计 5 000 张图像。其中,70% 作为训练集,30% 作为验证集和测试集。此外,实验环境的硬件配置为 Intel(R) Core i9-12900H CPU@ 2.50 GHz、32.0 GB RAM、NVIDIA GeForce GTX 3070Ti。模型训练时,设置学习率为 0.002、batch-size 为 2、epoch 迭代次数为 120。

3.3 消融实验

为验证本文模型中的改进模块对原网络模型的有效性,依次在基线网络模型的基础上进行了消融实验,实验结果如表 1 所示,可以看出,骨干网络采用 ViTAE 特征提取网络后, $mAP_{0.5}$ 和 $mAP_{0.75}$ 分别提高了 0.75% 和 7.99%,表明添加 ViTAE 骨干网络有助于提升桥梁裂缝的分割精度。将旋转可变窗口自注意力优化 ViTAE 模型后, $mAP_{0.5}$ 和 $mAP_{0.75}$ 分别提高了 3.7% 和 4.5%,表明模型通过旋转自注意力机制提升了对裂缝特征提取的关注度。最后,添加可变形卷积模块,其 $mAP_{0.5}$ 和 $mAP_{0.75}$ 再次提高了 0.4% 和 2.8%,表明可变形卷积相比传统卷积对裂缝能更有效地实现特征拟合。

表 1 消融实验结果

Table 1 Results of ablation experiments

基线模型	ViTAE	RVSA	DCN	$mAP_{0.5}/\%$	$mAP_{0.75}/\%$
				84.15	57.31
Mask-RCNN	✓			84.90	65.30
	✓	✓		88.60	69.80
	✓	✓	✓	89.00	72.60

注:“✓”表示模型添加模块; $mAP_{0.5}$ 表示当 IoU 阈值为 0.5 时,模型对裂缝检测精度的平均值; $mAP_{0.75}$ 表示当 IoU 阈值为 0.75 时,模型对裂缝检测精度的平均值。

3.4 对比实验结果

桥梁裂缝可以分为多种类型,包括垂直裂缝、水平裂缝、交叉 X 形、网状裂缝等。为了验证本文方法的有效性,针对垂直裂缝、水平裂缝、交叉 X 形、网状裂缝分别进行裂缝分割实验,并且将本文方法与 U-Net、Mask R-CNN、Swin-Transformer^[21]、Intern-Image^[22] 和 Mask2former^[23] 进行裂缝分割对比实验。

3.5 评价指标对比分析

不同方法在裂缝数据集上的分割对比实验结果见表 2,可以看出,U-Net 检测模型的检测准确率最低, Swin-Transformer 准确率高 Mask R-

表 2 对比实验结果

Table 2 Comparison of experimental results

检测方法	AP/%	AR/%	F_1 -score/%
U-Net	71.00	79.30	74.92
Mask R-CNN	84.15	80.54	82.30
Swin-Transformer	86.20	90.34	88.83
Intern-Image	87.40	91.23	89.27
Mask2former	87.29	90.27	88.75
本文网络模型	89.00	94.49	91.66

CNN 网络。在 4 种方法中,本文模型准确度最高,较 Mask R-CNN 提高了 4.85%,较 Swin-Transformer 提高了 2.80%,较 InternImage 和 Mask2former 分别提高 1.6% 和 4.22%。在召回率方面,较 Mask R-CNN 提高了 13.95%,较 Swin-Transformer 提高了 4.15%,较 InternImage 和 Mask2former 分别提高 3.26% 和 4.22%。从 F_1 -score 可以看出本文模型对裂缝的分割位置和精度更为准确,达到最高 91.66%。因此,从以上定量和定性的评价实验可以得出结论此模型对不同类型的混凝土桥梁裂缝具有更好的检测性能。

3.6 性能评价指标分析

3.6.1 垂直裂缝检测实验

首先,进行混凝土桥梁垂直裂缝的检测实验,垂直裂缝主要是桥梁间隙与结构组织之间的裂缝,不同方法实验结果如图 11 所示。其中,图 11(b)为 U-Net 网络检测结果,结果表明 U-Net 网络能基本实现对桥梁裂缝轮廓的分割,但全部实验的裂缝图像都显示裂缝间断点较多且抵抗背景噪声的干扰能力弱,大量背景噪声点影响检测结果。图 11(c)为 Mask R-CNN 网络检测结果,可以发现该方法相比 U-Net 能够对背景噪声点起到一定的抑制作用,但仍会出现裂缝检测不连续的断裂问题,如图 11(c)中圆圈处存在裂缝断裂情况,此外,还出现了裂缝漏检的问题,如该行第 1 幅右侧和第 3 幅下侧矩形框内存在裂缝漏检等问题。图 11(d)为 Swin-Transformer 网络检测结果,发现该方法相比 Mask R-CNN 网络分割的裂缝断裂处有所改善,但个别垂直裂缝仍存在不连续的问题,如该行第 1 幅和第 3 幅箭头处。图 11(e)为 InternImage 模型检测结果,该方法通过可变形卷积来拟合裂缝几何结构,一定程度上提高了裂缝检测的性能,但仍存在漏检的问题,如第 1 幅右侧裂缝分支未实现检测。图 11(f)为 Mask2former 模型裂缝分割结果,由于该模型自身有解码器序列间断缺陷,导致裂缝检测结果中存在检测不连续

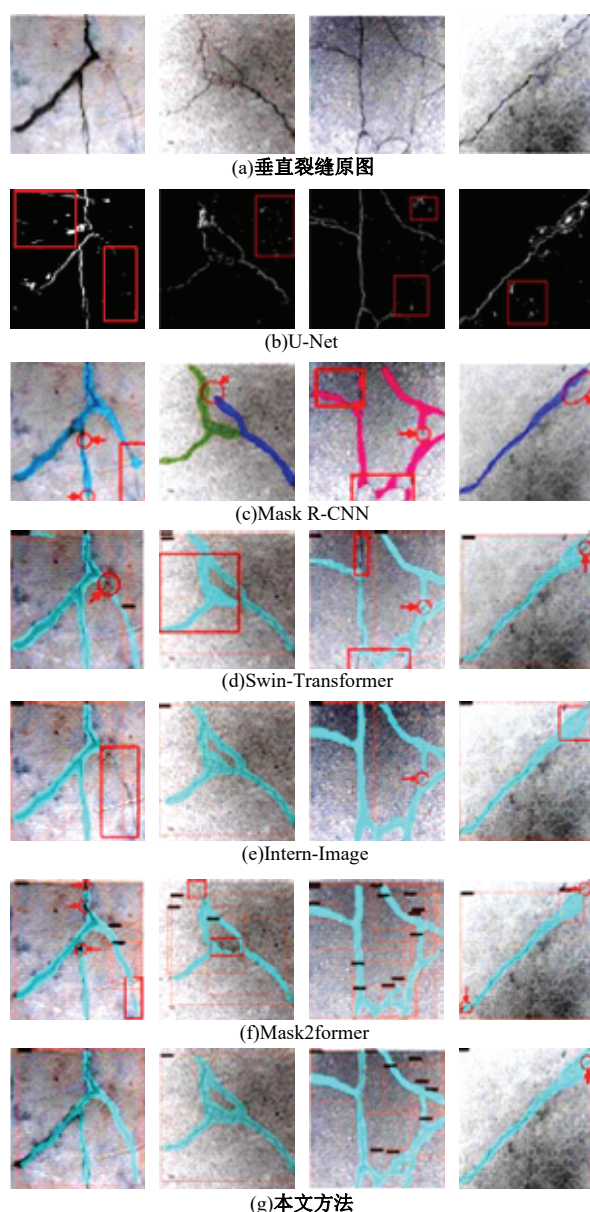


图 11 垂直裂缝实验对比图

Fig. 11 Comparison of vertical crack experiments

的问题。图 11(g)为本文方法结果,间断点、误检程度等都显示出对桥梁垂直裂缝的分割效果更好,裂缝分割结果优于其他对比方法。

3.6.2 水平裂缝检测实验

在完成垂直桥梁裂缝检测实验后,下面接着进行水平裂缝检测实验。水平裂缝一般出现在桥梁水平面上,形成裂缝病害,其主要原因为桥梁承受的荷载过大,不同检测方法结果如图 12 所示。图 12(b)为 U-Net 检测结果,可以发现对于水平裂缝同样存在较多的背景噪声干扰问题。Mask R-CNN 网络和 Swin-Transformer 方法较 U-Net 方法对噪声有所抑制,但都存在检测裂缝断裂的问题。

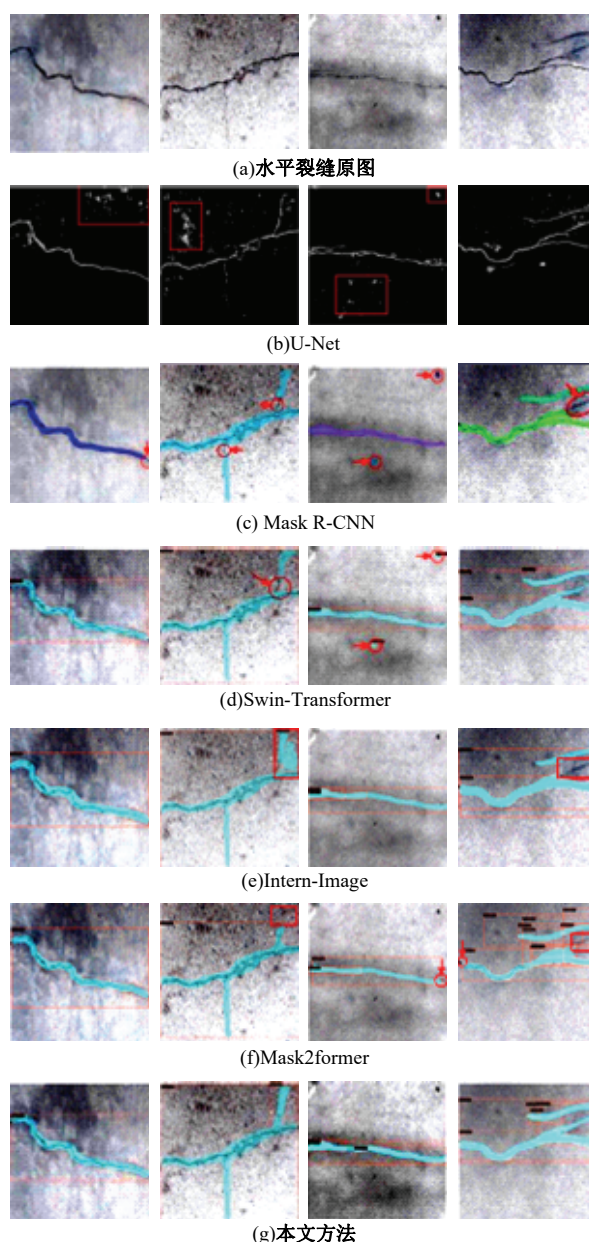


图 12 水平裂缝实验对比图

Fig. 12 Comparison of horizontal crack experiments

题。而 InternImage 和 Mask2former 方法均存在检测不彻底的问题,如第 4 幅右上角分支裂缝均未能完成实例分割检测。本文方法对水平裂缝的检测更加连贯,且对裂缝噪声干扰能够有效抑制。

3.6.3 交叉 X 形裂缝检测实验

交叉 X 形裂缝一般是指裂缝形状类似于字母 X 及其变体形状的裂缝,通常出现在混凝土材料不合格的桥梁中,X 形主要由于不均匀温度导致断裂。下面继续对 X 形裂缝进行对比检测实验,不同方法检测结果如图 13 所示,可以看出,在图 13(b)中,U-Net 网络对 X 形裂缝的检测依然受背景噪声

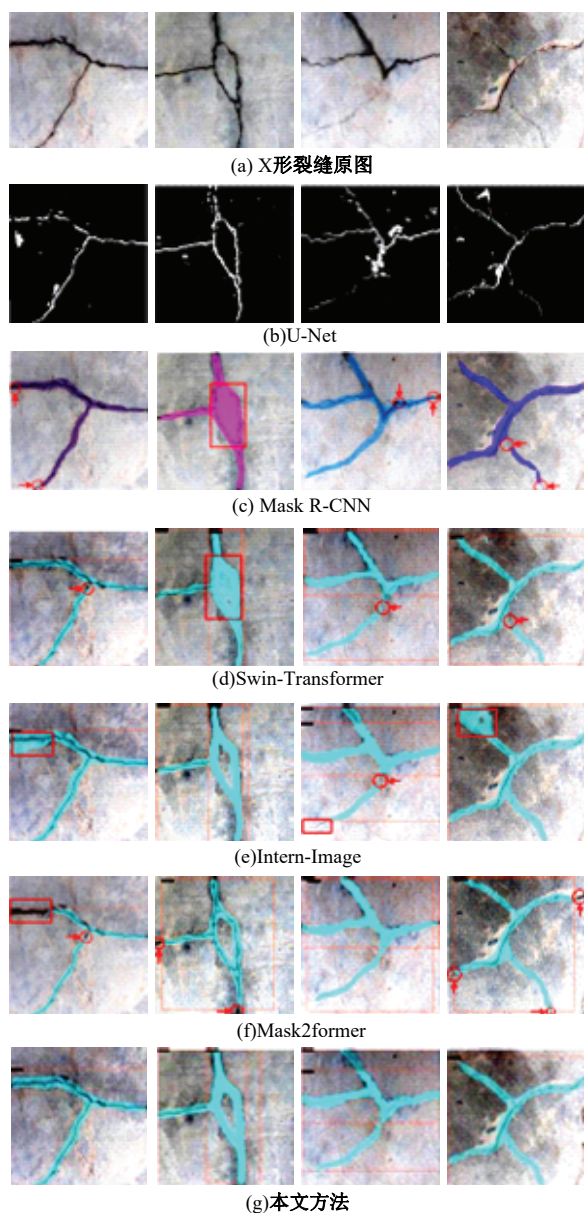


图 13 X形裂缝实验对比图

Fig. 13 Comparison of X-type crack experiments

影响非常大,容易出现断检的问题。从图 13(a)中对交叉X孔洞型裂缝(如第1行第2幅)的检测结果可以发现,Mask R-CNN网络和Swin-Transformer检测后,中心孔区域均出现了粘连检测的问题,此外,对于第1行第4幅裂缝,从检测结果可以发现,Mask R-CNN网络和Swin-Transformer检测后均出现了X右侧支路检测结果断裂的问题。图 13(e)中第3幅InternImage检测后也存在检测结果断裂的问题。而图 13(f)中的Mask2former方法在第1幅左侧裂缝出现了漏分割的问题。对比其他方法发现,本文方法对于X形裂缝能够实现完整分割检测,未出现断点的问题。

3.6.4 网状裂缝检测实验

最后进行网状裂缝检测实验,网状裂缝一般为交叉裂缝未经维护,逐步发展形成网状裂缝,车辆多次对裂缝施加载荷应力,导致裂缝始终处于活动状态,网状裂缝检测对比实验如图 14 所示。其中,图 14(b)为U-Net网状裂缝检测结果,可以看出U-Net模型对网状裂缝存在大量的断检的问题。图 14(c)中Mask R-CNN网络检测也存在大量裂缝无法连续检测的问题,如第4幅右侧分支只完成了部分孤立检测。图 14(d)中Swin-Transformer网络和图 14(e)中Intern-Image模型检测结果的第2幅均出现了严重的网状裂缝检测

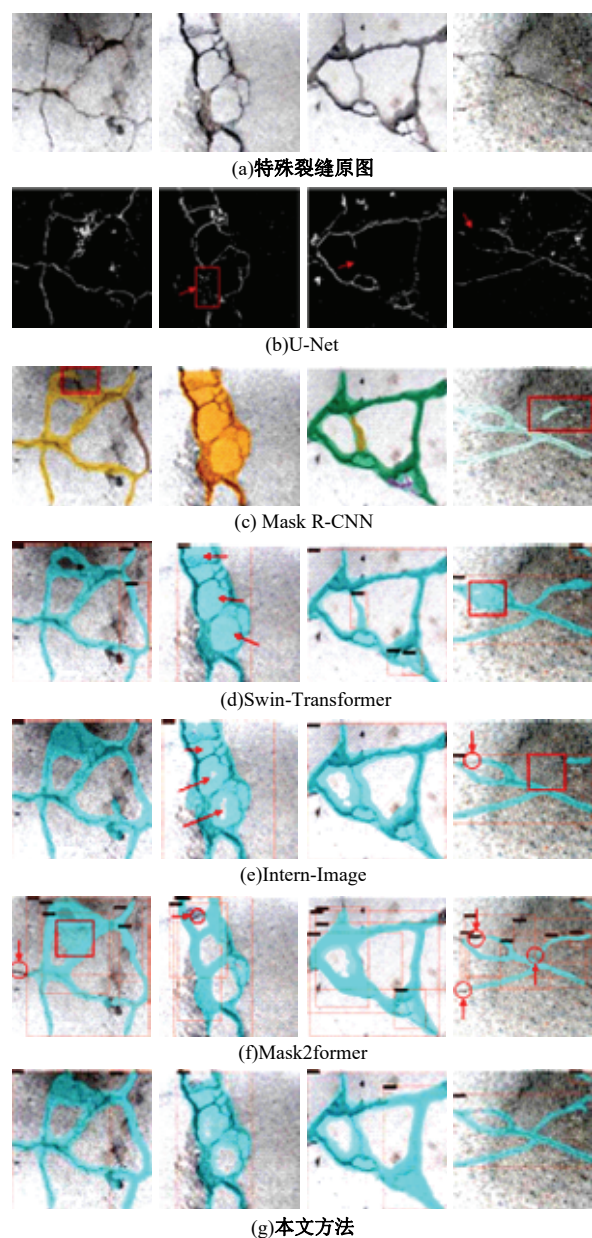


图 14 网状裂缝实验对比图

Fig. 14 Comparison of reticular crack experiments

后空洞粘连的问题,导致其分割精度较低。图 14(f)为Mask2former检测结果,该方法的粘连现象有所改善,但仍存在图像边缘裂缝和分支处裂缝分割的断点问题。图 14(g)为本文方法检测结果,可以看出,其分割性能更优,能够实现对网状裂缝的精细分割,粘连问题得到了改善,同时裂缝检测结果也更加连续。

4 结 论

(1)通过构建基于ViTAE网络的桥梁裂缝特征提取主干网络,促进不同尺度特征信息融合,提高了裂缝特征提取能力。

(2)设计了融合旋转可变窗口自注意力模块,提高了对裂缝旋转特征的利用率,克服了桥梁裂缝旋转特征提取不充分的缺点,使裂缝分割效果更好。

(3)提出使用可变形卷积提高对桥梁不规则裂缝几何形体的拟合能力,通过可变形卷积替换传统卷积运算,改善裂缝边缘的分割性能,减少了裂缝特征信息的丢失。

(4)对不同类型的混凝土桥梁裂缝检测实验表明:本文方法具有更高的检测和分割精度,相较原始Mask R-CNN准确率提高了4.85%,召回率提高了13.95%,本文方法能更准确地实现对混凝土桥梁裂缝病害的检测。

参考文献:

- [1] 杨国俊, 齐亚辉, 石秀名. 基于数字图像技术的桥梁裂缝检测综述[J]. 吉林大学学报: 工学版, 2024, 54(2): 313-332.
Yang Guo-jun, Qi Ya-hui, Shi Xiu-ming. Review of bridge crack detection based on digital image technology[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(2): 313-332.
- [2] 徐铖基, 王小虎, 戴雨晴, 等. 混凝土劈裂损伤和毛细传输过程的原位表征[J]. 土木工程学报, 2023, 56(8): 1-11.
Xu Cheng-ji, Wang Xiao-hu, Dai Yu-qing, et al. In-situ characterization of concrete splitting damage and capillary transport processes[J]. China Civil Engineering Journal, 2023, 56(8): 1-11.
- [3] Zoubir H, Rguig M, Aroussi M E, et al. Concrete bridge crack image classification using histograms of oriented gradients, uniform local binary patterns, and kernel principal component analysis[J]. Electronics, 2022, 11(20): 1-11.
- [4] Vivekananthan V, Vignesh R, Vasanthaseelan S, et al. Concrete bridge crack detection by image processing technique by using the improved OTSU method[J]. Materials Today: Proceedings, 2023, 7: 1002-1007.
- [5] 卢印举, 马芳, 戴曙光, 等. 融合多尺度特征的马尔可夫随机场路面裂缝分割算法[J]. 计算机辅助设计与图形学学报, 2022, 34(5): 711-721.
Lu Yin-ju, Ma Fang, Dai Shu-guang, et al. Markov random field road crack image segmentation algorithm integrating multi-scale features[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(5): 711-721.
- [6] Avendaño J C, Leander J, Karoumi R. Image-based concrete crack detection method using the median absolute deviation[J]. Sensor, 2024, 24(9): 2736.
- [7] Wan C F, Xiong X B, Wen B, et al. Crack detection for concrete bridges with imaged based deep learning[J]. Science Progress, 2022, 105(4): 1-21.
- [8] 朱苏雅, 杜建超, 李云松, 等. 采用U-Net卷积网络的桥梁裂缝检测方法[J]. 西安电子科技大学学报, 2019, 46(4): 35-42.
Zhu Su-ya, Du Jian-chao, Li Yun-song, et al. Method for bridge crack detection based on the U-Net convolutional networks[J]. Journal of Xidian University, 2019, 46(4): 35-42.
- [9] 谭国金, 欧吉, 艾永明, 等. 基于改进DeepLabv3+模型的桥梁裂缝图像分割方法[J]. 吉林大学学报: 工学版, 2024, 54(1): 173-179.
Tan Guo-jin, Ou Ji, Ai Yong-ming, et al. Bridge crack image segmentation method based on improved DeepLabv3+model[J]. Journal of Jilin University (Engineering and Technology Edition), 2024, 54(1): 173-179.
- [10] Xiao R Q. YOLOv5s-GTB: light-weighted and improved YOLOv5s for bridge crack detection[DB/OL]. [2024-01-30]. <https://arxiv.org/abs/2206.01498>
- [11] Deng L, Chu H H, Shi P, et al. Region-based CNN method with deformable modules for visually classifying concrete cracks[J]. Applied Sciences. 2020, 10(7): 2528.
- [12] Liu H J, Yang J, Miao X Y, et al. Crackformer network for pavement crack segmentation[J]. IEEE

- Transactions on Intelligent Transportation Systems, 2024, 24(9): 9240-9252.
- [13] He K M, Gkioxari G, Dollar P, et al. Mask R-CNN [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2018, 42(2): 386-397.
- [14] Zhang Q, Xu Y, Zhang J, et al. ViTAEv2: vision transformer advanced by exploring inductive bias for image recognition and beyond[J]. International Journal of Computer Vision, 2023, 131: 1141-1162.
- [15] Worrall D E, Garbin S J, Turmukhambetov D, et al. Harmonic networks: deep translation and rotation equivariance[C]//IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 7168-7177.
- [16] Wang D, Zhang Q M, Xu Y F, et al. Advancing plain vision transformer towards remote sensing foundation model[J]. IEEE Transactions on Geoscience and Remote Sensing, 2023, 61: 1-15.
- [17] Shish V, Noam S, Niki P, et al. Attention is all you need[DB/OL]. [2024-01-30]. <https://arxiv.org/abs/1706.03762>
- [18] 余家勇, 李锋, 薛现凯, 等. 基于无人机及 Mask R-CNN 的桥梁结构裂缝智能识别[J]. 中国公路学报, 2021, 34(12): 80-90.
- Yu Jia-yong, Li Feng, Xue Xian-kai, et al. Intelligent identification of bridge structural cracks based on unmanned aerial vehicle and Mask R-CNN[J]. China Journal of Highway and Transport, 2021, 34(12): 80-90.
- [19] Zhu X, Hu H, Lin S, et al. Deformable convnets v2: more deformable, better results[C] //IEEE/CVF Conference on Computer Vision and Pattern Recognition, Los Angeles, USA, 2019: 9300-9308.
- [20] 李良福, 马卫飞, 李丽, 等. 基于深度学习的桥梁裂缝检测算法研究[J]. 自动化学报, 2019, 45(9): 1727-1742.
- Li Liang-fu, Ma Wei-fei, Li Li, et al. Research on detection algorithm for bridge cracks based on deep learning[J]. Acta Automatica Sinica, 2019, 45(9): 1727-1742.
- [21] Liu Z, Yutong L, Yue C, et al. Swin-Transformer: hierarchical vision transformer using shifted windows [C]//IEEE International Conference on Computer Vision, Montreal, Canada, 2021: 9992-10002.
- [22] Wang W H, Dai J F, Chen Z, et al. InternImage: exploring large-scale vision foundation models with deformable convolutions[DB/OL]. [2023-8-14]. <https://arxiv.org/abs/2211.05778>
- [23] Bowen C, Ishan M, Alexander G S, et al. Masked-attention mask transformer for universal image segmentation[C]//IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, USA, 2022: 1280-1289.