基于注意力机制和多特征融合的裂缝检测算法

董晓宇[1] 肖佳彤[1] 马凤颖[1] 赵 琳[1]
DONG Xiaoyu XIAO Jiatong MA Fengying ZHAO Lin

摘 要 针对背景噪声较大的路面裂缝检测精度不高的问题，文章中提出一种基于注意力机制和多特征融合的路面裂缝检测算法。基于改进的 ResNet 网络，在编码阶段，使用嵌入双重注意力机制的多尺度扩张残差网络加强对裂缝像素的重点关注，提高模型的特征提取能力；在解码阶段，采用基于空间注意力机制的特征金字塔模块融合多层次和多尺度特征，提高裂缝检测的准确性。在 DeepCrack、CRKWH100 和 CrackTree260 数据集上所提算法表现出优越性能，F_1-score 值达到了 84.44%。

关键词 裂缝检测；注意力机制；扩张卷积；多尺度特征融合；残差网络

doi：10.3969/j.issn.1672-9528.2024.11.013

0 引言

裂缝作为衡量道路质量的重要指标，及时检测修复路面裂缝是道路养护工作的重要内容。传统基于数字图像处理的路面裂缝检测方法，通常假设裂缝具有良好的连续性和高对比度，但在现实生活中路面房屋、树木、杆子等物体的阴影遮挡会导致裂缝亮度不均匀。噪声背景下的全自动裂缝检测仍然是一个挑战。随着卷积神经网络（convolution neural network，CNN）在计算机视觉领域的快速发展，CNN 丰富的层次特征在像素级语义分割任务和裂缝检测方面取得了很大进展，很多研究者提出了基于卷积神经网络的裂缝检测网络模型，其检测精度均高于传统的基于数字图像处理的方法。

Liu 等人[1]提出了一种深层次的卷积神经网络 DeepCrack，该网络由全卷积神经网络和深度监督网络组成，训练过程聚合了从低卷积层到高卷积层的多层次特征，并为每一个卷积阶段提供集成的直接监督。Yang 等人[2]提出了一种端到端的裂缝检测模型，采用特征金字塔（feature pyramid）[3]结构将上下文多尺度特征进行融合，并采用分层推进算法调整样本权重，使得模型能够关注较难识别的样本。瞿中等人[4]提出了一种基于注意力机制和轻量级空洞卷积的裂缝检测算法，该算法通过引入注意力机制来提高模型识别精度。这些网络模型在检测速度和精度上都有较大提升，但对于拓扑结构复杂噪声较大的裂缝图像检测精度和抗噪性能仍有待提高。

本文提出了一种基于注意力机制和多特征融合的路面裂缝检测网络模型，该模型以改进的 ResNet[5] 作为骨干网络，在编码阶段采用嵌入双重注意力机制的扩张残差网络作为特征提取器，在增大网络感受野的同时加强对裂缝的重点关注，实现裂缝像素的准确提取。在解码阶段，采用多特征融合模块将各阶段得到的不同尺度特征图进行融合，形成最终的预测图，以提高裂缝检测的准确性。

1 基于注意力机制和多特征融合的路面裂缝检测

该网络模型需要对输入图像中每个像素点的类别进行区分，即判断裂缝像素和非裂缝像素，可将其视为图像的二分类操作。裂缝检测过程分为模型的训练过程和模型测试过程，保存训练好的网络模型，通过在不同的数据集上进行测试，验证模型是否具有较好的裂缝检测性能。

1.1 模型结构

本文所提出的裂缝检测网络模型结构如图 1 所示。

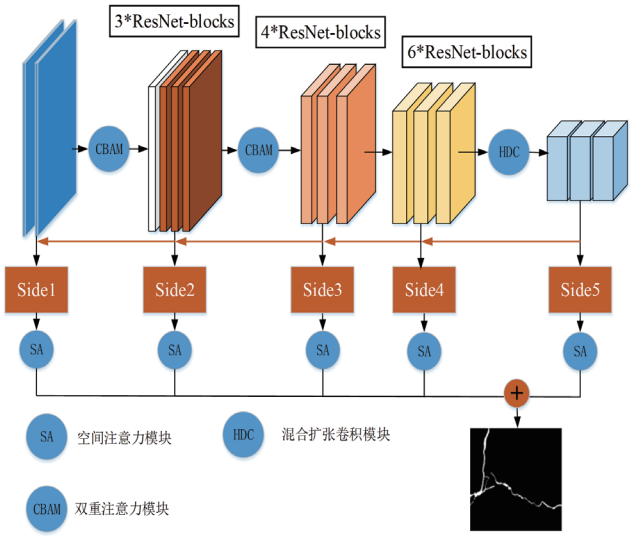


图 1 网络模型结构

1. 信阳学院大数据与人工智能学院 河南信阳 464000

该模型采用非对称的编码 - 解码结构，编码阶段采用改进的 ResNet 网络来提高特征提取能力，由于在浅层网络中提取到的噪声较大，因此在前两个卷积阶段的残差块中嵌入双重注意力机制（convolutional block attention module，CBAM）[6]，减少噪声影响提取更有效的裂缝信息。在最后一个卷积阶段采用混合扩张卷积来解决图像分辨率减小带来的细节信息丢失问题。在解码阶段，将各卷积阶段得到的多尺度多层次特征图进行融合，并再一次采用空间注意力（spatial attention，SA）抑制非裂缝特征，提高检测的准确性。

1.2 注意力机制

针对背景复杂的裂缝图像，在特征提取时常将噪声视为裂缝像素，导致检测精度不高。浅层网络具有丰富的细节信息和边缘信息，为提取到更有效的裂缝特征减少噪声干扰，本文在编码阶段的浅层网络中应用空间 - 通道注意力机制。通过通道注意力加强网络模型对有用通道的关注增大有用通道的权重，通过空间注意力让网络模型更加关注某一通道上的空间信息，提高对裂缝像素的定位能力。在 ResNet 前两个卷积阶段的残差块中嵌入注意力模型，图 2 可以看出特征图 side3 经过注意力模块处理前后的对比图。在第二卷积阶段后，提取到的特征图噪音信息明显减少，且边缘信息更加清晰，为后续的卷积层提供了更丰富的细节特征。

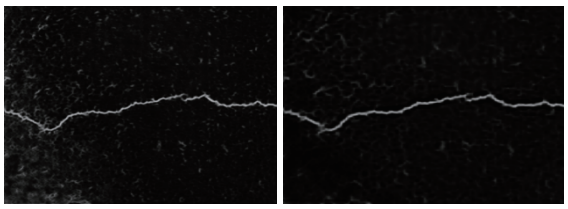
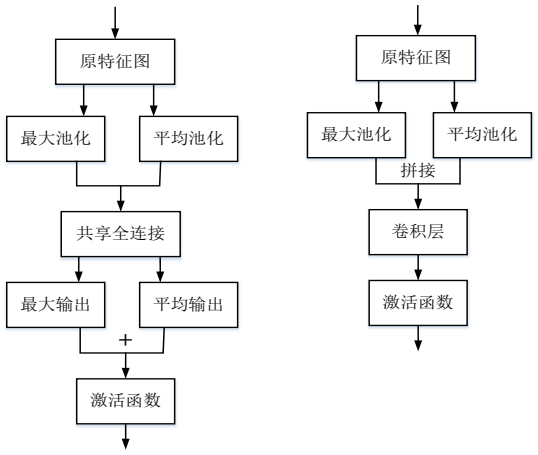


图 2 经注意力模块处理的前后对比图

空间 - 通道注意力模块如图 3 所示，采用平均池化和最大池化的方法对特征地图的空间信息和通道信息进行聚合。



(a) 通道注意力　　(b) 空间注意力

图 3 注意力机制

平均池化在图像中保留更多的全局信息，而最大池化保留更多的纹理信息。聚合两个池化操作可以大大提高网络模型对裂纹像素的聚焦能力。

1.3 混合扩张卷积

随着网络模型的加深，图像分辨率越来越小，会带来较多细节信息丢失的问题，因此在网络末端采用混合扩张卷积替换 ResNet 第五阶段的卷积操作。在第五阶段采用卷积核大小为 3×3，空洞率为 1、2、4 的卷积层，在每一卷积层加入 1、2、4 的空洞率后，感受野由原来 3、5、7 的变成 3、7、15，感受野达到原来的两倍，经过扩张卷积得到的特征图拥有更多的全局语义和细节信息。对于串联结构的卷积而言，裂缝特征会随着级联卷积被不断地稀释，为解决这一问题，本文采用串并联相结合的方式如图 4 所示，将经过 3 层扩张卷积得到的特征图进行拼接， 将拼接后的特征图并入到主干网级联的扩张卷积层中。通过该方式可以在既不增大网络参数的情况下又能捕获多尺度上下文信息。

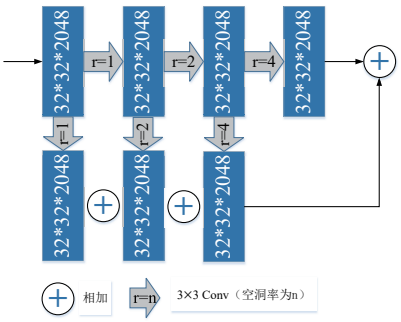


图 4 混合扩张卷积模块

1.4 多尺度特征融合

在编码阶段进行特征提取时，深层网络的语义信息表征能力强，但图像分辨率低，缺乏细节信息；而浅层网络的细节信息丰富，但噪声较大语义信息表征能力弱。为了获取丰富的裂纹信息，可以将深层语义信息和浅层细节信息进行融合。因此，在解码阶段，采用基于空间注意力机制的特征金字塔模块 (FPN-SA)，在将深层语义信息逐层与浅层网络融合时加入空间注意力模块，加强网络模型对深层语义信息的引导作用，加强深层语义信息的权重，从而解决深层语义信息丢失的问题。多尺度特征融合模块结构如图 5 所示。

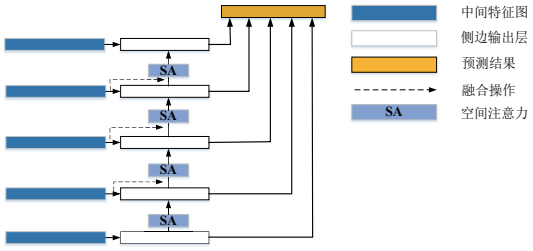


图 5 基于空间注意力机制的多尺度特征融合模块

在经过多尺度特征融合模块后得到 5 个侧边输出特征图。将 5 个侧边特征图分别经过卷积核为 1 的卷积层进行降维操作，然后经过转置卷积进行上采样到与输入图像相同尺寸，最终将上采样后的特征图进行融合形成最终的预测图。侧边特征融合模块网络结构如图 6 所示。融合后的特征图包含了更丰富的多层次多尺度裂缝信息，使预测结果更加贴合路面真实情况。

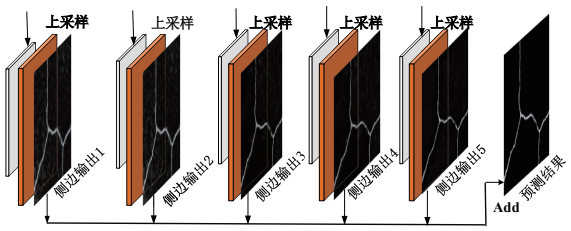


图 6 多尺度特征融合

2 实验结果分析

2.1 数据集

（1）DeepCrack：它是一个人工标注的多尺度、多场景的公开基准裂缝数据集。该数据集由 537 张 RGB 彩色图像组成，其中 300 张作为训练集，237 张作为测试集。本文使用 DeepCrack 的训练数据集来训练网络模型。

（2）CRKWH100[7]：包含了 100 张由线阵相机在地面采样距离为 1 mm 且可见光照明下拍摄的路面图像，其大小均为 512×512 的裂缝图像。本文将 CRKWH100 作为测试集。

（3）CrackTree260[8]：CrackTree260 数据集包含 260 张路面裂缝图像，采用区域阵列相机在可见光照射下拍摄，所有图像的大小为 800×600，且图像均受噪声和阴影的影响较大。CrackTree260 也作为测试集。

2.2 评价标准

为了对该网络性能进行评价引入裂缝检测领域常见的几个评价指标，精确率（precision，P）、召回率（recall，R）和 F_1-score（F_1），计算公式为：

$$P = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{1}$$

$$R = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{2}$$

$$F_1 = \frac{2PR}{P + R} \tag{3}$$

式中：P 是指在被所有预测为正的样本中实际为正样本的概率。R 是指在实际为正的样本中被预测为正样本的概率。F_1-score 是同时考虑正确率和召回率，当两者同时达到最大值时取平衡值。在本文中更加关注 F_1-score 的值，因为它同时集成了准确率和召回率，可以准确地评估网络。

2.3 对比方法

为了验证所提网络模型的有效性，本文将其与现有裂缝检测方法进行了比较。在所有比较方法中，均设置相同的网络参数。输入图像大小均设置为 256×256，Epoch 设置为 700，每 50 个 epoch 保存一次网络模型，学习率设置为 1e-4，权重衰减为 2e-4。

（1）U-Net[9]：U-Net 是基于全卷积网络下的一个语义分割的网络模型，采用典型端到端且对称的编码器 - 解码器网络结构。

（2）SegNet[10]：SegNet 是一个进行像素级语义分割的网络模型，在整体上运用了对称的编码器 - 解码器的结构。

（3）HED[11]：HED 在全卷积神经网络和 VGG16 的基础上进行改进，通过多个侧边输出不同尺度的边缘，然后通过一个训练的权重融合函数得到最终的边缘输出，是一个非常有效的边缘检测网络。

（4）CrackSeg[12]：一个端到端可训练的裂缝分割网络用于像素级的道路裂缝检测，该网络充分利用了层次卷积特征的语义信息，对于复杂场景下的裂纹检测非常有效。

2.4 实验结果与分析

将所提出的网络模型在 DeepCrack 数据集上进行消融实验，将 DeepCrack、SegNet、HED 和 CrackSeg 网络模型和本文所提出的模型在 DeepCrack 数据集、CRKWH100 数据集和 CrackTree260 数据集上进行测试。实验目的在于验证所提网络的有效性。

（1）消融实验：为了研究注意力机制和多尺度特征融合模块在本文所提网络模型中的有效性，以多尺度扩张残差网络为基准，实验结果显示在添加了两个模块以后网络取得最好的性能，表 1 为在 DeepCrack 数据集上的消融实验结果。

表 1 在 DeepCrack 数据集上的消融实验

Methods	DeepCrack		
	P	R	F_1-Score
ResNet	0.819 8	0.718 9	0.766 1
ResNet+CBAM	0.768 8	0.831 0	0.798 7
ResNet+FPN-SA	0.840 1	0.829 3	0.834 7
Ours	**0.835 4**	**0.853 6**	**0.844 4**

（2）在 DeepCrack 数据集上的结果：在 DeepCrack 数据集上其精确率召回率曲线如图 7 所示，本文所提出的方法表现出最优的效果。如表 2 所示本文所提的方法在 DeepCrack 数据集上 F_1-score 值达到 84.44%，其值高于第二好的 DeepCrack 方法。可视化结果如图 8 所示，第一行为在 DeepCrack

数据集上的结果，从视觉角度看本文所提的方法受噪声的影响最小，能够提取到较为清晰的裂缝。

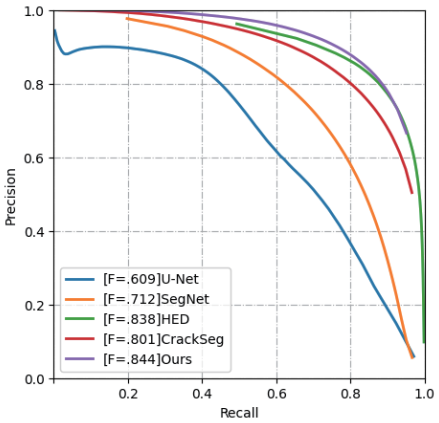


图 7 在 DeepCrack 数据集上的 Precision-Recall 曲线

表 2 在 DeepCrack 上的对比实验结果

Methods	DeepCrack		
	P	*R*	F_1-score
U-Net	0.637 1	0.582 3	0.608 5
SegNet	0.732 5	0.693 1	0.712 3
HED	0.825 0	0.852 1	0.838 3
CrackSeg	0.788 8	0.813 3	0.800 9
Ours	**0.835 4**	**0.853 6**	**0.844 4**

（3）在 CRKWH100 数据集上的结果：从图 9 和表 3 可以看出，本文提出的方法在 CRKWH100 数据集上检测精度是最高的，F_1-score 值达到 79.93%，均优于目前现有的裂缝检测方法。其可视化结果如图 8 第二行所示。

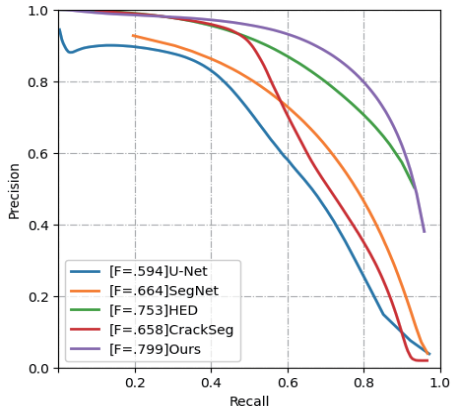
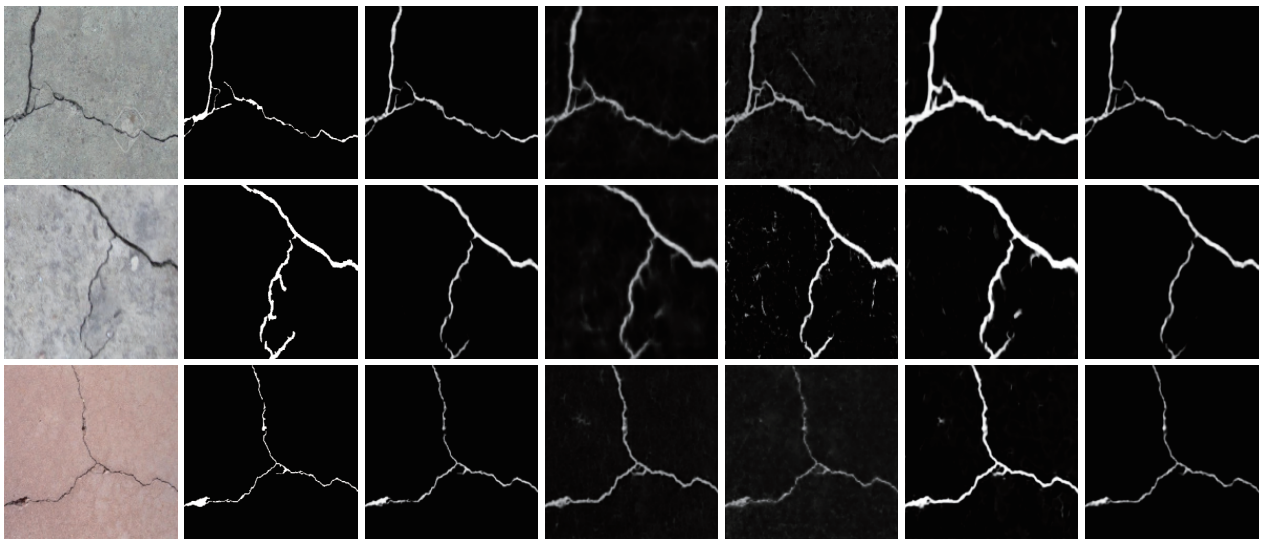


图 9 在 CRKWH100 数据集上的 Precision-Recall 曲线

表 3 在 CRKWH100 上各算法的对比实验结果

Methods	CRKWH100		
	P	*R*	F_1-score
U-Net	0.659 7	0.540 9	0.594 4
SegNet	0.683 5	0.645 9	0.664 2
HED	0.739 0	0.766 9	0.752 7
CrackSeg	**0.826 7**	0.546 2	0.657 8
Ours	0.816 5	**0.782 7**	**0.799 3**

（4）在 CrackTree260 数据集上的结果：如图 10 所示，本文提出的方法在 CrackTree260 数据集上的检测效果优于其他方法，在 Precision-Recall 曲线上获得了最高值。由表 4 可知本文方法的 F_1-score 值为 62.25%，较第二高提高了 1.67%，图 8 的第三行显示了几种方法在 CrackTree260 数据集上测试的结果，从视觉效果看本文提出的方法能够检测出较为连续的细裂缝，证明本文所提出的网络模型具有较好的泛化能力。



（a）原图 （b）标签图像 （c）Ours （d）U-Net （e）SegNet （f）HED （g）CrackSeg

图 8 三个数据集上不同网络的检测结果图

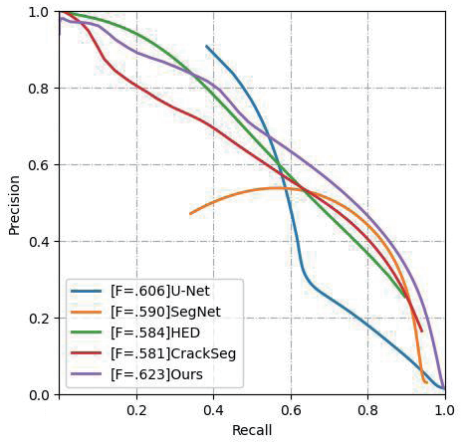


图 10 在 CrackTree260 数据集上的 Precision-Recall 曲线

表 4 在 CrackTree260 上各算法的对比实验结果

Methods	CrackTree260		
	P	*R*	F_1-score
U-Net	**0.721 1**	0.522 3	0.605 8
SegNet	0.501 6	**0.715 8**	0.589 8
HED	0.598 6	0.569 9	0.583 9
CrackSeg	0.527 0	0.646 9	0.580 9
Ours	0.585 0	0.665 1	**0.622 5**

3 结语

本文提出了一种基于注意力机制和多特征融合的路面裂缝检测方法。使用多尺度扩张残差网络作为特征提取器，利用混合扩张卷积解决了图像分辨率降低带来的细节信息丢失问题，同时为了降低噪声带来的影响，在浅层网络的每一个残差块中嵌入注意力模块，从通道和空间两个维度对图像不同区域的权重进行自适应调整。在解码阶段，采用基于空间注意力机制的特征金字塔模块将深层语义信息与浅层的细节信息进行充分融合，得到 5 个侧边输出特征图，将 5 个侧边输出特征图进行上采样与拼接，得到最终的预测结果图。本文方法相比于现有的裂缝检测方法在精确度方面有一定程度的提升，同时具有较好的泛化性能。

如何实现精确的裂缝自动检测，是今后应该继续研究的问题。从而进一步提高裂缝检测的精确度，不断提高网络模型的泛化能力，将其用于其他的目标检测领域。

参考文献：

[1] LIU Y H, YAO J, LU X H, et al. DeepCrack: a deep hierarchical feature learning architecture for crack segmentation[J]. Neurocomputing, 2019, 338(3): 139-153.

[2] YANG F, ZHANG L, YU S J, et al. Feature pyramid and hierarchical boosting network for pavement crack detection[J]. IEEE transactions on intelligent transportation systems, 2020, 21(4): 1525-1535.

[3] LIN T Y, DOLLÁR P, GIRSHICK R, et al. Feature pyramid networks for object detection[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2017: 936-944.

[4] 瞿中，王彩云．基于注意力机制和轻量级空洞卷积的混凝土路面裂缝检测 [J]. 计算机科学，2023(2): 231-236.

[5] HE K M, ZHANG X Y, REN S Q, et al. Deep residual learning for image recognition[C]//IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016: 770-778.

[6] WOO S, PARK J, LEE J Y, et al. CBAM: convolutional block attention module[C]//Computer Vision - ECCV 2018. Berlin: Springer, 2018: 3-19.

[7] ZOU Q, ZHANG Z, LI Q Q, et al. DeepCrack: Learning hierarchical convolutional features for crack detection[J]. IEEE transactions on image processing, 2019, 28(3): 1498–1512.

[8] ZOU Q, CAO Y, LI Q Q, et al. CrackTree: Automatic crack detection from pavement images[J]. Pattern recognition letters, 2012, 33(3): 227-238.

[9] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation[C]//Medical Image Computing and Computer-Assisted Intervention - MICCAI 2015. Berlin: Springer, 2015: 234-241.

[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] XIE S N, TU Z W. Holistically-nested edge detection[C]// IEEE International Conference on Computer Vision. Piscataway: IEEE, 2015: 1395-1403.

[12] SONG W D, JIA G H, ZHU H J, et al. Automated pavement crack damage detection using deep multiscale convolutional features[J]. Journal of advanced transportation, 2020, 2020(1): 6412562.1-6412562.11.

【作者简介】

董晓宇（1997—），女，河南信阳人，硕士，助教，研究方向：数字图像处理、人工智能。

（收稿日期：2024-08-06）