

[文章编号] 1003-4684(2022)02-0055-05

基于高语义特征与注意力机制的桥梁裂缝检测

王墨川¹, 王波^{2,3}, 王熊珏^{2,3}, 夏文祥^{2,3}, 阮小丽^{2,3}

(1 湖北工业大学电气与电子工程学院, 湖北 武汉 430068;

2 桥梁结构健康与安全国家重点实验室, 湖北 武汉 430034; 3 中铁大桥科学研究院有限公司, 湖北 武汉 430034)

[摘要] 为提高桥梁裂缝的检测效果, 提出了一种基于高语义特征与注意力机制的桥梁裂缝检测 (deep bridge crack detection, DBCD) 网络。该网络基于编码器-解码器网络架构, 首先编码器基于 SegNet, 在下采样过程中引入最大池化索引, 以避免提取桥梁裂缝特征时细节损失; 其次在编码器和解码器间加入高语义特征融合模块, 提高检测不同桥梁裂缝尺度的鲁棒性; 再次在解码器阶段引入注意力机制, 设计基于混合域注意力机制的解码器模块, 实现对桥梁裂缝的准确定位。最后在现场真实拍摄的桥梁裂缝图像数据集上, 与传统的网络进行了比较, DBCD 具有更丰富的细节、更准确的定位及更高的检测精度。

[关键词] 高语义特征; 注意力机制; 桥梁裂缝检测

[中图分类号] TP23, TP181, U8 [文献标识码] A

裂缝作为最主要的桥梁病害之一, 严重影响着桥梁的安全, 准确高效地检测裂缝对桥梁养护有着积极的意义。目前, 噪声背景下的全自动裂缝检测仍然是一个挑战^[1]。

桥梁裂缝检测方法分为边缘检测和图像分割, 在输入图像质量高、裂缝具有良好的连续性和高对比度情况下, 传统的边缘检测^[2]和图像分割^[3-4], 都可以高精度的检测出裂缝。然而, 在实践中, 背景噪声、阴影、曝光方向或裂缝本身形状多样、尺度多变等都可能影响传统方法的裂缝检测性能。近年来, 深度学习在图像分类^[5]、图像分割^[6]、语音识别^[7]等方面取得了巨大成功。基于深度学习的图像分割对图像整体分割, 通常采用编码器-解码器架构。编码阶段中池化层的索引可以提高边界定位的准确性^[8]。在编解码融合过程中, 全卷积网络^[9]对多尺度特征逐点相加 (add) 融合, 可以提高分割的准确性; 与全卷积网络不同, U-Net^[10]采用了对应层卷积通道拼接 (concat) 的方式来共享编码器的各层信息, 提高目标检测能力。解码阶段的卷积特征可以提高语义分割的性能, Chen 等^[8]提出了基于空洞卷积的空间金字塔池化 (ASPP), 获取不同尺度上的卷积特征。Gu 等^[11]提出在编码器-解码器网络中加入空间金字塔池化 (SPP), 降低在编码过程中因为卷积层和池化层造成的细节损失, 提高分割的准

确率。

由于 ASPP 易出现棋盘伪影, SPP 采用池化易出现局部信息的损失, 为了更好的检测桥梁裂缝, 本文提出了一种基于高语义特征与注意力机制的桥梁裂缝检测网络, 该网络采用编码器-解码器框架, 编码器基于 SegNet, 在下采样中利用最大池化索引捕获并记录特征图中的边界信息, 避免细节缺失; 在编码器和解码器间加入高语义特征融合 (HSFF) 模块, 该模块根据分层的卷积模块提取更高语义特征, 再根据融合后的更高语义特征提取编码器的高语义特征, 得到最终的高语义特征; 在解码器阶段, 不是简单的高低层特征融合, 而是引入混合域注意力机制来恢复桥梁裂缝细节, 实现对桥梁裂缝的准确定位, 从而提高桥梁裂缝检测的性能。

1 基于高语义特征与注意力机制的桥梁裂缝检测网络

为实现对桥梁裂缝的有效检测, 提出基于高语义特征与注意力机制的桥梁裂缝检测 (DBCB) 网络, 该网络采用编码器-解码器框架。该网络由三部分组成: 编码器、高语义特征融合模块和带有注意力机制的解码器。DBCB 基于 Segnet 编码网络, DBCB 利用高低层融合连接编码器和解码器网络 (图 1)。编码器网络和解码器网络的特征图在编码

[收稿日期] 2021-10-08

[第一作者] 王墨川 (1996-), 男, 湖北宜昌人, 湖北工业大学硕士研究生, 研究方向为图像处理与模式识别

[通信作者] 王波 (1980-), 男, 湖北宜昌人, 工学博士, 中铁大桥科学研究院有限公司教授级高工, 研究方向为桥梁结构振动控制, 桥梁智能建造

器的卷积阶段和解码器的混合域注意力机制(HDAM)阶段拼接(concat)融合,并且将所有尺度的融合图拼接(concat)融合,得到 DBCB 网络的输出。DBCBC 的特点表现在三个方面:1)在编码器的下采样步骤中利用最大池化索引捕获并记录特征图中的边界信息,避免细节缺失。2)在编码器和解码器之间引入高语义特征融合(HSFF)模块,使网络在进行特征提取时采用不同的卷积核提取高语义特征,更加专注于裂缝及其相关信息,可以更好地提取裂缝的特征,实现对不同尺度裂缝识别的鲁棒性。3)将混合域注意力机制(HDAM)模块引入到解码器中,可以保证解码器网络在解码过程中更加关注裂缝的细节信息,使得裂缝细节更加丰富、定位更准确。

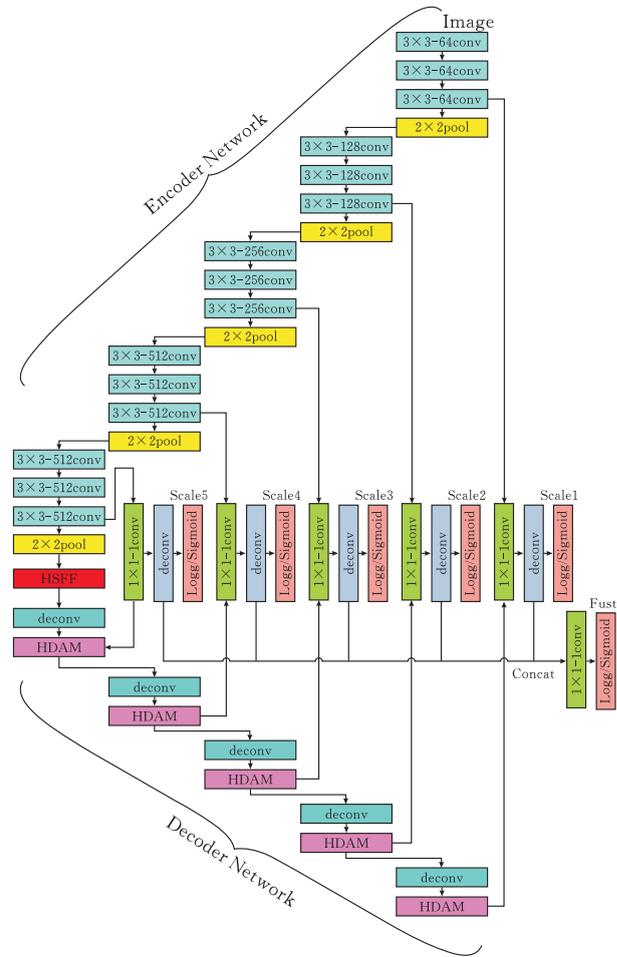


图1 DBCB结构

1.1 编码器模块

编码器网络结构与 VGG16^[14] 网络结构类似,编码器网络主要用于特征提取,采用 VGG16 的前 13 个卷积层和 5 个下采样池化层。丢弃完全连接层,有助于在最深的编码器输出处得到较高分辨率的特征图。每个编码器由卷积层、批归一化层、ReLU 构成。随后,执行最大池化层,窗口设置为 2×2 、步幅设置为 2(非重叠窗口),输出结果等价于下

采样(系数为二)。步长大于 1 的最大池化操作可以减少特征图的比例,同时不会在小的空间移位上引起平移变化,但是下采样将导致空间分辨率的损失,这可能导致边界的偏差。为避免细节的缺失,当执行下采样时,最大池化索引用于捕获并记录编码器特征图中的边界信息。每经过一个编码器,输出特征图尺寸缩小为输入特征图的一半,通道数变为输入特征图的两倍,通过编码器后输出的特征图尺寸为 $16 \times 16 \times 512$ 。

1.2 高语义特征融合模块

由于桥梁裂缝形状、长度和宽度在图像占比中不尽相同,给桥梁裂缝的识别带来了一定的困难。为了解决桥梁裂缝多尺度变化的问题,许多研究者提出了相关方法,例如采用 ASPP 或 SPP 结构,这些结构使得识别率有一定的提高,但是这些方法存在一些问题:ASPP 的空洞卷积操作易出现棋盘伪影效应,SPP 结构的池化操作以损失局部信息为代价^[12]。在缺乏相关信息情况下,融合各尺度特征图,提升的效果有限。为此,本文提出高语义特征融合(HSFF)模块,在编码器获得高层特征后,通过分层的连续卷积操作来提取更高层特征,然后将各层的高层特征统一尺寸后执行相加融合操作,最后将各层的融合结果与执行了 1×1 卷积操作的高层特征执行相乘融合,获得最终的高语义特征。通过高语义特征融合模块,利用不同的卷积核提取更高层特征,以实现对不同尺度桥梁裂缝识别的鲁棒性,解决因桥梁裂缝尺寸不同导致识别困难的问题。

HSFF 结构如图 2 所示。模块边的数字表示输出的特征图的尺寸。HSFF 与编码器直接相连,HSFF 的输入是编码器输出的高层特征(图像尺寸为 $16 \times 16 \times 512$),该特征同时输入到上下两条支路中:上条支路直接对输入特征进行 1×1 卷积操作;下条支路首先通过分层的卷积(conv)模块提取更高层特征,上一层 conv 的输出作为下一层 conv 的输入,一共 3 层。conv 模块采用不同的卷积核对特征图分别进行卷积,再将结果相加融合,其结构如图 2 所示。conv1、conv3、conv5 均采用 3×3 的卷积核,conv2、conv4、conv6 均采用 5×5 的卷积核,conv1-conv6 步长均为 3。前两层卷积层的输出不仅进入下一层卷积层,而且分别再进行一次 7×7 和 5×5 的卷积操作。第三层再进行一次 3×3 的卷积操作之后,经过 2 倍上采样,通过反卷积操作后与上一层(第二层)的卷积结果进行逐点相加(add)融合。同理,在第二层中,第一次融合的输出结果经过 2 倍上采样后与上一层(第一层)卷积结果逐点相加(add)融合,最后第二次的融合结果经过 2 倍上采样

后与上条支路的输出进行逐点相乘(求积)融合,从而得到 HSFF 的输出。

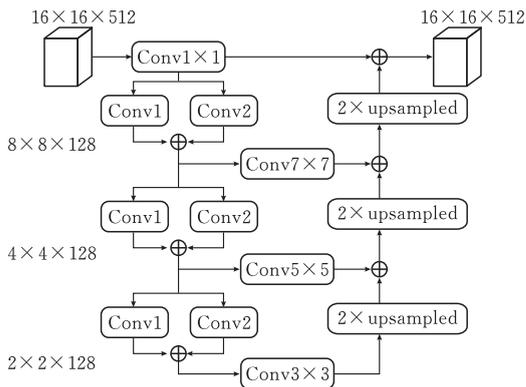


图 2 HSFF 结构

1.3 混合域注意力机制解码器模块

在编码器-解码器体系中,解码器主要作用是逐层恢复图像中目标的细节。传统的解码器工作原理是将高层特征进行上采样、卷积操作后与对应编码器的低层特征融合,实现目标细节的逐渐恢复。但是简单的高底层特征融合对桥梁裂缝细节恢复有限。人类的视觉注意力机制能够从图像中快速聚焦于重要的部分,忽略不重要信息。深度学习注意力机制源于人类的视觉注意力。深度学习注意力机制主要分为通道域、空间域和混合域注意力机制,通道域注意力机制主要关注学习通道域的权重,空间域注意力机制主要关注空间信息,而混合域注意力机制结合了两者特点。混合域注意力机制不仅可以作用于底层特征,而且可以作用于高层特征,更加适用于高低层融合特征^[13]。本文通过在高低层融合的过程中引入混合域注意力机制,使网络提高恢复裂缝细节的能力,抑制无关信息,准确定位,提高输出质量。高低层融合的结构如图 3 所示。

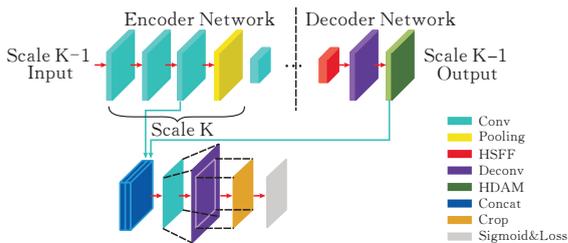


图 3 高低层融合的结构

首先,来自高语义特征融合模块的特征图经过反卷积操作后,引入混合域注意力机制(HDAM),HDAM 结构如图 4 所示,将通道注意力模块(CA)^[14]的输出作为空间注意力模块(SA)^[14]的输入,以着重关注裂缝的特征信息,去除冗余信息,提高输出特征图质量。其次,将 HDAM 的输出与解码器网络中相应尺度的最后一个卷积层拼接(concat)。再次,1x1 卷积层将多通道特征图减少到 1

个通道。最后,为了计算每个尺度中的像素级损失,加入反卷积层以对特征图进行上采样,并使用裁剪层将上采样结果裁剪成输入图像的大小。经过这些运算,可以得到与桥梁裂缝原始图像大小相同的各尺度的预测图。在五个不同尺度中生成的预测图进一步拼接(concat),并且添加 1x1 卷积层,用以融合所有尺度的输出,得到最终的多尺度融合图。

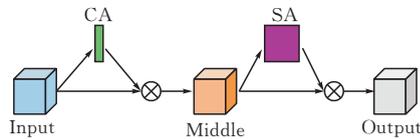


图 4 HDAM 结构

2 实验分析

为验证 DBCB 的有效性,采用现场真实拍摄的桥梁裂缝图像数据进行实验,桥梁裂缝图像是由大疆无人机经纬 M300 RTK、禅思 H20 系列云台相机采集的,要求无人机相机平行于桥梁构件表面拍摄,并且拍摄距离保持在 3m,角度偏差不大于 15°。一共采集了 1400 张原始桥梁裂缝图像,大小为 512x512,并将这 1400 张桥梁裂缝图像分为集合 A(700 张)、集合 B(200 张)、集合 C(500 张)。对集合 A 中的 700 张图像利用图像平移、水平镜像、垂直镜像,加入高斯随机噪声和随机噪声的方法进行数据扩增,用于构建 DBCB 的训练集。扩增后,挑选出 3100 张图像作为训练集。训练集样本的标签使用 labelme 的 Linstrip 标注,line_width 设置为 1。集合 B 作为验证集,集合 C 作为测试集。

2.1 实验设置

在编码器网络中的每个卷积层之后都使用批量归一化,加快训练过程中的收敛。整个网络中卷积层权重初始化的方法为“msra”,偏差初始化为 0。初始全局学习率设置为 0.00002,Batchsize 设置为 8,shuffle 设置为 True,最大训练轮数设置为 400,优化器选用 Adam,损失函数采用交叉熵。在训练时,每 20 轮训练,将学习率设置为原来的一半;验证集的损失函数值降低幅度不超过 5%,则提前结束训练。

实验环境为 NVIDIA GeForce TITAN RTX GPU,Intel(R) Xeon(R) Gold 6146 CPU,32GB 内存,Windows10 操作系统,在 Pytorch 上编程实现。

2.2 识别结果分析

为说明 DBCB 在桥梁裂缝检测中的有效性,将本网络与 U-Net^[15]、SegNet^[16]、DANet^[17] 和 PSP-Net^[18] 进行比较,图 5 各行展示了本文方法与对比方法的测试效果,图中第一列至第七列分别为桥梁

原始图像、人工标注图像、本文方法的识别结果、U-Net的识别结果、SegNet的识别结果、DANet的识别结果、PSPNet的识别结果。从图5可知,与其他网络相比,整体而言,本文网络检测出的裂缝细节信息更丰富,定位更准确,与人工标注的裂缝较吻合,其检测效果明显优于其它网络。

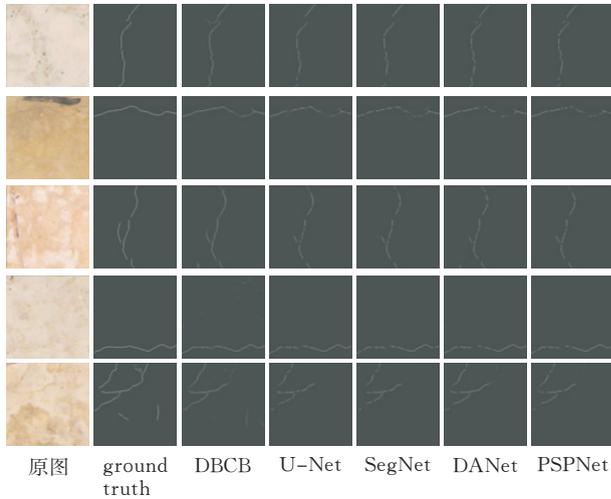


图5 识别效果对比

为对本文提出的桥梁裂缝检测网络实验结果进行量化分析,引入计算机视觉中常用的精确率 P (Precision)、召回率 R (Recall)、 F 值^[19] 以及平均交并比(mIoU)^[20] 进行评估,各指标定义如下

$$\begin{cases} P = \frac{TP}{TP + FP} \\ R = \frac{TP}{TP + FN} \\ F = 2 \cdot \frac{P \cdot R}{P + R} \\ mIoU = \frac{1}{k+1} \sum_{i=0}^k \frac{TP}{TP + FP + FN} \end{cases} \quad (1)$$

其中, TP 表示真阳性、 FP 表示假阳性、 FN 表示假阴性, $k+1$ 表示类别数,包含背景与裂缝两类, k 取为1。不同方法的识别结果与人工标注进行对比,各种网络在测试集中的检测结果见表1。

表1 不同网络的定量比较

网络	P	R	F	$mIoU$	时间/s
DBCBC	0.9032	0.9151	0.9091	0.8124	0.33
U-Net	0.8863	0.8912	0.8887	0.7847	0.26
SegNet	0.8786	0.8897	0.8841	0.7612	0.23
DANet	0.8965	0.9014	0.8989	0.8065	0.29
PSPNet	0.8894	0.9032	0.8962	0.7972	0.28

由表1可知,DBCBC的 P 分别比 U-Net、SegNet、DANet、PSPNet 高出 1.69%、2.46%、0.67%、1.38%, R 分别高出 2.39%、2.54%、1.37%、1.19%, F 分别高出 2.04%、2.50%、1.02%、1.29%, mIoU 分别高出 2.77%、5.12%、0.59%、1.52%。因此,

DBCBC 桥梁裂缝检测精度最好。在图像处理时间上,SegNet 效率最高,增加模块处理时间会增加。

3 结论

本文针对桥梁裂缝难以检测的问题,利用深度神经网络,通过在编码阶段加入最大池化索引,在编码器和解码器之间引入高语义特征融合模块,在解码阶段引入混合域注意力机制,构建了一种基于高语义特征与注意力机制的桥梁裂缝检测网络。实验结果表明,与传统网络相比,本网络具有更好的检测效果和更高的检测精度。

未来将结合 MMX、SSE 等优化算法,在提高算法检测准确率的情况下,进一步提高算法的处理效率,使得网络在实际应用中表现出更好的性能。

[参 考 文 献]

- [1] ZOU Q, ZHANG Z, LI Q, et al. Deepcrack: Learning hierarchical convolutional features for crack detection [J]. IEEE Transactions on Image Processing, 2018, 28(3): 1498-1512.
- [2] XIE S, TU Z. Holistically-nested edge detection[C]// Proceedings of the IEEE International Conference on Computer Vision. 2015: 1395-1403
- [3] MANINIS K K, PONT-TUSET J, ARBELÁEZ P, et al. Convolutional oriented boundaries [C]// European Conference on Computer Vision. Springer, Cham, 2016: 580-596.
- [4] KHOREVA A, BENENSON R, OMRAN M, et al. Weakly supervised object boundaries [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 183-192.
- [5] KOROT E, GUAN Z, FERRAZ D, et al. Code-free deep learning for multi-modality medical image classification [J]. Nature Machine Intelligence, 2021, 3(4): 288-298.
- [6] ISENSEE F, JAEGER P F, KOHL S A A, et al. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation [J]. Nature methods, 2021, 18(2): 203-211.
- [7] SANTHANAVIJAYAN A, KUMAR D N, DEEPAK G. A semantic-aware strategy for automatic speech recognition incorporating deep learning models [M]// Singapore: Springer, 2021: 247-254.
- [8] CHEN L C, PAPANDREOU G, KOKKINOS I, et al. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(4): 834-848.

- [9] LONG J, SHELHAMER E, DARRELL T. Fully convolutional networks for semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015: 3431-3440.
- [10] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2015: 234-241.
- [11] GU Z, CHENG J, FU H, et al. Ce-net: Context encoder network for 2d medical image segmentation[J]. IEEE Transactions on Medical Imaging, 2019, 38(10): 2281-2292.
- [12] CHEN T, CAI Z, ZHAO X, et al. Pavement crack detection and recognition using the architecture of segNet [J]. Journal of Industrial Information Integration, 2020, 18: 100144.
- [13] 曹锦纲,杨国田,杨锡运.基于注意力机制的深度学习路面裂缝检测[J].计算机辅助设计与图形学学报,2020, 32(08):1324-1333.
- [14] WOO S, PARK J, LEE J Y, et al. Cbam: Convolutional block attention module[C]//Proceedings of the European Conference on Computer Vision (ECCV). 2018: 3-19.
- [15] PASZKE A, GROSS S, MASSA F, et al. Pytorch: an imperative style, high-performance deep learning library[J]. Advances in Neural Information Processing Systems, 2019, 32: 8026-8037.
- [16] 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.
- [17] FU J, LIU J, TIAN H, et al. Dual attention network for scene segmentation[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3146-3154.
- [18] ZHAO H, SHI J, QI X, et al. Pyramid scene parsing network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2881-2890.
- [19] OLIVEIRA H, CORREIA P L. CrackIT—An image processing toolbox for crack detection and characterization[C]//2014 IEEE International Conference on Image Processing (ICIP). IEEE, 2014: 798-802.
- [20] LI G, LI X, ZHOU J, et al. Pixel-level bridge crack detection using a deep fusion about recurrent residual convolution and context encoder network[J]. Measurement, 2021, 176: 109171.

Bridge Crack Detection Based on High Level Semantic Feature and Attention Mechanism

WANG Mochuan¹, WANG Bo^{2,3}, WANG Xiongjie^{2,3}, XIA Wenxiang^{2,3}, RUAN Xiaoli^{2,3}

(1 School of Electrical and Electronic Engineering, Hubei Univ. of Tech., Wuhan 430068, China ;

2 State Key Laboratory for Health and Safety of Bridge Structures, Wuhan 430034, China ;

3 China Railway Bridge Science Research Institute, Ltd., Wuhan 430034, China)

Abstract: In order to improve the detection effect of bridge cracks, a Deep bridge crack detection (DBCD) networks based on high level semantic feature and attention mechanism is proposed. The detection networks is based on the encoder decoder network architecture. Firstly, the encoder is based on SegNet, and the maximum pool index is introduced in the downsampling process, which can avoid the loss of details when extracting bridge crack features. Secondly, by adding a high level semantic feature fusion (HSFF) module between the encoder and the decoder, the robustness of detecting different bridge crack scales is improved. Subsequently, by introducing an attention mechanism in the decoder stage, designing a Hybrid domain attention mechanism (HDAM) to achieve accurate positioning of bridge cracks. Finally, compared with the traditional model, DBCB has richer details, more accurate positioning and higher detection accuracy by using it on the real bridge crack image data set taken on site.

Keywords: high-level semantic feature; attention mechanism; bridge crack detection