Received 2 November 2022, accepted 24 December 2022, date of publication 29 December 2022, date of current version 5 January 2023.

Digital Object Identifier 10.1109/ACCESS.2022.3233072

APPLIED RESEARCH

Road Crack Detection Using Deep Neural Network Based on Attention Mechanism and Residual Structure

PENG JING, HAIYANG YU, ZHIHUA HUA, SAIFEI XIE, AND CAOYUAN SONG
School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo 454000, China

Corresponding author: Haiyang Yu (yuhaiyang@hpu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant U1304402.

ABSTRACT Intelligent detection of road cracks is crucial for road maintenance and safety. because of the interference of illumination and totally different background factors, the road crack extraction results of existing deep learning ways square measure incomplete, and therefore the extraction accuracy is low. we tend to designed a brand new network model, referred to as AR-UNet, that introduces a convolutional block attention module (CBAM) within the encoder and decoder of U-Net to effectively extract global and local detail information. The input and output CBAM features of the model are connected to increase the transmission path of features. The BasicBlock is adopted to replace the convolutional layer of the original network to avoid network degradation caused by gradient disappearance and network layer growth. we tested our method on DeepCrack, Crack Forest Dataset, and our own tagged road image dataset (RID). The experimental results show that our method focuses additional on crack feature info and extracts cracks with higher integrity. The comparison with existing deep learning ways conjointly demonstrates the effectiveness of our projected technique. The code is out there at: https://github.com/18435398440/ARUnet.

INDEX TERMS Residual structure, shortcut connection, CBAM attention mechanism, deep learning, road crack detection.

I. INTRODUCTION

Cracks are the foremost common kind of road illness. If cracks repair isn't disbursed in time, cracks can seriously endanger traffic safety. Therefore, finding and repairing cracks in time is an important responsibility of the transportation department. In recent years, with the event of road crack detection strategies for image and computer vision [1], deep learning has been wide used for crack detection [2], [3], [4]. Zhang et al. [5] first used deep learning for road crack extraction and planned and trained a supervised shallow neural network to find cracks. CrackForest [6] combined multi-level complementary features using structural information in crack patches to find and extract cracks. Yao et al. [7] planned a convolutional neural network for crack recognition, that suppressed the interference of background factors and considerably improved detection accuracy. Liu et al. [8] planned a pixel-level classification network combining native and global information to get richer multi-scale feature information and improve crack detection accuracy. Dorafshan et al. [9] reduced the interference of background factors on crack extraction by connecting edge detectors and deep convolutional neural networks. Li et al. [10] increased and extracted multi-scale crack features using dense connections. Finally, the feature maps at totally scales were amalgamate to attain crack extraction by complementing the options at different levels. However, these methods can less extract fine cracks in pavement images with many interfering factors.Lin H et al. [11] proposed LEDNet neural network for defect detection of LED chips, and achieved high detection results. Wu X et al. [12] generate small blocks centered on a pixel at several different scales and input the blocks into different convolution operations.The experimental results show that the method can learn more real fracture characteristics and the detection results are high precision.

The associate editor coordinating the review of this manuscript and approving it for publication was Yongjie Li.

Olaf et al. [13] proposed a U-Net-based medical image segmentation method to obtain contextual semantics by contracting the paths and determining the location by symmetrically expanding the trails. The encoder and decoder sub-networks of U-Net++ are connected by nested and dense jump paths [14] to reduce the semantic gap between the encoder-decoder sub-network feature mappings and Intersection over Union (IOU) is higher than the original U-Net network. Cheng et al. [15] treated the crack images as a whole; They also introduced a cost function based on distance transformation to improve the detection performance of the network. Fan et al. [16] proposed an encoder-decoder-based structured neural network U-HDN that integrates crack context information into a multi-expansion module to obtain more crack features. Drozdzal et al. [17] studied the importance of skip connections and introduced short skip connections in the encoder. ResNet34 residual network [18] was used, and the original convolution of the residual network was replaced with an expanded convolution [19] to extract crack information, and an attention mechanism was introduced to obtain the final crack detection results. these methods have poor detection accuracy in the presence of many background disturbing factors. Bang et al. [20] proposed a pixel-level detection method using an encoder-decoder to identify road cracks.The encoder consists of a convolution layer of a residual network for extracting crack features, and the decoder consists of a deconvolution layer for locating cracks in the input image.The experimental results are better than those of VGG-16, ResNet-50, ResNet-101 and ResNet-200.

U-Net neural network is a coding and decoding structure that can be trained end-to-end using fewer images to detect road cracks quickly. However, there square measure several distracting factors in road pictures, and also the U-Net network is low to extract the fine cracks within the pictures. when the introduction of the CBAM into the U-Net neural network, the structure of the neural network and also the variety of network layers increase, but network model shows network degradation. to solve the above issues, the add this paper focuses on the subsequent aspects:

1) we design a new network model called AR-UNet by introducing the convolutional block attention module (CBAM) in the U-Net neural network. The CBAM performs global averaging and global maximum hybrid pooling of channels and spaces of input features to focus on more global and local detail information. The performance of the neural network in detecting fine cracks is improved.

2) CBAM's input and output features are pooled using shortcut connections to increase the transmission path of crack features, and the network model can learn more about crack features.

3) BasicBlock replaces the convolutional layers of the U-Net network to avoid network degradation due to the increase in the number of network layers. Further, improve the accuracy of crack extraction.

II. RELATED WORK

Traditional road pavement crack detection principally has the subsequent categories: 1) manual detection, 2) threshold method, 3) wavelet transform, 4) morphological image processing and classification, 5) path method and 6) edge detection method. Manual detection is thru the pavement investigator driving on the road to record the situation of cracks, the degree of harm, and therefore the variety of data. Such a way is careful and comprehensive, however the quantity of human and assets consumption is giant and inefficient.

Thresholding-based image segmentation methods have an early origin and are widely used. The thresholding method detects cracks utilizing the feature that the gray value of crack image pixels is lower than the background [21]. Kirschke et al. [22] proposed a histogram-based threshold segmentation method, which can only be used for more apparent crack identification. Removal algorithms [23] using binary segmentation, morphological operations, and removal of isolated points and regions are prone to the presence of gaps in detected cracks. Segmentation using an improved adaptive iterative thresholding segmentation algorithm [24] can also yield crack images. Zhang et al. [25] took advantage of the significant difference between cracks and background to mark contours using FAST feature point recognition and used PYNQ for crack identification. However, the accuracy of those technique is poor once there's a great deal of noise within the background.

Ju et al. [26] use illumination compensation model (ICM) and k-means clustering algorithm to detect cracks, and use k-means clustering algorithm to extract crack area from road background after removing shadow in image.The proposed method has good performance in terms of average precision, recall and F-measure.

Algorithms like wavelet pavement crack detection [27], [28] use wavelet transform to convert cracks and noise into totally different wavelet coefficients. These strategies need high instrumentality necessities and are prone to disadvantages like over-segmentation and condition to interference by external factors.

Histogram statistics and shape analysis algorithms [29], morphological image processing and logistic regression statistical classification [30], and free-form path calculation methods [31], which combine brightness and connectivity to detect cracks. The detection is not practical under the influence of complex backgrounds and the presence of more background-interfering factors, etc. The median filtering algorithm [32] enhances grayscale pavement images using four structural element reconstructions and combines the morphological gradient operator and morphological closure operator to extract crack edges. However, these method can identify crack pixels with noticeable contrast changes in the crack image, and its crack extraction accuracy is poor for cracks with inconspicuous features.

Shah and Wang et al. [33] [34] studied crack segmentation based on edge detection. Still, the natural properties of road diseases were not considered, and the algorithm's applicability was less than ideal. The segmentation algorithm of edge detection is generally based on local grayscale and gradient information to identify crack edges, which is only applicable to cracks with complete edge information. It is easy to judge the background with strong edge information as crack information points. When there is more noise, the effect of edge detection is poor.

In traditional methods, the feature extraction is mainly dependent on the hand-designed extractor, which requires professional knowledge and complicated parameter adjustment process [35], and each method is specific to specific applications, with poor generalization ability.Deep learning is mainly data-driven feature extraction, learning from a large number of samples can be deep, dataset-specific feature representation, the expression of the dataset is more efficient and accurate, the extracted abstract features are more robust and have better generalization ability, and can be end-to-end training without complex parameters. Deep learning detection of cracks in the road can not only liberate people from the complicated work, but also achieve the accuracy of manual detection.Therefore, it is very important to realize automatic detection of road cracks by deep learning.

III. METHOD

A. OVERALL NETWORK STRUCTURE

The U-Net neural network is split into three parts: encoder, decoder, and prediction module. The encoder reduces the image size and extracts the initial image features by convolution and maximum pooling. The decoder obtains the deep features of the image by convolution (a ReLU perform follows every convolution). Finally, pixel classification is completed by 1×1 convolution.

The established network structure is shown in Figure 1. The network structure chiefly consists of a feature extraction network, residual module, and CBAM module. The BasicBlock module replaces the convolutional layer of the U-Net network. BasicBlock module will effectively solve the matter of network model degradation and gradient disappearance once the quantity of network layers will increase. The network introduces CBAM and sums the input and output of CBAM; the module is termed Res-CBAM. Res-CBAM makes the network pay a lot of attention to the channel and spatial dimensions crack information and assign a lot of weights to the network coefficients.

B. CONVOLUTIONAL BLOCK ATTENTION MODULE (CBAM)

CBAM is a light-weight module that contains spatial attention and channel attention. The module derives attention weights consecutive on two freelance dimensions, channel and space, so multiplies the output attention map with the input feature map for adaptative feature refinement. Since CBAM is a light-weight, general module, it is seamlessly integrated into any CNN design. It is trained end-to-end with the underlying CNN. Compared to attention modules specializing in only one facet, CBAM will beware of each side and extract additional information concerning the target.

As shown in Figure 2, assuming $F = C \times H \times W$ as the input feature map, the CBAM module computes the one-dimensional channel attention feature map $M_c \in C \times 1 \times 1$ and the two-dimensional spatial attention feature map $M_s \in 1 \times H \times W$ in turn, and finally outputs the weighted features with channel and space. The overall attention is calculated as follows:

$$F' = M_c(F) \times F \tag{1}$$

$$F'' = M_s\left(F'\right) \times F' \tag{2}$$

where F' denotes the input features after the channel attention operation, F'' is the final refined output.

C. CHANNEL ATTENTION MODULE (CAM)

The structure of the Channel Attention Module is shown in Figure 3; The two $M_c = 1 \times 1 \times C$ feature maps are obtained by feeding the input features into global max pooling and global average pooling, respectively. Then after two layers of the fully connected neural network, the number of neurons in the first layer is $\frac{C}{r}$ (r is the compression rate). ReLu is the activation function, and the number of neurons in the second layer is C. Then, the fully connected neural network's output features are summed and passed through the sigmoid activation function to generate the channel attention features (M_c). The channel attention is calculated as follows:

$$M_c(F) = \sigma(W_1(W_0(F^c_{avg})) + W_1(W_0(F^c_{max}))) \tag{3}$$

where σ denotes the sigmoid function,$W_0 = \frac{C}{r} \times C$, $W_1 = C \times \frac{C}{r}$.

D. SPATIAL ATTENTION MODULE (SAM)

The structure of the spatial attention module is shown in Figure 4. The spatial attention input features $F' = C \times H \times W$ are averaged and max pooling to obtain F'_{avg} and $F'_{\max}$. Then, the two feature maps are channel spliced. After a 7×7 convolution operation, it is compressed into $H \times W \times 1$. It generates M_s by the sigmoid activation function. Finally, the output feature map of this module is multiplied by the input feature map to get the final generated feature map. The spatial attention module is calculated as follows:

$$M_s(F') = \sigma\left\{f^{7\times7}[(F'_{avg} + F'_{\max})]\right\} \tag{4}$$

where σ denotes the sigmoid function and $f^{7\times7}$ denotes the convolution operation with a filter size of 7×7.

E. STRUCTURE DETAILS OF THE ENCODER

As shown in Figure 5, the input features enter the channel attention of CBAM after two convolution operations of size 3×3 to get the channel attention weight M_c. M_c is multiplied by the input feature map to get the input features required by the spatial attention module. Next, the spatial

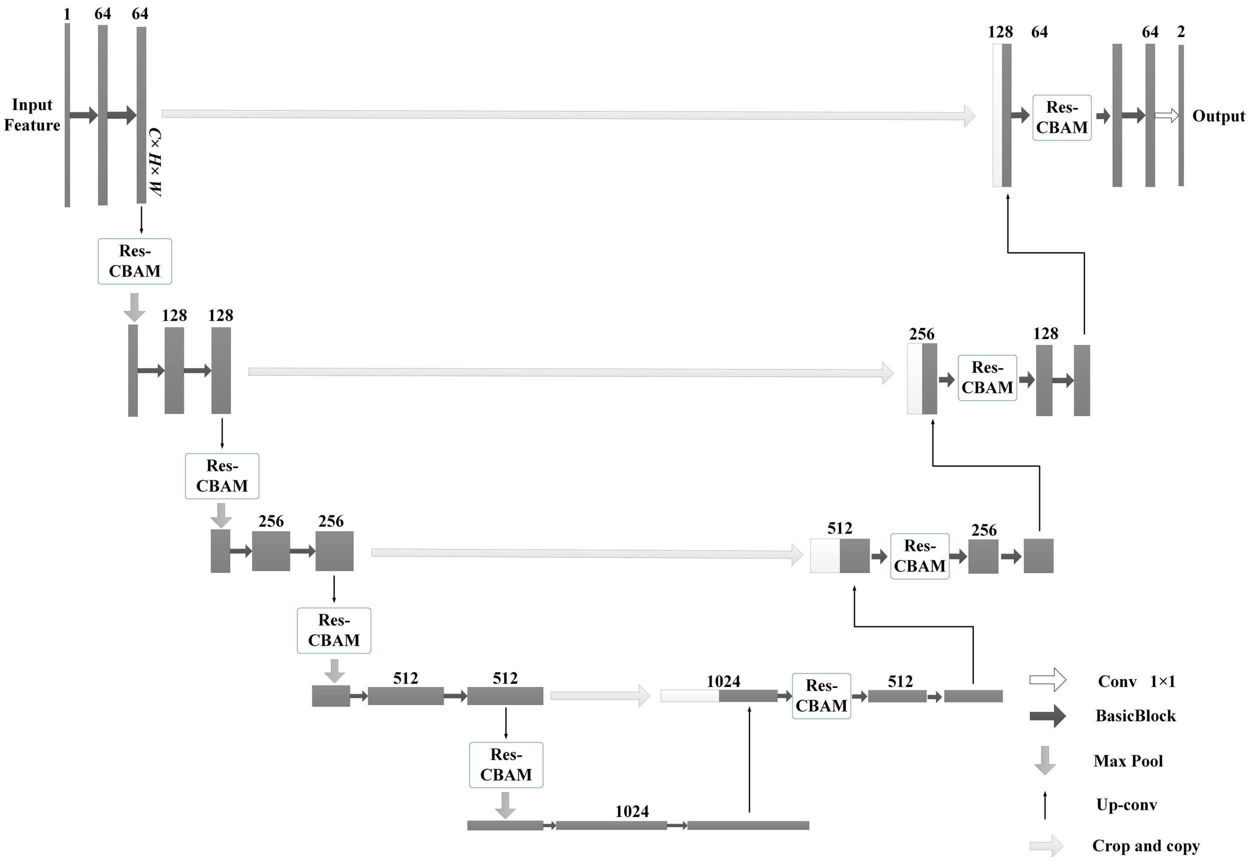


FIGURE 1. The overall structure of the neural network.

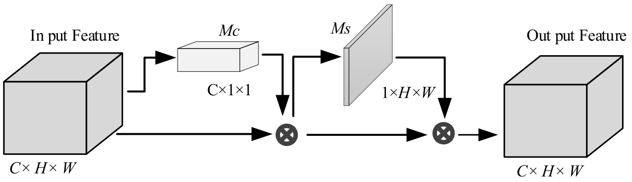


FIGURE 2. Convolutional block attention module.

attention weight M_s is obtained by the spatial attention operation, and the original input and M_s enter the 2×2 max pooling together after the shortcut connection to obtain the final feature map of size $C \times \frac{H}{2} \times \frac{W}{2}$.

F. STRUCTURE DETAILS OF THE DECODER

The residual-connected Res-CBAM is also introduced in the structure of the decoder, as shown in Figure 6. The feature map of size $C \times H \times W$ is deconvolved, and the corresponding CBAM input feature map of the encoder is copied and cut, and stitched with the deconvolved feature map to obtain the feature map of size $C \times 2H \times 2W$; The stitched feature map is input to the attention mechanism as the input feature map. The output feature map is connected with the input feature map and then convolved with a 3 × 3 convolution kernel to obtain the final feature map of size $\frac{C}{2} \times 2H \times 2W$.

G. RESIDUAL NETWORK

The residual network comes from the literature [36]. Typically, because the number of layers will increase, the training loss step by step decreases and then saturates, however the fact tells us that the training loss will increase when the network depth is increased again. this is often not overfitting because, in overfitting, the training loss endlessly decreases.

The deeper the network is, the harder it is to train. Therefore, it is essential to integrate shortcut connections in U-Net networks to cut back network degradation. Since the original convolutional layer is computationally long and unsuitable for pixel-level prediction. the original convolutional neural network layer is replaced by BasicBlock, whose structure is shown in Figure 7.

After the input feature map is passed through two convolutional layers and the ReLu function, it is summed with the original input features to obtain the final output feature map. A residual block can be expressed as:

$$x_{l+1} = x_l + f(x_l, w_l) \tag{5}$$

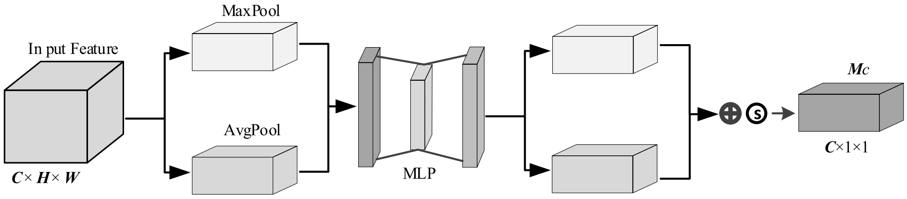


FIGURE 3. Channel attention module.

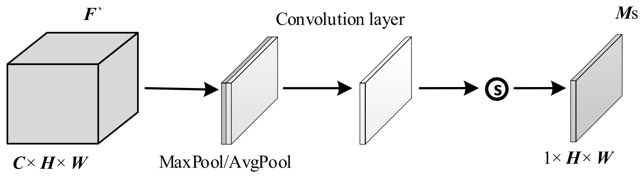


FIGURE 4. Spatial attention module.

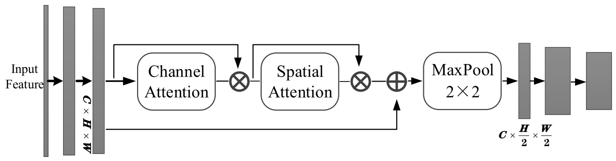


FIGURE 5. Partial structure of the encoder.

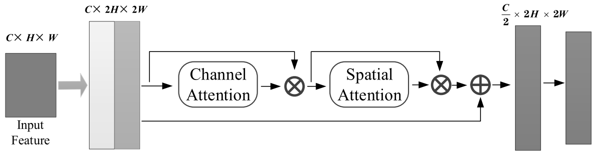


FIGURE 6. Partial structure of the decoder.

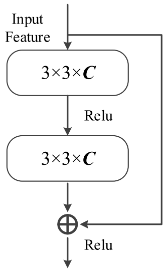


FIGURE 7. The structure of BasicBlock.

The residual block is divided into two parts: the direct mapping part and the residual part. $h(x_l)$ is the direct mapping, and the response is the curve on the right in Figure 7;$f\ (x_l, w_l)$ is the residual part, which consists of two convolution operations, and the part containing the convolution on the left in Figure 7.

The shortcut connections between the input and output feature maps will transfer the crack info extracted by the previous layer of the network to consequent layer. the information loss is avoided to a greater extent, and the network degradation caused by increasing the number of neural network layers is effectively prevented.

IV. EXPERIMENTS AND RESULTS

A. ROAD IMAGE DATA SET

The datasets used for the experiments are DeepCrack [37], Crack Forest Dataset [38], and our annotated onboard road image dataset, which we named RID. DeepCrack is a dataset containing 537 concrete pavement images of 544 × 384 pixels with multi-scene and multi-scale pavement cracks. The Crack Forest dataset is a dataset of asphalt pavement images, which contains 118 images of size 480 × 320 pixels with background noise such as white markers and shadows. These two datasets have fewer images and are enhanced using rotate, flip, and mirror operations. After enhancement, 2148 and 708 images were obtained from the DeepCrack and Crack Forest datasets, respectively. Then, we made a dataset with 548 images from the road images acquired by mobile LiDAR mapping system. The labeled images in these three datasets were manually labeled. To validate the established neural network models, we selected 80% of each dataset as training data and 20% as test data.

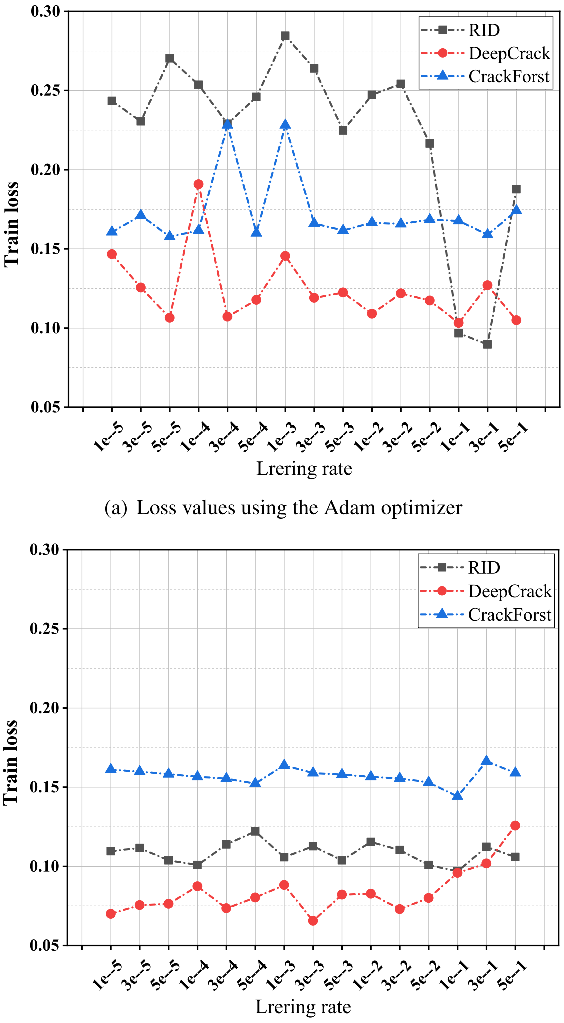
B. EXPERIMENTAL SETTINGS

1) ANALYSIS OF INITIAL LEARNING RATE AND OPTIMIZERS

In the first experiment, In order to obtain a suitable initial learning rate value and the optimization method, we set different learning rates and model optimization methods to analyze the training loss of the model. Figure 8 (a) indicates that we employed the Adam optimizer,The figure indicates that there are large fluctuations in the training loss for the three datasets, and the training loss values are large. Figure 8 (b) indicates that we employed the SGD optimizer. The figure shows that the training loss values of the three datasets are small and stable, therefore, we choose SGD as the network optimizer. The learning rates for the training RID and CrackForst datasets are set to 1e-1 and for the training DeepCrack datasets to 3e-3, because their corresponding loss values are the smallest.

2) OTHER EXPERIMENTAL SETTINGS

We implement all tests in Python 3.6, Pytorch 1.10.1, and CUDA 11.1 framework and use NVIDIA GeForce RTX2080 GPU for training. The model uses the SGD optimisation methodology to update the parameters by arbitrarily choosing



(a) Loss values using the Adam optimizer

(b) Loss values using the SGD optimizer

FIGURE 8. Statistical results of training loss values with different learning rates.

little batches of samples with the momentum optimisation algorithmic rule set to 0.9. The ReLu activation function suppresses gradient disappearance during training to accelerate the convergence rate of the model and maintain stability.

C. EXPERIMENTAL EVALUATION INDEXES

Neural network segmentation accuracy evaluation is performed using commonly used metrics, DICE (D), precision (P), recall (R), and F1-score are selected for assessment. Where DICE indicates the ratio of the area where the predicted and true results intersect with the total area, and the value of perfect segmentation is 1. The F1-score can better measure both the precision and the recall. The DICE and F1-score are calculated as follows:

$$D = \frac{2 \times (R_{seg} \cap R_{gt})}{R_{seg} + R_{gt}} \tag{6}$$

$$F_1 = \frac{2 \times P \times R}{P + R} \tag{7}$$

$$P = \frac{TP}{TP + FP} \tag{8}$$

$$R = \frac{TP}{TP + FN} \tag{9}$$

The exactitude indicates the proportion of properly detected crack pixels that were initially correct. wherever TP indicates the amount of properly classified crack pixels and FP indicates the amount of incorrectly classified crack pixels. Recall indicates the proportion of properly detected cracked pixels to all cracked pixels, wherever FN indicates the amount of pixels incorrectly classified as background.

D. THE RESULTS OF ABLATION EXPERIMENTS

1) VISUAL ANALYSIS OF EXPERIMENTAL RESULTS

To discuss the result of introducing Res-CBAM and BasicBlock within the neural network on crack feature extraction, we tend to validate it by ablation experiments. The tests were done in each of the three datasets. As Figure 9 shows the visualisation results of the experiments, rows 1-2 show the detection results of the DeepCrack dataset, that shows that the original neural network crack extraction is incomplete and the extraction accuracy is poor. after the introduction of Res-CBAM and BasicBlock, the network model can focus more on the crack region, and the crack completeness is higher. Rows 3-4 show the results of the crack forest dataset, and the extracted cracks are more realistic. Rows 5-6 show the results of RID, where the fine cracks are extracted to be more complete.

E. RESULTS OF ABLATION EXPERIMENTS

1) RESULTS ON DEEPCRACK

We explored the contribution of introducing every part on DeepCrack's test set. As shown in Table 1, we found that introducing Res-CBAM improved DICE from 65.39% to 68.72% and F1-scores from 67.26% to 75.64%. And then, we integrated BasicBlock into the original network and found that DICE and F1-scores improved further to 83.91% and 83.67%. we at the same time additional Res-CBAM and BasicBlock into the neural network, and therefore the DICE and F1-scores reached 84.09% and 85.82%, severally. we improve the structure of the encoder and decoder and yield higher extraction accuracy compared to U-Net.

2) RESULTS FOR THE CRACK FOREST DATASET

we can see that the DICE and F1-scores improve to 67.2% and 68.85%, respectively, after the introduction of Res-CBAM and BasicBlock in U-Net. The precision performance of the neural network is better after introducing Res-CBAM alone. The neural networks performed better in recall after introducing BasicBlock alone. But their F1-scores did not perform as well as the networks introduced simultaneously. The experimental results of the crack forest dataset show that the simultaneous introduction of Res-CBAM and BasicBlock can effectively improve the crack detection ability of U-Net.

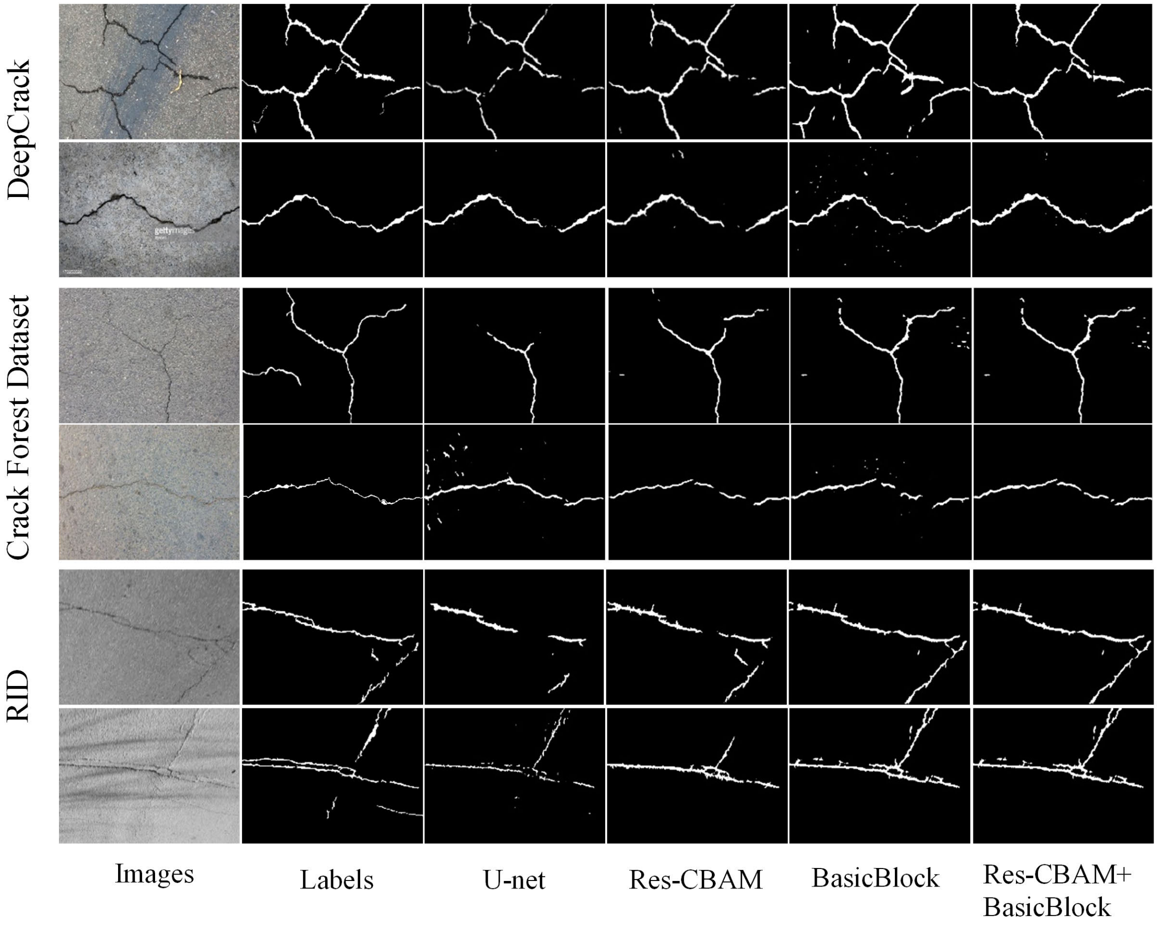


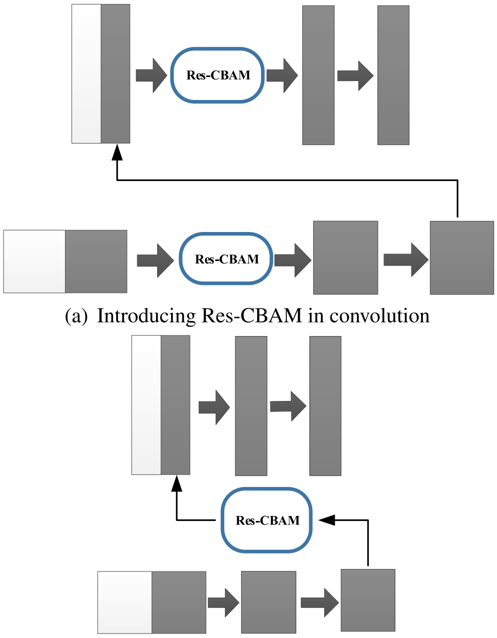
FIGURE 9. Experimental visualization results of three data sets.(where "Res-CBAM" means only Res-CBAM is introduced, "BasicBlock" means only BasicBlock is introduced, and "Res-CBAM+ BasicBlock" means all the two structures are introduced).

TABLE 1. Results on different datasets.(where "+" means the structure is introduced and "−" means the structure is not introduced).

Dataset	U-Net	Res-CBAM	BasicBlock	D(%)	P(%)	R(%)	F1-scores(%)
DeepCrack	+	-	-	65.39	64.20	70.63	67.26
	+	+	-	68.72	80.11	71.64	75.64
	+	-	+	83.91	84.05	**83.29**	83.67
	+	+	+	**84.09**	**89.24**	82.64	**85.82**
Crack Forest Dataset	+	-	-	58.57	61.74	54.85	54.85
	+	+	-	62.99	**70.10**	60.99	65.23
	+	-	+	66.25	57.85	**80.31**	67.25
	+	+	-	**67.22**	60.91	79.18	**68.85**
RID	+	-	-	37.77	57.41	32.89	41.82
	+	+	-	46.02	56.28	46.96	51.20
	+	-	+	39.52	56.31	36.00	43.92
	+	+	+	**50.39**	**58.97**	**52.36**	**55.47**

3) REGARDING THE RESULTS OF RID

we see that the network achieves the simplest performance by introducing attention and residual structure. The DICE and F1-scores reach 50.39% and 55.47%, severally. However, the obtained performance is under the performance on the other datasets. because the road image dataset (RID) has



(a) Introducing Res-CBAM in convolution

(b) Introducing Res-CBAM in up-convolution

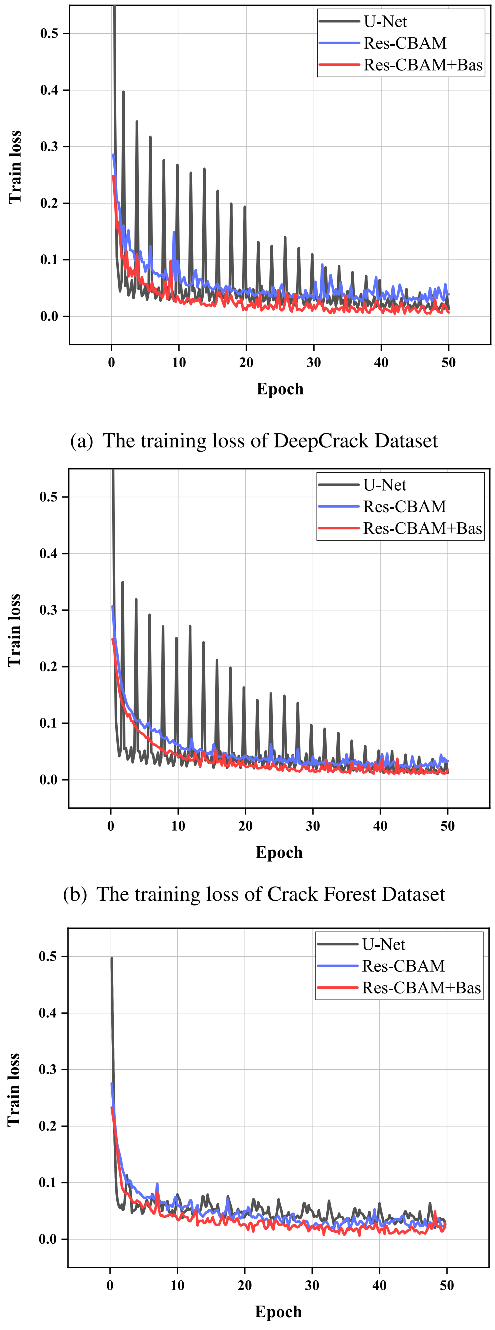
FIGURE 10. Arrangement of Res-CBAM at different positions in the decoder.

TABLE 2. Test results for CBAM in RID dataset with or without residual connections.

Methods	D(%)	P(%)	R(%)	F1-score(%)
CBAM	48.78	56.53	51.34	53.81
Res-CBAM	**50.39**	**58.97**	**52.36**	**55.47**

TABLE 3. Test results of different position arrangement methods.

Position	D(%)	P(%)	R(%)	F1-score(%)
Up-conv	48.84	57.24	52.09	54.54
Conv	**50.39**	**58.97**	**52.36**	**55.47**



(a) The training loss of DeepCrack Dataset

(b) The training loss of Crack Forest Dataset

(c) The training loss of RID

FIGURE 11. Training loss values in different data sets(Where "U-Net" indicates the original network, "Res-CBAM" indicates that the original network introduces Res-CBAM, and "Res-CBAM+Bas" indicates that all two structures are introduced).

uneven illumination and skew shooting angles. additionally, the ground labels of this dataset are just one or some pixels wide, that is one amongst the explanations for the low detection results.

V. DISCUSSION

A. EFFECTIVENESS OF SHORTCUT CONNECTIONS

We additional verified through ablation experiments whether or not adding shortcut connections in CBAM absolutely affects the extraction of cracks. The experimental results are shown in Table 2. we found that by adding shortcut connections, the crack extraction accuracy of the network was improved as a result of the shortcut connections enhanced the path of feature information propagation. The neural network learned more global and local crack information, proving our method's feasibleness.

Since Res-CBAM plays a vital role within the network structure, the position of Res-CBAM could have an effect on the neural network performance. we compare two position ways in which of Res-CBAM placement within the decoder, as shown in Figure 10 (a) and (b). the consequences of

TABLE 4. Results of comparison with other deep learning algorithms.

Dataset	Traditional Deep Learning Algorithms				Transformer Algorithm			
	Methods	**P(%)**	**R(%)**	**F1-score(%)**	**Methods**	**P(%)**	**R(%)**	**F1-score(%)**
DeepCrack	SegNet	73.2	81.2	77.0	VIT	82.6	83.7	83.2
	DeepCrack	53.5	55.5	54.5	Swin-UNet	85.7	83.6	84.6
	RCF	60.1	71.3	65.2	TransUNet	86.2	84.4	85.3
	Literature [41]	82.8	83.7	83.2	ours	**88.9**	**85.7**	**87.2**
	ours	**88.9**	**85.7**	**87.2**				
Crack Forest Dataset	SegNet	42.0	60.2	49.5	VIT	58.6	77.7	66.7
	DeepCrack	46.7	61.5	53.0	Swin-UNet	60.7	75.3	67.2
	RCF	41.5	49.5	45.2	TransUNet	**63.8**	79.8	70.9
	Literature [42]	61.7	72.5	65.4	ours	63.2	**81.2**	**71.1**
	ours	**63.2**	**81.2**	**71.1**				
RID	SegNet	37.2	51.6	43.2	VIT	48.6	51.7	50.1
	DeepCrack	39.3	51.4	44.5	Swin-UNet	52.8	50.9	51.8
	RCF	40.6	49.8	44.7	TransUNet	53.2	54.6	53.9
	ours	**58.9**	**52.3**	**55.4**	ours	**58.9**	52.3	**55.4**

introducing Res-CBAM in convolution and deconvolution on the neural network are discussed. within the same experimental surroundings, the neural networks with the two arrangement methods are tested individually. Table 3 summarizes the test results of different location arrangement methods. The results show that the neural network with the introduction of Res-CBAM in convolution performs higher because the input features of Res-CBAM embrace features from the encoder, that makes the input information richer. Introducing Res-CBAM into the position shown in Fig. 10(b), the DICE and F1-scores are lower because some feature information is lost after the input features are subjected to two convolution operations, leading to a degradation of the network detection performance.

B. NETWORK DEGRADATION IN TRAINING PROCESS

In addition, we also verified the network degradation during the training process by ablation experiments. And we recorded the changes in the training loss values during training of the three datasets. As shown in Figure 11 (a); (b) and (c), the U-Net with the introduction of Res-CBAM shows network degradation due to increased network layers. The figure shows that the loss values of the original U-Net are unstable, fluctuate greatly during the training process, and the neural network converges slowly. After the introduction of Res-CBAM, the neural network pays more attention to the crack features, converging faster. However, due to the increase in network layers, the neural network performance was slightly worse than the original network, and network degradation occurred. So we connected the input and output features of CBAM and replaced the convolutional layer of the original network with BasicBlock. The improved neural network converged faster and with higher accuracy.

C. COMPARISON WITH TRADITIONAL DEEP LEARNING ALGORITHMS

The comparison results with other commonly used methods are shown in Table 4. And our method has higher accuracy compared to SegNet [39], RCF [40], DeepCrack [37] and Literatures [41], [42]. The F1-scores in DeepCrack Dataset are 10.2% higher than SegNet, and also the preciseness and recall square measure 15.7% and 4.5% better, severally. In Crack Forest Dataset, the F1-score is improved by 18.1% compared to DeepCrack, and the precision and recall are improved by 16.5% and 19.7%, severally. In the RID dataset, our network outperforms other networks, with a 10.7% improvement in F1-score compared to RCF, 18.3%, and 2.5% improvement in preciseness and recall, severally. The experimental results show that integration CBAM and residual structure within the U-Net network will improve its crack detection performance and increase detection accuracy.

D. COMPARISON WITH TRANSFORMER ALGORITHM

To further demonstrate the advantages of the method proposed in this study, we also compare the method with the recently published Vision Transformer (VIT) [43], Swin-UNet [44], and TransUNet [45] algorithms. Our method also has some advantages. The comparison results are shown in Table 4; for the DeepCrack dataset, our method's overall accuracy is 87.2%, and the precision and recall are 88.9% and 85.7%, respectively. For Crack Forest Dataset, the precision of our method is lower than TransUNet by 0.6%, but our overall accuracy is 0.2% higher than TransUNet. And for the RID dataset, our method also outperforms other algorithms with an overall precision of 55.4%. Compared with Transformer, our method integrates the channel and spatial location information of cracks in the feature extraction stage, and the attention weight is tilted toward cracks. Transformer focuses more on global information and ignores local information. The proportion of crack pixels in the image is smaller, so ignoring local information will lead to lower detection accuracy.

VI. CONCLUSION

We introduced Res-CBAM and BasicBlock into the U-Net to ascertain a neural network model for crack detection. The experimental results show that the introduction of CBAM enhances the attention of the neural network to the crack region, improves the extraction ability of the neural network for fine cracks, and suppresses the interference of background factors. Meanwhile, The shortcut connections of Res-CBAM and the replacement of the convolutional layer within the network structure by BasicBlock make sure the transmission of crucial information as with efficiency as potential and effectively suppress the matter of network degradation. The created neural network learns a lot of features about cracks and improves the ability of the model to discover fine cracks. Compared with many other neural network methods, the neural network built in this study encompasses a considerably increased ability to extract cracks. the excellent accuracy and robustness of the neural network were verified through extensive experiments on completely different data sets.

REFERENCES

[1] S. C. Radopoulou and I. K. Brilakis, ''Automated detection of multiple pavement defects,'' *J. Comput. Civil Eng.*, vol. 31, no. 2, Mar. 2017, Art. no. 04016057, doi: 10.1061/(ASCE)CP.1943-5487.0000623.

[2] J. Jeong, H. Jo, and G. Ditzler, ''Convolutional neural networks for pavement roughness assessment using calibration-free vehicle dynamics,'' *Comput.-Aided Civil Infrastruct. Eng.*, vol. 35, no. 11, pp. 1209–1229, Mar. 2020, doi: 10.1111/mice.12546.

[3] H. Y. Ju, W. Li, S. Tighe, Z. C. Xu, and J. Z. Zhai, ''CrackU-Net: A novel deep convolutional neural network for pixelwise pavement crack detection,'' *Struct. Control Health Monitor.*, vol. 27, no. 8, Mar. 2020, Art. no. e2551, doi: 10.1002/stc.2551.

[4] E. H. Miller, ''Crack detection and segmentation using deep learning with 3D reality mesh model for quantitative assessment and integrated visualization,'' *J. Comput. Civil Eng.*, vol. 34, no. 3, May 2020, Art. no. 04020010, doi: 10.1061/(ASCE)CP.1943-5487.0000890.

[5] L. Zhang, F. Yang, Y. Daniel Zhang, and Y. J. Zhu, ''Road crack detection using deep convolutional neural network,'' in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 3708–3712, doi: 10.1109/ICIP.2016.7533052.

[6] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, ''Automatic road crack detection using random structured forests,'' *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3434–3445, Dec. 2016, doi: 10.1109/TITS.2016.2552248.

[7] G. Yao, F.-J. Wei, J.-Y. Qian, and Z.-G. Wu, ''Crack detection of concrete surface based on
convolutional neural networks,'' in *Proc. Int. Conf. Mach. Learn. Cybern. (ICMLC)*, Jul. 2018, pp. 246–250, doi: 10.1109/ICMLC.2018.8527035.

[8] Z. Liu, Y. Cao, Y. Wang, and W. Wang, ''Computer vision-based concrete crack detection using U-net fully convolutional networks,'' *Autom. Construct.*, vol. 104, pp. 129–139, Aug. 2019, doi: 10.1016/j.autcon.2019.04.005.

[9] S. Dorafshan, R. J. Thomas, and M. Maguire, ''Comparison of deep convolutional neural networks and edge detectors for image-based crack detection in concrete,'' *Construct. Building Mater.*, vol. 186, pp. 1031–1045, Oct. 2018, doi: 10.1016/j.conbuildmat.2018.08.011.

[10] H. Li, J. Zong, J. Nie, Z. Wu, and H. Han, ''Pavement crack detection algorithm based on densely connected and deeply supervised network,'' *IEEE Access*, vol. 9, pp. 11835–11842, 2021, doi: 10.1109/ACCESS.2021.3050401.

[11] H. Lin, B. Li, X. Wang, Y. Shu, and S. Niu, ''Automated defect inspection of LED chip using deep convolutional neural network,'' *J. Intell. Manuf.*, vol. 30, no. 6, pp. 2525–2534, Aug. 2019, doi: 10.1007/s10845-018-1415-x.

[12] X. Wu, J. Ma, Y. Sun, C. Zhao, and A. Basu, ''Multi-scale deep pixel distribution learning for concrete crack detection,'' in *Proc. 25th Int. Conf. Pattern Recognit. (ICPR)*, Jan. 2021, pp. 397–400, doi: 10.1109/ICPR48806.2021.9413312.

[13] O. Ronneberger, P. Fischer, and T. Brox, ''U-Net: Convolutional networks for biomedical image segmentation,'' in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, vol. 9351. Cham, Switzerland: Springer, Nov. 2015, pp. 234–241.

[14] Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. M. Liang, ''UNet++: A nested U-Net architecture for medical image segmentation,'' in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045. Cham, Switzerland: Springer, Sep. 2018, pp. 3–11.

[15] J. Cheng, W. Xiong, W. Chen, Y. Gu, and Y. Li, ''Pixel-level crack detection using U-Net,'' in *Proc. TENCON IEEE Region 10 Conf.*, Oct. 2018, pp. 462–466.

[16] Z. Fan, C. Li, Y. Chen, J. H. Wei, G. Loprencipe, X. P. Chen, and P. D. Mascio, ''Automatic crack detection on road pavements using encoder–decoder architecture,'' *Materials*, vol. 13, no. 13, p. 2960, May 2020, doi: 10.3390/ma13132960.

[17] Z. W. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. M. Liang, ''The importance of skip connections in biomedical image segmentation,'' in *Deep Learning and Data Labeling for Medical Applications*, vol. 10008. Cham, Switzerland: Springer, Sep. 2016, pp. 179–187.

[18] G. Xu, C. Liao, and J. Chen, ''Extraction of apparent crack information of concrete based on HU-ResNet,'' *Comput. Eng.*, vol. 46, no. 11, pp. 279–285, 2020.

[19] L. F. Li, N. Wang, B. Wu, and X. Zhang, ''Segmentation algorithm of bridge crack image based on modified pspnet,'' *Adv. Lasers Optoelectron.*, vol. 58, no. 22, pp. 101–109, 2021, doi: 10.3788/LOP202158.2210001.

[20] S. Bang, S. Park, H. Kim, and H. Kim, ''Encoder–decoder network for pixel-level road crack detection in black-box images,'' *Comput. Aided Civil Infrastruct. Eng.*, vol. 34, no. 8, pp. 713–727, 2019, doi: 10.3788/LOP202158.2210001.

[21] S. H. Hanzaei, A. Afshar, and F. Barazandeh, ''Automatic detection and classification of the ceramic tiles' surface defects,'' *Pattern Recognit.*, vol. 66, pp. 174–189, Jun. 2017, doi: 10.1016/j.patcog.2016.11.021.

[22] K. Kirschke and S. Velinsky, ''Histogram-based approach for automated pavement-crack sensing,'' *J. Transp. Eng.*, vol. 118, no. 5, pp. 700–710, 1992, doi: 10.1061/(ASCE)0733-947X(1992)118:5(700).

[23] W. Huang and N. Zhang, ''A novel road crack detection and identification method using digital image processing techniques,'' in *Proc. 7th Int. Conf. Comput. Converg. Technol.*, Dec. 2012, pp. 397–400.

[24] L. Peng, W. Chao, L. Shuangmiao, and F. Baocai, ''Research on crack detection method of airport runway based on twice-threshold segmentation,'' in *Proc. 5th Int. Conf. Instrum. Meas., Comput., Commun. Control*, Sep. 2015, pp. 1716–1720, doi: 10.1109/IMCCC.2015.364.

[25] Z. Yuhan, Q. Juan, G. Zhiling, J. Kuncheng, and C. Shiyuan, ''Detection of road surface crack based on PYNQ,'' in *Proc. IEEE Int. Conf. Mechatronics Autom. (ICMA)*, Oct. 2020, pp. 1150–1154.

[26] H. Y. J. W. L. S. Tighe, P. Eng, and R. Deng, ''Illumination compensation model with k-means algorithm for detection of pavement surface cracks with shadow,'' *J. Comput. Civil Eng.*, vol. 34, no. 1, 2020, Art. no. 04019049, doi: 10.1061/ASCECP.1943-5487.0000869.

[27] B. C. Sun, Y. J. Qiu, and S. Q. Liang, ''Research on wavelet-based pavement crack identification,'' *J. Chongqing Jiaotong Univ. Natural Sci. Ed.*, vol. 29, no. 1, pp. 69–72, 2010.

[28] P. Subirats, J. Dumoulin, V. Legeay, and D. Barba, ''Automation of pavement surface crack detection using the continuous wavelet transform,'' in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2006, pp. 3037–3040.

[29] Z. G. Xu, X. M. Zhao, and H. S. Song, ''Crack identification algorithm for asphalt pavement based on histogram estimation and shape analysis,'' *J. Instrum.*, vol. 31, no. 10, pp. 2260–2266, Oct. 2010.

[30] A. Landstrom and M. J. Thurley, ''Morphology-based crack detection for steel slabs,'' *IEEE J. Sel. Topics Signal Process.*, vol. 6, no. 7, pp. 866–875, Nov. 2012, doi: 10.1109/JSTSP.2012.2212416.

[31] T. S. Nguyen, S. Begot, F. Duculty, and M. Avila, ''Free-form anisotropy: A new method for crack detection on pavement surface images,'' in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2011, pp. 1069–1072, doi: 10.1109/ICIP.2011.6115610.

[32] Y. Maode, B. Shaobo, X. Kun, and H. Yuyao, "Pavement crack detection and analysis for high-grade highway," in *Proc. 8th Int. Conf. Electron. Meas. Instrum.*, Aug. 2007, p. 548, doi: 10.1109/ICEMI.2007.4351202.

[33] S. Shah, "Automatic cell segmentation using a shape-classification model in immunohistochemically stained cytological images," *IEICE Trans. Inf. Syst.*, vol. E91-D, no. 7, pp. 1955–1962, Jul. 2008, doi: 10.1093/ietisy/e91-d.7.1955.

[34] H. Wang, N. Zhu, and Q. Wang, "Segmentation of pavement cracks using differential box-counting approach," *J. Harbin Inst. Technol.*, vol. 39, no. 1, pp. 142–144, 2007.

[35] S. Bhat, S. Naik, M. Gaonkar, P. Sawant, S. Aswale, and P. Shetgaonkar, "A survey on road crack detection techniques," in *Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE)*, Feb. 2020, pp. 1–6, doi: 10.1109/ic-ETITE47903.2020.67.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[37] Y. Liu, J. Yao, X. Lu, R. Xie, and L. Li, "DeepCrack: A deep hierarchical feature learning architecture for crack segmentation," *Neurocomputing*, vol. 338, pp. 139–153, Apr. 2019.

[38] Y. Shi, L. Cui, Z. Qi, F. Meng, and Z. Chen, "Automatic road crack detection using random structured forests," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 12, pp. 3434–3445, Dec. 2016, doi: 10.1109/TPAMI.2016.2644615.

[39] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder–decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017, doi: 10.1109/TPAMI.2016.2644615.

[40] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, and X. Bai, "Richer convolutional features for edge detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3000–3009, doi: 10.1109/cvpr.2017.622.

[41] Z. Qu and W. CHEN, "Concrete pavement crack detection based on dilated convolution and multi-features fusion," *Comput. Sci.*, vol. 49, no. 3, pp. 192–196, Dec. 2022, doi: 10.11896/jsjkx.210100164.

[42] Z. Qu and W. Chen, "CrackU-Net: Towards high quality pavement crack detection," *Comput. Sci.*, vol. 49, no. 1, pp. 204–211, Dec. 2022, doi: 10.11896/jsjkx210100128.

[43] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. H. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16 × 16 words: Transformers for image recognition at scale," Jun. 2021, *arXiv:2010.11929*.

[44] H. Cao, Y. Wang, J. Chen, D. Jiang, X. Zhang, Q. Tian, and M. Wang, "Swin-UNet: UNet-like pure transformer for medical image segmentation," 2021, *arXiv:2105.05537*.

[45] J. Chen, Y. Lu, Q. Yu, X. Luo, E. Adeli, Y. Wang, L. Lu, A. L. Yuille, and Y. Zhou, "TransUNet: Transformers make strong encoders for medical image segmentation," 2021, *arXiv:2102.04306*.

PENG JING was born in Datong, Shanxi, China, in 1994. He is currently pursuing the master's degree with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo. His current research interests include deep learning object detection and semantic segmentation.

HAIYANG YU was born in Linyi, Shandong, China, in 1978. He received the Ph.D. degree from the Chain University of Geosciences. He is currently a Professor with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo. He is the author or coauthor of more than 50 papers published in academic journals and conferences. His main research interests include remote sensing theory and application and LiDAR data processing and application.

ZHIHUA HUA was born in Zhoukou, Henan, China, in 1998. He is currently pursuing the master's degree with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo. His current research interests include remote sensing image processing and change detection.

SAIFEI XIE was born in Xuchang, Henan, China, in 2000. She is currently pursuing the master's degree with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo. Her current research interest includes deep learning-based point cloud filtering.

CAOYUAN SONG was born in Xuchang, Henan, China, in 1997. He is currently pursuing the master's degree with the School of Surveying and Land Information Engineering, Henan Polytechnic University, Jiaozuo. His current research interest includes deep learning-based point cloud filtering.