



ARTICLE

A Novel Detection Method for Pavement Crack with Encoder-Decoder Architecture

Yalong Yang^{1,2,3}, Wenjing Xu^{1,2,3}, Yinfeng Zhu⁴, Liangliang Su^{1,2,3,*} and Gongquan Zhang^{1,2,3}

¹Anhui Province Key Laboratory of Intelligent Building and Building Energy Saving, Anhui Jianzhu University, Hefei, 230022, China

²Anhui Institute of Strategic Study on Carbon Dioxide Emissions Peak and Carbon Neutrality in Urban-Rural Development, Hefei, 230022, China

³School of Electronic and Information Engineering, Anhui Jianzhu University, Hefei, 230601, China

⁴School of Mechanical and Electrical Engineering, Anhui Jianzhu University, Hefei, 230601, China

*Corresponding Author: Liangliang Su. Email: llsu_yz@ahjzu.edu.cn

Received: 09 October 2022 Accepted: 16 January 2023

ABSTRACT

As a current popular method, intelligent detection of cracks is of great significance to road safety, so deep learning has gradually attracted attention in the field of crack image detection. The nonlinear structure, low contrast and discontinuity of cracks bring great challenges to existing crack detection methods based on deep learning. Therefore, an end-to-end deep convolutional neural network (AttentionCrack) is proposed for automatic crack detection to overcome the inaccuracy of boundary location between crack and non-crack pixels. The AttentionCrack network is built on U-Net based encoder-decoder architecture, and an attention mechanism is incorporated into the multi-scale convolutional feature to enhance the recognition of crack region. Additionally, a dilated convolution module is introduced in the encoder-decoder architecture to reduce the loss of crack detail due to the pooling operation in the encoder network. Furthermore, since up-sampling will lead to the loss of crack boundary information in the decoder network, a depthwise separable residual module is proposed to capture the boundary information of pavement crack. The AttentionCrack net on public pavement crack image datasets named CrackSegNet and Crack500 is trained and tested, the results demonstrate that the AttentionCrack achieves *F1* score over 0.70 on the CrackSegNet and 0.71 on the Crack500 in average and outperforms the current state-of-the-art methods.

KEYWORDS

Crack detection; deep learning; convolution neural network; pixel segmentation

1 Introduction

Crack formation on pavements poses a safety hazard to road users. Pavement is mainly made of asphalt or concrete [1]. Its performance will be affected by the surrounding environment and traffic load. It will gradually be damaged to varying degrees. Among them, cracks are one of the more



common and harmful road defects. Therefore, it is necessary to regularly check for cracks to ensure the safety of the infrastructure.

Cracks are a special image feature that appears as small, irregular black lines hidden in road texture. The crack image also contains complex background noise, such as uneven illumination, stains, and the texture of the road itself [2]. Since the background of the pavement crack image contains complex texture patterns, how to effectively segment the crack and improve the accuracy of the crack segmentation is a complex problem. Therefore, conducting in-depth research on crack detection is necessary to make up for existing algorithms' shortcomings.

In recent years, deep learning was used to solve many computer vision problems involving image classification, object detection and image segmentation. Using deep learning for image segmentation, such as U-Net, convolutional features in the encoder-decoder network have been found to help improve semantic segmentation performance [3].

Embedding the attention mechanism in the encoder-decoder network can effectively supervise the semantic segmentation of crack images and focus on learning local crack semantic features. What's more, it can accurately identify pseudo crack features to optimize the image segmentation effect of the network and enhance the encoder-decoder robustness and generalization of the network [4,5]. The deep learning algorithm proposed in this paper integrates the attention mechanism and makes the network focus on the crack region, so the network is named AttentionCrack.

In conclusion, a deep learning algorithm for pavement crack segmentation is proposed, which demonstrate outperforms other methods. The AttentionCrack network is constructed on the U-Net based encoder-decoder architecture, and it largely solves the inaccurate boundary localization problem of deep learning in crack segmentation. In the AttentionCrack, to make the network focus on crack regions and enhance the crack region recognition, the attention mechanism is fused into the same-scale convolutional features of the encoder network and decoder network. Meanwhile, to reduce the loss of crack details because of pooling operations in the encoder network, a dilated convolution module is introduced in the encoder-decoder architecture. Furthermore, since upsampling in the decoder network leads to the loss of crack boundary information, a depthwise separable residual module is proposed to capture the boundary information of pavement cracks.

The contribution of this work is as follows:

- (1) The main contribution is to design a new neural network for crack detection. Compared with the existing pavement crack detection methods, the proposed method has higher detection accuracy and achieves better performance on the benchmark dataset.
- (2) A residual network structure is designed and integrated into the decoder network, which can reduce the loss of pavement crack boundary information.
- (3) The impact of different modules and their combinations in crack detection is explored, the modules include attention mechanism, dilated convolution and depthwise separable residual module.
- (4) Extensive experiments are conducted on two publicly datasets, i.e., CrackSegNet and Crack500, and the results demonstrate the effectiveness of the proposed method.

2 Related Work

Pavement crack detection has been studied for many years as a fundamental task for maintaining road safety. With the advancement of digital image processing technology, various methods and

models have been applied to crack detection. Crack detection methods are generally divided into methods based on traditional image processing techniques and methods based on deep convolutional neural networks.

2.1 Crack Detection Using Traditional Image Processing Techniques

Over the past few decades, crack detection algorithms based on digital image processing techniques have been extensively studied. The gray value of crack pixels in crack images is lower than that of the background and propose threshold segmentation algorithms [6,7]. These algorithms use thresholds to separate cracks from the background. However, the detection accuracy of these threshold segmentation algorithms is low when there is much noise in the crack image. To eliminate the influence of noise, crack detection methods such as wavelet transform and NSCT transform have been proposed [8,9]. However, these methods do not handle fractures with poor continuity well. For the identification of pavement crack, the segmentation algorithm has relatively good accuracy, but its processing speed is slow [10]. If image texture considering brightness and connectivity to identify fractures or dynamic threshold method is used, relatively rough fracture morphology can only be obtained [11]. In addition, the crack recognition based on Canny edge detection is also prone to false recognition. It can be seen that the common problems of traditional image recognition technology are low accuracy and high false alarm rate. The traditional image recognition technology is unable to identify the crack at pixel level. Therefore, it needs to manually extract features, but the preprocessing method directly affects the recognition effect [12–14].

2.2 Crack Detection Using Deep Convolutional Neural Networks

Methods based on deep learning have been widely used in image classification, object detection, image segmentation, and other fields. Their detection accuracy far exceeds that of traditional methods based on digital image processing technology and even exceeds the detection level of humans. A network called DeepCrack is proposed by Zou et al., which fuses the down-sampled and up-sampled feature maps in the SegNet network to generate single-scale fusion features. Then, the fused feature maps of all scales are formed into a multi-scale fusion map to obtain better crack detection performance [2]. Lau et al. proposed a network architecture, which is based U-Net and replace the encoder with a pretrained ResNet-34 neural network [15]. The authors perform crack detection on the CFD and Crack500 dataset, which requires fewer features than other machine learning techniques. Chen et al. proposed a switch module named SWM to improve the efficiency of crack detection [16]. It judges whether the predicted image is positive and determines whether the decoder module needs to be skipped. An improved deep fully convolutional neural network named CrackSegNet is proposed to conduct dense pixel-wise crack segmentation [17]. Dilated convolution, spatial pyramid pooling, and skip connections are fused in the backbone feature extraction network modified from the classic convolutional network VGG-16. Zhong et al. [18] proposed a concrete crack detection network based on atrous convolution and multi-feature fusion. It adopts an encoding-decoding structure based on U-Net. In the encoding stage, an improved residual network, Res2Net, is used to improve feature extraction. In the middle part of the network, dilated convolutions with different dilation rates is used to increase the receptive field of feature points. To improve the representation of crack features effectively, the spatial channel combined attention mechanism was integrated into the encoder-decoder network [19].

Although these crack detection methods have achieved a certain degree of success, the edges of crack detection are blurred still, and the detection accuracy in complex backgrounds is not high.

3 The Proposed Method

3.1 Network Architecture

The AttentionCrack mainly includes encoder network and decoder network, which is improved based on the encoder-decoder architecture. The encoder network consists of four identical encoding blocks, each of which contains two convolutional layers with 3×3 convolution kernels and one Maxpooling layer with 2×2 convolutional kernels. Each encoding block generates feature maps with different channel numbers and resolutions. The encoder network will deepen the number of channels of the feature map and reduce the resolution of the feature map. The decoder network consists of four identical decoding blocks, each of which contains two convolutional layers with 3×3 convolution kernels. The feature map in the decoder network is upsampled by bilinear new interpolation, so that the resolution is restored to the original resolution. Each decoder network has a corresponding layer in the encoder network, so the encoder network and the decoder network are almost symmetrical.

In the AttentionCrack network, dilated convolutions is fused into the connected part of the encoder-decoder architecture to increase the receptive field of the model. Furthermore, the deep separable residual module is fused to the decoder network, which can capture the boundary information of pavement cracks. Additionally, the attention mechanism is combined with the encoder-decoder architecture to improve the representation of crack information. The overall model structure of the AttentionCrack is shown in Fig. 1.

Each convolution of the encoder network will deepen the number of channels of the feature map. Feature map with fewer channels contain more detail and location information, and feature map with more channels contain more semantic information. The decoder network upsamples the dense feature map progressively. The Maxpooling operation in the encoder network can reduce the size of the feature map, which leads to a loss of spatial resolution and a deviation in the location of the crack boundary. As a result, the dense feature map loses more location information for cracks. To solve this problem, many researchers believe that the resolution of feature map can be directly kept unchanged. However, this approach increases the number of parameter and takes more time during network training. Dilated convolution can increase the receptive field of the convolution kernel without the loss of information because of the reduction of feature map resolution. As illustrated in Fig. 2, the Dilated Convolution Module (DCM) is integrated in the middle of the encoder-decoder architecture. Dilated convolutions with dilation rates of 1, 2, and 4 is used in the AttentionCrack, which means that the receptive fields of the convolution kernels are $3 * 3 = 9$, $5 * 5 = 25$, and $9 * 9 = 81$, respectively. Our network makes full use of the information of different receptive fields to capture the location information of pavement cracks.

Bilinear interpolation is used in the encoder-decoder architecture to expand the size of the feature map, which makes the region boundary position of the sparse feature map blurred and results in the identification of pavement cracks discontinuity. To improve this problem, the Deep Separable Residual Module (DSRM) is applied to the decoder network, and the module structure is shown in the Fig. 3. The kernel size of 1×1 convolution is used in the DSRM to reduce the dimension of the feature map, and the kernel size of 5×5 convolution is used to extract feature. The DSRM reduces model size and computational complexity, which uses convolutions with larger receptive fields to capture boundary information of pavement cracks.

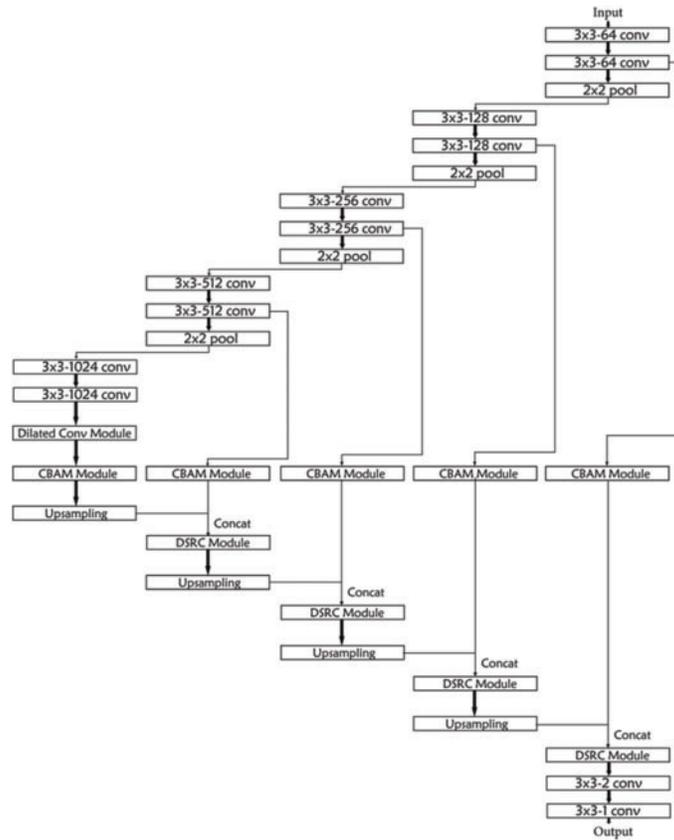


Figure 1: The AttentionCrack structure

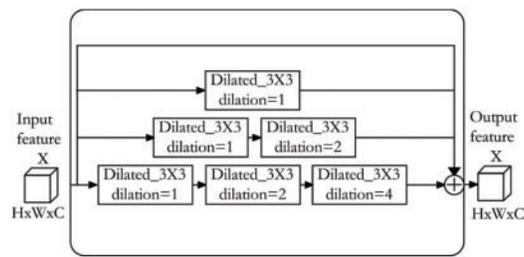


Figure 2: Dilated convolution module

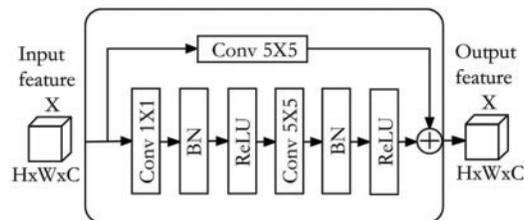


Figure 3: Depth separable residual module

Since multi-scale feature fusion is effective for improving network performance, coarse feature map is addressed by adding skip connections in the encoder-decoder architecture. However, skip connections cannot sufficiently integrate important global information, which results in discontinuities in crack prediction. Moreover, the encoder-decoder architecture lacks the ability to distinguish cracks and backgrounds on low-level feature maps, which is not conducive to guiding the learning of the network. The attention mechanism can refine crack features and effectively guide network training. Furthermore, the convolution operation extracts features by fusing the channel information between different feature map and the spatial information of the same feature map. Both channel information and spatial information are crucial for refining crack information. Therefore, the Convolutional Block Attention Module (CBAM) is introduced into the skip connections of the encoder-decoder architecture, which can improve the representation ability of the network [20]. The schematic diagram of the CBAM module is shown in Fig. 4.

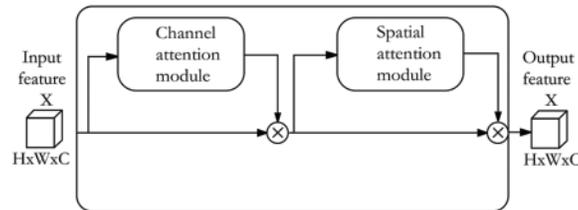


Figure 4: CBAM module

The CBAM is an attention module that combines channel attention module and spatial attention module. The purpose of the channel attention module is to selectively emphasize the relationship between channels, which can make the network pay more attention to the meaningful crack information. The channel attention module is shown in Fig. 5. The spatial attention module is designed to selectively emphasize the features of each spatial location, which can enable the network to extract more information about the spatial location of cracks. The spatial attention module is shown in Fig. 6.

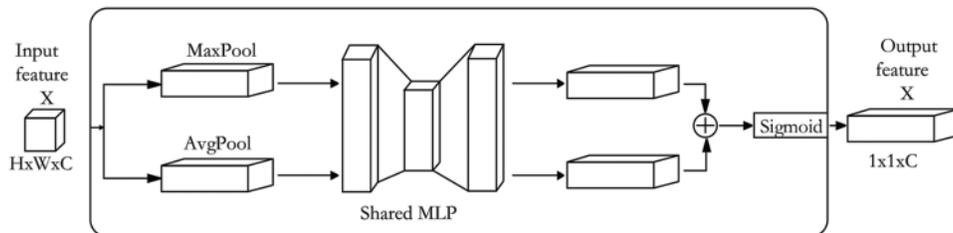


Figure 5: Channel attention module

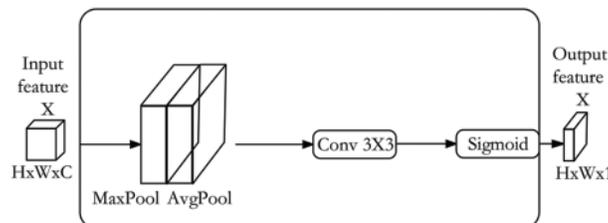


Figure 6: Spatial attention module

3.2 Loss Function

Since there are only two classes in crack detection, it can be viewed as a binary classification problem. Generally, the ground truth crack pixels stand as a minority class in the crack image, which makes it an imbalance classification or segmentation. Some works deal with this problem by adding larger weights to the minority class. However, in crack detection, larger weights adding to the cracks will result in more false positives. Thus, a binary cross-entropy loss (BCELoss) function is used for cost function in the training process of neural network. BCELoss is defines as

$$L_{BCE-LOSS} = -\frac{1}{N} \sum_i [L_i * \log(y_i) + (1 - L_i) * \log(1 - y_i)] \quad (1)$$

where N is the total number of pixels in a crack image, and L_i and y_i are the label value and predicted probability value of the i pixel, respectively.

4 Experiments and Results

4.1 Experimental Setup

During training, BN is used after each convolution layer of the encoder-decoder network to speed up convergence. Random weights in convolutional layers are initialized by built-in He normal initializers, and the bias is initialized to 0. In the same time, the bilinear interpolation method is used for up-sampling. The global learning rate is initialized to 1e-5, and it is divided by 10 after every 10 epochs. The stochastic gradient descent method is employed to update the network parameters with mini-batch size of 4 in each iteration. The training is completed until the evaluation in the test is optimal, and the corresponding number of epochs is 40. All experiments are performed by using on NVIDIA GeForce RTX 3090 GPU. We implement our network by using the publicly available PyTorch which is well-known in this community.

4.2 Dataset

- (1) CrackSegNet: The CrackSegNet dataset [17] contains 919 crack images, of which 735 crack images are used as training, and 184 images are used as test.
- (2) Crack500: The Crack500 dataset [21] contains 3368 crack images, which are cropped from 500 crack images captured by mobile phones, of which 2244 crack images are used for training, and 1124 images are used for testing. Crack500 is a large publicly accessible pavement crack dataset with corresponding labeled images. These images contain a variety of complex pavement backgrounds and various types of cracks.
- (3) Data augmentation has been performed to enlarge the size of the training set. In each training iteration, the crack image is rotated at eight different angles, 45° each time, and the crack image is also rotated horizontally and vertically. Therefore, the dataset is 16 times larger.

During training and testing, it is necessary to unify the image size as input of network model. Similar to [17] and [15], we resize the images of the CrackSegNet dataset to 512 × 512, and ones of the Crack500 dataset to 320 × 320, respectively.

4.3 Evaluation Indicators

All methods are tested on the above datasets. At test time, the detection result image is compared with the label image to calculate the *Precision*, *Recall* and *F1* score. These evaluation criteria can assess the accuracy of the semantic segmentation tasks, which are defined as shown in Eqs. (2)–(4):

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

Among them, TP is the number of pixels that are correctly detected and judged as cracks in the detection result. FP is the number of falsely detected background pixels as crack pixels, and FN is the number of falsely detected crack pixels as background pixels. Due to the limitations of $Precision$ and $Recall$, we use $F1$ score as an overall indicator to evaluate the detection effect.

4.4 Contrast Method

The proposed method is compared with existing crack detection methods, all of which are based on deep learning. These crack detection methods are trained and tested on the CrackSegNet dataset and Crack500 dataset, respectively.

- (1) U-Net [3]: U-Net is based on the encoder-decoder structure and realizes feature fusion through splicing. Its structure is simple but very effective.
- (2) SegNet [22]: SegNet is a fully convolutional network for pixel-level image segmentation with an encoder-decoder symmetric structure.
- (3) DeepCrack [2]: DeepCrack is a convolutional network for pixel-level crack segmentation, which is based on the SegNet and fusion jump connection.
- (4) CrackSegNet [17]: CrackSegNet is a convolutional network for dense pixel crack segmentation. The network consists of a backbone, dilated convolution, spatial pyramid pooling and skip connection modules.
- (5) CarNet [23]: CarNet is based on the encoder-decoder architecture. It is an efficient and high-quality crack detection method. In the decoder, a lightweight up-sampling feature pyramid module is introduced to learn rich hierarchical features for crack detection.

4.5 Experimental Results and Analysis

The U-Net, SegNet, DeepCrack, CrackSegNet, CarNet, and the AttentionCrack are trained and tested on the CrackSegNet and Crack500, respectively. The results on the CrackSegNet dataset is shown in Table 1, and the P-R (Precision-Recall) curves are shown in Fig. 7. It can be seen from the experimental results that the $F1$ score of the method is the highest, and the $F1$ scores of U-Net, SegNet, DeepCrack, CrackSegNet, and CarNet are lower than those of the method: 2.91%, 1.48%, 6.1%, 7.52%, 2.01%. In the P-R curves, our method is the best one of all methods.

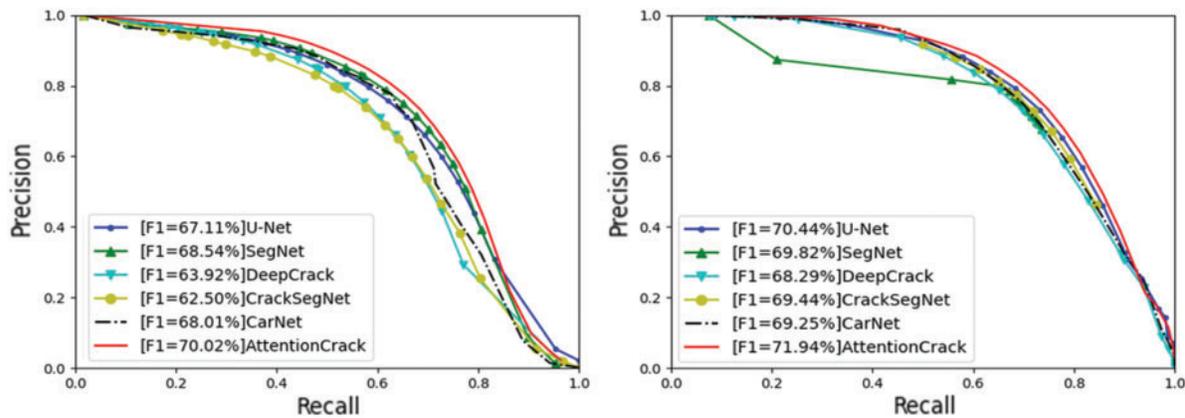
Table 1: Comparison of various methods on CrackSegNet dataset

Methods	Precision	Recall	F1
U-Net [3]	65.77%	71.31%	67.11%
SegNet [22]	65.02%	75.09%	68.54%
DeepCrack [2]	60.52%	70.93%	63.92%

(Continued)

Table 1 (continued)

Methods	Precision	Recall	F1
CrackSegNet [17]	64.11%	65.02%	62.50%
CarNet [23]	65.26%	73.41%	68.01%
AttentionCrack	65.81%	77.17%	70.02%

**Figure 7:** P-R curves on CrackSegNet dataset (left) and Crack500 dataset (right)

The quantitative results on the Crack500 datasets are shown in Table 2, *Recall* of our method is slightly lower than that of U-Net method, which is due to the influence of the imbalance of positive and negative samples in the dataset on model training. Although *Recall* is slightly lower than that of U-Net method, a single *Recall* cannot judge the model. The *F1* score and P-R curve are better indicators to measure model performance. It can be seen from Table 2 that *F1* score of AttentionCrack is the highest. Compared with U-Net, SegNet, DeepCrack, CrackSegNet, and CarNet, there are 1.5%, 2.12%, 3.65%, 2.5% and 2.69% performance improvement on *F1*, respectively. The P-R curves are shown in Fig. 7. The AttentionCrack achieves the best performance on Crack500.

Table 2: Comparison of various methods on Crack500 dataset

Methods	Precision	Recall	F1
U-Net [3]	68.39%	79.24%	70.44%
SegNet [22]	68.12%	77.15%	69.82%
DeepCrack [2]	69.74%	73.17%	68.29%
CrackSegNet [17]	68.51%	77.35%	69.44%
CarNet [23]	68.36%	75.82%	69.25%
AttentionCrack	71.76%	77.65%	71.94%

The experimental results show that the AttentionCrack can extract cracks from complex scenes and effectively remove the influence of shadows, stains and other interference. The AttentionCrack includes the advantages of U-Net, such as encoder-decoder path, skip connection and the attention

mechanism. Extended convolution expands the receptive field of the convolution kernel to include multi-scale context. In particular, the proposed extended convolution with different expansion rates can enhance the feature extraction ability without reducing the resolution of the feature map.

The partial segmentation results of various methods on the Crack500 dataset is shown in Fig. 8. Compared with other methods, the AttentionCrack is observed to be able to suppress more background artifacts than other methods. And the AttentionCrack has better performance in dealing with the problems of discontinuity of cracks, wrong segmentation of complex background, and accurate positioning of crack edges.

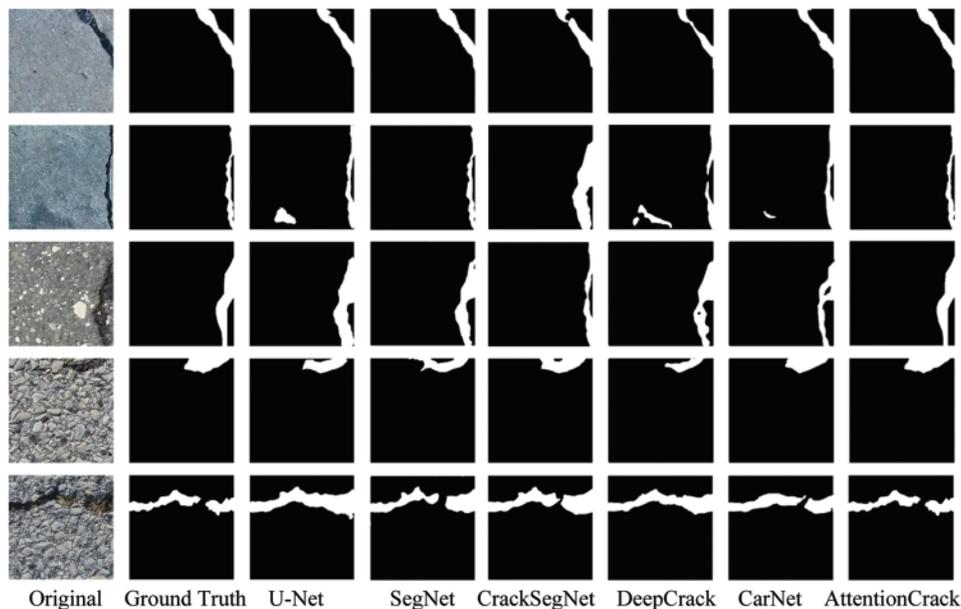


Figure 8: Images of detection results of different methods on the Crack500 dataset

4.6 Ablation Study

In Section 4.5, the experimental results show that this approach is superior to other advanced methods in crack detection. The effect of each module is studied in this part. Specifically, we investigated the importance of CBAM, DCM, and DSRM. Therefore, one or two of the three modules are removed each time before training the modified model with the same parameters. The reserved module is used as the method name. For example, Ours (CBAM) keeps the CBAM and remove the DCM and DSRM. And Ours (CBAM + DCM) keeps the CBAM and DCM and remove the DSRM. Finally, these models are tested on the CrackSegNet and Crack500 dataset, respectively.

The P-R curves on the CrackSegNet dataset and Crack500 dataset are shown in Fig. 9. The experimental results show that removing any module will lead to performance degradation.

As seen in Tables 3 and 4, the accuracy of Ours (CBAM) is slightly higher than AttentionCrack, because Ours (CBAM) misjudges the background of part of crack boundary as a crack. Although Ours (CBAM) accuracy improves, its *Recall* rate drops. Ours (CBAM) does not accurately locate the crack boundary information. The fusing of DCM and DSRM into the model increases the *F1* score of AttentionCrack, which indicates that AttentionCrack is the best compared with other methods.

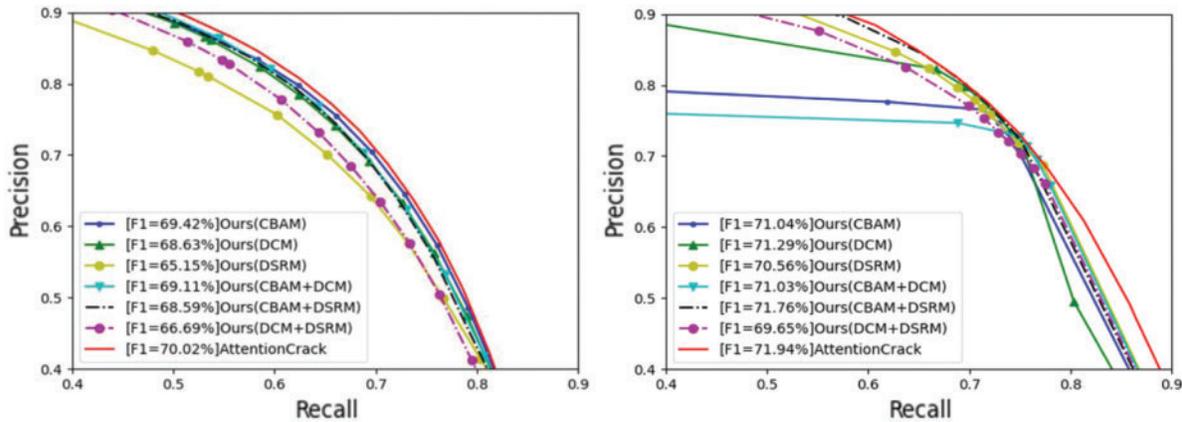


Figure 9: Comparison of the AttentionCrack with its modified versions by removing module on CrackSegNet dataset (left) and Crack500 dataset (right)

Table 3: Comparison between architectures for ablation study on CrackSegNet dataset

Methods	Precision	Recall	F1
Ours (CBAM)	66.11%	75.51%	69.42%
Ours (DCM)	65.97%	74.12%	68.63%
Ours (DSRM)	65.21%	70.09%	65.15%
Ours (CBAM + DCM)	64.31%	76.85%	69.11%
Ours (CBAM + DSRM)	65.81%	74.70%	68.59%
Ours (DCM + DSRM)	64.38%	73.14%	66.69%
AttentionCrack	65.81%	77.17%	70.02%

Table 4: Comparison between architectures for ablation study on Crack500 dataset

Methods	Precision	Recall	F1
Ours (CBAM)	72.61%	75.08%	71.04%
Ours (DCM)	71.53%	77.38%	71.29%
Ours (DSRM)	72.02%	75.82%	70.56%
Ours (CBAM + DCM)	75.47%	71.92%	71.03%
Ours (CBAM + DSRM)	71.36%	77.14%	71.76%
Ours (DCM + DSRM)	74.70%	70.96%	69.65%
AttentionCrack	71.76%	77.65%	71.94%

5 Conclusion

In this work, a novel end-to-end trainable convolutional network named AttentionCrack is proposed for crack detection. In AttentionCrack, the attention mechanism is integrated into multi-scale convolution features to enhance the recognition of crack areas. At the same time, the dilated convolution and the depth separable residual module are used to reduce the loss of crack details.

Compared with other state-of-the-art deep learning methods for crack detection, this method has achieved significant improvement in accuracy. The resulting method has good generalization and low data requirements, which can be enhanced by the inclusion of more powerful architectures or additional training of labeled images. Next, further investigation will be performed to improve the accuracy of crack detection, reduce the model's complexity and improve the model's generalization ability. In addition, exploration of other detection fields by using this method will be performed too.

Funding Statement: This work was supported by the National Natural Science Foundation of China under Grant No. 62001004, the Key Provincial Natural Science Research Projects of Colleges and Universities in Anhui Province under Grant No. KJ2019A0768, the Key Research and Development Program of Anhui Province under Grant No. 202104A07020017, the Research Project Reserve of Anhui Jianzhu University under Grant No. 2020XMK04, the Natural Science Foundation of the Anhui Higher Education Institutions of China, No. KJ2019A0789.

Author Contributions: Conceptualization, Yalong Yang and Liangliang Su; methodology, Wenjing Xu, Liangliang Su and Gongquan Zhang; formal analysis, Yalong Yang, Wenjing Xu and Liangliang Su; data curation, Wenjing Xu; writing—original draft preparation, Yalong Yang and Wenjing Xu; supervision, Yinfeng Zhu and Liangliang Su. All authors have read and agreed to the published version of the manuscript.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Adlinge, S., Gupta, A. (2013). Pavement Deterioration and its Causes. *Transport and Infrastructure*, 4(4), pp. 9–15.
2. Zou, Q., Zhang, Z., Li, Q., Qi, X. (2019). Deepcrack: Learning hierarchical convolutional features for crack detection. *IEEE Transactions on Image Processing a Publication of the IEEE Signal Processing Society*, 28(3), 1498–1512.
3. Jenkins, M. D., Carr, T. A., Iglesias, M. I., Buggy, T., Morison, G. (2018). A deep convolutional neural network for semantic pixel-wise segmentation of road and pavement surface cracks. *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2120–2124. Rome, Italy.
4. Huang, G., Zhu, J., Li, J., Wang, Z., Zhou, J. (2020). Channel-attention U-Net: Channel attention mechanism for semantic segmentation of esophagus and esophageal cancer. *IEEE Access*, 8, 122798–122810. <https://doi.org/10.1109/ACCESS.2020.3007719>
5. Jin, Y., Xu, W., Hu, Z., Jia, H., Shao, D. (2020). GSCA-UNet: Towards automatic shadow detection in urban aerial imagery with global-spatial-context attention module. *Remote Sensing*, 12(17), 2864. <https://doi.org/10.3390/rs12172864>
6. Hanzaei, S. H., Afshar, A. (2016). Automatic detection and classification of the ceramic tiles' surface defects. *Pattern Recognition*, 66(10), 174–189. <https://doi.org/10.1016/j.patcog.2016.11.021>
7. Oliveira, H., Correia, P. L. (2009). Automatic road crack segmentation using entropy and image dynamic thresholding. *2009 17th European Signal Processing Conference*, pp. 622–626. Glasgow, UK.
8. Nejad, F. M., Zakeri, H. (2011). An optimum feature extraction method based on wavelet-radon transform and dynamic neural network for pavement distress classification. *Expert Systems with Applications*, 38(8), 9442–9460. <https://doi.org/10.1016/j.eswa.2011.01.089>
9. Ma, C., Zhao, C., Hu, Y., Wang, H., Chen, H. (2009). Pavement cracks detection based on NSCT and morphology. *Journal of Computer-Aided Design and Computer Graphics*, 21(12), 1761–1767.

10. Tsai Y. C., Kaul V., Mersereau R. M. (2010). Critical assessment of pavement distress segmentation methods. *Journal of Transportation Engineering*, 136(1), 11–19. [https://doi.org/10.1061/\(ASCE\)TE.1943-5436.0000051](https://doi.org/10.1061/(ASCE)TE.1943-5436.0000051)
11. Nguyen, T. S., Bégot, S., Duculty, F., Avila, M. (2011). Free-form anisotropy: A new method for crack detection on pavement surface images. *IEEE International Conference on Image Processing*, pp. 1069–1072. Piscataway, NJ.
12. Amhaz, R., Chambon, S., Idier, J., Baltazart, V. (2016). Automatic crack detection on two-dimensional pavement images. *IEEE Transactions on Intelligent Transportation Systems*, 17(10), 2718–2729. <https://doi.org/10.1109/TITS.2015.2477675>
13. Wu, Z., Karimi, H. R., Dang, C. (2019). An approximation algorithm for graph partitioning via deterministic annealing neural network. *Neural Networks*, 117(5), 191–200. <https://doi.org/10.1016/j.neunet.2019.05.010>
14. Yeum, C. M., Dyke, S. J. (2015). Vision-based automated crack detection for bridge inspection. *Computer-Aided Civil and Infrastructure Engineering*, 30(10), 759–770. <https://doi.org/10.1111/mice.12141>
15. Lau, S., Chong, E., Yang, X., Wang, X. (2020). Automated pavement crack segmentation using U-Net-based convolutional neural network. *IEEE Access*, 8, 114892–114899. <https://doi.org/10.1109/ACCESS.2020.3003638>
16. Chen, H., Lin, H., Yao, M. (2019). Improving the efficiency of encoder-decoder architecture for pixel-level crack detection. *IEEE Access*, 7, 186657–186670. <https://doi.org/10.1109/ACCESS.2019.2961375>
17. Ren, Y. P., Huang, J. S., Hong, Z. Y., Lu, W., Yin, J. et al. (2020). Image-based concrete crack detection in tunnels using deep fully convolutional networks. *Construction and Building Materials*, 234, 117367–117379. <https://doi.org/10.1016/j.conbuildmat.2019.117367>
18. Zhong, Q., Wen, C. (2022). Crack detection of concrete pavement based on cavity convolution and multi feature fusion. *Computer Science*, 49(3), 192–196.
19. Sun, M., Zhao, H., Li, J. (2022). Road crack detection network under noise based on feature pyramid structure with feature enhancement (road crack detection under noise). *IET Image Processing*, 16(3), 809–822. <https://doi.org/10.1049/ipr2.12388>
20. Woo, S., Park, J., Lee, J. Y., Kweon, I. S. (2018). CBAM: Convolutional block attention module. *Proceedings of the European Conference on Computer Vision (ECCV)*, vol. 1651, pp. 3–19. <https://doi.org/10.1007/978-3-030-01234-2>
21. Yang, F., Zhang, L., Yu, S., Prokhorov, D., Mei, X. et al. (2019). Feature pyramid and hierarchical boosting network for pavement crack detection. *IEEE Transactions on Intelligent Transportation Systems*, 21(4), 1525–1535.
22. Badrinarayanan, V., Kendall, A., Cipolla, R. (2017). SegNet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 39(12), 2481–2495. <https://doi.org/10.1109/TPAMI.2016.2644615>
23. Li, K., Tian, Y., Qi, Z. (2021). CarNet: A lightweight and efficient encoder-decoder architecture for high-quality road crack detection. arXiv:2109.05707.